Sources of Wage Inequality:

Decomposing the Conditional Gini Coefficient

Carlos Hurtado♣
Nov 11, 2023

Abstract

This paper introduces a new econometric method to identify factors influencing the disparities within the distribution of a positive random variable, focusing on US wages between 1986 and 2015. I relate the conditional Lorenz curve to the conditional quantile function to additively decompose the conditional Gini index. Moreover, this paper presents a technique to disentangle the temporal changes in the distribution. The analysis shows that despite reduced impacts of race and gender on wages, persistent disparities require ongoing intervention, while higher education, especially college degrees, significantly reduces wage inequality during the analysis period.

Keywords: Gini Index, Wage Structure, Inequality, Quantile Regression

JEL Classification Numbers: C13, C21, C43, D63

♣ Economics Department, Robins School of Business, 102 UR Drive, University of Richmond, VA 23173, USA. Email: churtado@richmond.edu. Ph; 713-933-9027. The author would like to thank Roger Koenker for his advise and guidance throughout this research project. I thank comments and suggestions from Mark Borgschulte, George Deltas, Bhash Mazumder, Elizabeth Powers, and seminar participants at the Midwest Graduate Student Summit, LACEA-LAMES 2016, and the HCEO 2017 Summer School of Socioeconomic Inequality

1 Introduction

Since the early 1980s, the escalating issue of economic inequality in the US has increasingly captured scholarly attention. [1] An intricate interplay of sociodemographic attributes, individual endowments, and the returns on these attributes and endowments fundamentally shapes the final distribution of income and wages. However, it is crucial to recognize that these characteristics and endowments are interconnected, and their combined influence on economic inequality is complex and elusive. Consequently, applying advanced statistical techniques is indispensable for disentangling this complexity and precisely measuring the impact of these interconnected factors on inequality. While extensive research has focused on identifying sources of economic disparities by estimating income or wage distribution densities, a substantial knowledge gap persists in understanding how *inequality measures* respond to these multifaceted influencing factors.

In this study, I introduce a novel econometric methodology to systematically assess the influence of diverse determinants on the disparities within the distribution of a positive random variable, explicitly focusing on wages. I develop a method to decompose the conditional Gini index, a widely used inequality measure, by using conditional quantile regressions and leveraging the relationship between the Lorenz curve and the quantile function.[2] This approach enables the precise identification and quantification of the factors most significantly impacting distribution inequality. Using data from the Ongoing Rotation Group (ORG) of the Current Population Survey (CPS) for 1986 and 2015, I illustrate this

---

[1] See Levy and Murnane (1992), Katz (1999), Autor et al. (2008) , Guvenen et al. (2014), and Abel and Deitz (2019) for a review of the literature.

[2] The Gini coefficient is a measure of income distribution inequality within a population, ranging from 0 ("perfect equality") to 1 ("perfect inequality"). Preferred for its simplicity and ability to facilitate cross-population comparisons, it doesn't require parametric assumptions based on normative views, unlike some other inequality measures. For instance, the Atkinson Index depends on a parameter that reflects the society's aversion to inequality. This parameter essentially shapes how the index evaluates income distributions, making Atkinson's measure sensitive to the degree of inequality a society is willing to tolerate. Similarly, the Theil index and the Generalized Entropy measurements are subject to the selection of a parameter that assigns a weight to distances between incomes in different parts of the income distribution; the choice of the parameter, like in the Atkinson Index, relates to the society's aversion to inequality. See Hufe et al. (2020) for more contemporaneous measurements of inequality.

methodology and propose a decomposition of wage distribution changes, employing counterfactual scenarios based on estimated conditional Gini coefficients.

Building on seminal contributions by Oaxaca (1973) and Blinder (1973), the literature has consistently advanced methodologies for disentangling the intricate factors that underlie wage disparities.[3] The classical labor supply and demand model, featuring homogeneous agents, implicitly assumes a unique wage that clears the market.[4] However, recent developments have broadened our analytical horizons, prompting a shift towards a more profound examination of wage inequality across the entire distribution. For instance, using quantile regression, Buchinsky (1994) shows that the positive impact of education on wages is more pronounced for individuals in the upper echelons of the wage distribution compared with those at the lower end.[5] More recently, Bayer and Charles (2018) find that black men at higher percentiles, particularly those with college educations, have experienced significant advances in relative earnings; this improvement is primarily attributed to positional gains among higher-educated black individuals.

The progression of methodologies designed to model the entire wage distribution fuels optimism for a deeper understanding of the factors directly shaping the distribution.[6] Machado and Mata (2005) notably contribute to this endeavor by developing a counterfactual decomposition technique using conditional quantile regression. Their study estimates marginal (log) wage distributions in alignment with a conditional distribution derived from conditional quantile regressions. These estimations are a foundation for

---

[3] For a review of many of the decomposition methods, please refer to Fortin et al. (2011). For example, Kleven et al. (2019) use these decomposition methods to understand the gender wage gaps in Denmark.

[4] Following the classical model, part of the literature has focused on the average wage differences controlling for individual and institutional characteristics. See, for example, Katz and Murphy (1992), Bound and Johnson (1992), Blau and Kahn (1996), Card and Lemieux (2001)

[5] Angrist et al. (2006) find a similar result for a more contemporary subsample of the US population. Arellano and Bonhomme (2017) show similar findings for the UK.

[6] The first method that models the distribution of wages is in DiNardo et al. (1996). The authors developed an estimation procedure to analyze counterfactual (log) wage distributions using kernel density methods to appropriately weighted samples.

constructing counterfactual scenarios, where they compare the implied marginal wage distributions based on the assumed covariate distributions. Their method explores wage differentials at various individual quantiles to try to assess the dynamics of the entire wage distribution.

Traditional methods for analyzing inequality, such as visually inspecting kernel density estimates, focusing on isolated quantiles, or using ratios between quantiles, can overlook detailed dynamics and fail to reveal complex relationships within continuous parts of wage distributions. My proposed method addresses this by examining all quantiles of the distribution, and leveraging the inherent connection between the conditional quantile function and the conditional Lorenz curve.[7] This approach allows for a thorough analysis of intricate dynamics throughout the entire wage distribution, not just isolated points. This approach directly assesses the impact of log wage distribution determinants on inequality by connecting them with the conditional Gini index, thus providing a precise measure of inequality, bypassing the need for density modeling.

I employ comprehensive hourly wage data from the CPS ORG from 1986 to 2015 to exemplify the empirical application of my proposed methodology and contrast it with the approach of Machado and Mata (2005). I model the conditional quantile function of the logarithm of wages using a diverse array of individual, job-related, and demographic factors.[8] Several key findings emerge from my analysis. The manufacturing sector, which initially correlated with lower wages in 1986, shifted to increasing wages by 2015.

[7] A conditional quantile function estimates the value at a specific percentile within a subgroup of data under certain conditions. For instance, it might tell us the median earnings for college-educated individuals in a population. This function is closely linked to conditional quantile regression, which is a statistical technique used to predict the conditional quantile of a response variable. Conditional quantile regression models the relationship between a set of predictor variables and specific quantiles of the response variable, allowing for a nuanced understanding of how this relationship varies across different parts of the distribution.

[8] The model integrates individual controls (experience, urban living, education), job-related attributes (unionization, public sector employment, manufacturing job, part-time status), and demographic indicators (nonwhite, female, marital status). Crucially, it incorporates state and industry fixed effects, essential for capturing macroeconomic shifts and sectoral composition changes. These fixed effects are vital for accounting for global trends such as trade liberalization and technological advancements, and regional economic shifts.

Concerning race and gender, despite the persistence of negative impacts, a notable decline in their adverse effects was observed. However, the continuous presence of these disparities underscores the need for targeted interventions. Finally, higher education, particularly college education, emerged as a significant factor in reducing wage inequality. I find that an increase in the proportion of college-educated workers led to a substantial decrease in the conditional Gini index, highlighting the value of higher education in addressing wage disparities.

I implement the Machado and Mata (2005) method alongside my proposed technique to demonstrate the advantages of the later. My method reveals critical insights missed by the Machado and Mata approach. It uncovers a notable reduction in wage inequality at the lower wage spectrum, effectively balancing the rise at the higher end. This dual effect offers a more transparent view of the shifts in wage inequality over time. My method also detects the considerable influence of unionization and manufacturing across all wage levels, highlighting key economic shifts. Furthermore, it shows that urbanization consistently reduces wage inequality. This approach surpasses the Machado and Mata (2005) method by providing a holistic analysis of wage dynamics, encompassing various income levels and socio-economic factors, and particularly underscores the critical role of education in mitigating wage disparities.

The structure of this paper unfolds as follows. In Section 2, I establish the theoretical linkage between the conditional Lorenz curve and the conditional Gini index, elucidating the additive decomposition of the coefficient and its capacity to capture temporal changes in the distribution. Section 3 introduces the proposed estimation procedure, outlining its methodology and implementation. Section 4 offers an extensive account of the US hourly wage data employed in this analysis. Moving to Section 5, I delve into the empirical application, demonstrating the practicality and insights generated using the proposed method. Finally, Section 6 concludes.

2 Conditional Lorenz Curve and Gini Index

The Lorenz curve is a powerful instrument for illustrating the inequality present in the distribution of a positive random variable. Notably, this curve depicts the cumulative share of wages earned relative to the cumulative percentage of individuals, ranging from the lowest to the highest earners, thereby facilitating an insightful analysis of wage inequality. Adhering to the conceptual framework outlined by Koenker (2005), I define the Lorenz curve as follows:

$$L(\tau) = \frac{\int_0^\tau Q_Y(t)dt}{\int_0^1 Q_Y(t)dt} = \frac{1}{\mu}\int_0^\tau Q_Y(t)dt, \tag{1}$$

where Y is a continuous and positive random variable with a cumulative density function $F_Y(y)$, quantile function denoted as $Q_Y(t) = \inf\{y: F_Y(y) \geq t\} = F^{-1}(t)$, with $y_\tau = Q_Y(\tau)$ and mean μ satisfying $0 < \mu < \infty$. As delineated in Appendix A, the application of a monotonic transformation, denoted $h(\cdot)$, which satisfies $h(Y) \geq 0$ and $0 < \mu_h < \infty$, where $\mu_h = E[h(y)]$, culminates in a Lorenz curve of the *transformed* variable as given by

$$L_h(\tau) = \frac{1}{\mu_h}\int_0^\tau Q_{h(Y)}(t)dt = \frac{\tau E[h(y)|h(y) \leq h(y_\tau)]}{\mu_h}. \tag{2}$$

Drawing upon the fact that $0 \leq E[h(y)|h(y) \leq h(y_\tau)] \leq E[h(y)] = \mu_h$, and considering $\tau \in (0,1)$, it is clear that the Lorenz curve of the transformed variable lies between zero and one.

Let us consider $Q_{h(Y)}(t|x)$, with $t \in (0,1)$, to represent the t-th conditional quantile of the distribution of $h(Y)$, given a vector of covariates denoted by $x \in R^P$. I propose modeling this conditional quantile function as a linear combination of the covariates, illustrated in the Equation below:

$$Q_{h(Y)}(t|x) = x^T\beta(t) = \sum_{j=1}^{P} x_j\, \beta_j(t), \tag{3}$$

where each $\beta_j(t)$ is the coefficient aligned with j-th covariate at the t-th quantile. Next, let $\lambda_h(\tau) \in R^P$ be a vector where its j-th element is defined as $\lambda_{h,j}(\tau) = \frac{1}{\tau}\int_0^\tau \beta_j(t)dt$, representing, in essence, the mean of the j-th coefficient within the interval $(0, \tau)$[9]. From equations (2) and (3), the conditional Lorenz curve of the transformed variable is expressed as:

$$\mathrm{L_h}(\tau|x) = \frac{1}{\mu_h}\int_0^\tau Q_{h(Y)}(t|x)dt = \frac{1}{\mu_h}\sum_{j=1}^{P} x_j \int_0^\tau \beta_j(t)dt = \frac{\tau x^T \lambda_h(\tau)}{\mu_h}. \tag{4}$$

By comparing equations (2) and (4), it becomes clear that $E\big[h(y)|x \wedge \big(h(y) \leq h(y_\tau)\big)\big]$ equates to $x^T\lambda_h(\tau)$. By taking the limit when τ goes to one, I deduce that $E[h(y)|x]$ is given by $x^T\lambda_h(1) = x^T \int_0^1 \beta(t)dt$, provided that the integral exists for each characteristic j.

Based on the Lorenz curve, the Gini coefficient is widely used to summarize the disparity of the distribution of a positive random variable. The relationship between the coefficient and the curve is defined by the following Equation:

$$G = 1 - 2\int_0^1 L(\tau)\, d\tau, \tag{5}$$

[9] This convention implies that $\int_0^\tau \beta_j\,(t)dt = \tau\lambda_{h,j}(\tau)$, and $\lambda_h(\tau) = \frac{1}{\tau}\left(\int_0^\tau \beta_1(t)dt\,, \cdots, \int_0^\tau \beta_P(t)dt\right) = \frac{1}{\tau}\int_0^\tau \beta(t)dt$.

where G represents the Gini index value; this index quantifies the degree of deviation of a given random variable's Lorenz curve from the line that indicates perfect equality[10].

I compute the conditional Gini coefficient given a vector of covariates by integrating the conditional Lorenz curve, delineated in Equation (4), into the Gini index's definition. The formulation is given by

$$\begin{aligned} G_h(x) &= 1 - 2\int_0^1 L_h(\tau|x)d\tau \\ &= 1 - \frac{1}{\mu_h}\sum_{j=1}^{P} x_j \int_0^1 \int_0^\tau 2\beta_j(t)dt\,d\tau, \end{aligned} \tag{6}$$

where $x \in R^P$. This Equation (6) constitutes an additive decomposition of the conditional Gini index. This analytical tool is invaluable in scrutinizing the progression of variations in the distribution of $h(Y)$, contingent on the factor endowments and sociodemographic characteristics, x_j, as well as the returns (prices) associated with these endowments and characteristics, $\frac{1}{\mu_h}\int_0^1 \int_0^\tau 2\beta_j(t)dt\,d\tau$.

