IS EARNINGS NONRESPONSE IGNORABLE?

Christopher R. Bollinger and Barry T. Hirsch*

Abstract—Earnings nonresponse in the Current Population Survey is roughly 30% in the monthly surveys and 20% in the March survey. If nonresponse is ignorable, unbiased estimates can be achieved by omitting nonrespondents. Little is known about whether CPS nonresponse is ignorable. Using sample frame measures to identify selection, we find clear-cut evidence among men but limited evidence among women for negative selection into response. Wage equation slope coefficients are affected little by selection, but because of intercept shifts, wages for men and, to a lesser extent, women are understated, as are gender gaps. Selection is least severe among household heads.

I. Introduction

THE Current Population Survey (CPS) is used extensively by economists and other social scientists because of its large sample sizes, comprehensiveness, historical continuity, and timeliness. The monthly CPS Outgoing Rotation Group (ORG) files are widely used to analyze hourly earnings for wage and salary workers based on the principal job the previous week, and the American Social and Economic Supplement (ASEC) to the March CPS is similarly used to examine earnings reported across all wage and salary jobs during the previous calendar year.

Item nonresponse rates are low for most questions in the CPS, the notable exception being questions on income and earnings. Currently, about 30% of wage and salary workers sampled in the CPS ORG do not provide earnings information. Missing earnings are allocated to nonrespondents using a cell hot deck imputation procedure (Hirsch & Schumacher, 2004). In the March CPS, about 20% of individuals employed the previous year fail to report annual earnings. Their earnings are assigned using a sequential hot deck procedure (Lillard, Smith, & Welch, 1986).

Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006) establish that even if nonresponse is random, the imputation procedure can produce severe match bias. Wage regression coefficients on attributes that are not match criteria (for example, union, industry, foreign born) are biased toward 0 by a proportion close to the nonresponse (imputation) rate. Coefficients on imperfectly matched attributes such as education can also be severely biased. For example, returns to the GED are overstated because nonrespondents with a GED are typically assigned the earnings of donors with a regular high school diploma or some college.[1] Bollinger and Hirsch (2006) examine alternative methods to account for match bias, the simplest being removal of imputed earners (nonrespondents) from the estimation sample. But these approaches assume that nonresponse is either random or ignorable. Yet we have surprisingly little information on whether earnings nonresponse in the CPS (and other surveys) is ignorable.

The goal of this paper is to address the following important questions. Is earnings nonresponse in the CPS ignorable? If nonignorable, what is the nature and severity of the bias, and how might researchers account for it? We address these questions using CPS ORG and March ASEC data files for 1998 through 2008. Our principal approach is the estimation of selection-adjusted wage equations in which we use CPS sample frame measures to account for selection. We also examine the effect of proxy responses on reported earnings. In the CPS, a single household member generally provides responses for all household members. Thus, roughly half of earnings records are based on self-responses and half on the response of a proxy, often a spouse. Earnings nonresponse is far more likely when there is a proxy respondent, so the proxy status of an earnings record provides a potential measure to identify selection into response, assuming that proxy status does not affect the wage, conditional on other regressors.

II. Response Bias and CPS Earnings: What Is Known?

Surprisingly little is known about whether nonresponse in the CPS is ignorable and whether imputation does a good job, on average, in estimating earnings. A fairly sizable literature uses validation studies to evaluate the accuracy of measured earnings, and several of these use the CPS linked to administrative data.[2] These studies, however, typically exclude nonrespondents from the analysis.

Only a few studies have examined the quality of imputed values and the issue of response bias in the CPS. The work of which we are aware focuses on older March CPS files measuring annual earnings the previous year and not the widely used monthly earnings ORG files. Greenlees, Reese, and Zieschang (1982) examine the March 1973 CPS and compare wage and salary earnings the previous year with 1972 matched income tax records. They restrict their analysis to full-time, full-year heads of households in the private nonagricultural sector whose spouse did not work. They conclude that nonresponse is not ignorable, with response being negatively related to income (negative selection into response). The authors estimate a wage equation using administrative IRS earnings as the dependent variable for the sample of CPS respondents. Based on these estimates,

Received for publication November 9, 2009. Revision accepted for publication September 6, 2011.

* Bollinger: University of Kentucky; Hirsch: Georgia State University, IZA Bonn.

We thank editor Alberto Albadie and two anonymous referees for comments, as well as participants at seminars at Georgia State, Kentucky, Miami of Ohio, South Carolina, Syracuse, Tennessee, UNLV, Western Ontario, and at meetings of the Econometric Society, Society of Labor Economists (SOLE), and NBER Labor Studies Program Meetings.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00264.

[1] For evidence on earnings imputation and the GED, see Heckman and LaFontaine (2006) and Bollinger and Hirsch (2006).

[2] For a comprehensive review of studies examining measurement error, see Bound, Brown, and Mathiowetz (2001).

they impute earnings for the CPS nonrespondents. Their imputations understate administrative wage and salary earnings of the nonrespondents by .08 log points. The sample included only 561 nonrespondents, and earnings were censored at $50,000. Herriot and Spiers (1975) earlier reported similar results with these data, the ratio of CPS respondent to IRS earnings being .98 and of CPS imputed to IRS earnings being .91. These results suggest a downward bias in estimated earnings based either on samples of respondents or full samples with imputed values for nonrespondents. It is not known whether results from these studies can be generalized outside this survey and time period. The sequential hot deck procedure used in the March survey at that time was primitive, failing to use education as a match variable (Lillard et al., 1986). But the findings suggest the importance of knowing whether there exists nonignorable response bias, particularly given increasing nonresponse rates.

