Gender, price, and quantity effects in U.S. earnings inequality: revisiting counterfactual density estimates

Andrew Silva

The Gender Center
The Graduate Institute of International and Development Studies
Chemin Eugène-Rigot 2
1211 Geneva, Switzerland
andrew.silva@graduateinstitute.ch

Abstract

I decompose changes in the U.S. household earnings distribution from 1975 to 2018 to examine the labor market processes underlying its evolution over time. I model the distribution of earnings as a function of price effects (wages) and quantity effects (work hours and household employment), each of which are specified separately for men and women, and apply a semi-parametric density estimation technique to infer their contributions to inequality measures over time. Results indicate that changes to the male wage distribution explain much of the growth in earnings inequality, but that its contribution varied greatly over time, with peak contributions in the mid 1990s; while changes in female work hours have actually mitigated inequality growth, particularly by raising earnings in the lower and mid portions of the distribution, with very consistent effects over time. These results demonstrate the relevance of work hours in addition to wage rates in explaining earnings inequality, and the importance of gender differences therein.

JEL codes: D31 J31 R11

Keywords: gender, income, wages, inequality, decomposition

1 Introduction

This paper examines the growth in U.S. household earnings inequality over the past four decades by drawing a unique distinction between "price" and "quantity" effects, and identifying gender-specific components therein. A number of studies have confirmed the rapid growth in household income inequality over this period (Thompson and Smeeding, 2013), particularly with respect to earnings and wages (Katz et al., 1999; Autor et al., 2006; Autor et al., 2008). Less commonly examined, however, is the generating processes underlying these trends and how they differ by gender. This paper extends earlier research (Anonymous, 2019) on the role of auxiliary income sources in the development of household income inequality to focus solely on the growth of earnings inequality per se.

I model household earnings as a function of its most basic labor market components: hourly wages (price effects), work hours, and the number of employed individuals per household (quantity effects). A handful of studies examine the role of wages in household inequality growth, but few address the role of work hours or household employment, and no study (to my knowledge) addresses these components jointly in a unified decomposition framework. I consider these earnings components with respect to male and female workers, as earlier research has identified the salience of gender differences in labor market outcomes. In my analysis, I find that female annual work hours have increased nearly 50% since 1975, while male work hours and the number of employed males per household have gradually declined over the same period. These trends prompt investigation with regard to their potential role in the evolution of the household earnings distribution.

The approach begins with the simple mechanical construction of household earnings. Total household earnings y for household i, consisting of members with wages w and annual work hours h, may be expressed as:

$$y_i = w_{i1}h_{i1} + w_{i2}h_{i2} + ... + w_{iM_i}h_{iM_i} + w_{iF_i}h_{iF_i} = \sum_{m=1}^{M_i} w_{im}h_{im} + \sum_{f=1}^{F_i} w_{if}h_{if} \tag{1}$$

where M_i and F_i are the total number of males and females employed in the household, respectively. Unfortunately, there is no way to further simplify this expression into a generalizable multiplicative form representative of all households, i.e. a way to factor-out wages and work hours from the number of employed individuals, as M_i and F_i vary by household. I therefore turn to a distributional model of this relationship, expressing household earnings as the joint probability distribution of male and female wages, male and female work hours, and the number of males and females employed per household:

$$f\left(y\right) = g\left(w_m, w_f, h_m, h_f, M, F\right) \tag{2}$$

where wages and work hours refer to the household head and spouse of household head (if present).[1] The objective is to isolate the influences of these first-order components in determining the household earnings distribution over the last four decades, particularly in how they have contributed to the growth in earnings inequality.

To this end, I employ a semi-parametric density simulation procedure pioneered by DiNardo et al. (1996), which I find to be underutilized in the literature. The method allows for the es-

[1] The head is the only household member necessarily present, by survey design. Other prominent inequality studies similarly limit labor market considerations to the household head, e.g. Burtless (1999) and Daly and Valletta (2006).

timation of a counterfactual income distribution due to a change in the distribution of one (or more) household characteristics. In other words, the method allows one to answer questions such as "What would the earnings density be in 2018 had the male wage distribution remained as it were in 1975?" The way in which this counterfactual density differs from the true earnings density in 2018 can then be attributed to changes in the wage distribution between 1975 and 2018. This effectively allows me to attribute a proportion of the true change in earnings density over time to changes in the male wage distribution per se. I apply the method to the six household earnings components in Equation (2), in what amounts to a unified decomposition model for the household earnings distribution.

Because the method allows for the estimation of an entire counterfactual distribution, I can then characterize the distribution (and any changes to it) by any number of distributional metrics.[2] In this study I focus on measures of the Gini index, the 10$^{\text{th}}$ percentile, median, and 90$^{\text{th}}$ percentile. Additionally, I plot counterfactual time trends for distributional metrics by "freezing" a particular household characteristic in 1975 and projecting the resulting counterfactual trend forward through time. This latter exercise provides and important robustness check, ensuring that results are not dependent upon the particular years chosen for the decomposition period.

Several studies find a positive association between household income inequality and male and female wage contributions. (Shorrocks, 1983; Lerman and Yitzhaki, 1985; Karoly and Burtless, 1995; Burtless, 1999; Daly and Valletta (2006)). Conversely, another set of studies finds a negative association between household inequality and female wages (Cancian and Reed 1998, 1999; Reed and Cancian, 2001). However, these studies make no considerations for other labor market characteristics, and while all are based on CPS data, none addresses the potential influence of a 1994 CPS survey redesign (discussed in Section 2).

Only two studies, to my knowledge, address work hours in the context of earnings inequality—albeit using only descriptive methods. Gottschalk and Danziger (2005) find that while female wage inequality grew faster than male wage inequality, growth of female work hours at the base of the distribution resulted in a decline in female earnings inequality, which slowed the growth of household earnings inequality. Heathcote et al. (2010) also determine that increases in female work hours likely mitigated growing household inequality. A few additional studies find a negative association between female labor force participation and household income inequality, including Nielsen and Alderson (1997), McLaughlin (2002), and Chevan and Stokes (2000), each of which analyzes geographical observational units;[3] and Daly and Valletta (2006), who find a similar association in household observations.