It is interesting to mention that the coefficient can be reformulated by partitioning the interval (0,1) into n equally spaced sub-intervals, as illustrated by the Equation

$$G = 1 - \sum_{i=0}^{n-1} 2 \int_{\tau_i}^{\tau_{i+1}} L(\tau)d\tau, \tag{7}$$

[10] The line of perfect equitability is the Lorenz curve of a degenerate random variable δ_μ, which only takes the single value μ.

where $\tau_i = \frac{i}{n}$, for $i = 0, \cdots, n-1$. By definition, $\tau_{i+1} - \tau_i = \frac{1}{n}$, and noting that the area beneath the line of perfect equitability can be expressed in terms of rectangles and triangles, it becomes evident that

$$G = 2\left[\sum_{i=0}^{n-1}\left(\frac{i}{n^2} + \frac{1}{2n^2} - \int_{\tau_i}^{\tau_{i+1}} L(\tau)d\tau\right)\right]. \tag{8}$$

I wish to emphasize this connection as it provides a direct method to numerically approximate the Gini coefficient, making it a flexible tool in inequality analysis.

What is interesting about Equation (8) is its capability to pinpoint the sub-intervals most significantly contributing to the Gini index — essentially highlighting the quantiles that primarily augment the inequality within the distribution. Finally, combining equations (6) and (8), I re-express the coefficient, given a vector of covariates, as

$$\begin{aligned} G_h(x) &= 1 - \sum_{i=0}^{n-1}\sum_{j=1}^{P} x_j \frac{1}{\mu_h}\int_{\tau_i}^{\tau_{i+1}}\int_0^{\tau} 2\beta_j(t)dt\,d\tau \\ &= 2\left[\sum_{i=0}^{n-1}\left(\frac{i}{n^2} + \frac{1}{2n^2} - \sum_{j=1}^{P} x_j \frac{1}{\mu_h}\int_{\tau_i}^{\tau_{i+1}}\int_0^{\tau} \beta_j(t)dt\,d\tau\right)\right]. \end{aligned} \tag{9}$$

In this Equation, I am considering the additive decomposition of the conditional coefficient in relation to both the endowments and characteristics, x_j, as well as the specific sub-intervals that play a significant role in escalating the inequality of the distribution. This analysis might pave the way for a deeper understanding of the underlying factors driving income or wealth disparities in various economic settings.

2.1 Impact of Individual Characteristics on the Conditional Gini Index

I use Equation (6) to compute the variation in the conditional Gini index, resulting from a *small* positive change in a characteristic j from x_j to x_j', as shown below:

$$\frac{\Delta G_h(x)}{\Delta x_j} = \frac{G_h\left(x_j', x_{-j}\right) - G_h\left(x_j \ , x_{-j}\right)}{x_j' - x_j} = -\frac{1}{\mu_h}\int_0^1\int_0^\tau 2\beta_j(t)dt\,d\tau \stackrel{\text{def}}{=} -\frac{\Pi_j}{\mu_h}, \tag{10}$$

where $x = \left(x_j, x_{-j}\right) = (\,x_1, \cdots, x_j, \cdots, x_p) \in R^P$. I assume $\mu_h > 0$, which indicates that the direction of the change in the Gini coefficient exclusively relies on the sign of

$$\Pi_j = \int_0^1\int_0^\tau 2\beta_j(t)dt\,d\tau.$$

If Π_j is negative, a marginal positive shift in covariate j correlates with an uptick in the conditional Gini index, thereby pointing to augmented inequality in the distribution of $h(Y)$. Conversely, a slight positive adjustment in covariate j, paired with a positive Π_j, correlates with a decrease in the inequality found in the distribution of $h(Y)$.

Furthermore, μ_h serves as a positive scaling parameter that normalizes the Lorenz curve and the Gini coefficient within the range of zero to one. Consequently, the magnitude of Π_j imparts information about the extent of the alteration in the Gini index, following a slight positive adjustment in covariate j. Larger absolute values of Π_j correspond with more pronounced shifts in the Gini coefficient, also in absolute terms. I label the absolute value of $\frac{\Pi_j}{\mu_h}$ as the impact of covariate j on the distribution of $h(Y)$. Under this premise, it is evident that certain covariates exert a more substantial impact on the distribution of $h(Y)$ compared to others.

2.2 Temporal Changes in the Distribution of $h(Y)$

I aim to grasp the intricacies of the distribution of $h(Y)$ to dissect the influence of various factors on the *distributional changes over time*. This decomposition bears practical implications as it facilitates distinguishing between the effects stemming from shifts in individual characteristics and alterations in the returns to those attributes. Previous decomposition approaches also used this kind of analysis; for instance, DiNardo et al. (1996) apply kernel density methods on reweighted samples to scrutinize counterfactual wage distributions. Similarly, Machado and Mata (2005) craft a counterfactual decomposition technique leveraging quantile regression, a strategy that aligns closely with the one I introduce here. In all scenarios, including mine, the decomposition broadens the Oaxaca (1973) method, initially forged to investigate counterfactual disparities in average earnings.

Suppose I aim to scrutinize the changes in the distribution spanning two years, represented by $\Psi \in \{0,1\}$. I intend to explore two varieties of counterfactual scenarios. Firstly, I aspire to gauge the inequality in the distribution of $h(Y)$ in year $\Psi = 1$, aligned with the distribution of covariates in year $\Psi = 0$. Concurrently, I seek to calculate the disparity in the distribution of $h(Y)$ in year $\Psi = 1$, under the condition that only one covariate embodies the distribution witnessed in year $\Psi = 0$. By deploying these counterfactuals, I can understand the impacts on the distribution of $h(Y)$, which arise from changes in the covariates and shifts in the returns attributed to these covariates.

Let us model the conditional quantile function in year Ψ as

$$Q_{h(Y)}(t|x;\Psi) = x^T \beta_\Psi(t), \tag{11}$$

where $\beta_\Psi(t)$ represents the coefficients of the covariates in year Ψ at quantile t, and x is a vector of covariates. Now, let $X(\Psi)$ denote an $N_\Psi \times P$ matrix of data on these covariates in year Ψ with N_Ψ denoting the number of observations and P signifying the number of covariates. Also, denote by $\bar{X}_j(\Psi)$ the average of column j of the matrix $X(\Psi)$. Using the additive decomposition of the Gini coefficient in Equation (6), I propose an estimate for the conditional Gini index in year Ψ, expressed as

$$\hat{G}_h^{\Psi} = 1 - \sum_{j=1}^{P} \bar{X}_j(\Psi) \frac{\widehat{\Pi}_j^{\Psi}}{\hat{\mu}_h^{\Psi}}, \tag{12}$$

where $\hat{\mu}_h^{\Psi}$ and $\widehat{\Pi}_h^{\Psi}$ are the respective estimates for μ_h and Π_j in year Ψ.

Despite the potential for general equilibrium effects stemming from alterations in the distribution of the covariates—given that these changes can influence the returns to the characteristics— let me assume for simplicity that the changes in the covariates do not modify the returns of those characteristics[11]. Considering this assumption, I can compute an estimate of the conditional Gini index in year $\Psi = 1$, assuming all covariates were distributed as in year $\Psi = 0$, as follows:

$$\hat{G}_h^{1}\big(X(0)\big) = 1 - \sum_{j=1}^{P} \bar{X}_j(0) \frac{\widehat{\Pi}_j^{1}}{\hat{\mu}_h^{1}}. \tag{13}$$

For this discussion, let G_h^{Ψ} denote the Gini index computed from a sample in year Ψ. I calculate changes in the Gini coefficient to capture shifts in the distribution of $h(Y)$:

$$\begin{aligned} G_h^1 - G_h^0 &= \hat{G}_h^1 - \hat{G}_h^0 + \text{residual} \\ &= \underbrace{\hat{G}_h^1 - \hat{G}_h^1\big(X(0)\big)}_{\text{change in covariates}} + \underbrace{\hat{G}_h^1\big(X(0)\big) - \hat{G}_h^0}_{\text{change in returns}} + \text{residual}. \end{aligned} \tag{14}$$

The change in the distribution tied to shifts in individual characteristics is measured by $\hat{G}_h^1 - \hat{G}_h^1\big(X(0)\big)$, where returns remain constant and only the covariates vary. On the other hand,

[11] This is an inherent assumption of the Oaxaca (1973) decomposition that is also present in DiNardo *et al.* (1996) and Machado and Mata (2005).

the shift in the distribution of $h(Y)$ triggered by changes in returns to individual traits is encapsulated by $\hat{G}_h^1\big(X(0)\big) - \hat{G}_h^0$, where covariates stay the same and only returns change.

I define $X_{-j}^1(0) = \big(\bar{X}_1(1), \cdots, \bar{X}_j(0), \cdots, \bar{X}_P(1)\big)$ as a vector in R^P, where the j-th entry represents the average of characteristic j in year $\Psi = 0$, while all other entries are the averages of the respective covariates in year $\Psi = 1$. To pinpoint the effect of a single covariate changing from year $\Psi = 0$ to $\Psi = 1$, I introduce the impact on the change in the distribution of $h(Y)$:

$$\hat{G}_h^1 - \hat{G}_h^1\left(X_{-j}^1(0)\right) = -\left(\bar{X}_j(1) - \bar{X}_j(0)\right)\frac{\widehat{\Pi}_j^1}{\hat{\mu}_h^1}. \tag{15}$$

This Equation assumes a specific sequence of changes from $\Psi = 0$ to $\Psi = 1$, an assumption that, admittedly, is arbitrary. It's also worth noting that understanding what would happen to the distribution at $\Psi = 0$ if all covariates were as in $\Psi = 1$ can provide alternative measures of the effects of changes in both covariate returns and the covariates themselves.

2.2.1 Alternative Method

I consider another way to understand changes in the distribution of $h(Y)$ through the method proposed by Machado and Mata (2005). This approach estimates the entire distribution to isolate contributing factors to temporal changes. The technique relies on the probability integral transformation theorem, stating that if U is a uniform random variable on [0,1], then $F^{-1}(U)$ has distribution F. I model the conditional quantile of $h(Y)$ in year Ψ as given by equation (11):

$$Q_{h(Y)}(t|x;\Psi) = x^T\, \beta_\Psi(t),$$

Subsequently, I employ the following steps to estimate the implied marginal densities:

1. Generate a random sample of size m from a uniform random variable on [0,1]: $u_1, \cdots, u_m$

2. Estimate $Q_{h(Y)}(t|x;\Psi)$ yielding m estimates $\hat{\beta}_{\Psi}(u_i)$.

3. Generate a random sample of size m with replacement from $X(\Psi)$, the $N_{\Psi} \times P$ matrix of data on covariates, denoted by$\{x_i^*(\Psi)\}_{i=1}^{m}$.

4. Generate a random sample of $h(Y)$ that is consistent with the conditional distribution defined by the model: $\left\{\eta_i^*(\Psi) \stackrel{\text{def}}{=} x_i^*(\Psi)^T \hat{\beta}_{\Psi}(u_i)\right\}_{i=1}^{m}$.

For generating a random sample from the marginal distribution of $h(Y)$ as it would have been in $\Psi = 1$ —assuming all covariates had been as in $\Psi = 0$—I use $X(0)$ in the third step above. This procedure assumes that covariate changes do not modify their returns.

To construct a counterfactual where only one covariate, $x_i(1)$, is distributed as in year $\Psi = 0$, I introduce an additional step based on the method by Machado and Mata (2005). The authors defined a partition of the covariate $x_i(1)$ in J classes, $C_j(1)$, with relative frequencies $f_j(\cdot)$, for $j = 1, \cdots, J$, and propose the following procedure:

1. Generate $\{\eta_i^*(1)\}_{i=1}^{m}$, a random sample of $h(Y)$, with size m, that is consistent with the conditional distribution defined by the model.

2. Take the first class, $C_1(1)$, and select all elements of $\{\eta_i^*(1)\}_{i=1}^{m}$ that are generated using this class, $I_1 = \{i|x_i(1) \in C_1(1)\}$, that is $\{\eta_i^*(1)\}_{i \in I_1}$. Generate a random sample of size $m \times f_1(0)$ with replacement from $\{\eta_i^*(1)\}_{i \in I_1}$.

3. Repeat step 2 for $j = 2, \cdots, J$.

I generate random samples for various counterfactual scenarios using these two procedures. These approaches enable me to decompose changes in the density of $h(Y)$ based on these generated samples.

Let $\hat{f}\big(\eta(\Psi)\big)$ be an estimator of the marginal density of an observed sample of $h(Y)$ in year Ψ, and $\hat{f}\big(\eta^*(\Psi)\big)$ an estimator of the density of $h(Y)$ based on the generated sample $\{\eta_i^*(\Psi)\}_{i=1}^{m}$. I denote $\hat{f}(\eta^*(1); X(0))$ as an estimate of the counterfactual density in $\Psi = 1$ if the covariates had been distributed as in $\Psi = 0$. Similarly, $\hat{f}(\eta^*(1); x_i(0))$ is an estimate of the density in $\Psi = 1$ if only the i^{th} covariate is distributed as in $\Psi = 0$. For a summary statistic $\alpha(\cdot)$ (e.g., a quantile or scale measure), the decomposition of changes in α can be written as:

$$
\begin{aligned}
\alpha(\hat{f}(\eta(1))) - \alpha(\hat{f}(\eta(0))) \quad = \quad & \underbrace{(\hat{f}(\eta^*(1))) - \alpha\left(\hat{f}\left(\eta^*(1); X(0)\right)\right)}_{\text{change in covariates}} && (16) \\
& + \underbrace{\alpha\left(\hat{f}(\eta^*(1); X(0))\right) - \alpha\left(\hat{f}\left(\eta^*(0)\right)\right)}_{\text{change in returns}} && (17) \\
& + \text{residual} && (18)
\end{aligned}
$$

Likewise, the individual contribution of a covariate is:

$$\alpha(\hat{f}(\eta^*(1))) - \alpha(\hat{f}(\eta^*(1)); x_i(0)). \tag{19}$$

This alternative approach estimates the entire distribution to identify the factors affecting shifts in a particular quantile. In contrast, the method using the conditional Gini index directly evaluates inequality across the whole distribution.

