Estimating Jointly Determined Outcomes: How Minimum Wage affects Wages and Hours Worked

Yisroel Cahn*

September 1, 2021

[Most Recent Version]

Abstract

This paper goes beyond the standard single outcome framework for evaluating policy and addresses multiple jointly determined outcomes. It does so by examining the effects of minimum wage policy on both hourly wages and hours worked. While changes to the minimum wage might increase an individual's hourly wage, they might also decrease that individual's number of hours worked. Since wages and hours worked are contemporaneously determined, a conditional effect of minimum wage policy on wages or hours worked is conditioned on an endogenous variable. This obstacle is overcome by examining the joint distribution of outcomes. Using distribution regressions and an empirical copula to estimate the counterfactual joint distribution of wages and hours worked in the United States, this paper finds increased Federal minimum wage causes no wage-hour substitution phenomenon and that both wages and hours worked increased across the entire joint distribution for both men and women.

Keywords: Minimum Wages, Wages and Hours Worked, Joint Distribution, Copula, Distribution Regression

JEL Codes: J08, J22, J31, C31

---

*Emory University, Atlanta, GA 30322. Email: ycahn@emory.edu. I am deeply grateful to Esfandiar Maasoumi for his guidance and support. I would like to thank the Emory University faculty, participants of the Society of Labor Economists annual conference, and participants at the 2021 Summer School on Socioeconomic Inequality for their helpful comments and suggestions. All errors are my own.

1 Introduction

1.1 The Minimum Wage Puzzle

The main goal of minimum wage policy is to assist low-wage workers. However, the effectiveness of minimum wage as a poverty or inequality reducing measure is questionable. While increased minimum wage reduced wage inequality in the United States,[1] and had mild effects on employment,[2] little is known about its effects on hours worked[3] which might vary for individuals earning different wages and working different hours. By analyzing the joint distribution of wages and hours worked, this paper offers insight into whether increased minimum wage reduces poverty and inequality, and what mechanisms cause any such change.

Generally, papers studying the effects of minimum wage on inequality are only concerned with the change in the wage distribution and the change in employment statuses. Such studies usually include an analysis of various indices (e.g. differences between wage quantiles or the Gini coefficient) which address certain properties thought to define inequality or a level of income deemed to represent a level of poverty. However, there are different ways to consider an individual "poorer" or "worse off" besides wages. If, for example, a policy increases the wages of every individual in the population but has a negative effect on the health outcomes, cost of living, or educational attainment of some individuals, it is arguable that the policy increases inequality rather than decreasing it. While it is possible to attribute more or less "weight" (importance) to different measures of well-being at different parts of the distribution (e.g. one unit of health status at the bottom of the wage distribution is worth two units of wages, whereas one unit of health status at the top of the wage distribution is worth three units of wages at that level), such a weighting tends to be arbitrary and controversial. Additionally, Atkinson and Bourguignon (1982) point out another issue with this weighting scheme, namely, that transfers in one measure can change the marginal utility of another measure (e.g. if an individual gains or loses a unit of educational attainment, she might now value an increase in a unit of wages differently).

Although, for example, changes in health status are not likely to be caused by increased minimum wage, changes in educational attainment, cost of living, or hours worked due such increases are likely drastic and vary across the distribution. Consider Figure 1. It is possible for every individual's wage to be increased but for some individuals to experience a decrease in hours worked

[1] For example, see DiNardo et al., 1996; Lee, 1999; and Autor et al., 2016.

[2] Manning (2021) points out "[t]here is probably no economist who does not believe that there is some point at which higher minimum wages reduce employment." While that may be true, minimum wage increases in the United States at both the federal and state level have generally been small. As discussed in section 2, there is a growing consensus that these increases had little to no effect on employment and further minimum wage research should now focus on determining when minimum wage would cause a negative employment effect.

[3] Researchers studying the effects of Seattle's 2015 and 2016 minimum wage increases initially found the policies had negative impact on the hours worked of low-wage workers, but in a follow-up paper, the same researchers found little to no impact (Jardim et al., 2017; Jardim et al., 2018).

(assuming employment remains constant). Perhaps employers do not adjust for increases in minimum wage by discharging workers or reducing the wages paid to other employees. Instead, they may respond by increasing the responsibilities and hours of high-wage workers who are more productive than low-wage workers and reduce the hours of low-wage workers. Clearly, low-wage individuals value increased hourly wages less when hours worked decrease. Only considering the marginal distribution of hourly wages would lead to the conclusion that inequality was reduced, but considering the joint distribution of hourly wages and hours worked might lead to the conclusion that there was an increase in inequality! More details on comparing welfare of multivariate distributions can be found in Section 4.

Of course, the obvious question remains — why not try to estimate quantile treatment effects of wages conditional on hours worked, a more traditional approach? Doing so estimates the *ceteris paribus* effect of the policy on a quantile of wages assuming the policy does not contemporaneously affect hours worked, an endogenous variable. However, the policy could affect wages and hours worked of individuals very differently across the joint distribution (e.g. low-wage high-hour workers could be affected differently than low-wage low-hour workers). Therefore, the real object of interest for welfare analysis should be some comparison of the joint distribution of outcomes of interest with a policy and the counterfactual joint distribution without the policy.[4] The next subsection discusses in more detail the importance of considering the joint distribution of outcomes when outcomes are jointly determined.

Extending the method of Chernozhukov et al. (2013) for estimating counterfactual distributions using distribution regression, this paper employs a copula method and compares joint distributions of hourly wage and hours worked with and without a minimum wage change. Details on this method are laid out in Section 3. Additional robustness checks can be found in section 6.2 and a selection model accounting for possible employment effects can be found in section 6.3.

*******(expand discussion of results)******* This paper finds that both wages and hours worked increased across the entire joint distribution for both men and women.

This could possibly be due to the "backward bending" nature of the wage curve, which describes the labor-leisure trades-offs of workers. As a worker's wages increases, she prefers to work more hours and have less leisure time. However, at a certain wage level, this preference reverses and she would prefer to worker fewer hours as her wage increases. Since a large portion of minimum wage workers in the United States are paid by the hour,[5] a likely explanation

[4]Heckman (2010) formulates three classes of problems in policy evaluation: (P1) evaluating the impact of the policy; (P2) forecasting the impacts of the policy (constructing counterfactual states); and (P3) forecasting the impact of the policy never historically experienced (constructing counterfactual states associated with the policy). The class of problems (P1), which makes up a large portion of policy evaluation literature, generally focuses on means or conditional means (i.e. subgroups of the population). The argument here is that class of problems (P1) should also focus on joint distributions.

[5]See Figure 6. A large portion of US minimum wage workers are also part-time; see Figure

for this paper's findings is that low-wage workers are choosing to work more hours as their wages increase. Perhaps, it is not employers who are setting their workers' hours, but workers are choosing their own hours. This story is consistent with the two recent papers on Seattle's increased minimum wage (Jardim et al., 2017; Jardim et al., 2018) which suggested that workers were working additional hours at their secondary jobs.

1.2 Jointly Dependent Outcomes

Policy evaluation is chiefly concerned with comparing the observed outcomes after a policy has been implemented with the unobserved potential outcomes had the policy not been implemented.[6] Often, this is done by comparing the estimated mean of an observed outcome against the estimated mean of a counterfactual outcome had the policy not been implemented (i.e. average treatment effect). This can also be done conditioning on some group (e.g. conditional average treatment effect or average treatment effect on the treated). However, if the policy has heterogeneous effects on the distribution of outcomes—for example, a policy affects low-wage workers differently than high-wage workers, then simply comparing means masks the diversity of outcomes a policy maker might be interested in. This is particularly relevant if the policy maker is interested in a policy's effect on inequality or poverty.

While methods of comparing entire distributions[7] and quantiles[8] of outcomes have been employed, many applied researchers simply split the data into groupings of the distribution they are interested in (e.g. splitting the data by high- and low-wage workers) and estimate the mean effect in each grouping. An issue with using informal data splitting to account for outcome heterogeneity is that these groupings can be arbitrary, with results obtained from them possibly being an artifact of the arbitrary grouping decisions.

Both the sample splitting method and quantile treatment effects methods generally require a "rank invariance" assumption—i.e., the treatment preserves the ordering of individual outcomes—which is not plausible in most cases of interest.[9] Although these two methods are estimating different "objects," they are in essence capturing the same idea of heterogeneous outcomes, and data splitting is both commonly used and the simplest method.

However, simple data splitting cannot feasibly take into account multivariate heterogeneous outcomes. If a policy maker is interested in two outcomes—e.g., hourly wage and hours worked—then it is possible some individuals might see their wages increase and at the same time see their hours worked reduced while the opposite can be said about other individuals at a different part of the

---

5.

[6] This discussion precludes the case when deep structural parameters conveying mechanisms behind how the data generating process works are of interest. That is, this discussion is centered on "reduced-form" policy evaluation.

[7] See, for example, Maasoumi and Wang (2019).

[8] See Angrist and Pischke (2009) for discussion.

