Female earnings inequality: Understanding the Role of Family Factors on the Extensive and Intensive Margins

David Card* Dean Hyslop†

2018-09-01

Abstract: Although women make up nearly half the U.S. workforce, most studies of earnings dynamics focus on men. This is at least partly because of the complexity in modeling both the decision to work (i.e., the extensive margin) and the level of earnings condition on work (the intensive margin). In this paper we document a series of descriptive facts about female earnings inequality, using data for three cohorts in the PSID. We then fit earnings-generating models that incorporate both intensive- and extensive-margin dynamics to data for early and later cohorts. We show that inequality in annual earnings of women fell sharply between the late 1960s and the mid-1990s, with over half of this fall attributable to declines in the extensive margin components of inequality. Our models suggest that about 70% of the overall decline can be attributed to the weakening of the link between family-based factors – including spousal earnings and the presence of children – and the intensive and extensive margins of earnings determination.

Acknowledgments: This research was supported in part by the Royal Society of New Zealand Marsden Fund Grant MEP1301. We thank Dave Maré for many helpful discussions and Thomas Xiao for excellent research assistance.

*Address: Department of Economics, 549 Evans Hall, #3880 University of California, Berkeley, Berkeley, CA 94720-3880 and NBER E-mail: card@econ.berkeley.edu

†Address: Motu Economic and Public Policy Research, PO Box 24390, Wellington 6142, New Zealand. E-mail: dean.hyslop@motu.org.nz.

1 Introduction

Although female employment rates have risen substantially in the U.S. and other developed countries since the 1960s, women's employment rates remain lower and more intermittent than those of men.[1] In part because of the resulting difficulties in dealing with movements in and out of work, there has been surprisingly little research documenting the trends in female earnings inequality or decomposing the factors underlying these trends.[2] What is clear is that inequality trends differ by gender, and that the relative selectivity of working and non-working females has changed, potentially contributing to the differences.[3] There is also increasing recognition that variation at the extensive margin is potentially important for understanding aggregate movements in earnings and hours (e.g. Chetty et al., 2011; Hansen, 1985; Heckman, 1993; Keane and Rogerson, 2012; Ljungqvist and Sargent, 2011; Peterman, 2016), as male employment rates have dropped and women and older workers have become a larger share of the workforce.

In this paper we develop techniques for jointly model the extensive and intensive margins of earnings variation, and apply these to modeling female earnings

[1] For example the fraction of adult women working in any week rose from 35% in 1950 to 57% in 2000 (BLS, 2014). Eckstein and Lifshitz (2011) show that this increase was largely driven by increases for married women and a rising fraction of non-married women (see their Figures 2 and 7).

[2] The literatures on earnings instability (e.g., Gottshalk and Moffitt, 1994; Shin and Solon, 2011) and earnings dynamics (e.g., Haider, 2011; Baker and Solon, 2003; Meghir and Pistaferri, 2004; Guvenen, 2009) mainly focuses on men. Female workers have received more attention in the wage inequality literature, but because the main focus is on relative productivity trends for different groups of workers, most authors analyze full time-full year workers (e.g., Katz and Autor, 1999; Eckstein et al., 2014), or focus on hourly pay (e.g., Fortin and Lemieux, 1998; Blau and Kahn, 2016).

[3] Gottschalk and Moffitt (2009) recognize the potential importance of the changing composition of female workers and analyze trends in unconditional earnings (including non-workers), finding less systematic growth in transitory earnings dispersion for females than males since 1970. Mulligan and Rubinstein (2008) analyze the effects of selective participation on changes in female wage inequality since 1970.

dynamics in the US using data from the Panel Study of Income Dynamics (PSID).[4] Central to the analysis of earnings dynamics is the concept of a statistical earnings generating function (EGF), which links current and future earnings outcomes to a combination of observed and unobserved factors.[5] To date, the EGF literature has focussed exclusively on intensive-margin variation in earnings among constantly-employed men (e.g. Abowd and Card, 1989; MaCurdy, 1982; Meghir and Pistaferri, 2004).[6] We extend the literature by combining a relatively standard EGF specification for the intensive margin of earnings with a dynamic binary response model for the extensive margin of employment (Hyslop, 1999).

We begin with a descriptive empirical analysis of the relative contributions of the extensive and intensive margins to the total variation in individuals' annual earnings. We develop a simple decomposition of the squared coefficient of variation (CV^2) of earnings in a panel data setting that partitions overall earnings inequality into a within-person component, a between-person component, and an interaction term (which in our setting is small). We then show how both the within-person and between-person components can be further decomposed into intensive-margin and extensive-margin components. Much of the existing literature on earnings dy-

[4] Altonji et al. (2013) estimated a dynamic model of hours and employment with a simplified specification of behavioural responses to current wage opportunities. They find that the behavioural responses are quite small on both the intensive and extensive margins. This suggests it may be relatively costless to ignore these responses and focus on the extensive (employment) and intensive (wages and hours) margins of earnings.

[5] For example, Friedman (1957) posited that earnings are generated by a combination of permanent and transitory factors, and used this idea to explain the different relationships between income and consumption at the microeconomic and the macroeconomic levels.

[6] Two recent papers present relatively sophisticated intertemporal choice models of consumption and labour supply in which all the variation in earnings arises at the extensive margin. Low et al. (2010) study males, and allow heterogeneity across education and age groups but not within these groups, and thus cannot address issues related to overall earnings inequality. Eckstein and Lifshitz (2011) study females, and focus on explaining cohort trends in labour force participation, without directly addressing earnings.

namics focuses on the within-person/intensive-margin component. For individuals who are always employed, this is (approximately) the within-person variance of earnings, which Gottschlak and Moffit (1994) have called "earnings instability". The complimentary extensive-margin component of within-person inequality is a simple function of the average probability of employment: its value is 0 for those who always work. Finally, the between-person intensive- and extensive-margin components depend on the variance of average earnings when employed and the average employment rate, respectively. In a population that always works, the latter just the variance in "permanent earnings", defined as person-specific average earnings over all years in the panel.

Applying this framework to three cohorts of women in the PSID, each observed over 10 consecutive years, we reach three main conclusions. First, overall inequality in female earnings has declined remarkably (with a 50% decline in CV^2 and a 30% decline in CV). Second, this fall is attributable to declines in both the within-person and between-person components. Indeed, the relative shares of these components have remained quite stable, with the between-person share of CV^2 remaining at about 80%. Third, a relatively large share of the decline in between-person inequality is attributable to reductions in the extensive-margin component, reflecting the rise in mean employment rates of the women in our sample, from 58% per year for the cohort observed between 1968 and 1977 to 80% for the cohort observed between 1968 and 1977.

In contrast to these trends among women, we show that for men in the same three cohorts earnings inequality ***increased*** significantly, driven by rises in all dimensions, but especially in the between-person/intensive-margin component, which for men (whose average employment rate is around 95% across all three

PSID cohorts) is equivalent to a rise in the variation in "permanent earnings".

We then develop an EGF that jointly summarizes the extensive and intensive margins of earnings variation. Cogan (1981) and Mroz (1987) conclude that simple Tobit model restrictions on the employment and hours dimensions of female labour supply are rejected, and that more general models are required to analyze both margins adequately. Our approach combines a standard specification for the evolution of individuals' latent earnings (Abowd and Card, 1989; MaCurdy, 1982; Meghir and Pistaferri, 2011), with a dynamic discrete choice model of employment (Hyslop, 1999), allowing a fairly general correlation structure between the underlying error components in the two models. We show that such a model is able to capture the main features of the earnings data for women in our three cohorts through a combination of a state dependence in individuals' employment decisions and both permanent and transitory components of latent earnings.

A feature of our modeling approach is that we can flexibly model the channels through which observable factors (including family-related variables) affect the choice of whether to work or not in a given year and earnings conditional on work. In particular, building on a standard correlated random effects approach, we include a set of "average" characteristics of a person's family, like the averge number of children she has at home over the years she is observed in our sample or the average earnings of her spouse/partner, as well as corresponding period-specific variables, like whether she has any children under the age of 5 in the current year, or the deviation of her spouse's income in the current year from its longer-term average.

Using our model to simulate the effects of "turning off" the effects of family-related factors in each cohort, we show that a driving force in the decline of earnings inequality is the systematic weakening of family-related forces. In our earliest cohort of women (observed in the late 1960s and early 1970s), family related factors explain nearly one-half of the overall inequality in female earnings outcomes, with especially large contributions to the between-person component of inequality. In our latest cohort, family factors matter far less. The diminishing influence of family-related variables accounts for nearly three-quarters of the overall decline in female earnings inequality. In this regard, the earnings determination models for women have become much closer to those for men, for whom family-related influences were always relatively limited.

The remainder of the paper is organized as follows. We begin by briefly summarizing the two major strands of work on earnings dynamics that focus on intensive margin variation among men and extensive margin variation among women. Next we outline the broad trends in female earnings inequality in the US, using data from the Current Population Survey (CPS) and the PSID. We then develop a simple measurement framework for decomposing the components of variation in earnings, and apply this to the three cohorts of women from the PSID. With this background we then specify our earnings generating model, present the estimation results, and summarize the model's ability to describe earnings outcomes. In the final section we use the model to ask how the changing role of family-related factors has led to changes in the overall variation in female earnings, and the various components of this variation.

2 Modeling intensive and extensive margin dynamics

We begin by briefly discuss the existing literature on earnings and employment dynamics. We use salient features from each of these literatures to develop a framework for individual earnings dynamics that combines both the extensive and intensive margins of adjustment.

At the micro-econometric level there has been relatively little research combining the extensive margin with the intensive margin of earnings variation. Two recent papers present relatively sophisticated intertemporal choice models of consumption and labour supply in which all the variation in earnings arises at the extensive margin. Low et al., 2010 focus on men, allowing heterogeneity between education and age groups but not within these groups. Their approach thus cannot address issues related to overall earnings inequality. Eckstein and Lifshitz, 2011 study women, focusing almost exclusively on inter-cohort trends in labour force participation, with no direct attention on earnings. A third recent study by Altonji et al. (2013) models both the intensive and extensive margins, using a "semi-structural" model of hours and employment with a simplified specification of behavioural responses to current wage opportunities. Interestingly, Altonji et al. (2013) find that behavioural responses on both the intensive and extensive margins are quite small – a conclusion that suggests it may be relatively costless to ignore these responses, as is implicitly done in much of the consumption and inequality literature.

Instead, most of the literature has focused almost exclusively on either the intensive margin of male earnings dynamics, or on the extensive margin of female employment dynamics. Seminal empirical analyses of the longitudinal structure of US male earnings by MaCurdy (1982) and Abowd and Card (1989) have resulted in the, now standard, characterisation of earnings as consisting of an individual-specific non-stationary (random walk) permanent component of earnings, a low-order stationary autoregressive moving-average (ARMA) transitory component, and a purely transitory component which is typically interpreted to represent (classical) measurement errors (see Meghir and Pistaferri, 2011). Such non-stationary representations of permanent components are preferred both conceptually, as they capture the PIH notion of adjustment to individuals' permanent income via shocks, and statistically, because the variance of individuals' earnings and income tend to increase over the life cycle, at least in the US and UK.

Similarly, the extensive margin literature on female employment dynamics has built on a series of papers by Heckman (e.g., Heckman, 1978, Heckman, 1981), that established the now-standard approach to modeling employment dynamics. This typically involves specifying a dynamic binary response model that includes controls for observable covariates, state dependence in employment via a first-order Markov process, and persistent and transitory unobserved factors that affect employment. Typically in this literature researchers find that family-based influences, including the presence of young children and spousal earnings, exert some influence on extensive margin choices of women. We build on this in our model, but also allow these family-based factors to affect the determinants of earnings conditional on work.

3 Setting and Descriptive Overview

Next we turn to the salient empirical "facts" that motivate our analysis. We first provide some background information documenting trends in female earnings inequality in the US using data from the Current Population Survey. We then describe our main PSID samples, which consist of 3 panels of women who were continuously observed over three consecutive 10-year intervals: 1968-77, 1978-87, and 1988-97. Since the panels are relatively small we spend some time documenting that the main features of earnings we observe for all women are represented in the three PSID panels.