2.3 Inequality in the Distribution of Y

I compute the linear decomposition of the Gini index in Equation (6) for the transformed variable, $h(Y)$. However, I also find it compelling to gauge the influence of individual characteristics on the distribution of the positive random variable, Y. Studying the transformed variable instead of the variable in its original scale might pose a conflict between the statistical and economic objectives of this study. But, considering the assumed properties of the transformation $h(\cdot)$ and using the properties of the quantile function, I derive:

$$
\begin{aligned}
Q_Y(t|x) &= h^{-1}\left(Q_{h(Y)}(t|x)\right) \\
&= h^{-1}(x^T\beta(t)),
\end{aligned} \tag{20}
$$

This implies:

$$L(\tau|x) = \frac{1}{\mu}\int_{0}^{\tau} h^{-1}(x^T\beta(t))dt \tag{21}$$

and

$$G(x) = 1 - \frac{2}{\mu}\int_{0}^{1}\int_{0}^{\tau} h^{-1}\left(x^T\beta(t)\right)dt\,d\tau. \tag{22}$$

While the previous relationship is not inherently linear because $h^{-1}(\cdot)$ isn't linear, Equation (22) establishes a connection between the Gini coefficient of Y and a transformation of a linear combination of the quantile regression coefficients. I acknowledge that this link requires further exploration in future research.

3 Estimation Procedure

In the previous section, I measure the impacts and decompose the temporal changes in the distribution of $h(Y)$ by estimating $\Pi_j = \int_0^1 \int_0^\tau 2\beta_j(t)dt\,d\tau$ as detailed in Equation (6). A straightforward estimate would be $\widehat{\Pi}_j = \int_0^1 \int_0^\tau 2\hat{\beta}_j(t)dt\,d\tau$, where $\hat{\beta}_j$ represents the estimated quantile regression coefficient.

To clarify the estimation approach, I consider $Q_{h(Y)}(t|x)$ for $t \in (0,1)$ to be the t-th conditional quantile function of $h(Y)$, given a vector of covariates $x \in R^P$. I posit that the conditional quantile function can be represented as:

$$Q_{h(Y)}(t|x) = x^T\beta(t), \tag{23}$$

Here, $\beta(t)$ is a vector in R^P, with its entries being the quantile regression coefficients. Drawing from Koenker and Bassett (1978), for a specific $t \in (0,1)$, I can estimate $\beta(t)$ by:

$$\min_{b \in R^P} \sum_{i=1}^{N} \rho_t(h(y_i) - x^T b) \,, \tag{24}$$

where

$$\rho_t(u) = u\big(t - I(u < 0)\big) \,, \tag{25}$$

In the above, N signifies the number of observations and $I(\cdot)$ represents the indicator function.

Let $\hat{\beta}(t)$ be the solution to the optimization problem expressed in Equation (24). Further, let $\hat{\beta}_j(t)$ indicate the j-th component of this estimated vector, which functions based on the quantile t. My aim is to compute:

$$\widehat{\Pi}_j = \int_0^1 \int_0^\tau 2\hat{\beta}_j(t) dt \, d\tau = \sum_{i=0}^{n-1} \int_{\tau_i}^{\tau_{i+1}} \int_0^\tau 2\hat{\beta}_j(t) dt \, d\tau. \tag{26}$$

One feasible method to determine the double integrals in Equation (26) is to perform a numerical computation, using a set grid for the point evaluations of $\hat{\beta}_j(t)$. However, a challenge emerges when extending one-dimensional integration methods to multiple dimensions: the required function evaluations increase exponentially, which is often termed the "curse of dimensionality". For instance, if we opt for m evaluation points, the numerical approximation for the double integrals would demand an estimation proportional to m^2.

I propose an alternative. By deriving a smooth approximation for $\hat{\beta}_j(t)$ through a known functional form—one with a discernible antiderivative—I can then analytically calculate $\widehat{\Pi}_j$ using the established functional form. This method not only streamlines computations but also maintains the number of function evaluations at m, aligning with our original set of evaluation points.

There are various approximation procedures designed to smooth a continuous function. Splines, a commonly known method, approximate a function through a piecewise continuous polynomial. However, employing splines to estimate Π_j presents challenges. Firstly, it necessitates the designation of knots. [12] Secondly, it mandates numerous piecewise integrations contingent on the number of these knots.

Alternatively, I can utilize orthogonal polynomials to approximate any continuous function.[13] This technique produces a singular polynomial of order K that minimizes the error squared between the smoothing polynomial and the observed function values. One limitation when using orthogonal polynomials to compute Π_j is the need to define the polynomial's order, K. However, a notable advantage is the ease of computation for Π_j's estimate. I can determine the double integrals of Equation (26) in a single step, contrasting the piecewise integration needed with splines.

I assume that $\hat{\beta}_j(t)$ is continuous over the interval [0,1]. As outlined in Judd (1998), I can approximate this function using any set of orthogonal polynomials defined on this interval. The Weierstrass Approximation Theorem ensures that, given the aforementioned assumptions, $\hat{\beta}_j(t)$ can be uniformly approximated over [0,1] by polynomials, achieving any desired precision. Several orthogonal polynomial families exist, including Legendre, Chebyshev, Laguerre, and Hermite. The distinction among these families mainly lies in their weighting functions and polynomial domains. For functions with a bounded domain, the most straightforward weighting function is $w(x) = 1$, which aligns with the Legendre polynomials. For the sake of simplicity, I choose to use the Legendre polynomials to approximate $\hat{\beta}_j(t)$ over the interval [0,1].

---

[12] The knots are the places where the polynomial pieces connect in the jargon used for splines.

[13] A weighting function, $w(x)$, on $[a, b]$ is any function that is positive almost everywhere and has a finite integral on $[a, b]$. Given a weighting function, the inner product between the polynomials f and g is defined as $\langle f, g\rangle = \int_a^b f(x)g(x)w(x)dx$. A family of polynomials $\{p_n(x)\}$ is orthogonal with respect to the weighting function $w(x)$ if and only if $\langle p_m, p_n\rangle = 0$ for all $m \neq n$.

The domain of the Legendre polynomials is $[-1,1]$, yet our goal is to approximate $\hat{\beta}_j(t)$ on $[0,1]$. This requires reshaping the orthogonal Legendre polynomials to fit the $[0,1]$ interval, a process detailed in Appendix B.[14] With the reshaped polynomials, I can derive the least-square approximation of $\hat{\beta}_j(t)$ using a polynomial of order K. Specifically, let $\tilde{\beta}_{j,K}(t)$ represent a polynomial described as:

$$\tilde{\beta}_{j,K}(t) = \alpha_{0,j}p_0(t) + \alpha_{1,j}p_1(t) + \cdots + \alpha_{K,j}p_K(t).$$

Here, $\{p_k\}_{k=0}^{K}$ denotes the first $K+1$ Legendre polynomials defined over $[0,1]$. The objective is to minimize the sum of the squared errors between $\hat{\beta}_j(t)$ and $\tilde{\beta}_{j,K}(t)$, which is defined by

$$E(\alpha_{0,j},\cdots,\alpha_{K,j}) = \int_0^1 \left[\hat{\beta}_j(t) - \tilde{\beta}_{j,K}(t)\right]^2 dt.$$

For a given K, I define

$$\tilde{\tilde{\beta}}_{j,K}(t) = \underset{\alpha_{0,j},\cdots,\alpha_{K,j}}{argmin}\, E(\alpha_{0,j},\cdots,\alpha_{K,j});$$

the polynomial $\tilde{\tilde{\beta}}_{j,K}(t)$ is a smoothed approximation of $\hat{\beta}_j(t)$ based on $K+1$ known polynomials on $[0,1]$. One of the major benefits of this polynomial approximation is its closed-form antiderivatives, which streamline computation. By leveraging this smoothed approximation, I can efficiently determine the influence of the j-th covariate on the inequality of the distribution of $h(Y)$, as described in Equation (26).

To better grasp the implications of the polynomial approximation's order, K, consider a hypothetical scenario. Let's say a researcher postulates a model for the conditional quantile function of wage logarithms, using the transformation $h(\cdot) = \ln(\cdot)$. Further, let’s assume the researcher gauges a quantile regression coefficient, $\hat{\beta}_j(t)$, across a grid of $m = 69$ quantiles.

[14] Reshaping the Legendre orthogonal polynomials into the interval [0,1] requires a simple linear substitution that affects the limits of the integral.

Figure 1 illustrates a representative smooth least-square approximation of this assumed quantile regression coefficient, $\hat{\beta}_j(t)$. In each of the figure's panels, the dashed line represents $\tilde{\tilde{\beta}}_{j,K}(t)$. Panel (a) showcases the outcomes of employing a smoothed polynomial approximation of second degree. If our focus remains solely on the quantile regression's point estimate, a second-degree polynomial appears to inadequately represent the estimate. Yet, when accounting for the 95% confidence interval, it's evident that a second-degree polynomial might offer a reasonable approximation. On the other hand, Panel (b) depicts the outcome using a sixth-degree polynomial. Increasing the polynomial's degree enhances the approximation fidelity to the quantile regression coefficient's point estimate. Such an enhancement might be a preferable approach, depending on the research objectives.

The panels in Figure 1 elucidate that utilizing a higher polynomial degree improves the approximation's adherence to the quantile regression's point estimate. However, even with a higher degree, the polynomial might not completely smooth out abrupt variations in the estimates. Such variations could likely arise from data scarcity for specific quantiles—typically at the extreme top or bottom 1% of wage distributions. Selecting the polynomial's degree presents a balancing act. While a greater degree enhances the approximation's precision—something that may be sought after—it doesn't always align with the primary objective of the approximation. Additionally, as demonstrated in Figure 1, if the smoothing polynomial resides within the confidence bands, then the approximation might be deemed '*satisfactory*', despite any minor discrepancies.

I use bootstrapping to assess if the estimate for $\widehat{\Pi}_j$ significantly deviates from zero and to establish its confidence intervals. To detail the process, let N represent the sample size and $\mathbb{R}$ the number of bootstrap repetitions. During each repetition, one should resample N observations with replacement, estimate the quantile regression coefficients, $\hat{\beta}(t)$, using this resampled data, then determine the smooth polynomial approximations, $\tilde{\tilde{\beta}}_{j,K}(t)$. Using this smooth approximation, compute an estimate of $\beta_{j,K}$ based on Equation (26). The point estimate for the impacts, Π_j, is derived by averaging the $\mathbb{R}$ estimates from all repetitions, and the 95% bootstrap confidence intervals are constructed using the 2.5-th and 97.5-th quantiles from the $\mathbb{R}$ estimates.

The precision of my estimation procedure strongly depends on how accurately I model the conditional quantile function $Q_{h(Y)}(t|x)$, notably the linearity aspect of the quantile regression model. Relying on the consistency of the quantile regression estimate and considering regularity conditions, as outlined by Bantli and Hallin (1999), I am confident about the precision of my Π_j estimate. Assuming we know the quantile process $\beta(t)$ and therefore the true impact estimate Π_j, for a sizable sample and as detailed by Koenker (2005), we know that $\hat{\beta}_n(t) \rightarrow \beta(t)$. This leads me to the inference that $\widehat{\Pi}_j \approx \Pi_j$. I delve deeper into this rationale in Appendix C, where I assert that my estimation procedure offers an accurate assessment of the impact of the j-th covariate.

4 Hourly Wage Series From the CPS ORG

In this study, I scrutinize data from the Current Population Survey (CPS) to analyze the shifts in wage distribution in the United States from 1980 to 2015. Since 1979, the CPS's Outgoing Rotation Group (ORG) has actively engaged workers in a comprehensive survey, collecting detailed earnings data. This data facilitates the accurate estimation of hourly wages, which can be derived directly from reported hourly earnings or calculated by dividing weekly earnings by the corresponding hours worked per week. Employing this methodology provides a robust measure of labor costs and resonates well with established economic theories concerning wage determination, which are founded on the dynamics of market supply and demand.

Nevertheless, the task of utilizing the ORG data comes with its set of challenges, as highlighted by Acemoglu and Autor (2011). Over the years, the CPS has altered its methods of classifying and processing the earnings data for hourly and non-hourly paid workers, introducing variability in the dataset. To assemble a cohesive series of hourly wages, I make careful adjustments to account for the modifications in the top coding of weekly earnings and the changes in categorizing overtime payments, tips, and commissions for those earning hourly wages. Furthermore, I adapt to the transformations in responses concerning 'usual weekly hours' over various years. These shifts, which have occurred over a span greater than three decades, compel an attentive analysis of survey alterations to maintain data coherence and accuracy.

To ensure I overcome all these challenges, I utilize programs available under the GNU General Public License created by the Center for Economic and Policy Research (CEPR). This resource integrates data from the National Bureau of Economic Research (NBER) Annual Earnings Files with the CPS basic monthly files, establishing a consistent wage series utilizing the data from the CPS ORG. [15]

For the final sample, I first adjust every wage to mirror 2020's monetary value using the Consumer Price Index issued by the Bureau of Labor Statistics. [16] Following this, I focus on workers who report hourly wages between $1 and $100 (in 1979 dollars) and fall within the age bracket of 16 to 65 years. Subsequently, following Acemoglu and Autor (2011), I devise a potential experience variable, which is the result of subtracting the number of years of education from individuals' ages and deducting an additional five years to account for elementary schooling. Moreover, I incorporate an indicator variable to the series to signify female, nonwhite and unionized workers. I also adhere to a uniform classification concerning twenty industries and a consistent classification of manufacturing employees.

In this study, I meticulously categorize educational attainment, recognizing its crucial impact on earnings as established by seminal works in economics. I draw upon the insights of Card (1999), Autor, et al. (2003), Goldin and Katz (2007), and Acemoglu and Autor (2011),

---

[15] These programs have significantly contributed to my success in developing a consistent hourly wage series, extending from 1980 to 2015. I direct readers to Schmitt (2003) for a comprehensive description of the harmonized hourly wage series. Additionally, the methodology employed to derive this series aligns with that detailed in the Data Appendix of Acemoglu and Autor (2011). For those looking to conduct further analysis, I have made the relevant codes available on my website.

[16] Specifically, using the seasonally adjusted index for all items based on US city average (Series Id: CUSR0000SA0).

all of whom underscore education's pivotal role in shaping labor market outcomes and highlight its position as a key determinant of earnings and employment opportunities.

Building upon this foundational understanding, I incorporate a variety of educational categories in my analysis: non-school attendees or dropouts, high school graduates, individuals with some college education but no degree, those holding an associate degree, and individuals with a bachelor's degree or higher. Specifically, I include the associate degree category to acknowledge its proven positive impact on earnings. Recent research from Jepsen et al. (2014), Bahr et al. (2015), Stevens et al. (2019), and Grosz (2020), collectively affirm the significant earnings returns associated with community college programs, highlighting their vital role in the labor market and underscoring the importance of considering associate degrees in educational classifications for a thorough evaluation of their impact on wages.

