Robert M.
La Follette School of Public Affairs
at the University of Wisconsin-Madison

Working Paper Series
La Follette School Working Paper No. 2005-031
http://www.lafollette.wisc.edu/publications/workingpapers

Work and Earnings of Low-Skill Women: Do Survey Response and Administrative Records Provide Consistent Information?

Robert Haveman
La Follette School of Public Affairs and the Institute for Research on Poverty
at the University of Wisconsin-Madison

haveman@lafollette.wisc.edu

Geoffrey L. Wallace
La Follette School of Public Affairs and the Institute for Research on Poverty
at the University of Wisconsin-Madison

wallace@lafollette.wisc.edu

Robert M. La Follette School of Public Affairs
1225 Observatory Drive, Madison, Wisconsin 53706
Phone: 608.262.3581 / Fax: 608.265-3233
info@lafollette.wisc.edu / http://www.lafollette.wisc.edu

Work and Earnings of Low-Skill Women:
Do Survey Responses and Administrative Records Provide Consistent Information?

Robert Haveman
Institute for Research on Poverty
La Follette School of Public Affairs
Department of Economics
University of Wisconsin
haveman@lafollette.wisc.edu

Geoffrey L. Wallace
Institute for Research on Poverty
La Follette School of Public Affairs
Department of Economics
University of Wisconsin
wallace@lafollette.wisc.edu

November 2005

The authors would like to thank the La Follette School of Public Affairs and the Institute for Research on Poverty, both of the University of Wisconsin–Madison, for their support of this research. Research assistance by David Reznichek, Sangeun Lee, and Ben Winig is gratefully acknowledged, as are the helpful discussions at the PAWS workshop of the La Follette School, seminars at Dalhousie University, Australian National University, the University of Wisconsin, Syracuse University, Cornell University, and the University of Michigan.

Abstract

The employment and earnings effects of the state-oriented welfare reform legislation of 1996 have been extensively studied using either survey or administrative data. Because information may differ substantially between these sources, it is difficult both to identify the true effects of these interventions and to compare estimates among evaluations of these interventions that use these different data sources. This paper uses data gathered as part with the Child Support Demonstration Evaluation (CSDE) to examine the extent to which administrative (UI) and survey records on employment and earnings for a sample of low-skill women are congruent. Our findings suggest that there are substantial differences in both mean earnings and mean employment rates between survey and UI data. We identify the extent to which these disparities can be explained by differences between these data sources in the definition of earnings or the method of data collection. We also examine the differences between UI and survey sources in estimates of employment and earnings growth among low-skill women.

Work and Earnings of Low-Skill Women: Do Survey Responses and Administrative Records Provide Consistent Information?

I. INTRODUCTION

Since states began experimenting with welfare reform beginning in the early-1990s, it has become increasing difficult to use national data to study welfare-affected populations. At least part of this difficulty stems from the adoption by states of specific names to identify their TANF programs.[1] While in principle this need not pose a problem, in practice national surveys have been slow to change survey questionnaires to reflect the changing program nomenclature. For example, the Current Population Survey (CPS) makes no specific reference to state-specific TANF names, but rather asks respondents who indicate receiving public assistance whether they received AFDC/TANF or some other type of assistance.[2] Moreover, even if the national survey question did accurately measure welfare recipiency, it seems doubtful that current state level comparisons would be reliable given the dramatic decrease in caseloads since early-1994.[3]

Because of these difficulties, researchers have turned to state-level survey and administrative data to evaluate the impacts of the welfare reform and to monitor the post-reform outcomes for target populations. One common form of data are cross sectional or longitudinal surveys administered to a subset of a state's caseload, collecting information about earnings, employment, demographic characteristics and living arrangements. A second source of information is from state administrative data

---

[1]For example, Wisconsin has W-2 (Wisconsin Works), California has CALWORKS (California Work Opportunity and Responsibility to Kids) and Michigan has FIP (Family Investment Program). In all, 41 states have adopted state-specific names for their TANF programs.

[2]It was not until the 2001 CPS that the word TANF appeared in the questionnaire. The Survey of Income and Program Participation (SIPP) has also been slow in responding to the changing policy environment. It was only with the most recent SIPP panel (2001) that the survey began asking about state-specific TANF programs. The SIPP questionnaire now inquires whether respondents received AFDC, TANF or "insert state named program."

[3]National surveys were designed to provide accurate welfare population estimates for larger states prior to the national welfare reform. Since 1994, AFDC/TANF caseloads have decreased from over 5 million to about 2 million.

containing earnings information gathered as part of employer's reports to the Unemployment Insurance System (UI).

The potential existence of two sources of information on individual level earnings for individual states makes it difficult to compare results both within a state and across states. In this paper, we explore the extent of differences in individual employment and earnings for nearly 2,200 low-skill women in the state of Wisconsin between those observed in an extensive and careful survey and those reported by employers to the UI system. Both sources of data are available through a unique experimental research project undertaken at the Institute for Research on Poverty at the University of Wisconsin-Madison; the Child Support Demonstration Evaluation (CSDE).

In the CSDE project, single low-skill female resident parents in the state of Wisconsin who receive or have received welfare cash assistance are studied over time in an effort to assess their behavioral responses to a specific reform in child support policy.[4] The project surveyed both the female custodial parent (who was typically a welfare recipient at the start of the project) and the nonresident parent.[5] The project also obtained detailed information on the work and earnings of each covered person

[4] The study is designed as a social experiment in which 100 percent of the child support payments made by noncustodial parents are passed-through to the custodial parent.for the treatment group, with the control group subject to a partial pass-through (namely, a pass-through equal to the maximum of $50 per month or 50 percent of the monthly payment).

[5]The CSDE survey is a comprehensive one, inquiring about a variety of individual choices and living arrangements, in addition to socio-economic and demographic information. Information on the extent of work in a particular year (say 1998), and the earnings associated with that work is sought for each respondent. Uniquely, the survey also inquires about a detailed set of work-related attributes, such as the nature of the payments made (e.g., wages, tips, commissions, or the receipt of monetary payments from odd jobs), and the number of jobs held in a year. We use this information in analyzing the potential sources of differences in reports of work and earnings between the survey and the administrative UI data. The CSDE survey is unique in its efforts to secure reliable information on work and earnings responses. The special circumstances of low-skill women are reflected in explanations of the questions asked of survey respondents regarding the nature and extent of their work and earnings. For example, in asking about the number of jobs/employers, respondents were told to count each employer as one job, and that if employed by a temporary work agency and assigned to work at several different places, only one job should be indicated. In seeking information about earnings, it was explained that the question referred to "the total income you earned from all jobs combined during . . . [the year]." The respondent was explicitly told to exclude any money that was received from the public workforce/welfare agency, even though that payment required a specific amount of work. It was explained that money received in the form of salaries, tips, commissions, and as payment for jobs done on the side is to be included in the earnings response. Self-employment was explained, and respondents were told that income from this activity is also to be included. In cases where the respondent reported not knowing her income in a particular year, the interviewer was advised to "probe for the best estimate." In these cases a series of follow-up questions are asked involving the earnings interval in which earnings might fall, starting with high earnings intervals and working down.

included in the program from employer reports compiled by the Wisconsin Unemployment Insurance (UI) program.[6]

Our analysis proceeds as follows. In Section II-IV, we examine differences in the definitions of work and earnings between the survey and UI information sources, and in the data collection method. These differences suggest a number of reasons for discrepancies between work and earnings reports in the two data sources and some unknown 'true' value of earnings. Data from the two data sources reveal the extent of the discrepancies between them. In Section V, we use information available in the CSDE survey regarding the personal characteristics, location, welfare and work participation attributes of the workers in our sample to examine the correlates of the work and earnings discrepancies and the extent to which our conjectures regarding the sources of these discrepancies are able to explain the observed patterns. Finally, in Sections VI and VII, we explore the extent to which the use of survey or UI data affects empirical estimates of the determinants of employment and earnings, and estimates of the levels and changes in these variables across groups of workers.

II. SOURCES OF EARNINGS AND EMPLOYMENT DIFFERENCES IN SURVEY AND UI RECORDS

Relative to some unknown 'true' employment and earnings values there are reasons to suspect under and over-reporting in both survey data and UI reports.[7] UI earnings and employment may be underreported, reflecting potential incentives for both employees and employers to underreport earnings together with the difficulty in tracking some sources of income. For example, while the full amount of receipts of each employee's tips, bonuses, and commissions are required to appear in employer reports to the UI system, the incentives to underreport, combined with the difficulty of tracking income from these sources, make it likely that they are consistently underreported. Underreporting also exists because some

---

[6]These reports, which were compiled by the Wisconsin Unemployment Insurance (UI) program, indicate whether a person has worked during a quarter, and the quarterly earnings of the person.

[7] Appendix Table 1 summarizes the direction of the likely effect of these factors on the S and UI reports of employment (top bank) and earnings (bottom bank) relative to the unknown 'true' value.

employment categories are exempt from UI reporting requirements.[8] Workers may be falsely classified into these exempt categories, resulting in underreports of both earnings and employment in the UI data. Underreports in the UI data can also occur because the earnings of workers residing in one state and working in another, are unlikely to be reported by the employer to the UI system in the state of the employee's residence.[9] Finally, the UI reporting system may contain erroneous work and earnings information due to errors in recording Social Security numbers or in matching UI wage records. These errors may also reflect intentional or non-intentional noncompliance. Overall, the combined effect of these sources of potential bias suggest that UI employment and earnings measures are likely to be lower than ‘true’ earnings values.

