

IZA DP No. 5512

**Does It Matter Who Responded to the Survey?
Trends in the U.S. Gender Earnings Gap Revisited**

Jungmin Lee
Sokbae Lee

February 2011

Does It Matter Who Responded to the Survey? Trends in the U.S. Gender Earnings Gap Revisited

Jungmin Lee

*Sogang University
and IZA*

Sokbae Lee

*Seoul National University,
CeMMAP and IFS*

Discussion Paper No. 5512

February 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Does It Matter Who Responded to the Survey? Trends in the U.S. Gender Earnings Gap Revisited*

Blau and Kahn (JOLE, 1997; ILRR, 2006) decomposed trends in the U.S. gender earnings gap into observable and unobservable components using the PSID. They found that the unobservable part contributed significantly not only to the rapidly shrinking earnings gap in the 1980s, but also to the slowing-down of the convergence in the 1990s. In this paper, we extend their framework to consider measurement error due to the use of proxy/representative respondents. First, we document a strong trend of changing gender composition of household-representative respondents toward more females. Second, we estimate the impact of the changing gender composition on Blau and Kahn's decomposition. We find that a non-ignorable portion of changes in the gender gap could be attributed to changes in the self/proxy respondent composition. Specifically, the actual reduction in the gender gap can be smaller than what the estimates without taking into account the measurement error might suggest. We conclude that a careful validation study would be necessary to ascertain the magnitude of the spurious measurement error effects.

JEL Classification: J3

Keywords: gender earnings gap, survey response error, proxy response

Corresponding author:

Jungmin Lee
School of Economics
Sogang University
Shinsu-dong, Mapo-gu
Seoul, Korea 121-742
E-mail: junglee@sogang.ac.kr

* We would like to thank David Card, Daniel Hamermesh, participants at the 2008 SOLE meetings, and an anonymous referee for helpful comments. We also thank Lawrence Kahn for providing the data. Jungmin Lee's research was supported in part by the W.E. Upjohn Institute mini-grant and by the Sogang University Research Grant of 2010. Sokbae Lee's research was supported in part by the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice (RES-589-28-0001) and by the European Research Council research grant (ERC-2009-StG-240910-ROMETA).

1 Introduction

This paper examines trends in the U.S. gender earnings gap. Blau and Kahn (1997, 2006) decomposed the trends into changes in observable and unobservable components using the Michigan Panel Study of Income Dynamics (PSID, hereafter). They found that the unobservable gap significantly contributed to the rapidly shrinking earnings gap in the 1980s as well as to the slowing-down of the convergence in the 1990s. We extend their framework to take into account the potential bias due to non-classical measurement error in the PSID hourly earnings variable. In particular, we focus on the “proxy effect” on reported earnings—reporting discrepancies between self and proxy respondents—and changing gender composition of household representative respondents toward more females from 1980 to 1999.

<<< **Figure 1** >>>

While it is uncertain whether self- or proxy-reported data are better quality (Moore, 1988), it is clear that reporting discrepancies by the type of respondents should matter when we estimate trends in the gender earnings gap using the PSID.¹ The PSID, like most other household surveys, uses *household representative* respondents.² That is, a single person, who is usually the household head, provides information for all other members as well as for himself or herself. Figure 1 shows the proportion of head respondents among *male-headed married households* from 1973 to 2005. It is notable that there have been continuously fewer male/head respondents and more female/spouse respondents.³ In 1973, when the PSID started to replace personal interview with telephone interview, almost 85% of household representative respondents were male. However, in 1990, 57% of respondents were male.⁴ In 1999, only about 46% of representative respondents were male heads. One

¹As Blau and Kahn (2006) pointed out, the PSID is “the only nationally representative data base that contains information on workers’ actual labor market experience,” which is “an extremely important factor” in explaining the gender earnings gap.

²This is because this practice can save a substantial amount of survey costs.

³In the PSID, almost all proxy respondents are spouses. Since we focus on male-headed married households, more proxy respondents mean more female respondents.

⁴In 1976 and 1985, the PSID required, if possible, self response for heads and “wives.” This indicates that the PSID survey designers were also concerned about proxy-responded data quality. In his study on

possible explanation is that the PSID administrator initially put effort to have household heads as representative respondents but started to relax the rule later. Surprisingly, to our best knowledge, this striking fact and its potential impacts on data quality have never been examined before.

For male-headed married households, the trend in Figure 1 means more self-responded data for wives or women.⁵ On the contrary, the same trend means, for male heads, less self-responded data. In this paper, we focus on this monotonic trend of changing gender composition of respondents toward females and point out that the estimates for the trends in the gender earnings gap could be biased even if there is only a small difference in measurement error between self and proxy respondents. For example, suppose that male self respondents tend to report their own earnings higher (not necessarily more or less accurately) than their proxy respondents do. Then, when there are more and more proxy respondents for male earnings, without accounting for the self-proxy reporting gap, we will overestimate the actual convergence rate of the gender earnings gap.

One empirical question is whether there is any measurement gap in hourly earnings between self and proxy respondents in the PSID. In general, the relative quality of proxy-reported data is *a priori* unknown. The results vary by specific surveys and variables.⁶ It might seem to be intuitively appealing that the best information about a person can be obtained from the person himself or herself. This is, however, not always true. Proxy-responded data can be more accurate in many cases. For example, people would be reluctant to answer certain questions, particularly about sensitive personal issues, in which case other household members, if reasonably informed, may provide more accurate data.