David et al. (1986) conduct a related validation study using the March 1981 CPS matched to 1980 IRS reports. They conclude that the Census hot deck does a reasonably good job in predicting earnings as compared to alternative imputation methods. Their results are based on a broader sample and use of a more detailed Census imputation method than was present in Greenlees et al. (1982). David et al. note the many difficulties in comparing CPS and IRS measures of income, not regarding either measure as a true measure of earnings. They conclude that nonresponse is not ignorable; the earnings structure for respondents provides an unreliable basis for predicting the earnings of nonrespondents. In short, the limited evidence available suggests some degree of nonignorable response bias, possibly reflecting negative selection into response. It is hard to know how results based on March CPS records from thirty or more years ago apply to recent CPS surveys. We are unaware of prior work examining response bias in the widely used monthly ORG earnings files.[3]

Two CPS validation studies have examined the accuracy of proxy responses on earnings, pertinent here since earnings nonresponse is much higher among proxy than self-respondents. Bound and Krueger (1991) conclude that proxies are about as accurate as self-respondents, based on the 1977 and 1978 March CPS, measuring prior year annual earnings, matched to Social Security earnings records (imputed earners are excluded). Mellow and Sider (1983) compare earnings reported in a January 1977 CPS supplement with employer reports on earnings (the survey asked workers the name of their employer) and also conclude that self and proxy reports on earnings are broadly similar.[4] But they are not identical. Proxy reports of wages are lower than self-reports (Mellow & Sider, 1983, table 1), and both are lower than are employer reports. Both groups tend to over-report work hours as compared to employer reports, but proxy respondents do so by less than self-respondents.[5]

In short, there exists little evidence on CPS response bias. That which exists is from validation studies using dated March surveys from years when nonresponse was relatively low. We know of no response bias studies using the CPS ORG, a data source providing advantages over the March CPS for studies of wage determination (Lemieux, 2006), but which has high rates of earnings nonresponse.[6]

III. Data, Proxies, and Earnings Imputation among Nonrespondents

A. *Data*

The analysis uses the CPS Outgoing Rotation Group (ORG) monthly earnings files and the March CPS Annual Social and Economic Supplement (ASEC, previously known as the Annual Demographic File). Wage-level equations are estimated using multiple cross sections pooled across years.

The ORG files used are for January 1998 through December 2008. The ORG earnings supplement includes questions on, among other things, usual earnings at the principal job the previous week, usual hours worked per week in that job, and union status. We create a measure of average hourly earnings as follows. Hourly workers report their straight-time wage rate. For hourly workers who do not report tips, overtime, or commissions (and without an allocated "paid by the hour" flag), the straight-time wage is used as the wage measure. For all other workers, the wage is measured by usual weekly earnings, which includes tips, overtime, and commissions, divided by usual hours worked

[3] The ORGs began in January 1979. The 1973–1978 May CPS earnings supplements, a precursor to the ORGs, did not include imputed earnings values. About 20% of the May 1973–1978 records have missing earnings values, much of this presumed to be the result of nonresponse (Hirsch & Schumacher, 2004). Using recent ORG files, Hirsch and Schumacher estimate a selection wage equation model in which the proxy response variable is employed as the exclusion restriction. The purpose was to provide a robustness check of their union wage gaps estimates obtained by OLS with imputed earners omitted, not to address the more general issue of whether response bias is ignorable.

[4] In regressions of the employer-employee difference in reported wages on typical wage determinants, Mellow and Sider (1983) obtain no significant coefficients.

[5] Papers by Reynolds and Wenger (2012) and Lee and Lee (2012), the former using the CPS and the latter the PSID, show shifts over time in the use of proxies by women and men, affecting trends in the measured and unmeasured components of the gender wage gap.

[6] Korinek, Mistiaen, and Ravallion (2007) examine potential bias from unit rather than item nonresponse on earnings. CPS weights are designed to account for survey nonparticipation that is nonrandom across geographic areas (states) but random within states. Korinek et al. question the latter assumption. They show that response rates across states vary inversely with income, conditional on other covariates, and apply this relationship to adjust weights within states. It seems reasonable that negative selection in response might apply to item nonresponse as well as unit nonresponse. We find earnings response to be substantially higher in rural than in large metropolitan areas. The inverse relationship between response and income then found therefore may reflect in part the large earnings differences across area size if unit as well as earnings nonresponse varies with size. A separate literature considers various methods to deal with missing data (Little, 1988; Ibrahim & Lipsitz 1996; Durrant & Skinner 2006; Egel, Graham, & de Xavier Pinto, 2008). These methods often require strong distributional assumptions and appear to shed little light on whether there is nonignorable response bias in the CPS.

TABLE 1.—CPS IMPUTATION/RESPONSE RATES BY SAMPLE, WAGE MEASURE, SURVEY FRAME, PROXY STATUS, AND YEAR

	ORG		March Surveys	
Sample or Year	N	% Imputed	N	% Imputed
Full sample, unweighted	1,867,388		782,095	18.1%
Wage based on weekly earnings		29.8%	N.A.	N.A.
Wage based on hourly and weekly earnings		31.9%	N.A.	N.A.
Full sample, weighted	1,867,388		782,095	18.9%
Wage based on weekly earnings		31.2%	N.A.	N.A.
Wage based on hourly and weekly earnings		33.3%	N.A.	N.A.
Primary sample, unweighted	1,499,630		564,722	18.7%
Wage based on weekly earnings		30.4%	N.A.	N.A.
Wage based on hourly and weekly earnings		32.7%	N.A.	N.A.
Primary sample, weighted	1,499,630		564,722	19.6%
Wage based on weekly earnings		31.8%	N.A.	N.A.
Wage based on hourly and weekly earnings		34.1%	N.A.	N.A.
Primary sample, all years, weighted	1,499,630	34.1%	564,722	19.6%
1998	120,905	27.2%	40,464	17.2%
1999	126,269	31.0%	41,526	16.6%
2000	128,580	33.3%	40,779	19.6%
2001	136,088	35.0%	65,807	20.2%
2002	145,147	35.0%	63,757	21.4%
2003	142,438	36.4%	62,442	20.7%
2004	139,802	36.1%	61,878	20.8%
2005	141,171	35.7%	62,327	19.1%
2006	141,412	35.7%	62,749	20.0%
2007	139,990	34.9%	62,993	20.1%
2008	137,828	34.6%	N.A.	N.A.
Self-report	756,693	27.8%	281,887	15.2%
Proxy report[a]	742,937	40.5%	282,835	24.0%
Spouse	452,234	34.6%	185,813	18.6%
Nonspouse	290,703	49.0%	97,022	32.9%
February	123,985	30.5%	N.A.	N.A.
March	122,831	29.6%	N.A.	N.A.
January, April–December	1,252,828	34.9%	N.A.	N.A.
First interview	N.A.	N.A.	142,330	17.8%
Later interview	N.A.	N.A.	422,392	20.2%