Although most jobs are covered by the UI system,[10] both employment and wages for low-wage workers may be seriously underreported in UI reports. Relying on an extensive audit of a sample of 875 Illinois firms in 1987 (Q3), Blakemore et al. (1996) and Burgess et al. (1998) conclude that about 45 percent of employers failed to report earnings of some UI covered employees; 13.6 percent of their covered workers had no reports and 4.2 percent of wages were excluded. These underreports were concentrated among smaller firms.[11] The incorrect classification of some workers as uncovered independent contractors and high employee turnover accounted for much of the underreporting of work and earnings.[12]

---

[8]Categories of workers not covered by the UI reporting requirement include self-employed workers, independent contractors, farm laborers, domestic workers, military personnel, government workers, some part-time employees of nonprofit institutions, employees of religious orders, and some students employed by their schools. It is estimated that UI records cover about 91 percent of Wisconsin workers.

[9]Only the states of Missouri and Kansas have agreements to share information from UI reports.

[10] The Bureau of Labor Statistics (BLS) claims that 97.1 percent of jobs in 2001 were covered by the UI system.

[11]For firms with less than 5 workers, 56.5 percent of workers and 14.1 percent of earnings were unreported during the 3rd quarter of 1987.

[12]Nearly half of all unreported workers were improperly classified by their employers as independent contractors (Blakemore et al.). Because firms are responsible for paying UI taxes on employees up to an earnings threshold, those with high turnover must pay taxes on a larger portion of their total payroll; as a result, they are more likely to underreport workers and earnings. (Burgess et al.).

Individual survey responses regarding work and earnings may also have errors. Employment and earnings from illegal activities, irregular work, odd jobs or reciprocal tasks for friends, family and neighbors tend to be underreported in survey responses. To the extent that respondents view the survey as an instrument for obtaining information that may affect them adversely, survey information will understate the true level of employment and earnings.[13] Finally, error may arise from survey responses regarding work and earnings in the distant past or for periods of intermittent activity, and from the imputing of earnings values for workers who report that they do not know their earnings.[14]

Several studies have attempted to describe the extent of measurement problems in survey data by matching records from a survey (the Panel Study of Income Dynamics or March CPS) with 'true' earnings measures (Bound and Krueger 1991; Bound et al. 1994).[15] These studies indicate that there are substantial individual level differences between survey and ‘true’ earnings, but that this measurement error does not result in substantial biases in estimated coefficients from earnings regressions.

A number of studies have made direct comparisons between survey earnings measures and UI measures for low-skill populations. Using a sample of Job Training Partnership Act (JTPA) experiment participants that contained both UI and survey earnings, Kornfeld and Bloom (1999) found substantial differences in individual level and mean earnings. Twenty-six percent of adult men and nearly 15 percent of adult women had quarterly survey and UI earnings values that varied by more that $1,000; mean survey earnings were approximately 30 percent higher than mean UI earnings for both groups. Despite

---

[13]For example, welfare recipients may face implicit tax rates (of up to 100 percent) on earnings, and hence may be reluctant to report to a survey interviewer the existence or the extent of work and earnings. All of the women included in the sample were welfare recipients at some point during late 1997 or 1998, during which time Wisconsin had a 100 percent tax rate on the earnings of welfare recipients.

[14]In the CSDE survey, when individuals reported not knowing their earnings the interviewer was advised to “probe for the best estimate.” Then, individuals were asked to identify the $5000 dollar interval into which their earnings fell, starting with high intervals and working down. Because of this procedure, our assignment of the midpoint value of the interval as an estimate of their earnings is likely to lead to overstated survey earnings.

[15]In the context of these studies, the ‘true’ value of earnings is taken to be earnings from payroll records of a large unionized manufacturing firm (Bound et al.) or earnings from Social Security Administration (SSA) records (Bound and Krueger). We note that these ‘true’ earnings measures are themselves subject to error. For example, firm payroll records neglect earnings from second jobs, and SSA records exclude earnings from informal sector work.

large mean and individual level differences in survey and UI earnings, estimates of the impact of JTPA training were not substantially affected by which earnings measure was used.[16]

In addition to this study, there are a number of studies of women who exit state welfare programs conducted that relied on both survey and UI measures of employment and earnings.[17] As in the Kornfeld and Bloom study, survey-based employment rates and earnings exceed those from administrative data. One difficulty with many of these state-level studies is that the comparability survey and UI employment and earnings measures is questionable because the time frame covered by the surveys differ from those covered by the UI reports.[18] The CSDE survey that we analyze avoids this problem, as earnings are measured annually allowing comparability with UI records.

III. DISCREPANCIES IN EMPLOYMENT REPORTS

The most basic indicator of labor market performance is whether or not a person is employed during a specific period of time. For the 2,179 women in our sample, job-holding at any time during 1998 is recorded in both the survey and the UI data. For the UI data, we regard observations with positive UI earnings during any quarter of 1998 as working during that year. Table 1 reports a cross tabulation of survey and UI employment indicators for the 2,179 women in our sample. Eighteen percent have conflicting employment information from the two data sources. Eighty percent of these discrepancies are due to having UI, but no survey, reports of earnings. Because of these discrepancies, the survey and UI reports indicate quite different employment rates—83 percent using the UI data and 74 percent from the survey reports.

---

[16] Kornfeld and Bloom also review the findings of prior studies that have compared employment and earnings data from administrative records to those based on individual responses to survey questions. Such studies include Hotz and Scholz, 1999; Rodgers, Brown, and Duncan, 1993; Moore, Stinson, and Welniak, 1997; Baj, Trott, and Stevens, 1991; Baj, Fahey, and Trott, 1992; Burgess, Blakemore, and Low, 1998.

[17] Acs and Loprest (2001) review these studies; see also Issacs and Lyon (2000).

[18] For example, one study of welfare leavers compares quarterly UI employment and earnings (pre-exit to 4-months post-exit) to point-in-time survey records of employment and monthly earnings 12 to 18 months post-welfare exit. See Arizona Department of Income Security (2000).

It will be helpful for our further analysis to distinguish the groups in the various cells of Table 1. We label the 1,514 women in the first row/first column as *sure workers* because they are employed according to both data sources. Relying on the same rationale, we label the women in the second row/second column as *sure non-workers*. Because the women in the first row/second column report some earnings in the survey we classify them as *probable workers,* even though no employer report of earnings is recorded in the UI data. Because we know from employer reports that the 305 women in the second row/first column were working in 1998 in spite of their own reports of non-employment, we refer to them as *false non-workers*, and conclude that these women either forgot that they worked or misrepresented their earnings to survey interviewers.

IV. DISCREPANCIES IN EARNINGS REPORTS

Consistent with the disparities in alternative reports of employment, large disparities exist between earnings reported by CSDE sample respondents and earnings reported by employers in accordance with UI reporting requirements. Figure 1 presents a scatter-plot of the two earnings values for the entire sample of 2,179 women. The y-axis shows reports of earnings from the CSDE survey (S) and the x-axis employer reports of earnings actually paid (UI). The 272 *sure nonworkers* (zero earnings in both data sources) are concentrated at the origin of the figure. The 88 *probable workers* (zero UI earnings but positive S earnings) are shown along the x-axis, and the 305 *false nonworkers* (zero S earnings but positive UI earnings) are displayed along the x-axis. The 1,514 *sure workers* (those with positive earnings in both data sources) are shown in the interior of the figure. Were there no disparity between S and UI earnings, all of the observations would lie along the 45-degree line that divides the quadrant into two parts. Clearly such observations are a rare occurrence. While there is a substantial degree of nonconformity between S and UI earnings, there is a strong positive relationship between the series. The sample correlation between survey and UI earnings is 0.66 for the entire sample, and 0.65 among the *sure* workers.

Figure 2 provides another view of these disparities for the separate groups of women in our sample. Mean levels of S and UI, and the (S – UI) difference for each group are shown in the figure. *Sure workers* are shown in the positive quadrant of the figure, and we distinguish workers for whom the absolute value of the earnings difference exceeds $2500 from those for whom the difference lies within $2500 of the 45-degree line. The 88 *probable workers* are shown on the left side of the figure; they have positive S earnings but no UI earnings. Forty of these *probable workers* report S earnings of more than $2500, while having reported UI earnings of zero. Average S earnings for this group of 40 women are over $10,500. The 305 *false non-workers* are shown at the bottom of the diagram. There are 115 of these women who indicate no S earnings but for whom employers report average UI earnings of more than $2500. Employer-reported earnings for this group of *false non-workers* with UI earnings above $2500 average nearly $7,500.

In order to assess the degree of divergence between S and UI earnings we use two measures of the discrepancy between the two values–the mean absolute difference (MAD) and the mean squared difference (MSD), defined as the mean absolute difference and mean squared difference between S and UI earnings, respectively. MSD is also equal to the variance of the difference between S and UI earnings around zero.[19]

Table 2 reports average S and UI earnings by employment group, the fraction of the sample in each group, the two discrepancy indicators, and the percent of both the total absolute discrepancy ($\sum|S_i - UI_i|$) and the total squared discrepancy ($\sum(S_i - UI_i)^2$) attributable to each employment group. The top bank of Table 2 indicates significant variation in the extent of the earnings discrepancy across the groups of workers. For example, the MAD for *sure workers* is about $2900, compared to $3300 for *false non-workers,* and $5500 for those who report working but who have no employer reports of earnings (*probable workers*). For the average *sure worker* the mean value of S exceeds that of UI by nearly $1200,

[19]Like all measures of variance the MSD can be decomposed into systematic and random components, a property that we exploit below.

or by 18 percent. While *sure workers* comprise about 69 percent of all observations, they account for 75 percent of the total absolute discrepancy and 71 percent of the total squared discrepancy.

