Convergent Earnings Mobility in the U.S.: Is this Finding Robust?[1]

Lisa Marie Dragoset
Cornell University

October 2007

Abstract

Income mobility is defined as the change in income from one period to another for the same individual; earnings mobility concerns the change in earnings from the labor market. It is widely recognized that incomes and earnings are measured with error, the existence of which casts doubt on a major finding in the mobility literature - the finding of convergent mobility. Convergent mobility happens when low earners experience more positive earnings changes (i.e., they gain more in dollars or percentages) than high earners. This type of convergence is called unconditional convergence because it does not condition on other personal characteristics. Conditional convergent mobility occurs when, holding other things equal, low earners gain more in dollars or percentages than high earners. Measurement error in initial earnings causes a spurious link between earnings change and initial earnings level, possibly producing the appearance of convergent mobility when no convergence has truly taken place.

While measurement error is recognized as potentially important in the mobility literature, little is known about the degree to which the finding of convergent mobility is affected by measurement error. This study uses a new, confidential dataset from the U.S. Census Bureau that contains individually reported total annual labor earnings from the Survey of Income and Program Participation (SIPP) linked to employer-reported total annual labor earnings from the Social Security Administration's Detailed Earnings Record (DER) to explore the robustness of the convergent mobility finding for the United States from 1990 to1999. I ask: is the finding of convergence robust to non-linear functional forms (step functions and non-parametric functions) and alternate samples of the data? The parametric results have been completed, and they show that the findings of unconditional and conditional convergence are robust to alternate samples and to a non-linear functional

[1]This document reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This document is released to inform interested parties of ongoing research and to encourage discussion of work in progress. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01~AG018854-02, and the Alfred P. Sloan Foundation. The views expressed on statistical and methodological issues are those of the author(s) and not necessarily those of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers' use of these data through the Research Data Centers (see www.ces.census.gov). For other questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, LEHD Program, Demographic Surveys Division, FOB 3, Room 2138, 4700 Silver Hill Rd., Suitland, MD 20233, USA. (Jeremy.S.Wu@census.gov http://lehd.dsd.census.gov). I thank Gary Fields, George Jakubson, and John Abowd for their helpful comments, as well as the members of the Cornell University Research Seminar on Labor Markets in Comparative Perspective.

form of the equation, namely, a step function by initial earnings quintile. I then ask: does the finding of conditional convergence hold when neither SIPP nor DER earnings are treated as equal to "true" latent earnings? I use two different methods to obtain an estimate of conditional mobility that treats neither data source as being free of measurement error: a) instrumental variables (IV) and generalized method of moments (GMM), and b) maximum likelihood estimation (MLE). I find that conditional convergence does hold when neither earnings source is treated as equal to "true" latent earnings.

1 Introduction

Income mobility is defined as the change in income from one period to another for the same individual; compensation mobility concerns the change in incomes from the labor market (labor earnings plus benefits). Earnings mobility concerns only the change in labor earnings, excluding all benefits such as employer contributions to 401(k) plans and health insurance plans. The empirical literature on income and earnings mobility in various countries around the world is voluminous; see Atkinson, Bourguignon, and Morrisson (1992) and Baulch and Hoddinott (2000) for surveys.

It is widely recognized that incomes and earnings are measured with error, the existence of which casts doubt on some of the main conclusions in the mobility literature (Duncan and Hill, 1985; Deaton, 1997; Bound, Brown, and Mathiowetz, 2001; Fields et al., 2003). One problematic area has been that of determining the correlates of individual mobility within the earnings distribution ("micro mobility"), or which earnings groups experience the most positive or negative earnings changes. Measurement error in initial earnings produces a spurious link between earnings change and initial earnings level, possibly producing the appearance of convergent mobility, i.e. low earners gaining more in dollars or percentages than high earners. This type of convergence is called unconditional convergence because it does not condition on other personal characteristics. Conditional convergent mobility occurs when, holding other things equal, low earners gain more in dollars or percentages than high earners. Divergent mobility occurs when the opposite is true - high earners gain more in dollars or percentages than low earners.

Researchers have responded to the concern about measurement error in several ways. One is to note the concern and proceed to use survey-reported earnings despite it. This is by far the most common approach to the measurement error issue. A second response is to use administrative records rather than survey reports. This approach has dominated research on income mobility in France, in which a whole series of studies have been conducted using administrative data; see, for example, Bigard, Guillotin, and Lucifora (1998), Buchinsky, Fougère, and Kramarz (1998), and Buchinsky, Fields, Fougère, and Kramarz (2003), among others. A third response, found in the U.S. literature, is to measure the differences between results obtained using survey data compared with the results using administrative records. Such studies are called "validation studies" and are surveyed in Bound, Brown, and Mathiowetz (2001).

Following the tradition of these validation studies, in this paper, I explore the findings of unconditional and conditional earnings mobility for the U.S. during the 1990s using a dataset that contains both survey-based and administrative-based earnings. However, unlike most

validation studies, I do not treat either earnings source as being free of measurement error. Rather, I use the two earnings sources simultaneously to obtain mobility estimates that are adjusted for possible measurement error bias. The main research questions are: 1)is the finding of convergence (both conditionally and unconditionally) robust to non-linear functional forms and alternate samples of the data? and 2)does the finding of conditional convergence hold when neither the survey-based (SIPP) earnings nor the administrative-based (DER) earnings are treated as equal to "true" latent earnings? I use two different methods to obtain an estimate of conditional earnings mobility that treats neither data source as being free of measurement error: a) instrumental variables (IV) and generalized method of moments (GMM), and b) maximum likelihood estimation (MLE).

The results can be summarized as follows. In answer to the first research question, the findings of unconditional and conditional convergence are robust to alternate samples and to a non-linear functional form of the equation, namely, a step function by initial earnings quintile. In answer to the second research question, the finding of conditional convergence is robust to estimation methods which treat neither survey-based nor administrative-based earnings as being free of measurement error. However, the amount of convergence found using these methods is less than the amount of convergence found using either data source alone. In other words, measurement error in reported earnings causes us to find more conditional convergence than truly took place.

The balance of the paper is organized as follows. I motivate the paper in Section 2, review the previous literature in Section 3, describe the data in Section 4, discuss the empirical methodology in Section 4, present the results in Section 5, and conclude in Section 6.

2 Motivation

According to the U.S. Census Bureau, earnings inequality was either constant or rising in the U.S. during the 1990s. In particular, the Census Bureau reports constant earnings inequality in the U.S. for the early part of the 1990s and again in the later 1990s (U.S. Census Bureau, 2005).[2] From 1992 to 1993, earnings inequality jumped by three Gini points (the very same time when new methods were used to collect earnings data) (U.S. Census Bureau, 2004). Though it is impossible to tell whether using the old methods would have produced constant or rising earnings inequality, there is no evidence whatsoever

[2] These earnings inequality estimates were produced by the U.S. Census Bureau using cross-sectional data from the Current Population Survey, Annual Social and Economic Supplement (formerly known as the March Supplement), rather than from the SIPP panels.

suggesting that earnings inequality fell in the United States over the period 1990-1999; the Census Bureau evidence suggests that inequality either rose or remained constant. The 1990s was also a period of growth for the U.S.: real GDP per capita rose from $28,000 to $34,000 (Johnston and Williamson, 2006). The combination of growth with constant or rising inequality might lead one to expect that persons who start out with the highest earnings experienced the most positive earnings changes in dollars over time. However, this is not what Dragoset and Fields (2006) find. Using both survey-based and administrative-based earnings data, we find that the individuals in the lowest initial earnings quintile gained more in dollars from one year to the next over the period from 1990 to 1999.

These two results - constant or rising inequality and convergent mobility in dollars - may seem to contradict each other, but they can be reconciled (Dragoset and Fields 2006; Fields and Sanchez Puerta 2007). Inequality estimates such as the Gini coefficient treat the data as a series of cross sections, while mobility estimates employ the panel aspect of the data to look at mean earnings changes for named individuals whom we follow over time. Two things were happening at the same time for the U.S. during the 1990s. One is that the dollar differences between different percentiles of the earnings distribution were widening. The other is that the places in the different parts of the earnings distribution were being occupied by different individuals.

Many economists consider steadily rising inequality to be worrisome for the very reason mentioned above - that it may indicate a steady decrease in relative earnings for the poorest earners, while the highest earners continue to gain more in dollars over time. However, if mobility is very high, that is, if individuals move freely throughout the earnings distribution, then high inequality becomes less worrisome because we know that the poorest earners are not necessarily "stuck" at the bottom of the earnings distribution. Therefore, the finding of convergent mobility for the U.S. during the 1990s using survey-based data was taken as good news by many economists. However, this finding could be due to measurement error.

To see how measurement error affects earnings mobility estimates, let us look at a very basic earnings mobility model. Unconditional mobility is often estimated using one of the following equations:

$$\begin{aligned} (1)\ y_{it} - y_{it-1} &= \alpha + \beta y_{it-1} + \varepsilon_{it} \text{ or} \\ (2)\ y_{it} &= \alpha + \gamma y_{it-1} + \varepsilon_{it}, \end{aligned}$$

where y_{it} represents earnings for individual i at time t, and $\gamma = \beta + 1$. Both β and γ measure the persistence of earnings over time, or how closely earnings in the current time period depend upon earnings in the previous time period. A value of $-1 < \beta < 0$ or

$0 < \gamma < 1$ implies convergence and a value of $\beta > 0$ or $\gamma > 1$ implies divergence. It is well known that if measurement error in the right hand side variable follows the classic textbook model, the estimates of β and γ will be biased downwards. The same result holds true when measurement error takes a "mean-reverting" form, defined as a negative correlation between the measurement error and the value of earnings as given by the administrative earnings (Dragoset and Fields 2006). In other words, measurement error in initial earnings (y_{it-1}) can cause the appearance of convergent mobility when, in fact, no convergence has truly occurred.

Dragoset and Fields (2006) attempt to gauge the effect of measurement error on earnings mobility estimates by comparing estimates obtained from survey-based and administrative-based data. We find convergence (both conditionally and unconditionally) even when using administrative-based earnings, which are believed to be closer to underlying "true" earnings than the survey-based earnings. Recall that convergent mobility means that low earners experience more positive earnings changes than high earners. Strong convergence means that those individuals with the lowest initial earnings were gaining more in dollars than those with the highest initial earnings. Weak convergence means that those individuals with the lowest initial earnings were gaining more in percentages than those with the highest initial earnings. We used dollar earnings rather than log earnings because a finding of strong convergence implies a finding of weak convergence.

The finding of convergence using administrative-based earnings was another piece of evidence taken as good news by many economists, but the work by Dragoset and Fields (2006) suffers from a couple of restrictions. First, we used only linear functional forms to estimate earnings mobility. One might worry that the finding of convergence is dependent upon the functional form, because of the fact that equation (1) assumes a single rate of convergence (or divergence) for all individuals. This paper investigates whether the finding of convergence is robust to non-linear functional forms which do not assume a single rate of convergence for all individuals. Second, we did not utilize the SIPP and DER earnings simultaneously to obtain mobility estimates that were unbiased by measurement error, which this paper does. We believe that when equations (1) or (2) are estimated using either SIPP or DER earnings separately, measurement error in initial earnings causes the coefficient estimates to be biased because the observed initial earnings are correlated with the error term in the model. This paper uses SIPP and DER earnings to instrument for each other in an attempt to correct this bias. I instrument using two slightly different models - one specified in earnings levels (which assumes a classical measurement error model) and one specified in earnings changes (which assumes a "mean-reverting" measurement error model). This paper also uses a maximum likelihood technique to estimate an earnings model in

which the variance components of the error term are estimated directly. The SIPP and DER earnings are stacked and the measurement errors of each are identified simultaneously, while the earnings mobility estimate is identified off of the autocorrelation coefficients of the error terms.

3 Previous Evidence

There is very little evidence concerning how much mobility estimates may be affected by measurement error. A very large literature uses only survey-based data to study earnings mobility in the U.S. or other countries. See Atkinson, Bourguignon, and Morrisson (1992) for an excellent review of the earlier literature. Later studies include Abowd and Card (1989), Buchinsky and Hunt (1996), Burkhauser, Holtz-Eakin, and Rhody (1997), Gittleman and Joyce (1995, 1996), Gottschalk et al (1994), Gottschalk and Moffitt (2002), Haider (2001), Hause (1980), Hungerford (1993), Levy and Murnane (1992), Lillard and Weiss (1979), Lillard and Willis (1978), MaCurdy (1982), and Shorrocks (1981). A much smaller literature uses only administrative-based data to study mobility. In an attempt to work with an error-free measure of earnings, a number of researchers working on France have used administrative-based earnings measures rather than survey-based earnings measures, see, for example, Bigard, Guillotin, and Lucifora (1998), Buchinsky, Fougère, and Kramarz (1998), and Buchinsky, Fields, Fougère, and Kramarz (2003). Baker and Solon (2003) used Canadian tax records to study earnings dynamics and inequality.

The previous literature offers a small number of studies that make selective comparisons of survey-based versus administrative-based results; please see Bound, Brown, and Mathiowetz (2001) for a complete survey of this literature through the 1990s and Abowd and Stinson (2005) and Gottschalk and Huynh (2006) for more recent contributions. Duncan and Hill (1985), Bound and Krueger (1991), Bound et al. (1994), and Pischke (1995) all find evidence of "mean-reverting measurement error," defined as low earners tending to overstate earnings in surveys relative to administrative reports and high earners tending to understate them. Bound and Krueger (1991) report that for men nearly 65% of the observed variation in earnings changes is true variation, while for women the corresponding percentage is 80%. Pischke (1995) was the first of these studies to establish the relationship between measurement error and earnings dynamics. Pischke proposes a simple model in which annual earnings are composed of a permanent (random-walk) component and a transitory (white noise) component and measurement error is composed of a person-specific component which is constant over time, a component which is correlated with the transitory component of earnings, and white noise. When this model is applied to the Panel Study of Income Dy-

namics Validation Study (PSIDVS) data, the implications for earnings dynamics are that the white-noise error more than offsets the underreporting of transitory earnings, resulting in a slight understatement of the permanence in earnings changes in the survey-based data, relative to the administrative-based data. In other words, using this particular measurement error model, Pischke shows that the autocorrelations in earnings changes can be estimated correctly using only survey-based data, despite this measurement error. Pischke finds a slightly higher ratio of true variation to total variation in earnings changes than do Bound and Krueger - 0.8, to be exact. All of these authors use data for the U.S. that is earlier than the 1990s. In addition, the PSIDVS sample was small and not representative. This study builds on the previous work by using a larger and nationally representative sample to study the effect of measurement error on earnings changes estimates during the 1990s.

Abowd and Stinson (2005) created a person-job level dataset from the SIPP-SSA public use file used in this study by matching each SIPP respondent's reported jobs to his/her jobs from the Detailed Earnings Record (taken from Box 1 on the W-2 form) by employer name. Assuming that neither survey-based nor administrative-based earnings equal "true" earnings, but that both are measured with error, they estimated the ratio of true to total variance to be between 0.85 and 0.87 for survey-based earnings and between 0.73 and 0.80 for administrative-based earnings; the corresponding ratios for survey-based earnings changes and administrative-based earnings changes were 0.71 and 0.75, respectively. This study estimates a similar model for earnings, but does not use the imputed DER earnings and does not control for firm effects. Because this paper does not used matched SIPP and DER jobs, the estimates do not suffer from possible job mismatch errors.

As stated above, several studies find evidence of "mean-reverting" measurement error, or a negative correlation between the measurement error and the value of earnings as given by the employer-recorded or administrative earnings. To formalize how this finding will affect estimates of micro mobility (following Kim and Solon 2005), consider the textbook model of errors-in-variables:

$$(3)\ y_{it} = y^*_{it} + w_{it}$$

where y_{it} is observed earnings, y^*_{it} is "true" latent earnings, and the measurement error w_{it} is assumed to have zero mean and to be orthogonal to y^*_{it}. This model can be viewed as a restricted version of a more general model of measurement error:

$$(4)\ y_{it} = n_i + \lambda y^*_{it} + w_{it}$$

where n_i is an individual-specific effect for reporting error and w_{it} is again uncorrelated with y_{it} and each of its determinants. The textbook model of measurement error is the case

where $n_i = 0$ and $\lambda = 1$. The evidence of "mean-reverting" measurement error found in the literature corresponds to a value of λ that falls between 0 and 1. Differencing equation (4) leads to

$$(5)\ \Delta y = \lambda \Delta y^* + \Delta w.$$

Now suppose the earnings mobility equation we wish to estimate takes the following form:

$$(6)\ \Delta y^* = \rho \Delta x + \varepsilon$$

where x is a vector of determinants and ε is independently and identically distributed and orthogonal to Δx. What the researcher is actually able to estimate is the following:

$$(7)\ \Delta y = \rho_1 \Delta x + \varepsilon_2.$$

Least squares will provide a consistent estimate of $\rho_1$since both components of the error term (ε_2 and Δw) are orthogonal to the regressors. But if $0 < \lambda < 1$, then least squares provides estimates of ρ that are biased downward by λ (i.e., plim $\hat{\rho}_1 = \lambda\rho$). Bound et al. (1994) estimate equation (5) and obtain a value for λ of 0.779 (with standard error 0.041) using least squares.

