UNIVERSITY OF KENTUCKY
CENTER FOR POVERTY RESEARCH

Discussion Paper Series
DP 2020 - 01

ISSN: 1936-9379

Trends in earnings volatility using linked and administrative survey data

James P. Ziliak
Center for Poverty Research
University of Kentucky

Charles Hokayem
Kentucky Research Data Center
U.S. Census Bureau

Christopher R. Bollinger
Department of Economics
University of Kentucky

September 2020

Preferred citation:
Ziliak, J., Hokayem, C., and Bollinger, C. (2020, Sept.). Trends in earnings volatility using linked and administrative survey data. *University of Kentucky Center for Poverty Research Discussion Paper Series, DP2020-01*, Retrieved [Date] from http://ukcrp.org/research.

Author Correspondence
James P. Ziliak, jziliak@uky.edu

University of Kentucky Center for Poverty Research
550 South Limestone,
234 Gatton Building, Lexington, KY, 40506-0034
Phone: 859-257-7641. E-mail: ukcpr@uky.edu
ukcpr.org

Trends in Earnings Volatility using Linked Administrative and Survey Data*

May 2020

James P. Ziliak, University of Kentucky
Charles Hokayem, U.S. Census Bureau
Christopher R. Bollinger, University of Kentucky

Abstract: We document trends in earnings volatility separately by gender in combination with other characteristics such as race, educational attainment, and employment status using unique linked survey and administrative data for the tax years spanning 1995-2015. We also decompose the variance of trend volatility into within- and between-group contributions, as well as transitory and permanent shocks. Our results for continuously working men suggest that trend earnings volatility was stable over our period in both survey and tax data, though with a substantial countercyclical business-cycle component. Trend earnings volatility among women declined over the period in both survey and administrative data, but unlike for men, there was no change over the Great Recession. The variance decompositions indicate that nonresponders, low-educated, racial minorities, and part-year workers have the greatest group specific earnings volatility, but with the exception of part-year workers, they contribute least to the level and trend of volatility owing to their small share of the population. There is evidence of stable transitory volatility, but rising permanent volatility over the past two decades in male and female earnings.

Keywords: CPS ASEC, earnings volatility, nonresponse, administrative tax data

JEL Codes: J31 Wage Level and Structure, C8 Data Collection and Estimation Methodology

---

* Corresponding Author: James P. Ziliak, Center for Poverty Research, University of Kentucky, 550 South Limestone Street, Lexington, KY 40506-0034; jziliak@uky.edu. Bollinger and Ziliak are grateful for the financial support of the Washington Center for Equitable Growth and National Science Foundation grant 1918828. We appreciate the comments of Robert Moffitt, Michael Keane, Claudia Sahm, and seminar participants at the Winter Econometric Society Meetings, the Society of Labor Economists, the Western Economic Association, the JPMorgan Chase Institute Conference on Economic Research, the Longitudinal and Administrative Data for Distributional Analysis Workshop at University of Essex, University of Pittsburgh, and Washington Center for Equitable Growth Measuring Volatility Workshop. The analysis in this paper was conducted at the Kentucky Research Data Center. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed by the Disclosure Review Board to ensure that no confidential information is disclosed (DRB #2018-457, DRB #2019-322, DRB #2019-408, DRB #2019-449, DRB #2019-551, CBDRB-FY20-POP001-0076, CBDRB-FY20-POP001-0100).

Understanding the level and trend of earnings volatility is important both in its own right, and because of its potential contribution to rising inequality. If volatility is mostly transitory, then it is less likely to have long-term negative effects on lifetime mobility than if it is permanent, the latter of which is exacerbated by the fact that the tax and transfer system is less effective at insuring against permanent shocks (Kniesner and Ziliak 2002; Blundell et al. 2008). Thus, not only do we need to know the level and trend of volatility overall, but also its persistence. Much of what we know about volatility in the United States has come from survey data, the preponderance of which was obtained from the Panel Study of Income Dynamics (PSID). The general consensus from the PSID is that transitory instability increased from the early 1970s until the mid 1980s, and stabilized until 2000—a macroeconomic period known as "The Great Moderation" because of stability in policy, inflation, and the business cycle (Gottschalk and Moffitt 1994, 2009; Haider 2001; Stock and Watson 2003; Hacker and Jacobs 2008; Keys 2008; Dynan, Elmendorf, and Sichel 2012; Shin and Solon 2011). Permanent (lifetime) instability rose primarily in the 1980s. While there is corroborating evidence on the basic PSID trends up until 2000 from other surveys (Gittleman and Joyce 1996; Cameron and Tracy 1998; Dahl, DeLeire, and Schwabish 2011; Ziliak, et al. 2011; Celik et al. 2012; Carr and Wiemers 2018), beginning in the 2000's there is a sharp divergence between the PSID and some other survey-based estimates (Moffitt and Zhang 2018). More worrying is recent evidence from administrative data that calls into question the basic conclusion of whether volatility increased at any point since 1980 (Sabelhaus and Song 2010; Bloom et al. 2018). In this paper, and in conjunction with the other contributors to this special issue (Carr, Moffitt, and Wiemers 2020; McKinney and Abowd 2020; Moffitt and Zhang 2020), we aim to reconcile differences in volatility levels and trends in survey and administrative data by using linked data on the same individuals.

Survey data is generally advantageous because it offers a broad collection of variables, a long time series, population representativeness, and widespread availability to the research community. However, survey data suffers from data quality issues such as nonresponse and measurement error, the latter of which may include response error or survey reporting policy such as topcoding (Mellow and Sider 1983; Lillard, Smith, and Welch 1986; Bollinger, 1998; Bound, Brown and Mathiowetz, 2001;

Roemer, 2002; Hirsch and Schumacher 2004; Bollinger and Hirsch 2006; Kapteyn and Ypma, 2007; Meijer, Rohwedder, and Wansbeek, 2012; Abowd and Stinson, 2013; Bollinger, Hirsch, Hokayem, and Ziliak 2019). More recently, some scholars have turned to administrative data to examine volatility on the belief that it avoids some of the pitfalls of surveys (Sabelhaus and Song 2010; Bloom et al. 2018; Carr and Wiemers 2018). However, the standard assumption that administrative data serve as a "gold standard" has been challenged by some (Kapteyn and Ypma 2007; Abowd and Stinson 2013), and the populations covered between the survey and administrative samples are often quite different.

We present new estimates of earnings volatility using restricted-access survey data from the Current Population Survey Annual Social and Economic Supplement (ASEC) linked to the Social Security Administration's Detailed Earnings Record (DER) for the period spanning 1995-2015.[1] The ASEC is a large, nationally representative survey that serves as the source of official statistics on poverty and inequality, and is the workhorse dataset for research on earnings determinants. The DER reflects earnings reports provided by employers and the self-employed for purposes of payroll taxation and eligibility for Social Security retirement and disability programs. The advantage offered by the link to administrative data is that today nearly 45 percent of survey earnings responses in the ASEC are missing due to nonresponse from either refusing to respond to the ASEC in whole or failing to respond specifically to the earnings questions. The Census Bureau uses a randomly selected "donor" with similar characteristics to replace the entire ASEC (whole imputes) or missing earnings (item imputes), the so-called hot-deck imputation procedure. The Census process relies on the assumption that the data are missing at random, or that conditional on the observables, nonresponse is ignorable. Bollinger et al. (2019) find that the data are not missing randomly, and the current imputation procedure can lead to substantial bias in the tails of the earnings distribution, while Hirsch and Schumacher (2004) and

[1] The linked ASEC-DER were obtained as part of an internal-to-Census project (DSM1170) and analyzed in a secure federal facility at the Kentucky Research Data Center in Lexington, Ky. Researchers outside of Census interested in accessing such data must have their project approved by Census and the Social Security Administration for analysis conducted in a secure Federal Statistical Research Data Center. For more information see https://www.census.gov/fsrdc .

Bollinger and Hirsch (2006) show that the hot-deck procedure can lead to significant bias in estimated coefficients in standard Mincer-type earnings models. The link to administrative data offers an independent report on the worker and therefore does not require the use of imputed values from the Census Bureau. We are thus able to compare year-to-year estimates of earnings volatility based on two-year responders (i.e. those with non-imputed earnings in both years), two-year nonresponders, and those who switch response status.

Another crucial advantage of the linked data, and the primary focus of this paper, is that the current literature has reached differing conclusions on the trend in earnings volatility both across surveys and in comparing survey-alone to administrative-alone estimates. The survey data typically used may differ from the administrative data because of sampling frames, measurement, and survey response and non-response. Thus, it is difficult to know how much of the difference in trends is due to differences in measurement between survey and administrative reports, as opposed to differences in samples. The exact ASEC-DER link we use eliminates differences due to sample frames, and thus permits us to focus on differences in continued survey participation (attrition), item non-response, and measurement between survey and administrative data reports of earnings volatility. We find that much of the difference depends on how reports of zero earnings and item non-response to the ASEC earnings questions are handled. Surveys are crucially important because they bring together both earnings and socio-economic data unavailable from administrative data. However, careful handling of these data is necessary.

Our analysis begins with overall trends in earnings volatility, utilizing two measures of variance—the arc percent change and the first difference in log earnings. The arc percent change is bounded between ± 200%, and most important, it is still defined if earnings are zero in one of the two years, unlike the log growth measure which requires positive earnings in both years. Ziliak et al. (2011) found that earnings volatility was increasingly accounted for by employment transitions from the early 1970s to the mid 2000s, and these transitions are missed with the standard log-growth volatility measure. Like most of the literature, our estimates of volatility remove life-cycle factors, and we show how the series differ depending on whether and how we trim residual outliers. We conduct detailed sensitivity checks on the effects of earnings nonresponse, survey attrition, and reweighting. In the second part of the paper, we move beyond the aggregate volatility measures to explore trends across heterogenous subgroups. We decompose the variance into within- and between-group contributions, focusing on the roles of human capital, full-time work, race/ethnicity, and earnings response status. We follow with a further decomposition of the variance into trends in transitory and permanent volatility.

This work has important advantages over the recent contributions of Carr and Wiemers (2018), Sabelhaus and Song (2010), and Bloom et al. (2018). First, Carr and Wiemers had limited access to demographic data in the SIPP-DER file they used, and were unable to isolate the potential roles of human capital and family headship on earnings volatility. The PSID research is restricted to heads of households, and the SIPP-DER file does not release family structure information, making it difficult to do a direct comparison with the PSID. In our data we have access to the full ASEC, including education attainment, marital status, race, and relationship to head to examine whether some of the divergence in trends is from workers other than the household head. Second, while Sabelhaus and Song (2010) and Bloom et al. (2018) have access to a larger universe of DER workers, they have limited demographic information, notably missing information on education, race, and family structure, which we have with the ASEC-DER link. One important difference with the Sabelhaus and Song (2010) analysis is they pool men and women together, whereas most of the survey-based research focuses only on men. Ziliak et al (2011) conducted their analyses separately for men and women, showing that volatility of women declined from the 1970s to the 2000s. This could partially account for the declining volatility reported in Sabelhaus and Song, and thus we conduct our analyses separately for men and women.

Our results for men suggest overall earnings volatility among those workers reporting positive earnings across periods was stable in both the survey data and tax data and hence were essentially identical, at least when imputed survey observations are excluded. However, the results show declining volatility among women. When zeroes are included in the ASEC data, the results show that volatility among men increased in survey data, but was stable in tax data. For women, the survey data show stable volatility, while tax data indicate a secular decline. Importantly, for men there is a substantial countercyclical component to earnings volatility that is not present among women. The variance decompositions for both men and women indicate that nonresponders, low-educated, racial minorities, and part-year workers have the greatest group specific earnings volatility, but with the exception of part-year workers, they contribute least to the level and trend of volatility owing to their small share of the population. The finding that nonresponse in the ASEC has an upward bias in estimates of volatility adds to the evidence in Hokayem et al. (2015) and Bollinger et al. (2019) on the pitfalls of hot-deck imputations for estimation of parameters associated with earnings. This finding seems more binding for male earnings volatility. Finally, for both men and women, there has been a trend increase in permanent volatility, suggesting widening inequality and declining mobility in the American workforce.

II. Measuring Volatility

We adopt two summary measures of earnings volatility that are typical in the literature. The first is the variance of the arc percent change (Ziliak et al. 2011; Dynan et al. 2012), defined as

(1) $$varc_t = 100 * Var\left(\frac{y_{it} - y_{it-1}}{\bar{y}_i}\right),$$

where $\bar{y}_i$ is the average (absolute value) earnings across adjacent years, $\bar{y}_i = \frac{|y_{it}|+|y_{it-1}|}{2}$. The second measure is the variance of the change in log earnings (Shin and Solon 2011; Moffitt and Zhang 2018),

(2) $$vlog_t = 100 * Var(lny_{it} - lny_{it-1}) ,$$

where lny_{it} as the natural log of earnings for individual *i* in time period *t*.

A limitation of the change in logs measure of volatility is that log earnings are undefined if earnings are not positive. Employment rates, as depicted in Figure 1, have declined for men for the past 40 years, especially for low skilled males, while employment for women has declined since the peak in the late 1990s. Hence, a larger proportion of earners will have zero earnings in some years, and removing these earners, especially from first difference calculations, likely understates true earnings volatility. Moreover, a loose attachment to the labor force may lead to misreporting of earnings in survey data, or may lead to missing earnings from uncovered or informal labor markets. Both of these factors would contribute to differences in the earnings volatility measures between survey and administrative data. The arc percent measure has the advantage that those with zero or even negative earnings can be included in the computation without imposing adjustments to the levels to allow log transformations. In addition, the arc percent change is bounded between ± 200%, which facilitates interpretation. However, the symmetry property is violated if earnings are negative one year, say due to a business loss, and positive the next, and thus we take the absolute value of earnings in the denominator.

[Figure 1 here]

Because earnings volatility can be affected by life-cycle factors (Gottschalk and Moffitt 1994), we first regress the arc percent change (or log diff) on a quadratic in age year-by-year as

$$(3) \qquad \frac{y_{it}-y_{it-1}}{\bar{y}_i} = \alpha_t + \beta_t age_{it} + \gamma_t age_{it}^2 + \nu_{it}$$

and then use the estimated residuals $\hat{\nu}_{it}$ in equation (1) (equation (2) for log diff) prior to constructing the variance. In order to mitigate the influence of outliers, we trim the top and bottom 1 percent of the real annual cross-sectional ASEC and DER earnings distributions prior to estimating the age-adjustment in equation (3). We also present baseline estimates with a 5 percent trim, and without trimming.

III. Data

The data used in our analysis are restricted-access ASEC person records linked to the DER for survey years 1996-2016 (reporting earnings for tax years 1995-2015). The ASEC is a survey of roughly 90,000 households (60,000 from the usual CPS monthly rotation plus an additional 30,000 households oversampled as part of the Children's Health Insurance Program) conducted in March of each year. It serves as the source of official federal statistics on income, poverty, inequality, and health insurance coverage, and has been the primary survey dataset for earnings inequality research in the U.S. The difference between the internal ASEC and the public version is that the internal file has higher topcode values on income components.[2]

[2] The internal ASEC file has a topcode of $1.099 million for each earnings component (wage and salary, self employment), as opposed to $250,000 in the public ASEC. In the public files the Census Bureau replaces the topcoded value with a value obtained from rank proximity swapping, which is order preserving of the distribution above the topcode. Rank swapping was begun with the 2011 survey, but the Bureau released the corresponding values back to 1975 at https://www2.census.gov/programs-surveys/demo/datasets/income-poverty/time-series/data-

A. DER Linkage

We link the internal ASEC to the DER file, which is an extract of the Master Earnings File and includes data on total earnings as reported on a worker's W-2 form, wages and salaries and income from self-employment subject to Federal Insurance Contributions Act and/or Self-Employment Contributions Act taxation, as well as deferred wage (tax) contributions to 401(k), 403(b), 408(k), 457(b), and 501(c) retirement and trust plans, all of which we include in our earnings measure. Only positive self-employment earnings are reported in the DER because individuals do not make self-employment tax contributions if they have losses. In addition, some parts of gross compensation do not appear in the DER such as pre-tax health insurance premiums and education benefits, nor do "off-the-books" earnings appear in the DER, though they could be reported in the ASEC. Unlike the internal ASEC earnings records, DER earnings are not topcoded. This is important given substantial concerns regarding nonresponse and response bias in the tails of the distribution (Bollinger et al. 2019). Since a worker can appear multiple times per year in the DER file if they have multiple jobs, we collapse the DER file into one earnings observation per worker per year by aggregating total earnings, total self-employment earnings, and total deferred contributions across all employers. In this way, DER earnings are most compatible with ASEC earnings from all wage and salary jobs plus non-negative self-employment earnings.

The DER is linked to the ASEC using a unique Protected Identification Key (PIK) produced within the Census Bureau's Economic Reimbursable Surveys Division. The PIK is a confidentiality-protected version of the Social Security Number (SSN). Since the Census does not currently ask respondents for a SSN, Census uses its own record linkage software system, the Person Validation System, to assign a SSN. This assignment relies on a probabilistic matching model based on name, address, date of birth, and gender. The SSN is then converted to a PIK in order to link the ASEC and DER. The Census Bureau changed its consent protocol to link respondents to administrative data beginning with the 2006 ASEC. Prior to this the CPS collected respondent SSNs and an affirmative

extracts/asec-incometopcodes-swappingmethod-corrected-110514.zip . Bollinger et al. (2019) recommend replacing ASEC topcodes with rank swapped values for earnings research, especially that isolating the upper tail.

agreement allowing a link to administrative data; i.e., an "opt-in" consent option. Beginning with the 2006 ASEC, respondents not wanting to be linked to administrative data had to notify the Census Bureau through the survey field representative, website or use a mail-in response in order to "opt-out". This opt-out rate is a very small 0.5 percent of the ASEC sample. If the respondent doesn't opt out, they are assigned a PIK using the Person Validation System. Once a link to the ASEC is made, we have the entire time-series of DER earnings as far back as 1978 when the DER began. However, because our focus is on the exact match for overlapping years of the ASEC and DER, we do not use this additional information in the DER. To the extent that the ASEC is representative of the U.S. population, and the rate of non-PIKed is low, the volatility estimates obtained here with the DER should not differ from those using the broader sample from the Master Earnings File (e.g. Bloom et al).

B. CPS ASEC Panels

In order to construct measures of volatility we must follow the same individual over time. Here we exploit the fact that the ASEC has a rotating sample design whereby respondents are in-sample for 4 months, out-of-sample for 8 months, and then in-sample for 4 more months. This makes it possible to match up to one-half of the sample from one ASEC interview to the next, and thus creating a series of two-year panels.[3] Following the procedure recommended by the Census Bureau and extended by Madrian and Lefgren (1999), we initially match individuals based on four variables: month in sample (months 1–4 for year 1, months 5–8 for year 2); line number (unique person identifier); household identifier; and household number. Because the CPS sample domain is household addresses and not individuals, if a person moves between ASEC surveys then the Census Bureau interviews the new occupant at the address and does not follow the original respondent. Thus, we then do a cross check against four additional variables to make sure gender, race, and state of residence are unchanged, and that age changes by no more than two years (in case of staggered March interview, which actually spans February – April). This is the approach used by Cameron and Tracy (1998) and Ziliak et al. (2011), along with many others only

[3] The CHIP oversample in the ASEC is not eligible for the longitudinal follow-up, and thus we exclude it.

with access to public files. We link the ASEC to the DER prior to constructing the panel.[4] Appendix Table 1 contains the annual ASEC-DER linkage rate along with the two-year panel match rate. It is clear that the link to the DER improved substantially after Census adopted the opt-out default in 2006. We match about 72 percent of persons across March surveys. Below we discuss sensitivity of estimates to adjustments for nonlink and nonmatch.

C. Sample Summary Statistics

The principal sample used for the volatility measures is people between the ages of 25-59 who have positive earnings in at least one year, are respondents to the ASEC earnings questions in both years, and have a link to the DER in both years. We refer to this group as the linked respondent sample. We also remove individuals who are full-time students in any year or that have their entire ASEC supplement allocated. Some individuals respond to the monthly core of the CPS, but are unwilling or unable to provide a response to the ASEC supplement. For these cases, Census uses a sequential hot-deck procedure to replace the individual's entire ASEC supplement with a donor's supplement (called a whole imputation). During our sample period, roughly 12 percent of individuals had their entire ASEC imputed and so we drop these individuals, though we account for this in our attrition analysis below.

The linked respondent sample is intentionally restrictive because we wish to conduct an "apples-to-apples" comparison of volatility estimates from survey data against administrative data with a sample and measure as similar as possible. However, because the ASEC sample is much broader than the linked respondent sample, we also conduct our analyses with a sample of individuals who may be an ASEC earnings nonrespondent in one or both years or who may not have a link to the DER in either or both years, which we call the full sample. Similar to the whole imputes discussed above, Census also uses a sequential hot-deck procedure to impute earnings for individuals who otherwise responded to the ASEC, but did not provide a response to the earnings questions. The key assumption in the hot-deck procedure is missing at random (MAR). Bollinger et al. (2019) show that the economic consequences of the MAR

[4] Because we first link the internal ASEC to the DER prior to the longitudinal match, we also have the individual's unique PIK that can be used to match from one year to the next. Results are qualitatively similar using the PIK.

assumption for earnings levels is primarily in the tails of the distribution, and in this paper we extend that earlier analysis to earnings volatility. The full sample, and in particular the full sample of two-year respondents, provides a more comprehensive estimate of labor-market volatility because they may or may not have a DER link.[5]

One of our aims is to capture a broad measure of volatility in the labor market, including the impact of movements in and out of employment. As such, for much of our analysis we provide estimates from the arc percent measure with and without zero earnings. Notably, an earnings report of zero in the ASEC could be from nonwork, or it could be from misreporting by the respondent. That is, they could self-report zero earnings in the ASEC, but the firm could submit a positive earnings W2 that is included in the DER. There are reports of zero earnings in the DER, although this is rare, and likely reflects misreports on the part of the firm or self-employed worker.[6] Strictly the two-year linked respondent sample should be a sample in which earnings are positive in both the ASEC and DER; however, it is possible for there to be zero earnings in one or both the ASEC and DER if there exists measurement error in one or both surveys.