The literature thus far addresses price effects (wages) and quantity effects (work hours, employment) in a piecemeal fashion, over a series of disparate studies. This study not only jointly assesses wages and hours, but does so symmetrically for males and females in unified model of household earnings, providing for direct comparison of their relative contributions. Furthermore, measuring work hours per se provides much greater resolution than mere female labor force participation, as has primarily been the case in other studies. This analysis also takes

[2]Regression-based methods tend to rely on non-standard inequality metrics, such as R^2 (i.e. the Shapley decomposition; Shapley, 1953), or log-variance (Fields, 2003), which makes results incompatible with the larger literature. Rank-preserving income-replacement methods (Burtless, 1999; Reed and Cancian, 2001) essentially recreate household units, potentially resulting in implausible matches between income and household characteristics. Other density estimation techniques (Machado and Mata, 2005; Melly, 2005; Autor et al., 2005) remain faithful to household construction, yet rely on restrictive over-parameterization of the relationship between household income and labor market characteristics.

[3]Although based on geographical observation units, these studies apparently make no considerations for two-dimensional spatial autocorrelation, which likely biases estimates Anselin (2013).

precautions to adjust for income topcoding and a significant measurement transition due to the 1994 CPS survey redesign. Finally, the study provides a continuous decomposition over time, which ensures that beginning and ending years were not cherry-picked for more salient results.

The next sections describe the data and methodology for the study, followed by a brief discussion of results.

2 Data

The data come from the annual March Current Population Survey, Annual Social and Economic Supplement (CPS ASEC).[4] I restrict the sample to households with positive earnings values; I drop households in which both male and female heads have zero wages, and in which both heads work less than 5% of full-time equivalent (less than 104 hours per year). This eliminates extreme wage outliers due to unreasonably low hours recorded. Earnings and wage values are adjusted to 2018 dollars using the Personal Consumption Expenditures (PCE) index (U.S. Bureau of Economic Analysis, 2018).

To mitigate the bias introduced by income topcodes (Larrimore et al., 2008), I augment CPS sample data with Census rank-proximity swap values, which revise previously censored income values, and allow for a much more accurate reproduction of the distributional statistics derived from internal-use data. These values are available from 1975 to the present.

The CPS survey methodology underwent fundamental changes in 1994 (which affected the 1993 income year) when the Census Bureau adopted Computer Assisted Personal Interviewing and introduced a completely redesigned survey questionnaire. Polivka and Rothgeb (1993) and Cohany et al. (1994) provide detailed accounts of these changes, which introduce a measurable discontinuity in the Gini index between the 1992 and 1993 income years.[5] Following Burkhauser et al. (2012), I attribute all of the increase in Gini between 1992 and 1993 to the survey revisions, and add this amount (0.228 for household earnings) as an offset to the pre-1992 series when calculating percentage change and percentage shares in the Gini decompositions. This adjustment is reflected in Figure 3, while the unadjusted data are shown in Figure 1.

In the analysis I examine household earnings and six household labor market characteristics: hourly wages, annual work hours, and the number employed per household—each of which is recorded separately for males and females. Hourly wages and work hours refer to measures obtained from the household head and spouse of household head, if the latter is present (henceforth "household heads"). Values solely from household heads will not perfectly reconstruct total household earnings, but come close, constituting 92% of aggregate household earnings in 1975, and 89% in 2018. If more than one household head of the same gender is present, I use the maximum household head value for hourly wage and work hours. If one gender is missing for a particular characteristic, I set its value to zero to preserve the observation in the estimations, following Karoly and Burtless (1995), and Burtless (1999).

Earnings is defined as before-tax earned income, which includes wages and self-employment income (from both farm and non-farm sources). *Annual work hours* is defined as product of "usual hours worked per week" and "weeks worked per year", and *hourly wages* is defined annual earnings divided by annual work hours. *Household employment* is determined from any household member with non-zero earnings, which is then totaled separately for males and fe-

[4] I use data extracts provided by the Integrated Public Use Microdata Series (IPUMS) (Flood et al., 2018), which offers consistently coded variable names across survey years.

[5] The discontinuity is also present in published Census figures, which are based on internal-use data. Percentile measures, however, do not seem to have been affected.

Table 1: Descriptive statistics for labor market characteristics

	Obs.	Mean	Std. dev.	Min	Max
			1975		
Earnings	35099	55311	40464	4	724659
Male hours	35099	1667	1025	0	5148
Female hours	35099	859	927	0	5148
Male wages	35099	19.2	20.4	-213.0	1092.2
Female wages	35099	8.6	51.9	-2703.3	5399.1
Male employment	35099	1.0	0.6	0.0	5.0
Female employment	35099	0.7	0.7	0.0	5.0
			2018		
	Obs.	Mean	Std. dev.	Min	Max
Earnings	49868	96275	107843	1	2247998
Male hours	49868	1496	1101	0	5148
Female hours	49868	1196	1026	0	5148
Male wages	49868	27.8	202.9	-24.0	21153.8
Female wages	49868	18.3	77.4	-40.0	10000.0
Male employment	49868	0.9	0.6	0.0	6.0
Female employment	49868	0.8	0.6	0.0	6.0

Notes: Annual work hours are topcoded at 5148 hours per year. Wages for one household head can be zero or negative as long as they are positive for the other household head, leading to net positive household earnings. The same applies to work hours: values can be zero as long as there is another household head present with a positive value. Samples are limited to households with positive earnings and with at least one household head working more than 5% of full-time equivalent. All monetary values are adjusted to 2018 dollars.

males.[6] Table 1 provides descriptive statistics for these measures in the initial year of analysis, 1975, and final year, 2018.