Many earnings mobility studies in the United States and elsewhere seek to estimate the following type of conditional model which includes lagged earnings as an explanatory variable:

$$(8)\ \Delta y^* \equiv y^*_{it} - y^*_{it-1} = \rho x + \delta y^*_{it-1} + \varepsilon.$$

What the researcher is actually able to estimate is the following:

$$(9)\ \Delta y = \rho_1 x + \delta_1 y_{it-1} + \varepsilon,$$

where, for the univariate case, plim $\hat{\rho}_1 = \lambda\rho$ and

$$(10)\ \text{plim}\ \hat{\delta}_1 = \frac{\delta Var(y^*_{it-1})}{Var(y^*_{it-1}) + (1/\lambda^2)Var(w_{it-1})}.$$

See the appendix for the derivation of (10). It is easy to see that certain types of measurement error will cause biased mobility estimates.

Gottschalk and Huynh (2006) derive the analytical link between mean-reverting measurement error and two measures of mobility - the elasticity of log earnings with respect to lagged earnings and the correlation between current log earnings and lagged log earnings - and show that the various biases from mean-reverting measurement error act in offsetting

directions. Specifically, their decomposition equation is of the form

$$(11)\ \hat{\beta}_{yy_{-1}} = \beta_{yy_{-1}}(1 + \{(\beta_{wy^*} - \beta_{w_{-1}y^*_{-1}})\frac{var(y^*_{-1})}{var(y_{-1})}\}) + \{(\beta_{wy^*} - \beta_{w\varepsilon})\frac{var(\varepsilon)}{\beta_{yy_{-1}}var(y_{-1})}\} + \{[\beta_{ww_{-1}} + \beta_{\varepsilon w_{-1}} - \beta]\frac{var(w_{-1})}{var(y_{-1})}\},$$

where $\widehat{\beta}_{yy_{-1}}$ is the slope coefficient from a regression of log earnings on lagged log earnings:

$$(12)\ y_{it} = \beta_{yy_{-1}}y_{it-1} + \varepsilon,$$

and the measurement error in log earnings and lagged log earnings takes the following textbook model form:

$$(13)\ y_{it} = y^*_{it} + w_{it}$$

$$(14)\ y_{it-1} = y^*_{it-1} + w_{it-1},$$

where measurement errors w_{it} and w_{it-1} are assumed to have zero mean and to be orthogonal to y^*_{it} and y^*_{it-1}, respectively. Using the SIPP-SSA linked data, which is what I also use, Gottschalk and Huynh find that the mean-reverting measurement error in SIPP earnings almost completely offsets the bias of classical measurement error, resulting in very similar mobility estimates using survey-based and administrative-based earnings.

In summary, my review of the literature has found scattered evidence concerning how much estimates of earnings mobility may be affected by measurement error. To the best of my knowledge, no previous study uses both survey-based and administrative-based earnings to test whether the finding of convergence is robust to various functional forms and to models which use both earnings sources simultaneously to obtain mobility estimates that are unbiased by measurement error.

4 Data Description

In this research, I use a new dataset called the Survey of Income and Program Participation-Social Security Administration Public Use File (SIPP-SSA PUF), Version 4.0, which was created by the Longitudinal Employer Household Dynamics (LEHD) program at the U.S. Census Bureau. The dataset contains individually reported total annual labor earnings from the SIPP linked by Social Security Number (SSN) to employer reported total annual labor earnings subject to income tax from the Social Security Administration's Detailed Earnings Record (DER).

The SIPP-SSA person-level file contains five stacked SIPP panels (1990, 1991, 1992, 1993, and 1996). Some of the panels overlap chronologically, but each panel surveys a disjoint group of individuals. Together, the five panels cover all the years from 1990-1999. The 1990 and 1991 panels are two years long (e.g., the 1990 panel includes earnings data for 1990 and 1991), the 1992 and 1993 panels are three years long, and the 1996 panel is four years long. When studying earnings changes from one year to the next (called "one-year" mobility), I use all five panels. When studying three-year mobility from 1996-1999, I use the 1996 panel only. The instrumental variables estimation of Section 4.1 requires at least three years of consecutive earnings for every individual, in which case I use only the 1992, 1993, and 1996 panels.

Stacked together, the five panels include a total of 353,120 individuals. Each individual's record includes self-reported SIPP earnings for the years covered by the particular panel in which s/he was interviewed. The dataset also includes several key variables reported on the SIPP survey (race, age, gender, marital status, etc) and a flag variable indicating whether the individual has a validated social security number (SSN) and was thus able to be matched to his/her record in the SSA data. The method for validating SSNs for all five of the SIPP panels was as follows: If a SIPP respondent refused to provide an SSN, then no attempt was made to obtain a match for that person in the administrative data. If a respondent provided an SSN, then a clerk used their name, address, and personal information to look them up in the SSA master file of all applications for Social Security cards (called the Numident file). If the Numident SSN matched the self-reported SSN, then the record was labeled as having a validated SSN. In cases where the Numident SSN was different from the self-reported SSN, the clerk filled in the correct SSN from the Numident file and the record was labeled as having a validated SSN.

For those individuals who have a validated SSN, the person-level dataset includes annual earnings subject to FICA as reported on the Social Security Administration's Summary Earnings Record (SER), which are capped at the FICA taxable maximum, and the annual detailed earnings records (DER) as reported in the Social Security Administration's Master Earnings File, which are taken directly from Box 1 on the W-2 form and are not capped. The SER earnings are available for each individual for every year from 1951 to 2003. The DER earnings are available for each worker-employer combination for every year from 1978 through 2003. These job-level earnings are summed across employers to obtain total annual DER earnings for each individual.

If an individual does not have a validated SSN, then his/her SSA annual earnings (both SER and DER) were imputed using a multiple imputation technique for nonresponse in surveys. I exclude these individuals from the sample. Dropping individuals who do not have

a validated SSN assumes that SSNs are missing completely at random. Missing completely at random means that the probability of an observation being missing (being without a validated SSN) does not depend on observed or unobserved measurements. This is unlikely to be true. It is far more likely that the set of individuals who have a validated social security number differs systematically from the set of individuals who do not. Indeed the data confirm this hypothesis, which I will discuss shortly. However, I feel that the advantage of having reported (as opposed to imputed) DER earnings that is gained by using only individuals with a validated social security number far outweighs the disadvantage of having a sample that is only representative of the population of individuals with validated social security numbers, as opposed to being representative of the entire population. Therefore I use only those individuals with validated social security numbers and claim that my sample is representative of the civilian non-institutionalized U.S. population of individuals with validated SSNs.[3]

One would like to know whether the sample of individuals with validated SSNs is representative of the entire civilian non-institutionalized U.S. population of individuals, both with and without validated SSNs. One way to test this is to see whether the percentage of people with validated SSNs is the same for key personal characteristics as it is for the whole sample. Table 1 shows the percentage of observations who have validated SSNs broken into groups by demographic variables and other key variables in the data. For most groups, the percentage of observations with validated SSNs is close to 82.92%, which is the percentage of observations in the whole sample with validated SSNs. For a few groups (Hispanic, never married, and born outside the U.S.) the percentage is slightly smaller (around 75%). Because I am including only those individuals with validated SSNs, my sample likely includes fewer illegal immigrants than a representative sample would. The fact that there are fewer individuals born outside the U.S. in my sample than in the whole sample supports this hypothesis. A second way to test whether my sample is representative of the entire civilian non-institutionalized population is to compare the means and variances of several key variables across the two samples. These are shown in Table 2. For no variable do I reject the hypothesis that the means are equal for the two samples. Thus, Tables 1 and 2 together provide evidence that the set of individuals with validated Social Security numbers is mostly similar to the set of all individuals, both with and without validated SSNs, in the sample. The total number of individuals in the five SIPP panels with validated SSNs is 242,600.

All of the individuals with validated SSNs have non-missing SER and DER earnings.

[3] My dissertation will include an appendix in which I rerun all estimations from this paper using the full sample of individuals, both with and without validated SSNs.

However, some of these individuals have missing SIPP data. It is well known that ignoring missing data will result in a self-selection bias for regression coefficients because the missing data is not likely missing at random but is in fact correlated with other variables in the survey. Usual corrections for this self-selection bias include the maximum likelihood method of Griliches, Hall, and Hausman (1978) or Heckman's (1979) 2-stage procedure. However, when estimating earnings changes regressions for multiple time periods simultaneously, these method can become quite complicated because the self-selection condition for each respondent will depend on an entire sequence of stochastic disturbances. In addition, the likelihood function might involve cross products between these disturbances and individual demographic characteristics. An alternative method is to impute missing data. As part of the Census Bureau project which created this dataset, all SIPP data that were originally missing were completed using multiple imputation methods originally proposed by Rubin (1993) and updated by Raghunathan et al (2003). This imputation resulted in four completed datasets which each contain the "true" underlying microdata where they were available (or non-missing) and imputed missing data. These four completed datasets are analyzed by first analyzing each completed dataset separately and then combining results (such as regression coefficients) using formulas presented in Rubin (1987). Choosing to use multiply imputed data ensures that the mobility estimates presented in this paper do not suffer from attrition or self-selection biases; they could, however, include new biases resulting from poor imputation models. This paper is part of a larger Census Bureau project to assess the imputation quality of the multiply completed SIPP-SSA file.

I use three different samples in this paper, which I will refer to as the regular sample, the positive earners sample, and non-imputed earners sample. All three samples include only individuals ages 25-60 with validated SSNs who were labor force participants in all relevant consecutive years (e.g., two consecutive years for one-year mobility, four consecutive years for three-year mobility, etc). An individual was defined as a labor force participant if s/he either a) had positive SIPP earnings for the year, b) had positive DER earnings for the year, or c) reported in the SIPP that s/he was actively looking for work during at least one month of that year. The positive earners sample further restricts the sample to include only those individuals who had positive SIPP and DER earnings for all relevant consecutive years. The economic decisions and processes that determine whether an individual has any labor earnings at all are likely very different from those decisions and processes that determine how much an individual earns. Using the positive earners sample ignores the first set of decisions/processes and focuses instead on how much an individual earns given that s/he had positive earnings in a given year and in previous years. The non-imputed earners sample further restricts the sample to include only those individuals who had non-imputed, positive

SIPP and DER earnings for all relevant consecutive years. The SIPP collects information on earnings at a monthly level. Reported earnings are then summed across twelve months to create the annual earnings measure used in this study. The non-imputed earners sample includes only those individuals who had non-missing SIPP earnings in all twelve months for all relevant years and is thus much smaller than the other two samples.

Each of the three samples is trimmed using the following method. First, I estimated a mixed effect model for year-specific SIPP earnings with fixed personal characteristics and random person and employer effects using only SIPP earnings data that were within five standard deviations of the year-specific SIPP earnings mean. Then I created a residual for every observation, including those not used to fit the model. I repeated this process using DER earnings. Using the residual variances from these two models, I dropped year-individual observations with either the SIPP residual or the DER residual (or both) greater than five residual standard deviations.

Finally, depending on which time frame for mobility is being used, all sets of consecutive years of earnings are stacked. For example, to study one-year earnings changes, I stack all sets of two consecutive years from 1990-1991 up to 1998-1999. The reason for this is that I wish to study mobility throughout the 1990s, rather than mobility in any particular year. As a result of this stacking, some individuals appear more than once in my sample (for example, if they were labor force participants from 1997-1998 and were also labor force participants from 1998-1999).

A few final items should be noted. First, all earnings variables are expressed as real earnings in January 1995. Second, all of the SIPP panels are stratified multistage probability samples rather than simple random samples. The results presented in this paper take into account the SIPP sampling error resulting from this multistage sampling design by clustering on the primary sampling unit, which is the first-stage cluster in the SIPP sampling design.

4.1 SIPP earnings vs. DER earnings

I define "true" latent earnings as the earnings obtained from the labor market, exclusive of other compensation such as benefits. "True" earnings include pre-tax employee contributions to deferred compensation plans, such as 401(k) retirement plans, and pre-tax employee-paid health insurance plan premiums. "True" earnings do not include any type of benefits, such as employer contributions to health insurance plans and deferred compensation plans, Medical Savings Accounts, educational assistance above a certain monetary level, fringe benefits, etc.

I have several reasons to believe that the DER earnings measure is as close to "true"

latent earnings as it is possible to get, though I will not assume in this study that it is completely free of measurement error. First, the DER earnings measure is not capped at the FICA taxable maximum amount as is the SER earnings measure used in many previous earnings validation studies. Second, I am able to distinguish between self-employment DER earnings and employer DER earnings in the job-level dataset. This study will use only those jobs that represent wage and salary earnings and will exclude self-employment income. Hence, summing the DER earnings measure across jobs for each individual provides a measure of total employer-reported annual labor earnings from all jobs. This measure is directly comparable to the SIPP measured of annual labor earnings constructed by summing twelve monthly values of wage and salary earnings reported by the SIPP respondent.

There are several circumstances under which DER earnings may not equal "true" earnings. The first arises when an employee underreports tips and other earnings to the employer. I would prefer to drop all occupations which are likely to have large portions of their earnings in the form of tips, but the occupation variable available on the SIPP-SSA public use file is too coarse for this, with only five categories. Second, there are two items which may be reported under "gross earnings" on an employee's pay stub and which I include in my definition of "true" earnings, but which are not included in Box 1 on the W-2 form: pre-tax health insurance plan premiums and pre-tax contributions to deferred compensation plans, such as 401(k) retirement plans. Health insurance plan premiums are not likely to be missing from the DER earnings measure in a way that varies systematically with any of the explanatory variables, and hence will not bias the mobility estimates. Pre-tax contributions to deferred compensation plans are reported elsewhere on the W-2 form (for example in Box 13 in 1999) and I added them to Box 1 to obtain gross earnings. Thirdly, DER earnings can include the following items, all of which employers are required to report as part of taxable income: employer contributions to health insurance plans, Medical Savings Accounts, educational assistance above a certain monetary level, certain types of fringe benefits, etc.

DER earnings may differ from SIPP reported earnings in the following circumstances, even though these differences are not a result of measurement error in either SIPP or DER earnings. First, SIPP respondents are only asked to report earnings on at most two jobs in any given month. If the respondent held more than two jobs in that month, then the DER annual earnings measure will include earnings from all employers for that month, while the SIPP annual earnings measure will not include earnings from the additional jobs. Second, annual SIPP earnings are topcoded (at $150,000 for the 1996 panel and at $100,000 for the earlier panels) while DER earnings are not. However, the individuals affected by this topcoding are not included in our sample as a result of the trimming described above.

For a number of reasons - because the DER earnings are not capped, because I am not

including self-employment income, because I can add pre-tax contributions to deferred compensation plans onto Box 1 earnings, and because health insurance plan premiums missing from DER earnings are not likely to be correlated with other variables in the dataset - I believe that the DER earnings measure is as close to "true" earnings as it is possible to get. However, because the DER earnings may not include tips and health insurance plan premiums and may include employer contributions to health insurance plans and other such benefits, I will not assume in this study that DER earnings are without measurement error (i.e., equal to "true" earnings).

5 Empirical Methodology

The first research question asks: are the findings of unconditional and conditional convergence robust to non-linear functional forms and alternate samples of the data? I use two non-linear functional forms: 1) a step function by initial earnings quintile and 2) a non-parametric specification. I use three different samples, which I will refer to as the regular sample, the positive earners sample, and non-imputed earners sample. These samples are described in Section 3.

The question of unconditional convergence asks whether low initial earners or high initial earners experienced more positive earnings changes. I use the following linear base specification to study unconditional convergence:

$$\begin{aligned} (15)\ y_{it} - y_{it-1} &= \alpha + \beta y_{it-1} + \varepsilon_{it} \text{ and} \\ (16)\ y_{it} - y_{it-3} &= \alpha + \beta y_{it-3} + \varepsilon_{it}, \end{aligned}$$

where y_{it} represents dollar earnings for individual i in time period t. Equation (15) studies one-year mobility and equation (16) studies three-year mobility. To estimate equation (15), I stack all sets of two consecutive years of earnings from 1990-1991 up to 1998-1999. As a result of this stacking, some individuals appear more than once in the sample (for example, if they were labor force participants from 1997-1998 and were also labor force participants from 1998-1999). To estimate equation (16), I use only the 1996 SIPP panel to study earnings changes over the four-year period from 1996-1999. Both equations (15) and (16) are estimated separately for SIPP and DER earnings. The second specification is a step

function by initial earnings quintile:

$$
\begin{aligned}
(17)\ y_{it} - y_{it-1} &= \alpha + \beta_0 \text{Quintile2}_{it-1} + \beta_1 \text{Quintile3}_{it-1} + \beta_2 \text{Quintile4}_{it-1} \\
&\quad + \beta_3 \text{Quintile5}_{it-1} + \varepsilon_{it} \text{ and} \\
(18)\ y_{it} - y_{it-3} &= \alpha + \beta_0 \text{Quintile2}_{it-3} + \beta_1 \text{Quintile3}_{it-3} + \beta_2 \text{Quintile4}_{it-3} \\
&\quad + \beta_3 \text{Quintile5}_{it-3} + \varepsilon_{it},
\end{aligned}
$$

where Quintile1-Quintile5 are dummy variables equal to 1 if the individual's earnings in that time period were in that quintile of the earnings distribution. Quintile 1 is the lowest quintile and is the excluded group; Quintile 5 is the highest earnings quintile. The third specification estimates equations (15) and (16) using non-parametric regressions.