3.1 Trends in female earnings inequality

We begin by summarizing the main trends in female employment and earnings since the late 1960s. Figure 1 shows mean annual earnings (in real 2013 $) as reported in March Current Population Survey (CPS) for women age 25-60. We show both unconditional mean earnings (i.e., over workers and non-workers, assigning 0 to nonworkers) and conditional mean earnings (excluding the 0's), as well as the fraction who worked in the year and the standard deviation of conditional earnings. The figure confirms three well known "facts". First, mean real earnings of women in the U.S. have risen substantially over the past 5 decades. Second, part of the upward trend until year 2000 was a rising probability of work. Over the 20 years from 1968 to 1988 the growth in employment rates was particularly impressive, averaging about 1 percentage point per year. Thereafter the trend stalls, with a notable decline from the peak rate of around 78% in 2000 to 73% in the 2014 survey. Third, among those who work, the standard deviation of earnings also rose, at about the same rate as the conditional mean. Thus, the coefficient of variation of conditional earnings(CV^c) was roughly stable.

This is confirmed in Figure 2, where we plot three measures of earnings inequality: (1) the conditional coefficient of variation,CV^c; (2) the standard deviation of log(earnings) for those with positive earnings, which is a widely used metric of pay inequality among workers; and (3) the coefficient of variation for the full sample, including those with zero earnings, CV. We also show for reference the employment rate. The standard deviation of log(earnings) has gradually trended down during our sample period, from a value of 1.05 in the 1968 survey to about 0.95 in the 2014 survey. In contrast, CV^c was stable at around 0.8 from 1968 to 1993, rose slightly in the late 1990s, and has subsequently hovered around 0.85. Most remarkable, however is the trend in the coefficient of variation of unconditional earnings, which has fallen substantially. The downward trend was particularly strong from the late 1960s to the early 1990s, coincident with the era of rising employment rates.[7]

An important feature of all four series shown in Figure 2 is that most of the changes over the past 5 decades occurred during the 30 year period from the late 1960s to to the late 1990s. Fortunately, this coincides with the period during which the Panel Study of Income Dynamics (PSID) collected annual information on households and individuals. To take advantage of the rich information in the PSID without building a complex model that can accommodate the switch to an "every second year" interview schedule after 1997, we therefore limit our attention to this three-decade period.

[7]The simple correlation between CV and the employment rate is -0.98.

3.2 PSID Samples

From 1968 to 1997 the PSID conducted an annual survey of 5,000 or so families in the first half of the year that inquired about income and work during the previous calendar year (similar to the Annual Demographic Supplement to the March CPS). In selecting samples from the PSID for our analysis, we include individuals from both the nationally representative subsample of the PSID and the poverty subsample drawn from the Survey of Economic Opportunity (SEO) which made up about one-third of the original PSID sample.[8]

We have drawn separate panel samples of females and males from the annual survey over three non-overlapping 10-year periods (1968–77, 1978–87, and 1988–97). We require sample members to be between the ages of 25 and 60 in every year of the panel: thus each panel is (broadly) representative of "working age" adults during the particular years of the panel. Our female samples consists of women who were either household "heads" (if single) or "wives" (i.e., the married or unmarried female partner in a dual-headed household) for a 10 year period covering one of our three intervals.[9] Similarly, our male samples consist of men who were "heads" (single household heads or the married or unmarried male partner in a dual-headed household) for a 10 year period covering one of our three intervals.

Table 1 presents summary statistics on the demographic characteristics and labour market outcomes of females in each of our three panels.[10] We have between

[8]We have excluded the Latino sample, which was part of the PSID from 1990 until 1995, partly because the 1994 and 1995 surveys did not collect employment and earnings for this sample. In any case, individuals in the Latino sample would not qualify for the 10-year panel sample.

[9]The nomenclature used in the early waves of the PSID is anachronistic.

[10]Corresponding summary statistics for these male samples are presented in Appendix Table 1.

2,000 and 2,500 observations per panel. These modest sample sizes reflect the relatively high rates of family transition in the PSID that lead to substantial attrition out of "head" or "wife" status, combined with the age restrictions which limit our sample to women who are 25-50 at the start of each panel. We emphasize that we do not necassarily lose a woman who is initially a "wife" then becomes a single female "head" when her partner leaves the household. However, women who enter the sample as a partner of an original male "head", then split from that male, are not followed. The average age of the women is about 40 in each panel; average education is about 11 years for the first panel, rising to just over 13 years for the third panel; and the average number of children is 2.2 in the first panel, falling to 1.4 in the second and third panels. One somewhat unusual feature of our samples is the high fraction of African Americans (37% in the first panel, slightly lower in the second and third). This reflects inclusion of the SEO subsample.

Our measure of earnings is annual labor earnings, which includes wages and salaries, farm income, and self-employment income. In order to reduce the impacts of outliers, we have censored the top and bottom 5% of earnings, separately for males and females, and retained the censored observations.[11] The PSID also provides a measure of the annual hours worked which, together with annual earnings, can be used to derive an hourly wage. Although our primary focus is on annual earnings and employment, rather than hours worked, for consistency we define a person as "employed" if they have both positive annual earnings and positive hours worked. We reset earnings to 0 for any individuals with 0 hours but positive

[11]Between 7–8% of individuals have censored earnings at some stage over the 10-years, of which 4% of males and 5% of females are bottom censored, and 3.5% of males and 3% of females are top censored. Furthermore, about 80% of those with bottom censored earnings, and 45–50% of those with top censored earnings, are censored in a single year, which suggests the effect on within person earnings variation will be limited, particularly among lower earners.

earnings.

Average annual earnings (measured in 2013 dollars) rise from 19,000 for women in our first panel to 32,000 for those in our third panel. Some of the rise is driven by an increase in average employment rates (from 58% to 80% between the first and third panels), some is due to an increase in hours conditional on working (from an average 1,380 hours per year for the first panel to 1,660 hours per year for the third panel), and some is due to a rise in real hourly wages (from \$14.39 per hour for the first panel to \$20.29 for the third).

We also, in the bottom panel of the Table, show some mean characteristics of spouses for the 70-74 percent of women who are "married" (i.e., living in a dual-headed household). On average spouses are in their early 40's, have about a 95% chance of working, and work about 2,200 hours per year if they are employed. Spouses' annual earnings rise from an average of about 53,000 to 65,000 between the first and third panels, mainly reflecting a rise in real hourly wages.

3.3 Comparisons of PSID Samples with CPS

Given the small and somewhat unrepresentative nature of our PSID panels, an important question is whether the trends in earnings outcomes within in panel and betweeen the three panels broadly match the trends for females as a whole. Figures 3a-3d overlay estimates of four key earnings outcomes based on annual observations from each of our panels against the corresponding estimates derived from the CPS. Specifically, we examine the employment rate, mean earnings conditional on work, the unconditional CV of earnings, and the standard deviation of log earnings. Appendix Figures 1-3 similarly show the age, education, and racial compositions of the women in our three panels, compared to the characteristics of the CPS samples in the same years. Our assessment is that the trends in employment, earnings, and the dispersion in earnings are broadly similar across the two data sources, with the PSID panels capturing three main "facts": a rising employment rate; a falling coefficient of variation of overall earnings, and a relatively flat standard deviation of earnings conditional on working. Interestingly, the PSID comparisons suggest that a sizeable part of the increase in employment and reduction in the CV is attributable to cohort effects. For example, about 15 percentage points of the overall 20 percentage point increase in average female employment rates from 1968 to 1997 occurs at the between-cohort jumps. Similarly, 0.26 of the roughly 0.40 decline in the CV of unconditional earnings over the sample period occurs at the between-cohort jumps.

While our primary focus in this paper is on understanding **overall** earnings inequality and the changing contributions of intensive- and extensive-margin components, an important strand of recent research focuses more narrowly on the variability of within-person changes in earnings. This literature originated with Gottschalk and Moffit (1994), who argued that the variance of the transitory component of male earnings in the PSID had increased in the 1980s. Shin and Solon (2007) noted that this variance component is closely related to the cross-sectional variance of individual earnings changes, which is much easier to calculate. Many subsequent papers (e.g., Sabelhous and Song, 2010; deBacker et al. 2013) have used a variety of different data sources and samples to construct three main measures: (1) the standard deviation of the "arc-percentage" change in earnings between consecutive years;[12] (2) the standard deviation of the "arc-percentage" change in

[12] Letting y_{it} denote earnings of a given person i in year t, the arc-percent change is: $2(y_{it} -$

earnings for those who work in both years; and (3) the standard deviation of the change in log earnings for those who work in both years. We show these three measures in Figure 4 for our PSID samples. Consistent with the findings in recent papers that report data for female workers (e.g., Dahl and Schwabish, 2008, Figure A-2; Ziliak et al., 2011; Celik et al, 2012) we see a modest decline in the variability among those who work in both years. The decline in the measure that includes 0-earners in either the initial or later period is even larger, reflecting the decreasing importance of these extensive margin adjustments.

4 Quantifying the components of earnings inequality

In this section we present a simple statistical framework for quantifying the components of earnings inequality in a panel data setting with both intensive and extensive margin variation. Perhaps the most commonly used measure of earnings inequality is the variance of log earnings (e.g. Katz and Autor, 1999; Haider, 2001). However, in order to handle zero earnings associated with periods of non-employment in our inequality analysis, we use the squared coefficient of variation of earnings as our primary measure of inequality, which provides a first-order approximation to the variance of log earnings.[13]

We proceed by, first, deriving a decomposition of the aggregate earnings inequality over time into between-person and within-person contributions. We then decompose both the within-person and between-person components into sub-components

$y_{it-1})(y_{it} + y_{it-1})$. The variance of this is numerically equal to the mean squared coefficient of variation of within-person earnings changes – see below.

[13]See Levy and Murnane (1992) for a discussion of alternative measures of inequality.

attributable to intensive and extensive margin variation (i.e., differences in the probability of work versus differences in the amount of earnings conditional on working).

4.1 Within and between person contributions to inequality

Let y_{it} represesent the earnings of person i ($i = 1, ..., N$) observed in period t ($t = 1, ..., T$), and let $\overline{y}_i$ represent person i's mean earnings.[14] We define the within-person squared coefficient of variation, which measures the average percentage variation in their earnings over time, as

$$CV_i^2 = \frac{1}{(\overline{y}_i)^2} \frac{1}{T} \sum_t (y_{it} - \overline{y}_i)^2 = \frac{1}{T} \sum_t \left(\frac{y_{it} - \overline{y}_i}{\overline{y}_i} \right)^2 . \quad (1)$$

Notice that if there are only two observations per person this is the squared arc-percent change in earnings of person i. Let $\overline{y}_t$ represent mean earnings of all individuals in period t (including workers and nonworkers), and let $\overline{y}$ represent the grand mean of earnings across all years. In year t, the squared coefficient of variation in earnings (CV_t^2) is

$$CV_t^2 = \left(\frac{1}{\overline{y}_t} \right)^2 \frac{1}{N} \sum_i (y_{it} - \overline{y}_t)^2 = \frac{1}{N} \sum_i \left(\frac{y_{it} - \overline{y}_t}{\overline{y}_t} \right)^2 . \quad (2)$$

[14] For simplicity and consistency with our samples, we assume a balanced sample design with T constant across people.

while the average CV across all years in the sample is:

$$
\begin{aligned}
CV^2 =& \frac{1}{T}\sum_t CV_t^2 = \frac{1}{NT}\sum_t\sum_i\left(\frac{y_{it}-\overline{y}_t}{\overline{y}_t}\right)^2 \\
=& \frac{1}{NT}\sum_t\sum_i\left(\frac{\overline{y}_i}{\overline{y}_t}\right)^2\left(\frac{y_{it}-\overline{y}_i}{\overline{y}_i}\right)^2+\frac{1}{NT}\sum_t\sum_i\left(\frac{\overline{y}_i-\overline{y}_t}{\overline{y}_t}\right)^2 \\
&+\frac{2}{NT}\sum_t\left(\frac{1}{\overline{y}_t}\right)^2\sum_i(y_{it}-\overline{y}_i)(\overline{y}_i-\overline{y}_t).
\end{aligned} \tag{3}
$$

The first term in equation (3) is the "within-person" contribution to CV^2, associated with year-to-year fluctuations in individual earnings around the person-specific mean:

$$
\begin{aligned}
W &\equiv \frac{1}{NT}\sum_t\sum_i\left(\frac{\overline{y}_i}{\overline{y}_t}\right)^2\left(\frac{y_{it}-\overline{y}_i}{\overline{y}_i}\right)^2 \\
&= \frac{1}{N}\sum_i\left(\frac{\overline{y}_i}{\overline{y}}\right)^2\left\{\frac{1}{T}\sum_t\left(\frac{\overline{y}}{\overline{y}_t}\right)^2\left(\frac{y_{it}-\overline{y}_i}{\overline{y}_i}\right)^2\right\} \\
&\approx \frac{1}{N}\sum_i\left(\frac{\overline{y}_i}{\overline{y}}\right)^2 CV_i^2 = W^*
\end{aligned} \tag{4}
$$

Assuming that $\overline{y}_t$ is relatively stable over time (i.e., $\overline{y}_t \approx \overline{y}$), $W \approx W^*$and the within-person component can be expressed as a weighted sum of CV_i^2 terms, where the weight for person i is the squared ratio of their average earnings to overall average earnings.