I utilize large sample sizes for the analysis, averaging around 165,000 workers annually from 1980 to 2015. To highlight the gender gap, I delineate the CPS samples' summary statistics for men and women separately in Table 1. While the average real wages for men remain relatively constant at 3.2 ($24.4), I observe a systematic increase for women. A noticeable reduction in the gender wage gap occurred between 1980 and 2015, yet a discernible difference persists. Furthermore, Table 1 highlights an escalating trend in educational attainment for both genders, with women exhibiting a higher average number of years of education compared to men. During this period, I also note a decline in the unionization rate and the proportion of manufacturing employees, juxtaposed with a consistent increase in nonwhite workers' participation rate.

To elucidate the wage distribution disparity both intra- and inter-gender, I present weighted kernel density estimates of the hourly wages for men and women spanning from 1980 to 2015 in figures 2a and 2b.[17] These graphs feature a vertical line indicating the

[17] These figures bear resemblance to those found in DiNardo et al. (1996), though a key distinction lies in my application of the CPS sample weights, in contrast to their implementation of hours-weighted kernel estimates. As in DiNardo et al. (1996), I determine the bandwidth for this estimation employing the Sheather and Jones (1991) method.

respective (log) real minimum wage as referenced in column two of Table 1, which shows the concentration of wage distributions at the lower range. These representations clearly illustrate a significant expansion in the upper tail of the distribution in recent years compared to preceding periods. Moreover, I point out an evident broadening in the spread of hourly wages relative to the mean over time for both sexes, more pronounced in women's data. This visual representation aligns with previous findings by Levy and Murnane (1992), DiNardo et al. (1996), Katz (1999), and Autor et al. (2008).

5 Sources of Wage Inequality in the US

Building on the relationship between the conditional Gini coefficient and conditional quantile regression discussed in section 2, I employ the comprehensive hourly wage data from the CPS ORG for my empirical application. I start by setting $h(\cdot)$ as the natural logarithm. Then, I select 1986 ($\Psi = 0$) as my starting point to account for the onset of increasing wage inequality in the US during the early 1980s. The analysis extends to 2015 ($\Psi = 1$), encompassing three decades of rising wage disparity.

I model the conditional quantile function of the logarithm of wages as:

$$Q_{\ln w_{isj}}\left(t|x_{isj};\Psi\right) = x_{isj}^{T}\beta_{\Psi}(t) + \eta_s(t) + \gamma_j(t) + \varepsilon_{isj}(t) \tag{27}$$

Here, x_{isj} represents the characteristic for individual i, in state s, working in industry j. This vector encompasses job-related attributes such as unionization status, public sector employment, manufacturing job, and part-time work status. Demographic attributes include indicators for nonwhite, female, and marital status. Additional controls cover a quadratic in potential experience, urban living indicators, education categories, decade-based experience indicators, and their interactions with education classes. Finally, $\eta_s(t)$ and $\gamma_j(t)$ are state and industry fixed effects, respectively.

Importantly, the selected independent variables in my quantile regression model serve as robust controls for broader macroeconomic and structural shifts that have occurred over the study's timeframe. By incorporating industry fixed effects, inspired by the findings of Blum (2008), I effectively capture the sectoral composition of the economy. This inclusion is pivotal to elucidate transitions, such as those from manufacturing-heavy sectors to service-centric ones or the notable rise of the tech sector in more recent years.

Incorporating job-related attributes such as manufacturing job and state fixed effects, I indirectly tap into the effects of international trade. Chongvilaivan and Hur (2011) emphasize the nuanced relationship between sectors like manufacturing and the broader implications of trade liberalization, outsourcing, and global competition. Additionally, the state fixed effects account for regional disturbances, especially the 'China shock' highlighted by Autor et al. (2013).

Technological changes are significant drivers of wage inequality. In my model, I've incorporated the urban living indicators to identify regions more inclined towards technology adoption and innovation. Drawing from Hühne and Herzer (2017), I've combined education categories with decade-based experience indicators. This blend illuminates the changing job requirements amidst swift technological progress. Given that technology tends to benefit those with specialized education or skills, this integration provides insight into technology's impact on wages.

Figure 3 displays the quantile regression coefficients estimated of Equation (27) for a grid of 69 equally-spaced points over the (0,1) interval [18]. In each panel, the solid line corresponds to the estimates in 1986, while the dashed line depicts those from 2015. In each case, the shaded regions around the lines correspond to the 95% confidence interval obtained by computing a Huber sandwich estimate using a local estimate of the sparsity.

[18] The 69 equally spaced points create a grid that has constant step of around 1.4%.

Figure 3 offers a comprehensive view of the factors that affect wages from 1986 to 2015 across the entire distribution. Starting with union status, the positive wage premium associated with union membership was more pronounced in 1986 across all quantiles but has reduced by 2015. For public sector employees, there is a premium for lower wage workers, and a penalty for upper wage employees, and these premium and penalty remains similar in magnitude for both 1986 and 2015.

Figure 3 depicts the evolving impact of manufacturing job status on wages across three decades. In 1986 manufacturing job status had a notably negative influence on wages. Contrastingly, by 2015, manufacturing job status became a positive determinant for wages across almost all quantiles, although the strength of this influence declined among the highest earners. This shift from a predominantly negative to positive impact over the years highlights the dynamic nature of the manufacturing sector's role in wage determination, possibly influenced by structural changes, economic shifts, and external such as the China shock and trade agreements that over the studied period that led to keep only a more productive manufacturing sector withing the US.

Looking at demographic factors, nonwhite and women workers both show wage penalties, with the latter having a slightly more pronounced difference, especially in higher quantiles. Notably, this disparity has narrowed for both groups over the 30-year span, indicating progress, albeit limited, in wage equality. Moreover, married workers enjoy a wage premium, particularly in 1986 in the lower end of the distribution. However, the advantage at the lower wages diminishes by 2015.

Figure 3 underscores a shift in the urban wage premium between 1986 and 2015. In 1986, the figure shows an evident urban wage premium, especially prominent in the middle to higher quantiles. However, by 2015 this premium appears to diminish, suggesting a decrease in the urban wage advantage. This trend complements the findings of De la Roca and Puga (2017), emphasizing the dynamic nature of knowledge spillover in urban environments and its potential impact on wages over time.

Finally, Figure 3 reveals distinct patterns when evaluating the wage dynamics between 1986 and 2015 across various educational attainment levels. High school and some college-educated individuals from 1986 displayed a pronounced wage advantage over their 2015 counterparts, particularly in the lower quantiles. This advantage signified that those with less than a technical degree in 1986 had better wage outcomes, especially in the lower-paying roles. In contrast, those with an associate degree in both years showed nearly indistinguishable wage impacts across most quantiles, with only a slight edge for the 2015 cohort in the top earnings bracket. Remarkably, the trend inverts for individuals with a college degree or more. The 2015 cohort consistently outperformed the 1986 group, reflecting the escalating value of a college degree over these three decades, especially among the highest earners. This divergence underscores the shifting dynamics in the premium placed on education within the labor market over this period.

5.1 Impact of Individual Characteristics

I implement the procedure described in section 3 to estimate the sources of wage inequality in the US. I choose the order of the polynomial approximation to be 7 for year $\Psi = 0$ (1985) and $\Psi = 1$ (2015). Appendix C shows that the results are not sensitive to the choice of the order of the polynomial approximation. Additionally, I set the number of repetitions for the bootstrap estimation as $R = 1{,}000$. For each year, I compute $\hat{\mu}_{ln}$ as the weighted average of the logarithm of real wages in 2020 dollars. Then, I estimate $\widehat{\Pi}_j$ in Equation (26) for the years of analysis and compute the impact estimate $\frac{\widehat{\Pi}_j}{\hat{\mu}_{ln}}$.

The results of the estimation for selected covariates are presented in Table 2. As discussed in section 2.1, given a *small* positive change in covariate j, a positive sign of $\widehat{\Pi}_j$ is associated with a reduction in the inequality of the distribution of (log) wages. The first entry of each cell in Table 2 presents the impact estimation, whereas the second reports the 95% bootstrap confidence interval; each column exhibits the results for the corresponding year. All covariates shown in Table 2, except potential experience, are binary variables, which is relevant when thinking about small positive changes.

Evaluating the results in Table 2, one can discern the relative impacts of various factors on wage inequality between 1986 and 2015. Unionization, for instance, consistently reduced wage inequality. Holding other factors constant, an increase in the proportion of unionized workers can lead to a noticeable decline in wage disparities. However, its impact has lessened over the years, pointing to evolving workplace dynamics and the role of unions. Public sector jobs had a marginal positive effect in 1986, but by 2015, this influence disappeared. Conversely, the manufacturing sector, which initially exacerbated wage disparities in 1986, started to bridge this gap by 2015.

Table 2 shows that potential experience has a modest but steady role in reducing wage disparities. The implications tied to marital status and residence in urban areas also offer insights. The former consistently aids in narrowing the wage gap, while the influence of the latter waned slightly across the years. Factors such as race and gender, despite their negative implications, have witnessed a decline in their adverse impacts over the years. Nonetheless, the persistent effects tied to these demographics underline the need for continued interventions.

When examining education, higher levels have a resounding impact. Keeping everything else constant, an uptick in the proportion of college-educated workers significantly drops the conditional Gini index. This effect is more pronounced than that of high school-educated workers, indicating the premium attached to college education in wage determinants. The results in Table 2 not only highlight the variables impacting wage inequality in the US during the study period but also shed light on the evolving magnitude and direction of these effects. While some factors actively narrow the wage gap, others underscore the necessity for targeted interventions. Delving deeper, the following subsection offers insights into the temporal shifts in wage distribution.

5.2 Temporal Changes in the Distribution of (Log) Wages

In this subsection, I delve deeper into the temporal changes in the distribution of wages. I aim to compare the approach introduced in subsection 2.2 with the well-established algorithm by Machado and Mata (2005), hereafter referred to as MM. This comparison aims to underline the merits of the proposed method. For the execution of the proposed method from section 2.2, I partition the interval (0,1) into four evenly distributed sub-intervals, namely Q1, Q2, Q3, and Q4. I maintain the polynomial approximation order at 7 for the years 1986 and 2015 and set the repetition count at 1,000, consistent with my previous analyses. For the MM methodology, as detailed in subsection 2.2.1, I designate m = 4,500 and adopt $\alpha(\cdot)$ as the quantile statistic.

Table 3a presents the results of applying the MM decomposition to wage data. The first two columns feature the calculated quantiles for the estimated log-transformed wage distributions for the selected study years. These quantiles range from the 1st to the 99th, capturing the lower, median, and upper segments of the wage scale. The third column reports the changes in these quantiles between 1986 and 2015, including both the point estimates and the corresponding 95% bootstrap confidence intervals. I obtain these intervals from 1,000 bootstrap samples, using the 2.5th and 97.5th percentile markers of the bootstrap distribution, which provides solid statistical conclusions about the shifts in wage distribution across thirty years.

Column 3 of Table 3a indicates the overall conditional wage changes and reveals a critical limitation of the MM method in assessing inequality. From 1986 to 2015, while the data show substantial wage growth at the 1st quantile by 30.5% and at the 99th percentile by 29%, this approach fails to capture the nuance of inequality changes. The pronounced wage increases at the distribution's bottom may seem to be offset by similarly significant gains at the top; drawing conclusions from this is problematic since the analysis merely contrasts two discrete quantiles rather than a continuous evaluation of the entire wage distribution. Hence, it's not surprising that the Gini coefficient only shows a marginal increase from 11.04 to

11.09. This inconsistency points to a potential blind spot in the MM method for understanding wage inequality; it indicates sizable shifts in wages at the extremes but does not effectively translate these changes into a comprehensive understanding of inequality within the wage distribution.[19] Therefore, while the MM method can identify wage-level shifts, its ability to elucidate the broader picture of wage inequality is constrained.

Columns 4 to 6 in Table 3a decompose the total changes in the wage distribution into three distinct parts: changes due to covariates as described by equation (16), changes attributable to variations in returns as specified by equation (17), and the residual component as defined by equation (18). This decomposition dissects the factors influencing wage changes across different income levels, yet it yields results that are challenging to interpret concerning inequality. With some quantiles receiving positive impacts and others negative, the overarching effect on inequality remains unclear, underscoring the complexity of discerning clear trends from such diverse outcomes.

For the lower percentiles, specifically the 1st and 10th, the analysis shows that changes in covariates do not significantly affect wage changes. In stark contrast, a significant positive impact from covariates emerges at the 25th and upper levels of the wage distribution, suggesting that the characteristics measured by these covariates are particularly beneficial for individuals in higher income brackets.

The influence of returns on characteristics also shows variability across the distribution. The lower end and the top benefit from positive impacts, signifying that specific attributes or qualifications are highly valued for earners in these segments, perhaps due to unique market conditions or the premium placed on certain skills. At the median, however, the effect is negative, which could indicate that median earners benefit less from the measured characteristics, or that different factors are at play that are not captured by the covariates.

Finally, Columns 7 to 15 in Table 3a analyze the effects of individual covariates on wage changes across different income levels. Unionization appears to be a double-edged sword,

[19] In fact, one can realistically only report a few editorializing statistics in a table before a reader gets completely lost in numbers.

having a significantly negative impact at the 75th and 90th percentiles, which suggests that the benefits it provides may not be felt equally across the wage spectrum. Meanwhile, manufacturing does not significantly influence wage changes at any percentile, hinting at its diminished role or varied effects within the contemporary economic landscape. When it comes to demographics, nonwhite workers face a disadvantage with negative wage changes from the 10th to the 90th percentile. For women, the data presents a complex picture; there's a notable disadvantage at the higher end of the wage spectrum with a significant negative impact at the 90th percentile, whereas at the entry point of the 1st quantile, women tend to have a positive wage impact. Living in an urban area does not generally affect wage changes, except at the 90th percentile where its impact becomes significant. Finally, education stands out as a clear advantage; possessing a college degree is positively and significantly associated with wage increases across most of the percentiles, underscoring the critical value of higher education in promoting wage growth.

Table 3b provides a detailed breakdown of the shifts in the wage distribution, employing the additive decomposition approach to the Gini index as outlined in Equation (14). The format of this table mirrors that of Table 3a, where each cell offers two pieces of information: the initial entry denotes the point estimate, while the subsequent entry indicates the 95% bootstrap confidence intervals derived from 1,000 bootstrap samples. Within the context of this table, a *negative* point estimate suggests a *decrease* in the Gini coefficient, signaling a *decline* in wage inequality. Hence, any negative figures within the table can be interpreted as factors contributing to a more equitable wage distribution.