The quality of reported earnings by the type of respondents, self or proxy, has not been studied well. Most existing studies examine measurement error in self-reported earnings by comparing them with external and more objective data (such as company payroll

married women's labor supply, Mroz (1987) used the 1976 PSID data because of the same concern about proxy-responded data quality.

⁵This is not universally true for the entire sample we use in this paper because our sample includes unmarried individuals.

⁶Todorov (2003) demonstrates that differences between self- and proxy reports can be systematically related to measures of cognitive inferences.

or administrative records).⁷ One earlier exception is Hill (1987). He found a substantial reporting bias in the SIPP earnings by the type of respondent even after controlling for self-selectivity into respondents. Recently, using the CPS panel, Bollinger and Hirsch (2009) provided fixed-effect estimates that imply that both spouse and non-spouse proxy respondents report earnings about 2% less than do self respondents.⁸ It seems that there exist some proxy effects while the magnitude is likely to be small.

Very little is known for proxy effects in the PSID earnings. The quality of the PSID earnings has been examined in the context of measurement error literature (Duncan and Hill, 1985; Duncan and Mathiowetz, 1988; Rodgers, Brown, and Duncan, 1993; Bound, Brown, Duncan, and Rodgers, 1994; Pischke, 1995). The studies have found that reported earnings are not significantly different from validation (“true”) earnings in terms of mean and variance, although measurement error violates the assumptions of the classical error-in-variables model. However, they did not examine whether measurement error in proxy-reported earnings is worse than or different from that in self-reported earnings. Furthermore, the validation samples consist of predominantly males. Thus, they cannot examine whether the respondent’s gender makes a difference in measurement error.

The effects of measurement error on earnings gaps have been studied in different contexts. For example, Bollinger (2001) examined the effect of measurement error in reported union status on the union wage differential, and also Bollinger (2003) investigated the impact of measurement error in human capital accumulation on the black-white wage gap. Although measurement error is likely to be different between self and proxy respondents and also by their gender, to our best knowledge, it has not been examined whether the measurement error gap between self and proxy respondents may bias the estimated gender earnings gap.

The remainder of this paper is organized as follows. Section 2 presents our econometric framework and methodology. Section 3 describes our data and estimation results. Conclud-

⁷Some studies compared self and proxy data, although it was not their main focus. They found that proxy-reported data are not substantially different from self-reported data (Bound and Krueger, 1991; Mellow and Sider, 1983).

⁸It has been found that women are better reporters for their own earnings than men are for their own earnings (Bound and Krueger, 1991; Bollinger, 1998, Kristensen and Westergaard-Nielsen, 2007).

ing remarks are given in Section 4.

2 Analytic Framework

We build our analytic framework on Blau and Khan (1997), who used the decomposition method suggested by Juhn, Murphy and Pierce (1993) to analyze trends in the U.S. gender pay gap in the 1980s. In particular, we add the measurement error to their framework. As in Blau and Khan (1997), we start with a male wage equation:

$$Y_{it}^* = X_{it}B_t + \sigma_t\theta_{it}, \quad (1)$$

where Y_{it}^* is the logarithm of measurement-error-free wages, X_{it} is a vector of explanatory variables, B_t is a vector of corresponding coefficients, θ_{it} is a standardized residual for male log wages with mean zero and variance one for each year and σ_t is the residual standard deviation of male log wages.

Suppose that one only observes

$$Y_{it} = Y_{it}^* + U_{it}, \quad (2)$$

where U_{it} is the measurement error. For most years of the PSID, we observe individual earnings reported by one respondent for each household. Therefore, it is natural to consider different measurement errors depending on who responded the survey. Suppose that the measurement error U_{it} has the form

$$U_{it} = \mu_t + \delta_t R_{it} + \gamma_t V_{it}, \quad (3)$$

where R_{it} is a dummy variable that takes value one if earnings are self-reported and zero if proxy-reported, V_{it} is a standardized measurement-error with mean zero and variance one for each year, and $(\mu_t, \delta_t, \gamma_t)$ is a vector of unknown parameters. The parameter δ_t measures the difference between measurement errors in terms of whether earnings are self-reported or not.

Following the notation used in Blau and Khan (1997), let m and f subscripts refer to male and female averages, respectively, and let a Δ prefix denote the average male-female

difference for the variable immediately following. Then the male-female log wage gap for year t is

$$D_t \equiv Y_{mt} - Y_{ft} = \Delta X_t B_t + \sigma_t \Delta \theta_t + \Delta U_t, \quad (4)$$

where the last term ΔU_t , which measures the gender difference in terms of measurement errors, is an additional term, compared to equation (2) of Blau and Khan (1997).

Notice that

$$\Delta U_t = \delta_t \Delta R_t + \gamma_t \Delta V_t, \quad (5)$$

which implies that the average gender difference in measurement errors depends on ΔR_t if δ_t is nonzero. Furthermore, (5) implies that changes in ΔR_t might contribute to the trends in the gender wage gap if δ_t is nonzero. There is no *a priori* reason that ΔV_t be different across gender. Hence, we make the following simple assumption:

Assumption V. *The standardized measurement error V_{it} is mean-independent of gender. That is, $V_{mt} = V_{ft} = \Delta V_t = 0$.*