The second term in equation (3) represents the "between-person" contribution to CV^2 that arises from the variation in average earnings ($\overline{y}_i$) across people:

$$
B = \frac{1}{NT}\sum_t\sum_i\left(\frac{\overline{y}_i-\overline{y}_t}{\overline{y}_t}\right)^2 = B^* + \frac{1}{NT}\sum_t\sum\left(\frac{\overline{y}_t-\overline{y}}{\overline{y}_t}\right)^2 \tag{5}
$$

where

$$B^* = \frac{1}{NT}\sum_i \sum_t \left(\frac{\overline{y}_i - \overline{y}}{\overline{y}}\right)^2 \times \left(\frac{\overline{y}}{\overline{y}_t}\right)^2$$

Note that when $\overline{y}_t \approx \overline{y}$, $B \approx B^*$. In this case B is approximately the squared proportional deviation between $\overline{y}_i$ and $\overline{y}$, and provides a simple measure of the dispersion in "permanent earnings" across individuals in the sample.

The third, cross-product, term in equation (3) is:

$$C = \frac{2}{NT}\sum_t \left(\frac{1}{\overline{y}_t}\right)^2 \sum_{i \in t} (y_{it} - \overline{y}_i)(\overline{y}_i - \overline{y}_t)\,.$$

If individual earnings are procyclical, then years when $y_{it} - \overline{y}_i$ is relatively high will be years with relatively high values for $\overline{y}_t$, and C will be negative. As we show in our analysis below, C is in fact always negative but is relatively small for all groups except the earliest cohort of women.

4.2 Intensive and extensive margin contributions

Next, we consider how the within- and between-person components, W and B, can be decomposed into subcomponents representing intensive and extensive margin variation.

Within-person Let p_i represent the average employment rate of i (i.e., the fraction of years in which they have positive earnings), let $\overline{y}_i^c$ represent their mean earnings conditional on employment, and note that $\overline{y}_i = p_i \overline{y}_i^c$. As shown in the Appendix, a few simple substitutions establish that, for individuals with at least one year of positive earnings (i.e. $p_i > 0$):

$$CV_i^2 = \frac{1}{T}\sum_t \left(\frac{y_{it} - \overline{y}_i}{\overline{y}_i}\right)^2 = \frac{1}{p_i}\left(CV_i^c\right)^2 + \frac{1-p_i}{p_i} \quad , \tag{6}$$

where $(CV_i^c)^2$, the *conditional* squared coefficient of variation:

$$(CV_i^c)^2 = \frac{1}{T_i^+}\sum_{y>0}\left(\frac{y_{it} - \overline{y}_i^c}{\overline{y}_i^c}\right)^2$$

and $T_i^+ = p_i T$ is the number of years of positive work by i. (For people who never work, $CV_i^2 = 0$). Equation (6) decomposes CV_i^2 into an intensive margin component that depends on CV_i^c and an extensive-margin term is a simple decreasing function of the individual's average employment rate. For people who always work (i.e. with $p_i = 1$) the extensive margin component in (6) is 0 and $CV_i = CV_i^c$, which is approximately the within-person variance in log earnings:

$$\frac{1}{T}\sum_t \left(\frac{y_{it} - \overline{y}_i}{\overline{y}_i}\right)^2 \approx \frac{1}{T}\sum_t \left(\ln y_{it} - \overline{\ln y_{it}}\right)^2 .$$

Using the approximation for W in equation (4) above, we can write:

$$W \;\approx\; W^* = \frac{1}{N}\sum_i \left(\frac{\overline{y}_i}{\overline{y}}\right)^2 CV_i^2 = W^{int} + W^{ext} \tag{7}$$

where

$$W^{int} = \frac{1}{N}\sum_i \left(\frac{\overline{y}_i}{\overline{y}}\right)^2 \frac{1}{p_i}\left(CV_i^c\right)^2,$$
$$W^{ext} = \frac{1}{N}\sum_i \left(\frac{\overline{y}_i}{\overline{y}}\right)^2 \frac{1-p_i}{p_i}.$$

These are just weighed sums of the person-specific intensive and extensive margin components of CV_i^2.

Between-person Using another simple substitution we can we can decompose B^*, the main term in the between-person component of overall inequality, as:

$$B^* = B^{ext} + B^{int} + B^{cross} \tag{8}$$

where

$$B^{ext} = \frac{1}{NT}\sum_i\sum_t \left(\frac{\overline{y}}{\overline{y}_t}\right)^2 \left(\frac{\overline{p}_i - \overline{p}}{\overline{p}}\right)^2 \left(\frac{\overline{y}_i^c}{\overline{y}^c}\right)^2$$

$$B^{int} = \frac{1}{NT}\sum_i\sum_t \left(\frac{\overline{y}}{\overline{y}_t}\right)^2 \left(\frac{\overline{y}_i^c - \overline{y}^c}{\overline{y}^c}\right)^2$$

$$B^{cross} = \frac{2}{NT}\sum_i\sum_t \left(\frac{\overline{y}}{\overline{y}_t}\right)^2 \left(\frac{\overline{y}_i^c}{\overline{y}^c}\right)\left(\frac{\overline{y}_i^c - \overline{y}^c}{\overline{y}^c}\right)\left(\frac{\overline{p}_i - \overline{p}}{\overline{p}}\right).$$

Assuming that the gap between B and $B^*$is small (as it is in our samples), these three terms summarize the extension-margin component, the intensive margin component, and a joint "covariance" component of the overall between-person component of earnings inequality.

4.3 Measuring the changes in intensive and extensive margin contributions

With this background we turn to the estimates in Table 2, which quantify the various components of overall earnings inequality for women in our three 10-year panels. For comparative purposes we also show the same components for males in three parallel panels.

The top 5 rows of Table 2 show the overall value of CV^2 for each panel and the terms in equation (3) . We also show the approximation in equation (4) for the within-person component of inequality. Comparisons of the various terms for our three cohorts, and between women and men, lead to three main conclusions. First, total earnings inequality measured by the squared coefficient of variation is substantially greater for women than men. However, the respective levels converge somewhat over time, with male inequality rising steadily and female inequality falling dramatically. Second, for women all three components of equation (3) fall between the 1968-77 cohort and the 1988-97 cohort. In contast, for men the two main components representing within-person and between-person components rise, while the "cross term" is effectively zero in all cohorts. For both gender groups the between-person component accounts for a relatively constant $\approx 80\%$ share of the overall variation in annual earnings, while the within-person component accounts for a $\approx 20\%$ share. The cross-product term is always negative, but generally small, consistent with the notion of procyclical earnings.

A third useful finding is that the the approximation in equation (4) is quite good for all cohorts and genders except the 1968-77 cohort of females. Women in this cohort experienced relatively large increases in average earnings during the

10-year period of our panel, leading to a slight departure between W and W^*.

The rows in the second panel of Table 2 present the extensive and intensive margin components of W^* as specified by equation (7). We see that for women, both the extensive-margin and intensive-margin components of within-person inequality fell between the three cohorts, with a proportionally larger reduction in the extensive margin component. Consequently, the share of overall inequality attributable to within-person/extensive margin variation fell from 9% in the 1968-77 cohort to 6% in the 1988-97 cohort. In contrast, for men, the extensive margin component is very small across all three cohorts, contributing only 2-4% of overall earnings inequality. For men, however, we do see a notable rise in the within-person/intensive margin component, implying that earnings instability rose across these these cohorts of men.

The rows in the bottom panel of Table 2 present the three components of B^* specified by equation (8), as well as the deviation $B - B^*$ attributable to variation in $\overline{y}_t$ within each panel (see equation 5), which is relatively small. Here the contrasts between women and men are even larger. For women, we see that the between-person/extensive margin component of inequality fell in magnitude from 0.40 to 0.08. The decline in this single component accounts for 37% of the overall decline in CV^2 between the earliest and latest cohort. For men, the between-person/extensive margin component of inequality is very small and stable in size, accounting for 2% of overall earnings inequality in each panel.

For women we also see a relatively large decline between cohorts in B^{cross}, which depends on the covariance between $\overline{y}_i^c - \overline{y}^c$ and $\overline{p}_i - \overline{p}$. Arguably, we can interpret the declining magnitude of this term at least in part to the decline in extensive margin variation. For example, if we attributed the covariance term equally to the intensive and extensive margins, we could conclude that the between-person/extensive margin component of inequality accounts for about one-half of the overall decline in CV^2 between the 1968-77 and 1988-97 cohorts of women.

Finally, we also see a decline in the between-person/intensive margin component B^{int} for women between the three cohorts. Relative to the other components of CV^2,however, the proportional change in B^{int} was smaller. Consequently the share of overall earnings inequality among women attributed to differences in mean earnings conditional on working rose from 33% to 50%. Among men in our three cohorts B^{int} rose in magnitude, but at a slower rate than the within-person/intensive margin component, so the share of overall earnings variation attributable to B^{int} fell slightly, from 74% to 68%.

The relative size of the five main components of CV^2 for the 1968-77 and 1988-97 panels of women are summarized in Figure 5. The figure documents our two main conclusions from this descriptive exercise. First, overall female earnings inequality as summarized by the squared coefficient of variation fell by nearly 50% between these cohorts. By way of a benchmark, for comparable cohorts of men the same measure of earnings inequality rose by 40%. Second, a relatively large share of the decline in earnings inequality was due to declining magnitudes of the within-person and between-person extensive margin components.

5 Modeling earnings dynamics

5.1 A model for the extensive and intensive margins of earnings

We now turn to our model of earnings dynamics. We build on two existing literatures: one specifying dynamic models of participation for women; the other specifying earnings generating functions for men.

On the extensive margin, we specify a fairly standard dynamic panel data model for the event that individual i is observed working in year t ($E_{it} = 1$) that includes a first-order dynamic binary response model for $t = 2...T$ and a reduced-form specification for the initial condition in $t = 1$:

$$\begin{aligned} &E_{i1} = 1(X'_{ei1}\beta_{e0} + \epsilon_{ei1} > 0), \\ &E_{it} = 1(\gamma E_{it-1} + X'_{eit}\beta_e + \epsilon_{eit} > 0), t = 2, ..., T. \end{aligned} \tag{9}$$

Here the vector X_{eit} includes a set of characteristics that we take as exogenous deteminants of the probability of employment in each year. We assume that the latent error components are all normally distributed so the employment model amounts to a dynamic probit model.

On the intensive margin, we specify a standard log-linear specification for latent earnings (Y^*_{it}), which are realised as observed earnings (Y_{it}) conditional on employment:

$$\begin{aligned} &Y^*_{it} = X'_{yit}\beta_y + \epsilon_{yit}, t = 1, ..., T; \\ &Y_{it} = E_{it}.Y^*_{it}. \end{aligned} \tag{10}$$

Here the vector X_{yit} includes a set of characteristics that we take as exogenous deteminants of latent earnings in each year.