The question of conditional convergence asks whether, holding other things equal, low initial earners or high initial earners experienced more positive earnings changes. I use the following linear base specification to study unconditional convergence:

$$
\begin{aligned}
(19)\ y_{it} - y_{it-1} &= \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \beta_2 y_{it-1} + \varepsilon_{it} \text{ and} \\
(20)\ y_{it} - y_{it-3} &= \alpha + \beta_0 Z_i + \beta_1 Exp_{it-3} + \beta_2 y_{it-3} + \varepsilon_{it},
\end{aligned}
$$

where y_{it} represents dollar earnings for individual i in time period t, Z_i includes dummy variables for gender, race, age, education (all time-invariant variables), and Exp_{it} represents experience. Age is broken into three levels: ages 25-36, 37-48, and 49-60, with the youngest individuals being the excluded group. Education is also broken into three levels: no high school degree, high school degree, and college degree or higher, with the least educated individuals being the excluded group. Experience is calculated from the Social Security earnings history and equals the total number of years between age 25 and the current year during which the individual had positive social security earnings (SER earnings). I do not interpret this as a causal model of earnings changes, but rather a way of answering the question of which individuals (in terms of initial earnings level, race, gender, and broad levels of age and education) experienced the most positive earnings changes, holding other things equal. For this reason, notice that I do not include certain other variables that one would usually expect in an earnings equation, such as age squared, or interaction terms between race, gender, and education. The second specification is a step function by initial earnings

quintile:

$$
\begin{aligned}
(21)\ y_{it} - y_{it-1} &= \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \beta_2 \text{Quintile2}_{it-1} + \beta_3 \text{Quintile3}_{it-1} \\
&\quad + \beta_4 \text{Quintile4}_{it-1} + \beta_5 \text{Quintile5}_{it-1} + \varepsilon_{it} \text{ and} \\
(22)\ y_{it} - y_{it-3} &= \alpha + \beta_0 Z_i + \beta_1 Exp_{it-3} + \beta_2 \text{Quintile2}_{it-3} + \beta_3 \text{Quintile3}_{it-3} \\
&\quad + \beta_4 \text{Quintile4}_{it-3} + \beta_5 \text{Quintile5}_{it-3} + \varepsilon_{it},
\end{aligned}
$$

where Quintile1-Quintile5 are dummy variables equal to 1 if the individual's earnings in that time period were in that quintile of the earnings distribution. Quintile 1 is the lowest quintile and is the excluded group; Quintile 5 is the highest earnings quintile. The third specification estimates equations (19) and (20) using non-parametric regressions.

All of the regressions discussed in this section are weighted. As part of the Census Bureau project which created the SIPP-PUF file, a weight was created for all five stacked panels which makes the panels representative of the Decennial Census U.S. population of all civilian, non-institutionalized individuals as of April 1st, 2000. When studying one-year earnings changes, this Decennial weight is used. When studying three-year mobility, which uses only the 1996 SIPP panel, the person weight from that panel is used. Using weights can cause regression coefficients to be biased when earnings are used as an independent variable because income is included in the stratum for building the weights, which causes the weights to be endogenous in the model. As a robustness check, all of the OLS regressions are run once using weights and once without weights. I now turn to the second research question.

5.1 Instrumental Variables - Methodology

The second research question asks: does the finding of conditional convergence hold when neither SIPP nor DER earnings are treated as equal to "true" latent earnings? I use two different methods to obtain an estimate of conditional mobility that treats neither data source as being free of measurement error: 1) instrumental variables (IV) and generalized method of moments (GMM), and 2) maximum likelihood estimation (MLE). I will discuss the details of the IV and GMM estimation first.

Recall that when estimating equations (19) and (20), measurement error in earnings causes the lagged earnings variable on the right hand side of the equation to be correlated with the error term, thus producing biased estimates of β_2. The method of instrumental variables (IV) addresses this problem by replacing the lagged earnings variable on the right hand side of the equation with an instrumental variable which is correlated with the lagged earnings but uncorrelated with the error term in the equation. This study uses SIPP

and DER earnings as instruments for each other. Before moving on to the details of the estimation, let us rewrite the conditional convergence equation (19) separately for SIPP and DER earnings:

$$
\begin{aligned}
(23a)\ \ln(y_{SIPPit}) - \ln(y_{SIPPit-1}) &= \alpha^{SIPP} + \beta_0^{SIPP} Z_i + \beta_1^{SIPP} Exp_{it-1} + \beta_2^{SIPP} \ln(y_{SIPPit-1}) \\
&\quad + \varepsilon_{it}^{SIPP} \\
(23b)\ \ln(y_{DERit}) - \ln(y_{DERit-1}) &= \alpha^{DER} + \beta_0^{DER} Z_i + \beta_1^{DER} Exp_{it-1} + \beta_2^{DER} \ln(y_{DERit-1}) \\
&\quad + \varepsilon_{it}^{DER}.
\end{aligned}
$$

Earnings are now specified in logs rather than dollars, because I want to control for differences in scale when using IV and MLE methods. Let us make one further transformation as follows:

$$
\begin{aligned}
(24a)\ \ln(y_{SIPPit}) &= \alpha^{SIPP} + \beta_0^{SIPP} Z_i + \beta_1^{SIPP} Exp_{it-1} + \gamma^{SIPP} \ln(y_{SIPPit-1}) + \varepsilon_{it}^{SIPP} \\
(24b)\ \ln(y_{DERit}) &= \alpha^{DER} + \beta_0^{DER} Z_i + \beta_1^{DER} Exp_{it-1} + \gamma^{DER} \ln(y_{DERit-1}) + \varepsilon_{it}^{DER},
\end{aligned}
$$

where $\gamma^{SIPP} = 1 + \beta_2^{SIPP}$ and $\gamma^{DER} = 1 + \beta_2^{DER}$. One might worry that if ε_{it}^{SIPP} and ε_{it}^{DER} are not identically and independently distributed, that this transformation will not provide equivalent estimates of β_0 and β_1 across equations (23) and (24) and that γ will not equal $1+\beta_2$. I find in our data that this is not the case. Columns (1) and (2) of Tables 6 through 9 present estimates of equations (23) and (24). We see that indeed, β_0 and β_1 are equivalent across these two equations, and that $\gamma = 1 + \beta_2$ using both SIPP and DER earnings.

A value of $-1 < \beta_2 < 0$ or $0 < \gamma < 1$ corresponds to a finding of convergence, while a finding of $\beta_2 > 0$ or $\gamma > 1$ corresponds to a finding of divergence. In the language of the mobility literature, we are interested in the amount of time dependence - to what degree earnings in the current year depend upon earnings in the previous year. In the language of the broader labor economics literature, we are interested in how persistent earnings are over time - do they perform as a random walk from one year to the next or is there persistence in earnings over time? The amount of persistence is measured by the magnitude of γ. The closer γ is to 1 (or the closer β_2 is to 0), the less persistence there is in earnings over time, and the further away γ is from 1 (or the further away β_2 is from 0), the more persistence there is in earnings over time. I now turn to the IV estimation in detail.

Let us first assume that measurement error in SIPP and DER earnings is classical (I will drop this assumption shortly). Specifically, let us assume that measurement error follows

the textbook model in which we observe $\ln(y_{SIPPit})$ and $\ln(y_{DERit})$:

$$\begin{aligned}
(25a)\ \ln(y_{SIPPit}) &= \ln(y^*_{it}) + w_{it} \\
(25b)\ \ln(y_{DERit}) &= \ln(y^*_{it}) + v_{it} \text{ OR} \\
(26a)\ \ln(y^*_{it}) &= \ln(y_{SIPPit}) - w_{it} \\
(26b)\ \ln(y^*_{it}) &= \ln(y_{DERit}) - v_{it} \\
(27a)\ E(w_{it}) &= 0 \\
(27a)\ E(v_{it}) &= 0 \\
(28a)\ Cov(\ln(y^*_{it}), w_{it}) &= 0 \\
(28b)\ Cov(\ln(y^*_{it}), v_{it}) &= 0 \\
(29)\ Cov(w_{it}, v_{it}) &= 0,
\end{aligned}$$

where $\ln(y^*_{it})$ represents the log of true "latent" earnings, and w_{it} and v_{it} have zero mean and are orthogonal to $\ln(y^*_{it})$ and to each other. The equation I wish to estimate is:

$$(30)\ \ln(y^*_{it}) = \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \gamma \ln(y^*_{it-1}) + \varepsilon_{it}.$$

Plug (26a) into (30) and rearrange:

$$\begin{aligned}
(31a)\ \ln(y_{SIPPit}) - w_{it} &= \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \gamma(\ln(y_{SIPPit-1}) - w_{it-1}) + \varepsilon_{it} \\
(32a)\ \ln(y_{SIPPit}) &= \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \gamma \ln(y_{SIPPit-1}) + \varepsilon_{it} + w_{it} - \gamma w_{it-1}.
\end{aligned}$$

I will now instrument for lagged SIPP earnings with lagged DER earnings. The lagged DER earnings, which are equal to $\ln(y_{DERit-1}) = \ln(y^*_{it-1}) + v_{it-1}$, are not correlated with the error term in (32a). The 1st stage IV equation is:

$$(33a)\ \ln(y_{SIPPit-1}) = \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \delta \ln(y_{DERit-1}) + e_{it},$$

which is used to get predicted values $\ln(y'_{SIPPit-1})$. The 2nd stage IV equation is:

$$(34a)\ \ln(y_{SIPPit}) = \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \gamma \ln(y'_{SIPPit-1}) + u_{it}.$$

I also repeat this exercise using lagged SIPP earnings to instrument for lagged DER earnings.

Plug (26b) into (30) and rearrange:

$$
\begin{aligned}
(31b)\ \ln(y_{DERit}) - v_{it} &= \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \gamma(\ln(y_{DERit-1}) - v_{it-1}) + \varepsilon_{it} \\
(32b)\ \ln(y_{DERit}) &= \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \gamma \ln(y_{DERit-1}) + \varepsilon_{it} + v_{it} - \gamma v_{it-1}.
\end{aligned}
$$

I will now instrument for lagged DER earnings with lagged SIPP earnings. The lagged SIPP earnings, which are equal to $\ln(y_{SIPPit-1}) = \ln(y^*_{it-1}) + w_{it-1}$, are not correlated with the error term in (32b). The 1st stage IV equation is:

$$(33b)\ \ln(y_{DERit-1}) = \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \delta \ln(y_{SIPPit-1}) + c_{it},$$

which is used to get predicted values $\ln(y'_{DERit-1})$. The 2nd stage IV equation is:

$$(34b)\ \ln(y_{DERit}) = \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \gamma \ln(y'_{DERit-1}) + d_{it}.$$

It is unlikely that measurement error in earnings follows the textbook model. Duncan and Hill (1985), Bound and Krueger (1991), Bound et al. (1994), and Pischke (1995) all find evidence of "mean-reverting" measurement error, defined as low earners tending to overstate earnings in surveys relative to administrative reports and high earners tending to understate them. Furthermore, it is unlikely that w_{it} and v_{it} are uncorrelated. For example, underreported tips that are missing from the DER earnings measure are most likely missing from the SIPP earnings measure as well. I therefore redo the IV estimation using first differences and a "mean-reverting" model for measurement error. Specifically, let us assume that measurement error takes the following form, in which we observe $\ln(y_{SIPPit})$ and $\ln(y_{DERit})$:

$$
\begin{aligned}
(35a)\ \ln(y_{SIPPit}) &= \lambda^{SIPP} \ln(y^*_{it}) + n_i + p_i + w_{it} \\
(35b)\ \ln(y_{DERit}) &= \lambda^{DER} \ln(y^*_{it}) + n_i + q_i + v_{it} \text{ OR} \\
(36a)\ \ln(y^*_{it}) &= \frac{1}{\lambda^{SIPP}}[\ln(y_{SIPPit}) - n_i - p_i - w_{it}] \\
(36b)\ \ln(y^*_{it}) &= \frac{1}{\lambda^{DER}}[\ln(y_{DERit}) - n_i - p_i - w_{it}] \\
(37a)\ E(w_{it}) &= 0 \\
(37a)\ E(v_{it}) &= 0 \\
(38a)\ Cov(\ln(y^*_{it}), w_{it}) &= 0 \\
(38b)\ Cov(\ln(y^*_{it}), v_{it}) &= 0 \\
(39)\ Cov(w_{it}, v_{it}) &= 0,
\end{aligned}
$$

where $\ln(y_{it}^*)$ represents the log of true "latent" earnings, n_i is an individual-specific reporting error which is constant over time and which affects both the SIPP and DER earnings, p_i and q_i are individual-specific reporting errors which are constant over time and which are unique to the SIPP and DER, and where w_{it} and v_{it} have zero mean and are orthogonal to $\ln(y_{it}^*)$ and to each other. This model allows for measurement error components which are common to both SIPP and DER (captured by n_i), such as missing tips. This model also allows for measurement error components which are unique to SIPP and DER (captured by p_i and q_i), such as the pre-tax health insurance plan premiums which are missing from DER earnings but likely not missing from SIPP earnings. The key assumption of the following instrumental variables estimation is that the measurement error component which is common to both SIPP and DER (captured by n_i) is constant over time, so that taking first differences rids the equation of this component. To see this, first plug (36a) into (30) (the equation we wish to estimate) and rearrange:

$$
\begin{aligned}
(40a)\ \frac{1}{\lambda^{SIPP}}[\ln(y_{SIPPit}) - n_i - p_i - w_{it}] &= \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} \\
&\quad + \gamma \frac{1}{\lambda^{SIPP}}[\ln(y_{SIPPit-1}) - n_i - p_i - w_{it-1}] + \varepsilon_{it} \\
(41a)\ \ln(y_{SIPPit}) &= \lambda^{SIPP}\alpha + \lambda^{SIPP}\beta_0 Z_i + \lambda^{SIPP}\beta_1 Exp_{it-1} \\
&\quad + \gamma \ln(y_{SIPPit-1}) + (1-\gamma)n_i + (1-\gamma)p_i + w_{it} \\
&\quad - \gamma w_{it-1} + \lambda^{SIPP}\varepsilon_{it}.
\end{aligned}
$$

Now rewrite this equation for period t-1 and solve for $(1-\gamma)n_i + (1-\gamma)p_i$:

$$
\begin{aligned}
(42a)\ \ln(y_{SIPPit-1}) &= \lambda^{SIPP}\alpha + \lambda^{SIPP}\beta_0 Z_i + \lambda^{SIPP}\beta_1 Exp_{it-2} + \gamma \ln(y_{SIPPit-2}) \\
&\quad + (1-\gamma)n_i + (1-\gamma)p_i + w_{it-1} - \gamma w_{it-2} + \lambda^{SIPP}\varepsilon_{it-1} \\
(43a)\ (1-\gamma)n_i + (1-\gamma)p_i &= \ln(y_{SIPPit-1}) - \lambda^{SIPP}\alpha - \lambda^{SIPP}\beta_0 Z_i - \lambda^{SIPP}\beta_1 Exp_{it-2} \\
&\quad - \gamma \ln(y_{SIPPit-2}) - w_{it-1} + \gamma w_{it-2} - \lambda^{SIPP}\varepsilon_{it-1}.
\end{aligned}
$$

Next plug (43a) into (41a) for $(1-\gamma)n_i + (1-\gamma)p_i$ to get:

$$\begin{aligned}
(44a)\ \ln(y_{SIPPit}) &= \lambda^{SIPP}\alpha + \lambda^{SIPP}\beta_0 Z_i + \lambda^{SIPP}\beta_1 Exp_{it-1} + \gamma\ln(y_{SIPPit-1}) \\
&+ \ln(y_{SIPPit-1}) - \lambda^{SIPP}\alpha - \lambda^{SIPP}\beta_0 Z_i \\
&-\lambda^{SIPP}\beta_1 Exp_{it-2} - \gamma\ln(y_{SIPPit-2}) - w_{it-1} + \gamma w_{it-2} \\
&-\lambda^{SIPP}\varepsilon_{it-1} + w_{it} - \gamma w_{it-1} + \lambda^{SIPP}\varepsilon_{it} \\
(45a)\ \ln(y_{SIPPit}) - \ln(y_{SIPPit-1}) &= \lambda^{SIPP}\beta_1(Exp_{it-1} - Exp_{it-2}) + \gamma(\ln(y_{SIPPit-1}) \\
&- \ln(y_{SIPPit-2})) + w_{it} - w_{it-1} + \gamma w_{it-2} - \gamma w_{it-1} \\
&+\lambda^{SIPP}\varepsilon_{it} - \lambda^{SIPP}\varepsilon_{it-1}.
\end{aligned}$$

I will now instrument for lagged SIPP earnings changes with lagged DER earnings changes. The lagged DER earnings changes, which are equal to

$$(46a)\ln(y_{DERit-1}) - \ln(y_{DERit-2}) = \lambda^{DER}\ln(y^*_{it-1}) + v_{it-1} - \lambda^{DER}\ln(y^*_{it-2}) - v_{it-2},$$

are not correlated with the error term in (45a). The 1st stage IV equation is:

$$(47a)\ \ln(y_{SIPPit-1}) - \ln(y_{SIPPit-2}) = b(Exp_{it-1} - Exp_{it-2}) + \delta(\ln(y_{DERit-1}) - \ln(y_{DERit-2})) + e_{it},$$

which is used to get predicted values $\ln(y'_{SIPPit-1}) - \ln(y'_{SIPPit-2})$. The 2nd stage IV equation is:

$$(48a)\ \ln(y_{SIPPit}) - \ln(y_{SIPPit-1}) = b_1(Exp_{it-1} - Exp_{it-2}) + \gamma(\ln(y'_{SIPPit-1}) - \ln(y'_{SIPPit-2})) + u_{it}.$$