Using this assumption and (4), the difference in the gender wage gap between two years, say t and t' where $t' > t$, can be decomposed:

$$\begin{aligned} D_{t'} - D_t &= (\Delta X_{t'} - \Delta X_t) B_{t'} \quad [\text{Observed } X\text{'s Effect}] \\ &\quad + \Delta X_t (B_{t'} - B_t) \quad [\text{Observed Prices Effect}] \\ &\quad + (\Delta \theta_{t'} - \Delta \theta_t) \sigma_{t'} \quad [\text{Gap Effect}] \\ &\quad + \Delta \theta_t (\sigma_{t'} - \sigma_t) \quad [\text{Unobserved Prices Effect}] \\ &\quad + \delta_{t'} (\Delta R_{t'} - \Delta R_t) \quad [\text{R-Effect}] \\ &\quad + (\delta_{t'} - \delta_t) \Delta R_t \quad [\text{D-Effect}], \end{aligned} \quad (6)$$

where the first four terms are the same as those in Blau and Khan (1997) and the last two terms are new and coined as R and D effects. If δ_t is zero, or more generally if δ_t and ΔR_t are time constant, the changes in the gender wage gap would not be affected by measurement errors. Otherwise, trends in the gender wage gap can be contaminated by spurious effects due to changes in gender-specific averages of measurement errors.

The term $\delta_{t'}(\Delta R_{t'} - \Delta R_t)$, the “R-effect,” measures the contribution of changing male-female differences in proxy-response rates R . Recall that R_{it} takes value one if earnings are self-reported and zero if proxy-reported. In the PSID, there has been an increasing trend in female respondents for married couples. This means that R_f has increased, while R_m has decreased, thereby implying that ΔR_t has decreased. Thus, if $\delta_{t'}$ is nonzero, then the declining male-respondents can contribute the changes in the wage gap. On the one hand, if $\delta_{t'}$ is negative (i.e. self-respondents report lower earnings than proxy-respondents), then there might have been an even larger reduction in the wage gap in the absence of the declining male-respondents. On the other hand, if $\delta_{t'}$ is positive (i.e. self-respondents report higher earnings than proxy-respondents), then the declining male-respondents contributed to the reduction in the wage gap. In other words, some portion of the reduction in the wage gap is attributable to measurement errors.

The term $(\delta_{t'} - \delta_t)\Delta R_t$, the “D-effect,” reflects the effect of changing respondent-type measurement errors on the gender gap. For example, suppose that $\delta_{t'}$ and δ_t are positive but $\delta_{t'} < \delta_t$ and also that $\Delta R_t > 0$. In other words, self-respondents report higher earnings than proxy-respondents on both years but less so in t' and also that there are more self-respondents among males than females in t . In this case, one may conclude that again there might have been some spurious reduction in the wage gap due to measurement errors.

2.1 Estimation Methods

To obtain the decomposition results in (6), we only need to estimate the male earnings equation for each t . Suppose for the moment that both unobserved terms θ_{it} and V_{it} are uncorrelated with X_{it} and R_{it} for males. Then, one can estimate B_t and δ_t by OLS using male PSID data. Specifically, combining (1), (2) and (3), for each year, we estimate the following using only males:

$$Y_{it} = \mu_t + X_{it}B_t + \delta_t R_{it} + \varepsilon_{it}, \quad (7)$$

where $\varepsilon_{it} \equiv \sigma_t \theta_{it} + \gamma_t V_{it}$.

Once we have consistent estimates for male earnings equations, then it is straightforward to compute the observed X 's effect, observed prices effect, R-effect and D-Effect. In addition, both the gap and unobserved prices effects can be estimated as in Blau and Khan

(1997). For example, suppose that we decompose the change in the gender wage gap from 1979 to 1989. To figure out the impacts of the R- and D-effects on the observed change in the wage gap, we estimate them by:

$$\begin{aligned} \widehat{\text{R-Effect}}_{89-79} &= \widehat{\delta}_{89}(\Delta\bar{R}_{89} - \Delta\bar{R}_{79}) \\ \widehat{\text{D-Effect}}_{89-79} &= (\widehat{\delta}_{89} - \widehat{\delta}_{79})\Delta\bar{R}_{79}, \end{aligned} \tag{8}$$

where $\widehat{\delta}_t$ is the coefficient estimate for R_{it} in (7) for year t and $\Delta\bar{R}_t = \bar{R}_{mt} - \bar{R}_{ft}$. Here, \bar{R}_{mt} (or \bar{R}_{ft}) is the sample self response rate for men (or women). The sum of the R- and D-Effects represents the amount of the spurious change in the gender wage gap induced by proxy effects and changing gender composition of representative respondents (“measurement error effects”). The relative size of the measurement error effects on the observed trend is $(\widehat{\text{R-Effect}}_{89-79} + \widehat{\text{D-Effect}}_{89-79})/(\bar{D}_{89} - \bar{D}_{79})$, where \bar{D}_t is the unconditional sample average gender wage gap. Under the standard assumption that observations in 1989 and those in 1979 are independent of each other, standard errors for the estimated R- and D-effects and their sum can be obtained easily, using usual asymptotic approximations.

A potential econometric issue with estimating the equation with R_{it} as a regressor by OLS is the self-selection into the survey response. Household-representative respondents may not be randomly chosen within households even conditional on various household characteristics. In that case, the OLS estimates could give misleading results. It is likely that R_{it} might be correlated with θ_{it} , thereby variations in R_{it} being driven by changes in unobserved skills and their prices. According to the economic theory of time allocation, one’s time spent at home will be shorter when the person’s wage rate or income is higher (Gronau and Hamermesh, 2008). This suggests that the OLS estimate for the coefficient on the self-response indicator is downward biased.