Following the literature on EGF's for males (e.g. Abowd and Card, 1989; Meghir and Pistaferri, 2004), we assume that ϵ_{yit} consist of a permanent component (α_{yit}), which is person-specific random walk, and a stationary MA(1) component (u_{yit}). Recognizing that we observe individuals at different stages of their life cycles, we allow the variance in α_{yi1}, the initial period permanent component of latent earnings, to vary linearly with age. We adopt a parallel specification for the errors terms in the employment model: specifically we assume that ϵ_{eit} consists of a person-specific random walk component and a stational MA(1) component, and allow the variance in α_{ei1}, the initial period permanent component of the employment determination model, to vary linearly with age.

To account for potential non-random selectivity of employment and/or to allow for possibly correlated shocks to employment and earnings, we also allow each of the respective error components of the employment and earnings equations to be contemporaneously correlated. In particular, the various error components are specified as folllows:

$$\begin{aligned}
&\epsilon_{jit} = \alpha_{jit} + u_{jit},\ j = e, y; \\
&\alpha_{jit} = \alpha_{jit-1} + \eta_{jit} = \alpha_{ji1} + \Sigma_{s=2}^{t}\eta_{jis}; \\
&\alpha_{ji1} \sim N(0, \sigma^2_{\alpha j1}),\ \sigma^2_{\alpha j1} = \sigma^2_{\alpha j0} + \sigma^2_{\alpha j} * (age_{i1} - 24); \\
&\eta_{jis} \sim N(0, \sigma^2_{\eta j}); \\
&u_{jit} = \theta_j \omega_{jit-1} + \omega_{jit}; \\
&\omega_{jit} \sim N(0, \sigma^2_{\omega j}); \\
&\rho_\alpha = corr(\alpha_{ei1}, \alpha_{yi1}), \rho_\eta = corr(\eta_{eit}, \eta_{yit}), \rho_\omega = corr(\omega_{eit}, \omega_{yit}).
\end{aligned} \tag{11}$$

In addition, we specify the employment initial conditions ($t = 1$) equation error to consist of the employment initial permanent component with a factor loading (δ_{e1}), and the transitory error component from the subsequent MA(1) process:

$$\epsilon_{ei1} = \delta_{e1}\alpha_{ei1} + u_{ei1}.$$

For identification, we normalize the employment equation's initial-period total error variance to 1. Specifically, we set $\sigma^2_{\alpha e1} + \sigma^2_{ue} = 1$, so that $\sigma^2_{ue} = (1 + \theta^2_e)\sigma^2_{\omega e} = 1 - \sigma^2_{\alpha e1}$.[15]

Finally, because the log(earnings) distribution of females shows evidence of bi-modality (reflecting the presence of both full-time and part-time workers), in the estimation below we relax the assumption of normality and allow the earnings equation errors to be a bivariate mixture of normals.[16] In this mixture specifi-

[15]Note that this normalization is applied to the (t=1) (latent) structural error term, $\alpha_{ei1}+u_{ei1}$: these components are part of the permanent random walk and transitory MA(1) components respectively.

[16]An alternative would be to attempt to classify workers in each year as part-time or full-

cation, we adopt a relatively parsimonious formulation for the error components, allowing the means of the two underlying normal distributions to differ, and restricting the variances of the corresponding error components in these distributions to vary by a constant scale parameter. That is, we specify ϵ_{yit}:

$$\epsilon_{yit} = (1-\mu)\epsilon^{0}_{yit} + \mu\epsilon^{1}_{yit}, \tag{12}$$

where $\epsilon^{j}_{yit} = \alpha^{j}_{yit} + u^{j}_{yit}$ (j=0,1), as in equation (11); $\epsilon^{0}_{yit} \sim N(0, \sigma^{2}_{\epsilon y0})$; and $\epsilon^{1}_{yit} \sim N(\mu_{y1}, \delta_{y1}\sigma^{2}_{\epsilon y0})$, and the variance scale δ_{y1} is restricted to be the same on both the permanent (α_{yit}) and transitory (u_{yit}) error components. This adds three additional parameters to the model (μ,μ_{y1},δ_{y1}).

5.2 Specification of controls

A final specification issue is the choice of control variables to include in the pair of vectors (X_{eit},X_{yit}). We include education, a dummy for black race, and a quadratic function of age in both vectors. We also include a relatively rich set of family variables that are designed to characterize both the "long run" family characteristics for a given woman, and the current period characteristics. Specifically, both vectors include:

- a dummy for living with a partner ("married") in the current year, and the mean of this dummy over all 10 years
- a dummy for whether spouse is currently employed, the mean of this dummy over all years that a partner is present, and a dummy for having a spouse who is never employed

time and extend the discrete choice model for employment to model the choice between these alternatives, possibly allowing for part-year employment as well.

- the spouse's mean log earnings (over all years he is present) and the deviation of mean log earnings in the current year from this mean

The employment probability model also includes 6 child-related variables: the total number of children living with the woman and the average number of children present over the 10 years of the panel; a dummy for having a child under age 5 in the current year and the average number of children under age 5; and a dummy for having a child age 6-17 in the current year and the average number of such children. In addition, the employment model includes the aggregate unemployment rate, while the earnings model includes the log of median earnings in the economy and interactions of this variable with dummies for women with exactly 12 years of schooling, 13-15 years of schooling, and 16+ years of schooling. These variables are meant to capture economy-wide wage trends that may differ by education group. (We do not include year dummies so such trends are potentially important).

6 Results

6.1 Estimation results

The model is estimated using maximum simulated likelihood (MSL) estimation, with 20 simulation replications. Table 3 presents estimates from two specifications of the model for each of our 3 PSID panels. The first specification for each cohort (shown in the odd-numbered columns) adopts the usual normality assumption for the log(earnings) equation error. The second specification, which we use as our baseline model, allows the earnings equation errors to be a bivariate mixture of normals, as in equation (12. The first panel of the table presents the estimated coefficients (β_e) and error components from the model for employment in years 2-

10; the second panel shows the estimated coefficients (β_y) from the earnings models; the third panel shows the estimated variance components from the earnings models and the correlations between the error components in the employmenr and earnings models; and the fourth panel shows the estimated coefficients (β_{e0}) from the model for employment in the first year (i.e., the initial condition model).

Error structure Focusing first on the dynamic error structure of the model, we find large positive effects of lagged employment status on the probability of employment (i.e., the coefficient γ in equation (9), with slightly larger state dependence in the earliest panel (1868-77) than the two later panels. The magnitude of $\widehat{\gamma}$ for the first cohort suggests that other things equal, women who worked last year are about 70 percentage points more likely to work in the current year; for the later cohorts the impact is about 60 percentage points.

We also find statistically significantly permanent *and* transitory error components in both the employment and earnings equations, with a positively correlated MA1 transitory error in earnings ($\theta_y \approx 0.25$) but a negatively correlated MA1 transitory error in the latent determinants of employment ($\theta_e \approx -0.5$).[17] Shocks to the permanent components in the two equations are positively correlated, particularly for the first year observation (i.e., the initial condition). The estimates of $corr(\alpha_{ei1}, \alpha_{yi1})$ range from 0.55 to 0.82 depending on the cohort and specification, whereas $corr(\alpha_{ei}, \alpha_{yi})$ ranges from 0.12 to 0.39. Thus, unobserved factors that cause a woman to be more likely in a given year also increase her average earnings if she works in that year. The implication of this positive correlation is

[17] Although the negatively correlated MA1 errors components in employment are unusual, similar results have been found in the literature on female employment dynamics in models that include state dependence, and permanent and transitory errors (e.g. Hyslop, 1999)

that workers are a positively selected subset of the overall population.

In contrast, the shocks to the transitory components of latent employment and earnings are *negatively* correlated. Finally, our estimates imply that the variances of the inital permanent components of employment and earnings are both increasing with age (i.e., the estimates of the trend coefficients in the initial variances, $\sigma^2_{\alpha e}$ and $\sigma^2_{\alpha y}$, are positive) as is predicted by human capital models with unobserved investments (e.g., Mincer, 1974).

To help understand the dynamic contributions of the various permanent and transitory error components to the employment and log(earnings) dimensions, we conducted a series of simulations of the effects of shocks to each of the components of the error structure driving employment and log(earnings)[18] Given that the employment and earnings error components are correlated, a primitive shock to one dimension will generate a correlated shock in the other dimension that leads to persistent "cross effects". Appendix Figures 4-7 show the impacts of a one standard deviation shock to each of the 3 error components of employment and earnings (i.e., the initial permanent error components α_{ei1} and α_{yi1}; the permanent error component in year 2, α_{ei2} and α_{yi2}; and the transitory error components in year 1, ω_{ei1} and ω_{yi1}) on subsequent employment and earnings. In each case we show the response functions from the models for the three cohorts, allowing us to assess whether reactions to unobserved determinants of employment and earnings have changed between cohorts.

[18]In particular, we simulate the baseline model specification (2) using random draws for each component, and then add +/-0.5 standard deviation shocks for each component in turn, and estimate the average difference in the employment and earnings outcomes over time. The shocks were 1 standard deviation shocks to the age-adjusted initial permanent stock components; 1 standard deviation shocks to the period-2 random walk innovations; and 1 standard deviation shocks to the period-1 transitory MA innovations.

We find that overall shapes of the response functions are fairly similar for the three PSID cohorts. In particular, shocks to the error components in latent employment have very similar dynamic effects on future employment (see Appendix Figure 4), suggesting that the extensive margin responses to *unobserved* determinants of employment have not changed much across the cohorts. The "cross-effects" of shocks to employment on future earnings do differ slightly more between cohorts (Appendix Figure 5). We also see the effects of shocks to the permanent and transitory error components of earnings have very similar effects on future earnings (Appendix Figure 6), suggesting that the intensive margin responses to *unobserved* determinants of earnings are similar across cohorts, though again the "cross effects" of shocks to earnings on future employment differ slightly more between cohorts (Appendix Figure 7).

Effects of observed factors - child variables and spousal variables Next we briefly summarize the effects of the observed child- and spouse-related variables included in our models. Focusing first on the child-related variables, we see a general tendency for the coefficients of the presence of young children to become smaller for later cohorts. For example, in the models for the probability of employment in the first panel of Table 3, a dummy for the presence of a child under 5 has a coefficient of -0.26 in model 2 for the 1968-77 cohort and a coefficient of -0.13 in the parallel specification for the 1988-97 cohort. Translating these coefficients into marginal effects using the normal density for an individual with the sample average probability of work in each cohort, the implied effects are -10.4 percentage points for the 1968-77 cohort, and -4.3 percentage points for the 1988-97 cohort.[19]

[19] The conversion factors are about 0.40× the coefficients for the 1968-77 panel, 0.33× the coefficients for the 1978-87 panel, and 0.29× the coefficients for the 1988-97 model..

We see a similar dimunition in the effect of young children in the intial conditions model for employment in the first year of each panel, from an implied average marginal effect of -27.6 percentage points for the first cohort to -13.6 percentage points for the last cohort. The coefficients associated with the presence of a child age 6-17 in the model for the initial condition employment are also relatively large, but fall in magnitude across the cohorts: from an implied marginal effect of about -13 percentage points for the first cohort to essentially 0 for the last cohort.

The effects of the spousal-related variables are harder to assess because the presence of a spouse "turns on" a whole set of variables simultaneously, including the "marriage" dummy, the average employment probability of the spouse, the average log earnings of the spouse, and the dummy for whether the spouse is employed (which it equal to 1 95% of the time). Nevertheless, these effects also on average fall in magnitude. For example, the implied marginal effect on the probability of employment of moving from unmarried to married status with an "average" spouse is about -35 percentage points for women in the 1968-77 cohort, but -27 percentage points for those in the 1988-97 cohort.[20]

6.2 Model fit

How well can our relatively simple model capture the overall degree of earnings inequality in the three PSID cohorts, and the contributions of the extensive and intensive margin components to this overall inequality? Table 4 presents a comparison between the actual and fitted components of inequality, using the same

[20]The implied marginal effects of moving to married status with a spouse whose mean log earnings are only 50% of the average, and has a 85% average employment rate (rather than a 95% rate) are smaller: a -29 percentage point effect for our earliest cohort and a -21 perentage point effect for our latest cohort.

decomposition strategy as in Table 2, and focusing on our second and richer model which assumes that earnings are mixture of normals. We also show, in the two right-most columns of the table, the changes in the actual and predicted components of variance between the 1968-77 cohort and the 1988-97 cohort.