Table 3b breaks down the changes in wage inequality from 1986 to 2015 into four different quartiles and the entire distribution. The data shows that for the lower half of wage earners (Q1 and Q2), wage inequality decreased, with Q1 showing a statistically significant reduction. On the other hand, the third level (Q3) shows a slight, but not statistically significant, increase in inequality. The most significant change is seen in the top earners (Q4), where there is a definite increase in inequality, pointing towards heightened inequality at the higher end of wage earners. The first merit of the proposed method becomes apparent here. While overall inequality -when combining all quartiles- has gone up just a little, it highlights a significant decrease at the lower end that almost exactly offsets the increase at the higher end. This gives us a much clearer picture of how wage inequality has shifted over time.

Table 3b examines the effects of covariates and the associated returns on wage inequality. According to the figures in Column 4, the influence of covariates on wage distribution within the first quartile (Q1) is not statistically significant. In contrast, the impact is significant in the second, third, and fourth quartiles (Q2, Q3, and Q4), suggesting that these factors substantially influence the dynamics of wage distribution among middle- to upper-income brackets. Column 5 reveals a uniform and statistically significant increase in wage inequality due to shifts in the returns to specific characteristics across all quartiles. This pattern is particularly pronounced in the higher wage quartiles (Q3 and Q4), reflecting a labor market in which the prices of particular characteristics are commanding increasingly higher wages. This discrepancy significantly affects the upper tier of the wage distribution, contributing to a broadening wage gap.

Table 3b reveals that unionization and manufacturing have a uniform and significant impact on wage inequality across all quartiles. Specifically, Columns 7 and 8 show that changes in the returns to these sectors contribute to increased wage disparities. This phenomenon likely mirrors the broader economic shifts affecting these sectors, such as the decline in unionized jobs and the transformation of manufacturing due to factors like globalization and automation. Such economic transformations are creating a divergence in wage outcomes, where higher-skilled workers in these sectors may benefit from wage increases, while those with fewer skills could experience wage stagnation, thus intensifying wage inequality. The methodology I propose, in contrast to the MM method, is sensitive enough to isolate and identify the changes in returns from unionization and manufacturing jobs, offering a more precise understanding of the factors escalating wage disparities.

Regarding demographics, the analysis separates the impact of race and gender. Table 3b shows that racial disparities, represented by the Nonwhite variable in Column 9, contribute to increasing wage inequality consistently across all quartiles, suggesting that wages are diverging along racial lines. However, the gender factor, analyzed through the Women variable in Column 10, does not manifest a significant effect on wage inequality, indicating that gender by itself may not be a dominant factor in wage disparity within the scope of this analysis.

Urbanization, as covered in Column 11 of Table 3b, consistently presents a statistically significant reduction in wage inequality. The urban coefficient's negativity across all quartiles suggests that urbanization might serve as a leveler in wage distribution. Urban areas, with their broader spectrum of job opportunities and the potential for higher earnings, are inferred to facilitate a more equitable wage dispersion, which could mitigate the effects of wage inequality within cities. This is another characteristic that the proposed method isolates as a contributing factor to wage equality, in contrast to the MM method.

Lastly, the influence of educational attainment, particularly the possession of a College Degree shown in Column 15, displays a robust negative and significant impact on inequality, with the most substantial effects observable in the uppermost quartile and the aggregate. The importance of a college education in the contemporary job market is underscored by this trend; as college-educated workers are increasingly favored, the resulting wage premium compresses the upper tail of the wage distribution, culminating in a reduction of inequality at the higher wage levels. This underscores the critical role of higher education in the quest to diminish wage inequality.

In summary, my proposed method offers a more incisive analysis of the temporal changes in wage distribution compared to the MM method. The approach of the proposed method successfully captures the nuances of wage dynamics across different income levels and quantifies the influences of various socioeconomic factors. In particular, it highlights how unionization and shifts in manufacturing have differentially affected wage disparities, reflecting the complex interplay of globalization, technological advancement, and skill levels within the workforce. This method also reveals that urbanization tends to reduce wage inequality, while higher education emerges as a powerful equalizer, particularly for upper-tier earners. The detailed decomposition of wage changes afforded by the proposed method underscores its ability to discern the multifaceted drivers of wage inequality, which the MM method overlooks. Therefore, the proposed method stands out for its comprehensive analysis, providing clear insights into the factors that drive wage inequality and offering robust evidence that is crucial for informed policymaking aimed at reducing wage disparities.

5.2.1 Discussion

As I conclude my analysis, I find it crucial to integrate my research outcomes with the widely held views on how education affects wage inequality. Let me break it down simply: imagine an economy split into two types of workers, the low-skilled and the high-skilled, with education measuring their skill level. Even though there is a range of differences within each group, the average wages for low-skilled workers, which I will call w_L, contrast with the w_H earned by their high-skilled counterparts. The educational premium, reflected by the w_H/w_L ratio, is something one can roughly estimate by regressing log wages on years of education.

Now, as the proportion of high-skilled workers rises, it triggers a significant drop in wage inequality, and this happens through two main forces. The price effect kicks in first: the more these high-skilled workers flood the market, the more their relative wages start to fall. Then there is the composition effect: a larger slice of our workforce is climbing up to the high-skill, better-paid ranks, which naturally compresses the wage gap. These two effects combined lead to the prediction that shifts to the right in both demand and supply of skilled workers that would leave price unchanged would result in less wage inequality. The results show that educational progress is not just theoretical but a tangible lever for lessening wage disparities.

6 Conclusion

This paper introduces a novel econometric methodology to analyze factors influencing wage disparities in the US from 1986 to 2015. The study additively decomposes the conditional Gini index by linking the conditional Lorenz curve to the conditional quantile function, enabling precise quantification of factors affecting wage inequality. Utilizing data from the CPS ORG, the paper reveals key findings: the manufacturing sector's role in wage dynamics has evolved; race and gender impacts on wages have lessened but still necessitate intervention; and notably, higher education, particularly college degrees, significantly reduces wage inequality in the US during the analysis period. This methodological advancement offers new insights into wage distribution and inequality.

This study enhances the literature on decomposition methods by using the conditional Gini coefficient to measure inequality in log wage distribution directly, bypassing density modeling. Traditional methods relied on kernel estimates or selected quantile analyses, often overlooking comprehensive inequality measures. The proposed approach, assuming a linear conditional quantile function for log wages, reveals how the impacts of different factors on wages have evolved. Notably, when compared with the MM algorithm, this method not only aligns with MM's conclusions but also uncovers aspects of wage inequality that the MM method did not distinctly identify.

A limitation of this study, as an analog to Oaxaca (1973) decomposition, is the presumption that changes in the characteristics do not modify the returns of those characteristics. Moreover, the analysis only accounts for changes in the covariates from 1986 to 2015, but the proposed decomposition technique could have considered counterfactual scenarios in reverse order. More importantly, the linear decomposition works for a particular transformation of wages for which the conditional quantile functions are assumed to be linear in parameters (i.e., log wages), but this may not be a natural scale to analyze the distribution disparity.

Further research could usefully explore how to account for the general equilibrium effects given changes in the distribution of the covariates, because those changes will also affect the returns to the characteristics. Moreover, a future study investigating different counterfactual scenarios and more recent years of analysis would be very interesting. A natural progression of this work is to extend the proposed method to the untransformed variable (i.e., wages) to address questions related to the inequality of the distribution of the variable in levels.

Declarations

Ethical Approval: not applicable

Competing interests: not applicable

Authors' contributions: not applicable

Funding: Not applicable

Availability of data and materials: The Annual Earnings Files are also known as Merged Outgoing Rotation Groups (MORG) and can be downloaded from: http://www.nber.org/morg/annual/. The CPS basic monthly files are known as the Basic Monthly CPS, and the Bureau of Labor Statistics maintains those files available at: https://www.census.gov/data/datasets/time-series/demo/cps/cps-basic.html. The CEPR code can be downloaded from https://ceprdata.org/cps-uniform-data-extracts/cps-outgoing-rotation-group/cps-org-data/. The R code used to implement the methods is available upon request.

References

Abel, J. R., & Deitz, R. (2019). Why are some places so much more unequal than others? Federal Reserve Bank of New York Economic Policy Review, 25(1), 58-75.

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In Handbook of labor economics (Vol. 4, pp. 1043-1171). Elsevier.

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the US wage structure. Econometrica, 74(2):539–563.

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China syndrome: Local labor market effects of import competition in the United States. American economic review, 103(6), 2121-2168.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. The Quarterly journal of economics, 118(4), 1279-1333.

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

Bahr, P. R., Dynarski, S., Jacob, B., Kreisman, D., Sosa, A., & Wiederspan, M. (2015). Labor Market Returns to Community College Awards: Evidence from Michigan. A CAPSEE Working Paper. Center for Analysis of Postsecondary Education and Employment.

Bayer, P., & Charles, K. K. (2018). Divergent paths: A new perspective on earnings differences between black and white men since 1940. The Quarterly Journal of Economics, 133(3), 1459-1501.

Blau, F. and Kahn, L. (1996). International differences in male wage inequality: Institutions versus market forces. Journal of Political Economy, 104(4):791–837.

Blinder, A. (1973). Wage discrimination: Reduced form and structural estimates. Journal of Human Resources, 8:436–455.

Blum, B. S. (2008). Trade, technology, and the rise of the service sector: The effects on US wage inequality. Journal of International Economics, 74(2), 441-458.

Bound, J. and Johnson, G. (1992). Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations. American Economic Review, 82(3):371–92.

Buchinsky, M. (1994). Changes in the us wage structure 1963-1987: Application of quantile regression. Econometrica: Journal of the Econometric Society, pages 405–458.

Card, D. (1999). The causal effect of education on earnings. Handbook of labor economics, 3, 1801-1863.

Card, D. and Lemieux, T. (2001). Can falling supply explain the rising return to college for younger men? a cohort-based analysis. The Quarterly Journal of Economics, 116(2):705–746.

Chongvilaivan, A., & Hur, J. (2011). Outsourcing, labour productivity and wage inequality in the US: a primal approach. Applied Economics, 43(4), 487-502.

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

El Bantli, F. and Hallin, M. (1999). L 1-estimation in linear models with heterogeneous white noise. Statistics & probability letters, 45(4):305–315.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. Handbook of labor economics, 4:1–102.

Goldin, C., & Katz, L. F. (2007). The race between education and technology: The evolution of US educational wage differentials, 1890 to 2005.

Grosz, M. (2020). The returns to a large community college program: Evidence from admissions lotteries. American Economic Journal: Economic Policy, 12(1), 226-253.

Guvenen, F., Kuruscu, B., & Ozkan, S. (2014). Taxation of human capital and wage inequality: A cross-country analysis. Review of Economic Studies, 81(2), 818-850.

Hufe, P., Kanbur, R., & Peichl, A. (2022). Measuring unfair inequality: Reconciling equality of opportunity and freedom from poverty. The Review of Economic Studies, 89(6), 3345-3380.

Hühne, P., & Herzer, D. (2017). Is inequality an inevitable by-product of skill-biased technical change?. Applied Economics Letters, 24(18), 1346-1350.

Jepsen, C., Troske, K., & Coomes, P. (2014). The labor-market returns to community college degrees, diplomas, and certificates. Journal of Labor Economics, 32(1), 95-121.

Judd, K. (1998). Approximation methods. In Press, T. M., editor, Numerical Methods in Economics, volume 1. MIT Press Books, 1 edition.

Juhn, C., Murphy, K. M., & Pierce, B. (1993). Wage inequality and the rise in returns to skill. Journal of political Economy, 101(3), 410-442.

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

Katz, L. F. and Murphy, K. M. (1992). Changes in relative wages, 1963-1987: Supply and demand factors. Quarterly Journal of Economics, 107:35–78.

Koenker, R. (2005). Inequality measures and their decomposition. In Quantile regression. Cambridge University Press, Cambridge.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. Econometrica: journal of the Econometric Society, pages 33–50.

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from Denmark. American Economic Journal: Applied Economics, 11(4), 181-209.

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

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

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. International Economic Review, 14:693–709.

Roca, J. D. L., & Puga, D. (2017). Learning by working in big cities. The Review of Economic Studies, 84(1), 106-142.

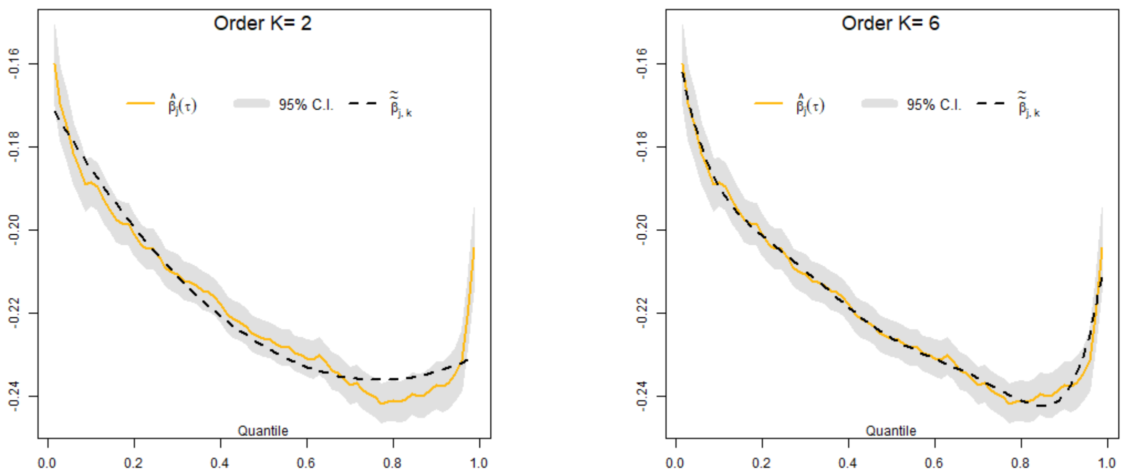
Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society. Series B (Methodological), pages 683–690.

Stevens, A. H., Kurlaender, M., & Grosz, M. (2019). Career technical education and labor market outcomes: Evidence from California community colleges. Journal of Human Resources, 54(4), 986-1036.