In this paper, to overcome the potential endogeneity problem, we use the instrumental variable approach. Instrumental variables are motivated by the PSID survey rules and practices. The first variable is the indicator of whether the survey was conducted during weekends. One may expect that *ceteris paribus*, males are more likely to respond during weekends. Hill (1987) also used the same variable to control for self-selectivity into respondents. Although this IV is intuitive, one may concern that the PSID administrators might take this into consideration and select the survey date to maximize the likelihood of getting

those household heads who are difficult to contact during weekdays. If we assume that those heads are likely to have higher hourly earnings, the IV estimate is likely to be upwardly biased. To address this concern, we compared working hours of those who had the survey during weekends with those during weekdays but found no significant difference.

The second instrument is the indicator of whether the survey was self-responded in the previous wave, $R_{i,t-1}$. The rationale is that the selection of the survey respondent is path dependent. This is in part because of the PSID’s survey practice. In the PSIDVS codebook, we found that “following standard PSID practice, interviews were conducted whenever possible with the head of the household.” We also found from personal communication with a research technician associate for the PSID that they attempt to contact the household head when the household enters the sample for the first time and, in the subsequent years, to keep the previous year’s respondent, *regardless of whoever it was*. This indicates that the IV should be relevant. The validity condition for the IV requires that the respondent selection at $(t - 1)$ is uncorrelated with unobserved determinants of hourly earnings at t . Our IV estimates will be upward biased if hourly earnings are autocorrelated. Thus, at the worst case, the OLS and IV estimates should bound the true coefficient.

2.2 Gender-Specific Self/Proxy Measurement Error

One important implicit assumption used in decomposition (6) is that coefficients for self/proxy measurement errors are the same between males and females. However, this assumption might be not valid, although it would not be possible to test this assumption without validation data.

Suppose now that coefficients μ_t and δ_t in (3) are gender-specific. Under Assumption V, we have

$$\Delta U_t = \Delta \mu_t + \delta_{mt} \Delta R_t + \Delta \delta_t R_{ft}. \quad (9)$$

Thus, if μ_t and δ_t in (3) are gender-specific, but Assumption V still holds, then

$$\begin{aligned} \Delta U_{t'} - \Delta U_t &= \delta_{m t'} (\Delta R_{t'} - \Delta R_t) + (\delta_{m t'} - \delta_{m t}) \Delta R_t \\ &+ (\Delta \mu_{t'} - \Delta \mu_t) + \Delta \delta_{t'} (R_{f t'} - R_{f t}) + (\Delta \delta_{t'} - \Delta \delta_t) R_{f t}. \end{aligned} \quad (10)$$

The last three terms are additional terms due to the possible differences in μ_t and δ_t between male and female respondents. To evaluate the importance of the potential bias under a simple setup, suppose that $\Delta\mu_t$ and $\Delta\delta_t$ are time invariant, so that (10) simplifies to

$$\Delta U_{t'} - \Delta U_t = \delta_{mt'}(\Delta R_{t'} - \Delta R_t) + (\delta_{mt'} - \delta_{mt})\Delta R_t + \Delta\delta_{t'}(R_{ft'} - R_{ft}). \quad (11)$$

As will be explained in the next section, following Blau and Kahn (2006), we will examine the gender earnings gap at three points of time, 1979, 1989, and 1998. In our data, male self-respondents have decreased all over the years (from 86.7% in 1979 to 73.6% in 1989 and to 62.2% in 1998). For females, self-respondents increased all over the years (from 51.2% in 1979 to 64.0% in 1989 and to 70.7% in 1998). Hence, $(R_{ft'} - R_{ft})$ is positive for every year. If $\Delta\delta_{t'}$ is large in terms of its magnitude, then we will have a substantial bias from omitting the third term in (11). Hence, in this case, our proposed estimators of the R- and D- effects in Section 2.1 may provide a misleading result. However, this concern may not be a serious issue if $\delta_{mt} \approx \delta_{ft}$, as suggested by the fixed effects estimates of Bollinger and Hirsch (2009, Table 8).

3 Data

We followed Blau and Kahn (2006) to obtain the same samples from three waves of the PSID, 1980, 1990, and 1999. Table 1 presents summary statistics of principal variables. They are very similar to those of Blau and Kahn's.⁹ In each wave, information on average hourly earnings in its previous year is available. After adjusting nominal wages in 1983 dollars using the Personal Consumption Expenditures deflator from the National Product Accounts, we exclude those earning less than \$1 or more than \$250 per hour. The samples are restricted to wage and salary workers (not self-employed) who were, as of the survey date, full-time, non-agricultural employees aged 18-65.

As Blau and Kahn pointed out, the main advantage of the PSID for a study of gender earnings gap is that actual experience variables can be created from longitudinal data. We constructed labor-market experience variables (years of full-time and part-time experiences

⁹The sample sizes are slightly different from those of Blau and Kahn's. This is presumably in part because we have some additional variables and in part because of slight differences in sampling procedures.

since age 18) in the same way Blau and Kahn did. We used the experience variables based on the responses to related questions in two waves, 1976 and 1985, as the reference. Note that in the two years, the PSID asked all households about the head’s labor-market experiences and, if any, the spouse’s. In addition, the PSID attempted to obtain self-responded individual data from the head and, if any, the head’s spouse partly because of the concern about the quality of proxy-responded data. For those households who joined the survey at some point after the two years (i.e. between 1977 and 1984 or after 1985), we used the experience variables in the first year they joined the survey when the PSID asked the experience questions. Having the reference experience variables, we updated the variables by using total annual working hours in the subsequent years. If total annual working hours are greater than 0, then we added one additional year of labor market experience. If annual working hours are greater than or equal to 1500, then we added one additional year of full-time experience. Starting in 1997, the PSID collected data biennially. Thus information on labor market experiences in 1998 is not available. Again, following Blau and Kahn, we predict labor market activity in the year between 1997 and 1999 and create complete labor market experience data. In Table 1, our sample is also compared with Blau and Kahn’s. Although they are not exactly matched, the means of variables are similar between the two samples.