The bottom bank of Table 2 shows the distribution of MAD and MSD across five categories of *sure workers*. *Sure workers* with absolute earnings differences of less than $2,500 have a MAD of only $852. While they comprise over 66 percent of the sample of *sure workers,* they account for only 20 percent of the total absolute discrepancy among these workers, and but 3 percent of the total squared discrepancy. On the other hand, the 23 percent of *sure workers* for whom S exceeds UI by more than $2,500 have a MAD of over $7000 and account for 59 percent of the *sure worker* total absolute discrepancy, and for 73 percent of the *sure worker* total mean squared discrepancy. The 10 percent of *sure workers* for whom UI exceeds S by more than $2500 have a MAD of $6000, and also account for a disproportionate share of both the total absolute discrepancy and the total squared discrepancy.

The last two categories divide *sure workers* into steady and non-steady workers. The 509 steady workers are those who worked at least 3 quarters in 1998 and had no more than two employers according to both UI records and the survey. As the numbers in Table 2 indicate, steady workers have higher average earnings than non-steady *sure workers, probable workers, or false non-workers.* Despite their high levels of earnings, steady workers have lower levels of earnings discrepancy than other groups of workers. The MAD for steady workers is roughly $2,600 compared with $3,000 for the non-steady workers*.* The difference in the MSD between steady and non-steady workers is more striking. Steady workers have a MSD of $18,830 compared with $27,640 for other *sure workers*.

The distribution of the algebraic difference between S and UI earnings (S – UI) among all sample members is shown in Table 3.[20] Also shown is an approximation to the distribution of the earnings difference from a sample of women who were Job Training Partnership Act trainees reported by Kornfeld

[20] The distribution of the (S – UI) discrepancy for the subsample of *sure workers* is similar to that for the entire sample, although the subsample of *sure workers* has larger fraction of observations with survey earnings substantially in excess of UI earnings.

and Bloom (1999).[21] There is substantial conformity between our estimates of the the S – UI discrepancies and those of Kornfeld-Bloom. For both samples, 45–50 percent of the observations have an earnings discrepancy of less than $800, and about 30 percent of the observations report survey earnings that exceed UI earnings by more than $2400. However, while about 8 percent of the women in our sample have UI earnings that exceed survey earnings by more than $4000, only about 3 percent of the observations in the Kornfeld-Bloom sample have (S – UI) values greater than $4000.[22]

V. EVALUATING SOME CONJECTURES CONCERNING THE SOURCES OF EARNINGS DISCREPANCY

As we have noted, a variety of differences in concept, definition, and reporting procedures between the survey and the UI data system may contribute to the discrepancies between S and UI earnings reports. Other factors also contribute to the discrepancies, such as the likelihood of working in the informal sector, being an irregular (nonsteady) worker, or respondent reports of difficulty in recalling earnings information. In Table 4, we indicate several conjectures regarding the source and magnitude of the earnings discrepancy between the S and UI data.

The CSDE survey provides detailed information that would allow us to explore the impact of a number of these factors on the survey-UI earnings discrepancy. For example, in addition to providing

---

[21]Kornfeld and Bloom provide quarterly earnings estimates for their sample of women. We have ‘annualized’ their estimates by multiplying their quarterly earnings class intervals by four, hence forcing comparability with our annual values.

[22]One possible explanation for the thicker and longer bottom tail of the distribution of the earnings difference for the workers in our sample may be the creation of annual S earnings values for the Kornfeld-Bloom observations by multiplying quarterly values by a factor of 4. Because problems of recall are smaller in quarterly than in annual survey data, the distribution of quarterly earnings is likely to show less dispersion than the distribution of annual earnings for these same observations. Hence, the distribution of discrepancies between S and UI is likely to be smaller for the annualized Kornfeld-Bloom data than for the annual data on women in our sample. Moreover, the types of errors made in quarterly reporting are likely to be compounded in annual surveys, and reports of annual earnings reflect larger intertemporal employment variation than is present over a quarter. Second, the Kornfeld -Bloom numbers are from a sample of women that volunteered for training associated with the Job Training Partnership Act, whereas the sample used in our analysis is composed of women who were on welfare at some point during late 1997 or early 1998. All else equal, the workers in the Kornfeld-Bloom sample are likely to have a greater attachment to the formal sector of the labor force than the workers in our sample, and hence a smaller discrepancy between S and UI.

extensive information on demographic variables, the CSDE survey identifies the receipt of income from odd jobs, tips, and commissions. We can also identify those respondents who indicate that they ‘don’t know’ their earnings, and those for whom estimates of these ‘unknown’ values must be imputed. Information describing the welfare and work histories of the observations obtained from administrative data has been merged to the survey data. This information includes the number of months respondents received cash welfare assistance through AFDC in the 2-years prior to being assigned to Wisconsin’s TANF program, and the fraction of 1998 calendar year that respondents received cash assistance. Finally, the survey also contains information on the county of residence for each respondent, the number and characteristics of jobs/employers during the year, and the nature of job and payment arrangements.

A. Correlates of Being a False Non-Worker

We first estimate a multinomial logit model to identify factors that are related to whether survey respondents are false non-workers (those with positive UI earnings but no survey earnings) or probable workers. We use the 1,907 observations comprising these groups, plus sure workers (included to provide a comparison group). False non-workers are of particular interest as they apparently either forgot that they worked in 1998, or intentionally misreported their work status. Because many of these women have perceived incentives to hide their earnings, some of them may have intentionally misreported their earnings.[23]

Table 5 presents the results of this estimation. The coefficients show the relative risk (or odds) ratio associated with a unit change in each of the independent variables; coefficients greater (less) than one indicate a larger risk of falling into the indicated group relative to the reference group by a factor

[23] Average UI earnings for *false non-workers* is $3010.

equal to the coefficient. Because false non-workers and probable workers are lacking reports of either S or UI earnings or employment, the specification is parsimonious.[24]

Women with low education levels are more likely to be false non-workers than are those with more schooling. Although Hispanics (and other non-whites and non-blacks) are much less likely to be in the sure worker category than whites, they are as likely to be a probable worker as a false non-worker. Aside from this effect, race-ethnicity does not appear to effect worker group status. Residing in an urban area (Milwaukee or another urban area) also appears to be associated with being a false non-worker, however, neither of these variables are statistically significant at standard levels. While residing outside of Wisconsin for part of the year is seems likely to reduce employer reports of work and earnings (and, hence, increase the likelihood of being a probable worker relative to the other categories), the effect of this variable is not statistically significant.

By far the largest determinate of having some missing source of earnings information is the fraction of 1998 that cash assistance is received. Being on cash assistance for all of the year, compared with none of it, increases the likelihood of being a false non-worker (relative to the other two categories), consistent with the perceived incentive to hide earnings for women receiving welfare. Welfare receipt also increases the likelihood of being a probable worker relative to a sure worker, consistent with the incentive to hide income, perhaps through the use of non-matched social security numbers. The magnitude of this effects is very large. If all the entire sample workers were on cash assistance for all of 1998, we predict that 35 percent of them would be false nonworkers and 6 percent of them would be probable workers, compared to 16 and 3 percent who are actually in these categories. Conversely, if none of the workers in this sample received welfare in 1998 we would expect 90 percent of them to be sure

[24]For example, in the case of *false non-workers*, the characteristics of the job obtained from the survey (e.g., hourly or salary basis, receipt of tips and commissions) are unreported. In the case of *probable workers* characteristics of employment obtained from the UI records, such as the number of employers, are not available.

workers, compared to the observed 80 percent, with bulk of this 10 percentage-point decrease being due to a reduced fraction of false non-workers.[25]

Because of the way Wisconsin treats the earnings of welfare recipients, the question arises whether the large impact of welfare recipiency on the likelihood of being a false non-worker or probable worker arises from intentional misrepresentation of earnings by respondents. While it is impossible to answer this question with certainty, we can shed some light on this issue. If false non-worker status is due to intentional misrepresentation, we would expect false non-workers (who report no survey earnings) to be more likely to have a positive UI earnings report in quarters that they received cash assistance than sure workers. However, this is not the case as false non-workers are in fact less likely to have UI earnings in quarters in which they received cash assistance (about 52 percent of such quarters) than are sure workers (66 percent of such quarters).[26]

B. Correlates of the Discrepancy between Earnings Reports among Sure Workers

We also estimate a multivariate regression to explore the consistency of our conjectures with earnings discrepancies between S and UI reports. The sample of *sure workers* is used in the estimation, in order to identify the sources of the discrepancy among workers who have positive earnings reports in both data sources. The dependent variable measuring the earnings discrepancy is the difference between S and UI earnings (S – UI); with this specification, each coefficient (x 1000) is interpreted as the change in the difference between S and UI earnings associated with a unit change in the indicated characteristic.

In Table 6, we present the results of our regression of (S – UI) on individual socio-economic characteristics, the work and welfare history variables and a set of variables designed to reflect our

---

[25]As noted, the percent of 1998 that the worker receives welfare also increases the likelihood of being a *probable worker* relative to being a *sure worker*. This effect is such that 2 percent of the sample would be *probable workers* conditional on none of the sample being on welfare in 1998, compared to 6 percent in the actual sample.