3 Methodology

The key to the DFL procedure is to derive a vector of counterfactual sampling weights, which, when applied to the sample, produce a counterfactual distribution. I deviate from DFL and follow the procedure of Biewen (2001) to use the counterfactual weights to augment existing sampling weights, and then compute counterfactual distributional metrics from the re-weighted discrete sample data.[7]

Consider a simplified model of household earnings income y and a single household char-

[6] In addition to the mentioned characteristics, conditional probability estimates include controls for: the four indicators for male and female age and education, and their four quadratics; the age-education interactions for male and female, the male-female interactions for age and education, and the four-way interaction between male and female age and education; indicators for state of residence; and total household size. Inclusion of these controls ensures convergence of probit estimates, and is similarly employed in the original implementation by DiNardo et al. (1996).

[7] In the original implementation, DFL compute a kernel density estimate as an intermediate step; however, this introduces additional sources of error in the choice of kernel function and kernel bandwidth parameter.

acteristic z. The density of household earnings in time t is

$$f_t(y) = f(y; t_y = t, t_z = t). \tag{3}$$

By applying the law of iterated expectation, I can express household earnings and household characteristics at independent time periods, t_y and t_z, respectively:

$$f_t(y) = \int_{z \in \Omega_z} \mathrm{d}F(y, z | t_y = t, t_z = t) = \int_{z \in \Omega_z} f(y | z, t_y = t) \, \mathrm{d}F(z | t_z = t) \tag{4}$$

in which Ω_z is the domain of household characteristics z. For example, $f(y; t_y = 2018, t_z = 1975)$ represents the earnings distribution in 2018 had the distribution of household characteristics remained as they were in 1975, maintaining the 2018 structural relation between earnings and household characteristics.[8] The counterfactual earnings density can then be written as a reweighted form of the actual 2018 density:

$$\begin{aligned} f(y; t_y = 2018, t_z = 1975) &= \int_{z \in \Omega_z} f(y | z, t_y = 2018) \, \mathrm{d}F(z | t_z = 1975) \\ &= \int_{z \in \Omega_z} f(y | z, t_y = 2018) \, \psi_z(z) \, \mathrm{d}F(z | t_z = 2018) \end{aligned} \tag{5}$$

in which $\psi_z(z)$ is a reweighting function, defined simply as the ratio of densities for household characteristics between the two years under consideration:

$$\psi_z(z) = \frac{\mathrm{d}F(z | t_z = 1975)}{\mathrm{d}F(z | t_z = 2018)}. \tag{6}$$

The intuitive reasoning underlying the reweighing expression is: observations in 2018 with characteristics (wages) that have a higher propensity of being found in 1975 receive a higher weight; observations with characteristics more likely to be found in 2018 receive a lower weight. An estimate for the reweighting vector, $\hat{\psi}_z(z)$, would yield the desired counterfactual weights.

Direct estimation of the reweighting equation (Equation 6) is difficult, and likely not even possible, due to lack of data at particular points in the density of z. The clever solution presented by DFL is to use Bayes' Rule to transform Equation 6 into

$$\psi_z(z) = \frac{P(t_z = 1975 | z)}{P(t_z = 2018 | z)} \cdot \frac{P(t_z = 2018)}{P(t_z = 1975)}. \tag{7}$$

The components of Equation 7 are much more readily estimated: t_z only takes on two possible values, regardless of the number of dimensions in z, and can be estimated with a simple proportion; while the probability of t_z given z can be estimated by a probit model, with z as a control.

To derive a reweighting function for only one of several household characteristics, we must breakup the dimensionality of z.[9] Consider a slightly more complex example in which household earnings is a function of three characteristics: hourly wages w, annual work hours h, and

[8] In other words, the procedure assumes that the conditional density of earnings $f(y|z, t_y = t)$ does not depend on the distribution of household characteristics $F(z)$, and reweighted marginal distributions of household characteristics in 2018 should match the true distributions of household characteristics in 1975. These counterfactual density estimates are ceteris paribus in nature, and do not account for general equilibrium effects.

[9] The original DFL implementation uses a sequential decomposition, the drawback of which being that decomposition results are highly dependent upon decomposition order.

number of employed household members m. To estimate a counterfactual earnings density in 2018 with *only* the distribution of wages set in 1975, the desired earnings function would be

$$f(y; t_y = 2018, t_w = 1975, t_h = 2018, t_m = 2018) \tag{8}$$

Applying the law of iterated expectation and rewriting the density formula in terms of 2018 characteristics yields the reweighting function

$$\psi_{w|h,m}(w, h, m) = \frac{\mathrm{d}F(w|h, m, t_w = 1975)}{\mathrm{d}F(w|h, m, t_w = 2018)}. \tag{9}$$

To transform this expression into a more readily estimated form, Fortin et al. (2011) demonstrate that the difference in the counterfactual density between reweighting the set $[w, h, m]$ versus reweighting the set $[h, m]$ should yield an unbiased estimate of the ceteris paribus contribution of w. This allows the reweighting function for wages to be written as

$$\psi_{w|h,m}(w, h, m) = \frac{\psi_{w,h,m}(w, h, m)}{\psi_{h,m}(h, m)}. \tag{10}$$

The numerator and denominator of Equation 10 are now each in the same form of Equation 6, derived earlier, only now the vector z is explicitly defined in each case. Applying the Bayes' Rule transformation to the numerator $\psi_{w,h,m}(w, h, m)$ yields:

$$= \frac{\mathrm{d}F(w, h, m|t_{w,h,m} = 1975)}{\mathrm{d}F(w, h, m|t_{w,h,s} = 2018)} = \frac{P(t_{w,h,m} = 1975|w, h, m)}{P(t_{w,h,m} = 2018|w, h, m)} \cdot \frac{P(t_{w,h,m} = 2018)}{P(t_{w,h,m} = 1975)} \tag{11}$$

and to the denominator $\psi_{h,m}(h, m)$ yields:

$$= \frac{\mathrm{d}F(h, m|t_{h,m} = 1975)}{\mathrm{d}F(h, m|t_{h,s} = 2018)} = \frac{P(t_{h,m} = 1975|h, m)}{P(t_{h,m} = 2018|h, m)} \cdot \frac{P(t_{h,m} = 2018)}{P(t_{h,m} = 1975)}. \tag{12}$$

Each of these expressions is readily estimated with probit models for the conditional means, and with simple proportions for the unconditional means. Taken together, they yield the desired reweighting vector, which provides for the estimation the counterfactual distributional statistics of choice. Any difference between the counterfactual and true statistic can then be attributed changes the household characteristic under consideration. Reweighting vectors for counterfactual annual hours worked, h, and counterfactual number of employed household members, m, can be derived analogously.