I also repeat this exercise using lagged SIPP earnings changes to instrument for lagged DER earnings changes. First plug (36b) into (30) (the equation I wish to estimate) and rearrange:

$$\begin{aligned}
(40b)\ \frac{1}{\lambda^{DER}}[\ln(y_{DERit}) - n_i - p_i - v_{it}] &= \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} \\
&+\gamma\frac{1}{\lambda^{DER}}[\ln(y_{DERit-1}) - n_i - p_i - v_{it-1}] + \varepsilon_{it} \\
(41b)\ \ln(y_{DERit}) &= \lambda^{DER}\alpha + \lambda^{DER}\beta_0 Z_i + \lambda^{DER}\beta_1 Exp_{it-1} \\
&+\gamma\ln(y_{DERit-1}) + (1-\gamma)n_i + (1-\gamma)p_i \\
&+v_{it} - \gamma v_{it-1} + \lambda^{DER}\varepsilon_{it}.
\end{aligned}$$

Now rewrite this equation for period t-1 and solve for $(1-\gamma)n_i + (1-\gamma)p_i$:

$$\begin{aligned}
(42b)\ \ln(y_{DERit-1}) &= \lambda^{DER}\alpha + \lambda^{DER}\beta_0 Z_i + \lambda^{DER}\beta_1 Exp_{it-2} + \gamma \ln(y_{DERit-2}) \\
&\quad +(1-\gamma)n_i + (1-\gamma)p_i + v_{it-1} - \gamma v_{it-2} + \lambda^{DER}\varepsilon_{it-1} \\
(43b)\ (1-\gamma)n_i + (1-\gamma)p_i &= \ln(y_{DERit-1}) - \lambda^{DER}\alpha - \lambda^{DER}\beta_0 Z_i - \lambda^{DER}\beta_1 Exp_{it-2} \\
&\quad -\gamma \ln(y_{DERit-2}) - v_{it-1} + \gamma v_{it-2} - \lambda^{DER}\varepsilon_{it-1}.
\end{aligned}$$

Next plug (43b) into (41b) for $(1-\gamma)n_i + (1-\gamma)p_i$ to get:

$$\begin{aligned}
(44b)\ \ln(y_{DERit}) &= \lambda^{DER}\alpha + \lambda^{DER}\beta_0 Z_i + \lambda^{DER}\beta_1 Exp_{it-1} + \gamma \ln(y_{DERit-1}) \\
&\quad + \ln(y_{DERit-1}) - \lambda^{DER}\alpha - \lambda^{DER}\beta_0 Z_i \\
&\quad -\lambda^{DER}\beta_1 Exp_{it-2} - \gamma \ln(y_{DERit-2}) - v_{it-1} + \gamma v_{it-2} \\
&\quad -\lambda^{DER}\varepsilon_{it-1} + v_{it} - \gamma v_{it-1} + \lambda^{DER}\varepsilon_{it} \\
(45b)\ \ln(y_{DERit}) - \ln(y_{DERit-1}) &= \lambda^{DER}\beta_1(Exp_{it-1} - Exp_{it-2}) + \gamma(\ln(y_{DERit-1}) \\
&\quad - \ln(y_{DERit-2})) + v_{it} - v_{it-1} + \gamma v_{it-2} - \gamma v_{it-1} \\
&\quad +\lambda^{DER}\varepsilon_{it} - \lambda^{DER}\varepsilon_{it-1}.
\end{aligned}$$

I will now instrument for lagged DER earnings changes with lagged SIPP earnings changes. The lagged SIPP earnings changes, which are equal to

$$(46b)\ln(y_{SIPPit-1}) - \ln(y_{SIPPit-2}) = \lambda^{SIPP}\ln(y^*_{it-1}) + w_{it-1} - \lambda^{SIPP}\ln(y^*_{it-2}) - w_{it-2},$$

are not correlated with the error term in (45b). The 1st stage IV equation is:

$$(47b)\ \ln(y_{DERit-1}) - \ln(y_{DERit-2}) = B(Exp_{it-1} - Exp_{it-2}) + \pi(\ln(y_{SIPPit-1}) - \ln(y_{SIPPit-2})) + c_{it},$$

which is used to get predicted values $\ln(y'_{DERit-1}) - \ln(y'_{DERit-2})$. The 2nd stage IV equation is:

$$(48b)\ \ln(y_{DERit}) - \ln(y_{DERit-1}) = B_1(Exp_{it-1} - Exp_{it-2}) + \gamma(\ln(y'_{DERit-1}) - \ln(y'_{DERit-2})) + d_{it}.$$

The sample used for the IV estimations differs from the sample used to answer the first research question in two ways. First, I use only the positive earners sample. The reason for this, as stated in Section 3, is that I wish to focus on the economic decisions and processes that determine how much an individual earns given that s/he had positive earnings in a given year and in previous years. I do also run all of the IV estimations using the non-

imputed earners sample. These results can be found in the appendix. Second, even though I are still interested in studying one-year earnings changes, the IV method which assumes "mean-reverting" measurement error and which uses first differences to rid the equation of the common measurement error component n_i requires that each individual have at least three consecutive years of earnings data available. Therefore, I use only the 1992, 1993, and 1996 SIPP panels, all of which provide at least three years of consecutive earnings for each individual. These three panels provide us with four different sets of three consecutive years of earnings: 1992-1994 (provided by the 1992 panel), 1993-1995 (provided by the 1993 panel), 1996-1998 (provided by the 1996 panel), and 1997-1999 (provided by the 1996 panel). I stack all four sets of three consecutive years of earnings. As a result of this stacking, some individuals from the 1996 panel appear more than once in the sample (for example, if they had positive earnings from 1996-1998 and also had positive earnings from 1997-1999). I include dummy variables indicating which set of three consecutive years an individual's earnings came from in all the IV, GMM, and MLE estimations in this paper. The IV method which assumes classical measurement error requires that each individual have only two consecutive years of earnings available. However, in order for these IV results to be comparable to the IV results obtained from the method which assumes "mean-reverting" measurement error, I use the same sample for both types of IV estimation.

All of the IV regressions are weighted. Because three SIPP panels are used in these regressions, the weight used is a pooled-panel version of the SIPP person weight from those three panels. This pooled panel weight is created according to the weighting rules for combining multiple SIPP panels found in the SIPP Users Guide (2001). As stated above, using weights can cause regression coefficients to be biased when earnings are used as an independent variable. Therefore, all of the IV regressions are run once using weights and once without weights.

5.2 Generalized Method of Moments - Methodology

The method of instrumental variables does not make full use of the information on earnings available in the data. In addition, the IV method makes strict assumptions on the error terms in the 1st and 2nd stage equations. The generalized method of moments (GMM) is an instrumental method that is superior to IV in both these respects. The efficiency gains of the GMM estimator relative to the traditional IV / 2SLS estimator derive from the use of an optimal weighting matrix, the overidentifying restrictions of the model, and the relaxation of the i.i.d. assumption on the error terms (Greene 2003). I discuss each of these efficiency gains in the next paragraph.

The IV estimation of the preceding section used the lagged DER earnings (and earnings changes) to instrument for lagged SIPP earnings (and earnings changes), and vice versa, but there is more information about earnings and earnings changes in the data which can be brought to bear on estimation. For example, in equations (32a), (32b), (45a), and (45b), not only are the disturbances at time t uncorrelated with the instrumental variables at time t, but they are also uncorrelated with the instrumental variables in other time periods. Therefore, the instrumental variables from other time periods can be added to the set of instruments. As a specific example, for equation (32a), when I instrument for SIPP earnings in 1998 using the IV method, the only instrument is DER earnings in 1998. When I use the GMM method to instrument for SIPP earnings in 1998, the set of instrumental variables includes not only DER earnings in 1998, but also DER earnings in 1997 and DER earnings in 1999. For equation (45a), when I instrument for SIPP earnings changes from 1997 to 1998 using the IV method, the only instrument is DER earnings changes from 1997 to 1998. When I use the GMM method to instrument for SIPP earnings changes from 1997 to 1998, the set of instrumental variables includes not only DER earnings changes from 1997 to 1998, but also DER earnings changes from 1998 to 1999. Because there is more than one instrumental variable in each equation, the model is overidentified, and thus the GMM method provides more efficient estimates than the IV method. Another flaw with the IV estimation of the preceeding section is that it makes strict assumptions on the error terms in the 1st and 2nd stage equations. In particular, the IV method assumes that the 1st and 2nd stage equations have disturbances that are distributed identically and independently. The GMM method relaxes this assumption. Finally, the IV method, like OLS, uses the identity matrix as a weighting matrix when solving the least squares minimum distance problem. In contrast, the GMM method uses an optimal weighting matrix in which the weights are inversely proportional to the variances of the moments. This optimization increases the efficiency of the GMM estimator over the IV estimator.

All of the GMM estimations in this paper use the same sample that was used for the IV estimations. This sample includes only three panels (1992, 1993, and 1996) and contains only those individuals who had at least three consecutive years of positive earnings. In addition, all of the GMM regressions are weighted using the same pooled panel weight that was used in the IV regressions. Because using weights introduces a possible endogeneity bias, all of the GMM regressions are run once using weights and once without weights.

5.3 Maximum Likelihood Estimation - Methodology

There are several reasons why the technique of maximum likelihood estimation (MLE) is superior to the methods of instrumental variables and generalized method of moments for estimating earnings equations. First, it is expected that earnings and earnings changes depend upon unobservable person effects. When instrumenting in levels, these person effects are ignored. When instrumenting after taking first differences, these person effects (if time-invariant) are excluded from the equation. For this reason, we may obtain very different results when instrumenting in levels versus instrumenting in differences. The maximum likelihood method allows us to estimate these unobserved person effects directly. Second, the MLE method makes weaker assumptions about the underlying data generating process than does the IV method. The IV method assumes that the 1st and 2nd stage equations have disturbances that are distributed identically and independently. In contrast, the maximum likelihood estimation provides unbiased and efficient estimates for functions of sample characteristics regardless of the underlying data generating process. Third, the IV and GMM methods require the assumption of a particular measurement error model. In contrast, the MLE method allows for the specification of a more general model of measurement error. Furthermore, the maximum likelihood method allows us to estimate the various components of measurement error directly. For all of these reasons, I now turn to estimating earnings mobility using the technique of maximum likelihood estimation.

The maximum likelihood estimation is similar to that done by Abowd and Stinson (2005). I specify the SIPP earnings equation for individual i in time period t as:

$$\begin{aligned}
(49a)\ \ln(y_{SIPPit}) &= \alpha^{SIPP} + \beta_0 Z_i + \beta_1 Exp_{it-1} + \theta_i + \eta_{it} + w_{it} \\
(50a)\ \eta_{it} &= \rho\eta_{it-1} + e_{it} \\
(51a)\ w_{it} &= \rho_{SIPP} w_{it-1} + u_{it},
\end{aligned}$$

and the DER earnings equation for the same individual as:

$$\begin{aligned}
(49b)\ \ln(y_{DERit}) &= \alpha^{DER} + \beta_0 Z_i + \beta_1 Exp_{it-1} + \theta_i + \eta_{it} + v_{it} \\
(50b)\ \eta_{it} &= \rho\eta_{it-1} + e_{it} \\
(51b)\ v_{it} &= \rho_{DER} v_{it-1} + q_{it},
\end{aligned}$$

where Z and Exp are defined in the beginning of Section 5, θ is an individual-specific, time-invariant intercept effect, η is a common error component, and w and v are the measurement errors of SIPP and DER earnings, respectively. The common error component, η represents economic shocks which influence true latent earnings. It is identified by the fact that there

are two earnings observations for each individual in each time period (SIPP and DER). The measurement errors, w and v, represent fluctuations in annual earnings reports which are due to things that effect SIPP and DER earnings separately, and which do not influence true latent earnings. These measurement errors are identified by the differences between SIPP and DER earnings for each individual in each time period. The random effects are modeled as follows:

$$(52) \text{ Person heterogeneity} = \theta \sim N(0, G_1)$$

$$(53) \text{ Common error component} = \eta \sim N(0, G_2)$$

$$(54) \text{ Measurement error, SIPP and DER} = (w, v) \sim N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, R),$$

where G_1, G_2, and R are defined below. The total number of individuals is I, the number of covariates included in X is k, and the total number of time periods is 3. The model can be written in matrix notation as:

$$(55)\ Y = X\beta + Zu + e,$$

where Y contains stacked SIPP and DER earnings, X contains covariates treated as fixed effects, β contains fixed effects coefficients, Z is a design matrix for the random effects contained in u, u contains the stacked random effects $\theta_1 ... \theta_I, \eta_{1t} ... \eta_{It+2}$, and e contains the stacked measurement error terms $w_{1t} ... w_{It+2}$, $v_{1t} ... v_{It+2}$. The sizes for these matrices are:

$$\begin{aligned} Y &: (I * 3 * 2) * 1 \\ X &: (I * 3 * 2) * k \\ \beta &: k * 1 \\ Z &: (I * 3 * 2) * (I * 3) \\ u &: (I * 3) * 1 \\ e &: (I * 3 * 2*) * 1 \end{aligned}$$

The variance matrices for the random person effects, common error component effects, and measurement errors, respectively, are:

$$(56)\ G_1 = I_{IxI} \otimes \sigma_\theta^2$$

$$(57)\ G_2 = I_{IxI} \otimes \sigma_\eta^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}_{3x3}$$

$$(58)\ R = I_{IxI} \otimes \begin{bmatrix} \sigma_w^2 \begin{bmatrix} 1 & \rho_{SIPP} & \rho_{SIPP}^2 \\ \rho_{SIPP} & 1 & \rho_{SIPP} \\ \rho_{SIPP}^2 & \rho_{SIPP} & 1 \end{bmatrix}_{3x3} & 0_{(Ix3)x(Ix3)} \\ 0_{(Ix3)x(Ix3)} & \sigma_v^2 \begin{bmatrix} 1 & \rho_{DER} & \rho_{DER}^2 \\ \rho_{DER} & 1 & \rho_{DER} \\ \rho_{DER}^2 & \rho_{DER} & 1 \end{bmatrix}_{3x3} \end{bmatrix},$$

where $\sigma_\eta^2 = \sigma_\varepsilon^2/(1-\rho^2)$, ρ is the autocorrelation parameter of the common error component η, and ρ_{SIPP} and ρ_{DER} are the autocorrelation parameters of the SIPP and DER measurement errors, respectively. This model specifies that the common error component, η, and the two measurement errors, w and v, each follow a first-order autoregressive process. Estimates of $\alpha^{SIPP}, \alpha^{DER}, \beta_0, \beta_1$, the variance components $(\sigma_\theta^2, \sigma_\eta^2, \sigma_w^2, \sigma_v^2)$, and realizations of the random effects (θ, η) and the residuals (w,v) are obtained by solving the equations:

$$(59)\ \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \widehat{\beta} \\ \widehat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}Y \\ Z'R^{-1}Y \end{bmatrix}.$$

The estimation is done by restricted maximum likelihood with a software package called ASReml, using an average information algorithm developed and programmed by Gilmour, Thompson, and Cullis (1995). As described by Abowd and Stinson (2005), the solution parameters to these equations are calculated iteratively by maximizing the log likelihood function to satisfy a set of first order conditions. The starting values for the variance components are: $\sigma_\theta^2 = \sigma_\eta^2 = \sigma_w^2 = \sigma_v^2 = 0.1$ and $\rho = \rho_{SIPP} = \rho_{DER} = 0.3$.

You will notice that the earnings equations (49a) and (49b) do not include lagged earnings as a regressor. This is another key difference between the IV/GMM methods and the MLE method. Lagged earnings are partially determined by observable personal characteristics like race, gender, and education. Including these personal characteristics and lagged earnings in an earnings equation causes the right hand side variables to be correlated with each other and dilutes the contribution of each variable's effect on current earnings. The MLE method avoids this problem by specifying current earnings as a time series. Because the error terms in the maximum likelihood equations (49a) and (49b) contain an autoregressive component, an individual's earnings in each year are not only a function of personal characteristics (both observable and unobservable), but also a function of the entire history of past earnings.