Figures

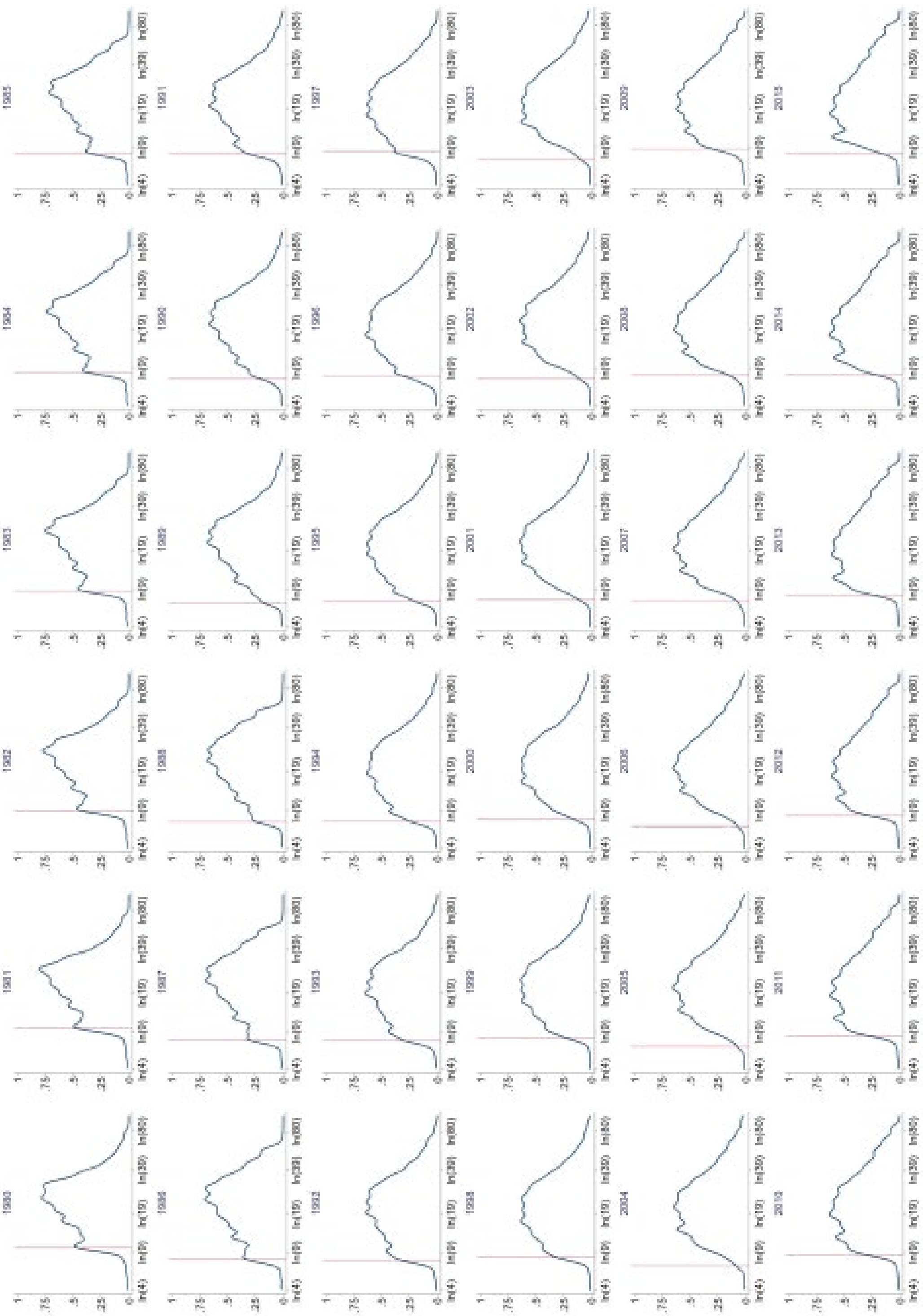
Figure 1: Example of approximation using Legendre polynomials

(a) Polynomial of degree 2 **(b)** Polynomial of degree 6



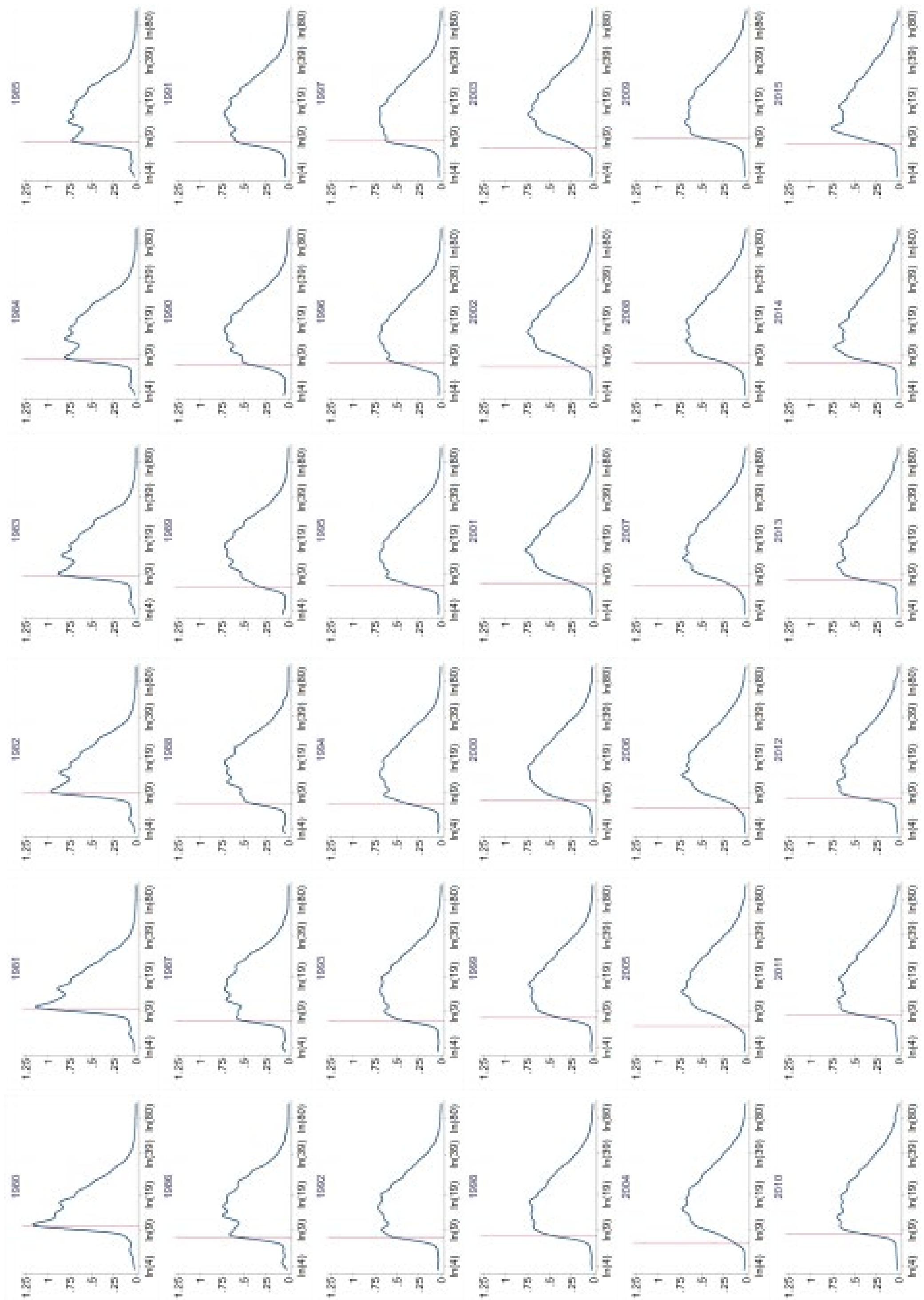
Note: The figure presents panels (a) and (b) to illustrative examples of least-square approximations on coefficients derived from a quantile regression, using Legendre polynomials of order two and six, respectively. In both panels, I showcase a generic quantile regression coefficient estimate, $\hat{\beta}_j(t)$, alongside its 95% confidence interval. The dashed line in each panel represents the least-square approximation using Legendre polynomials, $\tilde{\tilde{\beta}}_{j,K}(t)$, for $K = 2, 6$, respectively

Figure 2a: Kernel density estimates of men's (log) real wages 1980 - 2015 ($2020)



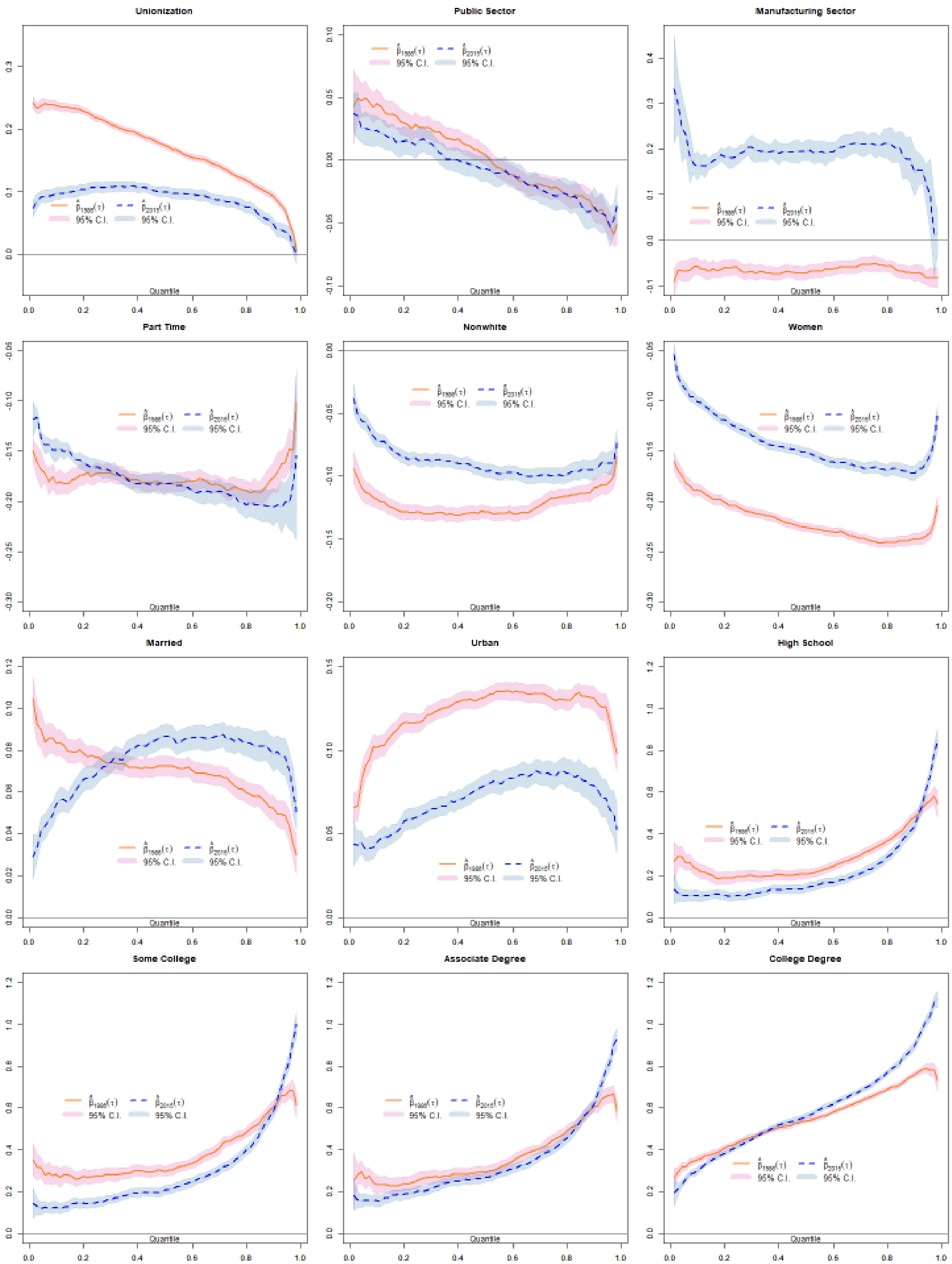
Note: The figure presents kernel density estimates for hourly (log) wages of men, covering the years 1980 through 2015. In each panel, a vertical line indicates the federal minimum wage for that specific year. I have converted all wage values to 2020 USD using the CPI series CUSR0000SA0. This figure draws on my sample from the CPS ORG, employing the CEPR Uniform Extracts to create a consistent hourly wage series. I have narrowed my focus to include workers aged 16 to 65, with hourly wages ranging from $1 to $100 in 1979 dollars.

Figure 2b: Kernel density estimates of women's (log) real wages 1980 - 2015 ($2020)



Note: The figure presents kernel density estimates for hourly (log) wages of women, covering the years 1980 through 2015. In each panel, a vertical line indicates the federal minimum wage for that specific year. I have converted all wage values to 2020 USD using the CPI series CUSR0000SA0. This figure draws on my sample from the CPS ORG, employing the CEPR Uniform Extracts to create a consistent hourly wage series. I have narrowed my focus to include workers aged 16 to 65, with hourly wages ranging from $1 to $100 in 1979 dollars.

Figure 3: Selected Coefficients Estimates From the Quantile Regression



Note: The figure shows selected quantile regression coefficient estimates for the model in Equation (27), computed for a grid of 69 equally-spaced points across the (0,1) interval. Each panel in the figure contrasts the estimates from two years, 1986 ($\Psi = 0$) and 2015 ($\Psi = 1$), with the solid line representing 1986 and the dashed line depicting 2015. The shaded areas surrounding each line denote the 95% confidence intervals, calculated using a Huber sandwich estimate with a local estimate of the sparsity. The regression uses comprehensive hourly wage data from the CPS ORG. The analysis incorporates a range of individual, job-related, demographic, and macroeconomic attributes, along with state and industry fixed effects. The figure demonstrates the shift in wage dynamics over time, reflecting broader economic and structural changes, including the impact of technological advancements and sectoral shifts in the economy.