[9] A way around this issue is to compare entire distributions, see Maasoumi and Wang (2019).

wage distributions. Splitting the sample into individuals with high-wage high-hours worked, high-wage low-hours worked, low-wage high-hours worked, and low-wage low-hours worked makes it difficult to obtain any meaningful policy conclusions. And, interpreting the results would quickly become infeasible as the number of outcomes of interest increases. Alternatively, comparing multi-dimensional distributions, discussed in section 4, leads to interpretable results.

While others have noted the importance of estimating multivariate heterogeneous effects in an interpretable way (e.g. Athey and Imbens, 2015; Wager and Athey, 2018), they focused on conditional means.[10] It might be tempting to, for example, simply condition mean wage on hours worked to capture the heterogeneity of wages. However, that would be taking hours worked as exogenously given whereas the policy might be affecting both wages and hours worked. Therefore, conditional mean methods are not well suited when the policy affects both outcomes of interest. Additionally, these multivariate heterogeneous effects get their heterogeneity from the covariates and only estimate an effect on the outcome's conditional mean, not a conditional quantile (i.e. the case in which outcomes are affected heterogeneously).

Hence, the main advantages of comparing joint distributions of outcomes proposed in this paper are its interpretability, the fact that it allows for outcomes to be affected heterogeneously, and the fact that it does not rely on exogeneity of any of the outcomes of interest.

1.3 The Remainder of the Paper

Section 2 reviews the literature on the predictions and empirical effects of minimum wage policy on low-wage workers as well as its distributional effects. Section 3 discussed the methodology of this paper. Section 4 discusses how to compare multidimensional distributions. Section 5 discusses the data used in this paper. Section 6 presents the results. Section 7 concludes.

2 Literature Review

2.1 Predictions of Minimum Wage Effects on Employment

Stigler (1946) states "...[minimum wage] reduces the earnings of those substantially below the minimum. These are undoubtedly the main allocational effects of a minimum wage in a competitive industry." Indeed, the standard model of a competitive market predicts increased minimum wage will cause an increase in unemployment. However, frictions in the labor market, a small elasticity of demand for labor, or a lack of binding minimum wage can drastically change this prediction.

If there is a large presence of monopsonies, increased minimum wage does not force firm profits to fall below the marginal cost of production (Robinson,

[10] Carlier et al. (2016) use optimal transport theory to develop vector quantile regression, however their results are difficult to interpret since the results are conditional on covariates.

1933; Stigler, 1946; Bhaskar and To, 1999; Manning, 2003; Azar et al., 2019). Hence, there could be little or positive effect on employment.

Flinn (2006) develops a Nash bargaining model in which increased minimum wage intensifies job search and improves employer-employee match quality. In turn, this increases productivity and offsets any negative employment effects.

Another explanation offered is the concept of an "efficiency wage." An efficiency wage is a wage offered by employers that is higher than the market-clearing wage in order to reduce costs associated with turnover (Shapiro and Stiglitz, 1984; Rebitzer and Taylor, 1995). Therefore, firms may be willing to pay higher wages to insure a consistent workforce.

Dessing (2002) suggests workers are "backward-bending" and take jobs below their real reservation wage and productivity level in order to earn some minimal income to feed their families.

Alternatively, unions could affect the labor market. With unions present, firms cannot terminate their employees at will and this generates a cascade effect over the entire wage structure (Lee, 1999; Autor et al., 2016; and Kearney and Harris, 2014).

With many possible alterations to the standard competitive market framework, there is no clear theoretical prediction of the effects of increased minimum wages on employment. Furthermore, some argue market frictions may be negligible and the magnitude of the theorized effects of market frictions has been disputed.

2.2 Empirical Results on Low-wage Workers

Traditional empirical work using observational data found an increase in minimum wage led to decreases in employment.[11] However, following the pathbreaking work of Card and Krueger (1994), the empirical consensus has somewhat shifted to supporting the view that an increase in minimum wage does not increase unemployment — at least in the United States, where minimum wages increases have remained modest. Card and Krueger used a "natural experiment" to compute the change in employment due to minimum wage and compare it to a counterfactual "control" state change in employment. Subsequent minimum wage papers generally used two-way fixed effect (e.g. Neumark and Wascher, 2008).

However, Card and Krueger's work was not without criticism and concerns. Using phone survey data of a sample similar to Card and Krueger's, Neumark and Wascher (2000) results led to opposite conclusions. Meer and West (2016) argue that minimum wage impacts happen over time and while immediate relative employment levels might remain stable, the growth rate of job openings likely decreased. Others have shown that the elasticity of employment decreases in the long-run (Sorkin, 2015; Aaronson et al., 2018). Jardim et al. (2017) argue the relevant market as well as the reduction in hours worked was not con-

[11] See Fernández-Villaverde, 2018 for discussion.

sidered.[12] While most studies—including Card and Krueger—use a proxy for low-wage industries such as teenagers or restaurant workers, they use Seattle data to identify low-wage industries and examine the average treatment effect of a minimum wage increase on both hourly wage and hours worked. They find the increased minimum wage reduced a low-wage worker's monthly earnings by an average of $74 per month. In a follow up paper, Jardim et al. (2018) find the minimum wage increase likely had a more modest or even negligible effect on hours worked citing the possibility that minimum wage workers took up additional outside work or new workers entered the workforce.

Additionally, Card and Krueger (1994) and similar papers have been the subject of methodological concerns. These papers assume "parallel trends" between treatment and control states, which has been criticized especially since the adoption of minimum wage laws appears to be clustered by geographical region (Allegretto et al., 2018). However, there is a large literature that attempts to address these concerns (see Neumark, 2018).

While the empirical effects of minimum wage are disputed, most empirical work does not draw a clear link to a theoretical prediction. If there is indeed no change in employment due to small minimum wage increases, then what theoretical market friction is causing it? Since different theoretical market frictions should have different effects at different parts of the wage distribution, considering distribution effects—discussed in the next section—is critical to understanding the effects of minimum wage.

2.3 Distribution Effects of Minimum Wage

Undoubtedly, minimum wage has heterogeneous effects on the distribution of wages. While some might benefit from increased minimum wage, others might see little benefit or even be hurt by it. The relevant policy question is whether increased minimum wage increased or reduced some welfare measure such as poverty or inequality.[13]

Following the work of Lee (1999), traditional analysis of distribution and inequality changes were concerned with "spillovers". Do workers earning below the minimum wage "spillover" into other parts of the wage distribution, earn the new minimum wage, become disemployed, or some combination of the previous possibilities?

Extending DiNardo et al. (1996), which did not account for spillovers, Lee (1999) compared the change in the ratio of 50th to 10th percentile of wage, calculated the reduction in real minimum wage, and concluded that minimum wage increases substantially increased inequality. Autor et al. (2016) considered a longer period of time and included state and time fixed effects and found similar but smaller effects of minimum wage on reducing inequality.

[12] Another notable paper that examines how minimum wage affects the average hours worked is Belman et al. (2015).

[13] Kearney and Harris (2014), MaCurdy (2015), and Harasztosi and Lindner (2019) consider the effectiveness of minimum wage relative to other antipoverty programs and the extent to which firm or consumers are paying for the minimum wage changes.

An additional concern to interpreting changes in the wage distribution due to increased minimum wage is the possibility of high-wage workers being substituted for low-wage workers. Cengiz et al. (2019) consider the bottom of the wages lost right below the new minimum wage before a minimum wage increase is implemented and found the new wages created right above the minimum wage after the policy is implemented was equivalent. This "bunching effect" explains the lack of job loss is not due to labor-labor substitutions.

The previous literature on distributional effects of minimum wages does not account for the possibility of firms substituting hours worked by low-wage workers with those of high-wage workers and does not take into account the labor-leisure decisions of the workers whereas this paper does.

3 Model and Estimation

3.1 Counterfactual Analysis Setting

Suppose we are interested in some outcome Y which has relevant characteristics X. In the spirit of Haavelmo (1944), assume Y and X are random variables with supports $\mathcal{Y} \subseteq \mathbb{R}$ and $\mathcal{X} \subseteq \mathbb{R}^{d_x}$, respectively, and have measurable density functions. Observations are therefore realizations that come from the joint probability density function of Y and X, $f_{Y,X}(y,x)$, and we are interested in the distribution of Y.

By the law of iterated probability

$$F_Y(y) = \int_{\mathcal{X}} F_{Y|X}(y|x) dF_X(x), \tag{1}$$

where $F_Y(y)$ is the distribution of Y, $F_X(x)$ is the distribution of X, and $F_{Y|X}(y|x)$ the conditional distribution of Y given X. Now, suppose there are two groups 0 and 1 (e.g. 0 is the control group and 1 is the treatment group, or 0 is one time period and 1 is another time period), then the outcome and relevant characteristics might be different for each group—i.e. the outcome is Y_t and the relevant characteristics is X_t with $t \in \{0,1\}$. However, observations would only be observed from $f_{Y_1,X_1}(x_1,y_1)$ and $f_{Y_0,X_0}(x_0,y_0)$. Equation (1) can be rewritten as

$$F_{Y\langle t,v\rangle}(y) = \int_{\mathcal{X}_v} F_{Y_t|X_t}(y|x) dF_{X_v}(x), \tag{2}$$

$t, v \in \{0,1\}$ and if $t \neq v$ then $F_{Y\langle t,v\rangle}$ is the *counterfactual distribution*—the distribution of the random variable Y_t had it come from the joint distribution $f_{Y_t,X_v}(y_t,x_v)$, for which observations are never observed.