[Table 1 here]

Table 1 provides pooled summary statistics for our full sample and sample of linked respondents, separately for men and women and weighted by the ASEC person supplement weight. In the full sample, the average person is 43 years old, and has an average of about 14 years of education. The majority are married with spouse present (64 percent of men; 62 percent of women), native born (83 percent of men; 85 percent of women), and White non-Hispanic (72 percent). Men work for pay an average of 48 weeks per year, and 42 hours per week, while women on average work for pay 46 weeks and 36 hours per week. Inflation adjusted ASEC total earnings for men are on average \$59,000 (\$38,470 women), while average

[5] Bollinger et al. (2019) show, for example, that non-citizen Hispanics have comparable response rates to native-born non-Hispanics, but are much less likely to be linked to the DER. These persons do not show up in DER volatility.

[6] We communicated with two separate staff members at the Social Security Administration and received conflicting responses on whether the \$0s are valid entries in the DER. As this affects only about 0.1 percent of the DER sample we have retained them as valid for completeness.

real DER earnings are a higher \$66,290 (\$41,630), likely reflecting the fact that the DER are not topcoded.[7] Among men, 84 percent have a DER link in both panel years, while 5 percent have a DER link in one but not the other panel year (85 and 7 percent, respectively, for women). For both men and women, the sample of linked respondents (linked in both panel years) is more educated, works more weeks and hours per week, is more likely to be White, and to be native born. Linked respondents have higher ASEC earnings than the full sample, but DER earnings are comparable.

[Figure 2 here]

Figure 2 depicts trends in selected percentiles of the male and female real earnings distributions. For this figure we require nonzero earnings and linked respondents, but do not trim the top and bottom for outliers. The figure shows that the earnings distribution for men in the DER is shifted leftward by 20-80 percent compared to the ASEC for percentiles below the 25th, but then the DER generally has a longer right tail than the ASEC. While percent differences in the left tail are large, the absolute dollar values are not, differing by \$1,000-\$2,000. In the upper quantiles the DER exceeds the ASEC on average by \$2,400 at the 95th percentile and just under \$19,000 at the 99th, which is consistent with the lack of a topcode in the DER. There does appear to be a more substantial decline at the 99th percentile of the ASEC leading up to the Great Recession, and more rapid recovery, but overall the growth in real earnings is comparable across the ASEC and DER (e.g. 34 percent in both sources at the 1st percentile, and 59 percent and 65 percent in the ASEC and DER, respectively at the 99th percentile). The story is different for women in that across most of the distribution the ASEC and DER differ little. There is considerable noise in the lowest percentile, with the ASEC exceeding the DER in some years, and the DER exceeding the ASEC in others, but overall there is little change. To see whether the differences in the survey and administrative distributions of men in the left tail might lead to different estimates in volatility, Appendix Figure 1 shows a parallel figure for the distribution of residuals from the log difference regression (which omits 0s by construction). In this case, the ASEC series are quite comparable to the DER. Thus, based on the log

[7] Earnings are inflation-adjusted using the personal consumption expenditure deflator with 2010 base year.

changes it does not appear that there are fundamental differences in the residualized distributions in the ASEC and DER for men that prima facie point to any potential differences in volatility.

IV. Results

We present results for several samples. The first is the narrowest sample—individuals who responded to the ASEC earnings question in both years and who were able to be linked to the DER data in both years. This should be the sample with most accurate ASEC earnings reports, and comparing volatility in the ASEC with that in the DER is the place to start. We then broaden the sample in stages, by first expanding the sample to include those who did not respond to the ASEC earnings in one or both years, but still requiring the DER linkage in both years, and then also expanding the sample further by including those who did not have a link. We then go even further by including those who were missing from the ASEC in the second year because of attrition, and examine ASEC and DER volatility in that sample. The baseline volatility estimates come from our preferred arc-percent change measure and with a 1% trim of outliers. The top panel of Figure 3 only requires men to have earnings in one of two years, while the bottom panel restricts to those with positive earnings in both years.

[Figure 3 here]

The key takeaway from Figure 3 is that while allowing for zero earnings has a substantive effect on the level and trend of male earnings volatility in the ASEC, zero earnings has little effect on the level of DER volatility, and there is no trend in DER volatility regardless of inclusion of zeros or level of trimming of residuals. The latter is unsurprising because of the rare reports of zeros in the DER. The top panel shows that the level of ASEC volatility in a typical year is about double compared to the bottom panel of continuous earners. Volatility increased between 30-40 percent in the sample period in panel A, compared to no trend in volatility for the series excluding ASEC zeros in panel B. In both panels, there is a notable uptick in volatility in the years surrounding the Great Recession, but there was a return to pre-recession levels among the subsample of continuous workers. There is a substantial increase in volatility around the Great Recession in the DER, which is even more pronounced than the ASEC with zeros excluded, but it returned to pre-recession levels by 2014. Thus, volatility over the last two decades is largely a business-cycle effect, and there is no substantive discrepancy between the ASEC and DER, unless one incorporates ASEC zeros into the analysis. Or, put more strongly, when only ASEC workers with non-imputed earnings are used, there is no difference in volatility trends in the survey and tax data. As discussed below, however, most of these zeros in the linked respondent ASEC sample report zero earnings in the ASEC but have a positive DER, suggesting these "0s" are reporting errors rather than periods of non-work. Appendix Figure 2 repeats the analysis of Figure 3, but with zeros and a 5% trim of outliers instead of 1% in panel A, and without zeros and no trim in panel B. The story is the same, but the level of volatility is notably lower in panel A, and higher in panel B when the trim is removed. In neither case is there a substantive effect on trend volatility.

A. Linking and Nonresponse

In this subsection we explore how the level and trend in volatility are affected sequentially by the requirement that men be linked to the DER across two years of the ASEC and the requirement that men respond to the ASEC earnings questions in both years.

[Figure 4 here]

In Figure 4 we present the arc percent change volatility series for the full ASEC in panel A, including those linked and not linked to the DER and those who both respond and do not respond to the earnings questions in the ASEC. For nonrespondents, the ASEC earnings are those imputed by the Census Bureau. The ASEC and DER samples in panel A are not the same, because the ASEC lines include both linked AND unlinked DER individuals, and the DER lines include those individuals who may have been linked in one or both years. In panel B we impose the requirement that sample members be linked to the DER both years, but still include earnings nonrespondents, while in panel C we restrict the sample to two-year respondents regardless of whether they have a DER link (the ASEC and DER samples are therefore again not the same) . For each panel we present the volatility series inclusive of zero earnings on the left side, and exclude zero earnings on the right side. We only present the 1% trim series since only the levels and not trends are affected with the 5% trim.

The left side of panel A of Figure 4 has the same shape as panel A of the linked responders in Figure 3, the difference being the level of ASEC volatility is much higher in Figure 4 when including both the unlinked and nonresponders, and the trend increase sharper. Interestingly, though, the right side of Figure 4 panel A without zero earnings is much different than panel B of linked responders in Figure 3. The former shows a wide gap between the ASEC and DER volatility, with volatility higher and rising in the ASEC, whereas Figure 3 depicted no trend in volatility in either the DER or the ASEC and if anything, volatility was higher in the DER during recessionary periods. When we impose the requirement that the individual have a DER link in both periods in panel B of Figure 4 this dampens the amplitude of volatility during the Great Recession in panel A with zero earnings included, but does not remove the trend increase in ASEC volatility with or without (right side of panel B) zeros included. Panel C, which requires the man to be an ASEC respondent in both periods, but not necessarily a DER link, shows that compared to panel A, excluding nonresponders reduces measured volatility considerably. This is particularly true for the right side with zero earnings excluded. That chart is nearly identical to panel B of Figure 3, suggesting that responders and nonresponders have similar earnings volatility in administrative data, but that the Census hot-deck method substantially amplifies that procedure in the ASEC and imparts bias much like Hirsch and Schumacher (2004) and Bollinger et al. (2019) show for wage levels.[8] Notably, in the left side of panel C, the ASEC zero earnings are much more likely to be "true nonworkers" and not those who report no earnings in the ASEC but have a DER report. The substantial run-up of volatility during the Great Recession, followed by the run-down during the recovery, captures the important movements in and out of employment. Thus the baseline volatility for the ASEC in Figure 3 is affected both by dropping nonlinked and nonresponders, and the covariance between the two groups.

[8] Hirsch and Schumacher (2004) show that there is bias in earnings regressions on coefficients not used in the Census earnings imputation process. Specifically, they show how estimates of the wage gain from union status, which is not a variable for the Census hot deck, are biased. Likewise, being matched across sample periods is not a variable used in the hot deck, thus leading to bias in volatility estimates using ASEC nonrespondents. Dahl et al. (2011) report a similar upward bias from imputation in the SIPP linked to the DER.

B. Attrition

A possible concern with matched ASEC is with sample attrition affecting our earnings series. The CPS sample domain is household addresses and not individuals, so that if a person moves between ASEC surveys then the Census Bureau interviews the new occupant at the address and does not follow the original respondent. This is why we use state of residence as one of the match criteria because if the state of residence changes for the household identifier then that signals an incorrect match across ASEC surveys. Moves are more likely among low-income families whose earnings are more volatile, which means we could understate the level and trends in volatility with our sample. Under the assumption that the probability of attrition is unobserved and time invariant (i.e., a fixed effect), or trending very slowly over time, then first differencing earnings as used in the volatility measures based on log-differences will remove the latent probability of attrition and our estimates will be purged of possible attrition bias (Wooldridge 2001). However, if there is time-variation in the factor loading on the unobserved individual-level heterogeneity then differencing will not eliminate potential attrition bias unless the factor loading is randomly distributed across the population. A conservative interpretation is that data from matched ASEC provides estimates of earnings volatility among the population of non-movers.

To examine the potential role of attrition on volatility, we expand our dataset to include not only those matched across years in the ASEC, but also those individuals observed in year 1 of the ASEC but not year 2. Appendix Table 2 reports the year 1 socioeconomic characteristics of attriters and non-attriters. Attriters are younger, more likely to be a member of a minority racial group, have fewer years of school, less likely to be married (though with a higher percentage of married but with spouse absent), work fewer weeks and hours per week, have lower earnings in both the ASEC and DER, and higher rates of earnings (item) nonresponse. These patterns hold for both men and women, and suggest that volatility is likely to differ between attriters and non-attriters.

[Figure 5 here]

Figure 5 presents trends in arc percent volatility of men by attrition status using a 1% trim at the top and bottom of the earnings distribution. In panel A we use the full sample of respondents and nonrespondents, while panel B is restricted to those who only responded to the ASEC earnings questions in year 1. Panel A has four series, two for attriters and two for non-attriters. For both groups we present DER volatility, while for attriters we also present a series that consists of the ASEC in year 1 and the DER in year 2 (for those linked to the DER). For non-attriters we present the ASEC series across both years. Focusing on the DER series, it is clear that attriters have higher earnings volatility than non-attriters, though the trends are largely similar (with a more Great Recession effect among attriters). This is also true for the subsample of year 1 earnings responders as depicted in panel B based on the DER alone. Interestingly, when we use the DER to "fill-in" for the missing year 2 ASEC among attriters in Panel A, volatility is noticeably higher in most years compared to the ASEC alone among non-attriters. Again, though, the trends are not substantively different. This suggests that our matched panels in the ASEC primarily affect the level, but not trend, in volatility.

[Figure 6 here]

To further explore the role of attrition and nonresponse on volatility, in Figure 6 we reproduce panel A of Figure 3 where we now reweight the data using inverse probability weighting (IPW). IPW is a general solution to attrition and nonresponse when the data are missing at random (Wooldridge 2007). Although there is evidence that the MAR assumption is violated in earnings levels (Bollinger et al. 2019), it may be a valid assumption with our volatility measures. We proceed by estimating a probit model of the probability that the person is (i) not a whole impute, (ii) is linked to the DER, (iii) is an earnings respondent, and (iv) is matched across ASEC waves as a function of a rich set of socioeconomic characteristics in both levels and interactions. We then divide the ASEC supplement weight by the fitted probability of response + link + match and estimate the IPW volatility series.[9] Figure 6 aligns with our priors from the evidence in Figure 5; namely, when we reweight the data, the original estimates of volatility are higher in each year, but with no effect on the trends.

[9] Specifically, in each time period *t* we estimate via probit $y_{it} = f(x_{it}\beta_t) + u_{it}$, where $y = 1$ if the person (i) is not a whole impute, (ii) is linked to the DER, (iii) is an earnings respondent, and (iv) is matched across ASEC waves. Let $\widehat{\Phi}_{it}$ be the fitted cdf from the probit, then the new weight is defined as $\widehat{w}_{it}^{ipw} = \frac{w_{it}^{ASEC}}{\widehat{\Phi}_{it}}$.

C. Comparison to Common Measures and Samples in the Literature

The most frequently utilized measure of volatility in the literature has been the variance of log earnings growth in equation (2). Figure 7 presents trends in volatility estimated with the change in log earnings for the sample of linked respondents with positive ASEC or DER earnings in each period. Comparing Figure 7 to panel B of Figure 3 shows that there is little difference in the patterns of volatility among those with positive earnings. The log difference increases the amplitude of volatility in the DER during the Great Recession compared to the arc percent change of earnings levels, and suggests a temporary greater separation with the ASEC, but no change in long-term trends.

[Figures 7 and 8 here]

Figure 8 returns to the arc percent volatility measure of equation (1) but restricts the sample to a variety of cuts using male heads of household only—the group observed in the PSID and most employed in the literature. The first row replicates panels A and B in Figure 3 with and without zeros included, and the only difference is a slight reduction in volatility levels but no change in trends when restricting to heads. The second row first drops those workers with any self-employment earnings, and then drops public-sector workers. The third row drops, in turn, those workers with real earnings levels below a quarter of a full-time full-year work at half the federal minimum wage, those workers with earnings below a fixed (in real terms) dollar value of $3,685, and those workers with earnings below the real value of the minimum Social Security earnings thresholds needed to qualify for retirement benefits credit, respectively. The latter three cuts have been used in various prior studies using administrative earnings records from Social Security (Kopczuk et al. 2010; Sabelhaus and Song 2010; Bloom et al. 2018). As the panels show, dropping self- employed or public-sector workers results in greater coincidence between survey and administrative data, while the alternative bottom cutpoints have little effect on volatility levels or trends, suggesting that there is nothing particularly restrictive about the PSID sample of heads of households in volatility analyses.

[Figure 9 here]

One concern in comparing our estimates with those studies using differing data sources may be differences in the underlying distribution of earnings. These differences may arise due to sampling frame differences or measurement approaches. In order to investigate this, we adopted a uniform weighting approach across the four papers in this project, matching the distribution of the PSID sample. Four different weighting approaches were examined. The first weighting approach takes the distribution of real (2010) earnings in each year of the PSID sample of males aged 25 to 59 working full time. The PSID data were broken into ventiles after trimming at the bottom and top 1%. We construct weights by measuring the proportion of each year's sample, p_v, which falls into the PSID ventile v. Then the weight for any observation i is given by $w_i = \sum_{v=1}^{20} 1[L_v \leq E_i < U_v] \frac{.05}{p_v}$, where E_i is the earnings (in that particular year) for observation i, $1[.]$ is the indicator function, and L_v, U_v are the lower and upper bounds of the year-specific earnings ventile. The second weighting scheme fixes the PSID distribution to the earnings distribution in the year 2000. The third and fourth approaches first regress log earnings on age and age squared. Residuals are then formed by $r_i = E_i - \eta e^{\widehat{lnE_i}}$, where $\widehat{lnE_i}$ is the predicted value from the log earnings regression. The final measure, $\tilde{r}_i = r_i - \bar{r}$, is re-centered on zero by subtracting the mean for that year. Again, ventiles for the residuals were based upon the PSID sample (similarly treated) and weights were constructed similarly.

[Figure 9 here]

The exercise resulted in six measures of the trend in earnings volatility, each of which removes imputations and zeros from the sample, as depicted in Figure 9. Permutation 1 is the same series as that in Figure 3B, except that instead of using residuals to construct the variance, we use earnings levels. Permutation 2 uses the year-by-year weights, while Permutation 3 uses the fixed year 2000 weights. Permutation 4 does not use any weights and is comparable to our Figure 3B, in that it first regresses the arc percent change in earnings on age and age squared, and constructs the variance series from the residuals. Permutation 5 returns to the level arc percent change variance, but uses the residual weights year-by-year. Finally, Permutation 6 again uses the level arc percent change variance but with the fixed year 2000 residual weights.

As can be seen in the left panel of Figure 9, these various permutations have no effect on the ASEC measures of volatility. The DER series, however, do show some differences. Most notably, we find that the weighting by the PSID data leads to a lower level, although similar trends, especially for the early years of the series. This could result from the longer left tail of earnings observed in the DER compared to the PSID. However, these differences are minor compared to the ones highlighted in our exercises above, and differences between the PSID and the ASEC or DER do not appear to be largely driven by the underlying distribution.

D. Heterogeneity in Volatility

We next explore whether the changes in volatility in the ASEC and DER over the past 21 years are widespread or isolated among specific demographic groups. We focus on heterogeneity across education attainment, race and ethnicity, and intensity of work effort as full-year or part-year.[10] This is an advantage of the ASEC-DER file because of the large sample sizes. For this exercise we focus on the arc percent change measure and within the linked respondent sample to make direct comparisons between the ASEC and DER. To assess the relative contributions of subgroups to total variance, we decompose the variance of volatility into within-group and between-group differences (Moffitt and Zhang 2020). Specifically, if we define $arc_{it} \equiv \frac{y_{it}-y_{it-1}}{\bar{y}_i}$, then for any time period t we can rewrite the variance of the arc percent change in equation (1) as

(4) $$varc_t = \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N_g}\left[\left(arc_i - \overline{arc}_g\right) + \left(\overline{arc}_g - \overline{arc}\right)\right]^2 = \sum_{g=1}^{G}\frac{N_g}{N}\left[\frac{1}{N_g}\sum_{i=1}^{N_g}(arc_i - \overline{arc}_g)^2\right] + \sum_{g=1}^{G}\frac{N_g}{N}\left(\overline{arc}_g - \overline{arc}\right)^2,$$

[10] Full year includes full-time full-year and part-time full-year workers, while part year includes full-time part-year and part-time part-year workers.

where $\overline{arc}_g$ is the group (g) mean, $\overline{arc}$ is the overall mean, and $\frac{N_g}{N}$ is the group-specific share of the total population. The first term in the second equality is the within-variance of the arc percent change at time t and the last term is the between-group variance. Because the between-group variance accounts for less than 10 percent of the total in any given year, we focus our discussion on the within-group variance. We only show the decomposition with a 1% trim, inclusive of zeros from (one-year) ASEC.

[Figure 10 here]

Figure 10 depicts trends in volatility separately in the ASEC and DER for linked-respondent men with less than high school education, high school, some college, and college or post-graduate work. The top panel contains the group-specific variances, while the bottom panel contains the within-variances weighted by the shares of the population. The trends in population shares are depicted in Appendix Figure 3. Panel A shows that volatility levels in the ASEC are much higher among high school dropouts than other education groups, and that the rise in ASEC volatility cuts across all education levels, except for those with a college degree or more. The bottom panel, however, shows that high-school dropouts contribute least to overall male earnings volatility, which is explained by their small and declining share of the population as depicted in panel A of Appendix Figure 3. The DER shows little difference in volatility levels or trends for those men with a high school diploma or more. There is an increase in DER volatility among dropouts around the Great Recession, but as with the ASEC, their weighted share of total variance is small and stable.

[Figures 11-12 here]

The panels of Figure 11 show trends by race and ethnicity for men in 4 groups: White Non-Hispanic, Black Non-Hispanic, Hispanic, and a combined Asian and American Indian group. Both whites and blacks experience an increase in ASEC volatility over the time period; there are no differences across race-ethnicity in DER volatility. However, once we weight by population shares, it becomes clear that White Non-Hispanic men contribute the most to volatility and their series mimics that depicted in Figure 3 for both the ASEC and DER. In Figure 12 we split the sample into whether the person is part-year in both years, part-year in year one and full-year in year two, full-year in year one and part-year in year two, and full-year in both years. As expected, men who are full-year both years have the lowest unweighted volatility, and those who are part-year in both periods have the highest unweighted volatility. Perhaps surprising, even though the two-year part-year workers are a small share of the total workforce, their weighted variance is the largest overall in the ASEC and the shape largely reflects that depicted in Figure 3.[11] Volatility in the last few years of the sample period does not fall as much as the weighted variance of two-year part-time workers, with overall volatility propped up by employment status switchers and the stability of the full-year group.

[Figure 13 here]

In Figure 13 we return to the full ASEC sample that includes both respondents and nonrespondents, and those with and without a link to the DER. Here we wish to explore how much of total volatility the standard user of the ASEC is likely to identify because of inclusion of earnings nonresponders. Panel A shows that those men who switch response status between years have the highest group-specific volatility, followed next by those who are ASEC nonresponders in both years. Two-year responders have the lowest volatility, but panel B shows they contribute the most to total volatility, which is consistent with their large share of the population. However, as panel D of Appendix Figure 3 shows, the latter groups share has declined in the last few years, which means total volatility did not fall as much as predicted by two-year respondents and was propped up by volatility of allocated earners. Crucially, while DER volatility among nonresponders exceeds that of responders in panel A, the differences are very small, especially in relation to the ASEC, underscoring pitfalls with the hot-deck imputation in the ASEC.

[11] Note that the ASEC w/ 0s measure is defined for this sample because there are some workers who report 0 earnings in the ASEC, but they have positive earnings in the DER. These persons fall mostly in the part-part group. In general the patterns described by the variance decompositions in the text agree with log-difference decompositions. The only difference is the weighted employment-status is now dominated by concurrent full-year workers, underscoring that the part-part group has many ASEC zeros but DER reports. Results available upon request.