The results from the MLE method are not directly comparable to those obtained from

running OLS or IV/GMM methods on equations (24a) and (24b). I will now algebraically convert equations (49a) and (49b) into equations similar to those estimated in Sections 5.1 and 5.2. Once this conversion is complete, I will have a new set of equations similar to, but not exactly like, equations (24a) and (24b). These new equations will be estimated alongside equations (49a) and (49b) so that I have something to compare our MLE results to, in order to tell how much the estimates change as a result of specifying the measurement error terms directly. In this way, we will know whether the finding of convergent mobility is robust to specifying a full measurement error model and treating neither SIPP nor DER earnings as equal to true latent earnings. I begin by rewriting equation (49a) for period t-1 and solving for η_{it-1} and w_{it-1}:

$$\begin{aligned}
(60)\ \ln(y_{SIPPit-1}) &= \alpha^{SIPP} + \beta_0 Z_i + \beta_1 Exp_{it-2} + \theta_i + \eta_{it-1} + w_{it-1}, \text{ or} \\
(61)\ \eta_{it-1} &= \ln(y_{SIPPit-1}) - \alpha^{SIPP} - \beta_0 Z_i - \beta_1 Exp_{it-2} - \theta_i - w_{it-1} \text{ and} \\
(62)\ w_{it-1} &= \ln(y_{SIPPit-1}) - \alpha^{SIPP} - \beta_0 Z_i - \beta_1 Exp_{it-2} - \theta_i - \eta_{it-1}.
\end{aligned}$$

Now plug (61) and (62) into (50a) and (51a):

$$\begin{aligned}
(63)\ \eta_{it} &= \rho(\ln(y_{SIPPit-1}) - \alpha^{SIPP} - \beta_0 Z_i - \beta_1 Exp_{it-2} - \theta_i - w_{it-1}) + e_{it} \\
(64)\ w_{it} &= \rho_{SIPP}(\ln(y_{SIPPit-1}) - \alpha^{SIPP} - \beta_0 Z_i - \beta_1 Exp_{it-2} - \theta_i - \eta_{it-1}) + u_{it}.
\end{aligned}$$

Now plug (63) and (64) into (49a) and rearrange:

$$\begin{aligned}
(65)\ \ln(y_{SIPPit}) &= \alpha^{SIPP} + \beta_0 Z_i + \beta_1 Exp_{it-1} + \theta_i + \rho(\ln(y_{SIPPit-1}) - \alpha^{SIPP} - \beta_0 Z_i \\
&\quad -\beta_1 Exp_{it-2} - \theta_i - w_{it-1}) + e_{it} + \rho_{SIPP}(\ln(y_{SIPPit-1}) - \alpha^{SIPP} \\
&\quad -\beta_0 Z_i - \beta_1 Exp_{it-2} - \theta_i - \eta_{it-1}) + u_{it} \\
(66a)\ \ln(y_{SIPPit}) &= (1-\rho-\rho_{SIPP})\alpha^{SIPP} + (1-\rho-\rho_{SIPP})\beta_0 Z_i + \beta_1 Exp_{it-1} \\
&\quad +(1-\rho-\rho_{SIPP})\theta_i + (\rho+\rho_{SIPP})\ln(y_{SIPPit-1}) \\
&\quad -(\rho\beta_1 + \rho_{SIPP}\beta_1)Exp_{it-2} - \rho w_{it-1} - \rho_{SIPP}\eta_{it-1} + e_{it} + u_{it}.
\end{aligned}$$

Using the same method for DER yields:

$$\begin{aligned}
(66b)\ \ln(y_{DERit}) &= (1-\rho-\rho_{DER})\alpha^{DER} + (1-\rho-\rho_{DER})\beta_0 Z_i + \beta_1 Exp_{it-1} \\
&\quad +(1-\rho-\rho_{DER})\theta_i + (\rho+\rho_{DER})\ln(y_{DERit-1}) \\
&\quad -(\rho\beta_1 + \rho_{DER}\beta_1)Exp_{it-2} - \rho w_{it-1} - \rho_{DER}\eta_{it-1} + e_{it} + u_{it}.
\end{aligned}$$

One can see that equations (66a) and (66b) contain twice-lagged experience, Exp_{it-2}, and a

person effect, θ_i, while equations (24a) and (24b) do not. Therefore, the maximum likelihood estimation of equations (49a)-(51a) and (49b)-(51b) is not comparable to the IV/GMM estimation of equations (24a) and (24b). However, the maximum likelihood estimation is comparable to the estimation of equations (66a) and (66b). Therefore, the maximum likelihood results will be compared to the coefficients obtained from estimating equations (66a) and (66b), rather than the coefficients from equations (24a) and (24b). Finally, it should be noted that none of the maximum likelihood estimations are weighted because many of the variables which were used in the stratum to create the weights are also included as regressors, causing the weights to be endogenous in the model.

6 Results

Overall, the finding of convergent mobility, both conditional and unconditional, is robust. The first research question asks whether the findings of unconditional and conditional convergence are robust to non-linear functional forms and alternate samples of the data. The parametric results have been completed and they show that the finding of unconditional convergence is robust to alternate samples and to a non-linear functional form of the equation, namely, a step function by initial earnings quintile. The non-parametric model still remains to be done and will be completed in my dissertation. The second research question asks whether the finding of conditional convergence is robust to methods which treat neither SIPP nor DER earnings as equal to "true" latent earnings. The IV, GMM, and MLE results all show that the finding of conditional convergence is robust to methods that use both earnings sources simultaneously to obtain mobility estimates that are unbiased by measurement error.

6.1 Non-Linear Functional Forms and Alternate Samples

The first research question asks: are the findings of unconditional and conditional convergence robust to non-linear functional forms and alternate samples of the data? I use two non-linear functional forms - a step function by initial earnings quintile and a non-parametric specification. The parametric results have been completed and are presented here. The non-parametric results will be completed in my dissertation. I use three different samples - the regular sample, the positive earners sample, and non-imputed earners sample. The results are presented in Table 3.

Table 3 presents results from the estimation of equations (15) through (22). The following items are varied in Table 3: a) the sample - page 1 of Table 3 contains the results for the

regular sample, page 2 contains the results for positive earners, and page 3 contains results for non-imputed positive earners, b) one-year mobility vs. three-year mobility - the left side of each page contains one-year results and the right side of each page contains three-year results, c) unconditional mobility vs. conditional mobility - the top section of each page contains unconditional results and the bottom section of each page contains conditional results, d) SIPP earnings vs. DER earnings, and e) the functional form (linear, step function by initial earnings quintile, non-parametric). Along with a finding of convergence or divergence, Table 3 reports the number of coefficients on the initial earnings variable(s) that were significant out of the total number of initial earnings coefficients in the regression. For example, when using a step function to estimate one-year unconditional mobility using the regular sample and SIPP earnings, I find that all four coefficients on the initial earnings quintile dummies in equation (17) were significant. All the regressions presented in Table 3 use earnings expressed in dollars. All results presented in Table 3 are significant at the 99% confidence level.

Table 3 shows that the finding of convergence is robust to alternate samples of the data and to a non-linear functional form of the data, namely, a step function by initial earnings quintile. A finding of convergence indicates that the coefficient(s) on initial earnings from one of the equations (15)-(22) was (were) significantly less than zero. For example, when using a step function to estimate one-year unconditional mobility using the regular sample and SIPP earnings, I find that the estimates of $\beta_0, \beta_1, \beta_2$, and β_3 from equation (17) were all significantly less than zero. I always find significant convergence whether I are using the regular sample, the positive earners sample, the non-imputed earners sample, SIPP earnings, DER earnings, one-year earnings changes, three-year earnings changes, a linear function, or a step function. In other words, even when varying the sample, the functional form, and the earnings measure, I find that it is the individuals with the lowest initial earnings in dollars who gain the most in dollar earnings over time. This is a finding of strong convergence, and it implies a finding of weak convergence - that those individuals with the lowest initial earnings gained the most in percentages over time. Furthermore, using a step function by initial earnings quintile demonstrates that even when I allow the rate of convergence to vary across quintile, it is the earners in the lowest quintile who experience the most positive earnings changes while those in the highest earnings quintile experience the least positive (or most negative) earnings changes. In summary, the findings of unconditional and conditional strong convergence are robust to a non-linear functional form and to alternate samples of the data.

6.2 Conditional Convergence - IV and GMM Results

The second research question asks: does the finding of conditional convergence hold when neither SIPP nor DER earnings are treated as equal to "true" latent earnings? I use two different methods to obtain an estimate of conditional mobility that treats neither data source as being free of measurement error: 1) instrumental variables (IV) and generalized method of moments (GMM), and 2) maximum likelihood estimation (MLE).

Tables 4 and 5 show the means and standard deviations of several key variables for the positive earners sample and the non-imputed earners sample, respectively. Table 6 shows the unweighted results of the IV and GMM estimations using DER earnings to instrument for SIPP earnings. The table columns contain the estimation of the following equations:

$$\text{Column 1:} \quad (23a)\ \ln(y_{SIPPit}) - \ln(y_{SIPPit-1}) = \alpha^{SIPP} + \beta_0^{SIPP} Z_i + \beta_1^{SIPP} Exp_{it-1} + \beta_2^{SIPP} \ln(y_{SIPPit-1}) + \varepsilon_{it}^{SIPP}$$

$$\text{Column 2:} \ (24a)\ \ln(y_{SIPPit}) = \alpha^{SIPP} + \beta_0^{SIPP} Z_i + \beta_1^{SIPP} Exp_{it-1} + \gamma^{SIPP} \ln(y_{SIPPit-1}) + \varepsilon_{it}^{SIPP}$$

$$\text{Column 3 Upper:} \ (34a)\ \ln(y_{SIPPit}) = \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \gamma \ln(y'_{SIPPit-1}) + u_{it}$$

$$\text{Column 3 Lower:} \ (33a)\ \ln(y_{SIPPit-1}) = \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \delta \ln(y_{DERit-1}) + e_{it}$$

$$\text{Column 4:} \ (34a)\ \ln(y_{SIPPit}) = \alpha + \beta_0 Z_i + \beta_1 Exp_{it-1} + \gamma \ln(y'_{SIPPit-1}) + u_{it}$$

$$\begin{aligned}\text{Column 5:} \quad (45a)\ \ln(y_{SIPPit}) - \ln(y_{SIPPit-1}) = \ & \lambda^{SIPP} \beta_1 (Exp_{it-1} - Exp_{it-2}) \\ & + \gamma(\ln(y_{SIPPit-1}) - \ln(y_{SIPPit-2})) + w_{it} \\ & - w_{it-1} + \gamma w_{it-2} - \gamma w_{it-1} + \lambda^{SIPP} \varepsilon_{it} \\ & - \lambda^{SIPP} \varepsilon_{it-1}\end{aligned}$$

$$\begin{aligned}\text{Column 6 Lower:} \quad (47a)\ \ln(y_{SIPPit-1}) - \ln(y_{SIPPit-2}) = \ & b(Exp_{it-1} - Exp_{it-2}) \\ & + \delta(\ln(y_{DERit-1}) \\ & - \ln(y_{DERit-2})) + e_{it}\end{aligned}$$

$$\begin{aligned}\text{Column 7:} \quad (48a)\ \ln(y_{SIPPit}) - \ln(y_{SIPPit-1}) = \ & b_1(Exp_{it-1} - Exp_{it-2}) \\ & + \gamma(\ln(y'_{SIPPit-1}) \\ & - \ln(y'_{SIPPit-2})) + u_{it}.\end{aligned}$$

Recall that the IV estimation uses only one instrumental variable, whereas the GMM estimation uses more than one. In other words, the GMM estimation in columns 4 and 7 estimates the same equation as the IV estimation in columns 3 and 6, but the predicted value of earnings (or earnings changes) is computed using more than one instrument. In column 1, we see that regressing earnings changes from one year to the next on earnings in the initial year results in a negative estimate of β_2^{SIPP}, i.e., a finding of convergence. When I transform the equation and regress earnings in the current year on earnings in the previous year in column 2, we see that $\widehat{\gamma^{SIPP}} = 1 + \widehat{\beta_2^{SIPP}} = 0.71 < 1$, again, a finding of convergence. In column 3, DER earnings in the previous year are now used to instrument for SIPP earnings in the previous year. I find that $\widehat{\gamma} = 0.83$, indicating that after controlling for measurement error in SIPP initial earnings, I find less convergence (or more persistence) in earnings than in column 2. The last row of Table 6 shows the corresponding $\widehat{\beta_2^{SIPP}}$ term of $-0.17 = 0.83 - 1$. The lower section of column 3 shows the first stage regression. We see that the instrumental variable is fairly good at predicting the SIPP initial earnings variable, with a T-statistic of 209.60 and an R^2 regression statistic of 0.54. In column 4, DER earnings from all three years are now used to instrument for SIPP earnings in the previous year. I find that $\widehat{\gamma} = 0.87$, indicating again that when I control for measurement error in SIPP initial earnings, I find less convergence (or more persistence) in earnings than I did using OLS. In column 5, SIPP earnings changes are regressed on lagged SIPP earnings changes. The γ estimate of -0.23 corresponds to a value of -1.23 for $\widehat{\beta_2^{SIPP}}$, which is very different in magnitude from the value for $\widehat{\beta_2^{SIPP}}$ when earnings are specified in levels. This large difference is likely due to the fact that there are time-invariant unobserved person effects which contribute both to earnings and earnings changes and which were ignored entirely during the estimation using earnings levels. Taking first differences of equation (24a) rids the equation of these effects, causing the estimate of γ^{SIPP} to change accordingly. I still find convergence, indicated by the fact that $-1.23 < 0$. In column 6, DER lagged earnings changes are now used to instrument for SIPP lagged earnings changes. I find that $\widehat{\gamma} = -0.13$ (or $\widehat{\beta_2^{SIPP}} = -1.13$), indicating that when I control for measurement error in SIPP earnings changes, I find less convergence (or more persistence) in earnings changes than I did using OLS. The lower section of column 6 shows the first stage regression. The instrumental variable is somewhat good at predicting the SIPP lagged earnings changes, with a T-statistic of 56.48 and an R^2 regression statistic of 0.08, though it is not as strong of a predictor as the DER earnings levels were of the SIPP earnings levels. In column 7, DER earnings changes from two different years are now used to instrument for SIPP lagged earnings changes. I find that $\widehat{\gamma} = 0.00$, indicating again that when I control for measurement error in SIPP lagged earnings changes, I find less convergence (or more persistence) in earnings changes than I did using OLS.

Table 7 shows the weighted results of the IV and GMM estimations using DER earnings to instrument for SIPP earnings. The overall results are the same as in Table 6. I find convergence using both earnings specified in levels and earnings specified in changes. I find that when using IV and GMM methods, specified either in levels or in changes, there is less convergence (or more persistence) in earnings than was found using OLS. The results using earnings changes are again very different in magnitude from those using earnings levels as a result of the unobserved person effects.

Tables 7 and 9 present the IV and GMM results (unweighted and weighted, respectively) using SIPP earnings to instrument for DER earnings. The main difference to note is that in column 1, the DER earnings show less convergence overall ($\widehat{\beta_2^{DER}} = -0.18$) than did the SIPP earnings ($\widehat{\beta_2^{SIPP}} = -0.29$). Other than this difference, DER earnings behave in much the same way as SIPP earnings. In particular, I find convergence using both earnings specified in levels and earnings specified in changes and I find that when using IV and GMM methods, there is less convergence (or more persistence) in earnings than was found using OLS. The weighted results are very similar to the unweighted results.

Tables 10 through 13 repeat the analysis from tables 6 through 9 using the non-imputed earners sample. The results are extremely similar to those obtained using the positive earners sample. Again, I find less convergence (or more persistence) in earnings once I have used IV and GMM methods to control for measurement error in initial earnings.

Overall, Tables 6 through 13 provide evidence that the finding of conditional convergence holds when neither SIPP nor DER earnings are treated as equal to "true" latent earnings. However, I find less convergence using IV and GMM methods than I do using OLS. Stated in slightly different terms, measurement error in SIPP and DER earnings makes earnings and earnings changes look less persistent over time than they truly are. Measurement error makes it appear that temporary earnings shocks die out faster than they actually do. Once I have controlled for measurement error in SIPP and DER earnings using IV and GMM methods, I find that earnings persist more over time, i.e., that the effect of temporary earnings shocks lasts longer than previously believed (using OLS on either SIPP or DER earnings separately). I now turn to the results of the maximum likelihood estimation.

6.3 Conditional Convergence - MLE Results

Table 14 shows the results of the maximum likelihood estimation of equations (49) through (51). Tables 15 and 16 show the estimation of equations (66a) and (66b), respectively. For the positive earners sample, the autocorrelation coefficient on the common error component in Table 14 equals 0.09 and the autocorrelation coefficient on the SIPP measurement error

equals 0.08; the sum of these is 0.17, which corresponds to the value of γ^{SIPP} from equation (24a). The corresponding $\widehat{\beta_2^{SIPP}}$ term from equation (23a) equals $0.17 - 1 = -0.83$. I wish to compare the value of 0.17 to the value of the coefficient on lagged earnings from equation (66a), which is reported in Table 15 as 0.72. Recall that equations (66a) and (49a)-(51a) are algebraically equivalent; the main difference is that equation (66a) is specified as a regression of current earnings on lagged earnings and does not estimate measurement error directly, while equations (49a)-(51a) are specified as an earnings model with an autocorrelation structure on the error term and measurement error in SIPP earnings is estimated directly. Both equations (66a) and (49a) contain a random person effect. We see that when measurement error is estimated directly, the coefficient on lagged earnings decreases greatly from 0.72 to 0.17. It appears that measurement error causes us to find less convergence (or more persistence) in earnings than truly occurred. In other words, I find less persistence in earnings when I control for measurement error by estimating it directly.

A similar result appears using the administrative-based earnings. The autocorrelation coefficient on the DER measurement error in Table 14 equals 0.63; adding to this the autocorrelation coefficient on the common error component (0.09) gives us 0.72, which corresponds to the value of γ^{SIPP} from equation (24a). I wish to compare the value of 0.72 to the value of the coefficient on lagged earnings from equation (66b), which is reported in Table 15 as 0.83. We see that when measurement error is estimated directly, the coefficient on lagged earnings decreases from 0.83 to 0.72. Again, it appears that measurement error causes us to find less convergence (or more persistence) in earnings than truly occurred. In other words, I find less persistence in earnings when I control for measurement error by estimating it directly. Overall, using the positive earners sample, I find that convergence holds when neither earnings source is treated as being free of measurement error.