To further demonstrate the usefulness of the counterfactual distributions, consider the case in which X is partitioned into two random variables, $X = (X_a, X_b)$.[14] Then by the law of iterated probability

$$F_Y(y) = \int_{\mathcal{X}_b} \int_{\mathcal{X}_a} F_{Y|X_a,X_b}(y|x_a,x_b) dF_{X_a|X_b}(x_a|x_b) dF_{X_b}(x_b). \tag{3}$$

[14]Note that X can easily be partitioned into as many dimensions as d_x.

For two groups 0 and 1, the counterfactual distribution is defined by

$$F_{Y\langle t|s,v\rangle}(y) = \int_{\mathcal{X}_{b,s}} \int_{\mathcal{X}_{a,v}} F_{Y|X_{a,t},X_{b,t}}(y|x_a, x_b) dF_{X_{a,s}|X_{b,s}}(x_a|x_b) dF_{X_{b,v}}(x_b), \tag{4}$$

$t, s, v \in \{0, 1\}$. The observed difference in distributions of the outcome of interest can be decomposed as follows:

$$F_{Y_{\langle 1|1,1\rangle}} - F_{Y_{\langle 0|0,0\rangle}} = \underbrace{F_{Y_{\langle 1|1,1\rangle}} - F_{Y_{\langle 1|0,1\rangle}}}_{\text{(i)}} + \underbrace{F_{Y_{\langle 1|0,1\rangle}} - F_{Y_{\langle 1|0,0\rangle}}}_{\text{(ii)}} + \underbrace{F_{Y_{\langle 1|0,0\rangle}} - F_{Y_{\langle 0|0,0\rangle}}}_{\text{(iii)}},$$

(i) is the effect of the change from $X_{a,0}$ to $X_{a,1}$ on the distribution of Y_1, (ii) is the effect of the change from $X_{b,0}$ to $X_{b,1}$ on the distribution of Y_1, and (iii) is the residual effect on the distribution of Y_1.[15][16]

There are several proposed methods for estimated the counterfactual distribution. The main practical concern is estimating the conditional distributions. DiNardo et al. (1996) propose an inverse propensity reweighting method. Alternatively, S. Firpo et al. (2009, 2018) use re-centered influence function (RIF) regressions. Chernozhukov et al. (2013) use quantile and distribution regressions and show valid inference can be done with an exchangeable bootstrap.[17]

Suppose now that there are two outcomes of interest, Y^1 and Y^2.[18] We are interested in the joint distribution of outcomes, $F_{Y^1,Y^2}(y_1, y_2)$. By Sklar's Theorem, if marginal distributions F_{Y^1} and F_{Y^2} are continuous, there exists a unique copula C such that

$$F_{Y^1,Y^2}(y_1, y_2) = C(F_{Y^1}(y_1), F_{Y^2}(y_2)). \tag{5}$$

For the remainder of this section, this paper will provide conditions for identification and use the method of Chernozhukov et al. (2013) to estimate the marginal counterfactual distributions of hourly wage and hours worked and then use an empirical copula to obtain an estimate of the joint counterfactual distribution of hourly wage and hours worked.

[15]Note an alternative decomposition is, for example, $F_{Y_{\langle 1|1,1\rangle}} - F_{Y_{\langle 0|0,0\rangle}} = (F_{Y_{\langle 1|1,1\rangle}} - F_{Y_{\langle 0|1,0\rangle}}) + (F_{Y_{\langle 0|1,0\rangle}} - F_{Y_{\langle 0|1,1\rangle}}) + (F_{Y_{\langle 0|1,1\rangle}} - F_{Y_{\langle 0|0,0\rangle}})$ or some decomposition of $F_{Y_{\langle 0|0,0\rangle}} - F_{Y_{\langle 1|1,1\rangle}}$. Therefore, the "sequential" ordering of the decomposition might have important implications and is a major drawback of this kind of decomposition analysis. Accordingly, it is important to check the reverse ordering for robustness (e.g. the effect of a change from $X_{a,0}$ to $X_{a,1}$ on the distribution of Y_1 should be adhere to the interpretation of a change from $X_{a,1}$ to $X_{a,0}$ on the distribution of Y_0).

[16]More generally, for the functional ϕ (e.g. Lorenz curve, Gini coefficient, quantile ranges, and more trivially the mean and variance), the observed differences, $\phi(F_{Y_{\langle 1|1,1\rangle}}) - \phi(F_{Y_{\langle 0|0,0\rangle}})$, can be decomposed as similarly.

[17]See Fortin et al. (2011) for more details on the decomposition method and methods for estimating the counterfactual distribution.

[18]Of course, this can easily be generalized to a case with there are more than two outcomes of interest.

3.2 Identification

Let $(Y_j^{1*}, Y_j^{2*} : j \in \mathcal{J})$ denote a vector of potential outcome variables for various values of a policy, $j \in \mathcal{J}$, and let X^1 and X^2 be vectors of covariates for Y_j^{1*} and Y_j^{2*}, respectively.[19] Let J be a random variable that denotes the realized policy with $Y^1 := Y_J^{1*}$ and $Y^2 := Y_J^{2*}$ the realized outcome. Let $F_{Y_j^{1*}, Y_j^{2*}|J}(y_1, y_2|k)$ denote the joint distribution of the potential outcome Y_j^{1*} and Y_j^{2*} in the population where $J = k \in \mathcal{J}$.

The causal effect of exogenously changing the policy from ℓ to j on the distribution of potential outcomes in the population with realized policy $J = k$ is

$$F_{Y_j^{1*}, Y_j^{2*}|J}(y_1, y_2|k) - F_{Y_\ell^{1*}, Y_\ell^{2*}|J}(y_1, y_2|k).$$

Assumption 1. Let $\mathcal{X}_k^1 \subseteq \mathcal{X}_j^1$, $\mathcal{X}_k^2 \subseteq \mathcal{X}_j^2$ for all $(j,k) \in \mathcal{J}\mathcal{K}$.

Assumption 2. Let the latent variables

$$(Y_j^{1*}, Y_j^{2*} : j \in \mathcal{J}) \perp\!\!\!\perp J|X^1, X^2 \quad \text{a.s.},$$

where $\perp\!\!\!\perp$ denotes independence.

Assumption 3. Let $F_{Y^1\langle j|k \rangle}(\cdot)$ and $F_{Y^2\langle j|k \rangle}(\cdot)$ with $j, k \in \mathcal{J}$ be continuous.

Theorem 3.2. Under Assumptions 1-3,

$$F_{Y^1, Y^2\langle j|k \rangle}(\cdot) = F_{Y_j^{1*}, Y_j^{2*}|J}(\cdot|k), \quad j, k \in \mathcal{J}.$$

Proof. See Appendix B.1.

Theorem 3.2 can be generalized to state that as long as the marginal distributions of the latent outcome variables are continuous and identified, the joint distribution of latent outcomes is identified.

3.3 Counterfactual Distribution

Let 0 denote a year with lower minimum wage and 1 denote a year with higher minimum wage (e.g. 0 denotes 1989 and 1 denotes 1992) such that $\mathbf{Y}_t = \left(Y_t^1, Y_t^2\right)$ denotes hourly wages and hours worked at year $t \in \{0, 1\}$, respectively. Let $\mathbf{X}_v$ denotes the job market-relevant characteristics affecting hourly wages and hours worked at year, $v \in \{0, 1\}$. For ease of notation, superscripts on the covariates denoting hourly wage or hours worked are dropped. Furthermore, let $\mathbf{X}_v$ be composed of a minimum wage variable m_v and all other characteristics $\mathbf{c}_v$, $\mathbf{X}_v = (m_v, \mathbf{c}_v)$, where

$$m_v = \begin{cases} 1 & \text{earning at or below minimum wage} \\ 0 & \text{otherwise} \end{cases}.$$

[19] Following the potential outcomes literature, general equilibrium effects are excluded in the definition of potential outcomes. However, since the *ceteris paribus* effects are on the joint distribution of outcomes, general equilibrium effects of the outcomes on each other are not excluded.

Let F_η, denote the distribution of the random variable η. The unconditional marginal distribution of Y^i, $i \in \{1,2\}$, can be computed by integrating over the conditional distributions as follows:

$$F_{Y^i\langle t|s,v\rangle}(y) := \int_{\mathcal{C}_v}\int_{\mathcal{M}_s} F_{Y^i_t|m_t,\mathbf{c}_t}(y|m,\mathbf{c})dF_{m_s|\mathbf{c}_s}(m|\mathbf{c})dF_{\mathbf{c}_v}(\mathbf{c}), \tag{6}$$

where $\mathcal{M}_s \subseteq \mathbb{R}$, $\mathcal{C}_v \subseteq \mathbb{R}^{d_c}$ denote the supports of m_s and $\mathbf{c}_v$, respectively. When $t = v = s$, (6) becomes the unconditional distribution $F_{Y^i\langle t|t,t\rangle} = F_{Y^i_t}$—the observed distribution of Y^i and time t—by the law of iterated probabilities. When t, s, and v are not equal, $F_{Y^i\langle t|s,v\rangle}$ is a counterfactual distribution. For example, When $t = v = 1, s = 0$, the $F_{Y^i\langle 1|0,1\rangle}$ is the distribution of outcomes that would prevail for time 1 had that period had the composition of minimum wage of time 0.