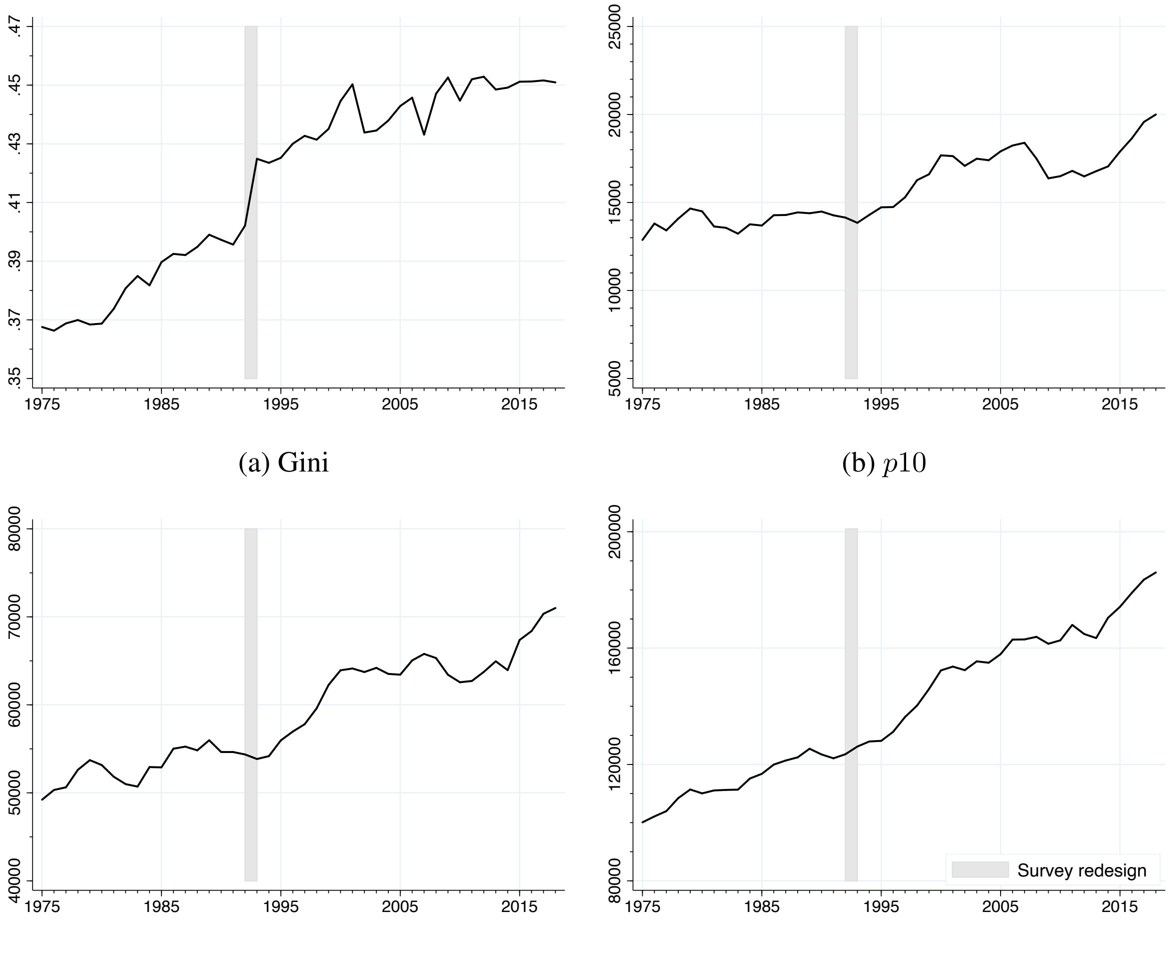
4 Results

4.1 Descriptive trends

A quick glance at the data makes it clear that earnings inequality has steadily grown between 1975 and 2018, with exceptional declines after the Dot-com boom of the late 1990s and the Financial crisis of 2007 (Figure 1a). The Gini index grew 22% over the period, with its most rapid rise occurring in 1980s.[10] Percentile measures grew most rapidly over the late 1990s. The 10$^{\text{th}}$ percentile grew 54% over the entire 44 year period (with almost zero growth between

[10] This includes the influence of the survey redesign in 1993, for which I compensate before reporting decomposition results.

Figure 1: Distributional measures for household earnings



(a) Gini

(b) $p10$

(c) Median

(d) $p90$

Notes: The Gini index is notably influenced by the survey redesign in 1993. Samples are limited to households with positive earnings and with at least one household head working more than 5% of full-time equivalent. All monetary values are adjusted to 2018 dollars. (See Section 2.)

1975 and 1993), which yet slightly outpaced the median, which grew at 47% (Figure 1b, 1c). Contrast this to the 90th percentile, which grew by 108% over the period (Figure 1d).

Male wages were consistently higher than female, but female wages have grown much faster, and over a much broader base. There was virtually no growth in mean male wages between 1975 and 1995, yet they saw substantial gains over the late 1990s and from 2014 onward, which in total amounted to a 47% increase between 1975 and 2018. Female wages, although lower, experienced more consistent growth at twice the rate, increasing by about 100% over the period (Figure 2a). Mean female wages were 45% that of male's in 1975, which rose to 64% by 2018.