E. Comparisons with Women

We conduct a full parallel set of analyses for women, but for ease of presentation we only present a subset here. To anchor with the baseline male volatility estimates in Figure 3, in Figure 14 we use the linked respondent sample to depict the arc percent change with one-year ASEC zero earnings included in panel A and with zero earnings excluded in panel B. Unlike men, the volatility of women is stable in the ASEC when zeros are included, while in the DER (and when zeros are excluded in the ASEC in panel B), there is evidence of declining volatility of women. The latter continues a trend first begun in the late 1970s (Ziliak et al 2011), and contrasts with the stability of male volatility in panel B of Figure 3. The other important contrast with men is the lack of business-cycle induced volatility of women's earnings. Appendix Figure 4 presents female earnings volatility using the log difference and confirms that declining female earnings volatility is robust across measures. This holds in Appendix Figure 5 with a 5% trim of outliers, in Appendix Figure 6 with no trimming of outliers, and in Appendix Figure 7 Panel B for both attriters and non-attriters.

[Figure 14 here]

In Figure 15 we unpack the volatility of women to examine the roles of imposing a DER link and earnings response, akin to that presented in Figure 4 for men. Panel A is for the full ASEC, those with and without a DER link and those with and without allocated earnings, and with a 1% trim of annual cross-sectional earnings. When we include women with no more than one-year ASEC nonemployment, ASEC volatility is at least three times the level that in the DER, and stable, and when we exclude zero earners in the right-hand side of panel A, ASEC volatility falls by half, but is still 50 percent higher than DER volatility. When we impose the two-year DER link requirement in panel B—but still include earnings responders and nonresponders—then ASEC volatility of women falls considerably (in part because many zeros fall out), but it is trending upward. In panel C we impose the requirement that earnings be reported in both years, with and without a DER link. While ASEC volatility is lower than in panel A, suggestive that nonresponse pushes volatility upward as we found with men, there is a yawning gap between the ASEC and DER when we include ASEC zeros, but they are coincident when we drop zeros in the right side of panel C. The fact that the baseline female ASEC volatility series in panel A of Figure 14 lies below the series in Figure 15 suggests important covariance in the joint imposition of having both a DER link and being a respondent.

[Figure 15]

We explore further the issue of nonresponse on women's earnings volatility in the variance decomposition based on equation (4) and presented in Figure 16. This parallels the decomposition for men in Figure 13. What is interesting here is that the group-specific volatility of two-year nonrespondents in panel A overlaps with the volatility of two-year respondents. For men the former was between 50-100 percent higher than the latter in most years, suggesting perhaps that the hot-deck may be more accurate in predicting missing earnings of women than men, or at least changes in earnings. However, as with men, once the variances are weighted by group-specific shares (panel B of Figure 16) then the volatility of two-year responders contributes most to annual volatility among women in both the ASEC and the DER, and is declining over time in both data sources.

[Figure 16 here]

Finally, in Appendix Figures 8-10 we present within-group variance decompositions (arc percent with zeros included) for women's earnings volatility by education attainment, race and ethnicity, and full-year/part-year employment status. In this case the decompositions for women are more similar to men in that the group-specific volatility of high school dropouts exceeds that of higher educated, but the weighted contribution of the less skilled is dominated by the other groups, the gross volatility of whites is less than non-whites, but the weighted variance of whites is greatest owing to their large population share, and the volatility of part-year workers dominates that of other groups, both unweighted and weighted by group shares.

F. Transitory and Permanent Volatility

Much of the early literature aimed to isolate whether earnings volatility stemmed from permanent or transitory changes in economic status. The classic model is a simple decomposition of earnings into a time-invariant permanent component, μ_i, and a time-varying transitory component, u_{it},

(5) $$y_{it} = \mu_i + u_{it}.$$

Identifying the transitory component is easily addressed in equation (5) using the arc measure or by modeling log earnings similarly (additive transitory and permanent components) because differencing removes the time-invariant permanent effect.

We note, however, that survey, y_{it}^S, and administrative, y_{it}^A, data differ in the data generating process of observed earnings. Researchers often treat administrative data as the "gold standard" or "truth", though Kapteyn and Ypma (2007) suggest that treating administrative data as error free may be unreasonable. The largest threat to treating administrative data as error free is the existence of an underground economy where workers' earnings are not reported formally. At its simplest, then, one can argue that measurement error potentially enters both the survey data and administrative data:

(6) $$y_{it}^S = \mu_i + u_{it} + \varepsilon_{it}$$

$$y_{it}^A = \mu_i + u_{it} + \upsilon_{it}.$$

Our key assumption is that while both measures may have measurement error, because the sources and causes of the measurement error differ, we assume that ε_{it} and υ_{it} are at least uncorrelated (independence seems appropriate, but is not necessary). We also assume—as is typically done—that the transitory shocks are also uncorrelated over time and uncorrelated with the two measurement error terms.

Under these two assumptions, then we can discuss various approaches to identification of the variance of the permanent component and the variance of the transitory component through autocorrelation and cross correlation terms:

(7) $$Cov\left(y_{it}^j, y_{it-1}^k\right) = V(\mu_i),$$

where *j* and *k* represent either survey (*S*) or administrative (*A*) measures. This produces four possible covariance terms which identify the permanent earnings variance under the simple additive measurement error terms above: $Cov(y_{it}^S, y_{it-1}^S)$, $Cov(y_{it}^A, y_{it-1}^A)$, $Cov(y_{it}^S, y_{it-1}^A)$, $Cov(y_{it}^A, y_{it-1}^S)$.

In order to identify the transitory term, we have to remove the permanent term. Typically, this is done with first differencing as

(8) $\Delta y_{it}^{S} = (u_{it} - u_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1})$

and

(9) $\Delta y_{it}^{A} = (u_{it} - u_{it-1}) + (\upsilon_{it} - \upsilon_{it-1})$.

The corresponding variances are then

(10) $V(\Delta y_{it}^{S}) = 2V(u_{it}) + 2V(\varepsilon_{it})$

and

(11) $V(\Delta y_{it}^{A}) = 2V(u_{it}) + 2V(\upsilon_{it})$.

However, as long as the measurement error terms ε_{it} and υ_{it} are uncorrelated, then the covariance of the two series identifies the transitory variance:

(12) $Cov(\Delta y_{it}^{S}, \Delta y_{it}^{A}) = 2V(u_{it})$.

This term also provides a test of the presence of measurement error. For example, if measurement error is present in the ASEC, but not in the DER data, then $V(\Delta y_{it}^{A})=Cov(\Delta y_{it}^{S}, \Delta y_{it}^{A})$, and both will be less than the $V(\Delta y_{it}^{S})$. If there is no measurement error, then $V(\Delta y_{it}^{S}) = V(\Delta y_{it}^{A}) = Cov(\Delta y_{it}^{S}, \Delta y_{it}^{A})$.

[Figure 17 here]

In Figure 17 we present the three sets of estimates of the transitory volatility based upon the two variances and the covariance term in equations (10) – (12), where Panel A is for men and Panel B is for women. The variances are labeled ASEC for $V(\Delta y_{it}^{S})$, DER for $V(\Delta y_{it}^{A})$, while the covariance term, $Cov(\Delta y_{it}^{S}, \Delta y_{it}^{A})$, is labelled Cov(ASEC,DER). Note, in all cases the estimated variances and covariances were divided by two to provide estimates of the underlying variance of the transitory income term.

As expected, we find that the covariance term is the lowest of the three, and thus provides evidence that there is measurement error in both the administrative and the survey data. The measurement error in the ASEC is clearly higher than in the DER, as one might expect. We note too, in comparing Panels A and B that measurement error appears largest for men both for the administrative and the survey earnings measures. The covariance-based estimate of transitory earnings variance is quite stable over the time period for both men and women. So too is the estimate of transitory earnings variance based on the administrative records. However, for men in particular, the estimates of transitory earnings variance from the ASEC records have fallen over the period, most notably around the Great Recession. However, it appears that all of this change is likely due to measurement error, rather than some significant change in the economy given the stability in the DER volatility.

[Figure 18 here]

In Figure 18 we present the four autocovariance measures of the permanent earnings volatility, with men in Panel A and women in Panel B. Here the four series are labeled Cov(ASEC(t), ASEC(t-1)), Cov(DER(t), DER(t-1)), Cov(ASEC(t), DER(t-1)), and Cov(DER(t), ASEC(t-1)). In all four cases for both men and women there is a distinct upward trend in permanent earnings variance. This would be consistent with an economy wide change increasing earnings variance. The upward trend is most pronounced for men, and most pronounced in the Cov(DER(t), DER(t-1)) series based only on the autocorrelation of the DER series for both men and women.

One explanation for why the Cov(DER(t), DER(t-1)) series differs from that of the ASEC series, is that the ASEC measures of earnings have measurement error which does not meet the assumptions of the simple model. A number of other authors have posited that

(13) $$y_{it}^{S} = \rho y_{it} + \epsilon_{it} = \rho \mu_i + \rho u_{it} + \varepsilon_{it}.$$

If this relationship holds, with $\rho<1$, then we would expect nearly the pattern we see. Simple exploration of this suggests an estimate of ρ around .9 or higher (nearly .99 for women). Note that this would produce all the patterns we have seen. It suggests that the autocovariance series using only the DER is correct (or closest), while the Cov(ASEC,DER) series in Figure 17 is a slight understatement of the transitory earnings variance.

V. Conclusion

The paper presents new estimates of earnings volatility of men and women using unique restricted-access survey and administrative tax data for the tax years spanning 1995-2015. Our results begin to reconcile differences between administrative estimates of earnings volatility and those found using survey data. Throughout the analysis, the administrative sample was linked to the survey sample, eliminating potential differences due to overall survey response or other sampling frame issues. As we varied the samples based on survey responses, including or excluding ASEC reports of zero earnings or imputed earnings, the lack of significant trend in the earnings volatility from the administrative data was unchanged. Thus, differences between survey and administrative data appear to be dominated by measurement and earnings item response issues.

Our results for men suggest that when ASEC zeros were omitted, male earnings volatility was stable over the period in both the ASEC and DER, though with a substantial business-cycle component. However, when ASEC zeros were included, earnings volatility as measured by the arc percent change increased in survey data, but was stable in tax data. These "nonworkers" in the linked respondent sample are primarily those who report zero earnings in the survey, but have a positive administrative earnings report, and thus reflect measurement error in the ASEC. When we focus on two-year ASEC respondents, but do not require a link to the DER, we capture more true nonworkers in the ASEC, but the main effect of this is to magnify the amplitude over the Great Recession with no remarkable change to the trend. We find no evidence of declining male earnings volatility in either survey or administrative data over the past two decades. For women, excluding zeros results in parity across the ASEC and DER with declining earnings volatility among women, but when zero earnings are included, survey data show stable volatility, while tax data indicate a secular decline. Importantly, for women there is no cyclical component to volatility. The variance decompositions for both men and women indicate that nonresponders, low-educated, racial minorities, and part-year workers have the greatest group specific earnings volatility, but with the exception of part-year workers, they contribute least to the level and trend of volatility owing to their small shares of the population. The finding that nonresponse in the ASEC has an upward bias in estimates of volatility lends additional evidence to that in Hokayem et al. (2015) and Bollinger et al. (2019) on the perils of use of the Census hot-deck imputations for earnings analyses. This finding seems more binding for male earnings volatility, but for both men and women, we recommend that users of the ASEC for earnings and income research drop earnings nonresponders, both whole supplement imputes as well as item nonresponders. Finally, in a decomposition of volatility into transitory and permanent components we find that measurement error is present in both administrative and survey data, though more pronounced in survey data. For both men and women there is evidence of stable transitory volatility, but rising permanent volatility over the past two decades.

We subjected our earnings volatility estimates to a number of additional robustness checks, both in comparison to the DER and to samples and modeling choices employed in the prior literature. Our results showed that levels and trends in volatility do not differ whether one uses the arc percent or log difference measure, once one drops zero earnings, but the arc percent measure is preferred to capture movements in and out of employment. In addition, we found that attrition from the ASEC only affects the level but not the trend in volatility. We also found that restricting to heads of household like in PSID studies had no discernable effect on the level and trends in volatility. Moreover, even though the distribution of earnings differs between the DER and ASEC, especially in the left tail, and potentially between the ASEC and PSID, we found no evidence that distribution differences impact volatility estimates. Our takeaway is that survey data and administrative data yield similar estimates of volatility of male and female earnings once one holds the sample frame and measurement constant.

References

Abowd, John M. and Martha H. Stinson. 2013. "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data." Review of Economics and Statistics, 95 (December): 1451-1467.

Bloom, Nicholas, Fatih Guvenen, Luigi Pistaferri, John Sabelhaus, Sergio Salgado, and Jae Song. 2018. "The Great Micro Moderation," Unpublished manuscript.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption Inequality and Partial Insurance," *American Economic Review*, 98(5): 1887-1921.

Bollinger, Christopher R. 1998. "Measurement error in the current population survey: a nonparametric look." *Journal of Labor Economics* 16(3): 576-94.

Bollinger, Christopher R., and Amitabh Chandra. 2005. "Iatrogenic Specification Error: A Cautionary Tale of Cleaning Data," *Journal of Labor Economics* 23(2): 235-257.

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

Bollinger, Christopher R., Barry Hirsch, Charles Hokayem, and James P. Ziliak. 2019. "Trouble in the Tails? What We Know About Earnings Nonresponse Thirty Years After Lillard, Smith, and Welch." *Journal of Political Economy*, 129(5):2143-2185

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data," in *Handbook of Econometrics*, Vol. 5, edited by E. E. Leamer and J. J. Heckman, Amsterdam: Elsevier: 3705-3843.

Cameron, Stephen, and Joseph Tracy. 1998. "Earnings Variability in the United States: An Examination Using Matched-CPS Data," Mimeo, Federal Reserve Bank of New York.

Carr, Michael, and Emily Wiemers. 2018. "New Evidence on Earnings Volatility in Survey and Administrative Data," *American Economic Review Papers & Proceedings*, 108:287-291.

Carr, Michael, Robert Moffitt, and Emily Wiemers. 2020. "Reconciling Rrends in Volatility: Evidence from the SIPP Survey and Administrative Data." Unpublished manuscript.

Celik, Sule, Chinhui Juhn, Kristin McCue, and Jesse Thompson. 2012. "Recent Trends in Earnings Volatility: Evidence from Survey and Administrative Data," *The B.E. Journal of Economic Analysis and Policy* 12(2): 1-26.

Dahl, Molly, Thomas DeLeire, and Jonathan Schwabish. 2011. "Estimates of Year-to-Year Variability in Worker Earnings and in Household Incomes from Administrative, Survey, and Matched Data." *Journal of Human Resources* 46(4): 750-774.

Dynan, Karen E., Douglas W. Elmendorf, and Daniel E. Sichel. 2012. "The Evolution of Household Income Volatility." *The B.E. Journal of Economic Analysis & Policy: Advances,* Volume 12, Issue 2: Article 3.

Gittleman, Maury, and Mary Joyce. 1996. "Earnings Mobility and Long-Run Inequality: An Analysis Using Matched CPS Data." *Industrial Relations* 35(2): 180-196.

Gottschalk, Peter and Robert Moffitt. 1994. "The Growth of Earnings Instability in the U.S. Labor Market." *Brookings Papers on Economic Activity* 1, 217–254.

Gottschalk, Peter, and Robert Moffitt. 2009. "The Rising Instability of U.S. Earnings." *Journal of Economic Perspectives* 23(4): 3-24.

Hacker, Jacob S., and Elisabeth Jacobs. 2008. "The Rising Instability of American Family Incomes, 1969-2004: Evidence from the Panel Study of Income Dynamics." EPI Briefing Paper 213, Economic Policy Institute

Haider, Steven. 2001. "Earnings Instability and Earnings Inequality of Males in the United States: 1967–1991." *Journal of Labor Economics* 19(4): 799–836.

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

Hokayem, Charles, Christopher R. Bollinger, and James P. Ziliak. 2015. "The Role of CPS Nonresponse in the Measurement of Poverty." *Journal of the American Statistical Association* 110(511): 935-45.

Kapteyn, Arie and Jelmer Y. Ypma. 2007. "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data." *Journal of Labor Economics* 25(3): 513-551.

Keys, Ben. 2008. "Trends in Income and Consumption Volatility, 1970–2000." In *Income Volatility and Food Assistance in the United States*, D. Jolliffe and J. P. Ziliak, eds., Kalamazoo, MI: W.E. Upjohn Institute.

Kniesner, Thomas J., and James P. Ziliak. 2002. "Tax Reform and Automatic Stabilization," *American Economic Review* 92(3): 590–612.

Kopczuk, Wojiech, Emmanuel Saez, and Jae Song. 2010. "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937." *Quarterly Journal of Economics* 125(1): 91-128.

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

Madrian, Brigitte, and Lars Lefgren. 1999. "A Note on Longitudinally Matching Current Population Survey (CPS) Respondents." NBER Working Paper 247.

McKinney, Kevin L, and John M. Abowd. 2020. "Male Earnings Volatility in the LEHD Before, During, and After the Great Recession." Unpublished manuscript.

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

Moffitt, Robert, and Sisi Zhang. 2018. "Income Volatility and the PSID: Past Research and New Results," *American Economic Review Papers & Proceedings* 108: 277-280.

Moffitt, Robert, and Sisi Zhang. 2020. "Estimating Trends in Male Earnings Volatility with the Panel Study of Income Dynamics." Unpublished manuscript.

Roemer, Mark. 2002. "Using Administrative Earnings Records to Assess Wage Data Quality in the Current Population Survey and the Survey of Income and Program Participation." Longitudinal Employer-Household Dynamics Program Technical Paper No. TP-2002-22, US Census Bureau.

Sabelhaus, John, and Jae Song. 2010. "The Great Moderation in Micro Labor Earnings." *Journal of Monetary Economics* 57(4): 391-403.

Shin, Donggyun, and Gary Solon. 2011. "Trends in Men's Earnings Volatility: What Does the Panel Study of Income Dynamics Show?" *Journal of Public Economics* 95(7): 973-982.

Stock, James, and Mark Watson. 2003. "Has the Business Cycle Changed and Why?," In M. Gertler and K. Rogoff (eds.) *NBER Macroeconomics Annual*, Volume 17, Cambridge, MA: MIT Press, 159-218.

Wooldridge, Jeffrey. 2001. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

Wooldridge, Jeffrey M. 2007. "Inverse Probability Weighted M-Estimation for General Missing Data Problems." *Journal of Econometrics* 141: 1281-1301.

Ziliak, James P., Bradley Hardy, and Christopher Bollinger. 2011. "Earnings Volatility in America: Evidence from Matched CPS." *Labour Economics* 18(6): 742-754.

Table 1. Sample Summary Statistics

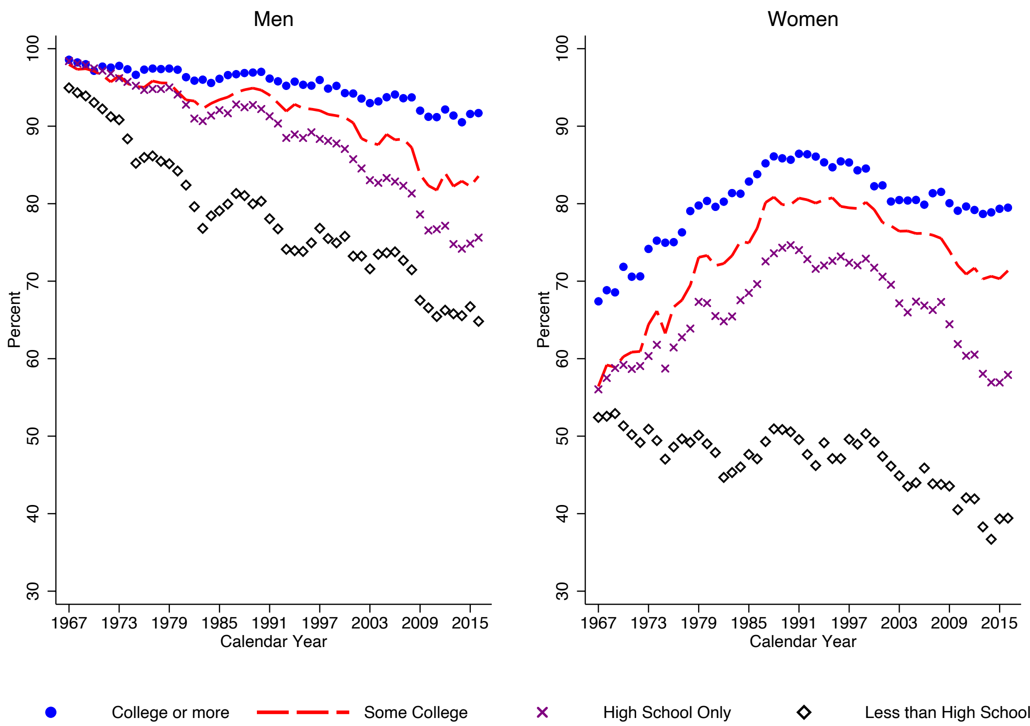
	A. Men			
	Full Sample		Linked Respondent Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Age	42.78	9.4	42.81	9.35
White	0.72	0.45	0.77	0.42
Black	0.08	0.27	0.07	0.26
Asian or American Indian	0.06	0.24	0.06	0.23
Hispanic	0.14	0.34	0.10	0.30
Years Education	13.87	2.8	14.17	2.66
Married, Spouse Present	0.64	0.48	0.67	0.47
Married, Spouse Absent	0.14	0.34	0.13	0.33
Never Married	0.22	0.42	0.20	0.40
Native	0.83	0.37	0.88	0.33
Foreign Citizen	0.07	0.26	0.07	0.25
Foreign Non-Citizen	0.09	0.29	0.06	0.23
Weeks Worked	47.65	11.93	48.97	9.60
Hours per Week	41.89	12.3	42.96	10.79
Nonrespond Yr1, Yr2	0.10	0.29		
Nonrespond Yr1, Respond Yr2	0.10	0.3		
Respond Yr1, Nonrespond Yr2	0.13	0.33		
Respond Yr1, Yr2	0.68	0.47	1.00	0.00
DER Non-Link Yr1, Yr2	0.10	0.3		
DER Non-link Yr1, Link Yr2	0.03	0.17		
DER Link Yr1, Non-link Yr2	0.02	0.15		
DER Link Yr1, Yr2	0.84	0.36	1.00	0.00
Proxy Response	0.50	0.5	0.47	0.50
Real ASEC Earnings ($2010 thou.)	59.00	74.64	64.35	75.93
Real DER Earnings ($2010 thou.)	66.29	129.00	67.42	134.00
Person-years (rounded)	381,000		222,000	
	B. Women			
Age	43.17	9.39	43.14	9.45
White	0.72	0.45	0.74	0.44
Black	0.11	0.31	0.10	0.30
Asian or American Indian	0.06	0.24	0.06	0.24
Hispanic	0.12	0.32	0.10	0.31
Years Education	14.27	2.62	14.48	2.57
Married, Spouse Present	0.62	0.49	0.63	0.48
Married, Spouse Absent	0.19	0.39	0.19	0.39
Never Married	0.19	0.39	0.19	0.39
Native	0.85	0.35	0.88	0.33
Foreign Citizen	0.08	0.27	0.07	0.26
Foreign Non-Citizen	0.07	0.25	0.05	0.22
Weeks Worked	45.62	14.33	47.56	13.89
Hours per Week	36.17	13.26	37.60	12.99
Nonrespond Yr1, Yr2	0.09	0.28		
Nonrespond Yr1, Respond Yr2	0.09	0.29		
Respond Yr1, Nonrespond Yr2	0.11	0.32		

Respond Yr1, Yr2	0.71	0.46	1.00	0.00
DER Non-Link Yr1, Yr2	0.08	0.28		
DER Non-link Yr1, Link Yr2	0.04	0.19		
DER Link Yr1, Non-link Yr2	0.03	0.18		
DER Link Yr1, Yr2	0.85	0.36	1.00	0.00
Proxy Response	0.41	0.49	0.38	0.49
Real ASEC Earnings ($2010 thou.)	38.47	47.20	41.81	46.72
Real DER Earnings ($2010 thou.)	41.63	94.60	41.74	58.41
Person-years (rounded)	367,000		213,000	

Note: The full sample consists of men and women ages 25-59 who work in at least one of two years, are not full-time students in either year, and do not have their entire ASEC supplement allocated (no whole imputes). The linked respondent sample imposes the further restriction that the individual does not have allocated earnings and it is possible to link their ASEC file to the DER in both years. Earnings are deflated by the personal consumption expenditure deflator with 2010 base year.

Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 1. Trends in Employment Rates by Education Attainment

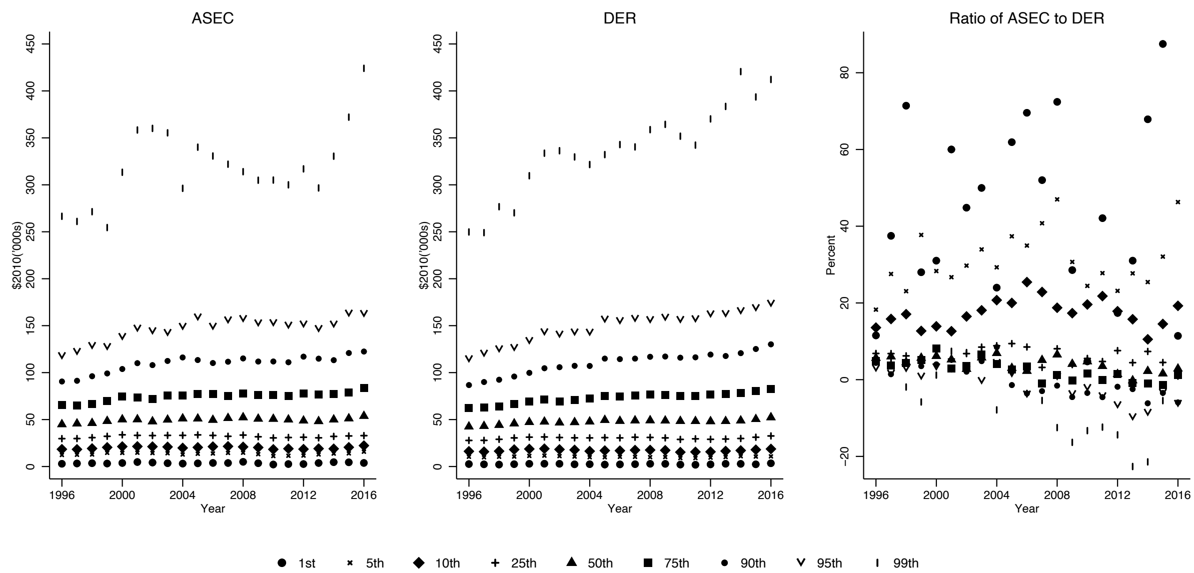


Note: Employment refers to any paid employment during the survey year.

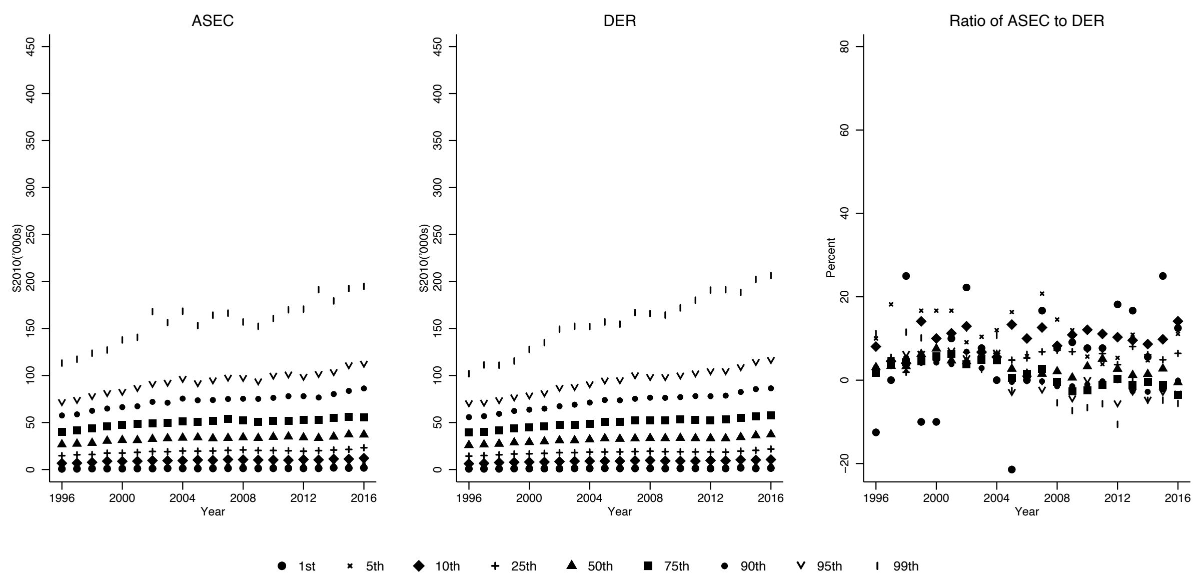
Source: U.S. Census Bureau, Current Population Survey, 1968-2017 Annual Social and Economic Supplement.

Figure 2. Percentiles of ASEC and DER Earnings Distribution of Linked Respondents

A. Men



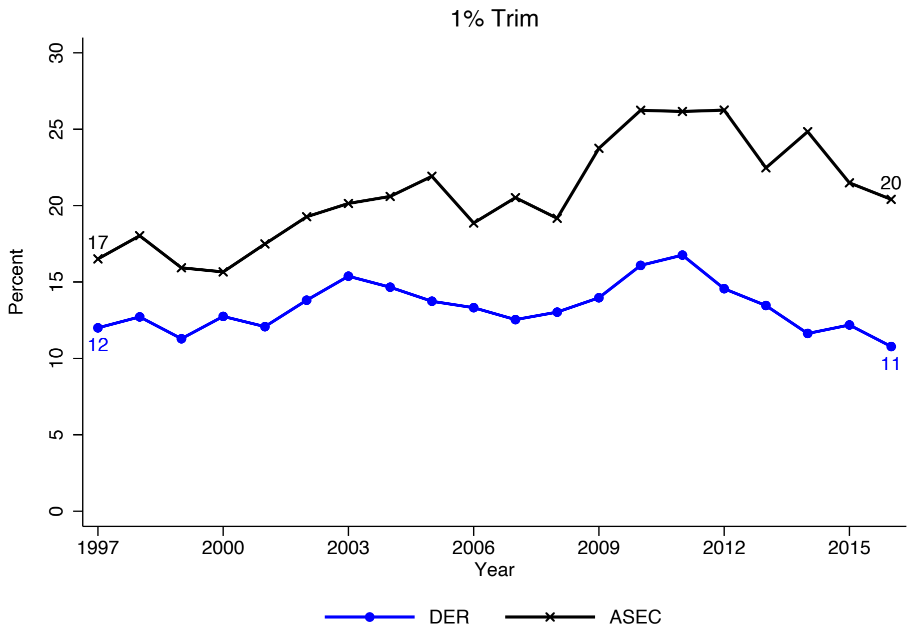
B. Women



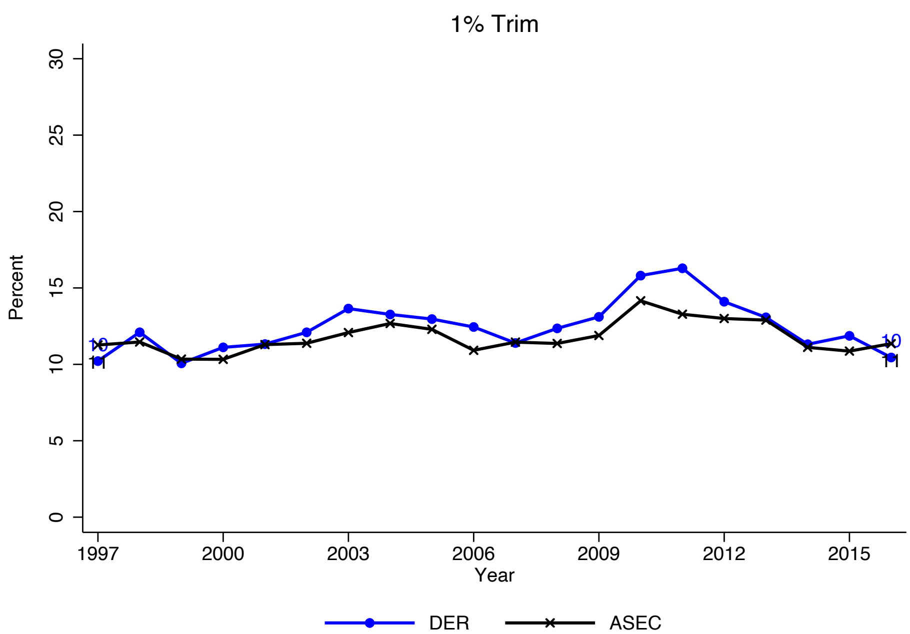
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 3. Arc Percent Earnings Volatility of Men

A. Linked Respondent Sample with Zeros



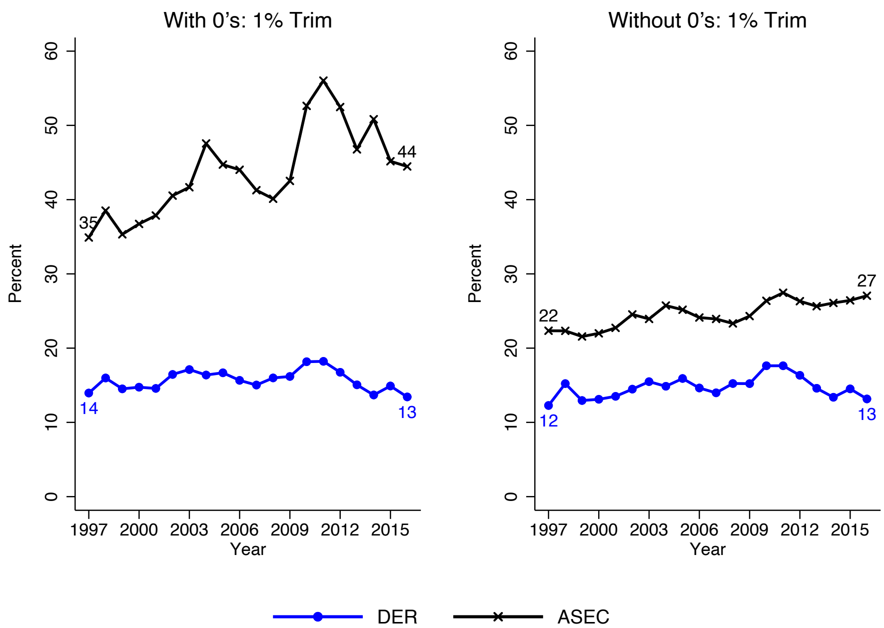
B. Linked Respondent Sample without Zeros



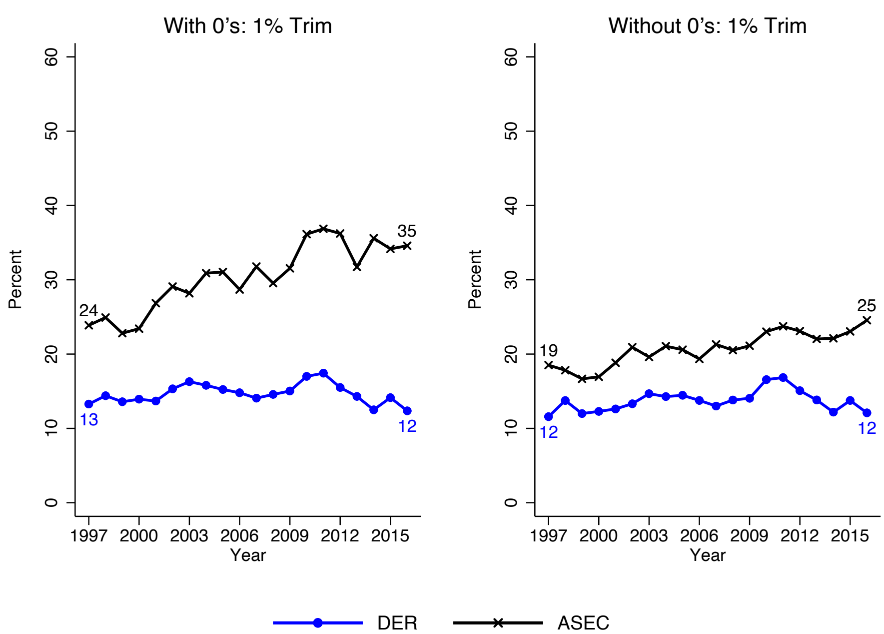
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 4. Arc Percent Earnings Volatility of Men by DER Link and ASEC Response

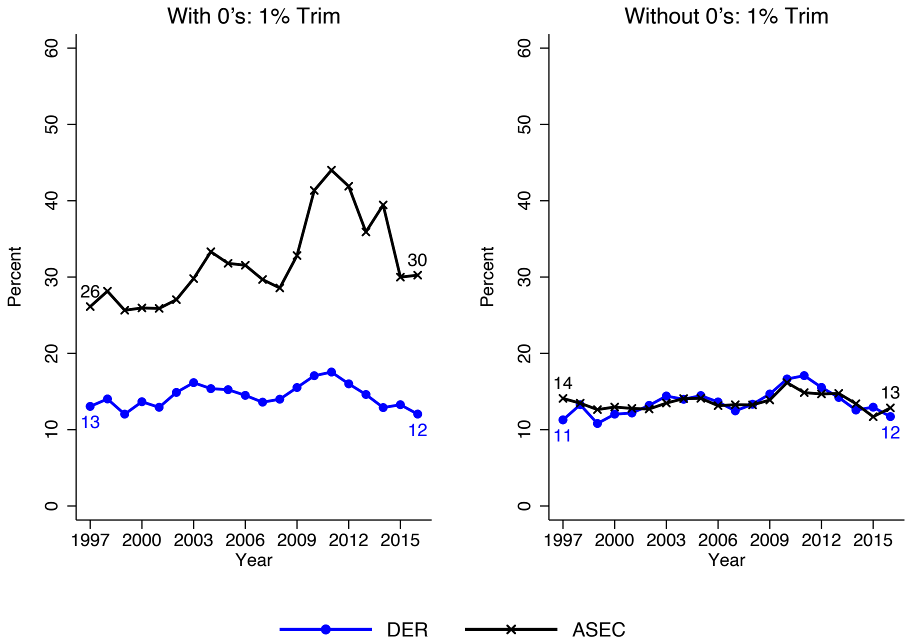
A. Full Sample with Zeros and without Zeros



B. Two-Year Linked Sample with and without Zeros



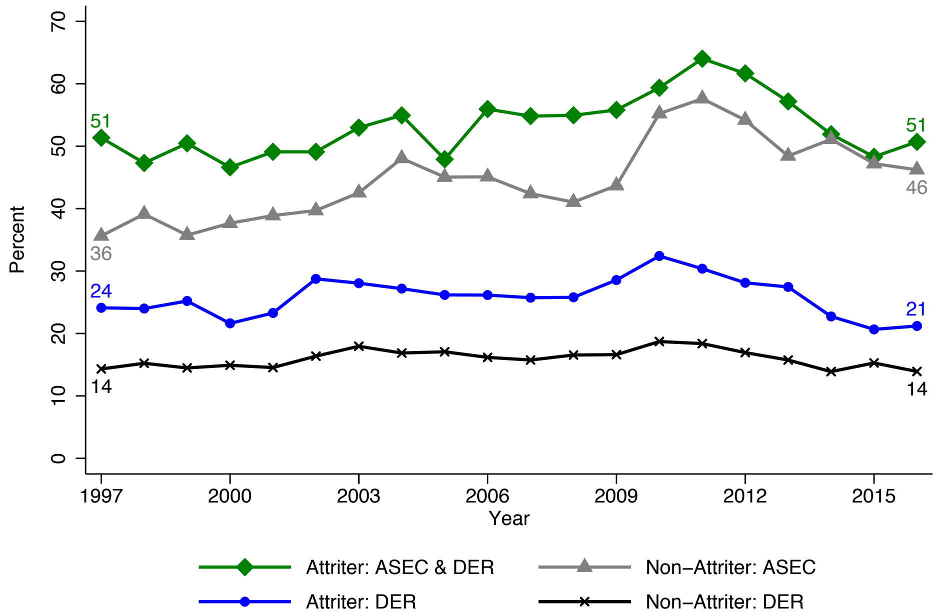
C. Two-Year Respondent Sample with and without Zeros



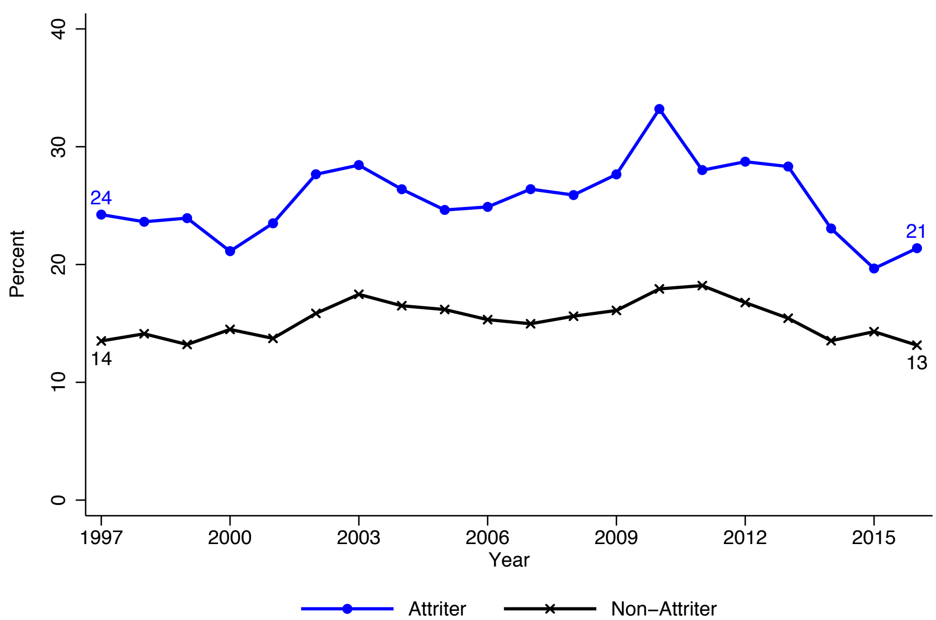
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 5. Arc Percent Earnings Volatility of Men by Attrition Status