I now turn to the results using the non-imputed earners sample. The autocorrelation coefficient on the common error component in Table 14 equals 0.27 and the autocorrelation coefficient on the SIPP measurement error equals 0.08; the sum of these is 0.35, which corresponds to the value of γ^{SIPP} from equation (24a). The corresponding $\widehat{\beta_2^{SIPP}}$ term from equation (23a) equals $0.35 - 1 = -0.65$. I wish to compare the value of 0.35 to the value of the coefficient on lagged earnings from equation (66a), which is reported in Table 16 as 0.78. We see that when measurement error is estimated directly, the coefficient on lagged earnings decreases from 0.78 to 0.35. Again, it appears that measurement error causes us to find less convergence (or more persistence) in earnings than truly occurred.

A different result appears using the administrative-based earnings and the non-imputed earners sample. The autocorrelation coefficient on the DER measurement error in Table 14 equals 0.70; adding to this the autocorrelation coefficient on the common error component

(0.27) gives us 0.97, which corresponds to the value of γ^{SIPP} from equation (24a). I wish to compare the value of 0.97 to the value of the coefficient on lagged earnings from equation (66b), which is reported in Table 16 as 0.84. Now we see that when measurement error is estimated directly, the coefficient on lagged earnings increases from 0.84 to 0.97. Now it appears that measurement error causes us to find more convergence (or less persistence) in earnings than truly occurred. In other words, I find more persistence in earnings when I control for measurement error by estimating it directly. Overall, using the non-imputed earners sample, I find that convergence holds when neither earnings source is treated as being free of measurement error.

The results from the maximum likelihood estimation do not always agree with each other or with the IV/GMM results on whether I find more or less convergence after controlling for measurement error. All of the IV/GMM results show that after controlling for measurement error, I find less convergence than truly occurred. The MLE results for administrative-based earnings using the non-imputed earners sample show this same result. The MLE results for survey-based earnings using the non-imputed earners sample and the MLE results for both earnings measures using the positive earners sample show the opposite - that after controlling for measurement error, I find more convergence than truly occurred. However, one thing remains clear. The finding of conditional convergence - i.e., the finding that, holding other things equal, the lowest earners experienced the most positive earnings changes - is robust to IV/GMM/MLE methods that use both earnings sources simultaneously to obtain mobility estimates that are unbiased by measurement error.

7 Conclusion

In this study, I have shown that for the U.S. during the 1990s, the finding of convergent mobility - that the individuals with the lowest initial earnings gained more over time in dollars or percentages than the individuals with the highest initial earnings - is robust. Convergent mobility is robust to alternate functional forms, alternate samples of the data, and alternate estimation techniques which control for measurement error bias. In particular, convergent mobility is robust to a non-linear function (a step function by initial earnings quintile) and to three different samples - one which includes individuals who had zero earnings in one or more years, a second sample which includes only those individuals who had positive earnings in all relevant years, and a third sample which includes only those individuals who had non-imputed, positive earnings in all relevant years. Convergent mobility is also robust to three estimation methods which treat neither survey-based nor administrative-based earnings as being free of measurement error - instrumental variables, generalized method of moments,

and maximum likelihood estimation. All of the IV/GMM results show that when one earnings measure is used to instrument for the other, there is less convergence than found using either data source alone. Stated differently, measurement error in SIPP and DER earnings makes earnings and earnings changes look less persistent over time than they truly are. Measurement error makes it appear that temporary earnings shocks die out faster than they actually do. Most of the MLE results show the opposite - that when measurement error is estimated directly by using survey-based and administrative-based earnings simultaneously, there is more convergence than found using either data source alone. However, when using the administrative-based earnings and the non-imputed earners sample, the MLE results confirm the IV/GMM results; I find less convergence than truly occurred.

In the face of rising earnings inequality in the U.S., the results of this study should be taken as good news. As stated in Section 2, many economists consider steadily rising inequality to be worrisome because it may indicate a steady decrease in relative earnings for the poorest earners, while the highest earners continue to gain more in dollars over time. However, if there is substantial earnings mobility, then high inequality becomes less worrisome because we know that the poorest earners are not necessarily "stuck" at the bottom of the earnings distribution. Several previous earnings mobility studies have found convergent mobility, both for the U.S. and for other countries. Because measurement error in initial earnings produces a spurious link between earnings change and initial earnings level, it can cause the appearance of convergent mobility when, in fact, no convergence has truly occurred. I find that convergent mobility for the U.S. during the 1990s is not due to measurement error, but rather, is robust to methods that control for measurement error. I believe this finding should be taken as good news by economists concerned with rising inequality and by researchers studying mobility in other countries.

It is impossible to tell whether the results of this study would hold for countries other than the U.S. However, mobility researchers who do not have access to administrative-based earnings data may take comfort in the fact that, at least in the U.S. data, the finding of convergent mobility is indeed robust, and not simply due to measurement error. It would also be worthwhile for this kind of study to be conducted for other countries that have matched survey-administrative earnings records.

8 References

1. Abowd, J. M. and D. Card. 1989. "On the Covariance Structure of Earnings and Hours Changes." *Econometrica* 57 (2): 411-445.

2. Abowd, J. M. and M. H. Stinson. 2005. "Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and SSA Administrative Data." LEHD Technical Paper, U.S. Census Bureau. <http://instruct1.cit.cornell.edu/~jma7/abowd-stinson-SOLE-2003.pdf>.

3. Atkinson, A. B., F. Bourguignon, and C. Morrisson. 1992. Empirical Studies of Earnings Mobility, Harwood Academic Publishers.

4. Baker, M. and G. Solon. 2003. "Earnings Dynamics and Earnings Inequality Among Canadian Men, 1976-1992: Evidence from Longitudinal Tax Records." *Journal of Labor Economics* 21: 289-321.

5. Baulch, B. and J. Hoddinott. 2000. "Economic Mobility and Poverty Dynamics in Developing Countries." *Journal of Development Studies* 36: 1-24.

6. Bigard, A., Y. Guillotin, and C. Lucifora. 1998. "Earnings Mobility: An International Comparison of Italy and France." *Review of Income and Wealth* 44: 535-554.

7. Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers. 1994. "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data." *Journal of Labor Economics* 12: 345-368.

8. Bound, J., C. Brown, and N. Mathiowetz. 2001. "Measurement Error in Survey Data" in J. J. Jeckman and E. Leamer, eds., Handbook of Econometrics, Chapter 59, Elsevier Science B.V., Amsterdam: 3707-3745.

9. Bound, J. and A. Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics* 9 (1): 1-24.

10. Buchinsky, M., D. Fougere, and F. Kramarz. 1998. "La mobilité salariale en France: 1967-1987." *Revue Economique* 49: 879-890.

11. Buchinsky, M., G. S. Fields, D. Fougere, and F. Kramarz. 2003. "Francs or Ranks? Earnings Mobility in France, 1967-1999." CEPR Discussion Papers 3937.

12. Buchinsky, M. and J. Hunt. 1996. "Wage mobility in the United States." *Review of Economics and Statistics* 81 (3): 351-368.

13. Burkhauser, R. V., D. Holtz-Eakin, and S. E. Rhody. 1997. "Labor Earnings Mobility and Inequality in the United States and Germany During the Growth Years of the 1980s." *International Economic Review* 38 (4): 775-794.

14. Deaton, A. 1997. The Analysis of Household Surveys. Baltimore: Johns Hopkins.

15. Dragoset, L. M. and G. S. Fields. 2006. "U.S. Earnings Mobility: Comparing Survey-Based and Administrative-Based Estimates." Working Paper, Cornell University.

16. Duncan, G. J. and D. H. Hill. 1985. "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data" *Journal of Labor Economics* 3(4):508-532.

17. Fields, G. S., P. L. Cichello, S. Freije, M. Menendez, and D. Newhouse. 2003. "Household Income Dynamics: A Four Country Story." *Journal of Development Studies* 40 (2): 30-54.

18. Gilmour, A. R., R. Thompson, and B. R. Cullis. 1995. "Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models." *Biometrics* 51 (4): 1440-1450.

19. Gittleman, M. and M. Joyce. 1995. "Earnings Mobility in the United States, 1967-91." *Monthly Labor Review* 118: 3-13.

20. Gittleman, M. and M. Joyce. 1996. "Earnings Mobility and Long-Run Inequality: An Analysis Using Matched CPS Data." *Industrial Relations* 35 (2): 180-196.

21. Gottschalk, P. and M. Huynh. 2006. "Impact of Non-Classical Measurement Error on Measures of Earnings Inequality and Mobility." Working Paper, Boston College and Social Security Administration.

22. Gottschalk, P., R. Moffitt, L. F. Katz, and W. T. Dickens. 1994. "The Growth of Earnings Instability in the U.S. Labor Market." Brookings Papers on Economic Activity 2: 217-272.

23. Gottschalk, P. and R. A. Moffitt. 2002. "Trends in the Transitory Variance of Earnings in the United States." *The Economic Journal* 112: 68-73.

24. Greene, W. H. 2003. Econometric Analysis: Fifth Edition. Prentice Hall, New Jersey.

25. Griliches, Z., B. H. Hall, and J. A. Hausman. 1978. "Missing Data and Self-selection in Large Panels." Annales de L'INSEE (30-31): 137-176.

26. Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153-161.

27. Haider, S. J. 2001. "Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991." *Journal of Labor Economics* 19 (4): 799-836.

28. Hause, J. C. 1980. "The Fine Structure of Earnings and the On-the-Job Training Hypothesis." *Econometrica* 48 (4): 1013-1029.

29. Hungerford, T. L. 1993. "U.S. Income Mobility in the Seventies and Eighties." *Review of Income and Wealth* 39: 403-417.

30. Johnston, L. D. and S. H. Williamson. "The Annual Real and Nominal GDP for the United States, 1970 - Present." Economic History Services, April 1, 2006.

    <http://www.eh.net/hmit/gdp/>.

31. Kim, B. and G. Solon. 2005. "Implications of Mean-Reverting Measurement Error for Longitudinal Studies of Wages and Employment." *Review of Economics and Statistics* 87 (1): 193-196.

32. Levy, F. and R. Murnane. 1992. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations." *Journal of Economic Literature* 30: 1333-1381.

33. Lillard, L. A. and Y. Weiss. 1979. "Components of Variation in Panel Earnings Data: American Scientists 1960-1970" *Econometrica* 47 (2):437-454.

34. Lillard, L. A. and R. J. Willis. 1978. "Dynamic Aspects of Earning Mobility." *Econometrica* 46 (5): 985-1012.

35. MaCurdy, T. E. 1982. "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis." *Journal of Econometrics* 18: 83-114.

36. Pishke, J. 1995. "Measurement Error and Earnings Dynamics: Some Estimates From PSID Validation Study." *Journal of Business and Economic Statistics* 13 (3): 305-314.

37. Raghunathan, T.E., J.P. Reiter, and D.B. Rubin. 2003. "Multiple Imputation for Statistical Disclosure Limitation." *Journal of Official Statistics* 19: 1-19.

38. Rubin, D. B. 1987. Multiple Imputation for Nonresponse in Surveys. Wiley, New York.

39. Rubin, D.B. 1993. "Satisfying Confidentiality Constraints Through the Use of Synthetic Multiply-Imputed Micro-Data." *Journal of Official Statistics* 9 (2): 461-468.

40. Shorrocks, A. F. 1981. "Income Stability in the United States" in N. A. Klevmarken and J. A. Lybeck eds., The Statics and Dynamics of Income, Clevedon, Tieto.

41. U.S. Census Bureau. "Gini Coefficient for Earnings of Men and Women, Full-Time Year-Round Workers: 1967-1998." May 13, 2004. <http://www.census.gov/hhes /income/incineq/p60204/fig2.html>.

42. U.S. Census Bureau. "Measures of Individual Earnings Inequality for Full-Time, Year-Round Workers by Sex: 1967 to 2001." May 13, 2005. <http://www.census.gov /hhes/www/income/histinc/ie2.html>.

Table 1: Representativeness of the "Regular" Sample: Pooled years from 1990-1999

The "regular" sample is defined as the set of individuals ages 25-60 with validated SSNs who were labor force participants in both years for each set of two consecutive years. This table shows the percentage of observations by category who have validated SSNs out of the entire set of individuals ages 25-60 who were labor force participants for each set of two consecutive years.

Category	Sample Size	Percentage with validated SSNs	Category	Sample Size	Percentage with validated SSNs
Total	279586	82.92	Received welfare payments	23545	81.21
			Did not receive welfare payments	256041	83.07
Male	145495	82.70			
Female	134091	83.15	Received disability payments	6418	86.15
			Did not receive disability payments	273168	82.84
Black	31306	79.84			
Non black	248280	83.30	Total net worth below $100,000	75467	83.84
			Total net worth at least $100,000	204119	82.57
Hispanic	26321	75.43			
Non Hispanic	253265	83.69	Homeowner	186170	84.49
			Not homeowner	93416	79.78
25-36 years old	116526	81.19			
37-48 years old	103887	84.43	Born in country other than U.S.	31166	72.82
49-60 years old	59173	83.66	Born in U.S.	248420	84.18
By Education			Had a defined contribution		
Primary or less	31671	81.83	pension plan	63950	85.68
Secondary	174657	82.07	Did not have a defined contribution		
Higher	73258	85.41	pension plan	167788	83.32
Married	177587	85.09	Had a defined benefit pension plan	87761	85.03
Widowed	4027	81.23	Did not have a defined benefit		
Divorced/Separated	44080	82.89	pension plan	143977	83.33
Never married	53892	75.89			
			Had health insurance coverage	206027	86.84
Reported job-limiting disability	20387	82.99	Did not have health insurance		
Did not report job-limiting disability	251339	83.07	coverage	30744	83.54
By Number of Children					
0	143602	80.85			
1	54006	84.28			
2	52469	86.02			
3	20404	85.71			
4	5343	84.81			
5 or more	2805	79.68			

Notes: The total sample size of 279586 corresponds to the set of individuals ages 25 to 60 who were labor force participants in both years for each set of two consecutive years from 1990-1999. All statistics are calculated from the first completed dataset.

Table 2: Representativeness of the "Regular" Sample: Pooled years from 1990-1999

This table shows the means and variances of several key variables for both the entire sample and for the "regular" sample used in the paper. The entire sample includes individuals ages 25-60 who were labor force participants in both years for each set of two consecutive years. The "regular" sample further restricts the entire sample to include individuals who have validated social security numbers. H_1: Means are equal for the two samples: ** reject at 1% level, * reject at 5% level.

	"Regular" Sample (231823)		Entire Sample (279586)		
Variable	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Test of H_1**
Male	0.52	0.50	0.52	0.50	
Black	0.11	0.31	0.11	0.32	H_1:
Hispanic	0.09	0.29	0.10	0.30	H_1:
Age (3 categories)	1.82	0.76	1.81	0.77	H_1:
Education (3 categories)	2.17	0.61	2.15	0.60	H_1:
Marital status	1.90	1.23	1.95	1.26	H_1:
Reported job-limiting disability	0.07	0.26	0.07	0.26	H_1:
Number of children	0.97	1.20	0.95	1.20	H_1:
Received welfare payments	0.09	0.28	0.09	0.28	H_1:
Received disability payments	0.02	0.15	0.02	0.14	H_1:
Total net worth	100459.04	246704.29	98753.90	327150.28	H_1:
Homeowner	0.67	0.47	0.64	0.48	H_1:
Born in country other than U.S.	0.11	0.31	0.13	0.33	H_1:
Had a defined contribution pension plan	0.29	0.46	0.29	0.45	H_1:
Had a defined benefit pension plan	0.39	0.49	0.38	0.49	H_1:
Had health insurance coverage	0.87	0.34	0.86	0.35	H_1:
Total annual SIPP reported real earnings	25751.47	20594.36	25227.26	20345.50	H_1:

Notes: The total sample size of 279586 corresponds to the set of individuals ages 25 to 60 who were labor force participants in both years for each set of two consecutive years. All calculations are weighted to reflect the corresponding Decennial Census population on April 1st, 2000. All calculations are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data.

Table 3: Unconditional and Conditional Earnings Mobility Regressions: Regular Sample

UNCONDITIONAL

One-year real earnings changes, stacked years from 1990-1999 Number of Observations = 231823			Three-year real earnings changes from 1996-1999 Number of Observations = 33401		
Functional Form	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**	**Functional Form**	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**
linear	SIPP	convergence, 1/1	linear	SIPP	convergence, 1/1
step function	SIPP	convergence, 4/4	step function	SIPP	convergence, 4/4
non-parametric	SIPP		non-parametric	SIPP	
linear	DER	convergence, 1/1	linear	DER	convergence, 1/1
step function	DER	convergence, 4/4	step function	DER	convergence, 4/4
non-parametric	DER		non-parametric	DER	

CONDITIONAL

One-year real earnings changes, stacked years from 1990-1999 Number of Observations = 231823			Three-year real earnings changes from 1996-1999 Number of Observations = 33401		
Functional Form	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**	**Functional Form**	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**
linear	SIPP	convergence, 1/1	linear	SIPP	convergence, 1/1
step function	SIPP	convergence, 4/4	step function	SIPP	convergence, 4/4
non-parametric	SIPP		non-parametric	SIPP	
linear	DER	convergence, 1/1	linear	DER	convergence, 1/1
step function	DER	convergence, 4/4	step function	DER	convergence, 4/4
non-parametric	DER		non-parametric	DER	

Notes: All results are significant at the 99% confidence level. The samples include all individuals ages 25 to 60 with validated SSNs who were labor force participants for all relevant consecutive years (two consecutive years for one-year mobility regressions and four consecutive years for three-year mobility regressions). All one-year regressions are weighted to reflect the corresponding Decennial Census population on April 1st, 2000. All three-year regressions are weighted using the 1996 SIPP person weight. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real dollar earnings in January 1995. The conditional mobility regressions include race, gender, age, education, and experience as additional right hand side variables.