Full sample includes all persons working during the earnings reference period. Primary sample restricted to persons ages 18 to 65 working full time (year round in ASEC) and not enrolled full time in school. N.A. = not applicable.

[a]Proxy information not available in 1998 March CPS.

per week on the principal job.[7] For workers whose weekly earnings are top-coded in the ORGs (at $2,885), we assign the estimated mean by year and gender above the cap assuming a Pareto distribution above the median.[8]

We use the March CPS for 1999 through 2008, administered to all CPS rotation groups. Earnings (and income) questions apply not to the previous week but to the previous calendar year (1998–2007). The March wage measure is calculated as annual earnings for all wage and salary jobs divided by annual hours worked (the product of week worked and hours worked per week). Industry and occupation designation is based on the longest job held the previous year. Union status is not reported.

In the ORGs and March CPS, we focus on full-time workers between the ages of 18 and 65 who are not enrolled in school full time. In the ORGs, full time is defined as usual hours worked per week on the primary job being at least 35 hours. In the March survey, full-time, full-year workers are defined as those who typically work at least 35 hours per week and were employed at least 50 weeks. These restrictions are meant to avoid issues with respect to selection into part-time work and retirement. These samples, similar to those used in numerous studies of wage determination, are referred to in table 1 as the "primary" samples. The "full" samples include part-time workers (and part-year workers) and no age or enrollment restrictions (apart from age 16 and over).

Rates of earnings nonresponse (%Imputed) in the CPS are shown in table 1. Due to more intensive efforts to contact and acquire responses for the March surveys, nonresponse rates for the ASEC are lower than for the ORG. In recent years, nonresponse in the ORG has been about 30% of the sample versus about 20% in ASEC. Nonresponse is about 1 percentage point higher if one applies employment weights to the sample. This difference results from lower response rates in large metropolitan areas than elsewhere, coupled with a smaller proportion of households sampled (hence larger weights) in such areas.

B. *Proxies*

Census usually interviews one individual (the "reference" person), most often the household head or

[7] For the few workers who do not report an hourly wage and report variable hours, the wage is calculated using hours worked the previous week.

[8] Estimates compiled by Barry Hirsch and David Macpherson are posted at www.unionstats.com. Estimated means above the cap for men (women) increased from 1.65 (1.55) times $2,885 in 1998 to 1.87 (1.68) in 2008.

TABLE 2.—SELF-REPORTS AND PROXY EARNINGS RESPONSES, BY GENDER AND MARITAL STATUS

	ORG Sample			March ASEC Sample		
	All	Men	Women	All	Men	Women
Self-reports	50.5%	42.9%	59.8%	49.9%	42.8%	59.1%
Proxy	49.5%	57.1%	40.2%	50.1%	57.2%	40.9%
Spouse	30.2%	36.6%	22.2%	32.9%	39.5%	24.5%
Nonspouse	19.4%	20.5%	18.0%	17.2%	17.8%	16.4%
%Proxies who are spouse	60.9%	64.1%	55.3%	65.7%	68.9%	59.9%

All results computed without sample weight using the primary sample (see table 1).

cohead, who typically provides responses for all household members. Roughly half of individuals have recorded responses that are self-reported and half responses reported by another household member. Among records based on proxy responses, over half are from a spouse. As seen in table 2, using our ORG primary sample, 57% of male earnings records are based on proxy respondents, 64% of whom are wives. For women, only 40% are based on a proxy, 55% of whom are husbands.

C. *Imputation*

Individuals for whom earnings are not reported have them imputed (allocated) by the Census. Different imputation procedures are used in the ORG and ASEC.[9] Earnings imputation in the ORG uses a "cell hot deck" method that has had only minor changes over time. During the 1998–2002 period, the Census created 14,976 ORG cells representing the possible combinations based on the product of the following seven categories: gender (two cells); age (six); race (two); education (three); occupation (thirteen); hours worked, including whether hours per week are variable (eight); and receipt of tips, commissions, or overtime (two). Occupation categories fell to ten in 2003 when new codes were adopted, reducing the number of hot deck cells to 11,520. Census keeps all cells "stocked" with a single donor, ensuring that an exact match is always found. The donor in each cell is the most recent earnings respondent surveyed previously by the Census with that exact combination of characteristics. As each surveyed worker reports an earnings value, the Census goes to the appropriate cell, removes the previous donor value, and refreshes the cell with a new respondent earnings value. If a cell is not stocked by a matching donor from the current survey month, Census uses donor earnings obtained in prior survey months (or years).

Bollinger and Hirsch (2006) provide analyses of coefficient match bias using the ORGs. The intuition is straightforward. Attributes that are not used in the imputation procedure are largely uncorrelated with imputed earnings. The wage equation coefficients estimated for these attributes are thus a rough weighted average of a value close to 0 and the true coefficient, the implicit weights being the respective proportions of observations that are and are not imputed. Attenuation of a union coefficient in their full sample exceeds 25%, nearly as large as the 28.7% of the sample imputed. Similar attenuation is found for coefficients on foreign born, marriage, Hispanic status, and others, as well as for dispersion in coefficients for industry, region, and city size dummies. Complex forms of bias are found for coefficients on imperfectly matched attributes such as schooling, age, and occupation.