[26]Ideally we would like to determine whether false non-workers were more likely to have UI earnings in the months in which they received cash assistance. Because employer reported UI earnings are only available on a quarterly basis we cannot compare monthly welfare receipt to monthly UI earnings.

conjectures (e.g., location, welfare receipt, intermittent or informal employment, job characteristics, and the elapsed time since last worked).[27]

Consider first the conjecture that individuals who have worked for an out-of-state employer have a large discrepancy because these workers will have some earnings that do not appear in the UI records. Two of the variables in the model allow us to assess this conjecture. If the conjecture is correct then respondents who report living out of state for some portion of 1998 are likely to have reduced UI earnings relative to survey earnings. Additionally, we might also expect that some respondents living in border counties would have some earnings that did not appear in UI records because these earnings were paid by out-of-state employers. The estimates in Table 6 support this conjecture, indicating that both living out of state in 1998 and residing in a border county lead to higher (S – UI) earnings differences. The effect of living out of state in 1998 is not statistically significant, but the effect of residing in a border county is.

With respect to the nature of work and the sources of compensation, we hypothesized that that earnings from tips, commissions, and odd jobs are likely to be undercounted in UI records relative to survey earnings. Consistent with this conjecture, the estimated effect of working an odd job in 1998 is to increase the (S – UI) earnings difference by approximately $670. The presence of income from tips and commissions is also estimated to increase the (S – UI) earnings difference, but the effect of this variable is not statistically different from zero at standard confidence levels.

Other conjectures concerned the role of being a steady worker (vs. a non-steady worker), difficulty in recalling earnings because of intermittent employment, or having a long gap between the time of employment and the date of the survey, all of which suggest error in survey earnings, but no particular directional bias in the S – UI earnings difference. In the model reported in Table 6, we included a dummy variable indicating being a steady worker.[28] The coefficient on this variable is large,

[27] We also estimated a model excluding the work and welfare history variables. The coefficient estimates for this model are similar to those in the model shown in the table; a joint significance test rejects the hypothesis that the coefficients in the two models are significantly different from each other.

[28] Recall that a steady worker is a *sure worker* who worked at least 3 quarters in 1998 according to UI records and who had no more than two employers according to both UI and survey records.

negative and statistically significant, suggesting that steady workers have a smaller (S – UI) earnings difference than do nonsteady workers. When the effects of this steady worker variable are netted out, there is little evidence that the number of quarters worked (reflecting intermittent work) or the time between the last quarter worked and the survey date[29] have a large impact on the (S – UI) earnings difference; the coefficients on both variables are relatively small and imprecisely estimated.

Although the number of quarters and last quarter worked in 1998 do not have statistically significant effects on the (S – UI) earnings difference, we predict higher survey earnings for a given level of UI earnings for workers that report that they 'don't know' their earnings; the coefficient on this variable is large and statistically significant. We are unable to identify the extent to which this effects reflects difficulty in remembering earnings, our crude imputation procedure, or intentional misrepresentation.[30]

Finally, consider the effect of the welfare receipt variable (the fraction of 1998 the respondent received cash assistance) on the difference between survey and UI earnings. Because the Wisconsin Works Program (W2) does not allow recipients to work for pay and receive cash assistance, there is an incentive for women who wish to work and receive program benefits to attempt to conceal their earnings, even in cases where confidentiality is promised. In the survey, these incentives toward concealment may lead to underreporting of earnings. In the UI data the incentives toward concealment may lead to women working under false social security numbers, working off the books or working odd jobs. Increased time on welfare in 1998 may lead to underreporting of both survey and UI earnings, meaning that its effect on

[29]The 1998 CSDE survey was administered in the spring of 1999 to correspond with the time that many workers would be looking back over their earnings records for the purpose of filling out tax returns. Because there was a uniform period of survey administration, the last quarter worked in 1998 is a proxy for the time between when a worker was last employed and when she is asked to recall her earnings on the survey.

[30]When a respondent reports that they 'don't know' their earnings in 1998, the survey administrator follows up with a series of questions designed to gain information about whether the respondent's earnings fell into certain intervals. The way that the survey is designed, the prospective earnings intervals start high, with each additional question inquiring if earnings fall into a lower interval. Because the questions start high and work their way to lower intervals, it may be more likely that respondents indicate that their earnings fell into a relatively high interval. In cases where respondents 'don't know' their earnings we interpolate their earnings as the midpoint of the indicated earnings interval. Because of this interpolation, and the design of the survey, we may obtain an upward biased estimate of survey earnings for respondents reporting that they 'don't know' earnings relative to the true value.

(S-UI) is ambiguous. The coefficient in the model indicates a significant and positive effect of spending time on cash assistance on the (S – UI) earnings difference.

In sum, with but few exceptions, our conjectures regarding the sources of the discrepancy between survey and UI earnings reports are confirmed in these estimates.

C. Simulated Effects of Selected Conjectures on the Total Discrepancy

The model reported in Table 6 can be used to simulate the quantitative contribution to the discrepancy between S and UI earnings of those factors expected to be related to this outcome. The discrepancy variable we use in this simulation is the MSD, and we simulate the percentage change in this variable attributable to each of the conjecture variables and to groups of these variables. Our simulation approach rests on the decomposition characteristics of the MSD measure, and is described in detail in Appendix A.

In our simulation, we set the variables of interest to values suggesting the absence of the expected effect (while holding all of the other variables at their observed levels) and record the estimated change in MSD. Our results, stated as the simulated percentage changes in MSD attributable to these alternative values of the conjecture variables are summarized below.

- Border County = 0: -1.12 percent
- Out of State in 1998 = 0: -0.0 percent
- Fraction of 1998 receiving cash assistance = 0: 3.91 percent
- Odd Job and Tips and Commissions = 0 -1.58 percent
- Steady worker = 1 -6.22 percent
- Overtime = 0 -0.19 percent
- Don’t Know Earnings = 0 -1.68 percent
- Last quarter worked was the third or fourth quarter -0.86 percent

Consistent with the conjectures, the presence of workers living in a border county increases the MSD, as does having an odd job, receiving tips and commissions, not being a steady worker, not knowing earnings, and last working prior to the third quarter of the year. Of the two conjecture variables for which the direction of the effect could be in either direction, having an overtime pay arrangement reduces the discrepancy, while assuming that no time is spent as a welfare recipient is associated with an increased MSD.

While this last finding seems counterintuitive, it has a sensible interpretation. Being on welfare during a year is associated with reduced earnings. For example, average survey and UI earnings for sample members who spent all of 1998 receiving cash assistance are $734 and $593 respectively, compared with $9025 and $8641 for sample members who spent none of 1998 receiving cash assistance. It follows that the absolute level of the survey-UI earnings discrepancy is also lower for those with low earnings relative to those with high earnings. Hence, simulating the effect of assuming that no worker was a welfare recipient results in both increased earnings levels and a greater level discrepancy between them.[31]

Using this same model, we also simulate the aggregate effect of two sets of conjecture variables on the mean squared discrepancy, one reflecting inadequacies in the UI measure and the other inadequacies in the survey measures; again, we set these variables at levels indicating the absence of the conjectured effect. The results are as follows:

- Variables reflecting the failure of UI to accurately capture earnings (living in a border county or out of state, having an odd job or receiving tips/commissions) — -2.65 percent

- Variables reflecting the failure of S to accurately capture earnings (due to forgetting or misrepresentation, including not being a steady worker, having overtime pay arrangements, not knowing earnings, or last working prior to the third quarter of the year) — -7.51 percent

---

[31] Given the importance of the level of survey and UI earnings in determining the discrepancy between the two values, it might be informative to estimate a model of the difference between log survey and log UI earnings. In such a model the independent variables would influence the percentage difference between survey and UI earnings, rather than the level difference between survey and UI earnings, by a constant amount. We estimated this model with the same set of independent variables as in the Table 7 specifications. While the estimated results were similar to those in Table 7, and the accompanying simulations, the fraction of 1998 spent on cash assistance had a lower simulated effect on the mean of the squared difference in log survey and UI earnings than on the MSD. The results of these log difference regressions are available from the authors upon request.

Overall, then, the variables that we have been able to study because of the detailed information available in the CSDE data account for about 10 percent of the total discrepancy. Factors that we are unable to measure—such as fixed difference in survey and UI reports not related to independent variables in the model[32], simple random variation, or nonsystematic effects of the conjecture variables—account for the bulk of the total discrepancy.

For example, in addition to their systematic impacts on the discrepancy, the conjecture variables might influence the discrepancy by affecting the random component of either survey or UI earnings reports. To explore the extent to which non-systematic effects of the conjecture variables increase the variability of earnings reports, and thus the discrepancy, we have regressed the squared residuals from the Table 7 regression on the independent variables. Two of the independent variables in this regression—the steady worker and the 'don't know' variables—are statistically different from zero at standard confidence levels.[33] Being a steady worker leads to survey and UI reports that are more consistent, while not knowing earnings in 1998 leads to survey reports that are substantially less consistent. The magnitude of these effects is large. We estimate that MSD would be reduced by 26 percent if the entire sample of sure workers were steady workers.[34] MSD would be reduced by an additional 8.5 percent if none of the sure workers in the sample reported not knowing their earnings. Thus, a substantial amount of the noisy reporting of both S and UI earnings can be explained the unsteady nature of work, problems of recall, or the misrepresentation of earnings by sample members.

VI. DO EMPLOYMENT AND EARNINGS FUNCTIONS VARY BY S AND UI?

Given the nature of these observed discrepancies between survey and UI data, an important question is whether there is a significant difference in the conclusions obtained from equations estimated

---

[32]The mean difference in survey and UI earnings accounts for about 6 percent of MSD.

[33]The results of this regression are available from the authors upon request.