Female wage growth between 1975 and 2018 nearly tripled male wage growth in the lower-mid portions of the wage distribution; at the 40th percentile, female wages grew at just over 60%, while male wages grew at 20% (Figure 2c). Growth was strongest in the upper-tail of the distribution, for both genders. Characterizing wage growth instead by household earnings percentile reveals even stronger female wage growth in the upper-mid portion of the distribution (Figure 2f).

Male work hours have declined while female work hours have grown, especially lower half of the distribution Male mean annual work hours steadily declined over the period, dropping about 10% between 1975 and 2018 (Figure 2b). Mean female hours increased over the series, growing by about 40% between 1975 and 2000, and stabilizing at around 1200 hours per year thereafter (about 60% of full-time equivalent). Female work hours were about 50% of that of males in 1975, which increased to 75% by the mid 1990s, and to 80% of male hours by the end of the series (in part due to declines in male hours). Growth in female work hours was also much stronger than males in the lower half of the annual hours distribution (Figure 2d). Characterizing growth in work hours by household earnings percentile reveals very strong growth for female hours in the middle of the distribution (Figure 2f).

Changes in household employment were less dramatic. Mean male household employment dropped by 12% across the series, while mean female employment grew 8% (figures omitted).

4.2 Counterfactual density estimates

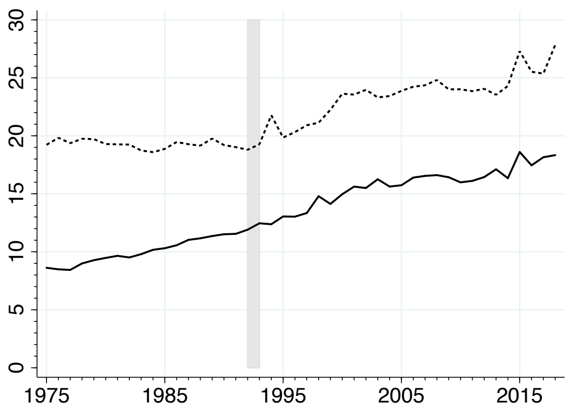
Changes in male wages have been the primary contributing factor to earnings inequality growth

Male hourly wages explain 47% of the increase in the earnings Gini index between 1975 and 2018 (0.028 of the observed 0.032 increase; Table 2). Keep in mind that these estimates take into account any potential artificial increase in the Gini due to the survey redesign in 1993 (see the notes in Table 2). Male wages also contributed substantially to the rise in the 90th percentile (19%) and the median (14%), but notably no significant contribution to the 10th percentile. So, although changes to male wages contributed to a widening of the earnings distribution, the process also contributed to important earnings gains overall.

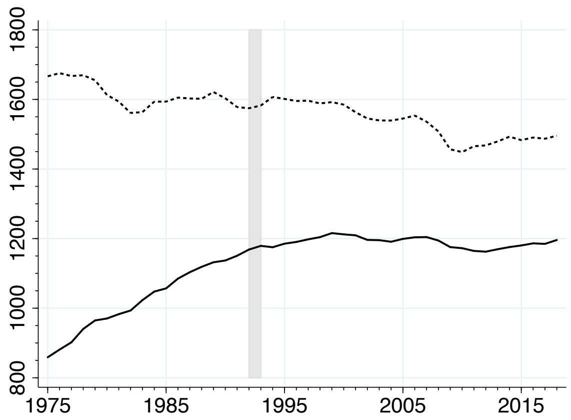
If we look at more nuanced decompositions over 10-year time periods, it becomes apparent the distribution male wages did not develop a strongly unequalizing effect until the late 1990s. Between 1975 and 1985, male wages actually reduced inequality: under the 1975 counterfactual, the Gini index would have been 12% higher–a finding overlooked by other studies (e.g. Burtless, 1999; Daly and Valletta, 2006). Male wages explains 57% of the growth in inequality between 1985 and 1995, and a full 98% of the increase between 1995 and 2005.

The counterfactual time-trend Gini plots make this point clear: holding 1975 male wages constant creates a time-trend that diverges markedly from the observed Gini starting around 1995 (Figure 3a). This corresponds precisely to the rapid growth in earnings and wages over the same time period (Figures 1b–1d, Figure 2a). By 2000, the 1975 male wage distribution would have reduced inequality to observed levels of 1985, and similarly would have reduced the 2018

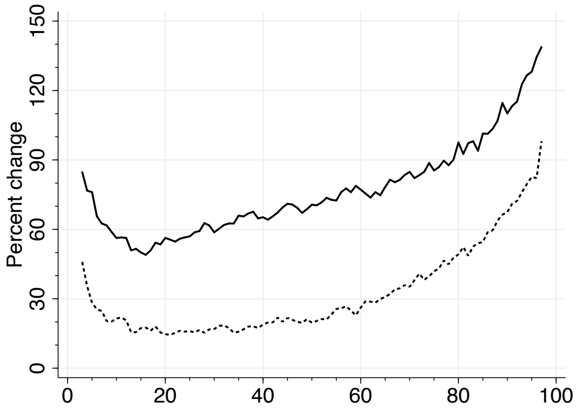
Figure 2: Time trends and percentage change for wages and hours, 1975-2018