4 Empirical Results

4.1 Log Hourly Earnings Equation

Tables 3 shows the results for the log hourly wage equation in three years, 1979, 1989, and 1998. We present the results for the “full” specification which, in addition to the standard human capital variables, includes a collective bargaining coverage indicator and a set of 19 occupation and 25 industry dummy variables.¹⁰ The specification is exactly the same as that of Blau and Khan (2006) except that we have an additional variable, i.e. the indicator for self response. The estimation results are overall not different from those of Blau and Khan (2006).

¹⁰The main conclusion of this paper is the same when using the “human capital” specification which includes only race, education variables, and experience variables.

For the 1979 data, we find that the coefficient on the self-response indicator is positive and significant. Specifically, holding other things constant, self respondents report own hourly earnings higher by about 5.6 percent (OLS) to 16.8 percent (IV).¹¹ The IVs are significant in the first stage. Those who self-responded in the previous wave are more likely to self-respond to the current survey. The autoregressive coefficient is 0.51 (p-value < 0.01).¹² Also, we find that male heads are 3.4% more likely to respond to the survey during weekends. The partial R-squared is 0.3. The overidentifying restriction cannot be rejected. For the 1989 data, the self-response coefficient becomes insignificant and virtually zero in magnitude. This is consistent with the finding that there is no proxy effect on the likelihood of the PSID assignment in the 1989 head labor income data (see footnote 11). For the 1998 data, the result is ambiguous depending on the estimation method. The IV estimate is significant and positive; self-reported hourly earnings are higher by about 8.9 percent. The OLS estimate is, however, small and insignificant. Given that the OLS estimate is a lower bound, we cannot determine whether there is a self-response effect.

It is worth emphasizing that estimated coefficient δ_t is significantly positive at least in 1979. This implies that changes in the average gender differences in response rates Δ_t should affect the trends in the gender earnings gap. Second, it is interesting to observe that estimated values of δ_t seems to be time varying. It is puzzling why this coefficient is changing over time. This is perhaps because information sharing within households might have changed over time. Also note that the PSID samples have changed. In particular, in 1997, the “core” sample was reduced by 2,332 families.¹³ This might make the proxy

¹¹It is *a priori* unknown whether self respondents are more accurate than proxy respondents or vice versa. However, there is indirect evidence that self respondents are more accurate in the PSID. The PSID staff evaluated the quality of survey responses, and made corrections in certain cases where reporting errors are presumably too large. The most common correction method is to assign the previous year’s value. This imputation is called as an “assignment.” We examined the effect of proxy response on the likelihood of receiving the PSID assignment. We find that proxy respondents’ reports about the head’s labor income are more likely to receive assignments in the 1979 and 1998 data. However, in the 1989 data, we find no proxy effect.

¹²The AR coefficient is larger in 1989 and 1998, about 0.8. This reflects the PSID’s survey rule of retaining the previous year’s respondent.

¹³The core sample in 1968 consists of a cross-sectional national sample (the “SRC sample”) and a national sample of low-income families (the “SEO sample”). There have been some changes in the sample design. In

response gap vary over time. Lastly, selection bias is likely to be more severe in earlier waves of the PSID when there were a relatively lesser number of households whose spouses responded to the survey.

4.2 Decomposition of Changes in the Gender Earnings Gap

Table 4 presents the decomposition results. The top panel shows the decomposition between 1979 and 1989 based on both OLS and IV estimates. For each estimation method, the first row of the table shows the gender log wage differential (a conventional measure of the gender earnings gap) and the second row presents the average male-female difference in self response rates (ΔR_t). The estimated coefficient (δ_t) is reproduced in the third row. The R- and D-effects are shown in the fourth and fifth rows and its sum, labeled as measurement error effects, is in the sixth row.

Our main empirical finding is that a non-ignorable part of the change in the gender earnings gap between 1979 and 1989 can be explained by changes in the self-response rate and the proxy-response bias in male earnings. Depending on the estimation methods, proxy effects and changing respondent composition can explain about 11% (OLS) to 34% (IV) of the gender gap change in the period. Specifically, we find that the result is driven by the D-effect. Between 1979 and 1989, we have that $\hat{\delta}_{89} < \hat{\delta}_{79}$ and also that $\Delta \bar{R}_{79} > 0$. In other words, self-respondents report higher earnings than proxy respondents in 1979 but much less so in 1989 and also there are more self respondents among males than females in 1979. As a result, there might have been some spurious reduction in the gender earnings gap due to measurement error. This suggests that the true reduction in the gender earnings gap might be smaller in its magnitude than the conventional measure reported in the first row of the table.¹⁴ The exact magnitude of this spurious reduction is harder to quantify since

1990 2,000 Latino households were added. The “Latino sample” was dropped after 1995. In 1997, there were two major changes. First, the core sample was reduced. Second, a refresher sample of immigrant families and their adult children was added (the “immigrant refresher sample”).