In general, our model does a reasonable job of fitting the between-person components of inequality, but is less successful at fitting the within-person components. The poorest fit is for the within-person/intensive margin component of earnings inequality, which is over-fit by a factor of 90% in the 1968-77 panel, 70% in the 1978-87 panel, and 62% in the 1988-97 panel. The positive bias in this component is driven by the inability of our normal-theoretic model of the components of earnings to adequately match the shape of the distribution of individual earnings (conditional on work). Because our models over-fit the within-person components for all three cohorts, however, the bias in predicting the *change* in overall inequality (i.e., the change in CV^2) between the first and third cohorts is relatively modest (around 7%)

6.3 Implications for inequality: the changing effects of family

We now use our estimated model to address the changing effects of family-related factors – specifically the presence of children in different ages and presence and characteristics of spouses/partners – on female earnings inequality. For simplicity, we focus on changes between our earliest cohort, who are observed between 1968 and 1977, and our latest cohort, who are observed between 1988 and 1997.

Effects of children Table 5a shows the results of a three alternative simulations of our model: the full simulation (already summarized in Table 4); a version of our model in which we "turn off" the direct effects of children in the current year (e.g., the dummy for having a child under 5 in the current year), but leave in place the effects of the average values of the child variables in our model, which capture differences in unobservable characteristics between women with no children and those with different average numbers of children during our panel; a version in which we turn off the effects of the average values of the child variables (which we call the "selection effects" of children) but leave in place the within-period effects; and a version in which we turn off all the child-related variables, effectively making each person childless in all years.

In general, both the direct and selection effects of children are larger in the early cohort. For example, in the first row of the table we see that in the 1968-77 cohort, the direct effects of children lower average employment by 8 percentage points (ppts) (from 67% to 59%), the selection effects lower average employment by 3 ppts, (from 62% to 59%) and the combination lowers average employment rates by 10 ppts (from 69% to 59%). In the 1988-97 cohort the corresponding impacts are -3 ppts for the direct effects of children, -1 ppt for the indirect effects, and -4 ppts for the combined effect.

The net impacts of these changes are summarized in Table 5b. Again focusing on the top row, we see that the actual change in average employment rates between the cohorts was 22 ppts, as was the predicted change in our full simulation. Taking away the direct effects of children the simulated increase would have been 17 ppts, taking away the selection effects of it would have been 20 ppts, and taking away all child factors it would have been 16 ppts. Thus, we infer that the changing effects of children account for 22-16=6 ppts of the increase, or a 27% share of the observed 22 ppt increase in employment.

Carrying out the same exercise for the change in CV^2 – our measure of overall earnings inequality – and its main components, the results in Table 5a and 5b suggest three conclusions. First, both the direct and selection effects of children increase earnings variation. Second, these effects were becoming smaller in the three decades we study. Third, as a result of the second factor, the gradual dimunition of child-related factors has contributed to a narrowing of most components of female earnings inequality, with a typical "explained share" around 30-50 percent.

Effects of spouse/partner Table 6 shows the a parallel analysis for the effects of the spousal variables included in our model. We simplify the analysis relative to Table 5a by simply comparing our full model simulation to a simulation from a model with all the spousal-related factors "turned off". In general, the results suggest that spouse-related factors are an important source of variation in female earnings, though again the effects are notably smaller for the later cohort. We emphasize that this is not a reflection of a fall in marriage rates: across our three cohorts the rate of marriage/partnering is very similar at around 70% (see Table 1). Instead it reflects falling magnitudes for the impacts of the spouse-related variables.

Looking at the right-most column of Table 6 we see that spousal variables explain about 40% of the rise in employment rates between the 1968-77 cohort and the 1988-97 cohort, and about 55% of the rise in overall earnings inequality, with an especially large explanatory role in the between-person/extensive margin component (64%).

Combined effects of family-related factors Finally, in Table 7 we show the results from a simulation in which we "turn off" all child-related and spouse-related variables. The combined effects of the two sets of variables are quite large for the 1968-78 cohort. For example, our model suggests that average employment rates of women would have been around 80% in the absense of these factors - nearly the same as the actual average employment rate of women in our 1988-97 cohort. By the third cohort the combined variables are still working in the same direction, but their net effects are far smaller. Thus, we conclude that the changing effects of children and spouse-related variables can account for about 63% of the rise in average employment rates.

The share of the fall in overall earnings inequality between the earlier and later cohort explained by family forces is even larger, at 73%, again with a particularly big impact on the between-person extensive margin component (85% explained).

7 Concluding discussion

incomplete

References

Abowd, J. M. and D. Card (1989). On the covariance structure of earnings and hours changes. *Econometrica 57*(2), 411–445.

Altonji, J. G., A. A. Smith, and I. Vidangos (2013). Modeling earnings dynamics. *Econometrica 81*(4), 1395–1454.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *The American Economic Review 101*(3), 471–475.

Cogan, J. F. (1981). Fixed costs and labor supply. *Econometrica*, 945–963.

Eckstein, Z. and O. Lifshitz (2011). Dynamic female labor supply. *Econometrica 79*(6), 1675–1726.

Friedman, M. (1957). The permanent income hypothesis. In *A theory of the consumption function*, pp. 20–37. Princeton University Press.

Gottschalk, P. and R. Moffitt (2009). The rising instability of us earnings. *The Journal of Economic Perspectives 23*(4), 3–24.

Haider, S. J. (2001). Earnings instability and earnings inequality of males in the united states: 1967–1991. *Journal of labor Economics 19*(4), 799–836.

Hansen, G. D. (1985). Indivisible labor and the business cycle. *Journal of monetary Economics 16*(3), 309–327.

Heckman, J. J. (1978). Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. *Annales de l'INSEE 30–31*, 227–269.

Heckman, J. J. (1981). Heterogeneity and state dependence. In S. Rosen (Ed.), *Studies in Labor Markets*, pp. 91–139. Chicago and London: University of Chicago Press.

Heckman, J. J. (1993). What has been learned about labor supply in the past twenty years? *The American Economic Review 83*(2), 116–121.

Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica 67*(6), 1255–1294.

Katz, L. F. and D. H. Autor (1999). Changes in the wage structure and earnings inequality. *Handbook of labor economics 3*, 1463–1555.

Keane, M. and R. Rogerson (2012). Micro and macro labor supply elasticities: A reassessment of conventional wisdom. *Journal of Economic Literature 50*(2), 464–476.

Levy, F. and R. J. Murnane (1992). Us earnings levels and earnings inequality: A review of recent trends and proposed explanations. *Journal of economic literature 30*(3), 1333–1381.

Ljungqvist, L. and T. J. Sargent (2011). A labor supply elasticity accord? *The American Economic Review 101*(3), 487–491.

Low, H., C. Meghir, and L. Pistaferri (2010). Wage risk and employment risk over the life cycle. *The American economic review 100*(4), 1432–1467.

MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of econometrics 18*(1), 83–114.

Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica 72*(1), 1–32.

Meghir, C. and L. Pistaferri (2011). Earnings, consumption and life cycle choices. *Handbook of Labor Economics 4*, 773–854.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 765–799.

Mulligan, C. B. and Y. Rubinstein (2008). Selection, investment, and women's relative wages over time. *The Quarterly Journal of Economics*, 1061–1110.

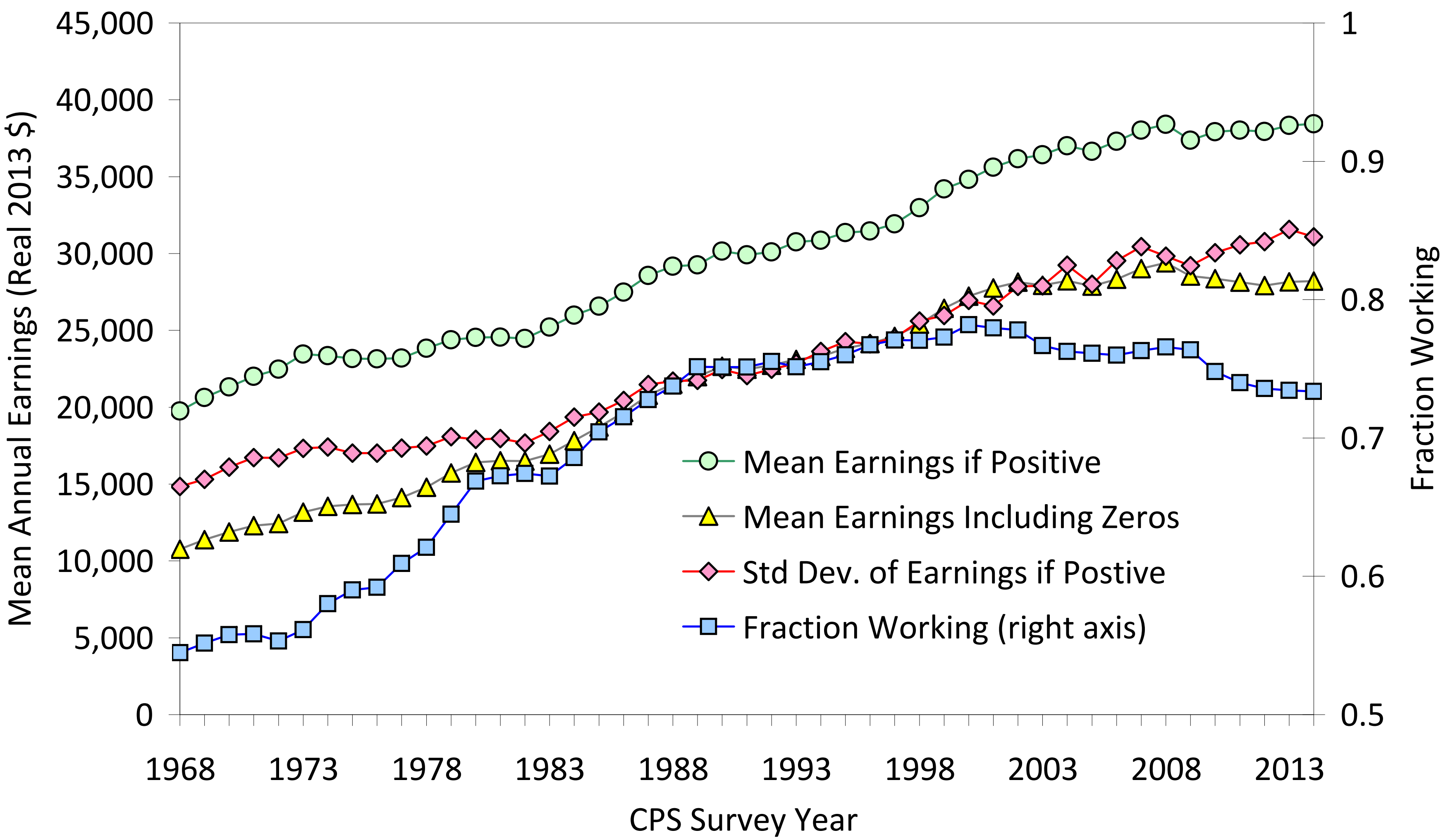
Peterman, W. B. (2016). Reconciling micro and macro estimates of the frisch labor supply elasticity. *Economic Inquiry 54*(1), 100–120.