The joint distributions can be obtained using Sklar's Theorem: there exists a copula C such that

$$F_{\mathbf{Y}\langle t|s,v\rangle}(y_1,y_2) = C(F_{Y^1\langle t|s,v\rangle}(y_1), F_{Y^2\langle t|s,v\rangle}(y_2)). \tag{7}$$

If $F_{Y^1\langle t|s,v\rangle}$ and $F_{Y^2\langle t|s,v\rangle}$ are continuous, C is unique.

3.4 Estimation of the Counterfactual Distribution

Chernozhukov et al., 2013 propose an algorithm for estimation $F_{Y^i\langle 1|0,1\rangle}$:

1. Estimate $F_{\mathbf{c}_1}(\mathbf{c})$ by empirical cdf to obtain $\hat{F}_{\mathbf{c}_1}(\mathbf{c})$
2. Estimate $F_{m_0|c_0}(m|\mathbf{c})$ by logistic regression to obtain $\hat{F}_{m_0|c_0}(m|\mathbf{c})$
3. Estimate $F_{Y^i_1|m_1,\mathbf{c}_1}(y|m,\mathbf{c})$ by distribution regression (discussed below) to obtain $\hat{F}_{Y^i_1|m_1,\mathbf{c}_1}(y|m,\mathbf{c})$
4. Obtain $\hat{F}_{Y^i\langle 1|0,1\rangle}(y) = \int_{\mathcal{C}_1}\int_{\mathcal{M}_0} \hat{F}_{Y^i_t|m_1,\mathbf{c}_1}(y|m,\mathbf{c})d\hat{F}_{m_0|\mathbf{c}_0}(m|\mathbf{c})d\hat{F}_{\mathbf{c}_1}(\mathbf{c})$

While it is possible to estimate $\hat{F}_{Y_1|m_1,\mathbf{c}_1}$ using quantile regression, Chernozhukov et al. (2013) show quantile regression does not preform well when there are large point masses in the distribution being estimated, such as the wage and hours worked distributions.

Therefore, Chernozhukov et al. (2013) propose the *distribution regression*, a modification of the method proposed by Foresi and Peracchi (1995). For n observations, $F_{Y_1|m_1,\mathbf{c}_1}(y|m,\mathbf{c})$ is estimated by

$$\hat{F}_{Y^i_1|\mathbf{X}_1}(y|\mathbf{X}) = \Lambda(P(\mathbf{X})\hat{\beta}(y)) \quad y \in \mathcal{Y}_i,$$

where $P(\cdot)$ is a vector of transformations of $\mathbf{X}$ (e.g. polynomials or basis splines), Λ is the link function, and $\hat{\beta}(y)$ is estimated using maximum likelihood

$$\hat{\beta}(y) = \arg\max_b \sum_{j=1}^n \{\mathbb{1}\{Y^i_j \le y\}\ln(P(\mathbf{X}_j)'b) + \mathbb{1}\{Y^i_j \ge y\}\ln((1-P(\mathbf{X}_j))'b)\}.$$

This paper uses a logit link function (a probit link function is also used in section 6.3).

Once the marginals are estimated, the joint distribution is obtained through

$$\hat{F}_{\mathbf{Y}\langle t|s,v\rangle}(y_1, y_2) = \hat{C}(\hat{F}_{Y^1\langle t|s,v\rangle}(y_1), \hat{F}_{Y^2\langle t|s,v\rangle}(y_2)), \tag{8}$$

where $\hat{C}$ is a consistent estimate of the copula.

This paper uses an "empirical copula" developed by Deheuvels (1979), which is a nonparametric method. However if there are more than two or three outcomes of interest, to avoid the *Curse of Dimensionality*, a parametric copula should be used.

Empirical Copula Method: For observations (Y_i^1, Y_i^2), $i = 1, ..., n$ we have copula observations $(U_i^1, U_i^2) = (F_{Y^1\langle t|s,v\rangle}(Y_i^1), F_{Y^2\langle t|s,v\rangle}(Y_i^2))$. Therefore $(\hat{U}_i^1, \hat{U}_i^2) = (\hat{F}_{Y^1\langle t|s,v\rangle}(Y_i^1), \hat{F}_{Y^2\langle t|s,v\rangle}(Y_i^2))$ and

$$\hat{C}(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\hat{U}_i^1 \leq u_1, \hat{U}_i^2 \leq u_2\}$$

is the empirical copula.

A simple simulation of this method can be found in Appendix A.1.

3.5 Inference

Let $\hat{\Delta}_{y_1,y_2} := \hat{F}_{\mathbf{Y}\langle 1|0,1\rangle}(y_1, y_2) - \hat{F}_{\mathbf{Y}\langle 1|1,1\rangle}(y_1, y_2)$. The algorithm for uniform bootstrap confidence bands is:

1. Obtain bootstrap draws $\left(\hat{F}^{*(j)}_{\mathbf{Y}\langle 1|0,1\rangle}(y_1, y_2) - \hat{F}^{*(j)}_{\mathbf{Y}\langle 1|1,1\rangle}(y_1, y_2)\right)_{y_1\in T_1, y_2\in T_2}$ for $j = 1, ..., B$
2. For each $y_1 \in T_1$, $y_2 \in T_2$ compute bootstrap variance $\hat{s}^2(y_1, y_2) = B^{-1}\sum_{j=1}^{B}\left(\left(\hat{F}^{*(j)}_{\mathbf{Y}\langle 1|0,1\rangle}(y_1, y_2) - \hat{F}^{*(j)}_{\mathbf{Y}\langle 1|1,1\rangle}(y_1, y_2)\right) - \hat{\Delta}_{y_1,y_2}\right)^2$
3. Compute the critical value $c(1-\alpha) = (1-\alpha)$-quantile of
$$\left\{\max_{y_1\in T_1, y_2\in T_2}\left|\left(\hat{F}^{*(j)}_{\mathbf{Y}\langle 1|0,1\rangle}(y_1, y_2) - \hat{F}^{*(j)}_{\mathbf{Y}\langle 1|1,1\rangle}(y_1, y_2)\right) - \hat{\Delta}_{y_1,y_2}\right| \Big/ \hat{s}(y_1, y_2)\right\}_{j=1}^{B}$$
4. Construct confidence band for $\left(\hat{\Delta}_{y_1,y_2}\right)_{y_1\in T_1, y_2\in T_2}$ as $[L(y_1, y_2), U(y_1, y_2)] = \left[\hat{\Delta}_{y_1,y_2} \pm c(1-\alpha)\hat{s}(y_1, y_2)\right]$

Chernozhukov et al. (2013) show the validity of the Kolmogorov-Smirnov confidence bands obtained through the algorithm above; see Appendix AS of Chernozhukov et al. (2013) for details.

4 Comparing Multidimensional Distributions

4.1 Measuring Inequality and the Social Welfare Function

Inequality comparisons between two distributions are often controversial because they depend on the *a priori* preferences of the policy maker. Therefore, comparing inequality requires some subjectivity and, accordingly, the goal of much of the work on inequality comparisons has been to create methods of comparing inequality that accommodate a large class of social welfare functions so that the result is widely accepted.

Define a social welfare function (SWF) for a population of n individuals as

$$\mathrm{SWF} = W(u_1(\mathbf{x}_1), ..., u_n(\mathbf{x}_n)), \tag{9}$$

where $u_i(\mathbf{x}_i)$ is the utility of individual i with bundle $\mathbf{x}_i = (x_i^1, x_i^2, ..., x_i^k)$ of k attributes, with $x_i^j \in \mathbb{R}_+$ for $j = 1, ..., k$.

Broadly speaking, there are two ways to compare inequality making assumptions about equation (9). First, using dominance criteria that place restrictions on the utility functions of individuals in the population. Second, using an index that satisfies certain properties thought to define inequality.

While dominance testing is generally less controversial because it makes fewer and less restrictive assumptions, it does not produce a complete ordering of distributions. In that way, showing first-order stochastic dominance of one distribution over another is considered the gold standard in inequality comparisons, but cannot always be achieved. However, an additional drawback with dominance testing methods is that they only produce an ordinal (and not cardinal) ranking of distributions. That is, they can determine which distribution is preferred, but not by how much.

While inequality indexes provide a complete ordering of distributions and are cardinal, many inequality indexes have been criticized since they measure inequality as the relative differences between individuals' allotments (i.e. they are relative inequality measures). For example, if a policy were to give all poor individuals a ten percent increase in income and at the same time give wealthy individuals a twenty percent increase in income, most inequality indexes would assign the latter distribution with a higher level inequality even though every individual is made better off. Furthermore, many do not find the goal of reducing inequality compelling and question the validity of the "ideal" properties of inequality indexes.