¹⁴Blau and Kahn (2006) also present trends in the gender earnings gap using the CPS data. The change in gender log wage differential is -0.1292 between 1979 and 1989 and -0.0677 between 1989 and 1998. The convergence rate is higher in the PSID data. It would be an interesting research topic to investigate how much of the difference in the convergence rate between the PSID and CPS data can be explained by proxy effects and changing respondent composition in the PSID; however, it is beyond the scope of this paper.

we believe that our OLS and IV estimates provide mainly some suggestive evidence rather than decisive evidence. Our decomposition results demonstrate that there would be a scope for future research. In particular, a validation study matched with administrative records would help us quantify the true measure of changes of the gender earnings gap.

The bottom panel of Table 4 presents the decomposition results between 1989 and 1998. The gender earnings gap between 1989 and 1998 decreased slowly by about 5.6 percent. The proportion of male self respondents continued to drop. The self/proxy measurement error accounts for 14% (IV) of the gender gap change in the period. There is almost no proxy effect based on the OLS estimate. Since the change in the raw gender gap is not large, contributions from measurement error effects are also quite small. We need to be careful to interpret results between 1989 and 1998 since we did not find any significant difference between self- and proxy-reported earnings. Recall that only the IV estimate for the 1998 data was significant.

5 Conclusion

We have decomposed the trends in the U.S. gender earnings gap into observable and unobservable components using the PSID, while taking into account differences between the self- and proxy-reported earnings. The differences between the self- and proxy-reports are particularly important since there has been a strong trend of changing gender composition of representative respondents. Our empirical results suggest that a significant portion of changes in the unmeasured gender gap could be attributed to spurious changes due to measurement error. Specifically, the measurement error can account for about 11%-34% of the gender earnings gap reduction from 1979 to 1989. In other words, the actual reduction seems to be smaller than what the estimates without taking into account the measurement error might suggest. The role of the measurement is weak for estimating the gender gap reduction from 1989 to 1998. But it is still possible that the decreasing trend is overestimated by 14% at the maximum. Further research such as a careful validation study would be needed to quantify the magnitude of the spurious measurement error effects.

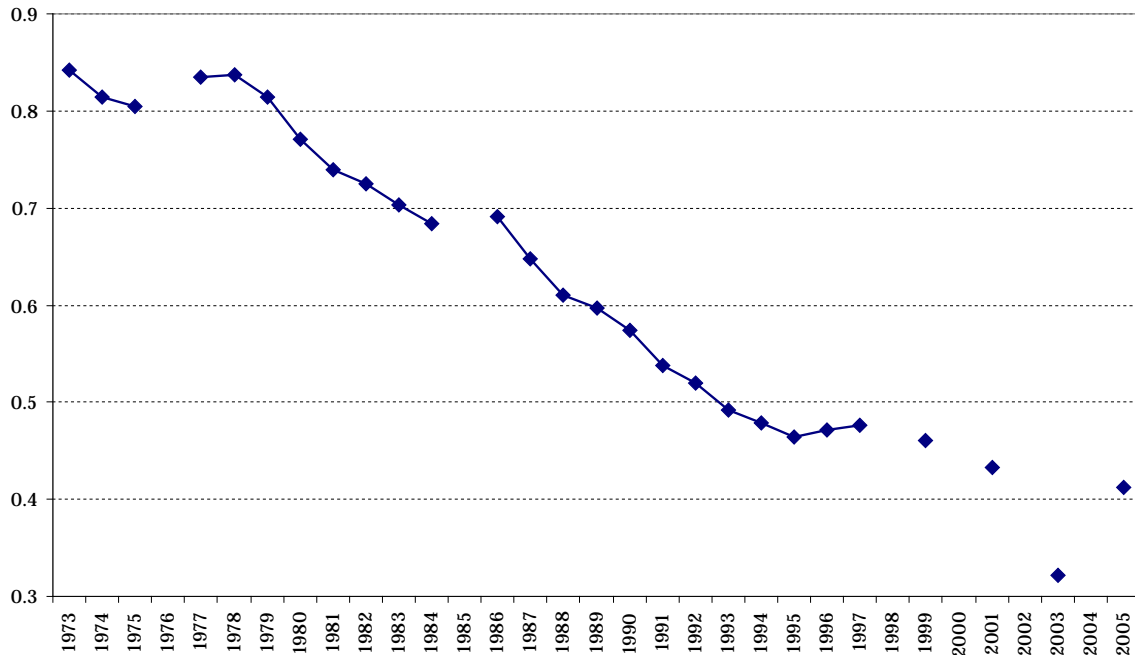
References

- Blau, Francine D. and Lawrence M. Kahn (1997) "Swimming Upstream: Trends in the Gender Wage Differential in 1980s." *Journal of Labor Economics*, Vol. 15, No. 1, pp. 1-42.
- Blau, Francine D. and Lawrence M. Kahn (2006) "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence." *Industrial and Labor Relations Review*, Vol. 60, No. 1, pp. 45-66.
- Bollinger, Christopher R. (1998) "Measurement Error in the Current Population Survey: A Nonparametric Look." *Journal of Labor Economics*, Vol. 16, No. 3, pp. 576-594.
- Bollinger, Christopher R. (2001) "Response Error and the Union Wage Differential." *Southern Economic Journal* 68:1, pp. 60-76.
- Bollinger, Christopher R. (2003) "Measurement Error in Human Capital and the Black-White Wage Gap." *Review of Economics and Statistics*, Vol. 85, No. 3, pp. 578-585.
- Bollinger, Christopher R. and Barry T. Hirsch (2009) "Wage Gap Estimation with Proxies and Nonresponse." Working Paper, University of Kentucky and Georgia State University, October 2009.
- Bound, John and Alan B. Krueger (1991) "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, Vol. 9, No. 1, pp. 1-24.
- Bound, John, Charles Brown, Gregg J. Duncan, and W. L. Rodgers (1994) "Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data." *Journal of Labor Economics*, Vol. 12, No. 3, pp. 345-368.
- Duncan, Gregg J. and Daniel H. Hill (1985) "An Investigation of the Extent and Consequences of Measurement Error in Labor-economic Survey Data." *Journal of Labor Economics*, Vol. 3, No. 4, pp. 508-532.
- Duncan, Greg J., and Mathiowetz, Nancy A. (1985) "A Validation Study of Economic Survey Data." monograph, University of Michigan, Survey Research Center.

