Women's Gains (or not) in Cumulative Earnings: Variations across Cohorts, Race, and Education[1]

Siwei Cheng (New York University)
Christopher R. Tamborini (U.S. Social Security Administration)

INTRODUCTION

The rise in women's employment and earnings constitutes one of the most significant social and economic changes in the U.S. labor market over the last half century. This progress happened in the context of substantial changes in women's work and family lives marked by the rise in women's labor force participation (Blau and Kahn 2000; England 2010; Goldin and Mitchell 2017; Percheski 2008; Pettit and Ewert 2009). A key aspect to these trends relates to age-graded shifts in women's labor force participation. According to recent estimates by Goldin and Mitchell (2017), women's labor force participation in recent decades has become persistently high for college-educated women in early life, but declines modestly by their 30s and early 40s, resulting in a more pronounced "sagging middle" in their life course trajectories of labor supply. Yet, among those without a college degree, the labor force participation rate is consistently lower throughout the life cycle, but there has been significant increase between ages 25 and 40 from earlier to later cohorts.

Beyond being an empirical pattern of interest, the life-cycle pattern of female employment raises an important methodological issue for research in this area. Most of what we know about women's earnings trends, and gender inequality more broadly, is based on analysis of cross-sectional data, in which the earnings distribution reflects a "snapshot" of the working population. However, recent research has begun to recognize

[1] This is a preliminary draft prepared for submission to the 2019 Meeting of the Population Association of America. The authors contributed equally to the work.

the complications that the selectivity of the working or non-working population can create for estimating earnings differences across groups, especially among women. Pettit and Ewert (2009), for example, show that the size of the female black-white wage gap is influenced by the composition of employed women. To the extent that selection processes into employment change over the life cycle as well as across cohorts, cross-sectional estimates of women's earnings at different ages will reflect, not just the level of income women earned, but also the selectivity of women who worked at each age. Depending on the selection mechanisms, the "snapshot" estimates of women's earnings may over- or under-state the between-group differences in women's economic position as well as change over time.

We argue in this paper that, in light of age-related changes in women's workforce participation across cohorts, an effective way to improve our empirical estimates of women's labor market outcomes lies in the utilization of longitudinal data to construct measures of cumulative earnings over the life course. A key advantage of long-term earnings is that it can incorporate both a person's employment and earnings levels over long periods of their lives. This feature is particularly important for women who may experience multiple years of nonemployment over their careers, which has direct (via forgone earnings) and indirect (via work experience) effects on their cumulative earnings. To date, little research has analyzed women's cumulative earnings levels. The conventional approach to studying women's work trends, instead, has concentrated on labor force participation or earnings in isolation, without formally linking both dimensions to women's trajectories.

In addition to being a crucial step for overcoming limitations of cross-sectional based estimates, the concept of cumulative earnings is important in its own right, as it is associated with outcomes including lifetime earnings, intrahousehold economic position, social status, and the accumulation of wealth and retirement income (Tamborini, Kim, and Sakamoto 2015; Sakamoto, Tamborini, Kim 2018). Hence, cumulative earnings provides a more comprehensive, longitudinal view of labor market stratification than common measures such as labor force participation or wages.

In this study, we aim to advance the current literature on women and the labor market by examining life course patterns of cumulative earnings. Drawing from a life course framework (Elder et al. 2003), we investigate the cumulative earnings of women across successive birth cohorts from 1955 to 1980, which encompasses the late baby boom through late Generation X. To do so, we assemble a novel data set that links respondents in the *Survey of Income Program and Participation* (SIPP) with restricted-use tax (W-2) records contained at the Social Security Administration (SSA). From the administrative data, we are able to leverage over three decades of annual earnings data (1980-2016) to construct the employment and earnings trajectories of real cohorts based repeated measurements of the same individuals' earnings over an extended period of time. The linked data also enable us to generate better predictions of the potential earnings of women not working using information on a person's own earnings trajectory in years when they worked.

Our analysis focuses on women's experiences from early to mid-career (ages 25 to 45) to compare late baby boomers (e.g. 1955-1964) with more recent birth cohorts (e.g., 1970-1980), which have not been well studied. Focusing on women's experiences through mid-adulthood is also well suited to capture cohort changes in the age-graded patterns of women's labor force participation (Goldin and Mitchell 2017). A person's earnings trajectory over mid-adulthood, moreover, can mediate important later life outcomes, including wealth accumulation (Killewald, Pfeffer, and Schachner 2017). In our analysis, we assess how women's work participation trends influences cumulative earnings by deriving an estimate that allocates potential earnings for women in a non-employed year.