For some of the reasons listed above, many find the goal of reducing poverty, an absolute measure, instead of inequality more reasonable. Poverty can be defined as the level at which individuals fall below or near a poverty threshold.

The next three subsections discuss stochastic dominance methods, inequality index methods, and poverty index methods.

4.1.1 Stochastic Dominance

Under the assumptions that individuals share the same utility function $u(\cdot)$ and that the SWF is *additively separable*, also known as a *utilitarian* SWF, equation (9) becomes

$$\text{SWF} = \sum_{i=1}^{n} u(\mathbf{x}_i). \tag{10}$$

With one attribute, i.e. $k = 1$, a distribution $x_1, ..., x_n$ "(strictly) dominates" the distribution $x'_1, ..., x'_n$ if $\sum_{i=1}^{n} u(x_i) > \sum_{i=1}^{n} u(x'_i)$. That is, one distribution (strictly) dominates another if its SWF defined in equation (10) is (strictly) larger than the other's.

It becomes easier to think of the distributions $x_1, ..., x_n$ and $x'_1, ..., x'_n$ as realizations of the random variables x and x', respectively.[20] With the class of utility functions $\mathcal{U}_1 \equiv \{u : u' \geq 0\}$, x first-order (strictly) stochastically dominates x' if $F_x(w) < F_{x'}(w)$ for all $w \in \mathbb{R}$.[21] Intuitively, as long as possessing more of the attribute does not cause disutility, a utilitarian SWF is higher for the distribution whose CDF is lower at all points, a testable condition.

Since first-order stochastic dominance cannot always be achieved, placing additional assumptions on the SWF can expand what pairs of distributions can be ranked. For the class of utility functions $\mathcal{U}_2 \equiv \{u : u' \geq 0, u'' \leq 0\}$, x second-order (strictly) stochastically dominates x' if $\int_{-\infty}^{x} F_x(w)dw < \int_{-\infty}^{x'} F_{x'}(w)dw$ for all $w \in \mathbb{R}$. Intuitively, for a "mean preserving spread," one distribution is preferred to another with the additional assumption that the attribute has decreasing marginal utility. So, taking one unit of the attribute away from someone who has more and giving it to someone who has less is now a preferred distribution as long as it does not change the mean of the distribution.

For the class of utility functions $\mathcal{U}_t \equiv \{u : u' \geq 0, u'' \leq 0, ..., (-1)^{t+1}u''^{\cdots\prime} \geq 0\}$, the distribution x is said to (strictly) t-order dominate x' if $\int_{-\infty}^{x} F_x'^{\cdots\prime}(w)dw < \int_{-\infty}^{x'} F_{x'}'^{\cdots\prime}(w)dw$ for all $w \in \mathbb{R}$. However, the normative implications of higher than second- or third-order dominance is generally not considered important enough to be a reasonable ranking criterion.

With more than one attribute (i.e. $k \geq 2$), additional restrictions need to be placed on the cross-derivatives of the individual's utility function. For example, with two attributes, the random vector $\mathbf{x}$ first-order (strictly) stochastically dominates $\mathbf{x}'$ if $\sum_{i=1}^{n} u(\mathbf{x}_i) > \sum_{i=1}^{n} u(\mathbf{x}'_i)$ with $u \in \mathcal{U}^- \equiv \{u_1, u_2 \geq 0, u_{12} \leq 0\}$ with a similar testable condition $F_{\mathbf{x}}(w_1, w_2) < F_{\mathbf{x}'}(w_1, w_2)$ for all $w_1, w_2 \in \mathbb{R}$.[22]

Stochastic dominance methods were first used in inequality comparisons by Atkinson (1970). Atkinson and Bourguignon (1982) extended univariate dominance to multidimensional distributions. McFadden (1989) provided early statistical tests for stochastic dominance. Linton et al. (2005) developed a test

[20] This also allows us to compare distributions that do not have the same number of individuals.

[21] See Whang (2019) for details.

[22] See Atkinson and Bourguignon (1982) for higher order dominance criteria with multiple attributes.

that uses a subsampling method and allows for general dependence amongst prospects to be ranked. Donald and Hsu (2016) use a re-centering method to provide a more powerful test than Linton et al. (2005). For an excellent survey on stochastic dominance tests and sample code, see Whang (2019).

4.1.2 Multidimensional Inequality

There are two approaches to constructing inequality indices — an *axiomatic* approach and an *information theoretic* approach. The axiomatic approach starts by defining a set of ranking rules and "desirable properties" the index should have *a priori*, and then constructs an index that satisfies those rules and properties. On the other hand, the information theoretic approach defines an ideal distribution, and then uses information theoretic arguments to quantify how close a distribution is to the ideal.

The axiomatic approach has been criticised because the cardinal values of such indices seem arbitrary as opposed the the entropy approach, which can be interpreted as "closeness" in the information sense. The information theoretic approach has been criticised because it does not make explicitly clear which properties and ranking rules are satisfied. However, common information theoretic approach indices have been shown to satisfy many of the commonly assumed desirable properties of the axiomatic approach. Additionally, it can be argued that the assumptions that go into defining entropy are more primitive and can more reasonably be considered *a priori*. In practice, it is best to use both approaches for robustness.

Either approach admits a large class of inequality indices. Inequality measures over multiple attributes inevitably make decisions on: (1) the extent to which each attribute contributes to an individual's well-being; (2) the degree of the policy maker's inequality aversion; (3) the degree of substitutability between attributes. These decisions show up as parameters in the inequality indices.

The extent to which each attribute contributes to an individual's well-being is quantified in one of three ways: (i) the *agnostic* approach (weighting all attributes equally); (ii) the *normative* approach (setting weighting according to some normative criteria); (iii) the *data-driven* approach (e.g. principal component analysis). Substitutability between attributes can been quantified similarly. Inequality aversion is usually quantified with the *normative* approach. In practice, a wide range (grid) of weights and parameter values are used to ensure the results are robust these subjective decisions.

Let $\mathcal{M}(n)$ be a $n \times k$ matrix whose elements are non-negative. A *multidimensional distribution* is a $n \times k$ matrix $X = (x_i^j) \in \mathcal{M}(n)$ with $k \geq 2$. The *multidimensional inequality index* is defined as the function $I^n(X) : \mathcal{M}(n) \to \mathbb{R}$.

One possible multidimensional inequality index is proposed by Maasoumi (1986). The index involves a "two-step" approach. The first step aggregates the attributes for each individual, obtaining a vector $S_i = f(\mathbf{x}_i)$ such that each element of S_i summarizes individuals i's marginal distribution of attributes ($f(\cdot)$ can be thought of as a utility function), and the second step applies a measure of inequality that is the same as the univariate case. The procedures for obtaining values in either step is made through "information theoretic" arguments.[23] Essentially, to construct a scalar that summarizes the attributes an individual has, Maasoumi (1986) minimizes the generalized cross-entropy measure

$$
\begin{aligned}
D_\beta(S, X; \alpha) &= \sum_{j=1}^{k} \alpha_j \left\{ \sum_{i=1}^{n} S_i \left[\left(\frac{S_i}{x_{ij}} \right)^{\beta} - 1 \right] \Big/ \beta(1-\beta) \right\}, \\
&= \sum_j \alpha_j \left\{ \sum_i S_i \log(S_i / x_{ij}) \right\} \quad \beta = 0, \\
&= \sum_j \alpha_j \left\{ \sum_i x_{ij} \log(x_{ij} / S_i) \right\} \quad \beta = 1.
\end{aligned}
$$

$S = (S_1, ..., S_n)$ is the "optimal" aggregation function which minimized $D_\beta(\cdot)$ (i.e. S_i is the "closest" aggregation of attributes for individual i). The parameters α_j is the weight of attribute j, and β is the degree of substitutability between attributes. The S_i that minimizes $D_\beta(\cdot)$ subject to $\sum_{i=1}^{n} S_i = 1$ is

$$
\begin{aligned}
S_i &\propto \left[\sum_{j=1}^{k} \delta_j x_{ij}^{-\beta} \right]^{-\frac{1}{\beta}} \qquad \beta \neq 0, \\
&\propto \prod_{j=1}^{K} x_{ij}^{\delta_j} \quad \beta = 0,
\end{aligned}
$$

where $\delta_j = \alpha_j / \sum_j \alpha_j$. For the second step, Maasoumi (1986) uses the univariate generalized entropy inequality measure[24] over S.

On the other hand, possible multidimensional inequality indices formed with *a priori* "desirable properties" are proposed in Tsui (1995, 1999) and Bourguignon (1999), as noted by Lugo (2005).[25]

4.1.3 Multidimensional Poverty

Foster et al. (1984) proposed a commonly used univariate class of poverty indices that is "subgroup decomposable," meaning the overall poverty in the population can be broken down in poverty in each subgroup. Duclos et al. (2006, 2007) extend the Foster et al. (1984) class of indices to the multivariate case and propose criteria for "poverty dominance," i.e. when one distribution have less poverty than another regardless of the poverty thresholds. In this way, their method is robust to the "union" (an individual is considered poor if she is

[23] See Maasoumi (1993) for more details on the axiomatic construction of the generalized entropy measure.