- Gronau, Reuben and Daniel S. Hamermesh (2008) "The Demand for Variety: A Household Production Perspective." *Review of Economics and Statistics*. Vol. 90, No. 3, pp. 562-572.
- Juhn, Chinhui, Murphy, Kevin M. and Pierce, Brooks (1993) "Wage Inequality and the Rise in Returns to Skill" *Journal of Political Economy* 101, pp. 410-42.
- Hill, Daniel (1987) "Response Errors in Labor Surveys: Comparisons of Self and Proxy." Survey of Income and Program Participation Working Paper No. 28.
- Kristensen, Nicolai and Niels Westergaard-Nielsen (2007) "A Large-Scale Validation Study of Measurement Errors in Longitudinal Survey Data." *Journal of Economic and Social Measurement*, Vol. 32, No. 2-3, pp. 65-92.
- Mellow, Wesley and Hal Sider (1983) "Accuracy of Response in Labor Market Surveys: Evidence and Implication." *Journal of Labor Economics*, Vol. 1, pp. 331-44.
- Mathiowetz, Nancy A. and Greg J. Duncan (1988) "Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment." *Journal of Business & Economic Statistics*, Vol. 6, No. 2, pp. 221-229.
- Moore, Jeffrey C. (1988) "Self/Proxy Response Status and Survey Response Quality." *Journal of Official Statistics*, Vol. 4, No. 2, pp. 155-172.
- Mroz, Thomas A. (1987) "The Sensitivity of An Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions." *Econometrica*, Vol. 55, No. 4, pp. 765-799.
- Pischke, Jorn-Steffen (1995) "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study." *Journal of Business and Economic Statistics*, Vol. 13, Issue 3, pp. 305-314.
- 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*, Vol. 88, No. 424, pp. 1208-1218.

Todorov, Alexander (2003) "Cognitive Procedures for Correcting Proxy-response Biases in Surveys." *Applied Cognitive Psychology*, 17, pp. 215-224.

**Figure 1. Composition of Household Representative Respondents in the PSID
(Proportion of Head Respondents)**



Notes: The figure depicts the proportion of head respondents among male-headed married households. In 1973, the PSID started telephone interviews. In 1976 and 1985, the PSID required, if possible, self response for both heads and spouses. Hence, these two years are not shown in the figure. Starting in 1997, the PSID collected data biennially.

Table 1. Means for the Regression Samples, PSID

	1979		1989		1998	
	Current Study	Blau & Kahn (2006)	Current Study	Blau & Kahn (2006)	Current Study	Blau & Kahn (2006)
Log Male Hourly Earnings	2.388	2.384	2.342	2.334	2.381	2.354
Log Female Hourly Earnings	1.902	1.926	2.030	2.039	2.125	2.126
Male (Head) Earnings Self Response	0.867		0.736		0.622	
Female Earnings Self Response	0.512		0.640		0.707	
Gender Self Response Differential	0.355		0.096		-0.085	
White	0.878	0.870	0.893	0.885	0.877	0.866
Years of Schooling	12.76	12.70	13.45	13.30	13.53	13.45
College Degree	0.156	0.160	0.199	0.203	0.239	0.219
Advanced Degree	0.074	0.069	0.085	0.069	0.083	0.066
Full Time (FT) Experience	18.80	18.31	18.82	18.28	20.95	19.84
Part Time (PT) Experience	1.268	1.204	1.735	1.667	2.143	1.423
Collective Bargaining	0.345	0.346	0.233	0.239	0.205	0.202
Number of observations =	2,734	2,816	2,752	2,894	2,403	2,336

Notes: Data were extracted from the PSID following Blau and Kahn (2006) to obtain samples from three waves of the PSID, 1980, 1990, and 1999. For the 1999 data, new immigrant family sample was excluded. Observations are weighted by cross-sectional family weights.