Table 3: Unconditional and Conditional Earnings Mobility Regressions: Positive Earners Sample

UNCONDITIONAL

One-year real earnings changes, stacked years from 1990-1999 Number of Observations = 192054			Three-year real earnings changes from 1996-1999 Number of Observations = 26211		
Functional Form	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**	**Functional Form**	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**
linear	SIPP	convergence, 1/1	linear	SIPP	convergence, 1/1
step function	SIPP	convergence, 4/4	step function	SIPP	convergence, 4/4
non-parametric	SIPP		non-parametric	SIPP	
linear	DER	convergence, 1/1	linear	DER	convergence, 1/1
step function	DER	convergence, 4/4	step function	DER	convergence, 4/4
non-parametric	DER		non-parametric	DER	

CONDITIONAL

One-year real earnings changes, stacked years from 1990-1999 Number of Observations = 192054			Three-year real earnings changes from 1996-1999 Number of Observations = 26211		
Functional Form	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**	**Functional Form**	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**
linear	SIPP	convergence, 1/1	linear	SIPP	convergence, 1/1
step function	SIPP	convergence, 4/4	step function	SIPP	convergence, 4/4
non-parametric	SIPP		non-parametric	SIPP	
linear	DER	convergence, 1/1	linear	DER	convergence, 1/1
step function	DER	convergence, 4/4	step function	DER	convergence, 4/4
non-parametric	DER		non-parametric	DER	

Notes: All results are significant at the 99% confidence level. The samples include all individuals ages 25 to 60 with validated SSNs who had positive earnings for all relevant consecutive years (two consecutive years for one-year mobility regressions and four consecutive years for three-year mobility regressions). All one-year regressions are weighted to reflect the corresponding Decennial Census population on April 1st, 2000. All three-year regressions are weighted using the 1996 SIPP person weight. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real dollar earnings in January 1995. The conditional mobility regressions include race, gender, age, education, and experience as additional right hand side variables.

Table 3: Unconditional and Conditional Earnings Mobility Regressions: Non-Imputed Earners Sample

UNCONDITIONAL

One-year real earnings changes, stacked years from 1990-1999 Number of Observations = 99102			Three-year real earnings changes from 1996-1999 Number of Observations = 3695		
Functional Form	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**	**Functional Form**	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**
linear	SIPP	convergence, 1/1	linear	SIPP	convergence, 1/1
step function	SIPP	convergence, 4/4	step function	SIPP	convergence, 4/4
non-parametric	SIPP		non-parametric	SIPP	
linear	DER	convergence, 1/1	linear	DER	convergence, 1/1
step function	DER	convergence, 4/4	step function	DER	convergence, 4/4
non-parametric	DER		non-parametric	DER	

CONDITIONAL

One-year real earnings changes, stacked years from 1990-1999 Number of Observations = 99102			Three-year real earnings changes from 1996-1999 Number of Observations = 3695		
Functional Form	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**	**Functional Form**	**Earnings measure**	**Result, #Significant Coefficients / #Total Coefficients**
linear	SIPP	convergence, 1/1	linear	SIPP	convergence, 1/1
step function	SIPP	convergence, 4/4	step function	SIPP	convergence, 4/4
non-parametric	SIPP		non-parametric	SIPP	
linear	DER	convergence, 1/1	linear	DER	convergence, 1/1
step function	DER	convergence, 4/4	step function	DER	convergence, 4/4
non-parametric	DER		non-parametric	DER	

Notes: All results are significant at the 99% confidence level. The samples include all individuals ages 25 to 60 with validated SSNs who had non-imputed, positive earnings for all relevant consecutive years (two consecutive years for one-year mobility regressions and four consecutive years for three-year mobility regressions). All one-year regressions are weighted to reflect the corresponding Decennial Census population on April 1st, 2000. All three-year regressions are weighted using the 1996 SIPP person weight. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real dollar earnings in January 1995. The conditional mobility regressions include race, gender, age, education, and experience as additional right hand side variables.

Table 4: Means and Standard Deviations: Positive Earners

Number of observations = 87699

	Weighted		Unweighted	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Male	0.53	0.50	0.52	0.50
Black	0.11	0.31	0.11	0.31
High school	0.62	0.49	0.63	0.48
College	0.29	0.46	0.28	0.39
Ages 49-60	0.18	0.39	0.19	0.45
Ages 37-48	0.40	0.49	0.40	0.49
Ages 25-36	0.41	0.49	0.42	0.49
Initial year is 1993	0.14	0.35	0.18	0.38
Initial year is 1996	0.36	0.48	0.32	0.47
Initial year is 1997	0.37	0.48	0.33	0.47
Experience in year0	18.75	8.71	18.56	8.71
Experience in year1	19.70	8.75	19.51	8.75
Experience in year2	20.66	8.79	20.47	8.79
Change in experience from year1 to year2	0.96	0.21	0.96	0.21
Change in experience from year2 to year3	0.96	0.21	0.96	0.21
SIPP log earnings in year2	10.01	0.79	9.98	0.80
DER log earnings in year2	10.03	0.89	9.99	0.91
SIPP log earnings in year3	10.06	0.78	10.03	0.79
DER log earnings in year3	10.03	0.95	9.99	0.97
SIPP log earnings change from year1 to year2	0.002	0.55	0.01	0.58
DER log earnings change from year1 to year2	0.07	0.55	0.07	0.58
SIPP log earnings change from year2 to year3	0.05	0.51	0.05	0.53
DER log earnings change from year2 to year3	0.01	0.57	0.01	0.59

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995.

Table 5: Means and Standard Deviations: Non-Imputed Earners

Number of observations = 26366

	Weighted		Unweighted	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Male	0.52	0.50	0.51	0.50
Black	0.09	0.28	0.09	0.29
High school	0.61	0.49	0.61	0.49
College	0.32	0.46	0.31	0.46
Ages 49-60	0.20	0.40	0.21	0.41
Ages 37-48	0.41	0.49	0.42	0.49
Ages 25-36	0.38	0.49	0.38	0.48
Initial year is 1993	0.08	0.26	0.11	0.31
Initial year is 1996	0.35	0.48	0.32	0.47
Initial year is 1997	0.50	0.50	0.47	0.50
Experience in year0	19.31	8.67	19.29	8.67
Experience in year1	20.26	8.71	20.25	8.71
Experience in year2	21.22	8.76	21.20	8.76
Change in experience from year1 to year2	0.95	0.21	0.95	0.21
Change in experience from year2 to year3	0.95	0.21	0.95	0.21
SIPP log earnings in year2	10.06	0.77	10.04	0.78
DER log earnings in year2	10.09	0.86	10.07	0.86
SIPP log earnings in year3	10.08	0.78	10.06	0.79
DER log earnings in year3	10.11	0.88	10.09	0.88
SIPP log earnings change from year1 to year2	0.06	0.49	0.06	0.49
DER log earnings change from year1 to year2	0.07	0.49	0.07	0.50
SIPP log earnings change from year2 to year3	0.02	0.48	0.02	0.48
DER log earnings change from year2 to year3	0.02	0.49	0.02	0.49

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had positive, non-imputed earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995.

Table 6: Using DER to instrument for SIPP: Positive Earners, Unweighted

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
Dependent variables	OLS		OLS		IV		GMM		OLS		IV		GMM	
SIPP log earnings in year3			X		X		X							
SIPP log earnings change from year2 to year3	X								X		X		X	
Independent variables	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
Intercept	2.74**	0.04	2.74**	0.04	1.67**	0.05	1.29**	0.07						
Male	0.11**	0.02	0.11**	0.02	0.06*	0.02	0.04	0.02						
Black	-0.02	0.04	-0.02	0.04	-0.01	0.04	0.00	0.04						
High school	0.09**	0.01	0.09**	0.01	0.04**	0.01	0.03	0.02						
College	0.24**	0.02	0.24**	0.02	0.13**	0.02	0.10*	0.02						
Ages 49-60	-0.06**	0.01	-0.06**	0.01	-0.05**	0.01	-0.04**	0.01						
Ages 37-48	-0.01*	0.01	-0.01*	0.01	-0.01*	0.01	-0.02**	0.01						
Experience in year2	0.00**	0.00	0.00**	0.00	0.00**	0.00	0.00	0.00						
SIPP log earnings in year2	**-0.29****	0.00	**0.71**	0.00										
Predicted SIPP log earnings in year2					**0.83****	0.00	**0.87****	0.01						
Change in experience from year1 to year2									0.06**	0.01	0.06**	0.01	0.06**	0.01
SIPP log earnings change from year1 to year2									**-0.23****	0.01				
Predicted SIPP log earnings change from yr1 to yr2											**-0.13****	0.02	**0.00**	0.03
R-squared	0.15		0.61		0.44		0.59		0.07		0.01		0.01	
N	87669		87669		87669		87669		87669		87669		87669	
1st stage:														
SIPP log earnings in year2					X									
SIPP log earnings change from year1 to year2											X			
					Est.	**S.E.**					**Est.**	**S.E.**		
Intercept					4.11**	0.02								
Male					0.18**	0.01								
Black					-0.06**	0.02								
High school					0.16**	0.01								
College					0.39**	0.01								
Ages 49-60					-0.04**	0.01								
Ages 37-48					0.00	0.01								
Experience in year2					0.01**	0.00								
DER log earnings in year2					**0.55****	0.00								
Change in experience from year1 to year2											-0.01	0.01		
DER log earnings change from year1 to year2											**0.30****	0.01		
T-stat on IV					209.60						56.48			
R-squared					0.54						0.08			
N					87669						87669			
Corresponding Beta Term	-0.29		-0.29		-0.17		-0.13		-1.23		-1.13		-1.00	

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. All the regressions specified in earnings levels (rather than changes) include dummy variables indicating which set of three consecutive years an individual's earnings came from. * indicates significance at 5%; ** indicates significance at 1%.

Table 7: Using SIPP to instrument for DER: Positive Earners, Unweighted

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
Dependent variables	OLS		OLS		IV		GMM		OLS		IV		GMM	
DER log earnings in year3			X		X		X							
DER log earnings change from year2 to year3	X								X		X		X	
Independent variables	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
Intercept	1.65**	0.03	1.65**	0.03	0.79**	0.05	0.44**	0.06						
Male	0.09**	0.00	0.09**	0.00	0.05**	0.01	0.03**	0.00						
Black	-0.02**	0.01	-0.02**	0.01	-0.01	0.01	-0.01	0.01						
High school	0.08**	0.01	0.08**	0.01	0.04**	0.01	0.03**	0.01						
College	0.18**	0.01	0.18**	0.01	0.10**	0.01	0.07**	0.01						
Ages 49-60	-0.05**	0.01	-0.05**	0.01	-0.04**	0.01	-0.03**	0.01						
Ages 37-48	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01						
Experience in year2	0.00**	0.00	0.00**	0.00	0.00	0.00	0.00**	0.00						
DER log earnings in year2	**-0.18****	0.00	**0.82****	0.00										
Predicted DER log earnings in year2					**0.91****	0.01	**0.95****	0.01						
Change in experience from year1 to year2									0.02**	0.00	0.01**	0.00	0.00	0.00
DER log earnings change from year1 to year2									**-0.19****	0.00				
Predicted DER log earnings change from yr1 to yr2											**-0.07***	0.02	**0.08***	0.03
R-squared	0.07		0.65		0.41		0.64		0.04		0.0007		-0.02	
N	87699		87699		87699		87699		87699		87699		87699	
1st stage:														
DER log earnings in year2					X									
DER log earnings change from year1 to year2											X			
					Est.	**S.E.**					**Est.**	**S.E.**		
Intercept					2.12**	0.16								
Male					0.10**	0.01								
Black					-0.01	0.01								
High school					0.11**	0.02								
College					0.23**	0.02								
Ages 49-60					-0.02*	0.01								
Ages 37-48					0.02*	0.01								
Experience in year2					0.01**	0.00								
SIPP log earnings in year2					**0.75****	0.02								
Change in experience from year1 to year2											0.07**	0.00		
SIPP log earnings change from year1 to year2											**0.29****	0.01		
T-stat on IV					43.53						24.00			
R-squared					0.52						0.10			
N					87669						87669			
Corresponding Beta Term	-0.18		-0.18		-0.09		-0.05		-1.19		-1.07		-0.92	

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. All the regressions specified in earnings levels (rather than changes) include dummy variables indicating which set of three consecutive years an individual's earnings came from. * indicates significance at 5%; ** indicates significance at 1%.

Table 8: Using DER to instrument for SIPP: Positive Earners, Weighted

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
Dependent variables	OLS		OLS		IV		GMM		OLS		IV		GMM	
SIPP log earnings in year3			X		X		X							
SIPP log earnings change from year2 to year3	X								X		X		X	
Independent variables	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
Intercept	2.64**	0.05	2.64**	0.05	1.64**	0.04	1.29**	0.07						
Male	0.11**	0.02	0.11**	0.02	0.06*	0.02	0.04	0.02						
Black	-0.02	0.03	-0.02	0.03	-0.01	0.03	-0.01	0.04						
High school	0.09**	0.01	0.09**	0.01	0.05**	0.01	0.03*	0.01						
College	0.23**	0.02	0.23**	0.02	0.13**	0.02	0.10**	0.02						
Ages 49-60	-0.06**	0.01	-0.06**	0.01	-0.05**	0.01	-0.04**	0.01						
Ages 37-48	-0.01**	0.01	-0.01**	0.01	-0.01**	0.01	-0.02**	0.01						
Experience in year2	0.00**	0.00	0.00**	0.00	0.00**	0.00	0.00*	0.00						
SIPP log earnings in year2	**-0.28****	0.00	**0.72****	0.00										
Predicted SIPP log earnings in year2					**0.83****	0.00	**0.87****	0.01						
Change in experience from year1 to year2									0.05**	0.01	0.05**	0.01	0.06**	0.01
SIPP log earnings change from year1 to year2									**-0.23****	0.01				
Predicted SIPP log earnings change from yr1 to yr2											**-0.12****	0.02	**0.01**	0.03
R-squared	0.14		0.63		0.58		0.61		0.07		0.02		0.004	
N	79579		79579		79579		79579		79579		79579		79579	
1st stage:														
SIPP log earnings in year2					X									
SIPP log earnings change from year1 to year2											X			
					Est.	**S.E.**					**Est.**	**S.E.**		
Intercept					3.80**	0.03								
Male					0.17**	0.01								
Black					-0.06**	0.01								
High school					0.15**	0.01								
College					0.36**	0.01								
Ages 49-60					-0.03**	0.01								
Ages 37-48					0.01	0.01								
Experience in year2					0.00**	0.00								
DER log earnings in year2					**0.58****	0.00								
Change in experience from year1 to year2											-0.02*	0.01		
DER log earnings change from year1 to year2											**0.33****	0.00		
T-stat on IV					232.09						73.65			
R-squared					0.47						0.11			
N					79579						79579			
Corresponding Beta Term	-0.28		-0.28		-0.17		-0.13		-1.23		-1.12		-0.99	

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. All the regressions specified in earnings levels (rather than changes) include dummy variables indicating which set of three consecutive years an individual's earnings came from. * indicates significance at 5%; ** indicates significance at 1%.

Table 9: Using SIPP to instrument for DER: Positive Earners, Weighted

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
Dependent variables	OLS		OLS		IV		GMM		OLS		IV		GMM	
DER log earnings in year3			X		X		X							
DER log earnings change from year2 to year3	X								X		X		X	
Independent variables	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
Intercept	1.57**	0.03	1.57**	0.03	0.81**	0.05	0.47**	0.06						
Male	0.09**	0.00	0.09**	0.00	0.05**	0.01	0.03**	0.01						
Black	-0.02*	0.01	-0.02*	0.01	-0.01	0.01	-0.01	0.01						
High school	0.08**	0.01	0.08**	0.01	0.04**	0.01	0.03**	0.01						
College	0.17**	0.01	0.17**	0.01	0.10**	0.01	0.08**	0.01						
Ages 49-60	-0.05**	0.01	-0.05**	0.01	-0.04**	0.01	-0.04**	0.01						
Ages 37-48	0.00	0.01	0.00	0.01	0.00	0.01	-0.01	0.01						
Experience in year2	0.00**	0.00	0.00**	0.00	0.00	0.00	0.00**	0.00						
DER log earnings in year2	**-0.18****	0.00	**0.82****	0.00										
Predicted DER log earnings in year2					**0.91****	0.01	**0.95****	0.01						
Change in experience from year1 to year2									0.02**	0.00	0.01**	0.00	0.00	0.00
DER log earnings change from year1 to year2									**-0.19****	0.00				
Predicted DER log earnings change from yr1 to yr2											**-0.07****	0.02	**0.07****	0.02
R-squared	0.07		0.66		0.44		0.65		0.03		0.0007		-0.02	
N	79579		79579		79579		79579		79579		79579		79579	
1st stage:														
DER log earnings in year2					X									
DER log earnings change from year1 to year2											X			
					Est.	**S.E.**					**Est.**	**S.E.**		
Intercept					1.87**	0.12								
Male					0.10**	0.01								
Black					-0.01	0.01								
High school					0.09**	0.01								
College					0.19**	0.02								
Ages 49-60					-0.03**	0.01								
Ages 37-48					0.01	0.01								
Experience in year2					0.01**	0.00								
SIPP log earnings in year2					**0.78****	0.01								
Change in experience from year1 to year2											0.07**	0.00		
SIPP log earnings change from year1 to year2											**0.33****	0.01		
T-stat on IV					61.19						27.92			
R-squared					0.56						0.12			
N					79579						79579			
Corresponding Beta Term	-0.18		-0.18		-0.09		-0.05		-1.19		-1.07		-0.93	

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. All the regressions specified in earnings levels (rather than changes) include dummy variables indicating which set of three consecutive years an individual's earnings came from. * indicates significance at 5%; ** indicates significance at 1%.