[24] See Shorrocks (1980).

[25] As opposed to Maasoumi (1986), these multivariate axiomatic approach indices are "one-step" methods. That is, both the aggregation of attributes and inequality measure construction are done in one step.

below the poverty threshold in any dimension), "intersection" (an individual is considered poor if she is below the poverty threshold in all dimensions) or some "intermediate" (an intermediate condition for being poor) approaches to constructing multidimensional poverty indices.

For two dimensions, define z_1 and z_2 as the univarite poverty lines for their respective dimensions. The Duclos et al. (2006, 2007) multivariate version of the Foster et al. (1984) poverty measure is

$$P^{a_1 a_2}(z_1, z_2) = \int_0^{z_2} \int_0^{z_1} (z_1 - x_1)^{a_1} (z_2 - x_2)^{a_2} dF(x_1, x_2),$$

where $\alpha_1, \alpha_1 \geq 0$ capture the aversion to inequality in the x_1 and x_2 dimensions, respectively. Poverty gaps in each dimension are captured by $(z_1 - x_1)$ and $(z_2 - x_2)$. $P^{0,0}(z_1, z_2)$ is the intersection headcount poverty index.

Maasoumi and Lugo (2008) offer a information theoretic approach for constructing poverty indices. See Alkire and Foster (2011) for discussion on other approaches for obtaining intermediate poverty indices, including the fuzzy set approach and the latent variables approach.

5 Data

5.1 Data Cleaning

This paper uses the Current Population Survey Merged Outgoing Rotation Group (CPS MORG) for the years 1979 to 2019 and was obtained though the NBER website.[26] CPS MORG is different than the March CPS[27] because survey participants in the CPS MORG extracts were asked about their hourly wage and hours worked from that week (as opposed to imputed weekly hours worked and hourly wage from yearly earnings and usual hours worked). These "point-in-time" measures are arguably more reliable because participants are more likely to accurately remember their hourly wage and hours worked that week.[28]

The data is cleaned using the specifications of Autor et al. (2016) and state minimum wage data was obtained through Vaghul and Zipperer (2016).[29] Wages are in 2019 dollars and Consumer Price Indexes were obtained form Frederal Reserve Bank of Minneapolis.[30] The sample includes individuals ages 18 through 64 and excludes those who are self employed. Top-coded values are multiplied by 1.5 and the top two wage percentiles for each state, year, and sex grouping are "Winsorized" (replaced with the ninety-seventh percentile's value).[31]

[26] Available at https://data.nber.org/morg/annual/.

[27] Commonly referred to as IPUM CPS, since it is maintained by the Minnesota Population center at the University of Minnesota.

[28] See Lemieux (2006).

[29] Available at https://github.com/benzipperer/historicalminwage/releases.

[30] Available at https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913-

[31] See Autor et al. (2016) for more details on the data cleaning.

However, unlike Autor et al. (2016), this paper uses CPS individual weights and does not multiply these weights by hours worked in the previous week. DiNardo et al. (1996) states "These 'hours-weighted' estimates put more weight on the wages of workers who supply many hours to the labor market. This gives a better representation of the dispersion of wages for each and every hour worked in the labor market, regardless of who is supplying this hour". While this approach may have originally deviated from using weekly earning as the dependent variable to account for labor market participation decisions, it makes welfare comparisons more difficult since there is no way of knowing if individuals actually have higher wage and hours worked bundles (incomes). Additionally, it treats hours as exogeously given. The paper's approach accounts for both of these shortcomings.

Following DiNardo et al. (1996) and Chernozhukov et al. (2013), control variables included in this study are union status, marital status, race, an indicator for part time worker, educational and experience dummy variables, occupation dummy variables, industry dummy variables, Standard Metropolitan Statistical Area (SMSA). As is common in the literature, this paper uses number of children under 5 as an exclusion restriction that potentially affects hours worked but is uncorrelated with hourly wage.

5.2 Data Visualization

Figure 2 shows real U.S. minimum wage over time. Throughout the 1980's virtually all states shared the declining real federal minimum wage. Around the turn of the century, many states started adopting a minimum wage that was higher than the federal minimum wage, causing more variation in minimum wage. Figure 3 shows the US states' minimum wages in 1979, 1999, and 2019. Table 1 shows the timeline of nominal federal minimum wage increases.

Figure 4 shows the percent of workers earning at or below the minimum wage. The percent of men earning at or below the minimum wage is lower than that of women. Figure 7 shows the evolution of hourly wage and weekly hours worked by five quantiles ranges for hourly wage. The gap between the top and bottom quatiles range of hourly wages in the pooled sample widens over time but seems to be mainly driven by the gap between the top and bottom quatiles range of hourly wages for women widening. Additionally, Figure 7 shows the "backward bending" of the labor-leisure curve, i.e. low-wage workers work the fewest hours, mid-wage workers work the most hours, and high-wage workers work fewer hours than mid-wage workers. This is true in all samples, but is most prominent for women. Hours worked seems to be more volatile for workers at the bottom of the wage distribution.

6 Results

6.1 Main Results

Tables 2 shows the effect of having 1992's and not 1989's minimum wage on 1992's Distribution of Hours and Wages ($\hat{F}_{\mathbf{Y}_{\langle 1|1,1\rangle}} - \hat{F}_{\mathbf{Y}_{\langle 1|0,1\rangle}}$) for men. Tables 2-7 show similar counterfactual effects for both men and women and in different years. Since these tables are differences in CDFs, a way to interpret the tables is, for example in Table 2, there was a 7% increase in men earning up to \$9.35 per hour and working up to 40 hours per week from 1989 to 1992 due to minimum wage changes. The tables suggest that, for both men and women, workers were better off across the entire joint distribution of wages and hours worked with increased minimum wage.

6.2 Robustness Checks

Since the main identifying assumption to estimate the counterfactual distribution is an unconfoundedness assumption (i.e. assuming the model is correctly specified), the paper includes additional robustness check for unobserved heterogeneity. To do so, redefine the minimum wage variable (MW) as the minimum wage in the region an individual works. Hence, MW varies by state and time, and therefore the years 2003 through 2019 are used because MW has more variation in those years.

Breaking the sample into five quantile ranges of hourly wages, Table 8 shows the results of the fixed effects models

$$\text{hours}_{its} = \alpha_s + \gamma_t + \delta^H X_i^H + \beta^H \text{MW}_{ts} + \epsilon_{its}^H,$$

and

$$\log(\text{wages})_{its} = \alpha_s + \gamma_t + \delta^W X_i^W + \beta^W S\text{MW}_{ts} + \epsilon_{its}^W,$$

where α_s and γ_t are state and year fixed effects (respectively), X_i^H and X_i^W are vectors of control variables for hours worked and wages (respectively), and ϵ_{its}^H and ϵ_{its}^W are error terms. The results from Table 8 are similar to Tables 2-7 in that minimum wages seemed to either increase the hours worked or not be statistically significant at any wage quantile range for both men and women.

Table 9 estimates the model

$$\text{hours}_{its} = \alpha_s + \gamma_t + \delta X_i^H + \beta_1 \text{MW}_{ts} + \beta_2 \text{MW}_{ts}^2 + \epsilon_{its},$$

which allows for the effect of minimum wages to change based on the level of minimum wage. That is, the marginal effect of minimum wage is $\beta_1 + 2\beta_2 MW_{ts}$, and hence the point at which the effect changes from a negative to a positive or from a positive to a negative is $\frac{\beta_1}{2\beta_2}$. Overall, it seems that any minimum wage up to \$15.90 increases the number of hours worked for men, whereas for women, any minimum wage above \$3.98 increases their number of hours worked.

Roth and Sant'Anna (2021) formally state criteria for when partitioning the data into quantile ranges and running fixed effects regressions is valid.[32]

6.3 Selection and Employment Effects

While Cengiz et al. (2019) compellingly showed most minimum wage increases in the United States had little to no employment effects, it is possible that even minor employment effects are driving the results of this paper. Feasibly, minimum wage workers are being terminated from their jobs and only high-wage workers working longer hours are left in the sample. To account for this possibility, we use the method developed in Chernozhukov et al. (2020) to account for the possibility of selection due to employment effects.

Maasoumi and Wang (2019) used the method of Arellano and Bonhomme (2017) to account for selection in the wage distribution for women using quantile regressions, however since this paper is using distribution regression, the Chernozhukov et al. (2020) method is a more natural choice.

Chernozhukov et al. (2020) use local Gaussian representation (LGR) of the joint distribution of the latent outcome and selection variables to account for selection with distribution regressions. The algorithm for estimating counterfactual distribution with selection is

1. Run a probit for the selection equation
$$\hat{\pi} = \arg\max_{c \in \mathbb{R}^{d_\pi}} \frac{1}{n} \sum_{j=1}^{n} \left[D_j \ln \Phi(\mathbf{Z}_j' c) + (1 - D_j) \ln \Phi(-\mathbf{Z}_j' c) \right],$$
where D is the selection variable (i.e. $D_j = 1$ if the outcome variable for observation j is observed), $\mathbf{Z}_j$ are covariates of labor force participation, and $\Phi(\cdot)$ is a standard normal CDF.