[34]This is above and beyond the 6.22 percent reduction in MSD resulting from the systematic effect of steady worker status reported above

with these alternative variables. To answer this question, we estimated simple models of employment and earnings, using both data sources.[35] The independent variables in each of these equations are age, age squared (divided by 100), indicators of educational attainment (high school dropout, high school graduate, and some college), and indicators of race (white, black, Hispanic or other). These models are estimated over two samples: all women who appeared in the 1998 and 1999 surveys and sure workers that appeared in the 1998 and 1999 surveys.

We conducted an F-test of the equivalence of the coefficients (or sets of coefficients) across the two regressions for each group of women. Table 7 summarizes the results, with cells in bold indicating significant differences between the coefficients using survey and UI information. For the models using all of the observations, a substantial number of estimated relationships differ significantly between the survey and UI measures of earnings. In particular, significant differences between the coefficients on the education variable estimated using the alternative earnings variables are indicated; the entire set of coefficients also fail to pass the relevant F-test indicating significant differences in estimated effects depending on the earnings data used in estimation. These differences do not exist when the earnings functions are fit over only all sure workers; in no cell of Table 7 estimates fit over this group of workers are significant differences indicated. We conclude that estimates of the determinant earnings and employment are somewhat sensitive to the source of data of the dependent variable, especially for estimates fit over full samples of observations.

VII. DO ESTIMATES OF TOTAL AND GROUP SPECIFIC EMPLOYMENT, EARNINGS, AND POVERTY STATUS VARY BY S AND UI?

Studies of low-income women, especially the numerous studies of welfare leavers,[36] monitor the employment and earnings of these workers over time in order to assess the effects of policy reform

---

[35]The equation for employment was estimated as a logit model, while the earnings regression was estimated using OLS. Employment in both the survey and UI data was defined by the presence of positive earnings.

[36]Cancian, Haveman, Meyer, and Wolfe, 2003, presents such results, and reviews other studies of this nature.

efforts. Both UI and survey information are used in these assessments of the performance of leavers. Our data allow us to estimate the extent to which these patterns vary by the source of the information used, ultimately aiding in reconciling information across different studies.

Because the monitoring studies often emphasize subgroup differences in employment and earnings, we present the (S – UI) differences in these variables for race and education subgroups using data on all respondents and all *sure workers* included in both the 1998 and 1999 surveys. We also show subgroup differences in estimates of earnings growth between the two sources of information.

Consider first the comparisons of employment rates shown in Table 8. For all of the subgroups, the employment rate based on UI information exceeds that based on the survey data. The S/UI ratios range from .81 to .97 suggesting quite different patterns among the groups based on the source of information. For all subgroups, the patterns of change in employment rates from 1998 to 1999 based on survey data are larger (suggesting more growth or smaller decreases) than those based on UI information.

A similar pattern of differences in the level of earnings is shown in Tables 9 and 10 for all workers and *sure workers*, respectively. For all workers, the S/UI ratio of earnings ranges from 1.03 to 1.23 across the subgroups. For *sure workers*, S exceeds UI even more, and S/UI ranges 1.12 to 1.18 across the subgroups. For *sure workers*, earnings growth for all of the subgroups is greater when measured using UI information, but for all workers the S and UI differences in growth patterns vary across the subgroups.[37]

Overall, the use of survey information tends to understate employment levels but overstate earnings among low-skill female workers, relative to information from administrative records. Survey-based employment rates tend to be about 90 percent of those based on administrative records. Conversely, earnings for *sure workers* estimated from survey-based information are about 16 percent higher than those based on UI data, and about 12 percent higher for all women. For *sure workers*, the ratios of survey to UI

[37] For example, earnings growth among educated nonwhites of 35 percent is suggested by the survey data, compared to 33 percent when UI earnings are used. For less educated nonwhites the pattern is reversed, with 35 percent growth indicated by the UI information and 14 percent when survey data are used.

earnings are similar among the race-education subgroups, but vary substantially among by subgroups when all workers are studied.

In terms of employment growth, use of survey information yields larger employment increases for all of the subgroups. However, earnings increases based on UI information tend to be larger than when survey information is used. For all workers the patterns of earnings growth among nonwhites without a high school diploma differ substantially across the S and UI data, but other groups show similar patters of earnings growth across the two sources of earnings information. These results suggest that analysts tracking the employment and earnings growth of welfare leavers or other populations of low skill workers need to interpret carefully the patterns that they observe.

VIII. CONCLUSIONS

Using data on a large sample of low skill women with children, we find substantial disparities in employment and earnings reports between a uniquely high-quality survey of low-skill workers and employer-based reports of earnings. These differences exist for both steady workers and those who work intermittently and on several jobs. Some of these differences are to be expected because of differences between the two data sources in coverage, definition, and the process of data collection.

We proposed several ‘conjectures’ for these discrepancies reflecting both the differences in definition and data collection between survey and UI information sources, and the location or job-related characteristics of the workers. Using information available in the survey, we measured the relationship of these worker and employment characteristics to the work and earnings discrepancies among the workers in the sample. Although the survey data are unique in the extent of detailed information regarding work and earnings patterns they provide, we were able to account for only about 10 percent of the total discrepancy; the great bulk of the discrepancy is due to random error in data reporting or recording or definitional or employment differences for which we are unable to account.

Our estimates of the effect of the alternative data sources on both econometric estimates of the determinants of work and earnings outcomes and the reliability of measures of employment and earnings levels and trends (such as those reported in studies designed to monitor the labor market success of low skill women, such as welfare leavers) suggest the need for caution by researchers in interpreting results from such studies.

Appendix A

Our simulation approach relies on the least squares decomposition of the sum of the squared dependent variable. Assuming that

$$\left(S_i - UI_i\right) = x_i' \cdot \beta + e_i ,$$

the average squared discrepancy can be decomposed as follows:

$$MSD = \frac{1}{n}\sum_{i=1}^{n}\left(S_i - UI_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\beta' \cdot x_i \cdot x_i' \cdot \beta + \frac{1}{n}\sum_{i}\left(S_i - UI_i - x_i' \cdot \beta\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\beta' \cdot x_i \cdot x_i' \cdot \beta + \frac{1}{n}\sum_{i} e_i^2 \qquad (1)$$

This decomposition indicates that the mean squared discrepancy (MSD) is due in part to systematic effects of the x variables and in part to random differences in S and UI. The ratio of the first right hand side term in (1) to the MSD is the fraction of the variance of survey less UI earnings around zero that is due to the x variables (including the intercept term). One minus this fraction is attributable to random errors in reporting.

The coefficients in Table 7 provide estimates of the β used in this simulation. With estimated coefficients ($\hat{\beta}$) replacing the actual coefficients and actual residuals ($\hat{e}$) replacing the error terms, we obtain the following decomposition

$$MSD = \frac{1}{n}\sum_{i}\left(S_i - UI_i\right)^2 = \frac{1}{n}\sum_{i}\hat{\beta}' \cdot x_i \cdot x_i' \cdot \hat{\beta} + \frac{1}{n}\sum_{i}\hat{e}_i^{\,2}$$

This decomposition yields an estimate of the variance of (S-UI) around zero explained by the independent variables.

We simulated the effect of the conjecture variables (as a subset of the x variables) by setting them to alternative values and measuring the estimated change in MSD. For example, we assume that no worker lived in a border county, no worker lived out of state, and no worker received income from tips and/or commissions. Letting MSD_0 be the actual value of MSD, the percentage simulated change in the MSD is

$$\frac{100 \cdot \left(MSD\,|_{x=x_1} - MSD_0 \right)}{MSD_o} ,$$

where $MSD\,|_{x=x_1}$ is the MSD with a subset of the independent variables set to specific values.

REFERENCES

Acs, Gregory, and Pamela Loprest. 2001. "Studies of Welfare Leavers: Data, Methods, and Contributions to the Policy Process." In *Studies of Welfare Populations: Data Collection and Research Issues*, National Research Council: National Academy Press, 385–414.

Arizona Department of Economic Security. 2000. *Arizona Cash Assistance Exit Study: First Quarter 1998 Cohort.*

Baj, John, Charles Trott, and David Stevens. 1991. "A Feasibility Study of the Use of Unemployment Insurance Wage-Record Data as an Evaluation Tool for JTPA: Report on Project Phase 1 Activities." National Commission on Employment Policy.

Baj, John, Sean Fahey and Charles E. Trott. 1992. "Using Unemployment Insurance Wage-Record Data for JTPA Performance Management." Research Report 91-07. National Commission for Employment Policy, Chapter 4.

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

Burgess, Paul L., Arthur E. Blakemore, and Stuart A. Low. 1998. "Using Statistical Profiles to Improve Unemployment Insurance Tax Compliance." *Research in Employment Policy* 1: 243–271.

Cancian, Maria, Robert Haveman, Daniel Meyer, and Barbara Wolfe. 2003. "Before and After TANF: The Economic Well-Being of Women Leaving Welfare." *Social Service Review* 76(4): 603-641.

Hotz, Joseph V., and John Karl Scholz, 1999. "Measuring Employment and Income Outcomes for Low-Income Populations with Administrative and Survey Data." In *Studies of Welfare Populations: Data Collection and Research Issues*, National Research Council: National Academy Press, 2002, 275–315.

Issacs, Julia and Matthew Lyon. 2000. "A Cross-State Examination of Families Leaving Welfare: Findings from the ASPE-Funded Leavers Studies." United States Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation: Washington, D.C. *http://aspe.hhs.gov/hsp/leavers99/cross-state00/index.htm*

Kornfeld, Robert, and Howard S. Bloom. 1999. "Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals?" *Journal of Labor Economics* 17(1): 168–197.