To examine heterogeneity in the experiences of women, we analyze how cumulative earnings differ by education and race within (intra) and across (inter) cohorts. Patterns by motherhood and marriage are also considered using event histories. Previous research shows important differences in women's employment and earnings by educational attainment, race-ethnicity, and family context (Damaske 2011; Damaske and Frech 2016; Goldin and Mitchell 2017; Percheski 2008; Pettit and Ewert 2009; Tamborini and Iams 2011), but how these factors interact to shape the life-course

patterns of women's cumulative earnings overall and across cohorts is not well established.

Applying a life course perspective to trends in women's paid work helps clarify cohort changes in women's standing in the labor market taking a long-term view. Our preliminary findings document how the life course pattern of women's cumulative earnings through mid-adulthood has changed over recent cohorts. We also show important differences in the cohort changes among particular subgroups based on stratifying characteristics, namely race and education. Among college graduates, black women have greater cumulative earnings than white women in all cohorts due in large part to having greater years of employment. In recent cohorts, however, white college graduates are catching up to their black counterparts in terms of employment, while black college graduates are narrowing the earnings gap with whites. Among non-college graduates, we find evidence that the black-white gap in earnings has attenuated in recent cohorts relative to the late baby boom generation.

The results also are supportive of the notion that selectivity into employment influences the magnitude of racial and education gaps observed in women's earnings within and across cohorts. By providing simulations that allocate potential earnings for years of nonemployment among women in our sample, we show that regardless of whether they have a college degree, black women's disadvantage in cumulative earnings relative to their white counterparts would have been greater if all women worked. But there are also substantial educational differences. Among college-educated women, black women receive similar, if not higher, cumulative earnings as their white counterparts without an employment adjustment. Adjusting for employment by hypothetically putting everyone to work, however, results in patterns in which black college-educated women have consistently lower levels of cumulative earnings compared to their white counterparts. Among those without a college degree, black women's disadvantage in cumulative earnings has lessened steadily across cohorts without adjusting for the racial differences in employment. This is due, in part, to the fact that black women make up for their disadvantage by staying more attached to the labor force than their white counterparts. These results highlight the importance of accounting for the cohort changes not just in

the level of earnings, but also in the changing employment rates, in studying the cohort trends in women's economic standings over the life cycle, as well as their between-group variations.

DATA AND SAMPLE

We use confidential data that combines the 2004 and 2008 panels of *Survey of Income and Program Participation* (SIPP) to administrative tax records (W-2) contained at SSA. Administered by the U.S. Census Bureau, the SIPP is a nationally representative household panel survey that collects detailed demographic, income, and program participation information (https://www.census.gov/programssurveys/sipp/). We pool the two panels to increase sample sizes. For each panel, we draw our main sample from wave 1 to avoid any attrition. For some analyses, we use information contained in wave 2 Topical modules containing women's retrospective marital and fertility histories.

The linked tax data (W-2) come from SSA's Detailed Earnings Record (DER) file, which provides SIPP respondents' annual earnings over the years 1980 to 2016 (McNabb et. al. 2009).[2] Annual earnings reflect a person's labor income for all jobs over the calendar year based on the W-2 forms that employers are required to submit to the Internal Revenue Service (IRS). These data are uniquely well suited for this study because they provide the earnings of the same individual for many years, before and after the actual SIPP panel, which is crucial for estimating cumulative earnings. The tax data also allow us to expand the number of cohorts we can observe resulting in a wider number of birth years that can be assessed relative to other cumulative panel surveys such as the PSID and NSLY. Other advantages include a measure of earnings that is observed frequently (annually), contains less measurement error than self-reports, are

[2] Linked data are confidential, anonymous, and analyzed by a user with Special Sworn Status at a Census-approved secure site. All estimates are reviewed by SSA's Disclosure Review Board for confidentiality before their public release. Our statistical programs are available upon request for researchers with access to these data.

not capped, and experience no attrition so long as individuals are matched to the administrative records (Kim and Tamborini 2014).