2. Run multiple distribution regressions with the sample selection correction to estimate $\theta_y^i = (\beta(y), (y))$ for each $y \in \bar{\mathcal{Y}}_i$ (a grid of y values)
$$\hat{\theta}^i = \arg\max_{(b,d) \in \Theta} \frac{1}{n} \sum_{j=1}^{n} D_j [\mathbb{1}\{Y_j^i \leq y\} \ln \Phi_2(-\mathbf{X}_j' b, \mathbf{Z}_j' \hat{\pi}; -\rho(\mathbf{X}_j' d))$$
$$(1 - \mathbb{1}\{Y_j^i \leq y\}) \ln \Phi_2(\mathbf{X}_j' b, \mathbf{Z}_j' \hat{\pi}; \rho(\mathbf{X}_j' d)],$$
where $\Phi_2(\cdot, \cdot; \rho)$ is the joint CDF of a bivariate normal random varibale with parameter ρ, $\Theta \in \mathbb{R}^{d_\Theta}$ is a compact parameter set, and $\rho(u) = \tanh(u)$.

3. Use the Chernozhukov et al. (2013) algorithm from section 3.4, replacing step 3 with selection correct $\hat{F}_{Y_1^i|m_1,\mathbf{c}_1}(y|m, \mathbf{c})$ (i.e. using the $\hat{\beta}(y)$ estimated in the previous step) to obtain $\hat{F}_{Y^i\langle 1|0,1\rangle}$.[33]

[32] They require a "parallel trends" condition on the CDF of the untreated potential outcomes so that treatment in the partitioned subgroups is effectively randomly assigned.

[33] An estimator for the probability of selection (being in the labor force) is $\widehat{Prob}(D = 1) = \frac{1}{n} \sum_{j=1}^{n} \Phi(z' \hat{\pi})$.

********Insert Selection Results***********

6.4 $15 Federal Minimum Wage

In the United States, public interest in minimum wage policy, and in particular interest in raising the federal minimum wage from $7.25 to $15 per hour, notably increased.[34] *******add 2019 counterfactual effect of $15 minimum wage with the possibility of employment effects for men and women*********

7 Conclusion

Considering multivariate outcomes heterogeneously affected by a policy is important since individuals could be affected positively in one dimension and negatively in another. This has been a common criticism of minimum wage literature when applied to US data—while there seems to be little employment effect due to small minimum wage increases and wages increase across the wage distribution, some workers might see their hours reduced. However, this paper finds that increased minimum wage had positive effects on both hourly wage and hours worked for all individuals and no such substitution is seems to be taking place.

[34]See, for example, https://www.wsj.com/articles/biden-wants-a-15-minimum-wage-heres-what-people-say-it-would-do-to-the-economy-11612348201.

References

Aaronson, D., French, E., Sorkin, I., & To, T. (2018). Industry dynamics and the minimum wage: A putty-clay approach. *International Economic Review*, *59*(1), 51–84.

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, *95*(7-8), 476–487.

Allegretto, S., Dube, A., Reich, M., & Zipperer, B. (2018). Credible research designs for minimum wage studies: A response to neumark, salas, and wascher: Corrigendum. *ILR Review*, *71*(2), 562.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton; Oxford:

Arellano, M., & Bonhomme, S. (2017). Quantile selection models with an application to understanding changes in wage inequality. *Econometrica*, *85*(1), 1–28.

Athey, S., & Imbens, G. (2015). Recursive partitioning for heterogeneous causal effects.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*(3), 244–263.

Atkinson, A. B., & Bourguignon, F. (1982). The comparison of multi-dimensioned distributions of economic status. *Review of Economic Studies*, *49*(2), 183–201.

Autor, D. H., Manning, A., & Smith, C. L. (2016). The contribution of the minimum wage to us wage inequality over three decades: A reassessment. *American Economic Journal: Applied Economics*, *8*(1), 58–99.

Azar, J., Huet-Vaughn, E., Marinescu, I., Taska, B., & von Wachter, T. (2019). Minimum wage employment effects and labor market concentration. *National Bureau of Economic Research Working Paper 26101.*

Belman, D., Wolfson, P., & Nawakitphaitoon, K. (2015). Who is affected by the minimum wage?. *Industrial Relations*, *54*(4), 582–621.

Bhaskar, V., & To, T. (1999). Minimum wages for ronald mcdonald monopsonies: A theory of monopsonistic competition. *Economic Journal*, *109*(455), 190–203.

Bourguignon, F. (1999). Multidimensioned approaches to welfare analysis: Comment. *Handbook of income inequality measurement* (pp. 477–484). With a foreword by Amartya Sen.

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *American Economic Review*, *84*(4), 772–793.

Carlier, G., Chernozhukov, V., & Galichon, A. (2016). Vector quantile regression: An optimal transport approach. *Ann. Statist.*, *44*(3), 1165–1192.

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs: Evidence from the united states using a bunching estimator. *Quarterly Journal of Economics (forthcoming).*

Chernozhukov, V., Fernandez-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, *81*(6), 2205–2268.

Chernozhukov, V., Fernández-Val, I., & Luo, S. (2020). Distribution regression with sample selection, with an application to wage decompositions in the uk.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. *Acad. Royl. Belg., Bull. C1 Sci. 5ième sér*, *65*, 274–292.

Dessing, M. (2002). Labor supply, the family and poverty: The s-shaped labor supply curve. *Journal of Economic Behavior and Organization*, *49*(4), 433–458.

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

Donald, S. G., & Hsu, Y.-C. (2016). Improving the power of tests of stochastic dominance. *Econometric Reviews*, *35*(4), 553–585.

Duclos, J.-Y., Sahn, D. E., & Younger, S. D. (2006). Robust multidimensional poverty comparisons. *Economic Journal*, *116*(514), 943–968.

Duclos, J.-Y., Sahn, D. E., & Younger, S. D. (2007). Robust multidimensional poverty comparisons with discrete indicators of well-being. In S. Jenkins & J. Micklewright (Eds.). Oxford University Press.

Fernández-Villaverde, J. (2018). The economic consequences of labor market regulations. *University of Chicago Legal Forum: Vol. 2017, Article 6.*

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953–973.

Firpo, S. P., Fortin, N. M., & Lemieux, T. (2018). Decomposing wage distributions using recentered influence function regressions. *Econometrics*, *6*(2), 1–40.

Flinn, C. J. (2006). Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates. *Econometrica*, *74*(4), 1013–1062.

Foresi, S., & Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association*, *90*(430), 451–466.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Chapter 1 - decomposition methods in economics. In O. Ashenfelter & D. Card (Eds.). Elsevier.

Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, *52*(3), 761–766.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, *12*, 1–115.

Harasztosi, P., & Lindner, A. (2019). Who pays for the minimum wage? *American Economic Review*, *109*(8), 2693–2727.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature*, *48*(2), 356–98. https://doi.org/10.1257/jel.48.2.356

Jardim, E., Long, M. C., Plotnick, R., van Inwegen, E., Vigdor, J., & Wething, H. (2017). *Minimum wage increases, wages, and low-wage employment: Evidence from seattle* (Working Paper No. 23532). National Bureau of Economic Research.

Jardim, E., Long, M. C., Plotnick, R., van Inwegen, E., Vigdor, J., & Wething, H. (2018). *Minimum wage increases and individual employment trajectories* (Working Paper No. 25182). National Bureau of Economic Research.

Kearney, M. S., & Harris, B. H. (2014). *The "ripple effect" of a minimum wage increase on american workers*. Discussion paper, Brookings.

Lee, D. S. (1999). Wage inequality in the united states during the 1980s: Rising dispersion or falling minimum wage?. *Quarterly Journal of Economics*, *114*(3), 977–1023.

Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *American Economic Review*, *96*(3), 461–498. https://doi.org/10.1257/aer.96.3.461

Linton, O., Maasoumi, E., & Whang, Y.-J. (2005). Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies*, *72*(3), 735–765.

Lugo, M. (2005). Comparing multidimensional indices of inequality: Methods and application. *ECINEQ, Society for the Study of Economic Inequality, Working Papers*, *14*.

Maasoumi, E. (1986). The measurement and decomposition of multi-dimensional inequality. *Econometrica*, *54*(4), 991–997.

Maasoumi, E. (1993). A compendium to information theory in economics and econometrics. *Econometric Reviews*, *12*(2), 137–181.

Maasoumi, E., & Lugo, M. A. (2008). The information basis of multivariate poverty assessments. In N. Kakwani & J. Silber (Eds.), *Quantitative approaches to multidimensional poverty measurement* (pp. 1–29). Palgrave Macmillan UK.

Maasoumi, E., & Wang, L. (2019). The gender gap between earnings distributions. *Journal of Political Economy*, *127*(5), 2438–2504.