A. Full Sample of Respondents and Nonrespondents with Zeros

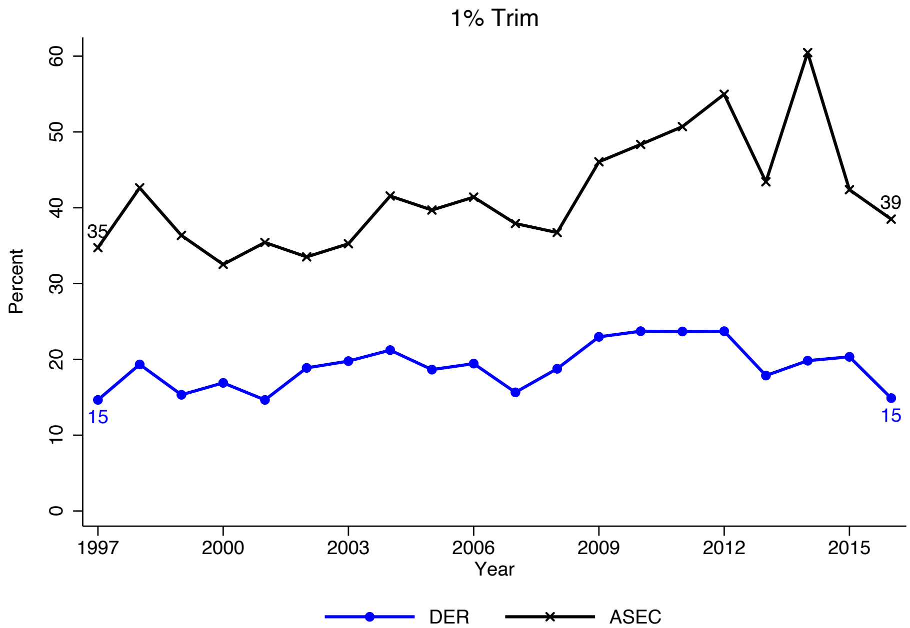


B. Sample of Year 1 Respondents with Zeros (DER only)



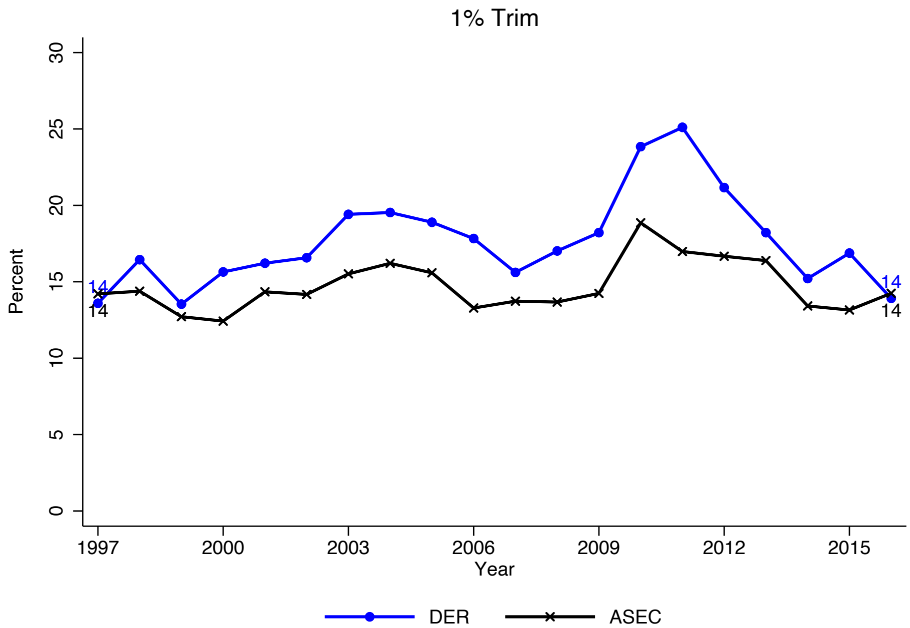
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 6. Inverse Probability Weighted Earnings Volatility of Men, Arc Percent with 0s



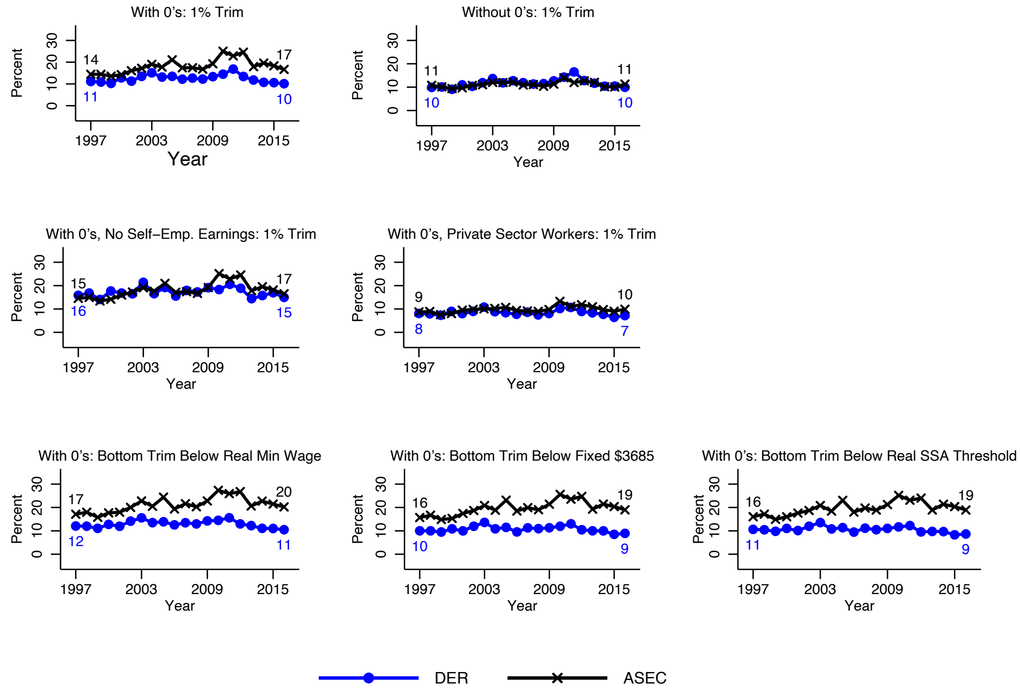
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 7. Log Difference Earnings Volatility of Men, Linked Respondents



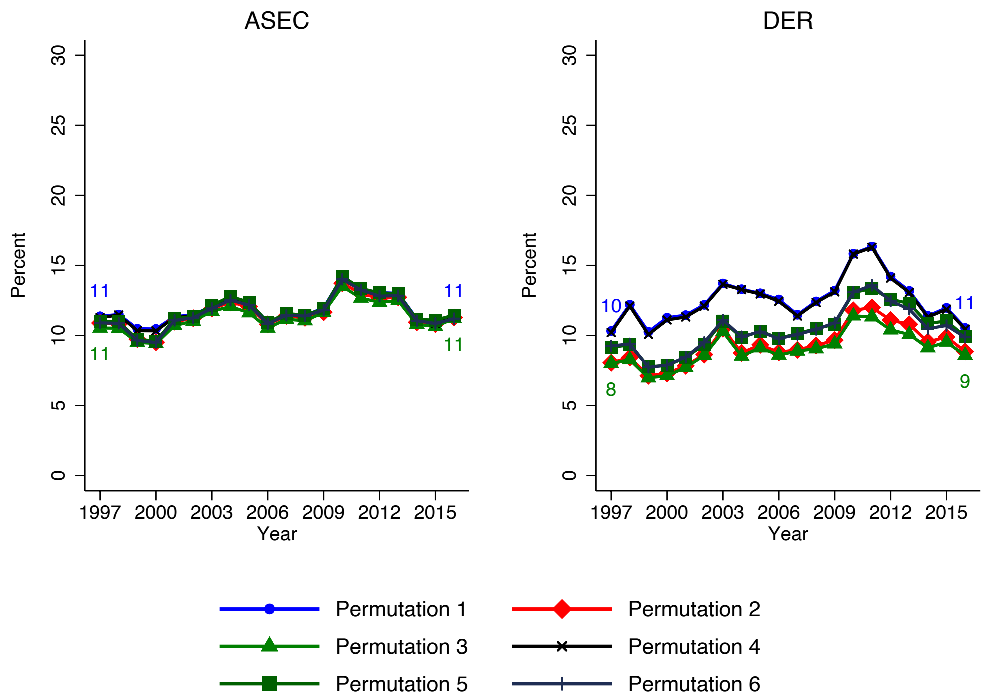
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 8. Arc Percent Earnings Volatility of Male Heads of Household, Linked Respondents



Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

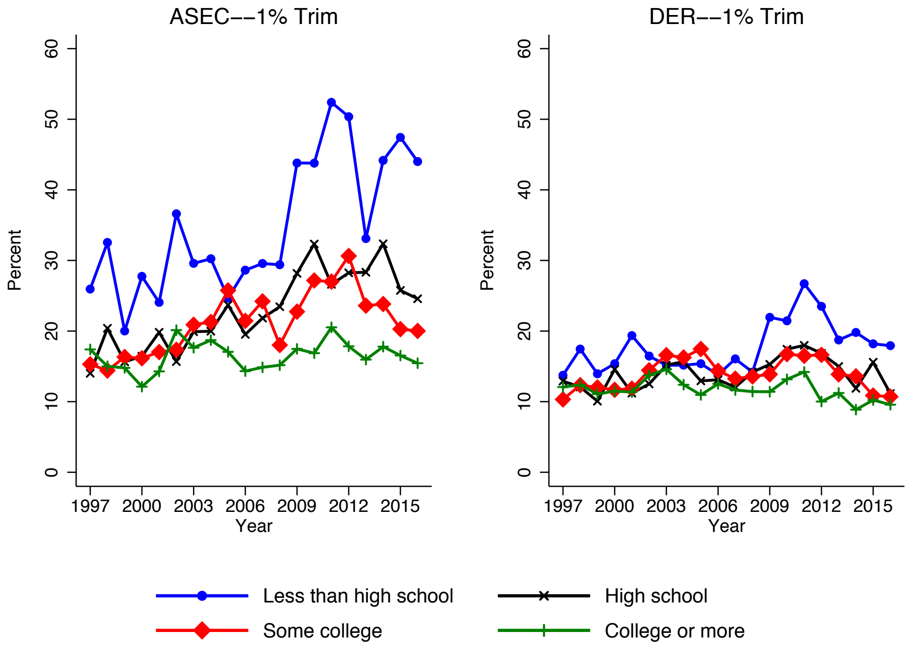
Figure 9. Arc Percent Earnings Volatility under Alternative Weighting, Linked Respondents with Zeros



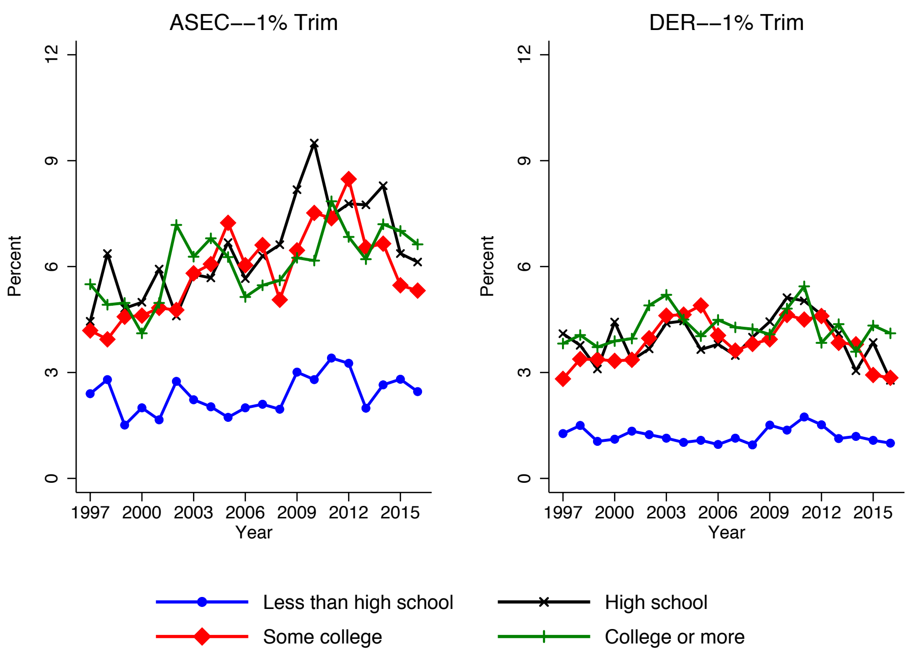
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 10. Within-Education Group Volatility of Men: Arc Percent Linked Respondents

A. Education-Specific Volatility



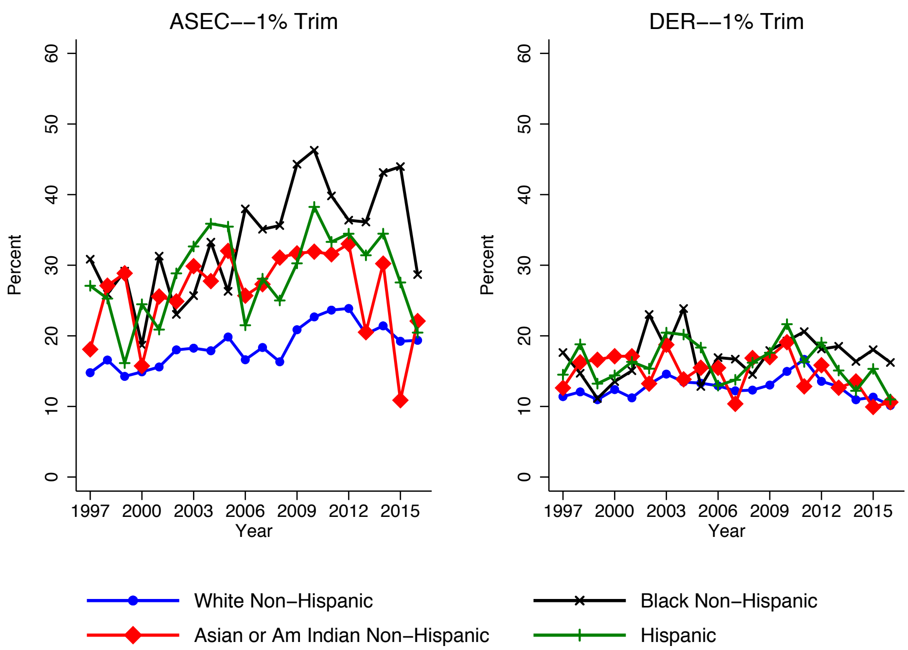
B. Weighted Education-Group Volatility



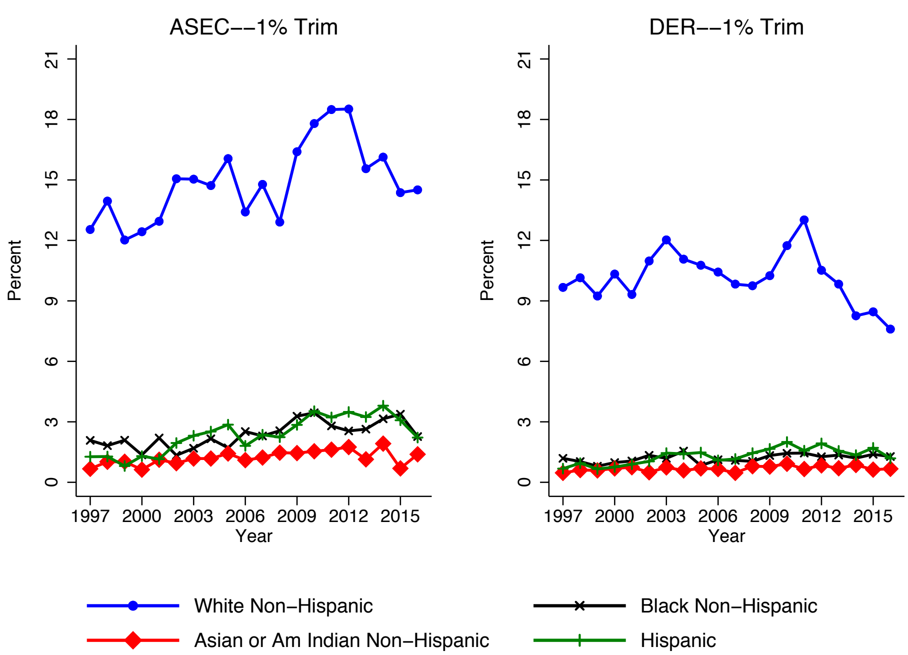
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 11. Within-Race/Ethnicity Group Volatility of Men, Arc Percent Linked Respondents

A. Race/Ethnicity-Specific Volatility



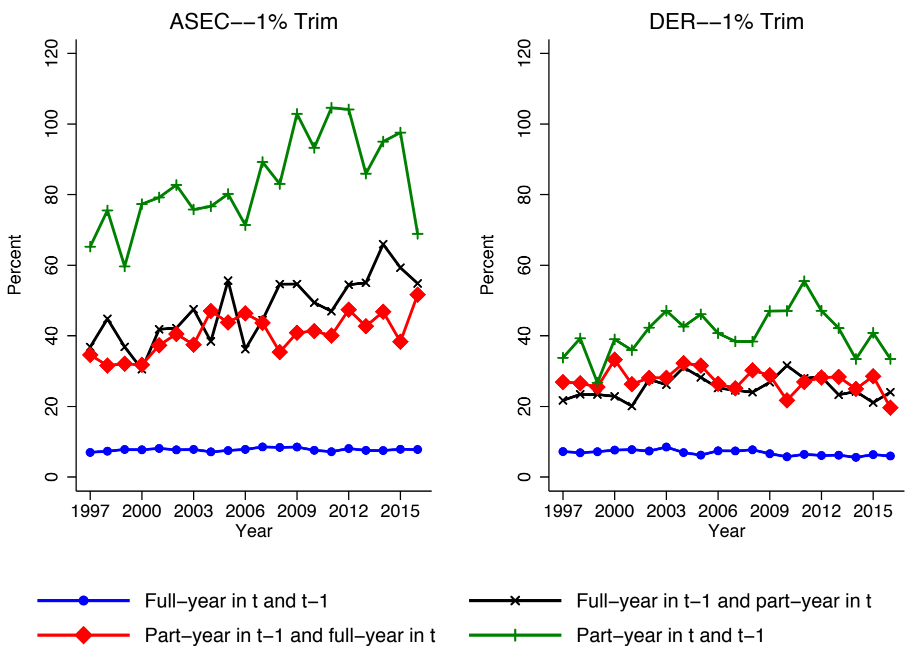
B. Weighted Race/Ethnicity-Group Volatility



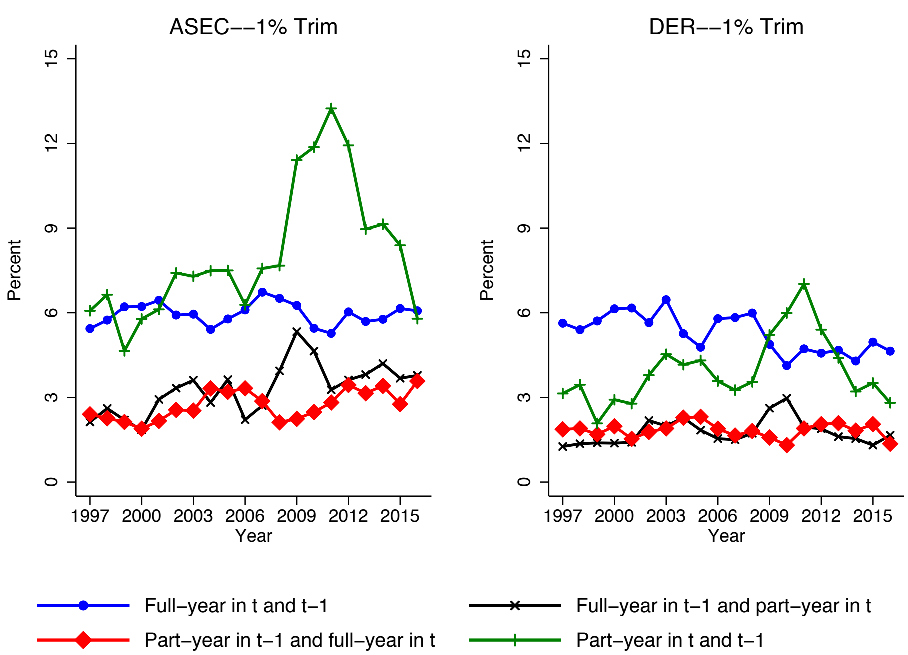
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 12. Within-Employment Status Group Volatility of Men, Arc Percent Linked Respondents

A. Employment-Specific Volatility



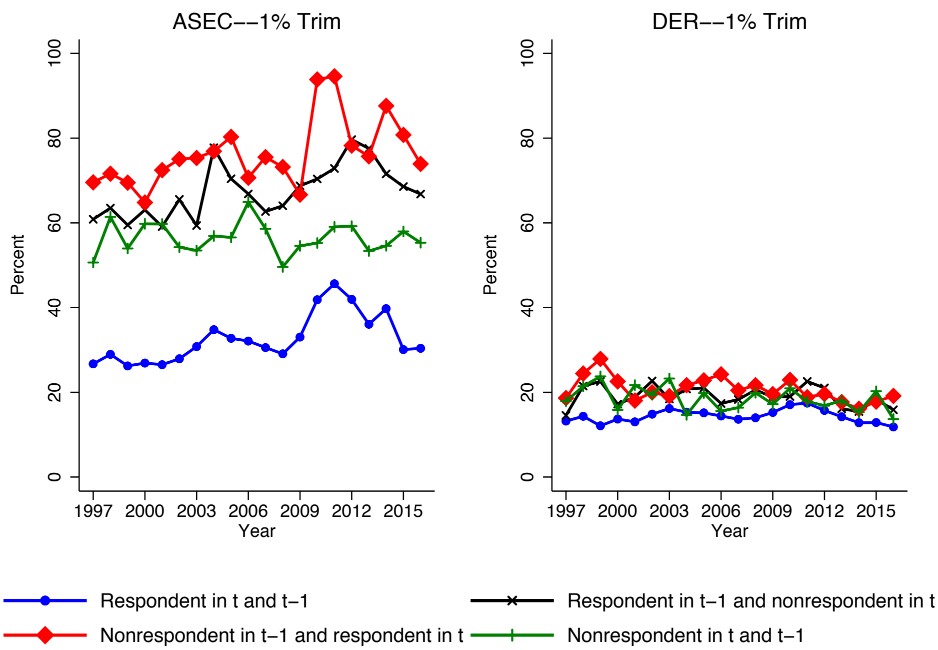
B. Weighted Employment-Group Volatility



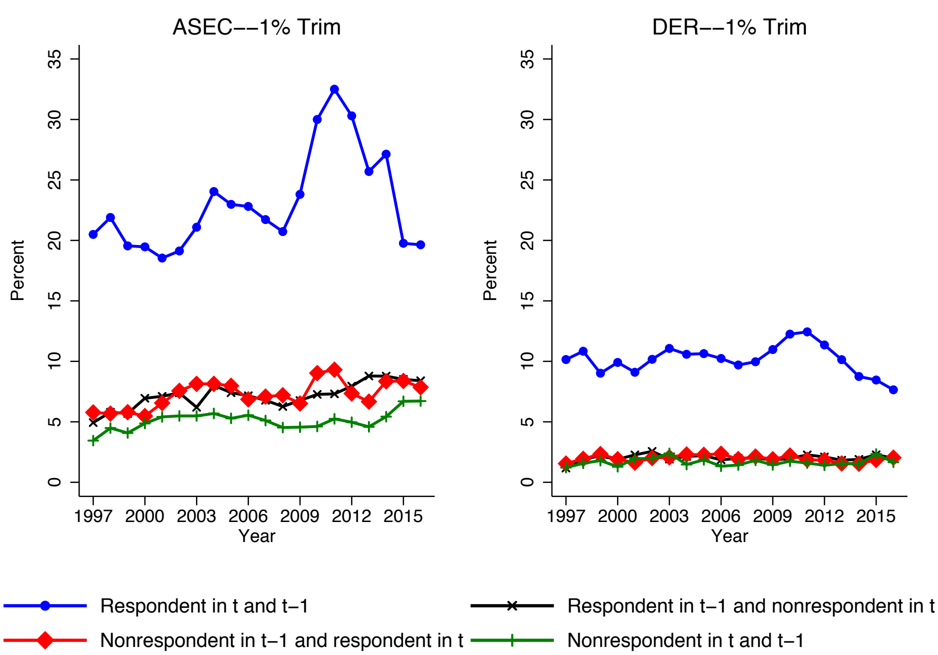
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 13. Within-Response Status Group Volatility of Men, Arc Percent Full Sample