Table 2. Log Hourly Earnings Regression

	1979		1989		1998	
	OLS	IV	OLS	IV	OLS	IV
Self Response (δ_i)	0.056 (0.028)	0.168 (0.049)	0.001 (0.022)	0.010 (0.029)	0.005 (0.023)	0.081 (0.030)
White	0.048 (0.025)	0.046 (0.025)	0.093 (0.027)	0.093 (0.027)	0.090 (0.036)	0.090 (0.035)
Years of Schooling	0.044 (0.006)	0.042 (0.006)	0.060 (0.007)	0.059 (0.007)	0.065 (0.009)	0.063 (0.009)
College Degree	0.052 (0.041)	0.048 (0.041)	0.116 (0.042)	0.116 (0.041)	0.089 (0.039)	0.092 (0.039)
Advanced Degree	0.155 (0.055)	0.151 (0.055)	0.241 (0.067)	0.241 (0.066)	0.171 (0.057)	0.172 (0.056)
Full-time Experience	0.040 (0.003)	0.041 (0.003)	0.040 (0.004)	0.040 (0.004)	0.041 (0.005)	0.042 (0.005)
FT Exp Squared	-0.0007 (0.0001)	-0.0007 (0.0001)	-0.0006 (0.0001)	-0.0006 (0.0001)	-0.0007 (0.0001)	-0.0007 (0.0001)
Part-time Experience	-0.009 (0.008)	-0.009 (0.008)	-0.0004 (0.008)	-0.0005 (0.0001)	0.002 (0.006)	0.002 (0.006)
PT Exp Squared	0.0005 (0.0004)	0.0005 (0.0004)	0.0004 (0.0004)	0.0004 (0.0004)	-0.0001 (0.0002)	-0.0001 (0.0002)
Union	0.210 (0.021)	0.208 (0.021)	0.263 (0.024)	0.263 (0.023)	0.293 (0.030)	0.293 (0.030)
Constant	0.941 (0.095)	0.866 (0.099)	0.530 (0.122)	0.525 (0.122)	0.468 (0.141)	0.441 (0.142)
<i>IVs in the First Stage</i>						
Self Response at (t - 1)		0.515 (0.028)		0.818 (0.018)		0.767 (0.016)
Weekend Interview		0.034 (0.013)		0.038 (0.016)		0.028 (0.019)
Occupation Dummies	YES	YES	YES	YES	YES	YES
Industry Dummies	YES	YES	YES	YES	YES	YES
Partial R-squared		0.301		0.616		0.605
Hansen's J Statistic p-value		0.743		0.585		0.189
Observations	2,734	2,734	2,752	2,752	2,403	2,403
R-squared	0.396	0.390	0.447	0.447	0.416	0.412

Notes: Robust standard errors are in parentheses. Observations are weighted by cross-sectional family weights. IV estimates were obtained by 2SLS with two instrumental variables: self response at the previous wave and weekend interview. The estimates for the IVs in the first stage regression are presented.

Table 3.A. Decomposition, 1979-1989

Decomposition results using OLS estimates			
	Year 1979 (<i>t</i>) (A)	Year 1989 (<i>t'</i>) (B)	Change (1979-1989) (B) minus (A)
Gender Log Wage Differential	0.486 (0.018)	0.312 (0.020)	-0.174 (0.027)
Gender Self Report Differential	0.355 (0.017)	0.096 (0.016)	-0.259 (0.024)
Estimated Coefficient for Self Report	0.056 (0.028)	0.001 (0.022)	-0.055 (0.036)
R-Effect			-0.0003 (0.006)
D-Effect			-0.020 (0.013)
Measurement Error Effects			-0.020 (0.010)
Decomposition results using IV estimates			
	Year 1979 (<i>t</i>) (A)	Year 1989 (<i>t'</i>) (B)	Change (1979-1989) (B) minus (A)
Gender Log Wage Differential	0.486 (0.018)	0.312 (0.020)	-0.174 (0.027)
Gender Self Report Differential	0.355 (0.017)	0.096 (0.016)	-0.259 (0.024)
Estimated Coefficient for Self Report	0.168 (0.049)	0.010 (0.029)	-0.158 (0.057)
R-Effect			-0.003 (0.008)
D-Effect			-0.056 (0.020)
Measurement Error Effects			-0.059 (0.018)

Notes: Results of decomposition of changes in the gender earnings gap are presented. Standard errors are presented in parentheses. Observations are weighted by cross-sectional family weights. Gender Log Wage Differential = $\Delta \bar{D}_t$; Gender Self Report Differential = $\Delta \bar{R}_t$; Estimated Coefficient for Self Report = $\hat{\delta}_t$; R-Effect = $\hat{\delta}_{t'}(\Delta \bar{R}_{t'} - \Delta \bar{R}_t)$; D-Effect = $(\hat{\delta}_{t'} - \hat{\delta}_t)\Delta \bar{R}_t$; Measurement Error Effects = R-Effect + D-Effect.

Table 3.B. Decomposition, 1989-1998

Decomposition results using OLS estimates			
	Year 1989 (<i>t</i>) (A)	Year 1998 (<i>t'</i>) (B)	Change (1989-1998) (B) minus (A)
Gender Log Wage Differential	0.312 (0.020)	0.256 (0.022)	-0.056 (0.030)
Gender Self Report Differential	0.096 (0.016)	-0.085 (0.017)	-0.181 (0.024)
Estimated Coefficient for Self Report	0.001 (0.022)	0.005 (0.023)	0.004 (0.032)
R-Effect			-0.001 (0.004)
D-Effect			0.0004 (0.003)
Measurement Error Effects			-0.001 (0.003)
Decomposition results using IV estimates			
	Year 1989 (<i>t</i>) (A)	Year 1998 (<i>t'</i>) (B)	Change (1989-1998) (B) minus (A)
Gender Log Wage Differential	0.312 (0.020)	0.256 (0.022)	-0.056 (0.030)
Gender Self Report Differential	0.096 (0.016)	-0.085 (0.017)	-0.181 (0.024)
Estimated Coefficient for Self Report	0.010 (0.029)	0.081 (0.030)	0.071 (0.042)
R-Effect			-0.015 (0.006)
D-Effect			0.007 (0.004)
Measurement Error Effects			-0.008 (0.004)

Notes: See the notes for Table 3.A.