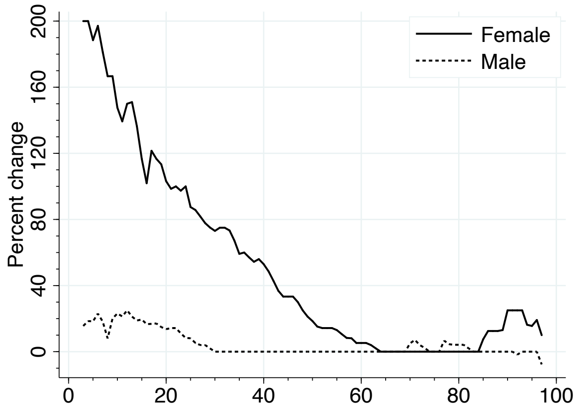
(a) Real hourly wage



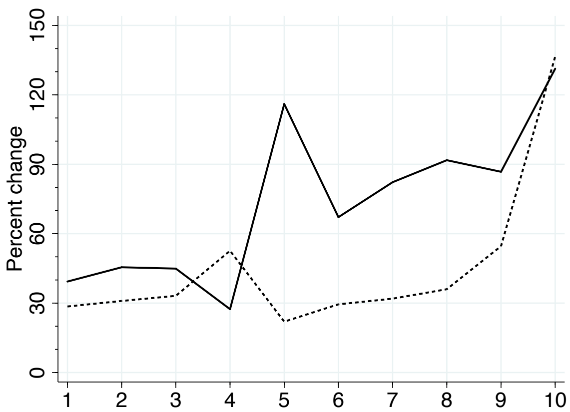
(b) Annual work hours



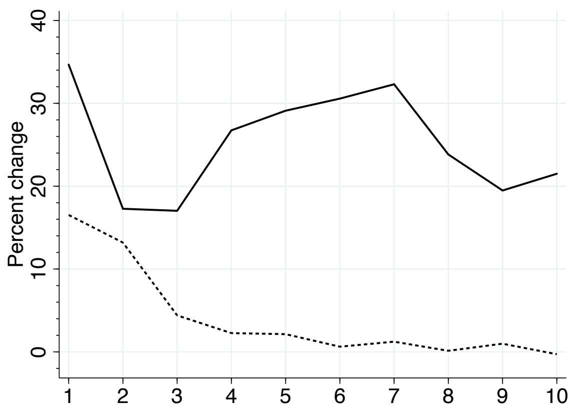
(c) Percent change in hourly wage



(d) Percent change in work hours



(e) Percent change in hourly wage,
by earnings decile



(f) Percent change in work hours,
by earnings decile

Notes: Hourly wages and work hours refer to household heads. Note that flat growth for males above the 30th percentile corresponds to households that report working 2080 hours (full-time equivalent) in both time periods. In Figures 2e-2f, hourly wages and work hours are averaged by earnings deciles in their respective year before percentage change is calculated. Samples are limited to households with positive earnings and with at least one household head working more than 5% of full-time equivalent. All monetary values are adjusted to 2018 dollars.

Table 2: Decomposition of changes in earnings densities, 1975-2018

	Gini			***p10***		
	Δ 1975-2017	Δ exp.	% exp.	Δ 1975-2017	Δ exp.	% exp.
Actual	0.061	-	-	7127	-	-
Male hours	0.058	0.003	5	7127	0	0
Female hours	0.074	-0.013	-22	3327	3800	53
Male wages	0.032	0.028	47	7127	0	0
Female wages	0.060	0.000	0	7127	0	0
Male emp.	0.060	0.001	1	7127	0	0
Female emp.	0.060	0.001	1	7127	0	0
	p50			***p90***		
	Δ 1975-2017	Δ exp.	% exp.	Δ 1975-2017	Δ exp.	% exp.
Actual	21780	-	-	85894	-	-
Male hours	20780	1000	5	83894	2000	2
Female hours	17780	4000	18	79894	6000	7
Male wages	18780	3000	14	69894	16000	19
Female wages	21032	748	3	84894	1000	1
Male emp.	21780	0	0	85894	0	0
Female emp.	20980	800	4	84898	996	1

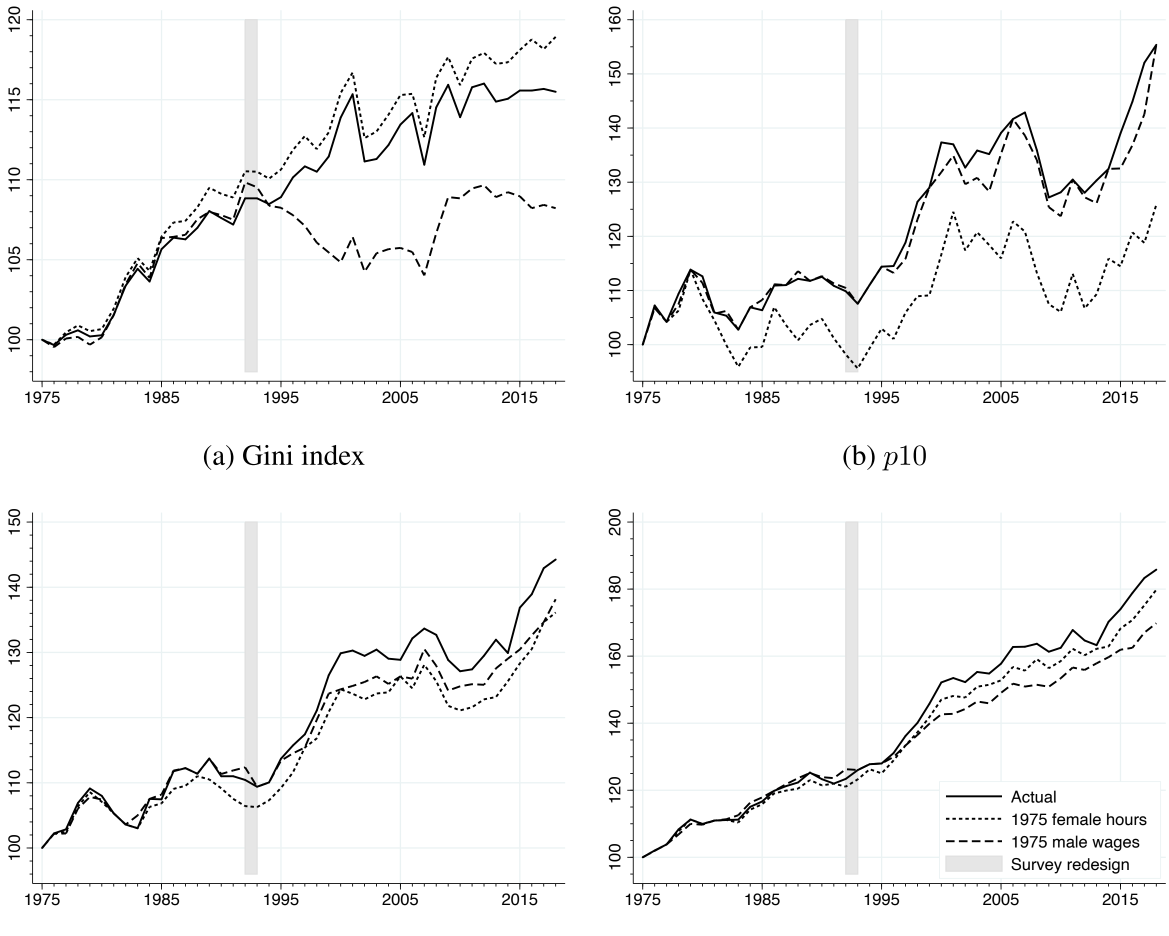
Notes: The left column indicates the counterfactual characteristic x fixed at 1975 levels. Change explained (Δ exp.) is the actual index minus counterfactual index, e.g. $I_{2018} - I_{2018,x1975}$. Percentage explained (% exp.) is change explained as a percentage of actual change, e.g. $100 \times (I_{2018} - I_{2018,x1975})/(I_{2018} - I_{1975})$. Negative results indicate that the counterfactual index was higher than the actual index in the final year, e.g. 2018 (assuming a positive increase in the actual measure, which is always the case). The Gini index from 1975-1992 was adjusted by adding an offset of 0.0228. For each counterfactual estimate, ceteris paribus controls include all other covariates and controls listed in Section 2. Samples are limited to households with positive earnings and with at least one household head working more than 5% of full-time equivalent. All monetary values are adjusted to 2018 dollars.