A. Response-Specific Volatility



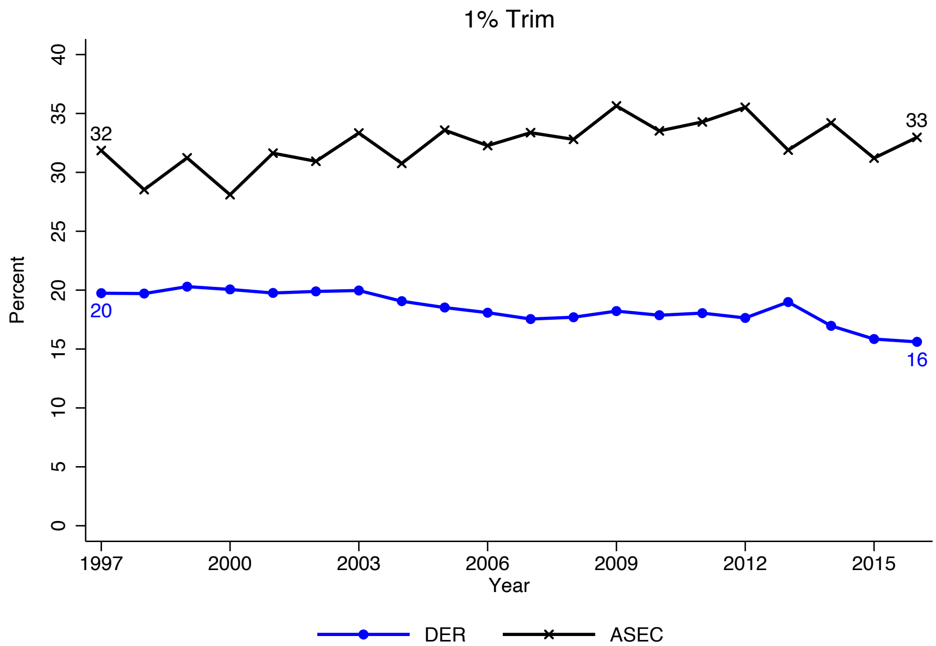
B. Weighted Response-Group Volatility



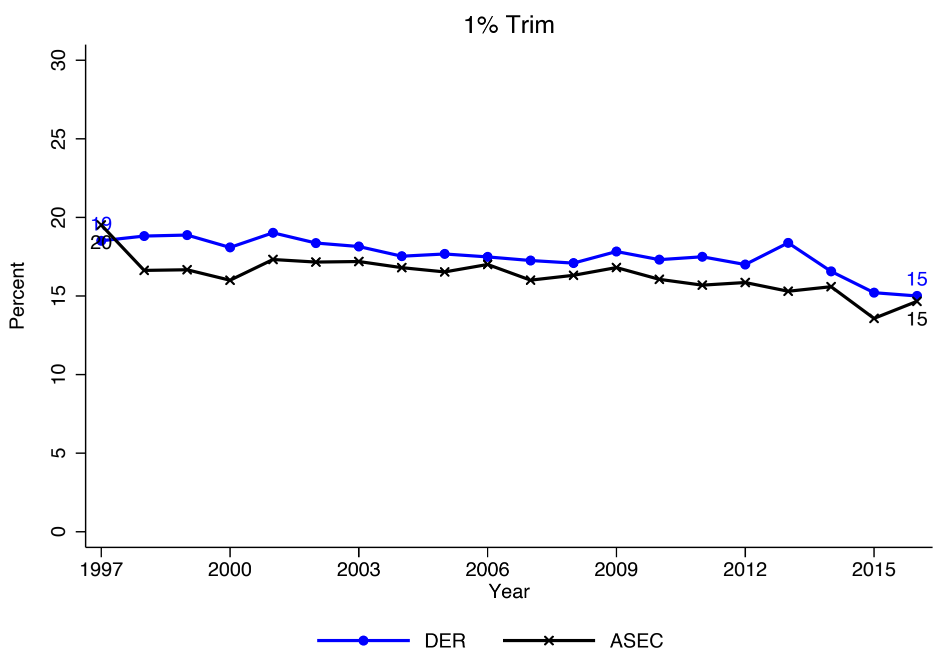
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 14. Arc Percent Earnings Volatility of Women

A. Linked Respondent Sample with Zeros



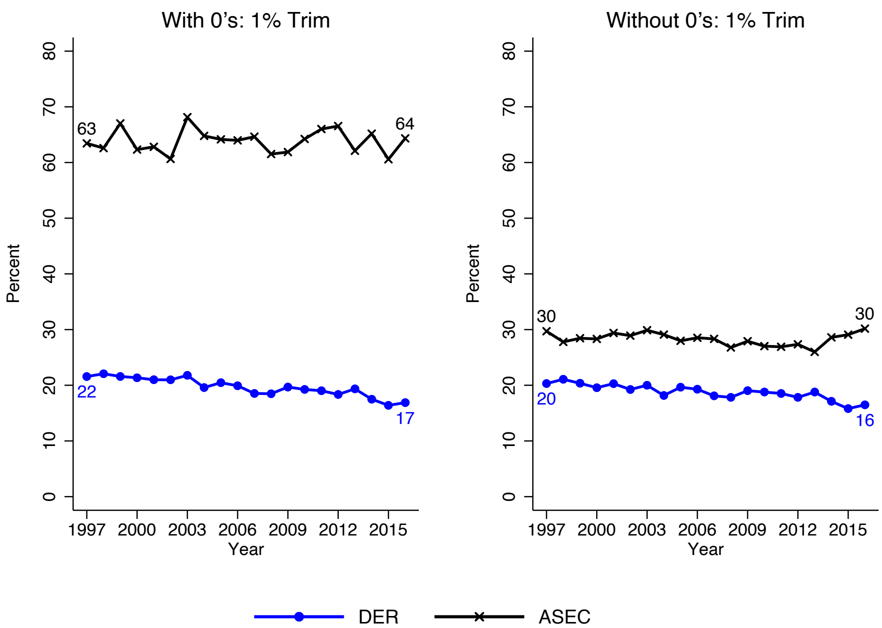
B. Linked Respondent Sample without Zeros



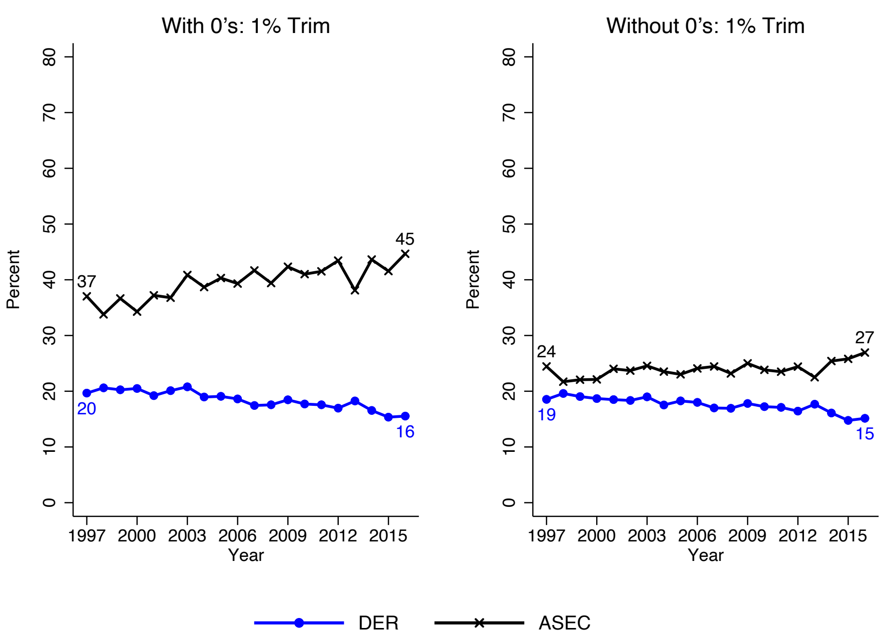
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 15. Arc Percent Earnings Volatility of Women by DER Link and ASEC Response

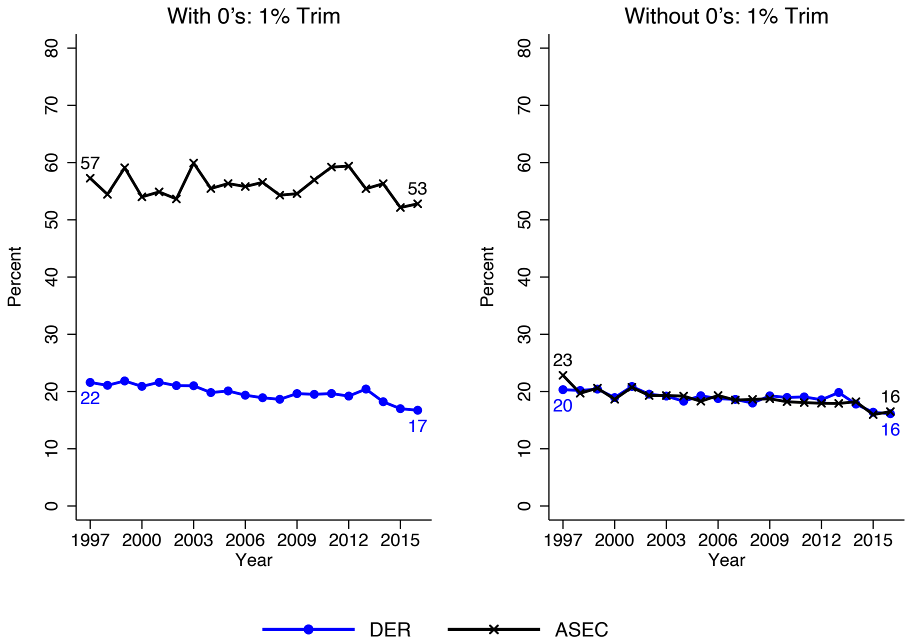
A. Full Sample with Zeros and without Zeros



B. Two-Year Linked Sample with and without Zeros



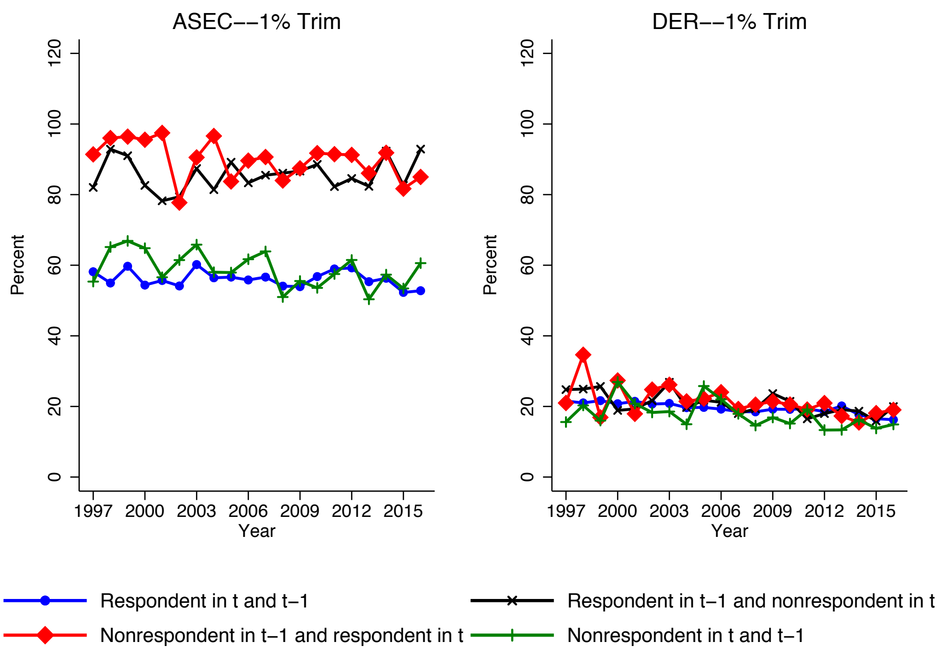
C. Two-Year Respondent Sample with and without Zeros



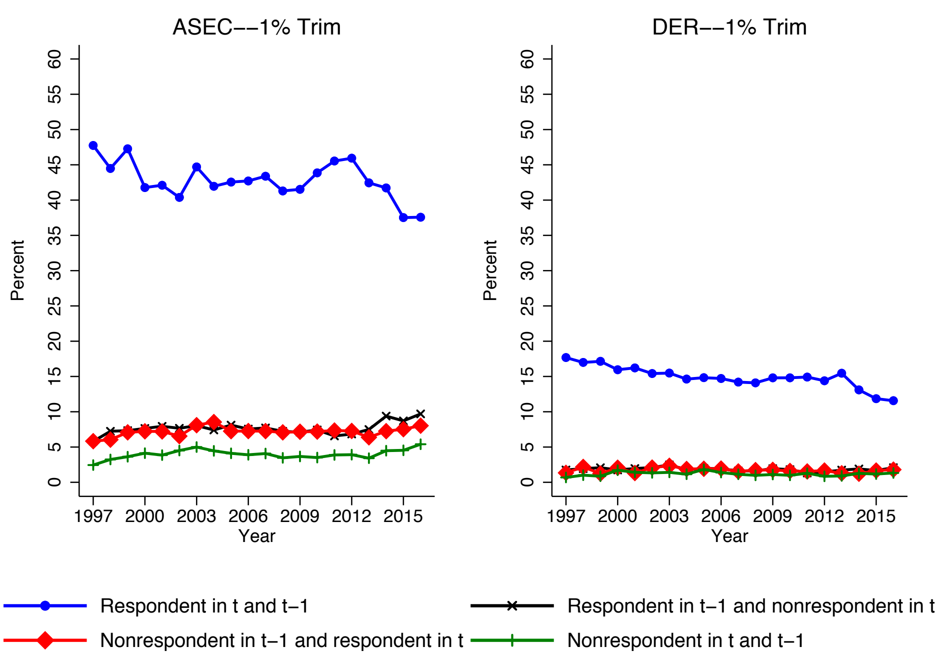
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 16. Within-Response Status Group Volatility of Women, Arc Percent Full Sample

A. Response-Specific Volatility



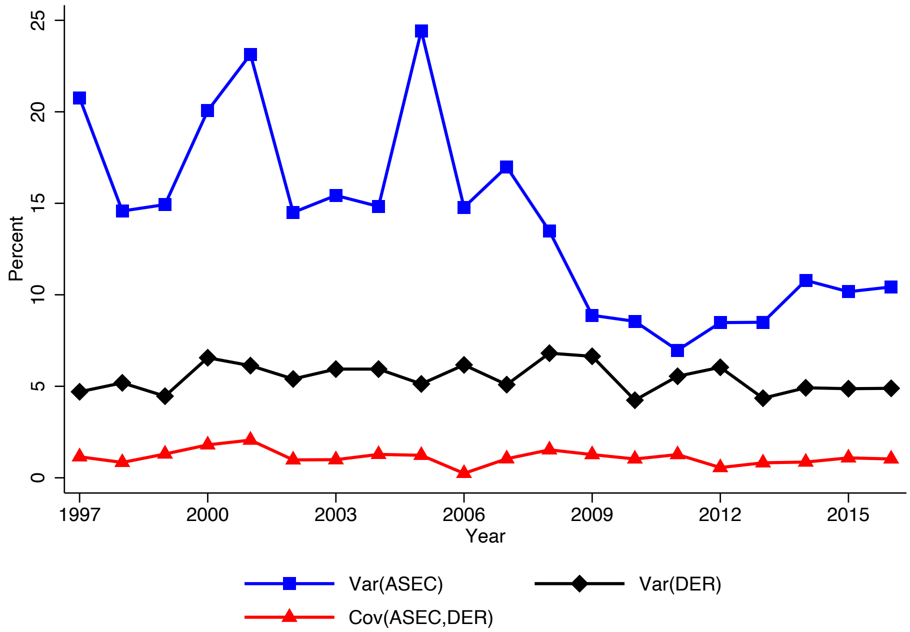
B. Weighted Response-Group Volatility



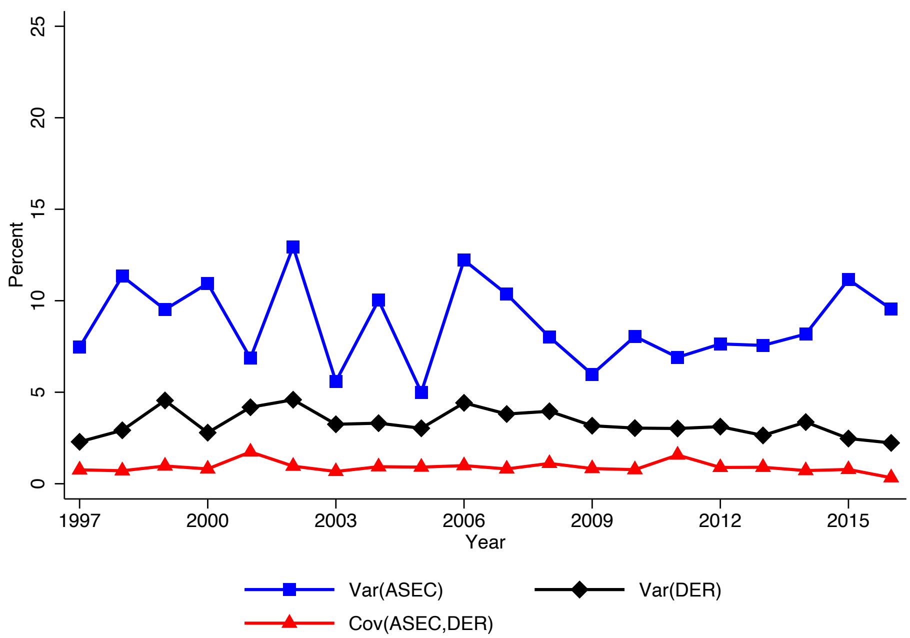
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 17. Transitory Earnings Volatility

A. Men



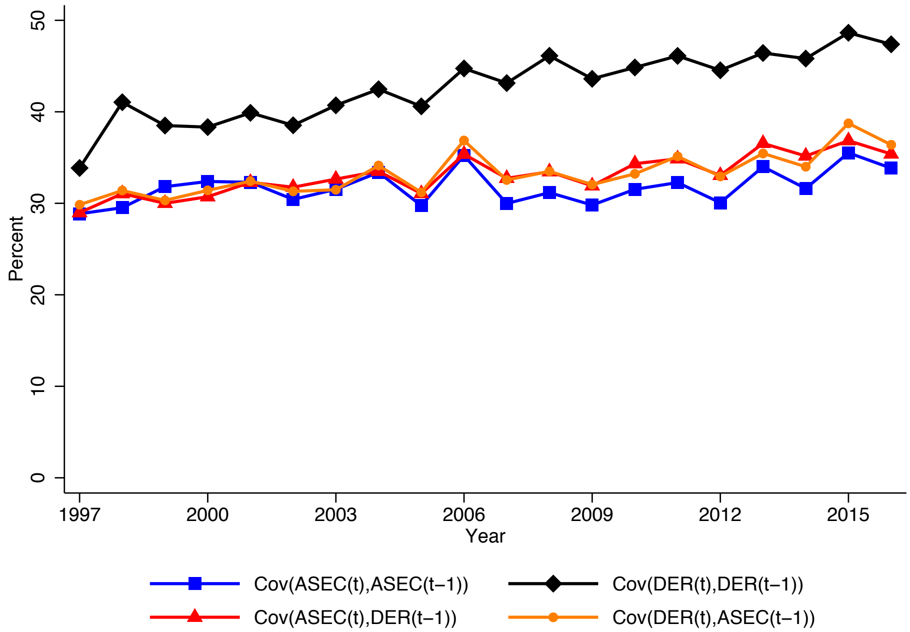
B. Women



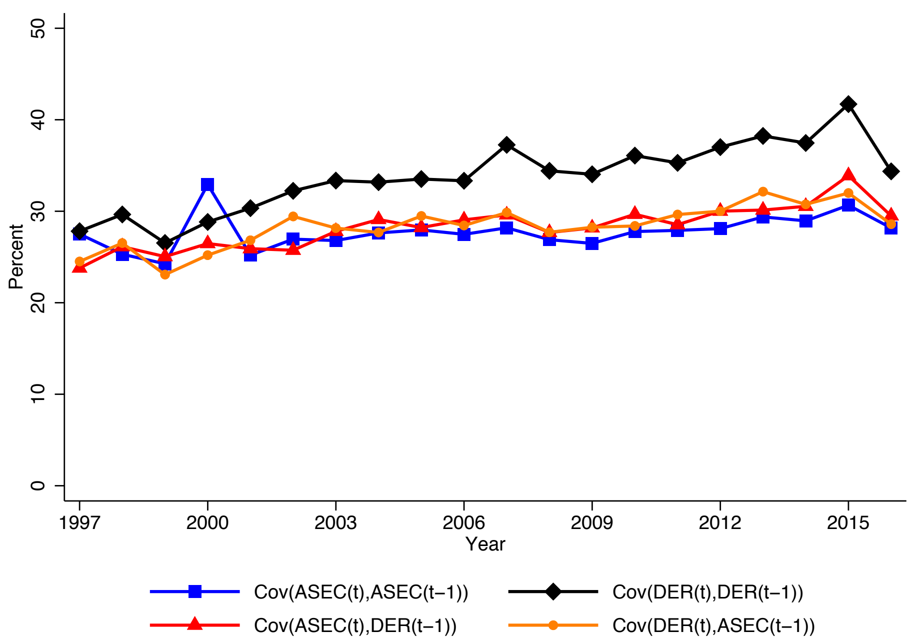
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Figure 18. Permanent Earnings Volatility