The CPS-ASEC uses a sequential hot deck imputation procedure. Nonrespondents are matched to donors from within the same March survey in sequential steps, each step involving a less detailed match requirement. The procedure first attempts to find a match on the exact combination of variables using the full set of match characteristics (similar to those used in the ORG). Absent a successful match at that level, matching advances to a new step with a less detailed breakdown, for example, broader occupation and age categories. As Lillard et al. (1986) emphasized, the probability of a close match declines the less common an individual's characteristics.

Bollinger and Hirsch (2006) examine alternative estimation procedures to correct for match bias, the simplest being estimation based on the sample of respondents.[10] Suggested corrections, however, rely on the assumption that earnings are conditional missing at random, that is, response bias is ignorable. Thus, a principal contribution of this paper is the guidance it provides on how to deal with imputed earners and match bias. If response bias is largely ignorable, match bias can be easily addressed. If response bias is nonignorable, more nuanced implications follow.

IV. Who Fails to Report Earnings?

In this section, we examine correlates of earnings nonresponse, focusing on variables that might provide an exclusion restriction in a selection model—determinants of response not correlated with a wage equation error term. For both the ORG and ASEC, we consider use of proxy

[9] Details on ORG imputation procedure are provided by Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006). Lillard et al. (1986) provide a detailed discussion of the March imputation method.

[10] Other approaches include inverse probability weighting (IPW) of the respondent sample to correct for its changed composition due to nonresponse and estimation using the full sample coupled with application of a (complex) bias correction formula for the estimated coefficients.

Table 3.—Marginal Effects of Potential Selection Identifiers in Probit Response Model

	ORG		ASEC	
	Male	Female	Male	Female
Nonspouse proxy	−0.238*	−0.254*	−0.201*	−0.195*
Spouse proxy	−0.0618*	−0.0818*	−0.0385*	−0.0196*
February	0.0434*	0.0411*	N.A.	N.A.
March	0.0500*	0.0461*	N.A.	N.A.
Month in sample 1 or 5	N.A.	N.A.	0.0229*	0.0251*
Sample size	827,531	672,099	318,119	246,603

Dependent variable = 1 if respondent. Unweighted estimates shown (weighted estimates available on request). Other variables and coefficients included are potential experience in quartic form and detailed dummies for education, marital status, race and ethnicity, foreign-born status, metropolitan area size, region, public sector, industry, occupation, year, and (in the ORG) union status. Complete results are posted at the authors' websites: (Bollinger: http://gatton.uky.edu/faculty/bollinger/; Hirsch: http://www2.gsu.edu/hirsch). Significant at *1%. N.A.= not applicable.

variables. For the ORGs, we consider calendar month of the survey and for ASEC the CPS rotation group.[11]

As seen in table 1, nonresponse rates in the ORG are 27.8% among earnings records based on self-reports and 40.5% among records relying on proxies. For the latter group, nonresponse is 34.6% when the proxy is a spouse but a far higher 49.0% otherwise. A similar pattern is found in the March supplements, where nonresponse rates are 9 percentage points higher for proxy than self-respondents.

For the ORGs, we conclude that dummies for survey months February and March are attractive exclusion restrictions. Nonresponse rates of about 30% seen in the February and March ORG interviews are substantially lower than the 34.9% average rate the rest of the year (there is little variation in rates across the other ten months). Moreover, earnings are not found to differ in February and March from other months, conditional on other covariates. Discussion with personnel at the Bureau of Labor Statistics revealed that enumerators are evaluated based largely on interview performance at that time of year. This coincides with the March ASEC being in the field and is done to ensure higher responses and more diligence during the ASEC. We speculate that enumerators do not distinguish among the various parts of the survey, so additional effort affects response rates for all aspects of the survey. Consistent with this explanation is the higher earnings response rate seen for ASEC than for the ORG.

Alternative explanations for the February and March differences exist, although we find them less convincing. At that time, household members are more likely aware of income amounts because of tax documents, leading to a higher response rate in the February and March ORG (indeed, ASEC is administered early in the year because it is during tax season). Knowledge of tax documents, however, is less critical for the ORG than for ASEC since ORG questions concern hours worked and rates of pay at the principal job during the prior week, not earnings from the prior calendar year. We also considered whether the ORG response rates might be affected by seasonal factors (such as bad weather) that reduce participation costs and improve earnings response during February and March, but monthly response patterns were found to be highly similar across states with very different seasonal weather patterns.

Turning to ASEC, households in either their first or fifth month in the sample display nonresponse rates about 2 percentage points lower than in the other six months. The first and fifth month interviews, which take place the same month one year apart, typically are done in person (CAPI), whereas rotation groups 2 to 4 and 6 to 8 in the months following the first and fifth month interviews are administered by phone (CATI). It is reasonable to assume, and is generally accepted in the survey literature (see, for example, Lyberg & Kasprzyk, 2004), that use of an in-person interviewer results in a higher earnings response.

Because of space constraints, we summarize but do not provide descriptive evidence on the correlates of response or the coefficients on the controls included in our probit selection models (coefficients on the potential selection identifiers are shown in table 3).[12] In both the ORG and March data, response is less likely for those over age 55. Respondents are more likely to have college degrees, and nonrespondents are more likely to have their highest educational attainment be high school graduation. Response among women exceeds that for men. Respondents are more likely to be white, while nonrespondents are more likely black or Asian. Workers residing outside of metropolitan areas are most likely to be respondents, while those who live in the largest metropolitan areas are least likely to respond. None of these differences is particularly large. Not surprisingly, those who do not report earnings demonstrate much higher nonresponse rates for such variables as industry, occupation, and union status.