Francine D. Blau & Lawrence M. Kahn, 2017. "The Gender Wage Gap: Extent, Trends, and Explanations," Journal of Economic Literature, vol 55(3), pages 789-865

Gottschalk, Peter and Moffitt, Robert, (1994), The Growth of Earnings Instability in the U.S. Labor Market, Brookings Papers on Economic Activity, 25, issue 2, p. 217-272

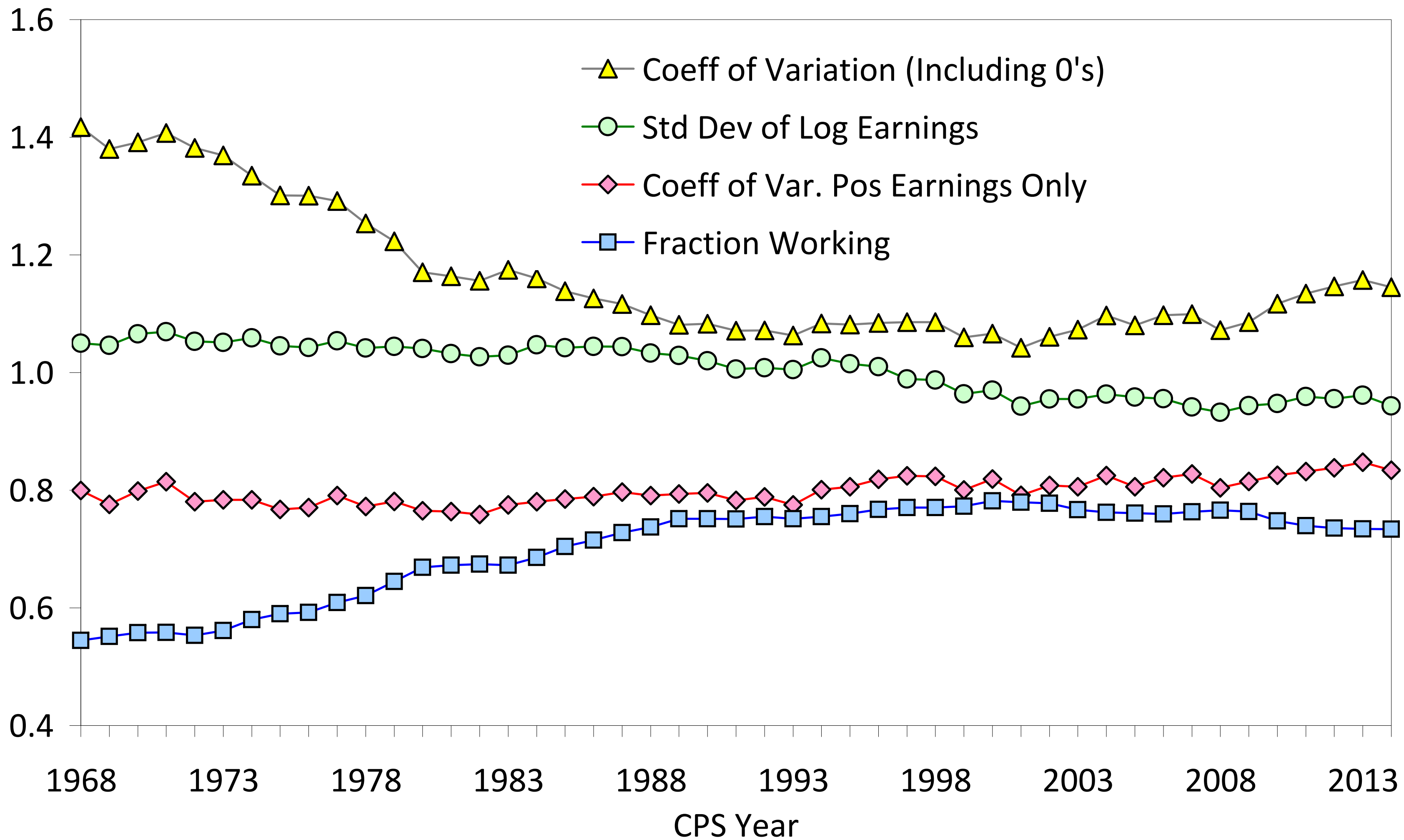
Dean R. Hyslop, 1999. "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," Econometrica, Econometric Society, vol. 67(6), pages 1255-1294, November.

Figure 1: Real Mean Annual Earnings of Females in March CPS



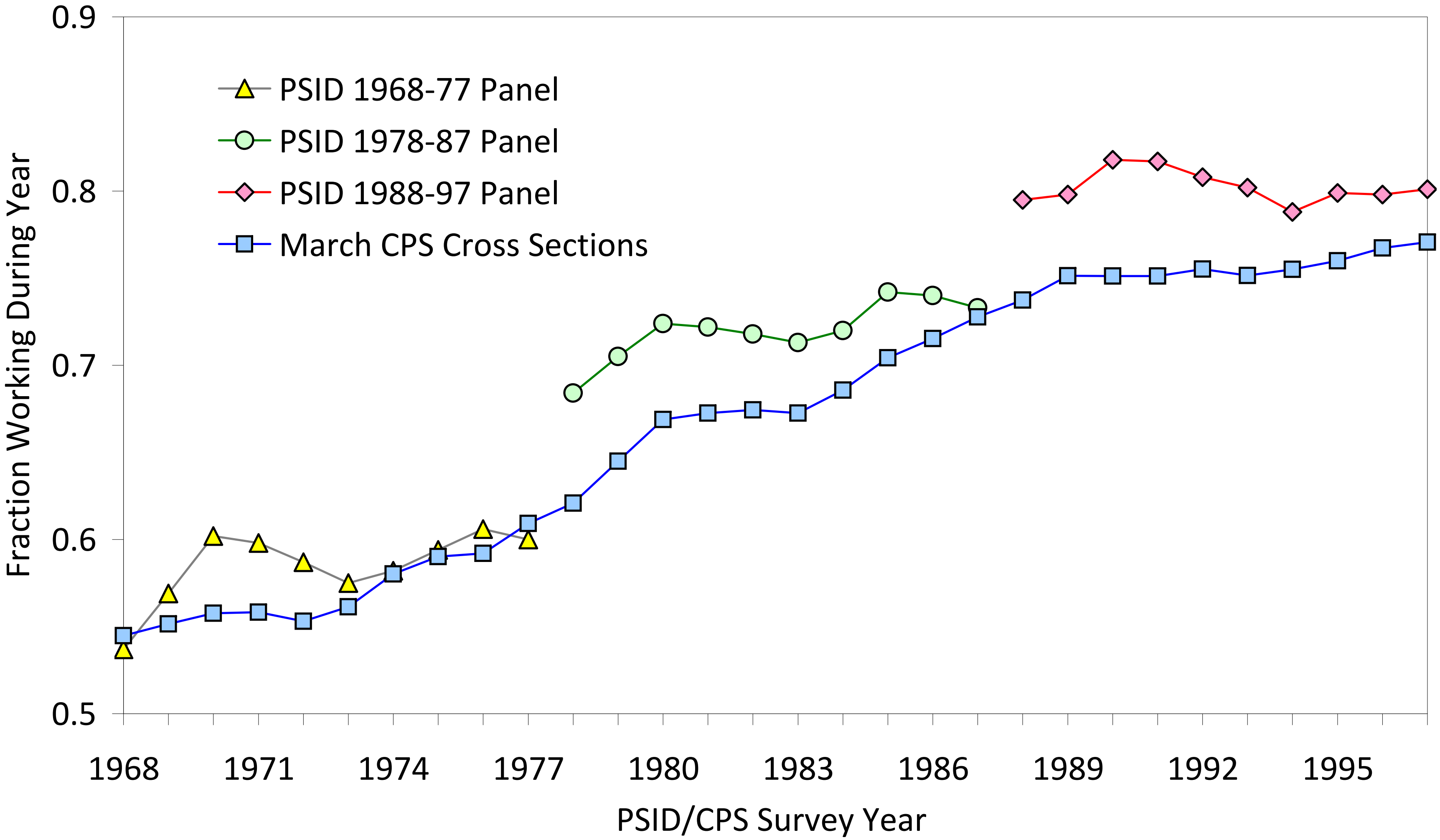
Note: sample restricted to females age 25-60 in all years. Sample weights are not used. Earnings are Winsorized at 5th and 99th percentiles.

Figure 2: Trends in Female Earnings Inequality - Annual Data from March CPS



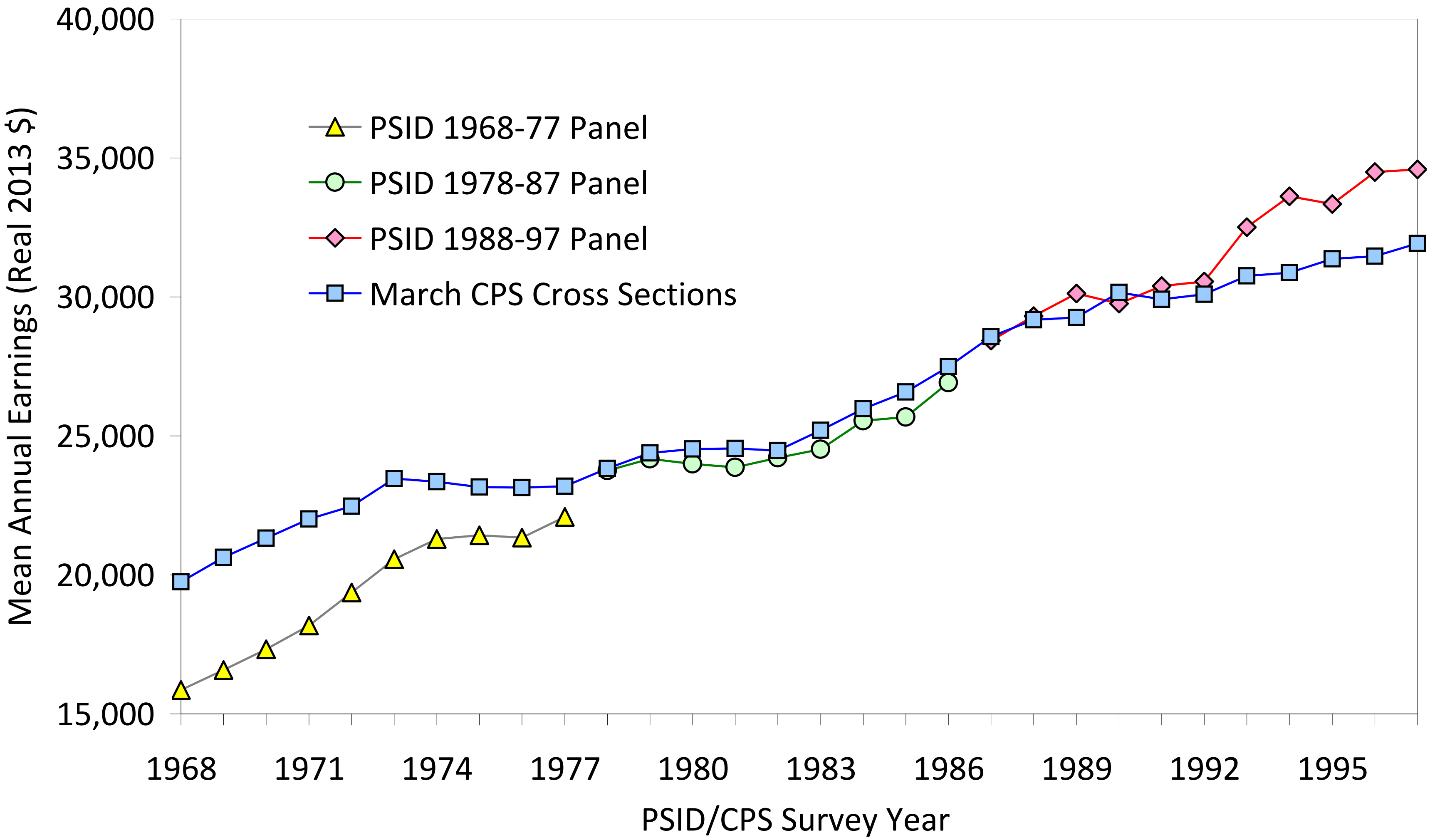
Note: sample restricted to females age 25-60 in all years. Earnings in real 2013 dollars are Winsorized at 5th and 99th percentiles.