A limitation of the administrative data is the absence of information on outcomes such as hours worked, occupation, household income. Further, not all SIPP respondents are successfully matched to their records, although fortunately, the match rate is high (80% for 2004 SIPP panel; 90% for 2008 SIPP panel). Following past research (Couch, Tamborini, Reznik 2015), we adjust SIPP weights for unmatched respondents based on a logistic regression model that predicts a match probability from key characteristics. This adjustment helps maintain national representation of the samples and is applied throughout our analyses.

The birth years we use to anchor our cohort analyses span 35 years (1955-1980). These years contain the late baby boom generation (1955-1964) and early and later segments of Generation X (1965-1980).[3] This approach not only aligns well with available data but also provide the basis to compare the cumulative earnings of women who came of age during sharp changes in women's work and family lives (baby boom) with the patterns experienced by recent cohorts of women, which are not well established in the literature.

We assess women's annual earnings from ages 25 up to a maximum of age 50 for earlier cohorts. This choice allows us to compare cumulative earnings across cohorts over similar life course stages. For example, to assess earnings form ages 25 to 45, we observe annual earnings from 1980 to 1999 for women born in 1955 and from 1990 to 2010 for women born in 1970.

Our sample contains a number of restrictions. First, we exclude a small share of individuals with less than 2 years of positive earnings over their observation period.

[3] We acknowledge that generally naming cohorts can be arbitrary, but our definitions are in line with conventional definitions. In our empirical analysis, we will also construct the concept of "moving cohorts" to capture the continuous changes across birth years rather than assuming that the cohort shifts have occurred across arbitrarily specified cutoff years.

Immigrants are excluded for two reasons: we may lack their complete earnings history and their work patterns may reflect different selection processes. As noted above, we also only consider respondents who were successfully matched to the administrative records. Furthermore, based on a matched administrative variable, we exclude a very small fraction of women who received a disability benefit prior to age 25 from Social Security's disability insurance (DI) or needs-based Supplemental Security Income (SSI) programs. For most analyses, our sample is based on wave 1 of each SIPP panel to maximize sample size. When marital and fertility history variables are included, the sample is based on wave 2 data. The final baseline sample contains 620,184 person-years, which reflect the experiences of 28,445 workers.

Dependent Variables

The main dependent variable is the natural logarithm of cumulative earnings defined as the sum of annual earnings over various intervals of the life course. Earnings include all gross wages, salaries, tips, commissions, deferred compensation, and bonuses for all jobs. Taxable earnings from self-employment are excluded. To minimize outliers, we cap earnings at the 99.5 percentile. All earnings are inflation adjusted to 2016 dollars using the CPI-U.

We follow Goldin and Mitchell (2017) in generating an estimate of "labor force participation" by assuming that individuals are labor force participants if they earned more than some minimum amount—equivalent to 10 hours a week for 52 weeks at the federal minimum wage. Our sensitivity analysis will also examine the results under alternative threshold definitions, such as half of the federal minimum wage level specified above.

Independent variables.

Key independent variables are birth year, race-ethnicity, educational attainment and age. Race-ethnicity are self-reported (Non-Hispanic white, non-Hispanic black, Hispanic, other). Education reflects highest attainment completed by time of survey.

Note that the measure of educational attainment here reflects respondents' attainment at time of SIPP not of a given time in the past. We start the analysis at age 25 as most women would have completed their education by this point. We do not separate those with a graduate degree from those with a BA degree only because of lack of sufficient sample size, particularly among blacks. Sensitivity tests are undertaken that controls for year of final degree using SIPP's education history module which contains information on years of degree acquisition. We will also include field of study as a control variable.

Our analysis employs the concept of "moving cohorts" to avoid assuming arbitrary cut-off years for defining birth cohorts and to better capture a continuum of changes across cohort. These cohorts consist of a moving window of consecutive 5-year birth cohorts from 1955 to 1979. We use the middle year of each moving window to label the moving cohorts. For example, the 5-year moving window of individuals born between 1955 and 1959 will be labeled the "1957 moving cohort," because 1957 is the middle year in this window.

In additional analysis, we explore potential mechanisms behind the cohort trends in cumulative earnings. Part of this analysis will rely on SIPP's wave 2 marital and fertility history module, which allow us to construct retrospective measures of a person's marital biography (e.g. year and duration of marriage(s)) and fertility histories (e.g. parenthood status, age at first child's birth, number of children).