V. Estimation Models

We begin with a standard log linear model of wages:

$$\ln\text{Wage} = w = X\beta + u.$$

Given our large sample, we choose a rich set of regressors including fourth-order polynomial in potential experience, plus multiple categorical variables for education, marital status, race, immigrant status, metropolitan size, census region, public sector, two-digit industry, and occupation categories and, in the ORG, union status. Although the genesis of the Mincerian wage equation is as a supply-side

[11] Nicoletti and Peracchi (2005), based on analysis of the European Community Household Panel, provide evidence justifying inclusion of variables that characterize the data collection process in models of response, while excluding them from the outcome model of interest. We also examined using as identifier variables information from CPS supplements on voting behavior and volunteer activity, expecting that "public spirit" might increase the likelihood of survey response but be uncorrelated with the wage. Volunteer activity but not voting was found to be associated with higher earnings response. Each of these potential identifiers was significantly correlated with the wage.

[12] Tables can be viewed in the online supplement and are available from the authors.

human capital model, as employed here it should be regarded as a reduced-form equation including demand as well as supply-side wage determinants. In conjunction, we posit a threshold crossing model of nonresponse,

$$R = \begin{bmatrix} 1 & \lambda w + Z\delta + v > 0 \\ 0 & \text{otherwise} \end{bmatrix},$$

where w is the labor market log wage, Z represents all observable characteristics, including those in the wage equation, and ν are unobservable terms independent of both the determinants of the wage and variables in Z. The term λ allows this model to be linked to the wage equation with either positive (response correlated with high wage) or negative (response correlated with low wage) selection. By substituting the wage equation into the above model, we establish a reduced-form model for response:

$$R = \begin{bmatrix} 1 & Z\gamma + \varepsilon > 0 \\ 0 & \text{otherwise} \end{bmatrix}.$$

The parameter, $\gamma = \lambda\beta + \delta$, while $\varepsilon = \lambda u + v$. We further impose the assumption of standard normality on ε and require that ε be strictly independent of components of Z for which the corresponding γ term is not 0. We recognize that these are strong assumptions. Consistent estimation of selection models using Heckman's two-step approach typically requires these assumptions. While it may be possible to relax them, the computational burden, given our large sample sizes, becomes problematic. The two-step approach is well known (see Vella, 1998) to be less sensitive to violations of normality and strict independence than maximum likelihood approaches.[13]

We first turn to estimates from the reduced-form response probits. The marginal effects (evaluated at the mean of all variables) are shown in table 3 for the sample frame variables we consider as potential identifier variables in Z. Other results are not shown, but are available in the online supplement. Variables with large marginal effects include black, Asian, large metropolitan, and selected regions. The reported estimates do not use sample weights (differences are minor). Because the weights do not account for sample selection, there is not a strong conceptual argument for using weights in the selection-corrected wage equation or in the corresponding first-stage probit. Qualitatively, results are largely comparable for men and women and across the ORG and March samples despite some differences in regressors. Marginal effects are generally larger for the ORG than for the March survey due to higher ORG nonresponse.

[13] Other authors have used selection models to analyze nonresponse data sets other than the CPS. For example, Hamermesh and Donald (2008) consider a selection model for earnings in a survey of college graduates. De Luca and Peracchi (2007) consider a selection model for unit and item nonresponse in a study estimating Engel curves for consumption expenditures. Johansson (2007) considers alternative methods, including sample selection, to address nonresponse in Swedish data.

The multivariate probit analysis reinforces support for the potential selection identifiers shown previously in table 1. As evident in table 3, the proxy and the interview timing variables (MIS for the March ASEC and February/March for the ORG) are good potential exclusion restrictions for the selection models. Proxy respondents are substantially less likely to respond to the earnings questions. All else constant, a proxy respondent other than a spouse decreases the likelihood of response by about 20%, while a spouse proxy decreases response by somewhat more than 5% in the ORG and less than 5% in the March sample. In the ORG, response rates in February and March are roughly 5 percentage points higher than during the rest of the year. And in the ASEC data, response rates are about 2 percentage points higher for people in their first of four months in the survey during each of two years (rotation groups 1 and 5).

VI. Evidence for Selection into Response: Significance and Importance

In order to investigate whether nonresponse in the CPS is ignorable, we begin by estimating rich log-linear wage models of the type seen in the literature, both with and without imputed earners. As Bollinger and Hirsch (2006) emphasized, inclusion of imputed earners leads to severe coefficient match bias even if nonresponse is random. For example, absent inclusion of imputed earners, the male sample coefficient on an associate degree (relative to high school) is 0.127. When imputations are included, the OLS estimate falls to 0.093, reflecting the fact that nonrespondents with an associate degree are assigned the earnings of donors with education ranging between high school (including the GED) and some college short of a B.A. If nonresponse is neither random nor ignorable, OLS coefficients without imputed earners included also will be biased. Yet including imputations in an OLS equation is not a valid solution for response bias since the imputations are simply predicted values from respondents.