Moore, Jeffrey C., Linda L. Stinson, and Edward J. Welniak, Jr. 1997. "Income Measurement Error in Surveys: A Review." U.S. Bureau of the Census, Statistical Research Report.

Rodgers, Willard L., Charles Brown, and Greg J. Duncan. 1993. "Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages." *Journal of the American Statistical Association* 88(424): 1208–1218.

Table 1
Survey versus UI Reports of Employment Status
(Percentages in parentheses)

	UI Employment Status		
Survey Employment Status	Employed	Not Employed	Total
Employed	1,514	88	1,602
	(69.48)	(4.04)	(73.52)
	sure workers	**probable workers**	
Not Employed	305	272	577
	(14.00)	(12.48)	(26.48)
	false non-workers	**sure non-workers**	
Total	1,819	360	2,179
	(83.48)	(16.52)	(100.00)

Table 2
Survey/UI Earnings Discrepancies by Earnings Groups and Among Sure Workers (N=2,179)

Earnings Groups	Frequency (Percent of Sample/Percent of Sure Workers)	Mean Survey Earnings ($000)	Mean UI Earning ($000)	Mean Absolute Discrepancy ($000) (Percent of Total Absolute Discrepancy)	Mean Squared Discrepancy ($000) (Percent of Total Squared Discrepancy)
Discrepancies by Earnings Groups					
Zero Earnings in Survey and UI Records **(sure nonworkers)**	272 (12.48/ NA)	—	—	0.00 (0.00)	0.00 (0.00)
Positive Earnings in Survey and UI Records **(sure workers)**	1,514 (69.48/ NA)	7.76	6.59	2.89 (74.60)	24.68 (71.42)
Earnings in Survey/ No Earnings in UI **(probable workers)**	88 (4.04/ NA)	5.48	—	5.48 (8.21)	75.26 (12.66)
Earnings in UI/ No Earnings in Survey **(false nonworkers)**	305 (14.00/ NA)	—	3.31	3.31 (17.19)	27.30 (15.91)
Discrepancies Among Sure Workers				[Percent of Sure Worker Discrepancy]	
By Earnings Difference Levels					
(Survey-UI)>$2500	354 (16.25/23.38)	13.22	5.89	7.33 (44.19), [59.24]	77.41 (52.20), [73.09]
-$2,500<(Survey-UI)<$2500	1,005 (46.12/66.38)	6.25	6.14	0.85 (14.56), [19.51]	1.23 (2.37), [3.32]
(Survey-UI)<-$2500	155 (7.11/10.24)	1.11	5.10	6.00 (15.85), [21.24]	56.86 (16.85), [23.59]
By Steady Worker Status					
Steady Worker	509 (23.36/33.62)	9.85	9.61	2.63 (22.76),[30.51]	18.83 (18.32),[25.65]
Unsteady Worker	1,005 (76.64/66.38)	6.71	5.06	3.03 (51.84),[69.49]	27.64 (53.10),[74.35]

Table 3
Distribution of Survey Minus UI Earnings

Survey Less UI Earnings	Frequency	Percent	Cumulative Percent	Bloom and Kornfeld Percent
$8,001 or more	141	6.47	6.47	3.5
$4,001 to $8,000	147	6.75	13.22	7.7
$2,401 to $4,000	133	6.10	19.32	7.9
$1,601 to $2,400	107	4.91	24.23	6.9
$800 to $1,600	147	6.75	30.98	10.4
$1 to $800	283	12.99	43.97	17.3
$0	285	13.08	57.04	14.3
-$800 to -$1	395	18.13	75.17	16.2
-$1,600 to -$801	157	7.21	82.38	6.2
-$2,400 to -$1,601	105	4.84	87.20	3.3
-$4,000 to -$2,401	99	4.54	91.74	3.4
-$8,000 to -$4,000	110	5.05	96.79	2.3
-$8,001 or less	70	3.21	100.00	0.9
Total	2,179	100.00	100.00	100.0

Table 4
Conjectures Regarding the Source and Magnitude of Discrepancy between S and UI Reports of Earnings

Conjecture	Rationale for Conjecture
1. On average, respondents who report earnings from odd jobs, tips, and commissions will have larger discrepancies, all else equal.	UI records do not include earnings from odd jobs, tips, or commissions, so earnings from these sources increase S relative to UI, leading to larger discrepancies.
2. On average, respondents who report holding multiple jobs, or for whom the UI database indicates multiple employers, will have larger discrepancies, all else equal.	Individuals with multiple jobs or employers may have more difficulty accurately recalling their annual income in a survey, resulting in larger discrepancies.
3. On average, respondents who were on welfare for more months during 1998 will have larger discrepancies, all else equal.	Because Wisconsin reduces welfare benefits dollar for dollar with earned income, there are strong incentives for working welfare recipients to conceal their work activities. This may in reduced reports survey or UI earnings depending on the nature of concealment.*
4. Respondents who lived out of state for some portion of 1998 or lived in a border county will have larger discrepancies, all else equal.	Because individuals living out of state, or in a border county, are more likely to be working out of state, and because non-Wisconsin employers are not likely to report earnings to Wisconsin UI system, survey earnings will be higher than UI earnings, leading to larger discrepancies.
5. Steady workers—those with earnings in at least 3 quarters of the year, and with no more than two employers or jobs during the year—will have smaller discrepancies, all else equal.	Workers with continuous work on a few jobs or for few employers are more likely to provide reliable earnings reports to survey interviewers.
6. Respondents who indicate that they ‘don’t know’ their earnings (and for whom an imputed value is provided) will have larger discrepancies than those who do respond to the earnings question, all else equal.	Respondents indicating that they don’t know their earnings in 1998 are more likely to report their earnings to surveyors with error (either intentionally or unintentionally). Additionally, errors introduced in the imputation process may increase the discrepancy.
7. On average, those whose last quarter of employment was early in the year will have larger discrepancies than those who worked in the last quarter of the year, all else equal.	Those whose last quarter of employment was far from the date of the survey are likely to forget their prior years earnings, and to report them with error.

*Concealing work activities may take two forms. Recipients may be inclined to underreport (or to not report) their earnings to surveyors if they believe this information will result in adverse administrative decisions such as loss of benefits. In this case UI earnings would be higher than survey earnings, all else equal. Alternatively, recipients may provide false social security numbers to employers in an attempt to disguise their earnings. Assuming these earnings are reported to surveyors, such concealment would lead to higher survey than UI earnings, all else equal.

Table 5
Multinomial Logit Estimates of the Probability of Being a False Non-Worker, Sure Worker, or Probable Worker
(Relative Odds Ratios Reported with Standard Errors in Parentheses, N=1,907)

	Relative to Sure Worker		Relative to Probable Worker	
Independent Variable	False Non-Worker	Probable Worker	Sure Worker	False Non-Worker
AGE	1.0273 (0.0701)	1.0853 (0.1227)	0.9214 (0.1042)	0.9465 (0.1184)
AGE SQUARED	0.9826 (0.1123)	0.9154 (0.1695)	1.0924 (0.2023)	1.0734 (0.2209)
HIGH SCHOOL GRADUATE	0.6925** (0.0989)	0.8471 (0.2114)	1.1806 (0.2946)	0.8176 (0.2235)
SOME COLLEGE	0.5016** (0.1375)	1.2544 (0.4324)	0.7972 (0.2748)	0.3998** (0.1672)
BLACK	1.1439 (0.2264)	0.8287 (0.2652)	1.2066 (0.3861)	1.3802 (0.4936)
HISPANIC OR OTHER	1.9938** (0.5303)	2.1172** (0.8015)	0.4723** (0.1788)	0.9417 (0.4043)
MILWAUKEE	1.1724 (0.3563)	0.6264 (0.2503)	1.5965 (0.6381)	1.8718 (0.8884)
OTHER URBAN	1.6285 (0.4971)	0.9086 (0.3424)	1.1006 (0.4148)	1.7924 (0.8264)
BORDER COUNTY	1.1254 (0.3390)	1.6508 (0.6317)	0.6058 (0.2318)	0.6818 (0.3150)
OUT OF STATE IN 1998	0.5828 (0.6298)	2.3213 (2.5009)	0.4308 (0.4641)	0.2511 (0.3593)
FRACTION OF 1998 ON CASH ASSISTANCE	8.0127** (1.7935)	5.2900** (2.0033)	0.1890** (0.0716)	1.5147 (0.6289)

Table 6
OLS Estimates of the Determinates of the Survey-UI Earnings Difference for Sure Workers
(Standard Errors in Parentheses, N=1,514)