Figure 3a: Fraction of Females Working During Year - PSID vs. CPS



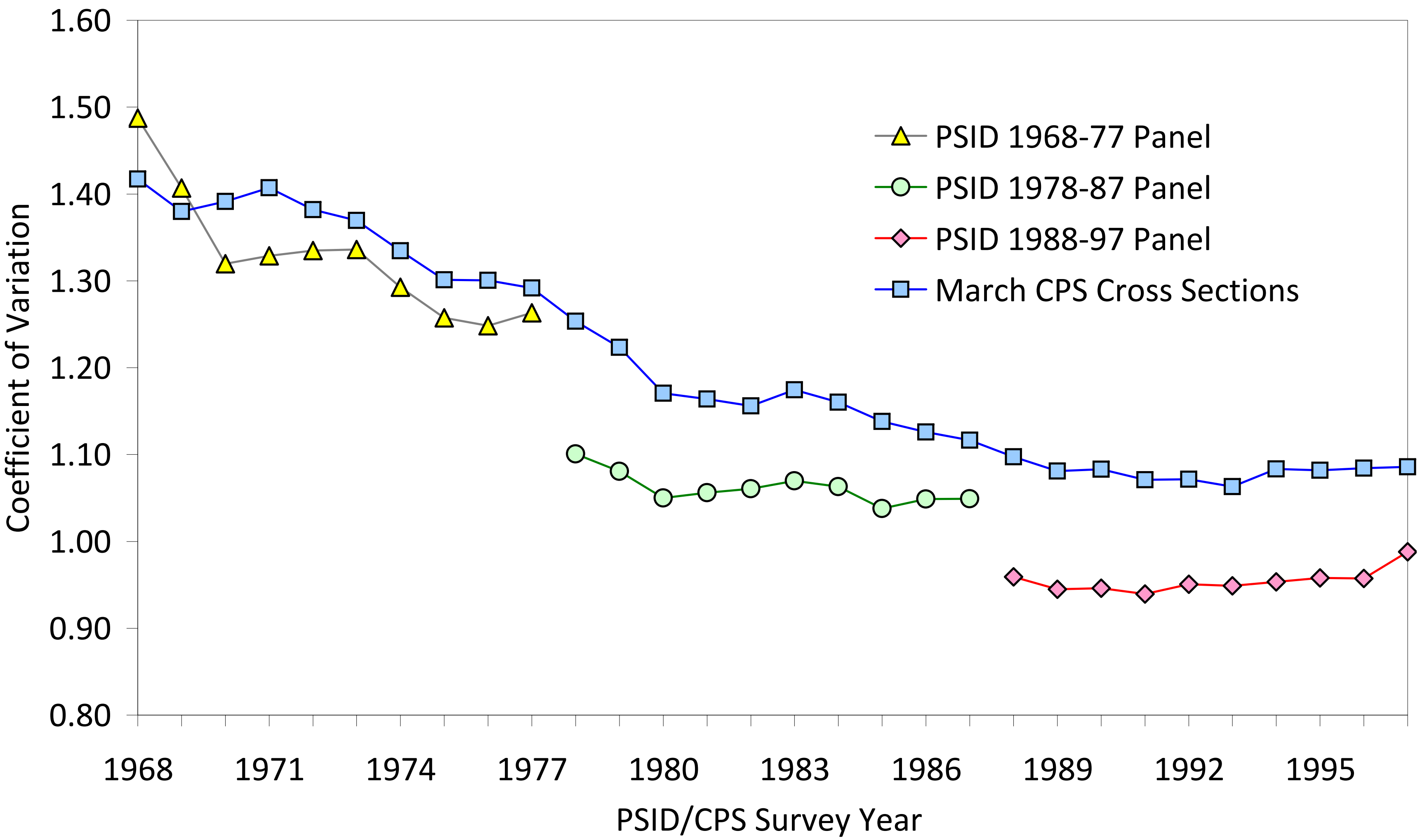
Note: sample restricted to females age 25-60 in all years. PSID and CPS sample weights are not used.

Figure 3b: Mean Earnings Conditional on Working - PSID vs. CPS



Note: sample restricted to females age 25-60 in all years. PSID and CPS sample weights are not used. Low earnings are Winsorized at $1500/year in real 2013 dollars.

Figure 3c: Coefficient of Variation of Female Earnings During Year - PSID vs. CPS



Note: sample restricted to females age 25-60 in all years. PSID and CPS sample weights are not used. Earnings are Winsorized at 5th and 99th percentiles.

Figure 4: Measures of Year-to-Year Volatility in Earnings

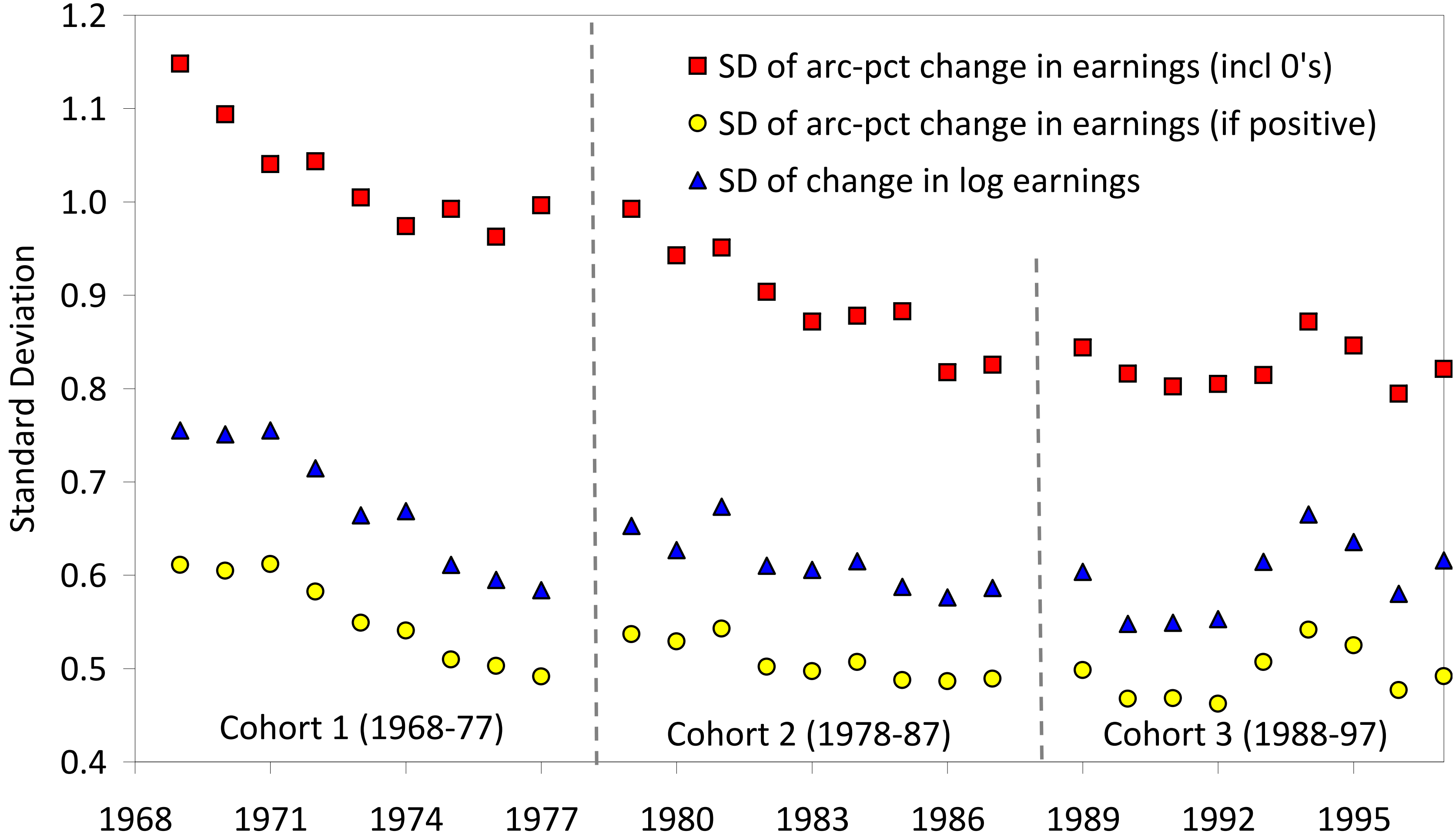


Figure 5: Components of Overall Female Earnings Inequality - Early vs Late Cohorts in PSID

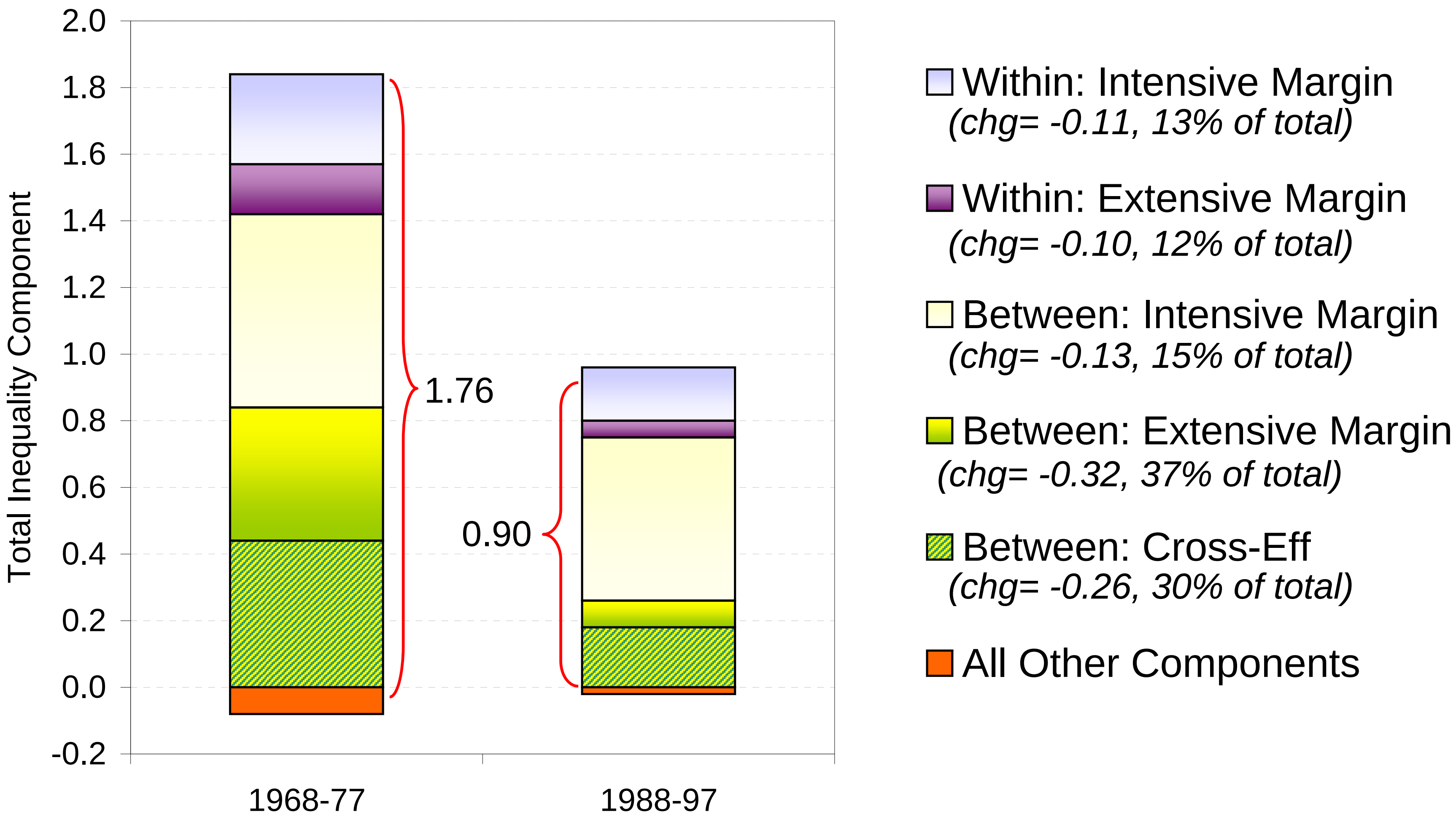


Table 1: Descriptive statistics: Females over 10-year balanced panels

	Sample period		
	1968-77	1978-87	1988-97
Age	41.6	39.8	39.2
	(8.3)	(8.9)	(7.7)
Education	11.2	12.6	13.2
	(2.7)	(2.4)	(2.2)
Black	0.37	0.33	0.31
	(.48)	(.47)	(.46)
Married	0.71	0.72	0.74
	(.45)	(.45)	(.44)
No. children	2.21	1.45	1.42
	(2.0)	(1.4)	(1.3)
Age of	7.54	7.81	7.29
youngest child	(4.6)	(5.1)	(4.9)
Employed	0.58	0.72	0.80
(Annual hours & earnings)	(.49)	(.45)	(.40)
Annual hours	1,384	1,512	1,661
	(750)	(717)	(703)
Annual earnings	19,422	25,129	31,889
	(15,241)	(18,351)	(23,309)
Average hourly earnings	14.39	17.06	20.29
	(11.5)	(19.8)	(27.4)
Spouse's:			
Age	44.5	42.3	41.7
	(9.3)	(10.2)	(8.8)
Employed	0.95	0.93	0.94
	(.22)	(.25)	(.23)
Annual hours	2,245	2,185	2,257
	(675)	(675)	(642)
Annual earnings	53,222	58,948	65,127
	(31,800)	(35,582)	(45,496)
Average hourly earnings	25.37	28.85	31.63
	(19.2)	(26.7)	(45.4)
No. Individuals	2,076	2,500	2,286

Notes: Each sample consists of balanced 10-year panels over the period. Average annual hours and earnings, and hourly earnings, are conditional on employment. Earnings are censored annually at the 5th and 95th percentiles, and adjusted using the CPI-U index to be in constant (2013) $-values. Standard deviations are in parentheses.