The top half of table 4 presents wage equation estimates for the selection-related variables, separately for men and women and for the ORG and ASEC primary samples. The first column for each data set shows OLS estimates based on respondents only. The second and third columns present the wage model estimated using the two-step Heckman correction with the coefficient on the inverse Mills ratio reported in the first row. The selection models in all cases rely on the sample-based identifier variables—February and March dummies for the ORG and a first interview dummy (rotation group 1 or 5) for ASEC. The second column results are based on the use of the proxy variables as additional selection identifiers, with inclusion of the proxy variables in the selection but not wage equations. The third column includes the proxy variables as wage regressors. Proxy is such a strong predictor of response that it is natural to consider it to identify the selection model, given that it has no causal impact on realized (as opposed to reported) earn-

Table 4.—Wage Equation Selection Effect Estimates for Men and Women Using CPS ORG and ASEC

	ORG			ASEC		
	OLS	Selection	Selection	OLS	Selection	Selection
Male primary sample						
Inverse Mills ratio	–	−0.167*	−0.166*	–	−0.267*	−0.276*
Nonspouse proxy	–	–	−0.002	–	–	0.00008
Spouse proxy	–	–	0.008*	–	–	0.012*
Intercept	2.436*	2.515*	2.516*	1.802*	1.921*	1.925*
Sample size (OLS respondents only)	553,727	827,531	827,531	258,552	318,119	318,119
Female primary sample						
Inverse Mills ratio	–	−0.114*	−0.034	–	−0.142*	−0.062
Nonspouse proxy	–	–	−0.037*	–	–	−0.024
Spouse proxy	–	–	0.010*	–	–	0.023*
Intercept	2.310*	2.355*	2.345*	1.673*	1.728*	1.711*
Sample size (OLS respondents only)	454,991	672,099	672,099	200,826	246,603	246,603
Male head/cohead sample						
Inverse Mills ratio	–	−0.078*	−.095*	–	−0.108*	−0.089
Nonspouse proxy	–	–	0.005	–	–	−0.008
Spouse proxy	–	–	0.003	–	–	0.003
Intercept	2.454*	2.480*	2.485*	1.807*	1.841*	1.837*
Sample size (OLS respondents only)	470,354	681,555	681,555	223,939	269,093	269,093
Female head/cohead sample						
Inverse Mills ratio	–	−0.069*	−.004	–	−0.069*	−0.086
Nonspouse proxy	–	–	−.038*	–	–	0.002
Spouse proxy	–	–	.006*	–	–	0.024*
Intercept	2.340*	2.358*	2.347*	1.702*	1.730*	1.711*
Sample size (OLS respondents only)	393,078	565,387	565,387	174,322	209,989	209,989

Estimates are unweighted. The wage equations include potential experience in quartic form and detailed dummies for education, marital status, race and ethnicity, foreign-born status, metropolitan area size, region, public sector, industry, occupation, year, and (in the ORG) union status. Complete results are posted at the authors' websites (see table 3). Significant at *1%.

ings. Our concern is that the proxy measures may be correlated with the wage equation error term if proxy respondents report higher or lower earnings than do self-respondents.

We first examine the coefficients on the inverse Mills ratio selection terms. The coefficients on the Mills ratios for men using the ORG and ASEC are negative, highly significant, and quite stable with respect to the inclusion or exclusion of the proxy variables as regressors in the wage equation (results in columns 2 and 3 are similar). Based on these results, we conclude that men exhibit negative selection into response, consistent with earlier research based on men in the 1973 March CPS matched to 1972 IRS records (Herriot & Spiers, 1975; Greenlees et al., 1982). We also note that nonspouse proxy responses have no apparent correlation with unobservable wage determinants, while proxy reports by wives have a very small positive correlation, reported earnings being about 1% higher. Proxy indicators can serve as a reasonable selection identifier variable for CPS male wage equations.

In contrast to the results for men, the inverse Mills ratios for women seen in table 4 are sensitive to the use of the proxy variables in the wage equation. In regressions with the proxy variables used as an exclusion restriction, negative and significant coefficients on the inverse Mills ratios are obtained in both the ORG and ASEC. When the proxy variables are included in both the response and wage equations, the inverse Mills ratio coefficients become insignificant and small in magnitude, although they remain negative. As with men, a spouse proxy response is correlated with a slightly higher reported wage. The nonspouse proxy response in the ORG is associated with 3.7% lower reported wage for women, while in the ASEC, the effect is 2.4% lower (although not statistically significant). For women, partial correlation of the proxy variables with the wage, coupled with changes in the inverse Mills results, appears sufficient to reject using them as exclusion restrictions. Whereas results for men clearly indicate negative selection into response, the evidence for women is weaker.

Because selection into response may differ substantially across different populations, we separately examine a sample that is restricted to household heads and coheads (principal householders), a sample that includes primary individuals, individual heads with relatives, and husbands and wives. Excluded are children of principal householders, partners or roommates, and all other relationships. This grouping was done with a focus on the largest categories and their relative imputation rates. Heads and coheads have substantially lower nonresponse rates than the other groups, with partners and roommates second, followed by other relations and adult children living with their parents.

In the lower half of table 4, we provide estimation results from the samples of heads and coheads. Switching from our primary samples to the head and cohead samples sharply reduces estimates of negative selection, as measured by the inverse Mills coefficients. In both the ORG and March samples of men, the sample selection coefficients fall sharply in absolute value, from −0.166 to −0.095 in the ORG and from −.276 to −.089 in the March CPS. There remains clear evidence for negative selection, but it is smaller for heads than for other household males. Among women, coefficients remain negative but are statistically insignifi-

Table 5.—Mean Log Wage Differences from OLS versus Selection Estimates

	(1) Full Sample Means	(2) Respondent Means	(3) Full Sample Means, OLS Respondent Coefficients	(4) Selection Coefficient Means	(5) Overall Bias (3) − (4)
ORG primary sample					
Male	2.973	2.982	2.981	3.069	−0.088
Female	2.781	2.792	2.790	2.806	−0.016
M-F difference	0.192	0.190	0.191	0.262	−0.072
ASEC primary sample:					
Male	3.014	3.021	3.016	3.105	−0.089
Female	2.769	2.780	2.776	2.795	−0.019
M-F difference	0.244	0.242	0.240	0.310	−0.070
ORG head and cohead sample					
Male	3.042	3.051	3.056	3.104	−0.048
Female	2.816	2.827	2.827	2.828	−0.001
M-F difference	0.225	0.224	0.229	0.276	−0.047
ASEC head and cohead sample					
Male	3.092	3.089	3.091	3.117	−0.026
Female	2.808	2.812	2.812	2.838	−0.026
M-F difference	0.284	0.277	0.279	0.279	0.000

Column 1 presents overall mean log wages inclusive of nonrespondents' imputed wages. Column 2 presents means for respondents only. Column 3 presents predicted mean earnings for all observations (including nonrespondents) using coefficients from the OLS respondent-only model. Column 4 reports mean earnings predicted using the selection models that include all observations, reported in column 3 of table 4. The selection term is not used in the prediction; hence, this represents the estimated mean of all wages were they to be reported. The difference between the two (column 5) provides a measure of the bias due to selection into response.

cant. An implication of these results is that selection effects from nonignorable nonresponse in the CPS can be reduced by limiting samples to household heads and coheads. The obvious downside is that the narrower sample is no longer representative of the larger working population.