Estimation Strategy

We use descriptive and regression techniques to assess cohort changes in the pattern of cumulative earnings across the life course. To calculate cumulative earnings, we structure the data as person-year format, where each year is distinct observation. Cumulative earnings between age p and age q are constructed as $Y^{cum} = \sum_{p}^{q} Y_k$, where Y_k denotes the person's earnings at age k. If at a given age a person does not meet the minimal threshold for "labor force participation" as described earlier, we code $Y_k = 0$ for that year. We break out patterns by race and education to examine potential

heterogeneity in the cohort trends. Specifically, we will examine whether the cohort trends in the racial gap in cumulative earnings have unfolded differently for women with and without a college degree.

To further examine the "counterfactual" trajectories of cumulative earnings under the condition that every individual has worked in all years, we need to impute the "potential" earnings for person i at age k. We do this by estimating a multilevel growth curve model, with the first level being a person-year level earnings regression:

$$LogY_{ik} = \beta_{0i} + \beta_{1i} \cdot k + \beta_{2i} \cdot k^2 + \epsilon_{ik}$$

where the random intercept, random slope, and random coefficient on the quadratic term of age k are specified in the Level 2 models below:

$$\beta_{0i} = \gamma_{00} + \sum_1^m \gamma_{0j} X_j + u_{0i}$$

$$\beta_{1i} = \gamma_{10} + \sum_1^m \gamma_{1j} X_j + u_{1i}$$

$$\beta_{2i} = \gamma_{10} + \sum_1^m \gamma_{1j} X_j + u_{2i}$$

The X's contain a set of person-level characteristics, such as education, age, and cohort. Intuitively, this multilevel model allows each individual person to have a unique intercept, slope, and earnings growth deceleration rate, which together defines the person's earnings trajectory over life. Because these random coefficients are assumed to follow normal distributions within groups defined by the X's, we can consider the model as shrinking the person-level trajectories towards their group means.

Based on the multilevel models above, we predict the potential earnings for the labor force non-participants, termed $Y_k^{imputed}$, and calculate the employment-adjusted

cumulative earnings as: $Y^{cum,imputed} = \sum_p^q Y_k + I_{missing} Y_k^{imputed}$, where $I_{missing}$ is an indicator variable which equals 1 if the person is a non-participant at year k.

PRELIMINARY EMPIRICAL RESULTS

Figure 1 presents the observed and employment-adjusted cohort trends in logged cumulative earnings for white and black women among those with (left panels) and without a college degree (right panels). The cumulative earnings are calculated for age 25-35 and 25-40 respectively. The figures reveal remarkable differences in the trajectories of cohort trends by college degree. Among college-educated women, cumulative earnings for both races have been steadily rising from those born in the late 1950s to those born in the 1970s. But the rate of increase appears to be relatively similar for whites and blacks. Adjusting for employment by hypothetically putting everyone to work at all ages increases white women's cumulative earnings slightly more than it does for black women's earnings. But overall, there does not seems to be much progress in terms to the narrowing of the racial earnings gap.

Turning to women without a college degree, the figure suggests that the cumulative earnings for white women have increased moderately from those born in the late 1950s to those born in the early 1960s, but the cohort trend has stayed relatively flat since then. In contrast, black women's cumulative earnings have increased more dramatically from those born in the late 1950s to those born in the 1960s, resulting in a steady narrowing of the black-white earnings gap among this group. The narrowing slowed down in more recent cohorts (i.e. those born after the 1970s), with the gap staying largely unchanged since then. Adjusting for labor force participation by (hypothetically) putting everyone to work at all ages increases the cumulative earnings for both races, but the increases is greater for whites than for blacks. This suggest that black women in this group have partly made up for their earnings disadvantage relative to their white counterparts by staying more attached to the labor force over the early and middle stages of their career.