Variable	Model I	Model II
AGE	0.0337 (0.1285)	0.0287 (0.1342)
AGE*AGE/100	-0.1260 (0.2149)	-0.1127 (0.2230)
HIGH SCHOOL GRADUATE (vs. Less than High School)	0.2826 (0.2678)	0.2603 (0.2699)
SOME COLLEGE (vs. Less than High School)	0.9013 (0.4170)	0.8642 (0.4220)
BLACK (vs. White)	-0.4415 (0.3495)	-0.4366 (0.3516)
HISPANIC OR OTHER (vs. White)	-0.6692 (0.5204)	-0.6363 (0.5219)
MILWAUKEE (vs. Rural)	-0.0757 (0.4650)	-0.0177 (0.4747)
OTHER URBAN (vs. Rural)	-0.9659 (0.4507)	-0.9377 (0.4533)
BORDER COUNTY (=1)	0.9148 (0.4712)	0.9306 (0.4719)
OUT OF STATE IN 1998 (=1)	0.6054 (1.6906)	0.5999 (1.6920)
FRACTION OF 1998 ON CASH ASSISTANCE	-0.9246 (0.4916)	-0.8544 (0.4940)
STEADY WORKER (=1)	-1.2313 (0.2919)	-1.1653 (0.2937)
HOURLY WORKER (vs. Salaried Worker)	-0.8905 (0.7488)	-0.8892 (0.7491)
ARRANGEMENTS FOR OVERTIME PAY (=1)	0.0426 (0.2530)	0.0349 (0.2540)
WORKED AN ODD JOB (=1)	0.6765 (0.3440)	0.6632 (0.3442)
HAD INCOME FROM TIPS AND COMMISSIONS (=1)	0.4042 (0.4933)	0.4216 (0.4934)
DIDN'T KNOW EARNINGS (=1)	0.9128 (0.3431)	0.9390 (0.3434)
LAST QUARTER WORKED WAS QUARTER 1 OR 2 (=1)	0.6574 (0.5259)	0.6104 (0.5289)
Quarters Worked in 1998 Controls	Yes	Yes
Work History Controls	No	Yes
Welfare History Controls	No	Yes
Adjusted R-Squared	0.042	0.042

Table 7
Tests of Significance of the Difference between Coefficients in Models Fit to S and UI Data

	Intercept	Age	Education	Race	All Except Intercept	All	(Pseudo) R^2
Employment 1998	F = .13 P = .71	F = 1.42 P = .49	F = 3.34 P = .19	F = 1.93 P = .38	F = .16 P = .23	**F = 60.3** **P = .00**	.06
Employment 1999	F = .01 P = .92	F = 1.22 P = .54	**F = 5.18** **P = .07**	F = 4.27 P = .12	**F = 13.1** **P = .04**	**F = 39.5** **P = .00**	.06
Earnings All 1998	F = .16 P = .69	F = .28 P = .75	**F = 2.38** **P = .09**	F = 2.29 P = .10	**F = 1.85** **P = .09**	NA	.07
Earnings All 1999	F = .93 P = .33	F = .75 P = .47	**F = 3.59** **P = .03**	F = 1.91 P = .15	**F = 2.22** **P = .04**	NA	.06
Earnings Sure 1998	F = 1.14 P = .29	F = 1.15 P = ..31	F = .09 P = .91	F = .07 P = .91	F = .43 P = .86	NA	.07
Earnings Sure 1998	F = 1.42 P = .23	F = 1.29 P = .28	F = .21 P = .81	F = .90 P = .41	F = .80 P = .57	NA	.08

Table 8
S and UI Employment, 1998, and Change in Employment from 1998 to 1999:
All Workers in both 1998 and 1999 Data

	Survey	UI	S/UI	S99/S98	UI99/UI98
NW, High School or more (382)	.86	.89	.97	.99	.95
NW, Less than High School (212)	.75	.83	.90	.96	.94
W, High School or more (597)	.78	.86	.91	1.03	1.00
W, Less than High School (691)	.64	.79	.81	1.09	1.03
Total (1882)	.72	.86	.86	1.03	.99

Table 9
S and UI Earnings, 1998, and Change in Earnings from 1998 to 1999:
All Workers in both 1998 and 1999 Data

	Survey	UI	S/UI	S99/S98	UI99/UI98
NW, High School or more (382)	$7604	$6427	1.18	1.35	1.33
NW, Less than High School (212)	$5085	$4127	1.23	1.14	1.35
W, High School or more (597)	$6975	$6327	1.10	1.32	1.37
W, Less than High School (691)	$3843	$3717	1.03	1.39	1.39
Total (1882)	$5740	$5141	1.12	1.33	1.36

Table 10
S and UI Earnings, 1998, and Change in Earnings from 1998 to 1999:
Sure Workers in both 1998 and 1999 Data

	Survey	UI	S/UI	S99/S98	UI99/UI98
NW, High School or more (277)	$9150	$7727	1.18	1.34	1.36
NW, Less than High School (115)	$7094	$6084	1.17	1.28	1.35
W, High School or more (376)	$9489	$8284	1.15	1.29	1.36
W, Less than High School (335)	$9358	$8141	1.12	1.26	1.32
Total (1103)	$8316	7153	1.16	1.30	1.35