A. Men



B. Women



Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Appendix Table 1. ASEC-DER Linkage Rate and 2-Yr Panel Match Rate

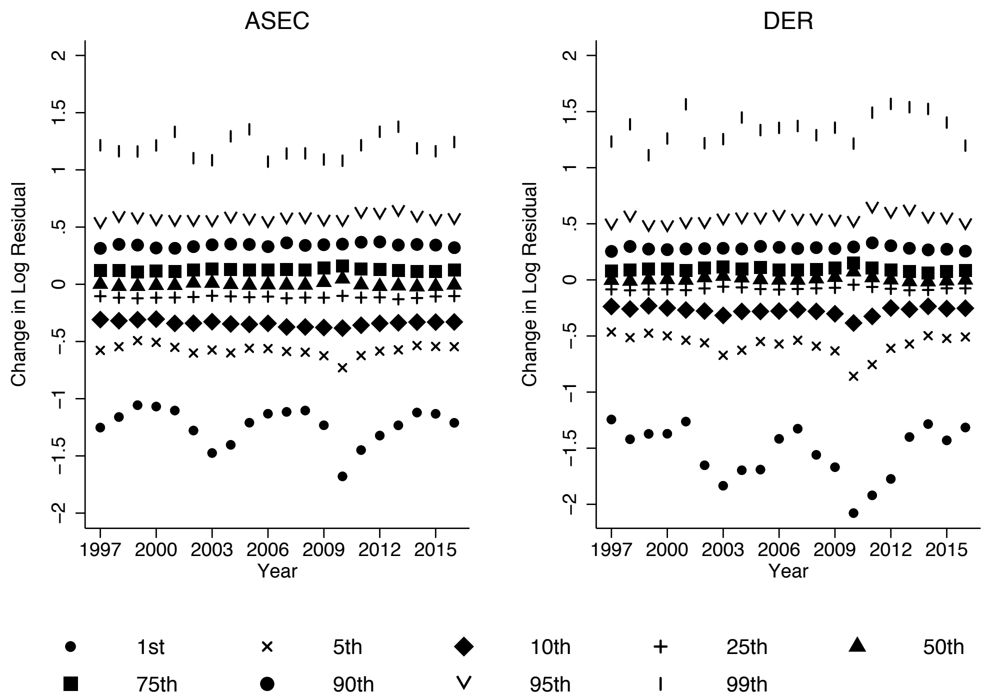
Year	Linkage Rate	Panel Match Rate
1996	79.22%	
1997	76.64	69.52%
1998	71.79	70.10
1999	66.43	69.96
2000	66.74	62.54
2001	69.20	64.20
2002	74.26	62.60
2003	71.39	74.44
2004	64.17	76.38
2005	62.48	67.26
2006	86.58	73.62
2007	86.48	74.65
2008	86.12	75.76
2009	85.73	75.73
2010	84.94	75.75
2011	85.67	76.89
2012	85.39	76.41
2013	85.14	76.05
2014	84.40	74.82
2015	84.53	62.26
2016	84.16	72.71

Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Appendix Table 2. Sample Summary Statistics by Attrition Status

	A. Men			
	Non-Attriters		Attriters	
	Mean	Std. Dev.	Mean	Std. Dev.
Age	42.65	9.47	38.62	9.63
White	0.71	0.46	0.64	0.48
Black	0.09	0.29	0.13	0.34
Asian or American Indian	0.06	0.24	0.06	0.24
Hispanic	0.14	0.35	0.16	0.37
Years Education	13.66	2.80	13.37	2.82
Married, Spouse Present	0.61	0.49	0.44	0.50
Married, Spouse Absent	0.15	0.35	0.20	0.40
Never Married	0.25	0.43	0.36	0.48
Native	0.84	0.37	0.83	0.37
Foreign Citizen	0.07	0.26	0.06	0.23
Foreign Non-Citizen	0.09	0.29	0.11	0.31
Weeks Worked	42.95	18.12	40.27	19.44
Hours per Week	37.97	17.22	35.92	17.48
Proxy Response	0.50	0.50	0.51	0.50
Earnings Nonresponse	0.18	0.38	0.23	0.42
Real ASEC Earnings ($2010 thou.)	51.72	71.34	41.56	64.65
Real DER Earnings ($2010 thou.)	64.78	151.10	48.28	54.86
Persons in Year 1 (rounded)	204,000		79,000	
	B. Women			
Age	42.95	9.39	38.79	9.82
White	0.69	0.46	0.63	0.48
Black	0.11	0.31	0.14	0.35
Asian or American Indian	0.06	0.25	0.07	0.26
Hispanic	0.13	0.34	0.16	0.37
Years Education	13.95	2.72	13.67	2.72
Married, Spouse Present	0.62	0.49	0.46	0.50
Married, Spouse Absent	0.19	0.39	0.26	0.44
Never Married	0.19	0.39	0.29	0.45
Native	0.83	0.37	0.83	0.37
Foreign Citizen	0.08	0.27	0.06	0.24
Foreign Non-Citizen	0.09	0.28	0.11	0.31
Weeks Worked	36.43	22.28	34.94	22.67
Hours per Week	28.91	18.68	28.57	19.11
Proxy Response	0.41	0.49	0.42	0.49
Earnings Nonresponse	0.15	0.35	0.19	0.39
Real ASEC Earnings ($2010 thou.)	29.68	37.57	26.79	36.10
Real DER Earnings ($2010 thou.)	38.65	48.13	32.92	31.58
Persons in Year 1 (rounded)	223,000		81,000	

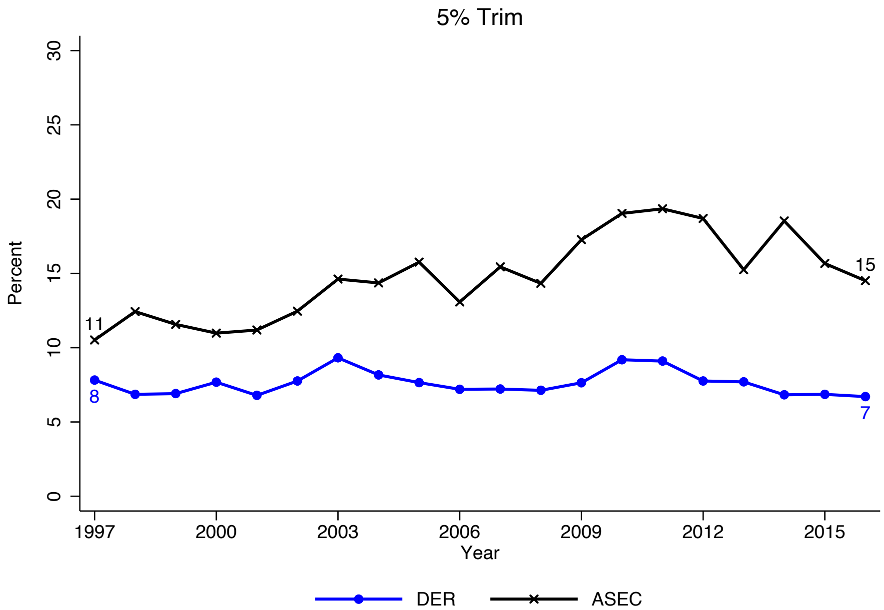
Appendix Figure 1. Percentiles of Log Difference Residuals of Men



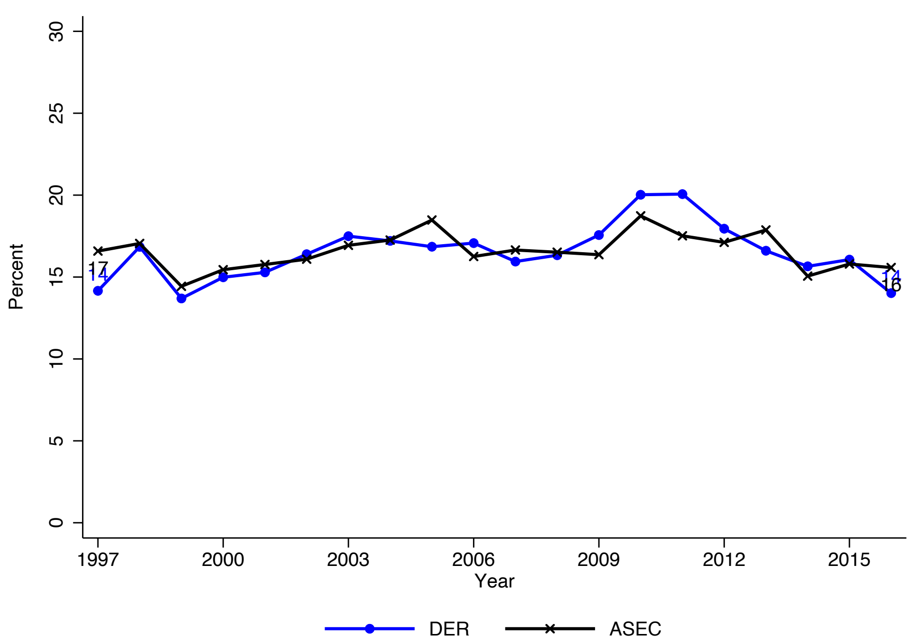
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Appendix Figure 2. Earnings Volatility of Men, Alternative Trim

A. Arc Percent Change with 0s and 5% Trim



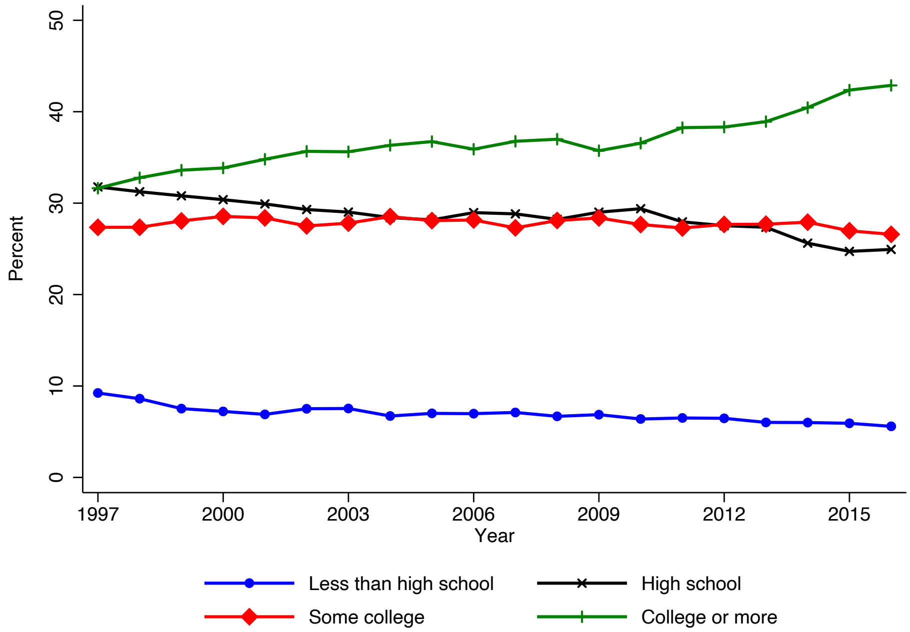
B. Arc Percent, Linked Respondents without 0s and No Trim



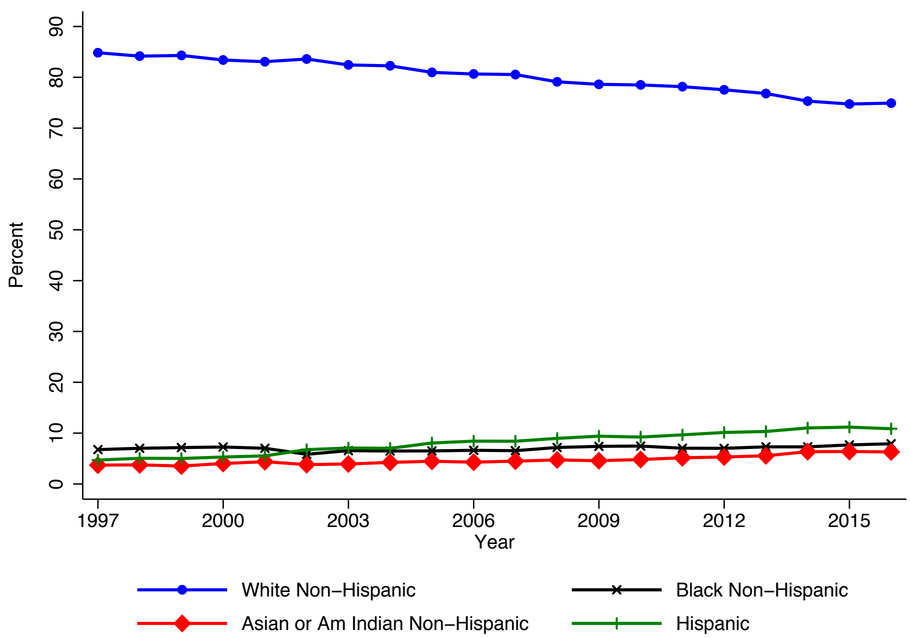
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Appendix Figure 3. Population Shares of Men used in Variance Decompositions

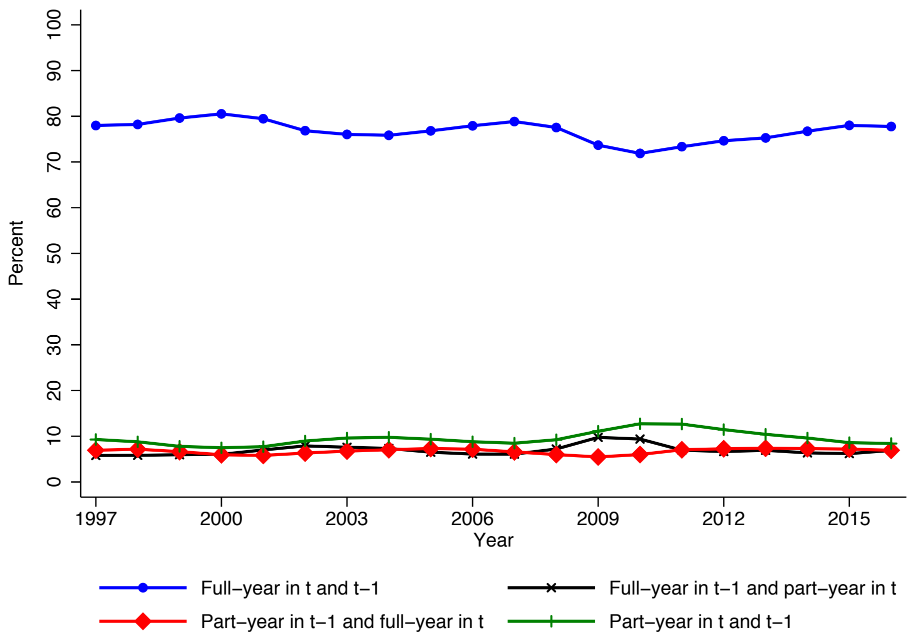
A. Education Attainment



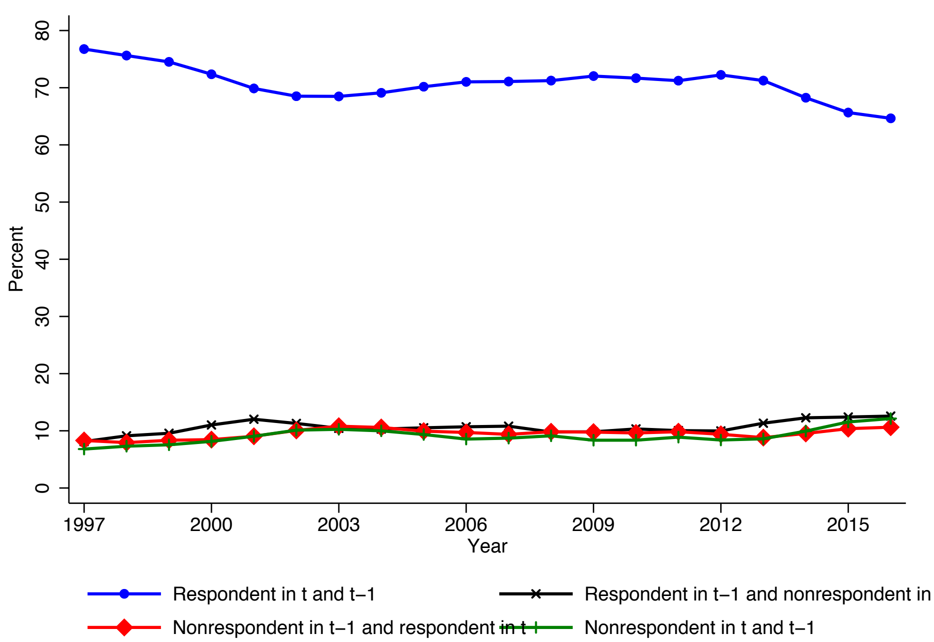
B. Race and Ethnicity



C. Full-time/Part-time Employment Status

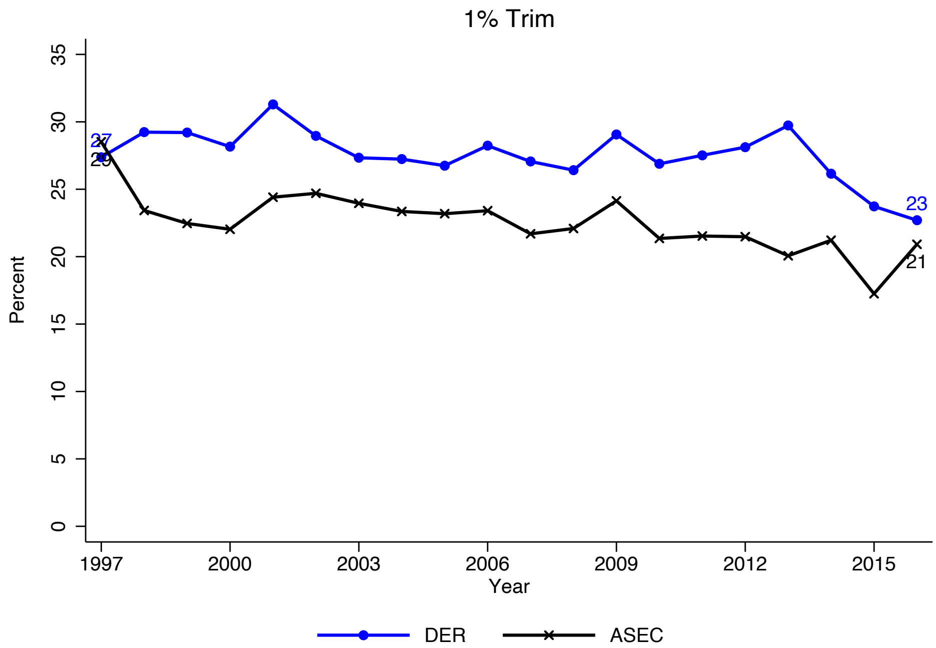


D. ASEC Earnings Response Status



Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

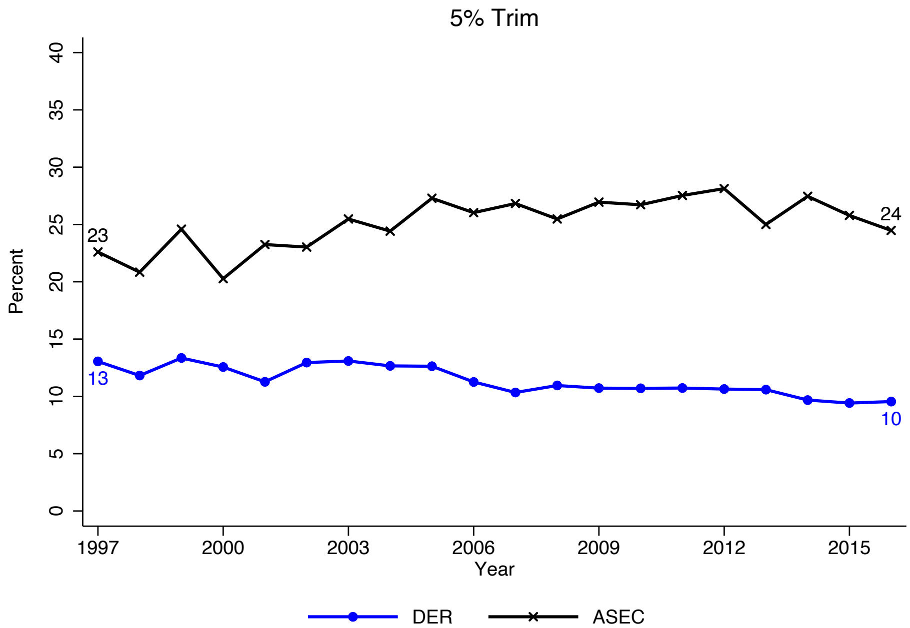
Appendix Figure 4. Log-Difference Earnings Volatility of Women