Tables

Table 1: Summary Statistics for the CPS 1980-2015

			Men						Women					
Year	Log Real Minimum Wage^A	Manufact. Employ.	Log Real Wage^A	Union^B	Nonwhite	Education	Experience^C	No. Obs.	Log Real Wage^A	Union^B	Nonwhite	Education	Experience^C	No. Obs.
1980	2.28	0.26	3.20		0.17	12.75	18.26	106,932	2.79		0.18	12.76	17.64	87,888
1981	2.26	0.25	3.18		0.18	12.82	18.19	99,527	2.78		0.19	12.80	17.68	83,229
1982	2.20	0.24	3.18		0.18	12.92	18.23	92,249	2.80		0.19	12.91	17.68	79,594
1983	2.16	0.23	3.18	0.28	0.18	12.99	18.11	91,053	2.81	0.18	0.19	12.99	17.59	79,017
1984	2.12	0.23	3.17	0.26	0.18	13.02	17.91	92,732	2.82	0.17	0.19	13.03	17.53	80,796
1985	2.09	0.23	3.17	0.25	0.20	13.03	18.02	93,768	2.83	0.16	0.20	13.07	17.58	82,833
1986	2.07	0.22	3.19	0.24	0.20	13.07	17.94	92,085	2.86	0.16	0.20	13.12	17.64	83,459
1987	2.03	0.22	3.18	0.23	0.21	13.08	18.00	92,014	2.87	0.15	0.21	13.15	17.69	84,693
1988	1.99	0.21	3.17	0.23	0.22	13.11	17.99	88,096	2.87	0.15	0.21	13.19	17.78	81,205
1989	1.95	0.21	3.19	0.22	0.22	13.14	18.09	89,459	2.88	0.15	0.22	13.23	18.02	82,998
1990	2.02	0.20	3.17	0.21	0.23	13.12	18.05	93,498	2.88	0.15	0.23	13.27	18.01	87,323
1991	2.09	0.20	3.16	0.21	0.24	13.18	18.24	90,154	2.89	0.15	0.23	13.32	18.26	85,381
1992	2.06	0.19	3.15	0.21	0.24	13.00	18.55	88,364	2.90	0.15	0.23	13.14	18.64	84,530
1993	2.03	0.19	3.14	0.20	0.24	13.06	18.59	86,820	2.91	0.15	0.23	13.20	18.78	83,925
1994	2.00	0.19	3.14	0.20	0.24	13.09	18.62	82,362	2.92	0.15	0.24	13.24	18.81	80,375
1995	1.98	0.19	3.13	0.19	0.24	13.12	18.74	82,515	2.90	0.14	0.24	13.26	18.96	79,910
1996	2.06	0.18	3.12	0.19	0.26	13.12	18.98	73,043	2.90	0.14	0.25	13.29	19.08	71,486
1997	2.12	0.18	3.14	0.18	0.27	13.10	19.09	74,587	2.92	0.14	0.26	13.30	19.27	72,771
1998	2.10	0.18	3.18	0.18	0.27	13.13	19.22	75,596	2.95	0.13	0.27	13.32	19.34	73,483
1999	2.08	0.17	3.21	0.18	0.27	13.18	19.34	76,758	2.97	0.13	0.27	13.35	19.45	74,452
2000	2.05	0.17	3.22	0.17	0.28	13.18	19.47	77,723	2.98	0.13	0.28	13.36	19.60	75,175
2001	2.02	0.16	3.23	0.17	0.28	13.23	19.73	82,354	3.00	0.13	0.28	13.41	19.85	80,052
2002	2.00	0.15	3.24	0.16	0.28	13.26	20.01	87,802	3.03	0.13	0.28	13.46	20.09	86,481
2003	1.98	0.14	3.24	0.16	0.31	13.23	20.16	85,508	3.03	0.13	0.30	13.50	20.46	85,314
2004	1.95	0.13	3.24	0.15	0.32	13.25	20.25	84,298	3.03	0.13	0.30	13.54	20.55	83,302
2005	1.92	0.13	3.22	0.15	0.32	13.24	20.46	84,895	3.03	0.13	0.31	13.58	20.66	83,789
2006	1.89	0.13	3.22	0.14	0.33	13.26	20.51	85,018	3.03	0.12	0.31	13.60	20.79	82,926
2007	1.99	0.13	3.23	0.14	0.33	13.31	20.60	83,742	3.04	0.13	0.32	13.67	20.86	82,382
2008	2.06	0.13	3.23	0.15	0.33	13.39	20.81	82,316	3.05	0.13	0.32	13.75	21.03	81,683
2009	2.17	0.12	3.26	0.15	0.33	13.46	21.15	78,900	3.07	0.13	0.32	13.80	21.29	80,037
2010	2.15	0.12	3.24	0.14	0.33	13.49	21.26	78,101	3.07	0.13	0.32	13.85	21.43	78,973
2011	2.12	0.12	3.22	0.14	0.34	13.52	21.19	77,791	3.06	0.13	0.32	13.89	21.50	77,681
2012	2.10	0.12	3.22	0.13	0.35	13.56	21.35	78,104	3.05	0.12	0.35	13.92	21.51	76,849
2013	2.09	0.12	3.22	0.13	0.36	13.58	21.34	78,071	3.06	0.12	0.35	13.99	21.44	76,426
2014	2.07	0.12	3.21	0.13	0.37	13.59	21.35	78,762	3.06	0.12	0.36	14.02	21.34	76,565
2015	2.07	0.12	3.24	0.13	0.37	13.63	21.29	77,812	3.08	0.12	0.37	14.05	21.28	75,642

^{A} 2020 Constant Dollars

^{B} Union status of workers was not collected in the outgoing rotation group supplements from 1980 to 1982. However, using the May pension supplement it may be possible estimate this summary statistic for a subsample of the population

^{C} Potential experience is computed as age - years of education - 5

Note: The table presents summary statistics for my refined sample from the CPS ORG, utilizing the CEPR Uniform Extracts to generate a consistent series of hourly wages. All wages have been adjusted to 2020 USD using the CPI series CUSR0000SA0. My focus is on workers aged 16 to 65, earning hourly wages between $1 and $100, adjusted to 1979 dollars. Potential experience is calculated by subtracting the number of years of education and an additional five years for elementary schooling from each individual's age. Both education and potential experience are expressed in years. The columns labeled "manufacturing," "union," and "nonwhite" represent the proportion of workers in manufacturing jobs, unionized positions, or those who did not identify as white, respectively. All summary statistics are weighted by the CPS sample weights.

Table 2: Impact Estimates of Selected Covariates

	1986	2015
Unionization status	0.071	0.032
	0.069; 0.073	0.029; 0.035
Public sector job	0.006	0.001
	0.003; 0.010	-0.002; 0.005
Manufacturing job	-0.022	0.066
	-0.028;-0.017	0.059; 0.074
Potential Experience	0.0078	0.0052
	0.0072;0.0084	0.0047;0.0058
Part time employee	-0.059	-0.056
	-0.062;-0.055	-0.059;-0.052
Nonwhite	-0.044	-0.028
	-0.046;-0.042	-0.030;-0.026
Female	-0.07	-0.041
	-0.072;-0.068	-0.043;-0.040
Married	0.027	0.023
	0.025; 0.028	0.021; 0.025
Urban area	0.041	0.021
	0.039; 0.043	0.018; 0.023
High school	0.081	0.04
	0.072; 0.090	0.034; 0.047
Some college	0.109	0.058
	0.098; 0.119	0.051; 0.065
Associate degree	0.099	0.076
	0.088; 0.111	0.069; 0.084
College degree	0.161	0.153
	0.157; 0.166	0.149; 0.158

Note: The table presents the impact of individual characteristics on wage inequality for years 1986 and 2017. A positive sign of the reported impact is associated with a reduction in the inequality of the distribution of (log) wages given a small increase in the corresponding characteristic. The first entry of each cell in the table presents the impact estimation, whereas the second reports the 95% bootstrap confidence interval. Each column exhibits the results for the corresponding year. I computed the figures using the estimation procedure described in section 3. The polynomial approximation order is 7 for both years. The bootstrap uses 1,000 repetitions for each year. For the bootstrap, in each iteration, I calculated the $\hat{\mu}_{ln}$ using the weighted average of the logarithm of real wages, adjusted to 2020 dollar value; Then, I compute the estimate $\widehat{\Pi}_j$ from Equation (26) and compute the impact estimate $\widehat{\Pi}_j/\hat{\mu}_{ln}$

Table 3a: Decomposition of Wage Changes (1986-2015) Using the Machado and Mata Method

	Marginals			Aggregate Contributions			Individual Covariates								
	1986 (1)	2015 (2)	Change (3)	Covariates (4)	Returns (5)	Residual (6)	Unionization (7)	Manufacturing (8)	Nonwhite (9)	Women (10)	Urban (11)	High School (12)	Some College (13)	Associate Degree (14)	College Degree (15)
1st quant.	1.56	1.86	0.305	-0.027	0.288	0.044	0.030	0.038	-0.012	0.043	0.042	0.050	0.040	0.029	0.087
			0.284;0.328	-0.075;0.009	0.242;0.347		-0.014;0.083	-0.033;0.078	-0.057;0.058	0.001;0.092	-0.020;0.084	-0.004;0.086	-0.019;0.089	-0.030;0.071	0.017;0.146
10th quant.	2.15	2.29	0.132	0.015	0.082	0.035	-0.007	-0.010	-0.034	0.026	0.006	0.031	0.008	0.017	0.051
			0.132;0.136	-0.008;0.035	0.043;0.104		-0.034;0.017	-0.032;0.014	-0.063;-0.005	-0.003;0.051	-0.022;0.027	-0.001;0.053	-0.020;0.038	-0.009;0.040	0.027;0.073
25th quant.	2.47	2.53	0.066	0.029	0.005	0.032	-0.020	-0.016	-0.046	0.004	0.014	0.014	0.012	-0.004	0.046
			0.063;0.071	0.006;0.047	-0.020;0.028		-0.053;0.005	-0.043;0.009	-0.074;-0.021	-0.024;0.031	-0.011;0.038	-0.012;0.041	-0.016;0.035	-0.031;0.025	0.019;0.070
Median	2.87	2.94	0.065	0.087	-0.041	0.019	-0.019	-0.004	-0.040	0.004	0.025	0.032	0.005	-0.001	0.067
			0.053;0.066	0.062;0.107	-0.063;-0.013		-0.049;0.006	-0.033;0.022	-0.075;-0.015	-0.025;0.029	-0.004;0.051	0.006;0.058	-0.028;0.034	-0.031;0.026	0.039;0.098
75th quant.	3.30	3.39	0.089	0.131	-0.025	-0.016	-0.050	-0.034	-0.068	-0.022	-0.027	0.040	-0.023	-0.001	0.092
			0.079;0.098	0.101;0.157	-0.055;0.003		-0.091;-0.013	-0.068;0.002	-0.105;-0.036	-0.056;0.015	-0.065;0.009	0.004;0.077	-0.063;0.013	-0.039;0.033	0.051;0.125
90th quant.	3.63	3.81	0.179	0.130	0.023	0.026	-0.061	-0.045	-0.096	-0.047	-0.041	-0.010	-0.039	-0.016	0.067
			0.173;0.195	0.086;0.178	-0.026;0.066		-0.107;-0.026	-0.080;0.003	-0.139;-0.052	-0.085;-0.013	-0.077; 0.000	-0.056;0.038	-0.075;0.006	-0.059;0.028	0.023;0.121
99th quant.	4.08	4.37	0.290	0.134	0.105	0.052	0.004	0.009	-0.033	0.016	0.024	0.021	0.002	0.022	0.090
			0.290;0.290	0.058;0.197	0.040;0.186		-0.048;0.058	-0.041;0.081	-0.079;0.024	-0.038;0.078	-0.034;0.094	-0.034;0.107	-0.057;0.060	-0.036;0.080	0.025;0.150
Gini of LogW	11.04	11.09	0.05	-1.7019	2.8591	-1.106	-0.231	-0.096	-0.173	-0.321	-0.246	-0.176	-0.305	-0.088	0.013
			-0.554;0.664	-2.767;-0.577	2.394;3.327		-0.494;0.033	-0.356;0.169	-0.441;0.093	-0.590;-0.053	-0.511;0.020	-0.446;0.097	-0.567;-0.041	-0.349;0.173	-0.250;0.276

Note: The table presents the results of the Machado and Mata decomposition applied to wage data. The first two columns list the estimated log-transformed wages in 1986 and 2015 for selected quantiles. The third column reports the changes in these quantiles over the period. Columns four to six break down the total changes in the wage distribution into segments associated with covariates, variations in returns, and the residual component. The final columns, seven to fifteen, illustrate the effects of specific individual covariates. All point estimates come with 95% bootstrap confidence intervals, derived from 1,000 bootstrap samples and displayed beneath each computed value.

Table 3b: Wage Distribution Shifts (1986-2015) Using an Additive Decomposition of Gini Index

	Marginals			Aggregate Contributions			Individual Covariates								
	Gini 1986 (1)	Gini 2015 (2)	Change (3)	Covariates (4)	Returns (5)	Residual (6)	Unionization (7)	Manufacturing (8)	Nonwhite (9)	Women (10)	Urban (11)	High School (12)	Some College (13)	Associate Degree (14)	College Degree (15)
Q1	1.82	1.68	-0.1460	-0.0127	0.0756	-0.2088	0.0143	0.0462	0.0223	0.0034	-0.0064	0.0237	-0.0155	0.0012	-0.0685
			-0.278;-0.021	-0.062;0.039	0.041;0.108		0.008;0.021	0.030;0.063	0.017;0.028	-0.006;0.013	-0.009;-0.003	0.015;0.032	-0.024;-0.007	-0.010;0.011	-0.094;-0.042
Q2	3.70	3.61	-0.0867	-0.2067	0.4840	-0.3640	0.0466	0.1263	0.0814	0.0127	-0.0236	0.0735	-0.0543	0.0042	-0.2783
			-0.301;0.126	-0.380;-0.023	0.394;0.573		0.026;0.068	0.083;0.171	0.061;0.104	-0.021;0.048	-0.035;-0.011	0.047;0.099	-0.085;-0.026	-0.034;0.039	-0.380;-0.173
Q3	3.69	3.81	0.1212	-0.5305	0.9510	-0.2993	0.0787	0.2128	0.1471	0.0241	-0.0468	0.1439	-0.1101	0.0084	-0.5660
			-0.113;0.355	-0.863;-0.179	0.807;1.096		0.044;0.115	0.139;0.288	0.111;0.188	-0.040;0.092	-0.068;-0.023	0.093;0.195	-0.172;-0.052	-0.068;0.077	-0.773;-0.351
Q4	1.83	1.99	0.1629	-0.9520	1.3486	-0.2336	0.1042	0.3024	0.2155	0.0366	-0.0730	0.2862	-0.2123	0.0152	-0.9412
			0.041;0.280	-1.470;-0.421	1.151;1.547		0.058;0.153	0.198;0.410	0.162;0.275	-0.060;0.139	-0.107;-0.035	0.185;0.387	-0.331;-0.100	-0.122;0.139	-1.285;-0.583
Total	11.04	11.09	0.0514	-1.7019	2.8591	-1.1058	0.2438	0.6877	0.4663	0.0768	-0.1499	0.5274	-0.3922	0.0291	-1.8540
			-0.554;0.664	-2.767;-0.577	2.394;3.327		0.136;0.357	0.450;0.932	0.351;0.596	-0.126;0.293	-0.219;-0.073	0.340;0.713	-0.611;-0.184	-0.234;0.266	-2.532;-1.149

Note: The table details the shifts in wage distribution from 1986 to 2015, employing an additive decomposition of the Gini index as detailed in Equation (14). Each entry in Table 3b comprises two parts: the point estimate followed by the 95% bootstrap confidence intervals, obtained from 1,000 bootstrap samples. Negative point estimates in this table imply a reduction in the Gini coefficient, indicating a decline in wage inequality. The table spans four quartiles and the entire wage distribution, offering insights into the dynamics of wage changes and the impact of diverse socioeconomic factors.