Table 2: Decomposition of Total Earnings Inequality into Within/Between Person and Extensive/Intensive Margins

	Female Heads:			Male Heads:		
	1968-1977	1978-1987	1988-1997	1968-1977	1978-1987	1988-1997
Total CV^2 of annual earnings	1.76	1.12	0.90	0.37	0.41	0.51
Within-person component	0.46	0.24	0.21	0.07	0.10	0.12
(share of total)	(0.26)	(0.22)	(0.23)	(0.20)	(0.23)	(0.23)
Wtd. avg of CV_i^2 (approx. within)	0.42	0.24	0.21	0.07	0.10	0.12
Between-person component	1.44	0.91	0.71	0.31	0.32	0.39
(share of total)	(0.82)	(0.81)	(0.79)	(0.84)	(0.77)	(0.78)
Cross-term	-0.15	-0.03	-0.02	-0.01	0.00	-0.01
Decomposition of within-person component:						
Extensive margin component	0.15	0.07	0.05	0.01	0.01	0.02
(share of total)	(0.09)	(0.06)	(0.06)	(0.03)	(0.02)	(0.04)
Intensive margin component	0.27	0.17	0.16	0.06	0.08	0.10
(share of total)	(0.15)	(0.15)	(0.18)	(0.16)	(0.19)	(0.20)
Decomposition of between-person component:						
Extensive margin component	0.40	0.16	0.08	0.01	0.01	0.01
(share of total)	(0.23)	(0.14)	(0.09)	(0.02)	(0.02)	(0.02)
Intensive margin component	0.58	0.49	0.45	0.27	0.28	0.35
(share of total)	(0.33)	(0.43)	(0.50)	(0.74)	(0.68)	(0.68)
Cross-term	0.44	0.26	0.18	0.03	0.03	0.04
(share of total)	(0.25)	(0.23)	(0.20)	(0.07)	(0.07)	(0.08)
Time variation	0.02	0.01	0.00	0.00	0.00	0.00

Notes: See text for details of decomposition. Samples are desribed in Table 1 (females) and Appendix Table 1 (males).

Table 4: Actual and Predicted Components of Total Earnings Inequality of Females in PSID

	1968-77		1978-87		1988-97		Change: 1968-77 to 1988-97	
	Actual	Simulated	Actual	Simulated	Actual	Simulated	Actual	Simulated
Mean employment rate	0.58	0.59	0.72	0.72	0.80	0.81	0.22	0.22
Total CV^2 of annual earnings	1.76	2.11	1.12	1.40	0.90	1.19	-0.85	-0.92
Within-person component	0.46	0.80	0.24	0.41	0.21	0.35	-0.26	-0.45
Wtd Avg. of CV_i^2 (approx. within)	0.42	0.75	0.24	0.41	0.21	0.35	-0.21	-0.40
Between-person component	1.44	1.53	0.91	1.04	0.71	0.89	-0.73	-0.64
Cross-term	-0.15	-0.22	-0.03	-0.05	-0.02	-0.05	0.13	0.16
Decompostion of within-person component								
Extensive margin component	0.15	0.24	0.07	0.12	0.05	0.08	-0.10	-0.16
Intensive margin component	0.27	0.51	0.17	0.29	0.16	0.26	-0.11	-0.24
Decompostion of between-person component								
Extensive margin component	0.40	0.41	0.15	0.17	0.08	0.10	-0.33	-0.32
Intensive margin component	0.58	0.64	0.49	0.58	0.45	0.59	-0.12	-0.06
Cross term	0.44	0.44	0.26	0.28	0.18	0.20	-0.26	-0.24
Time-variation	0.02	0.03	0.01	0.01	0.00	0.01	-0.02	-0.02

Note: see notes in Table 2.

Table 5a: Actual and Simulated Changes in Inequality Between 1968-77 and 1988-97 Cohorts

	Early Cohort (1968-77)					Late Cohort (1988-97)				
	Actual	Full Simulation	Remove Direct Effs. of Kids	Remove Selection Effs. of Kids	Remove All Effs. of Kids	Actual	Full Simulation	Remove Direct Effs. of Kids	Remove Selection Effs. of Kids	Remove All Effs. of Kids
Mean Employment Rate	0.58	0.59	0.67	0.62	0.69	0.80	0.81	0.84	0.82	0.85
Total CV^2 of annual earnings	1.76	2.11	1.78	1.98	1.71	0.90	1.19	1.12	1.17	1.09
Within-person component	0.46	0.80	0.67	0.71	0.62	0.21	0.35	0.32	0.33	0.31
Wtd Avg. of CV_i^2 (~within)	0.42	0.75	0.63	0.70	0.61	0.21	0.35	0.32	0.34	0.31
Between-person component	1.44	1.53	1.28	1.39	1.20	0.71	0.89	0.85	0.88	0.82
Cross-term	-0.15	-0.22	-0.17	-0.13	-0.11	-0.02	-0.05	-0.05	-0.04	-0.04
Decomposition of within-person component										
Extensive margin component	0.15	0.24	0.17	0.21	0.15	0.05	0.08	0.07	0.08	0.06
Intensive margin component	0.27	0.51	0.46	0.49	0.45	0.16	0.26	0.25	0.26	0.25
Decomposition of between-person component										
Extensive margin component	0.40	0.41	0.25	0.34	0.22	0.08	0.10	0.07	0.08	0.06
Intensive margin component	0.58	0.64	0.64	0.63	0.62	0.45	0.59	0.60	0.59	0.59
Cross term	0.44	0.44	0.37	0.41	0.35	0.18	0.20	0.17	0.20	0.17
Time-variation	0.02	0.03	0.03	0.01	0.01	0.00	0.01	0.01	0.00	0.00

Notes: see Table 4 and text for description of alternative simulations.

Table 5b: Actual and Simulated Changes in Components of Earnings Inequality, Early to Late Cohort

	Simulated Changes, 1968-77 (Early) to 1987-97 (Late) Cohort					Share of Change Attributable to Change Effect of Kids
	Actual	Full Simulation	Remove Direct Effs. of Kids	Remove Selection Effs. of Kids	Remove All Effs. of Kids	
Mean Employment Rate	0.22	0.22	0.17	0.20	0.16	0.27
Total CV^2 of annual earnings	-0.85	-0.92	-0.66	-0.81	-0.61	0.33
Within-person component	-0.26	-0.45	-0.35	-0.38	-0.31	0.32
Wtd Avg. of CV_i^2 (approx. within)	-0.21	-0.40	-0.31	-0.36	-0.29	0.28
Between-person component	-0.73	-0.64	-0.44	-0.52	-0.37	0.41
Cross-term	0.13	0.16	0.13	0.09	0.07	0.60
Decomposition of within-person component						
Extensive margin component	-0.10	-0.16	-0.11	-0.13	-0.09	0.43
Intensive margin component	-0.11	-0.24	-0.21	-0.23	-0.20	0.18
Decomposition of between-person component						
Extensive margin component	-0.33	-0.32	-0.18	-0.26	-0.16	0.51
Intensive margin component	-0.12	-0.06	-0.04	-0.04	-0.03	0.51
Cross term	-0.26	-0.24	-0.19	-0.21	-0.18	0.24
Time-variation	-0.02	-0.02	-0.02	-0.01	-0.01	0.64

Notes: Entries are changes between 1968-77 cohort and 1988-97 cohort. See Table 5a.

Table 6: Actual and Simulated Components of Earnings Inequality, Early to Late Cohort

	Early Cohort (1968-77)			Late Cohort (1988-97)			Change from Early to Late Cohort			Share of Change Attributable to Changing Effs. of Spouse
	Actual	Full Simulation	Remove Effects of Spouse	Actual	Full Simulation	Remove Effects of Spouse	Actual	Full Simulation	Remove Effects of Spouse	
Mean Employment Rate	0.58	0.59	0.71	0.80	0.80	0.84	0.22	0.21	0.13	0.41
Total CV^2 of annual earnings	1.76	2.11	1.47	0.90	1.20	1.08	-0.85	-0.90	-0.39	0.55
Within-person component	0.46	0.80	0.51	0.21	0.35	0.30	-0.26	-0.45	-0.21	0.18
Wtd Avg. of CV_i^2 (~within)	0.42	0.75	0.50	0.21	0.35	0.31	-0.21	-0.40	-0.19	0.11
Between-person component	1.44	1.53	1.06	0.71	0.91	0.82	-0.73	-0.62	-0.24	0.67
Cross-term	-0.15	-0.22	-0.11	-0.02	-0.05	-0.05	0.13	0.17	0.06	0.51
Decomposition of within-person component										
Extensive margin component	0.15	0.24	0.13	0.05	0.08	0.06	-0.10	-0.16	-0.07	0.31
Intensive margin component	0.27	0.51	0.37	0.16	0.26	0.25	-0.11	-0.24	-0.12	-0.08
Decomposition of between-person component										
Extensive margin component	0.40	0.41	0.18	0.08	0.10	0.06	-0.33	-0.32	-0.12	0.64
Intensive margin component	0.58	0.64	0.57	0.45	0.59	0.57	-0.12	-0.05	0.00	1.02
Cross term	0.44	0.44	0.29	0.18	0.21	0.18	-0.26	-0.23	-0.11	0.57
Time-variation	0.02	0.03	0.02	0.00	0.01	0.01	-0.02	-0.02	-0.01	0.40

Note: see Table 4 and text for decription of alternative simulations.

Table 7: Actual and Simulated Components of Earnings Inequality, Early to Late Cohort

	Early Cohort (1968-77)			Late Cohort (1988-97)			Change: Early to Late Cohort			Share of Change Attributable to Changing Effs. of Spouse/Kids
	Actual	Full Simulation	Remove Effects of Spouses/Kids	Actual	Full Simulation	Remove Effects of Spouses/Kids	Actual	Full Simulation	Remove Effects of Spouse	
Mean Employment Rate	0.58	0.59	0.80	0.80	0.81	0.88	0.22	0.21	0.08	0.63
Total CV^2 of annual earnings	1.76	2.11	1.24	0.90	1.19	1.01	-0.85	-0.90	-0.23	0.73
Within-person component	0.46	0.80	0.43	0.21	0.35	0.28	-0.26	-0.45	-0.15	0.41
Wtd Avg. of CV_i^2 (~within)	0.42	0.75	0.42	0.21	0.35	0.28	-0.21	-0.40	-0.14	0.35
Between-person component	1.44	1.53	0.88	0.71	0.89	0.76	-0.73	-0.62	-0.12	0.84
Cross-term	-0.15	-0.22	-0.07	-0.02	-0.05	-0.04	0.13	0.17	0.03	0.77
Decomposition of within-person component										
Extensive margin component	0.15	0.24	0.08	0.05	0.08	0.05	-0.10	-0.16	-0.04	0.66
Intensive margin component	0.27	0.51	0.34	0.16	0.26	0.24	-0.11	-0.24	-0.10	0.10
Decomposition of between-person component										
Extensive margin component	0.40	0.41	0.09	0.08	0.10	0.04	-0.33	-0.32	-0.05	0.85
Intensive margin component	0.58	0.64	0.57	0.45	0.59	0.58	-0.12	-0.05	0.01	1.08
Cross term	0.44	0.44	0.22	0.18	0.20	0.14	-0.26	-0.23	-0.08	0.69
Time-variation	0.02	0.03	0.01	0.00	0.01	0.01	-0.02	-0.02	0.00	0.78

Note: see Table 4 and text for decription of alternative simulations.