Figure 2 expresses black women's cumulative earnings as a percentage of white women's cumulative earnings. If the percentage equals one, then the two races have equal cumulative earnings on average. A value lower than one indicates an earnings disadvantage among black women, and vice versa. Among college graduates, the racial difference in cumulative earnings between age 25 and 35 and between age 25 and 40 followed similar trends across cohorts: it fluctuates around one for most of the cohorts, but black women's earnings relative to white women decreased slightly from those born in the late 1950s to early 1960s, followed by a moderate climb back later. But overall, without adjusting for employment, black college-educated women seem to accumulate similar, if not higher, cumulative earnings as their white counterparts during early and middle career stages. In contrast, among those without a college degree, black women's relative earnings decreased among those born in the last few years of the 1950s, but they climb back steadily among those born around 1960 and later, up from around 80% to about 95%, only 5% lower than their white counterparts. The closing of the racial gap in cumulative earnings among those without a college degree is partly attributable to black women's higher labor force participation in early and middle career – adjusting for the racial difference in employment rate indeed lowers black women's relative earnings.

CONCLUSION AND FUTURE STEPS

Most prior studies on the improvements in women's economic standings have relied on cross-sectional data, in which earnings are observed among the snapshot of women who are employed at a certain age. We argue that, given that women have experienced dramatic changes not just in terms of earnings but also in their life-cycle profile of labor force participation over the past half-century, cross-sectional estimates may not be adequate to fully understand the differential economic positions experienced by different cohorts of women. In contrast, the measure of cumulative earnings based on longitudinal earnings and employment trajectories can provide a better understanding of the long-term economic positions of women over the life course.

Specifically, our analysis takes advantage of administrative earnings data linked to national surveys to examines the cohort trends in women's cumulative earnings by simultaneously considering their changes in earnings levels as well as changes in labor force participation. Taken together, our preliminary analysis suggests that the racial gap in cumulative earnings has persisted among women with a college degree, but has closed steadily across cohorts among women without a college degree. This narrowing gap in cumulative earnings is driven in part by black women's greater increase in labor force participation relative to their white counterparts.

In future analysis, we will consider the impact of non-employment on subsequent earnings as an additional mechanism through which labor force participation affects cumulative earnings. We will also take advantage of additional information from the SIPP data to examine the extent to which these cohort changes can be attributed to changes in women's marital and fertility histories.

Works Cited

Blau, F. D., & Kahn, L. M. (2000). Gender differences in pay. *Journal of Economic perspectives*, *14*(4), 75-99.

Couch, Kenneth A., Christopher R. Tamborini, and Gayle L. Reznik. 2015 "The Long-Term Health Implications of Marital Disruption: Divorce, Work Limits, and Social Security Disability Benefits Among Men." *Demography* 52, no. 5: 1487-1512.

Damaske, S. (2011). *For the family? How class and gender shape women's work*. New York, NY: Oxford University Press.

Damaske, Sarah, and Adrianne Frech. 2016. "Women's work pathways across the life course." *Demography* 53(2): 365-391.

Elder, Glen H., Jr. 1994. "Time, Human Agency, and Social Change: Perspectives on the Life Course." *Social Psychology Quarterly* 57:4–15.

Elder, Glen H., Jr., Monica Kirkpatrick Johnson, and Robert Crosnoe. 2003. "The Emergence
and Development of Life Course Theory." Pp. 3–19 in *Handbook of the Life Course*, edited by J.
T. Mortimer and M. J. Shanahan. New York: Kluwer.

England, Paula (2010). The gender revolution: Uneven and stalled. *Gender & society*, *24*(2), 149-166.

Goldin, Claudia, and Joshua Mitchell. 2017. "The New Life Cycle of Women's Employment: Disappearing Humps, Sagging Middles, Expanding Tops." *Journal of Economic Perspectives* 31(1): 161-82.

Goldin, Claudia. 2006 "The quiet revolution that transformed women's employment, education, and family." *American Economic Review* 96(2): 1-21.

Goldstein, Joshua and Catherine Kenney. 2001. "Marriage Delayed or Marriage Forgone? New
Cohort Forecasts of First Marriage for U.S. Women." *American Sociological Review* 66:506–519.

Killewald, Alexandra, Fabian T. Pfeffer, and Jared N. Schachner. 2017. "Wealth inequality and accumulation." *Annual Review of Sociology* 43: 379-404.

Kim, ChangHwan, and Christopher R. Tamborini. 2014. "Response error in earnings: An analysis of the Survey of Income and Program Participation matched with administrative data." *Sociological Methods & Research* 43(1): 39-72.

McNabb, Jennifer, David Timmons, Jae G. Song, and Carolyn Puckett. 2009. “Uses of Administrative Data at the Social Security Administration.” *Social Security Bulletin* 69(1):75–84.

