**An Evaluation of the Gender Earnings Gap Using Survey and Administrative Data[1]**

Marta Murray-Close, U.S. Census Bureau (Presenter)
Thomas B. Foster, U.S. Census Bureau
Liana Christin Landivar, U.S. Census Bureau Special Sworn Status
Mark deWolf, U.S. Census Bureau Special Sworn Status

**Abstract**

On average, women earn about 80 percent of what men earn. Gender differences in characteristics like education, labor-market experience, occupation, and industry explain much of the gender earnings gap, but even after controlling for these differences, most studies find an unexplained and relatively stable residual gap of 4 to 8 percent. The primary objective of this study is to improve estimates of the gender earnings gap by combining survey and administrative data. A major limitation of existing studies is that the large datasets required to effectively analyze occupation and industry lack measures of actual labor-market experience. This study will augment data from the Current Population Survey with measures of labor-market experience constructed from administrative earnings histories. The study will use these linked data to compare estimates of the gender earnings gap based on potential versus actual experience and to better understand the impact of occupation selection on the earnings gap.

**Extended Abstract**

The existence of a gender earnings gap is firmly established in the economic and sociological literature. On average, women earn about 80 percent of men's earnings (Semega et al. 2017). As women have attained higher levels of human capital and have increased their employment continuity, the relative importance of variables such as industry and occupation has grown over time. Men and women remain segregated across industries and occupations, and new research shows this segregation contributes to about half of the explained variance in earnings between men and women (Chamberlain

---

2016; Blau and Kahn 2017). When studies control for characteristics such as occupation, industry, tenure, and human capital, much of the earnings gap is explained. Yet, even after controlling for a substantial number of characteristics, most studies find a relatively stable unexplained earnings gap of 4 to 8 percent (Chamberlain 2016; Blau and Kahn 2017). This unexplained earnings gap could stem from some combination of unobserved gender differences in labor-market productivity, gender differences in employer characteristics or bargaining power (Card, Cardoso, and Kline 2015), and discrimination against women workers.

The primary objective of this study is to provide an improved measurement of the adjusted and residual gender earnings gap using an innovative dataset that combines survey data and administrative records. One of the major limitations to existing studies on the gender earnings gap is a tradeoff, in nationally representative datasets, between sample size and the availability of key measures. In the U.S., large-scale surveys like the American Community Survey and the Current Population Survey allow researchers to estimate the (descriptive) effects of industry and occupation on earnings with reasonable precision but lack measures of job tenure and labor-market experience (CONSAD Research 2009; Goldin 2014). (The Survey of Income and Program Participation is a partial exception, but it contains employment histories only in select years.) Smaller longitudinal surveys like the Panel Study of Income Dynamics and the National Longitudinal Surveys allow researchers to construct measures of job tenure and labor-market experience (Anderson et al. 2003) but contain few observations in many industry and occupation categories.

Omitting either employment histories or measures of industry and occupation from analyses of the gender wage gap may produce misleading results. Simulated measures of labor-market experience, such as age-years of education-6 years (Mincer 1974), will overestimate the experience of workers who take time out of the labor force after finishing their education (Regan and Oaxaca 2009). Since women are more likely than men to take time out due to unequal caregiving responsibilities, empirical models

that include simulated rather than actual experience will tend to overestimate the gender differential in returns to experience and hence the unexplained component of the gender wage gap. In addition, labor force disruptions due to the Great Recession have made measurements of actual work experience increasingly important among men (Blau and Kahn 2017). With respect to industry and occupation, recent evidence from the Panel Study of Income Dynamics shows that the contribution of industry and occupation to the gender earnings gap is substantial and has grown considerably over time (Blau and Kahn 2017).

To overcome these limitations, and to improve the measurement of the gender earnings gap, we use administrative data from the Social Security Administration (SSA) linked with the Current Population Survey (CPS). The CPS linked with SSA administrative earnings histories allows us to evaluate demographic and economic characteristics jointly with earnings histories going back to 1978. We estimate the gender earnings gap using pooled data from the 2004 to 2013 survey years (which correspond to the 2003 to 2012 tax years), controlling for demographic characteristics, actual work history, current industry, and current detailed occupation. We use an Oaxaca-Blinder decomposition of male-female differences in log earnings (Blinder 1973; Oaxaca 1973) which shows the relative contributions of each characteristic in the model to explaining the gender earnings gap. The residual, or remaining unexplained component, has been partly attributed to gender discrimination. To the extent that unmeasured differences in labor-market productivity may account for the residual, this would overstate the effect of discrimination. On the other hand, discrimination may affect measured characteristics (e.g., occupation selection, time out of the labor force), and by attributing the differential to the measured characteristics, the effect of discrimination would be underestimated (Oaxaca 1973; Fortin et al. 2010). We show the decomposition for the measured characteristics and unexplained residual and how the contributions of each characteristic have changed over time.