Table 3: Decomposition of changes in earnings Gini index, 10-year periods

	Δ ***1975-1985***	Δ exp.	% exp.	Δ ***1985-1995***	Δ exp.	% exp.
Actual	0.022	-	-	0.013	-	-
Male hours	0.022	0.000	-1	0.013	0.000	2
Female hours	0.025	-0.003	-14	0.015	-0.003	-20
Male wages	0.025	-0.003	-12	0.006	0.007	57
Female wages	0.022	0.000	0	0.011	0.002	13
Male emp.	0.022	0.000	-1	0.011	0.001	10
Female emp.	0.022	0.000	-1	0.012	0.000	3
	Δ ***1995-2005***	Δ exp.	% exp.	Δ ***2005-2018***	Δ exp.	% exp.
Actual	0.018	-	-	0.008	-	-
Male hours	0.018	0.000	0	0.008	0.000	-2
Female hours	0.020	-0.003	-14	0.009	-0.001	-14
Male wages	0.000	0.017	98	0.006	0.002	25
Female wages	0.019	-0.001	-5	0.006	0.002	20
Male emp.	0.017	0.000	2	0.008	0.000	0
Female emp.	0.017	0.001	5	0.009	-0.001	-9

Notes: Left column indicates the counterfactual characteristic x fixed at the initial year (the first year of the 10-year period). Change explained (Δ exp.) is the actual index minus counterfactual index, e.g. $I_{2018} - I_{2018,x1975}$. Percentage explained (% exp.) is change explained as a percentage of actual change, e.g. $100 \times (I_{2018} - I_{2018,x1975})/(I_{2018} - I_{1975})$. Negative results indicate that the counterfactual index was higher than the actual index in the final year, e.g. 2018 (assuming a positive increase in the actual measure, which is always the case). The Gini index from 1975-1992 was adjusted by adding an offset of 0.0228. For each counterfactual estimate, ceteris paribus controls include all other covariates and controls listed in Section 2. Samples are limited to households with positive earnings and with at least one household head working more than 5% of full-time equivalent. All monetary values are adjusted to 2018 dollars.

Figure 3: Distributional measures for counterfactual household characteristics



(a) Gini index

(b) $p10$

(c) Median

(d) $p90$

Notes: Measures are for the indicated year with counterfactual characteristics held at their 1975 distribution. Values are expressed as a percentage of the 1975 level. The Gini index from 1975-1992 was adjusted by adding an offset of 0.0228. Samples are limited to households with positive earnings and with at least one household head working more than 5% of full-time equivalent. All monetary values are adjusted to 2018 dollars.

level of inequality to that of 1992. This result would have come about by reducing earnings at the median and 90^{the} percentile, rather than increasing earnings at the 10th percentile, as evident from the attenuated counterfactual male wage percentile plots (Figures 3b–3d).

Changes in the female wage distribution only contributed substantially to inequality growth between 2005 and 2018, during which explains 20% of the increase in the Gini index.

Changes in female work hours have had the largest mitigating effect on earnings inequality growth

The rapid growth in female work hours, particularly in the lower and middle portions of the earnings distribution (recall Figure 2f), had an unambiguous mitigating effect on inequality growth between 1975 and 2018. Under the counterfactual of the 1975 distribution of female work hours, the Gini index would have grown 22% *higher* than the observed growth between 1975 and 2018 (Table 2). This is apparently because changes to female hours explain a large portion of rising household earnings in the lower percentiles of the distribution—and yet contribute modestly to the middle and upper percentiles as well. The growth of female hours explains a full 53% of the rise in the 10th percentile of earnings, 18% of the rise in median, and 7% of the rise in the 90th percentile.

The equalizing effect of increasing female work hours is remarkably consistent across annual samples, reducing the Gini index by between 14% and 20% for each of the shorter 10-year (and one 13-year) decomposition periods (Table 3). Counterfactual time series plots of these measures make this phenomenon clear: under the 1975 distribution of female hours, the Gini would have been 10-20% higher as early on as 1986, and would have continued on a divergent trend of higher inequality into 2018 (Figure 3a). Counterfactual female hours also depresses the 10th percentile trend line by a visible 50% from 1980 to the end of the series (Figure 3b), and depresses the other percentile measures to a lesser extent.