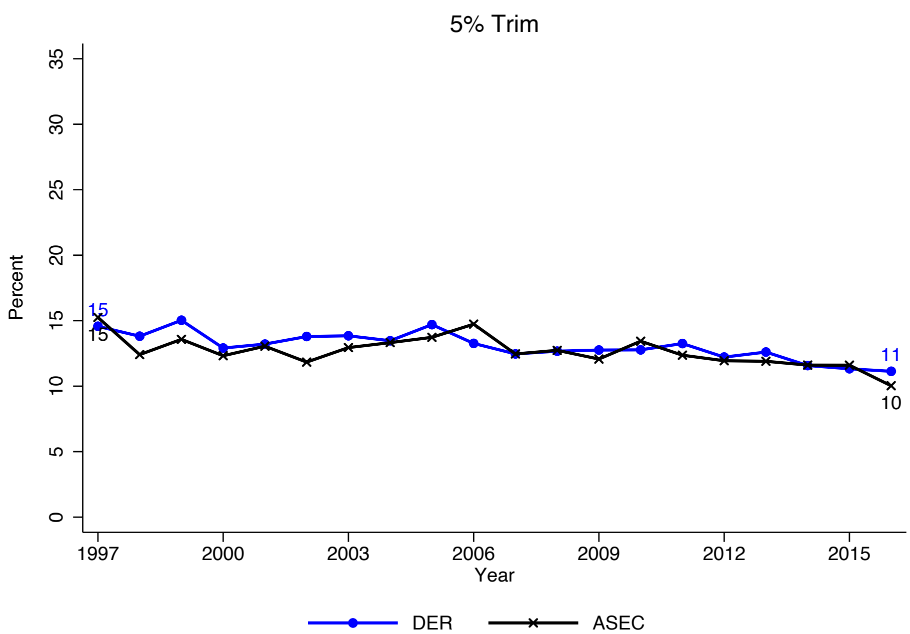
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Appendix Figure 5. Earnings Volatility of Women, Alternative Trim

A. Arc Percent Change with 0s



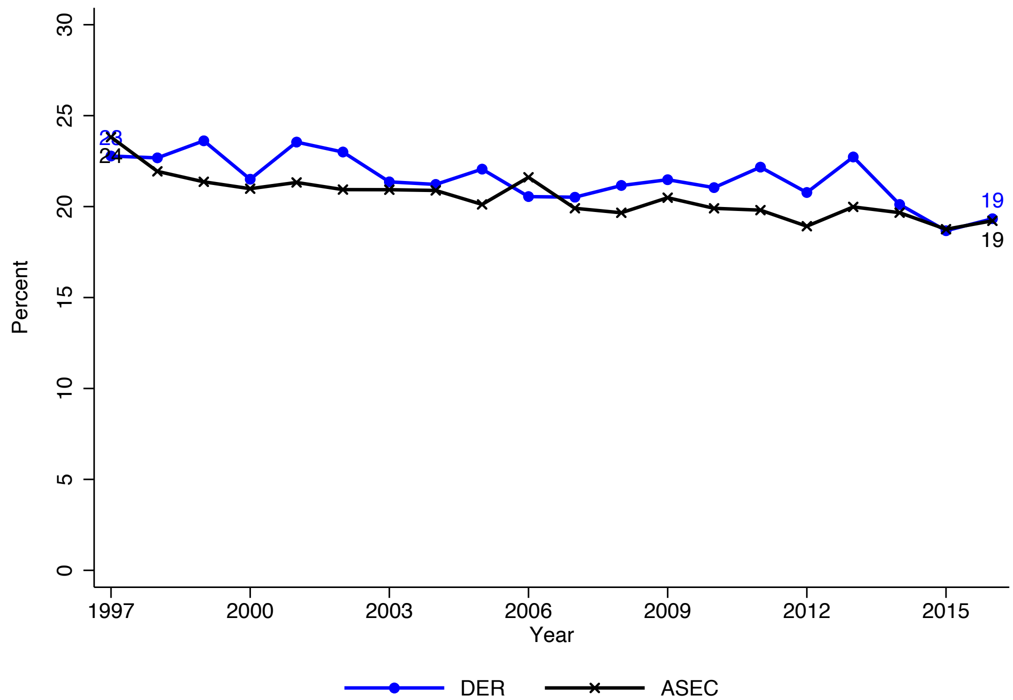
B. Log Difference



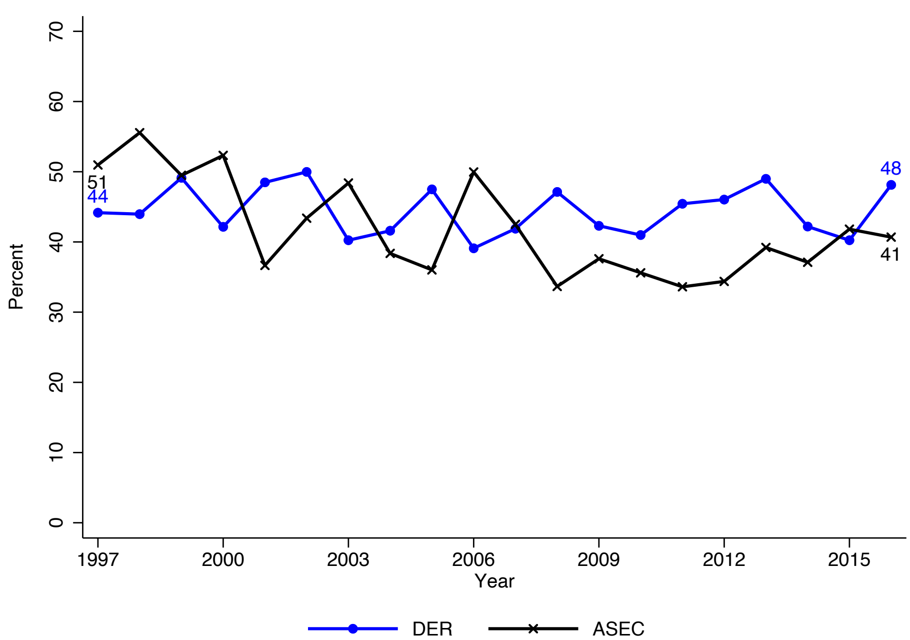
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Appendix Figure 6. Earnings Volatility of Women, No Trim

A. Arc Percent Change without 0s



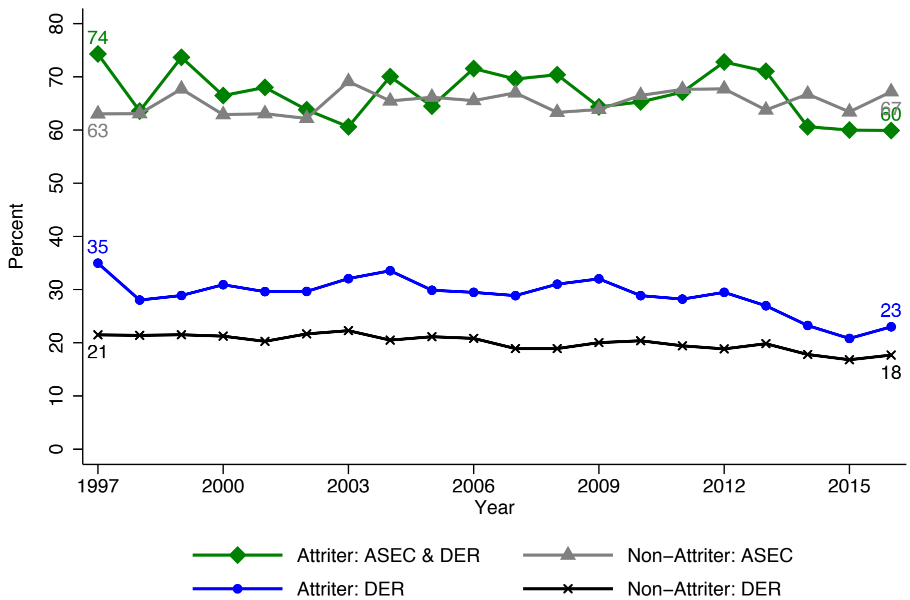
B. Log Difference



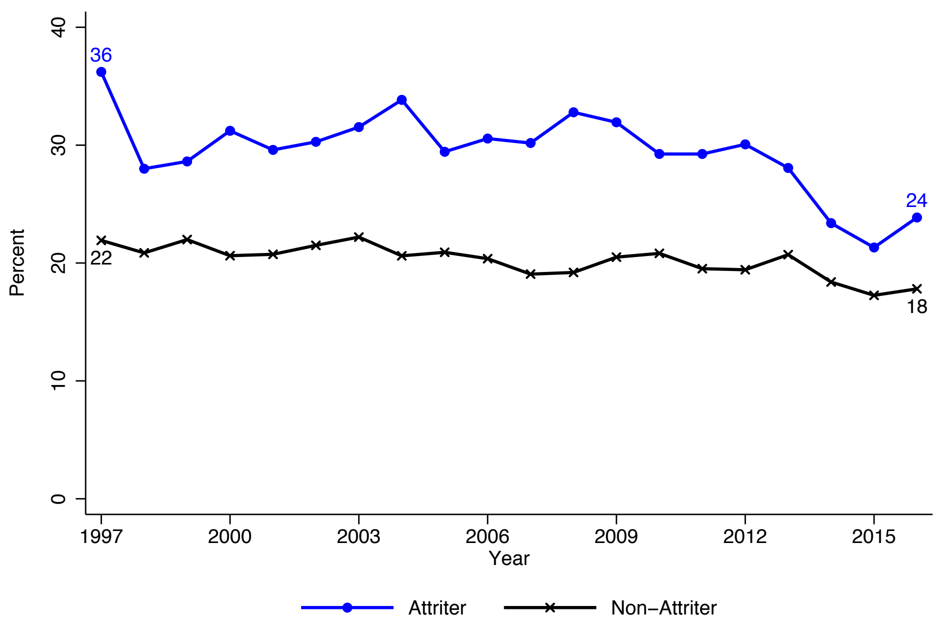
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Appendix Figure 7. Arc Percent Volatility of Women by Attrition Status

A. Full Sample of Respondents and Nonrespondents with Zeros (1% trim)

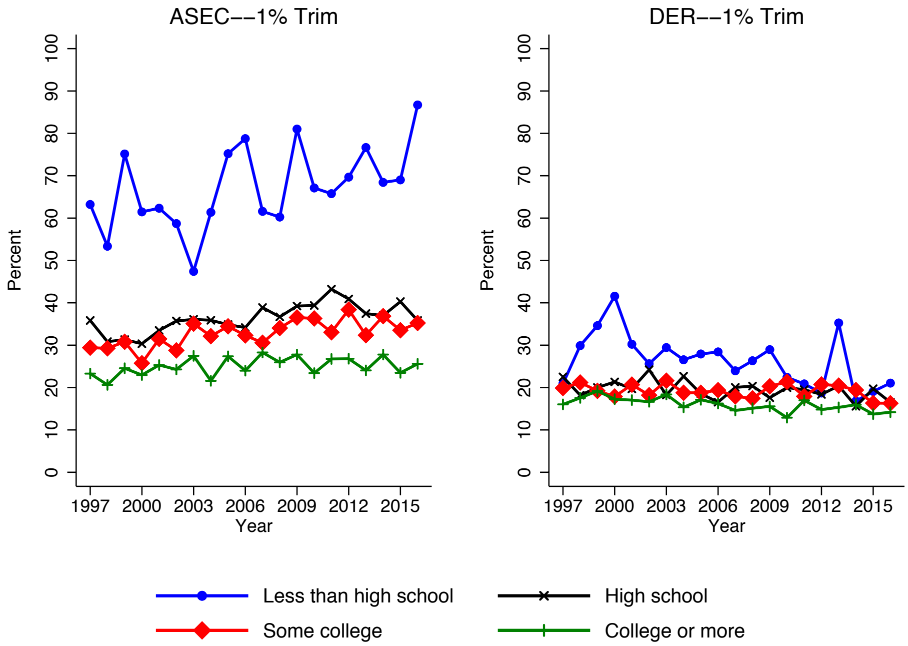


B. Sample of Year 1 Respondents with Zeros (DER 1% trim)

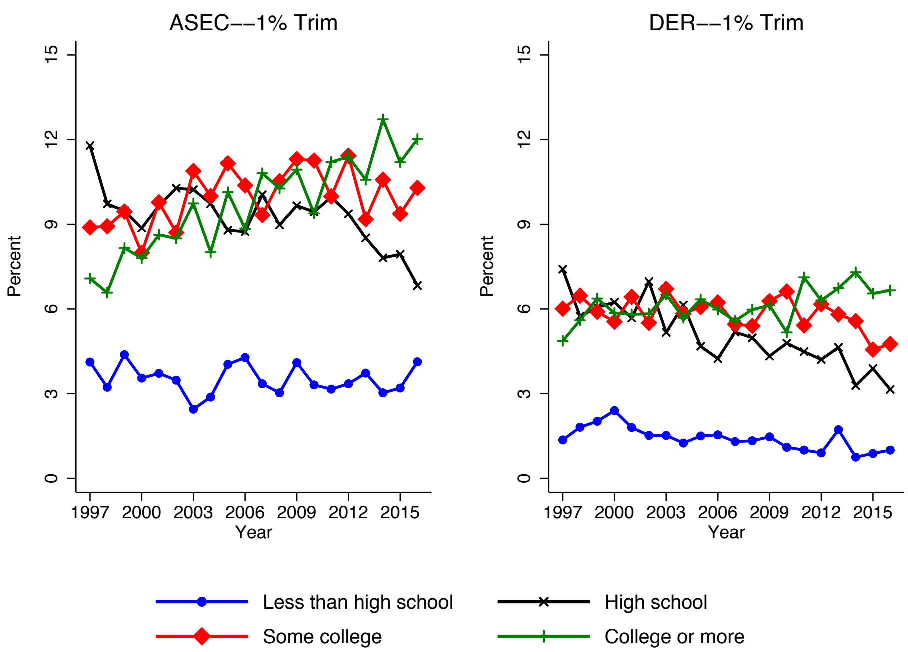


Appendix Figure 8. Within-Education Group Volatility of Women: Arc Percent Linked Respondents

A. Education-Specific Volatility



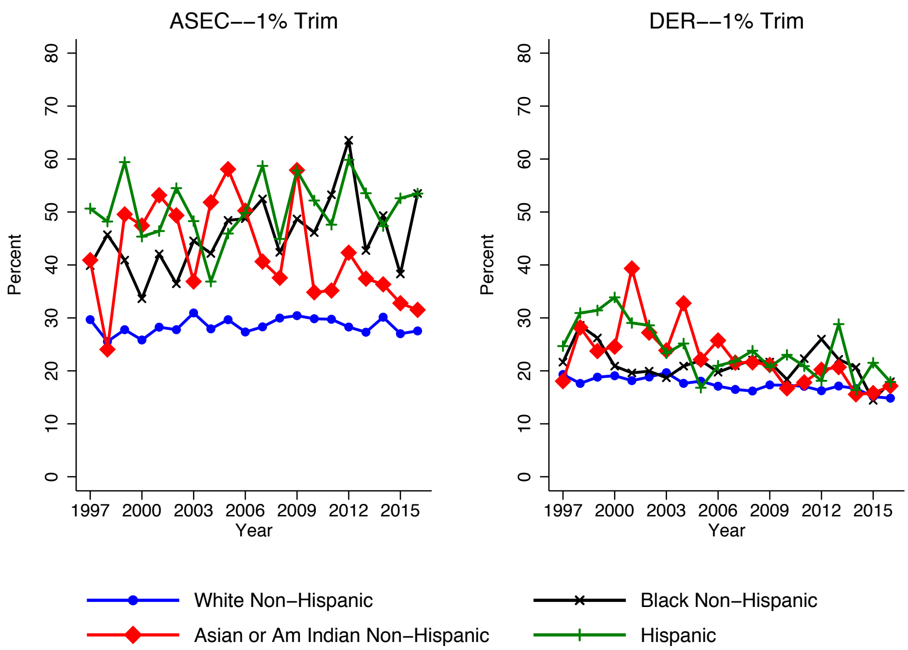
B. Weighted Education-Group Volatility



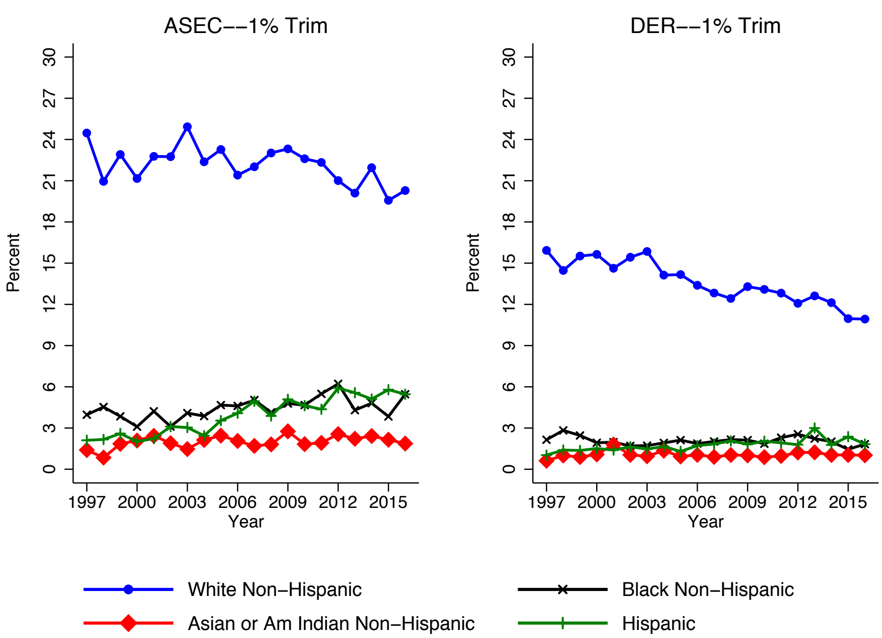
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Appendix Figure 9. Within-Race/Ethnicity Group Volatility of Women, Arc Percent Linked Respondents

A. Race/Ethnicity-Specific Volatility



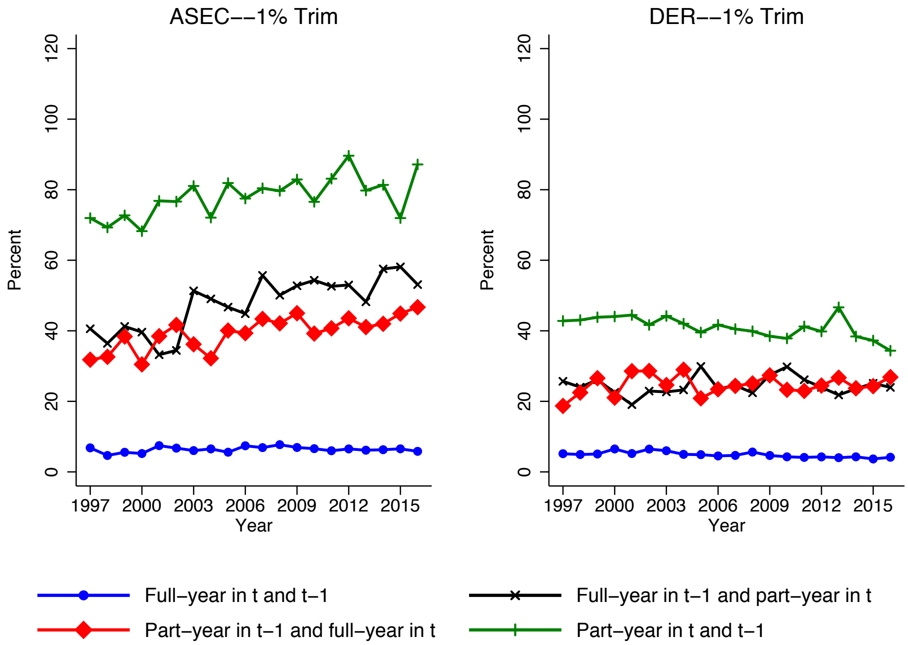
B. Weighted Race/Ethnicity-Group Volatility



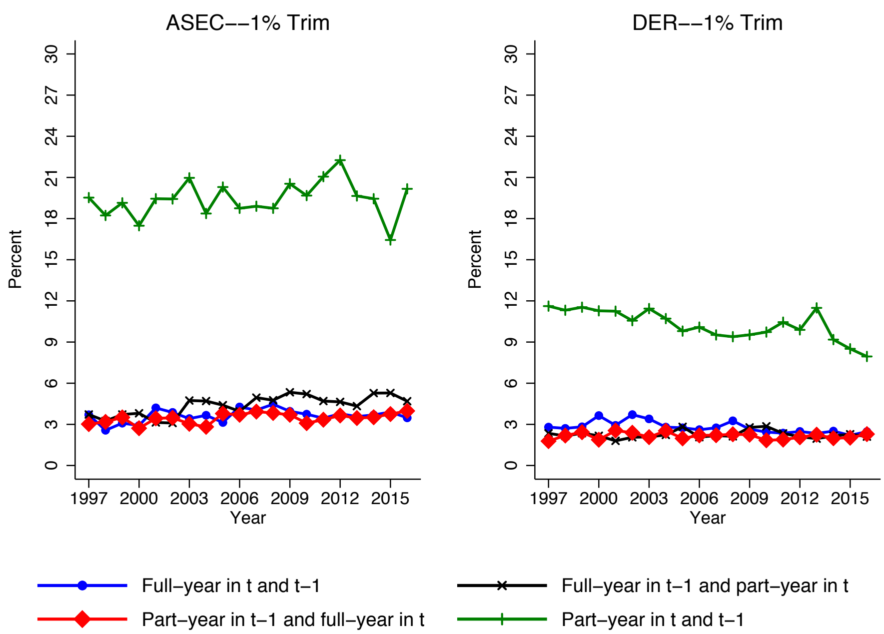
Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.

Appendix Figure 10. Within-Employment Status Group Volatility of Women, Arc Percent Linked Respondents

A. Employment-Specific Volatility



B. Weighted Employment-Group Volatility



Sources: U.S. Census Bureau, Current Population Survey, 1996-2016 Annual Social and Economic Supplement. Social Security Administration, Detailed Earnings Record, 1995-2015.