Percheski, Christine. 2008 "Opting out? Cohort differences in professional women's employment rates from 1960 to 2005." *American sociological review* 73(3): 497-517.

Pettit, Becky, and Stephanie Ewert. 2009. "Employment gains and wage declines: The erosion of black women’s relative wages since 1980." *Demography* 46(3): 469-492.

Sakamoto, Arthur, Christopher R. Tamborini, and ChangHwan Kim. (2018). "Long-Term Earnings Differentials Between African American and White Men by Educational Level." *Population Research and Policy Review* 37(1): 91-116.

Stevenson, B., & Wolfers, J. (2007). Marriage and divorce: Changes and their driving forces. *Journal of Economic perspectives*, *21*(2), 27-52.

Sweeney, Megan M. 2002. “Two Decades of Family Change: The Shifting Economic Foundations of Marriage.” *American Sociological Review* 67:132–147.

Tamborini, Christopher R., ChangHwan Kim, and Arthur Sakamoto. (2015). "Education and lifetime earnings in the United States." *Demography* 52(4): 1383-1407.

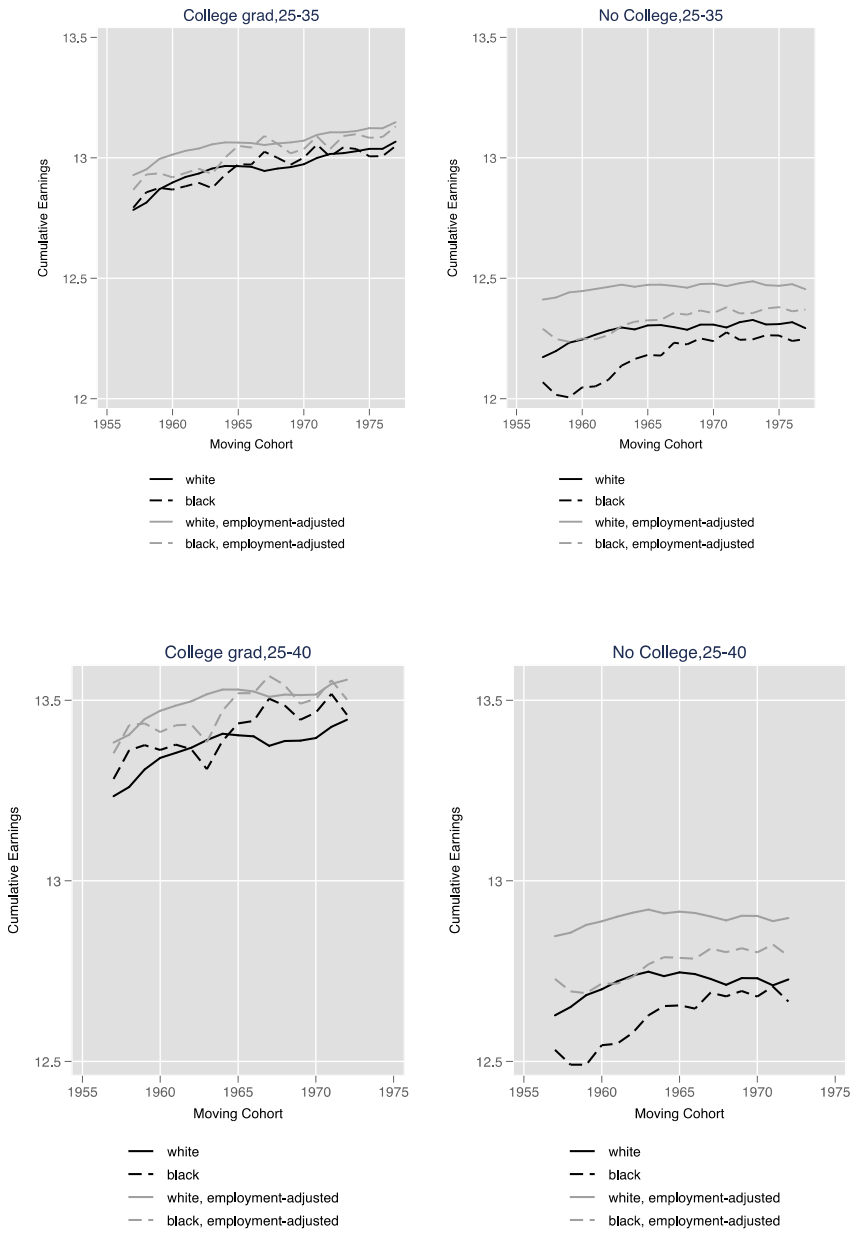
Tamborini, C. R., & Iams, H. M. (2011). Are generation X'ers different than late Boomers? Family and earnings trends among recent cohorts of women at young adulthood. *Population Research and Policy Review*, 30, 59–79.