Changes to distribution of male hours, and male and female employment, has had a much smaller influence on the earnings distribution. This follows, as these characteristics underwent relatively less change over the period.

5 Conclusion

In this study I decompose changes in U.S. household earnings inequality between 1975 and 2018 using a semi-parametric method to simulate counterfactual earnings densities. The method allows for the joint quantification price and quantity effects, and their symmetrical male and female components.

Results reveal that changes male hourly wages (the price effect) have been the primary contributing factor to earnings inequality growth between 1975 and 2018. Male wages actually mitigate inequality growth between 1975 and 1985, but this quickly reverses, and in fact, from 1995 to 2005 male wages explain 98% of the growth in earnings inequality. On the other hand, the steady growth in female work hours over the past few decades (the quantity effect) has had a strong mitigating effect on inequality growth, doing so by raising earnings in the lower and mid portions of the earnings distribution. This latter result is remarkably consistent in the full-period decomposition, in each of the four short-term decompositions, and in plots of counterfactual statistics over time.

These results not only assemble the existing piecemeal findings in the literature into a unified decomposition framework, with directly comparable percentage contributions for each labor market component, but also add the crucial quantification of work hours—which up to this point has been missing from inequality decomposition studies.

References

Anonymous (2019). Title blinded for review. *Journal title blinded for review*.

Anselin, L. (2013). *Spatial Econometrics: Methods and Models*, Volume 4. Springer Science & Business Media.

Autor, D. H., L. F. Katz, and M. S. Kearney (2005). Rising wage inequality: the role of composition and prices. *National Bureau of Economic Research*.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006). The polarization of the US labor market. *The American Economic Review 96*(2), 189–194.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008). Trends in US wage inequality: Revising the revisionists. *The Review of Economics and Statistics 90*(2), 300–323.

Biewen, M. (2001). Measuring the effects of socio-economic variables on the income distribution: An application to the East German transition process. *Review of Economics and Statistics 83*(1), 185–190.

Burkhauser, R. V., S. Feng, S. P. Jenkins, and J. Larrimore (2012). Recent trends in top income shares in the United States: reconciling estimates from March CPS and IRS tax return data. *Review of Economics and Statistics 94*(2), 371–388.

Burtless, G. (1999). Effects of growing wage disparities and changing family composition on the US income distribution. *European Economic Review 43*(4), 853–865.

Cancian, M. and D. Reed (1998). Assessing the effects of wives' earnings on family income inequality. *Review of Economics and Statistics 80*(1), 73–79.

Cancian, M. and D. Reed (1999). The impact of wives earnings on income inequality: Issues and estimates. *Demography 36*(2), 173–184.

Chevan, A. and R. Stokes (2000). Growth in family income inequality, 1970–1990: Industrial restructuring and demographic change. *Demography 37*(3), 365–380.

Cohany, S. R., A. E. Polivka, and J. M. Rothgeb (1994). Revisions in the Current Population Survey effective January 1994. *Employment & Earnings 41*, 13.

Daly, M. C. and R. G. Valletta (2006). Inequality and poverty in United States: the effects of rising dispersion of men's earnings and changing family behaviour. *Economica 73*(289), 75–98.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica 64*(5), 1001–1044.

Fields, G. S. (2003). Accounting for income inequality and its change: A new method, with application to the distribution of earnings in the United States. *Research in Labor Economics 22*(3), 1–38.

Flood, S., M. King, R. Rodgers, S. Ruggles, and J. R. Warren (2018). *Integrated Public Use Microdata Series, Current Population Survey: Version 6.0. (dataset).* Minneapolis, MN: IPUMS, 2018. https://doi.org/10.18128/D030.V65.0. Cited 10 October 2018.

Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4, pp. 1–102. Elsevier.

Gottschalk, P. and S. Danziger (2005). Inequality of wage rates, earnings and family income in the United States, 1975–2002. *Review of Income and Wealth 51*(2), 231–254.

Heathcote, J., F. Perri, and G. L. Violante (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006. *Review of Economic Dynamics 13*(1), 15–51.

Karoly, L. A. and G. Burtless (1995). Demographic change, rising earnings inequality, and the distribution of personal well-being, 1959–1989. *Demography 32*(3), 379–405.

Katz, L. F., D. Autor, O. Ashenfelter, and D. Card (1999). Changes in the wage structure and earnings inequality. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3. Elsevier.

Larrimore, J., R. V. Burkhauser, S. Feng, and L. Zayatz (2008). Consistent cell means for topcoded incomes in the public use March CPS (1976–2007). *Journal of Economic and Social Measurement 33*(2), 89–128.

Lerman, R. I. and S. Yitzhaki (1985). Income inequality effects by income source: a new approach and applications to the United States. *The Review of Economics and Statistics 67*(1), 151–156.

Machado, J. A. and J. Mata (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics 20*(4), 445–465.

McLaughlin, D. K. (2002). Changing income inequality in nonmetropolitan counties, 1980 to 1990. *Rural Sociology 67*(4), 512–533.

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics 12*(4), 577–590.

Nielsen, F. and A. S. Alderson (1997). The kuznets curve and the great U-turn: income inequality in US counties, 1970 to 1990. *American Sociological Review 62*(1), 12–33.

Polivka, A. E. and J. M. Rothgeb (1993). Redesigning the CPS questionnaire. *Monthly Labor Review 116*(9), 10–28.

Reed, D. and M. Cancian (2001). Sources of inequality: Measuring the contributions of income sources to rising family income inequality. *Review of Income and Wealth 47*(3), 321–333.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games 2*(28), 307–317.

Shorrocks, A. F. (1983). The impact of income components on the distribution of family incomes. *The Quarterly Journal of Economics 98*(2), 311–326.

Thompson, J. P. and T. M. Smeeding (2013). Inequality and poverty in the United States: the aftermath of the Great Recession. *FEDS Working Paper*.

U.S. Bureau of Economic Analysis (2018). *Personal Consumption Expenditures (PCE)*. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/PCE. Cited 27 July 2018.