Figure 1

Scatter Plot of Survey Versus UI Earnings

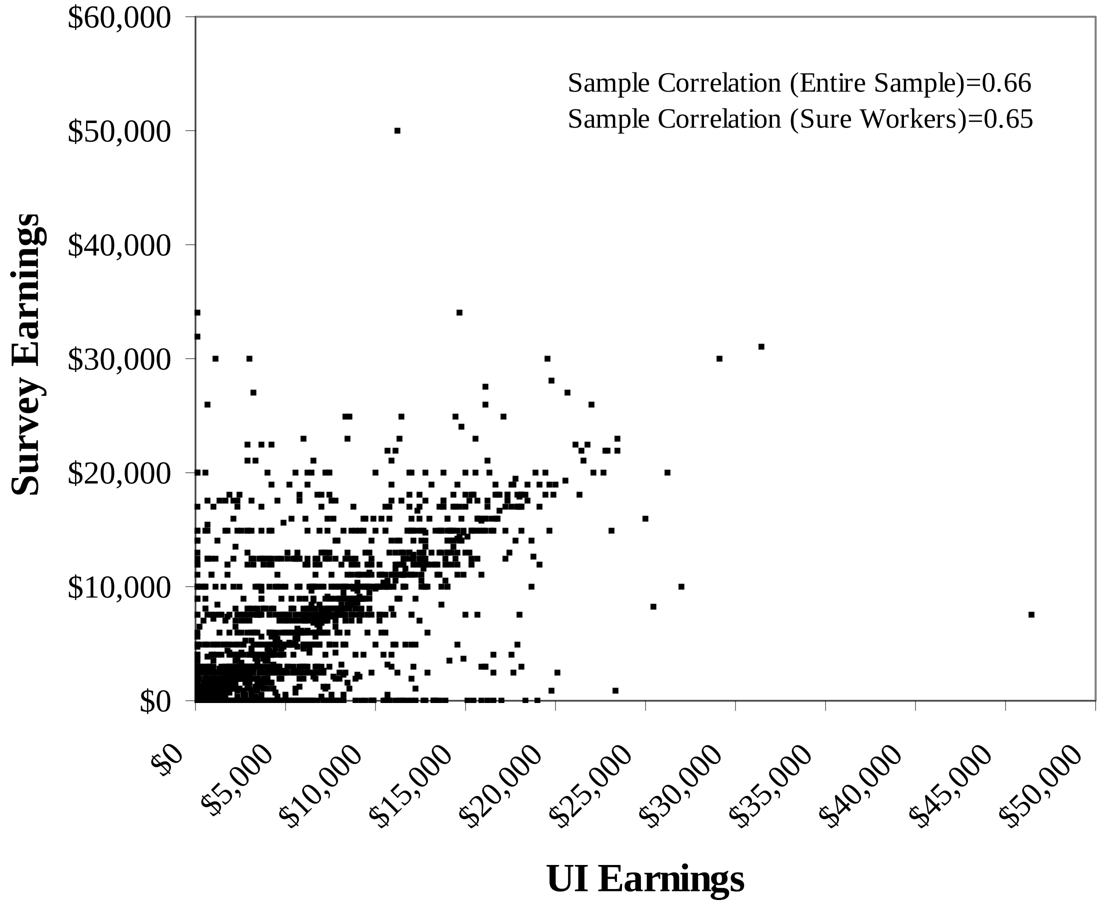

Figure 2

Summary Measures by of S and UI Earnings by Earnings Group

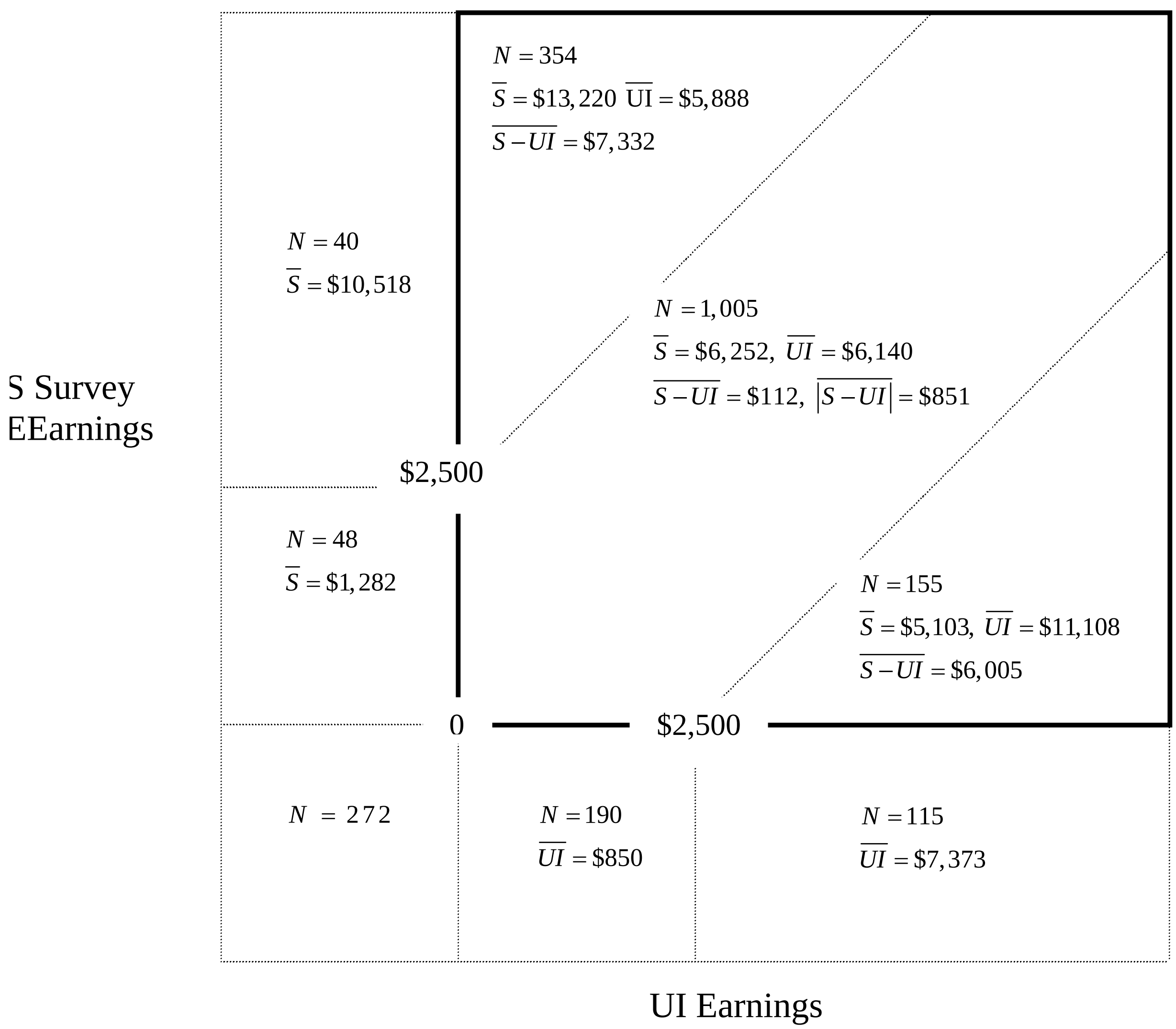

Table A1
Sample Means and Standard Deviations by Earnings Group

Variable	Sure Workers	False Non-Workers	Probable Workers	Sure Non-Workers
SURVEY EARNINGS	7.764	—	5.480	—
(In 1000s of Dollars)	(5.112)		(6.764)	
UI EARNINGS	6.590	3.310	—	—
(In 1000s of Dollars)	(5.345)	(4.050)		
EARNINGS DIFFERNCE	1.1741	-3.310	5.480	—
(Survey less UI earnings)	(4.828)	(4.050)	(6.764)	
SQUARED DISCREPANCY	24.677	27.303	75.263	—
	(4.968)[1]	(5.225)[1]	(8.675)[1]	
ABSOLUTE DISCREPENCY	2.894	33.00	5.480	—
(In 1000s of $)	(4.039)	(4.050)	(6.764)	
AGE	26.615	26.875	28.034	29.562
	(6.854)	(7.116)	(7.291)	(8.534)
HIGH SCHOOL DROPOUT (=1)	0.439	0.607	0.466	0.643
	(0.496)	(0.489)	(0.502)	(0.480)
HIGH SCHOOL GRADUATE (=1)	0.439	0.334	0.375	0.276
	(0.496)	(0.473)	(0.487)	(0.448)
SOME COLLEGE (=1)	0.122	0.059	0.159	0.081
	(0.328)	(0.236)	(0.368)	(0.273)
WHITE (=1)	0.341	0.207	0.352	0.221
	(0.474)	(0.406)	(0.480)	(0.415)
BLACK (=1)	0.588	0.695	0.511	0.673
	(0.492)	(0.461)	(0.503)	(0.470)
HISPANIC OR OTHER (=1)	0.072	0.098	0.136	0.107
	(0.257)	(0.298)	(0.345)	(0.309)
MILWAUKEE (=1)	0.680	0.793	0.625	0.805
	(0.447)	(0.406)	(0.487)	(0.397)
OTHER URBAN (=1)	0.182	0.144	0.205	0.103
	(0.386)	(0.352)	(0.406)	(0.304)
RURAL (=1)	0.138	0.062	0.170	0.092
	(0.345)	(0.242)	(0.378)	(0.289)
BORDER COUNTY (=1)	0.010	0.066	0.148	0.048
	(0.300)	(0.248)	(0.357)	(0.214)
OUT OF STATE IN 1998 (=1)	0.005	0.003	0.011	0.007
	(0.072)	(0.057)	(0.107)	(0.086)
FRACTION OF 1998 ON CASH ASSIST	0.353	0.586	0.471	0.748
	(0.315)	(0.336)	(0.371)	(0.298)

(table continues)

Table A1, continues

Variable	Sure Workers	False Non-Workers	Probable Workers	Sure Non-Workers
NUMBER EMPLOYERS – UI	2.563 (1.678)	2.089 (1.343)	—	—
NUMBER OF JOBS - SURVEY	1.945 (1.312)	—	1.068 (0.814)	—
SALARIED WORKER (=1)	0.028 (0.166)	0.000 (0.000)	0.115 (0.321)	—
HOURLY WORKER (=1)	0.971 (0.168)	1.000 (0.000)	0.875 (0.333)	—
WOULD BE PAID FOR OVERTIME (=1)	0.511 (0.500)	0.007 (0.081)	0.216 (0.414)	—
HAD ODD JOB (=1)	0.151 (0.358)	0.000 (0.000)	0.170 (0.378)	—
HAD TIPS AND COMMISSIONS (=1	0.067 (0.251)	0.003 (0.057)	0.057 (0.232)	—

Table A2
OLS Estimates of the Determinates of the Log(Survey Earnings)-Log(UI Earnings) for Sure Workers
(Standard Errors in Parentheses, N=1,514)

Variable	Model I	Model II
AGE	0.0213 (0.238)	0.0197 (0.0248)
AGE*AGE/100	-0.0469 (0.0397)	0.0435 (0.0413)
HIGH SCHOOL GRADUATE (vs. Less than High School)	0.0771 (0.0495)	0.0724 (0.0500)
SOME COLLEGE (vs. Less Than High School)	0.0896 (0.771)	0.0842 (0.0781)
BLACK (vs. White)	-0.1311 (0.647)	-0.1319 (0.0651)
HISPANIC OR OTHER (vs. White)	-0.0860 (0.963)	-0.0833 (0.0966)
MILWAUKEE (vs. Rural)	-0.0097 (0.860)	-0.0022 (0.0879)
OTHER URBAN (vs. Rural)	-0.1113 (0.834)	-0.1088 (0.0839)
BORDER COUNTY (=1)	0.1823 (0.872)	0.1860 (0.0874)
OUT OF STATE IN 1998 (=1)	0.3111 (0.3127)	0.3143 (0.3132)
FRACTION OF 1998 ON CASH ASSISTANCE	-0.1125 (0.909)	-0.1033 (0.0914)
STEADY WORKER (=1)	-0.1618 (0.0540)	-0.1517 (0.0544)
HOURLY WORKER (vs. Salaried Worker)	-0.1435 (0.1385)	-0.1431 (0.1387)
ARRANGEMENTS FOR OVERTIME PAY (=1)	0.0200 (0.468)	0.0185 (0.0470)
WORKED AN ODD JOB (=1)	0.1579 (0.0636)	0.1559 (0.0637)
HAD INCOME FROM TIPS AND COMMISSIONS (=1)	0.0879 (0.0912)	0.0895 (0.0913)
DIDN'T KNOW EARNINGS (=1)	0.3719 (0.0635)	0.3755 (0.0636)

(table continues)

Table A2, continued

Variable	Model I	Model II
LAST WORKED IN QUARTER 1 OR 2 (=1)	0.1342 (0.0973)	0.1231 (0.0978)
Quarters Worked in 1998 Controls	Yes	Yes
Work History Controls	No	Yes
Welfare History Controls	No	Yes
Adjusted R-Squared	0.123	0.120

Table A3
OLS Estimates of the Determinates of the Squared Residuals from Table 8 Model I Survey-UI Earnings Difference Equation
(Standard Errors in Parentheses, N=1,514)

Variable	Coefficient Estimate
AGE	-0.2169 (2.0323)
AGE*AGE/100	-0.0747 (3.3987)
HIGH SCHOOL (vs. High School Dropout)	-2.7249 (4.2362)
SOME COLLEGE (vs. High School Dropout)	12.3174 (6.5965)
BLACK (vs. White)	0.5845 (5.5292)
HISPANIC OR OTHER (vs. White)	-5.2562 (8.2325)
MILWAUKEE (vs. Rural)	7.6341 (7.3563)
OTHER URBAN (vs. Rural)	13.1720 (7.1288)
BORDER COUNTY (=1)	-10.6786 (7.4545)
OUT OF STATE IN 1999 (=1)	-1.2294 (26.7428)
FRACTION OF 1999 ON CASH ASSISTANCE	-13.7477 (7.7767)
STEADY WORKER (=1)	-9.6603 (4.6167)
HOURLY WORKER (vs. Salaried worker)	-8.0662 (11.8446)
ARRANGEMENTS FOR OVERTIME PAY (=1)	0.1935 (4.0015)
WORKED AN ODD JOB (=1)	6.8370 (5.4413)
HAD INCOME FROM TIPS AND COMMISSIONS (=1)	-7.5991 (7.8027)
DIDN'T KNOW EARNINGS (=1)	13.7097 (5.4281)

(table continues)

Table A3, continued

Variable	Coefficient Estimate
LAST QUARTER WORKED WAS 1 OR 2 (=1)	3.9424 (8.3189)
Quarters Worked in 1998 Controls	Yes
Work History Controls	No
Welfare History Controls	No
Adjusted R-Squared	0.007