Table 10: Using DER to instrument for SIPP: Non-Imputed Earners, Unweighted

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
Dependent variables	OLS		OLS		IV		GMM		OLS		IV		GMM	
SIPP log earnings in year3			X		X		X							
SIPP log earnings change from year2 to year3	X								X		X		X	
Independent variables	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
Intercept	2.12**	0.04	2.12**	0.04	1.23**	0.06	0.87**	0.09						
Male	0.10**	0.01	0.10**	0.01	0.06**	0.01	0.02**	0.01						
Black	-0.01	0.01	-0.01	0.01	0.00	0.01	0.00	0.01						
High school	0.08**	0.01	0.08**	0.01	0.04**	0.01	0.03**	0.01						
College	0.19**	0.01	0.19**	0.01	0.11**	0.01	0.08**	0.02						
Ages 49-60	-0.06*	0.01	-0.06*	0.01	-0.05**	0.01	-0.04**	0.01						
Ages 37-48	-0.01	0.01	-0.01	0.01	-0.01	0.01	-0.02*	0.01						
Experience in year2	0.00**	0.00	0.00**	0.00	0.00	0.00	0.00	0.00						
SIPP log earnings in year2	**-0.23****	0.00	**0.77****	0.00										
Predicted SIPP log earnings in year2					**0.87****	0.01	**0.91****	0.01						
Change in experience from year1 to year2									0.03**	0.00	0.02**	0.00	0.02**	0.00
SIPP log earnings change from year1 to year2									**-0.22****	0.01				
Predicted SIPP log earnings change from yr1 to yr2											**-0.07****	0.01	**0.11****	0.02
R-squared	0.11		0.66		0.55		0.65		0.05		0.003		-0.06	
N	26366		26366		26366		26366		26366		26366		26366	
1st stage:														
SIPP log earnings in year2					X									
SIPP log earnings change from year1 to year2											X			
					Est.	**S.E.**					**Est.**	**S.E.**		
Intercept					2.87**	0.04								
Male					0.14**	0.01								
Black					-0.03**	0.01								
High school					0.15**	0.01								
College					0.31**	0.01								
Ages 49-60					0.00	0.01								
Ages 37-48					0.01	0.01								
Experience in year2					0.00**	0.00								
DER log earnings in year2					**0.68****	0.00								
Change in experience from year1 to year2											0.02**	0.00		
DER log earnings change from year1 to year2											**0.52****	0.01		
T-stat on IV					141.67						100.74			
R-squared					0.69						0.29			
N					26366						26366			
Corresponding Beta Term	-0.23		-0.23		-0.13		-0.09		-1.22		-1.07		-0.89	

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had non-imputed, positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. All the regressions specified in earnings levels (rather than changes) include dummy variables indicating which set of three consecutive years an individual's earnings came from. * indicates significance at 5%; ** indicates significance at 1%.

Table 11: Using SIPP to instrument for DER: Non-Imputed Earners, Unweighted

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
Dependent variables	OLS		OLS		IV		GMM		OLS		IV		GMM	
DER log earnings in year3			X		X		X							
DER log earnings change from year2 to year3	X								X		X		X	
Independent variables	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
Intercept	1.55**	0.04	1.55**	0.04	1.09**	0.06	0.62**	0.08						
Male	0.08**	0.01	0.08**	0.01	0.06**	0.01	0.02**	0.01						
Black	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01						
High school	0.06**	0.01	0.06**	0.01	0.04**	0.01	0.03**	0.01						
College	0.15**	0.01	0.15**	0.01	0.10**	0.02	0.08**	0.01						
Ages 49-60	-0.05**	0.01	-0.05**	0.01	-0.04**	0.01	-0.04**	0.01						
Ages 37-48	0.00	0.01	0.00	0.01	-0.01	0.01	-0.02*	0.01						
Experience in year2	0.00**	0.00	0.00**	0.00	0.00	0.00	0.00	0.00						
DER log earnings in year2	**-0.17****	0.00	**0.83****	0.00										
Predicted DER log earnings in year2					**0.88****	0.01	**0.94****	0.01						
Change in experience from year1 to year2									0.03**	0.00	0.02**	0.00	0.02**	0.00
DER log earnings change from year1 to year2									**-0.18****	0.01				
Predicted DER log earnings change from yr1 to yr2											**-0.08****	0.01	**0.12****	0.02
R-squared	0.07		0.71		0.54		0.70		0.04		0.003		-0.06	
N	26366		26366		26366		26366		26366		26366		26366	
1st stage:														
DER log earnings in year2					X									
DER log earnings change from year1 to year2											X			
					Est.	**S.E.**					**Est.**	**S.E.**		
Intercept					1.15**	0.05								
Male					0.06**	0.01								
Black					0.00	0.01								
High school					0.01	0.01								
College					0.06**	0.01								
Ages 49-60					-0.04**	0.01								
Ages 37-48					0.00	0.01								
Experience in year2					0.01**	0.00								
SIPP log earnings in year2					**0.87****	0.00								
Change in experience from year1 to year2											0.04**	0.01		
SIPP log earnings change from year1 to year2											**0.54****	0.01		
T-stat on IV					185.20						100.67			
R-squared					0.67						0.30			
N					26366						26366			
Corresponding Beta Term	-0.17		-0.17		-0.12		-0.06		-1.18		-1.08		-0.88	

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had non-imputed, positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. All the regressions specified in earnings levels (rather than changes) include dummy variables indicating which set of three consecutive years an individual's earnings came from. * indicates significance at 5%; ** indicates significance at 1%.

Table 12: Using DER to instrument for SIPP: Non-Imputed Earners, Weighted

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
Dependent variables	OLS		OLS		IV		GMM		OLS		IV		GMM	
SIPP log earnings in year3			X		X		X							
SIPP log earnings change from year2 to year3	X								X		X		X	
Independent variables	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
Intercept	2.05**	0.04	2.05**	0.04	1.16**	0.06	0.82**	0.10						
Male	0.10**	0.01	0.10**	0.01	0.06**	0.01	0.02**	0.01						
Black	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.01						
High school	0.08**	0.01	0.08**	0.01	0.05**	0.01	0.03**	0.02						
College	0.19**	0.01	0.19**	0.01	0.10**	0.01	0.08**	0.01						
Ages 49-60	-0.06**	0.01	-0.06**	0.01	-0.05**	0.01	-0.04**	0.01						
Ages 37-48	-0.01	0.01	-0.01	0.01	-0.01	0.02	-0.02**	0.01						
Experience in year2	0.00**	0.00	0.00**	0.00	0.00	0.00	0.00	0.00						
SIPP log earnings in year2	**-0.22****	0.00	**0.78****	0.00										
Predicted SIPP log earnings in year2					**0.88****	0.01	**0.92****	0.01						
Change in experience from year1 to year2									0.03**	0.00	0.02**	0.00	0.02**	0.00
SIPP log earnings change from year1 to year2									**-0.21****	0.01				
Predicted SIPP log earnings change from yr1 to yr2											**-0.06****	0.01	**0.13****	0.03
R-squared	0.10		0.67		0.55		0.65		0.05		0.003		-0.08	
N	26189		26189		26189		26189		26189		26189		26189	
1st stage:														
SIPP log earnings in year2					X									
SIPP log earnings change from year1 to year2											X			
					Est.	**S.E.**					**Est.**	**S.E.**		
Intercept					2.88**	0.05								
Male					0.14**	0.01								
Black					-0.03**	0.01								
High school					0.16**	0.01								
College					0.31**	0.01								
Ages 49-60					0.00	0.01								
Ages 37-48					0.01*	0.01								
Experience in year2					0.00**	0.00								
DER log earnings in year2					**0.68****	0.01								
Change in experience from year1 to year2											0.02**	0.00		
DER log earnings change from year1 to year2											**0.52****	0.01		
T-stat on IV					123.28						100.16			
R-squared					0.68						0.29			
N					26189						26189			
Corresponding Beta Term	-0.22		-0.22		-0.12		-0.08		-1.21		-1.06		-0.87	

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had non-imputed, positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. All the regressions specified in earnings levels (rather than changes) include dummy variables indicating which set of three consecutive years an individual's earnings came from. * indicates significance at 5%; ** indicates significance at 1%.

Table 13: Using SIPP to instrument for DER: Non-Imputed Earners, Weighted

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
Dependent variables	OLS		OLS		IV		GMM		OLS		IV		GMM	
DER log earnings in year3			X		X		X							
DER log earnings change from year2 to year3	X								X		X		X	
Independent variables	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
Intercept	1.51**	0.04	1.51**	0.04	1.03**	0.06	0.59**	0.08						
Male	0.08**	0.01	0.08**	0.01	0.06**	0.01	0.03**	0.01						
Black	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.01						
High school	0.06**	0.01	0.06**	0.01	0.04**	0.01	0.04**	0.01						
College	0.15**	0.01	0.15**	0.01	0.10**	0.01	0.09**	0.01						
Ages 49-60	-0.05**	0.01	-0.05**	0.01	-0.05**	0.01	-0.04**	0.01						
Ages 37-48	-0.01	0.01	-0.01	0.01	-0.01	0.01	-0.02**	0.01						
Experience in year2	0.00*	0.00	0.00*	0.00	0.00	0.00	0.00*	0.00						
DER log earnings in year2	**-0.16****	0.00	**0.84****	0.00										
Predicted DER log earnings in year2					**0.89****	0.01	**0.94****	0.01						
Change in experience from year1 to year2									0.03**	0.00	0.02**	0.00	0.02**	0.00
DER log earnings change from year1 to year2									**-0.19****	0.01				
Predicted DER log earnings change from yr1 to yr2											**-0.07****	0.01	**0.11****	0.03
R-squared	0.07		0.71		0.53		0.70		0.04		0.003		-0.05	
N	26189		26189		26189		26189		26189		26189		26189	
1st stage:														
DER log earnings in year2					X									
DER log earnings change from year1 to year2											X			
					Est.	**S.E.**					**Est.**	**S.E.**		
Intercept					1.18**	0.05								
Male					0.06**	0.01								
Black					0.00	0.01								
High school					0.01	0.01								
College					0.06**	0.01								
Ages 49-60					-0.04**	0.01								
Ages 37-48					0.00	0.01								
Experience in year2					0.01**	0.00								
SIPP log earnings in year2					**0.87****	0.00								
Change in experience from year1 to year2											0.04**	0.00		
SIPP log earnings change from year1 to year2											**0.54****	0.01		
T-stat on IV					180.53						99.86			
R-squared					0.67						0.29			
N					26189						26189			
Corresponding Beta Term	-0.16		-0.16		-0.11		-0.06		-1.19		-1.07		-0.89	

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had non-imputed, positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. All the regressions specified in earnings levels (rather than changes) include dummy variables indicating which set of three consecutive years an individual's earnings came from. * indicates significance at 5%; ** indicates significance at 1%.

9 Appendix

Derivation of equation (10) for the univariate case: From equation (8),

$$(67)\ \delta = \frac{Cov(\Delta y^*, y^*_{it-1})}{Var(y^*_{it-1})}.$$

From equation (9),

$$(68)\ \hat{\delta}_1 = \frac{Cov(\Delta y, y_{it-1})}{Var(y_{it-1})}$$

$$(69) = \frac{Cov(\lambda\Delta y^* + \Delta w, n_i + \lambda y^*_{it-1} + w_{it-1})}{Var(n_i + \lambda y^*_{it-1} + w_{it-1})} \text{ (plugging in from equations 4 and 5)}$$

$$(70) = \frac{Cov(\lambda\Delta y^*, \lambda y^*_{it-1})}{Var(\lambda y^*_{it-1}) + Var(w_{it-1})} = \frac{\lambda^2 Cov(\Delta y^*, y^*_{it-1})}{\lambda^2 Var(y^*_{it-1}) + Var(w_{it-1})} * \frac{Var(y^*_{it-1})}{Var(y^*_{it-1})}$$

$$(71) = \frac{Cov(\Delta y^*, y^*_{it-1})V(y^*_{it-1})}{[Var(y^*_{it-1})] * [Var(y^*_{it-1}) + (1/\lambda^2)Var(w_{it-1})]} = \frac{\delta Var(y^*_{it-1})}{Var(y^*_{it-1}) + (1/\lambda^2)Var(w_{it-1})}.$$

Table 14: Maximum Likelihood Estimation for Earnings Model in Equations (49)-(51)

		Positive Earners N=175398		Non-Imputed Earners N=52732	
Fixed Effects	**Term**	**Parameter**	**S.E.**	**Parameter**	**S.E.**
SIPP Intercept	α^{SIPP}	8.90**	0.01	8.88**	0.02
DER Intercept	α^{DER}	8.87**	0.01	8.91**	0.02
Male	β_0	0.41**	0.01	0.44**	0.01
Black	β_0	-0.11**	0.02	-0.07**	0.01
High school	β_0	0.37**	0.01	0.37**	0.02
College	β_0	0.87**	0.01	0.85**	0.02
Ages 37-48	β_0	-0.01	0.01	-0.02*	0.01
Ages 49-60	β_0	-0.13**	0.01	-0.16**	0.01
Initial year is 1993	β_0	0.00	0.01	0.00	0.02
Initial year is 1996	β_0	0.00	0.01	0.00	0.01
Initial year is 1997	β_0	-0.01	0.01	0.02	0.01
Experience	β_1	0.02**	0.00	0.02**	0.00
Random Effects					
Person effect	σ^2_θ	0.33		0.33	
Common error	σ^2_η	0.05		0.09	
Common error	ρ	**0.09**		**0.27**	
SIPP error	σ^2_w	0.13		0.06	
SIPP error	ρ_{SIPP}	**0.08**		**0.08**	
DER error	σ^2_v	0.35		0.19	
DER error	ρ_{DER}	**0.63**		**0.70**	
Corresponding Gamma Term - SIPP		**0.17**		**0.35**	
Corresponding Gamma Term - DER		**0.72**		**0.97**	
Corresponding Beta Term - SIPP		-0.83		-0.65	
Corresponding Beta Term - DER		-0.28		-0.03	

Notes: The sample includes all individuals ages 25 to 60 with validated SSNs who had either positive or non-imputed earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. * indicates significance at 5%; ** indicates significance at 1%.

Table 15: Estimation of Equations (66a) and (66b): Positive Earners

Dependent variable: Log earnings in year3

	SIPP		**DER**	
	Parameter	**S.E.**	**Parameter**	**S.E.**
Intercept	2.66**	0.07	1.57**	0.03
Male	0.10**	0.02	0.08**	0.01
Black	-0.02	0.04	-0.02*	0.01
High school	0.08**	0.01	0.07**	0.01
College	0.22**	0.02	0.16**	0.01
Ages 37-48	-0.07**	0.01	-0.06**	0.01
Ages 49-60	-0.02**	0.01	-0.01	0.01
Initial year is 1993	-0.03*	0.01	-0.01	0.01
Initial year is 1996	-0.07*	0.02	0.04**	0.01
Initial year is 1997	-0.03	0.03	0.02**	0.01
Experience in year2	-0.05**	0.01	-0.07**	0.01
Experience in year1	-0.05**	0.01	0.07**	0.01
Log earnings in year2	**0.72****	0.01	**0.83****	0.00
N	87699		87699	
Corresponding beta term	-0.28		-0.17	

Notes: Random person effects are included in both regressions. The sample includes all individuals ages 25 to 60 with validated SSNs who had positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. * indicates significance at 5%; ** indicates significance at 1%.

Table 16: Estimation of Equations (66a) and (66b): Non-Imputed Earners
Dependent variable: Log earnings in year3

	SIPP		DER	
	Parameter	**S.E.**	**Parameter**	**S.E.**
Intercept	2.08**	0.04	1.57**	0.01
Male	0.10**	0.01	0.08**	0.01
Black	-0.01	0.01	0.00	0.01
High school	0.08**	0.01	0.05**	0.01
College	0.18**	0.01	0.14**	0.01
Ages 37-48	-0.07**	0.01	-0.07**	0.01
Ages 49-60	-0.02*	0.01	-0.02*	0.01
Initial year is 1993	-0.01	0.01	0.01*	0.01
Initial year is 1996	0.03**	0.01	0.05**	0.01
Initial year is 1997	0.02**	0.01	0.03**	0.01
Experience in year2	-0.07**	0.01	-0.07**	0.01
Experience in year1	0.07**	0.01	0.07**	0.01
Log earnings in year2	**0.78****	0.00	**0.84****	0.00
N	26366		26366	
Corresponding beta term	-0.22		-0.16	

Notes: Random person effects are included in both regressions. The sample includes all individuals ages 25 to 60 with validated SSNs who had positive earnings for three consecutive years. All results are averaged across four completed datasets using Rubin's (1987) formulas for computing statistics from multiply imputed data. All earnings are expressed as real log earnings in January 1995. * indicates significance at 5%; ** indicates significance at 1%.