We next examine the practical importance of selection on coefficient estimates. Because of large samples, trivial differences in coefficients can be statistically significant. We instead focus on the size of coefficient differences between wage equations with and without accounting for selection into response (column 3 selection results using the sampling frame but not proxy dummies as identifiers versus column 1 OLS results based on respondents only). The key result of these comparisons is that changes in slope coefficients are quite minor.[14] Consider the coefficient for associate degree mentioned earlier. When the response selection correction is included, the coefficient on associate degree becomes 0.120 compared to 0.127 for the uncorrected respondent-only sample and 0.093 for the sample with imputations included. Bias from including the imputations is far more severe than bias from failing to correct for selection once imputations are excluded.[15] Changes in coefficients are noticeable (but not large) only for the variables most highly correlated with earnings nonresponse, for example, Asian, black, large metropolitan, and several regions. In all cases, OLS estimates with imputations included appear far more biased than OLS estimates from the respondent sample, as compared to estimates from selection-corrected models.

Although response bias has little effect on wage equation slope estimates, this need not imply that selection into response is not substantive. To assess the magnitude of response bias, we compare predicted earnings based on both the OLS coefficients for respondents and the selection-corrected estimates. These results are seen in table 5. The first column presents overall mean log wages in the ORG and ASEC samples, inclusive of the nonrespondents' imputed wages, while the second column presents means for respondents only. The third column presents predicted mean earnings using coefficients from the OLS respondent model but for all observations, including nonrespondents. The fourth column reports the mean earnings prediction using the selection models reported in column 3 of table 4 that include all observations. The selection term is not used in the prediction; hence, this represents the estimated mean of all wages were they to be reported.[16] The difference between the two provides a measure of the magnitude of bias due to selection into response.

Focusing first on the primary sample results, we find that the difference for men is sizable in both the ORG and ASEC. Negative selection into response among men is predicted to result in average earnings being understated by 9%. Estimated downward bias in earnings for women is much smaller (about 2%), as expected given the weak evidence among women of selection bias. Taken at face value, the implication is that conventional estimates of the gender gap in earnings are understated by some 7%. Additionally,

[14] As emphasized previously, OLS coefficient estimates from a sample including imputed earners would differ substantially from those shown in table 4 as a result of imputation match bias.

[15] Full results for all estimated wage equations are available in the online supplement or from the authors. Some coefficients, such as those on bachelor's degree, are not affected very much at all (see Bollinger & Hirsch, 2006, for a full explanation). In this case, the estimate with imputations is 0.296, while the estimate using respondents is .301 and the selection-corrected estimate is .294.

[16] Selection predicted means for respondents, not shown in the table, are highly similar to those shown in column 4.

increases in imputation over time may have caused narrowing of the gender gap to be overstated. Importantly, whatever the biases are due to nonresponse, these show up mainly as differences in the intercepts and not slopes, the latter typically being the principal concern of researchers. We acknowledge that the estimated intercepts may rely more highly on the selection correction normality assumption than do the slope estimates.

Restricting the sample to heads and coheads appears to reduce response bias, with negative selection less severe and male-female differences small or even 0 (table 5, bottom portion). In the ORG sample of heads and coheads, the gender gap is estimated to be understated by 4.5%, smaller than the full sample estimate but still of concern. In the ASEC head and cohead sample, nonresponse bias is equal for men and women, resulting in no bias in gender gap estimates.

To recap, our conclusion is that selection into response is negative for men, perhaps substantially so, and modest for women. Regression coefficients, apart from intercepts, are not sensitive to selection, with the exception of those on variables highly correlated with nonresponse. Negative selection among men is less severe when the sample is restricted to household heads and coheads. More broadly, there appears to be considerable heterogeneity in the degree of nonignorable response bias.[17]

VII. Conclusion

Earnings nonresponse and imputation are common in the CPS. We examine the issue of response bias on earnings using the CPS ORG monthly earnings files and March CPS ASEC for 1998–2008. Although wage studies by labor economists typically include imputed earners, their inclusion introduces substantial bias due to mismatch in the imputation process. Simple corrections for match bias, including removal of imputed earners from the estimation sample, largely eliminate the first-order distortions resulting from imperfect matching. But this and other approaches to correct for match bias rest on the assumption that response bias is ignorable (see Bollinger & Hirsch, 2006). Absent a definitive validation study based on recent matched CPS household and administrative earnings records, it is difficult to know with great certainty the existence and degree of nonignorable response bias.

Nonetheless, using selection wage equations in which selection is identified by measures on the timing of surveys, we find clear-cut evidence of negative selection into response among men but much weaker evidence among women. Understatement of men's earnings due to nonresponse, coupled with a small effect on women's earnings, results in an understatement of the gender wage gap. The response bias is largely a fixed effect, introducing bias into estimates of wage equation intercepts but not slopes, with the exception of a few attributes most highly correlated with nonresponse. Negative selection among men and bias in the gender gap are far less evident when samples are restricted to heads and coheads. For empirical labor economists, a key conclusion is that for wage analyses in which the principal interest is slope coefficient estimates, omitting imputed earners from OLS wage equations is generally sufficient to avoid major bias.