Tamborini, Christopher R. (2007). "The Never-Married in Old Age: Projections and Concerns for the Near Future." *Social Security Bulletin* 67(2):25–40.

Thornton, Arland and Linda Young-DeMarco. (2001). "Four Decades of Trends in Attitudes toward Family Issues in the United States: The 1960s through the 1990s." *Journal of Marriage and the Family* 63(4):1009–1037.

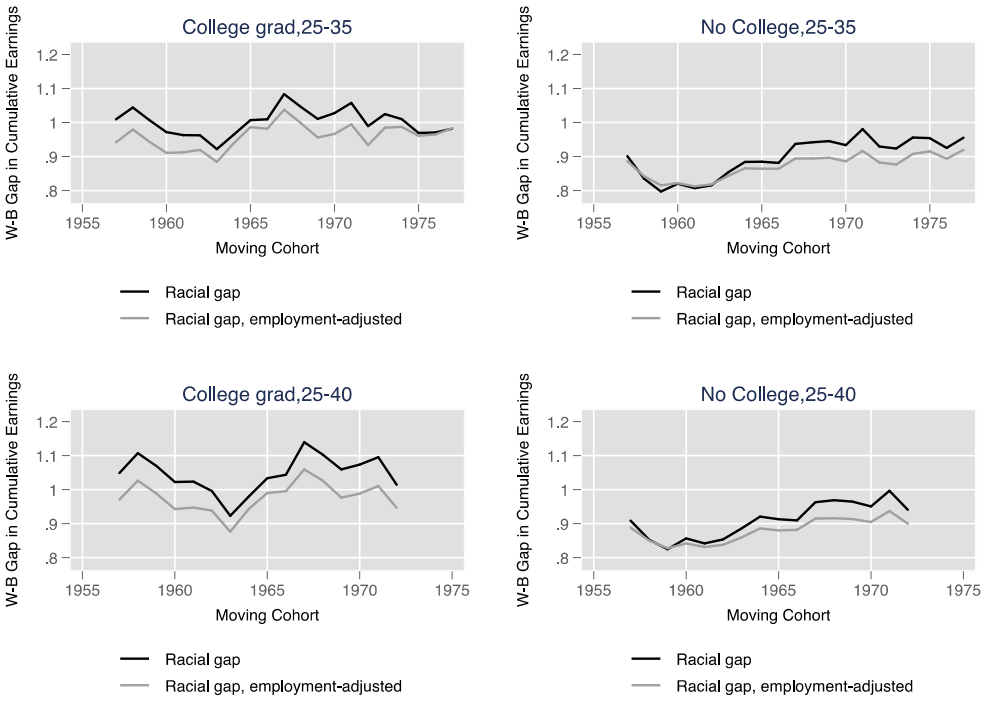
Vere, J. P. (2007). Having it all no longer: Fertility, female labor supply, and the new life choices of generation X. *Demography*, 44(4), 821–828.

Figure 1. Observed and Employment-adjusted Cohort Trends in Logged Cumulative Earnings by Race and Education.



NOTE: The moving cohorts consist of a moving window of consecutive 5-year birth cohorts from 1955 to 1979. We use the middle year of each moving window to label the moving cohorts. For example, the 5-year moving window of individuals born between 1955 and 1959 will be labeled the "1957 moving cohort," because 1957 is the middle year in this window.

Figure 2. Observed and Employment-adjusted Black-white Earnings Ratio in Cumulative Earnings by Education



NOTE: The moving cohorts consist of a moving window of consecutive 5-year birth cohorts from 1955 to 1979. We use the middle year of each moving window to label the moving cohorts. For example, the 5-year moving window of individuals born between 1955 and 1959 will be labeled the "1957 moving cohort," because 1957 is the middle year in this window. The W-B gap in cumulative earnings is measured by expressing the average cumulative earnings among black women as a percentage of those among their white counterparts.