Our dependent variable is the log of annual earnings. Our explanatory variables include, age, marital status, presence and age of children, race, Hispanic origin, educational attainment, industry (aggregate level), occupation (detailed), class of worker, work hours, tenure, work history and experience, and timing of career interruptions. We use an Oaxaca-Blinder decomposition to obtain estimates of the unexplained male-female difference in log earnings. Next, we compare estimates of the unexplained difference from decomposition models with three different measures of experience: (1) potential experience based on age and educational attainment, (2) actual cumulative years of experience based on the administrative earnings histories, and (3) detailed career history variables capturing the timing and duration of labor-market breaks, in the spirit of Light and Ureta (1995). In addition to capturing *when* women work through work history measures, we examine *how much* women work using their current work hours. We specifically evaluate nonlinearity in returns to hours worked and whether this has become a more significant contributor to the gender earnings gap. To the extent that women work fewer hours than men and the returns to work hours are increasingly nonlinear (Goldin 2014), the gender earnings gap may be disproportionately large in long-hours occupations.

This study offers three main contributions. First, using combined survey and administrative data sources, we offer a more precise decomposition of the gender earnings gap through our use of a large survey sample linked with actual work history measures captured in administrative sources. Second, we compare our results using combined survey and administrative data to previous studies that rely solely on survey data. We evaluate simulated work experience measures used in cross-sectional analyses by comparing these with actual work experience measures available from administrative records. Third, we show with greater measurement precision how occupation selection and nonlinearities in hours worked within each occupation impact the gender earnings gap. Together, these analyses will offer an improved measurement of the adjusted and residual gender earnings gap.

**References**

Anderson, Deborah J., Melissa Binder, and Kate Krause. 2003. "The Motherhood Wage Penalty

   Revisited: Experience, Heterogeneity, Work Effort, and Work Schedule Flexibility."

   *Industrial and Labor Relations Review* 56(2): 273-274.

Blau, Francine D. and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and

   Explanations." *Journal of Economic Literature* 55(3): 789-865.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *The Journal*

   *of Human Resources* 8(4): 436-455.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2015. "Bargaining, Sorting, and the Gender Wage Gap:

   Quantifying the Impact of Firms on the Relative Pay of Women." The Quarterly Journal of

   Economics 13(2): 633-686.


Chamberlain, Andrew. 2016. "Demystifying the Gender Pay Gap: Evidence From Glassdoor Salary

   Data." Retrieved July 27, 2018:

   https://www.glassdoor.com/research/app/uploads/sites/2/2016/03/Glassdoor-Gender-Pay-

   Gap-Study.pdf.

CONSAD Research. 2009. "An Analysis of the Reasons for the Disparity in Wages Between Men and

   Women." Issued January 12. Available online: https://www.shrm.org/hr-today/public-policy/hr-

   public-policy-issues/Documents/Gender%20Wage%20Gap%20Final%20Report.pdf.

Fortin, Nicole, Thomax Lemieux, and Sergio Firpo. 2010. "Decomposition Methods in Economics."

   NBER Working Paper 16045. Issued June.

Goldin, Claudia. 2014. "A Grand Convergence: Its Last Chapter." *American Economic Review* 104(4):

   1091-1119.

Light, Audrey and Manuelita Ureta. 1995. "Early-Career Work Experience and Gender Wage

   Differentials." *Journal of Labor Economics* 13(1): 121-154.

Mincer, Jacob A. 1974. "The Human Capital Earnings Function." In *Schooling, Experience, and Earnings.*

Cambridge, MA: National Bureau of Economic Research.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14(3): 693-709.

Regan, Tracy L. and Ronald L. Oaxaca. 2009. "Work Experience as a Source of Specification Error in Earnings Models: Implications for Gender Wage Decompositions." *Journal of Population Economics* 22(2): 463-499.

Semega, Jessica L., Kayla R. Fontenot, and Melissa A. Kollar. 2017. "Income and Poverty in the United States: 2016." Current Population Reports P60-259. Washington, DC: U.S. Census Bureau.