A final point warrants repeating: even were selection bias nonignorable and severe, inclusion of imputed earners is not a solution. Imputations are based entirely on respondent donors and generated under the assumption of conditional missing at random. Their inclusion in OLS wage regressions does not alleviate response bias but does introduce substantial match bias in coefficient estimates.

[17] In results not reported, we find differences in selection by race. In contrast to the full male or white male samples, black men exhibit little if any selection (small and insignificant inverse Mills ratios). Black women exhibit substantial positive selection into response, in contrast to no or weak negative selection found in the full female and white female samples. Because study of racial differences is beyond the scope of the paper and the white-only and full samples produce similar results, we present results from the combined samples throughout the paper. We also estimate selection models for different portions of the predicted earnings distribution: the bottom 25th, middle 50th, and top 25th percentiles, plus the top 10th and top 5th percentiles. For men, we find negative selection throughout the predicted earnings distribution, the magnitude being quite modest over most of the distribution but with stronger effects in the top percentiles. For women, there is no clear-cut pattern, with estimates of positive as well as negative selection into response, particularly in the top and bottom tails of the distribution.

REFERENCES

Bollinger, Christopher R., and Barry T. Hirsch, "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching," *Journal of Labor Economics* 24 (2006), 483–519.

Bound, John, Charles Brown, and Nancy Mathiowetz, "Measurement Error in Survey Data" (pp. 3705–3843), in E. E. Leamer and J. J. Heckman (eds.), *Handbook of Econometrics*, vol. 5 (Amsterdam: Elsevier, 2001).

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

David, Martin, Roderick J. A. Little, Michael E. Samuhel, and Robert K. Triest, "Alternative Methods for CPS Income Imputation," *Journal of the American Statistical Association* 81 (1986), 29–41.

De Luca, Giuseppe, and Franco Peracchi, "A Sample Selection Model for Unit and Item Nonresponse in Cross-Sectional Surveys," CEIS working paper 99 (2007), http://ssrn.com/abstract=967391.

Durrant, Gabriele B., and Chris Skinner, "Using Data Augmentation to Correct for Non-Ignorable Non-Response When Surrogate Data Are Available: An Application to the Distribution of Hourly Pay," *Journal of the Royal Statistical Society A* 169: pt. 3 (2006), 605–623.

Egel, Daniel, Bryan S. Graham, and Cristine Campos de Xavier Pinto, "Inverse Probability Tilting and Missing Data Problems," NBER working paper 13981 (April 2008).

Greenlees, John, William Reece, and Kimberly Zieschang, "Imputation of Missing Values When the Probability of Response Depends on the Variable Being Imputed," *Journal of the American Statistical Association* 77 (1982), 251–261.

Hamermesh, Daniel S., and Stephen G. Donald, "The Effect of College Curriculum on Earnings: An Affinity Identifier for Non-Ignorable Non-Response Bias," *Journal of Econometrics* 144 (2008), 479–491.

Heckman, James J., and Paul A. LaFontaine, "Bias Corrected Estimates of GED Returns," *Journal of Labor Economics* 24 (2006), 661–700.

Herriot, R. A., and E. F. Spiers, "Measuring the Impact on Income Statistics of Reporting Differences between the Current Population Survey and Administrative Sources" (pp. 147–158), in *Proceedings of the American Statistical Association Social Statistics Section* (Alexandria, VA: American Statistical Association, 1975).

Hirsch, Barry T., and Edward J. Schumacher, "Match Bias in Wage Gap Estimates due to Earnings Imputation," *Journal of Labor Economics* 22 (2004), 689–722.

Ibrahim, Joseph G., and Stuart R. Lipsitz. "Parameter Estimation from Incomplete Data in Binomial Regression When the Missing Data Mechanism Is Nonignorable," *Biometrics* 52 (1996), 1071–1078.

Johansson, Fredrik, "How to Adjust for Nonignorable Nonresponse: Calibration, Heckit or FIML?" Uppsala University working paper 2007:22 (2007).

Korinek, Anton, Johan A. Mistiaen, and Martin Ravallion, "An Econometric Method of Correcting for Unit Nonresponse Bias in Surveys," *Journal of Econometrics* 136 (2007), 213–235.

Lee, Jungmin, and Sokbae Lee, "Does It Matter Who Responded to the Survey? Trends in the U.S. Gender Earnings Gap Revisited," *Industrial and Labor Relations Review* 65 (2012), 148–160.

Lemieux, Thomas, "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill," *American Economic Review* 96 (2006), 461–498.

Lillard, Lee, James P. Smith, and Finis Welch, "What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation," *Journal of Political Economy* 94 (1986), 489–506.

Little, Roderick J. A., "Missing Data Adjustments in Large Surveys," *Journal of Business and Economic Statistics* 6 (1988), 287–296.

Lyberg, Lars E., and Daniel Kasprzyk, "Data Collection Methods and Measurement Error: An Overview" (pp. 237–257), in Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman (eds.), *Measurement Errors in Surveys* (Hoboken, NJ: Wiley, 2004).

Mellow, Wesley, and Hal Sider, "Accuracy of Response in Labor Market Surveys: Evidence and Implications," *Journal of Labor Economics* 1 (1983), 331–344.

Nicoletti, Cheti, and Franco Peracchi, "Survey Response and Survey Characteristics: Microlevel Evidence from the European Community Household Panel," *Journal of the Royal Statistical Society A* 168: pt. 4 (2005), 763–781.

Reynolds, Jeremy, and Jeffrey B. Wenger. "He Said, She Said: The Gender Wage Gap according to Self and Proxy Reports in the Current Population Survey," *Social Science Research* 41 (2012), 392–411.

Vella, Francis, "Estimating Models with Sample Selection Bias: A Survey," *Journal of Human Resources* 33 (1998), 127–169.