Appendices

A. The Lorenz Curve as an Expected Value

Consider Z, a continuous random variable with support in R. Let its cumulative distribution function be $F_Z(z)$ and its probability density function be $f_Z(z)$. Assume that the conditional expectation $E[z|z \leq a]$ exists and is finite for every $a \in R$. For a given τ in the interval (0,1), define the quantile function $Q_Z(t) = \inf\{z: F_Z(z) \geq t\} = F_Z^{-1}(t)$, and denote $z_\tau = Q_Z(\tau)$.

Then, the integral of $Q_Z(t)$ from 0 to τ can be expressed as follows:

$$\begin{aligned}\int_0^\tau Q_Z(t)dt &= \int_{-\infty}^{z_\tau} z f_Z(z)dz \\ &= \tau \int_{-\infty}^{z_\tau} z \frac{f_Z(z)}{F_Z(z_\tau)} dz \\ &= \tau \int_{-\infty}^{z_\tau} z f_{Z<z_\tau}(z)\, dz \\ &= \tau E[z|z \leq z_\tau]. \end{aligned} \tag{28}$$

From Equation (28), it becomes evident that the integral of $Q_Z(t)$ from 0 to 1 is equal to the expected value of z.

Moreover, $\forall \tau \in (0,1)$

$$E[z|z \leq z_\tau] = \int_{-\infty}^{z_\tau} z \frac{f_Z(z)}{F_Z(z_\tau)} dz \leq z_\tau \int_{-\infty}^{z_\tau} \frac{f_Z(z)}{F_Z(z_\tau)} = z_\tau$$

and,

$$z_\tau = z_\tau \int_{z_\tau}^{\infty} \frac{f_Z(z)}{1 - F_Z(z_\tau)} dz \leq \int_{z_\tau}^{\infty} z \frac{f_Z(z)}{1 - F_Z(z_\tau)} dz = E[z|z \geq z_\tau].$$

Then, $\forall \tau \in (0,1)$

$$0 \leq (1-\tau)(E[z|z \geq z_\tau] - E\ [z|z \leq z_\tau]),$$

which implies

$$E[z|z \leq z\tau] \leq \tau E[z|z \leq z\tau] + (1-\tau)E[z|z \geq z\tau] = E[z]. \tag{29}$$

Let Y be a continuous and positive random variable, with cumulative density function $F_Y(y)$, quantile function denoted by $Q_Y(t) = \inf\{y: F_Y(y) \geq t\} = F_Y^{-1}(t)$, and $y_\tau = Q_Y(\tau)$. Assume that $0 < E[y] < \infty$. Let $h(\cdot)$ be a continuous and monotone function. Define $Z = h(Y)$ and $\mu_h = E[h(y)] = E[z]$. Assume that $h(\cdot)$ is such that $h(Y) \geq 0$ and $0 < \mu_h < \infty$. By the properties of the quantile function, $Q_{h(Y)}(t) = h\big(Q_Y(t)\big)$. Then, using Equation (28), the Lorenz curve of the transformed variable is given by

$$L_h(\tau) = \frac{1}{\mu_h}\int_0^\tau Q_{h(Y)}(t)dt = \frac{\tau E[h(t)|h(t) \leq h(y_\tau)]}{E[h(y)]}.$$

Using the inequality in (29), the transformed Lorenz curve takes values between 0 and 1.

By the definition of the Gini coefficient, we have

$$\begin{aligned} G_h &= 1 - 2\int_0^1 L_h(\tau)d\tau \\ &= 1 - \frac{2}{\mu_h}\int_0^1 \tau E[h(y)|h(y) \leq h(y_\tau)]\, d\tau. \end{aligned}$$

The Gini index, G_h, is always less than or equal to 1 because $h(Y) \geq 0$. Furthermore, based on the inequality presented in Equation (29), we can deduce that

$$E[h(y)|h(y) \leq h(y_\tau)] \leq E[h(y)],$$

which implies

$$\int_0^1 \tau E[h(y)|h(y) \leq h(y_\tau)]d\tau \leq \int_0^1 \tau E[h(y)]d\tau = \frac{1}{2}E[h(y)] = \frac{\mu_h}{2}.$$

In other words, the Gini index, G_h, is always positive.

B. Integral Approximation

The Gini index takes the form

$$G_h(x) = 1 - \frac{1}{\mu_h}\sum_{j=1}^{P} x_j \int_0^1 \int_0^\tau 2\beta_j(\tau) dt\, d\tau.$$

I use the family of Legendre polyonomy of degree K to approximate each quantile regression coefficient estimate, $\hat{\beta}_j(t)$:

$$\hat{\beta}_j(t) \approx \tilde{\tilde{\beta}}_{j,K}(t) = \hat{\alpha}_0 p_0(t) + \cdots + \hat{\alpha}_K p_K(t) = \sum_{i=0}^{K} \hat{\alpha}_i p_i(t).$$

Finally, I approximate the impact of characteristic j as

$$\frac{2}{\mu_h}\sum_{i=0}^{K} \hat{\alpha}_i \int_0^1 \int_0^\tau p_i(t)\, d\tau.$$

To compute the vector $\hat{\alpha} = (\hat{\alpha}_0, \cdots, \hat{\alpha}_K)$, first create a fixed grid t_{n_g} with n_g equally-spaced points on (0,1). Using statistical software, compute n_g quantile regression coefficient estimates, one for each point on the grid. Represent these estimates as a $1 \times n_g$ vector $\hat{\beta}_j\left(t_{n_g}\right)$. In a similar fashion, define $p_i\left(t_{n_g}\right)$ as the $1 \times n_g$ vector that computes the i-th Legendre polynomial at each grid point. Subsequently, define the $n_g \times (K+1)$ matrix $P\left(t_{n_g}\right)$ with each $p_i\left(t_{n_g}\right)$ as its columns, resulting in $P\left(t_{n_g}\right) = \left[p_0\left(t_{n_g}\right), \cdots, p_K\left(t_{n_g}\right)\right]$. The values of $\hat{\alpha}_i$ are then computed as the scalars that minimize the squared error between $\hat{\beta}_j\left(t_{n_g}\right)$ and $P\left(t_{n_g}\right) \times \hat{\alpha}$. In other words, the entries of the vector $\hat{\alpha}$ are the OLS coefficient estimates from the model with $\hat{\beta}_j\left(t_{n_g}\right)$ as the dependent variable and the design matrix $P\left(t_{n_g}\right)$.

C. Accuracy of the Estimating Procedure

Set $h(\cdot) = ln(\cdot)$ and assume that the conditional quantile function of the logarithm of y can be modeled by a simple linear relation:

$$Q_{\ln(y)}(t|x) = \beta_0(t) + x\beta_1(t) + \varepsilon(t).$$

For simplicity, let's assume that the intercept is constant, e.g., $\beta_0 = 0.5$, and the slope parameter increases linearly with quantiles, e.g., $\beta_1(t) = 0.2 + 0.05t$. Under these simplifying assumptions, the numerator of the impact estimates would be:

$$\Pi_0 = 2\int_0^1 \int_0^\tau 0.5 dt\, d\tau = 0.5$$

and

$$\Pi_1 = 2\int_0^1 \int_0^\tau 0.2 + 0.05t dt\, d\tau = 0.21667$$

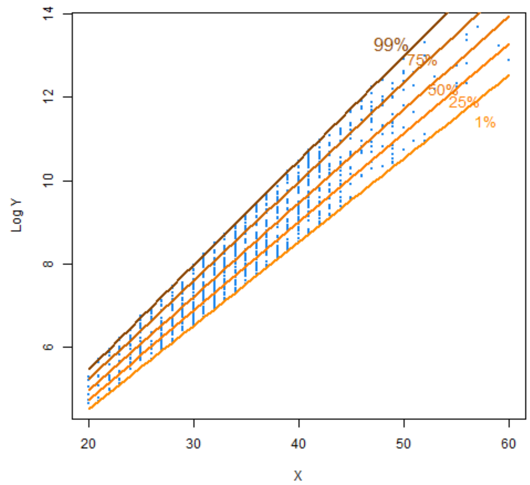
To evaluate the precision of the estimation procedure, I generated a simulated dataset comprising $N = 1{,}000$ observations. For this dataset, the variable x is normally distributed, truncated at 20, and characterized by a mean of 35 and a standard deviation of 8. I set the error term to be uniformly distributed between zero and one, and to increase proportionally with x by 0.05 to achieve a slope that linearly rises with quantiles. Panel (a) of Figure C.1 showcases the simulated dataset, featuring selected estimated lines representing the conditional quantile functions. Panel (b) of the same figure displays the estimated conditional quantile regression coefficients marked with dots, in contrast to the actual conditional quantile regression coefficient, depicted as a solid line.

I estimate the numerator of the impact using the integral approximation detailed in Appendix B. Varying the order of polynomial approximation from K =2 to 10, I compute the 95% bootstrap confidence intervals using the 2.5th and 97.5th quantiles from 1,000 repetitions. Table C.1 presents the results of this performance test. The first column of the table displays the exact numerators of the impact estimates. Columns two through ten reveal the estimated results using Legendre polynomials of the corresponding orders. Each estimate's cell includes the 95%

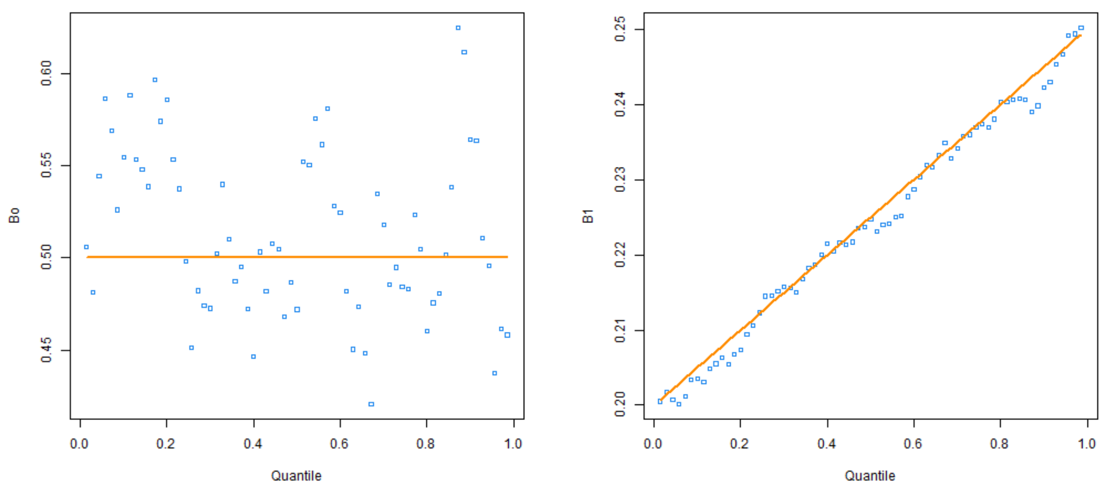
bootstrap confidence interval. The final two rows of the table feature the results of the hypothesis tests $\widehat{\Pi}_i \neq \Pi_i$ for $i = 0,1$. The key insight from this table is the accuracy of the procedure and the minimal effect of the polynomial approximation's order on the accuracy of the estimation. This exercise validates the estimation procedure's effectiveness in quantifying the covariates' impact.

Figure C.1:

(a) Simulated dataset and selected conditional quantile relationships



(b) Coefficient estimates and actual parameters



Note: Panel (a) showcases the simulated dataset, and panel (b) displays the estimated quantile regression coefficients. The dataset in panel (a) comprises $N = 1{,}000$ observations with the variable x following a normal distribution, truncated at 20. This distribution has a mean of 35 and a standard deviation of 8. The simulation's error term is uniformly distributed between zero and one, incrementally increasing in proportion to x by 0.05 to replicate a slope that linearly ascends with the quantiles. Panel (b) compares the estimated conditional quantile regression coefficients marked with dots, against the actual conditional quantile regression coefficient, shown as a solid line.

Table C.1: Performance Test Results Using Various Orders of Polynomial Approximation

	Exact Impact (1)	Order 2 (2)	Order 3 (3)	Order 4 (4)	Order 5 (5)	Order 6 (6)	Order 7 (7)	Order 8 (8)	Order 9 (9)	Order 10 (10)
Intercept	0.50000	0.51978	0.51994	0.51938	0.51909	0.51782	0.51833	0.51898	0.51822	0.51843
		0.3712;0.6683	0.3715;0.6684	0.3708;0.668	0.3712;0.667	0.3701;0.6655	0.3698;0.6669	0.3705;0.6675	0.3701;0.6663	0.3699;0.667
x	0.21667	0.21576	0.21575	0.21576	0.21577	0.21582	0.21580	0.21577	0.21580	0.21580
		0.21132;0.2202	0.21131;0.22019	0.21132;0.2202	0.21135;0.22019	0.2114;0.22023	0.21136;0.22025	0.21133;0.22022	0.21137;0.22023	0.21136;0.22024
$\hat{\Pi}_0 \neq \Pi_0$		Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
$\hat{\Pi}_1 \neq \Pi_1$		Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

Note: The table shows the results from the performance test. The numerator of the impact is estimated using various orders of polynomial approximation, ranging from $K = 2$ to 10. The first column provides the exact numerators of the impact estimates, while columns two through ten display the estimated results using Legendre polynomials of each specified order. Below each estimate, the corresponding 95% bootstrap confidence intervals are provided. The bootstrap confidence intervals are derived from 1,000 repetitions. The last two rows of the table show the results of the hypothesis test $\hat{\Pi}_i \neq \Pi_i$ for $i = 0,1$, underlining the accuracy of the procedure and the minimal impact of the polynomial approximation's order on this accuracy.