MaCurdy, T. (2015). How effective is the minimum wage at supporting the poor? *Journal of Political Economy*, *123*(2), 497–545.

Manning, A. (2003). *Monopsony in motion: Imperfect competition in labor markets*. Princeton; Oxford:

Manning, A. (2021). The elusive employment effect of the minimum wage. *Journal of Economic Perspectives*, *35*(1), 3–26. https://doi.org/10.1257/jep.35.1.3

McFadden, D. (1989). Testing for stochastic dominance. *Studies in the economics of uncertainty: In honor of josef hadar* (pp. 113–134). New York; Berlin; London; Tokyo:

Meer, J., & West, J. (2016). Effects of the minimum wage on employment dynamics. *Journal of Human Resources*, *51*(2), 500–522.

Neumark, D., & Wascher, W. (2008). *Minimum wages*. MIT Press.

Neumark, D. (2018). The econometrics and economics of the employment effects of minimum wages: Getting from known unknowns to known knowns. *National Bureau of Economic Research Working Paper 25043*.

Neumark, D., & Wascher, W. (2000). Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania: Comment. *American Economic Review*, *90*(5), 1362–1396.

Rebitzer, J. B., & Taylor, L. J. (1995). The consequences of minimum wage laws: Some new theoretical ideas. *Journal of Public Economics*, *56*(2), 245–255.

Robinson, J. (1933). *The economics of imperfect competition.* McMillan.

Roth, J., & Sant'Anna, P. H. C. (2021). When is parallel trends sensitive to functional form?

Sen, A. K. (1976). Poverty: An ordinal approach to measurement. *Econometrica*, *44*(2), 219–231.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Shapiro, C., & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review*, *74*(3), 433–444.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica*, *48*(3), 613–625.

Sorkin, I. (2015). Are there long-run effects of the minimum wage?. *Review of Economic Dynamics*, *18*(2), 306–333.

Stigler, G. J. (1946). The economics of minimum wage legislation. *American Economic Review*, *36*, 358–365.

Tsui, K.-y. (1995). Multidimensional generalizations of the relative and absolute inequality indices: The atkinson-kolm-sen approach. *Journal of Economic Theory*, *67*(1), 251–265.

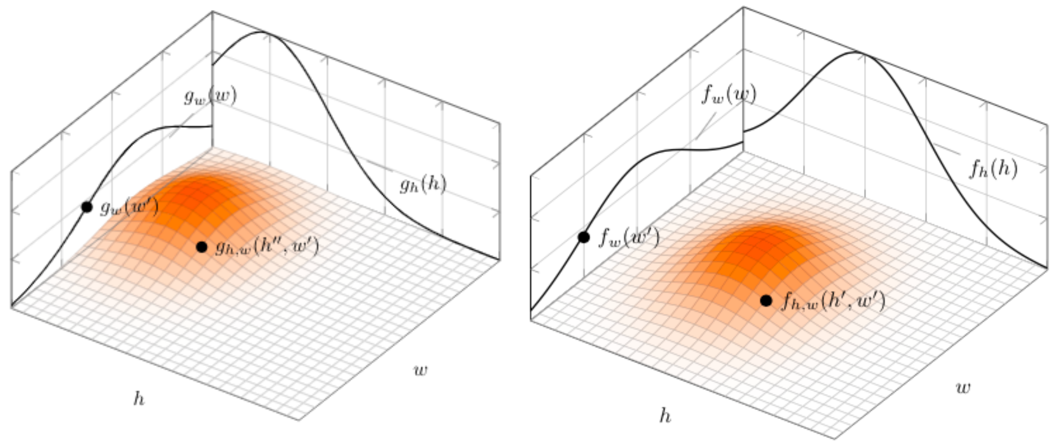
Tsui, K.-y. (1999). Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation. *Social Choice and Welfare*, *16*(1), 145–157.

Vaghul, K., & Zipperer, B. (2016). *Historical state and sub-state minimum wage data* (Working Paper). Washington Center for Equitable Growth.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Whang, Y.-J. (2019). *Econometric analysis of stochastic dominance: Concepts, methods, tools, and applications*. Cambridge University Press.

Figures



(a) With Minimum Wage Change

(b) Without Minimum Wage Change

Figure 1: Let $g_{h,w}$ be the observed frequency distributions after a change in minimum wage for hours worked and hourly wage, with respective marginal distributions g_h and g_w. Let f be analogous for the counterfactual distribution had minimum wage not changed. $f_{h,w}(h', w')$ is the number of individuals making wage w' and working h' hours whereas $f_w(w')$ is the number of individuals making wage w' for any number of hours worked. Clearly, just because every individual might earn higher wages with a minimum wage change — $G(w) < F(w)$ for all $w \in \mathbb{R}$ — does not mean $g_{h,w}$ is "preferable" to $f_{h,w}$. Indeed, in this example, while there are fewer individuals working a low wage with the minimum wage change (e.g. $g(w') < f(w')$), those same individuals are working fewer hours (e.g. $f_{h,w}(h', w') > g_{h,w}(h'', w')$).

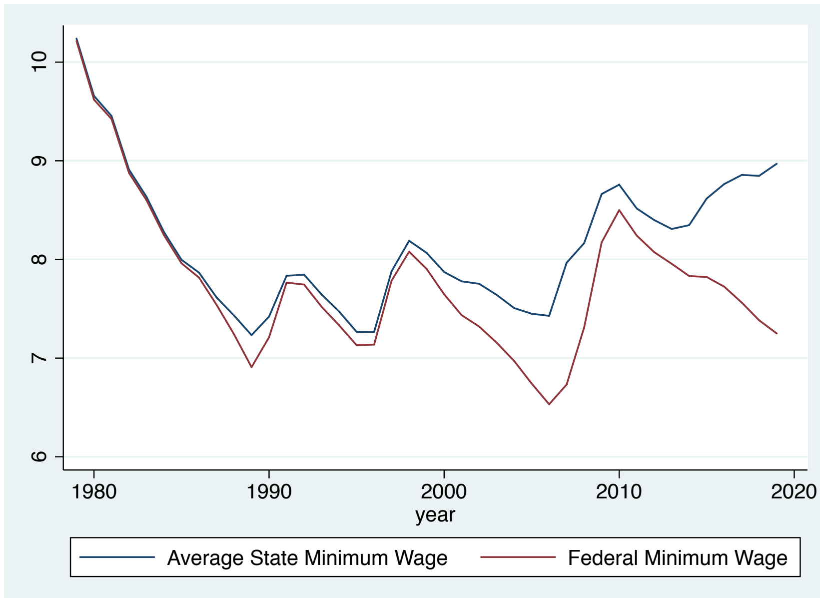
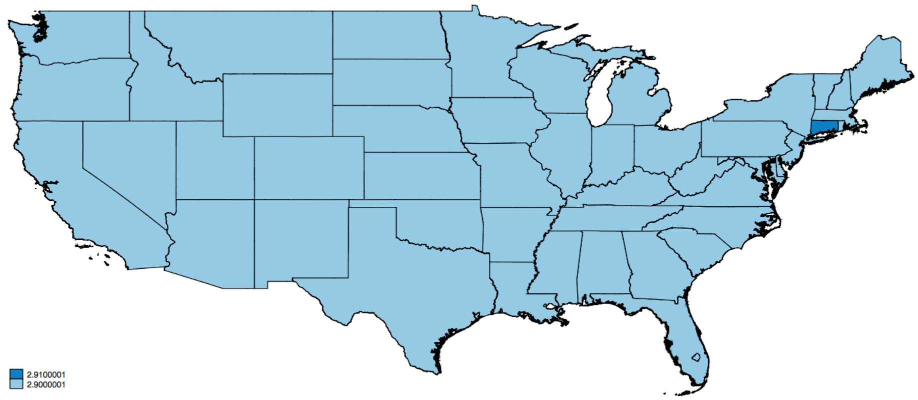
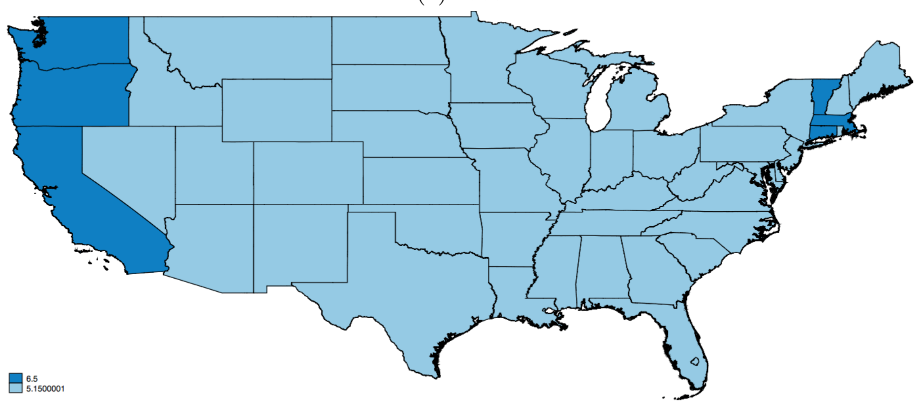


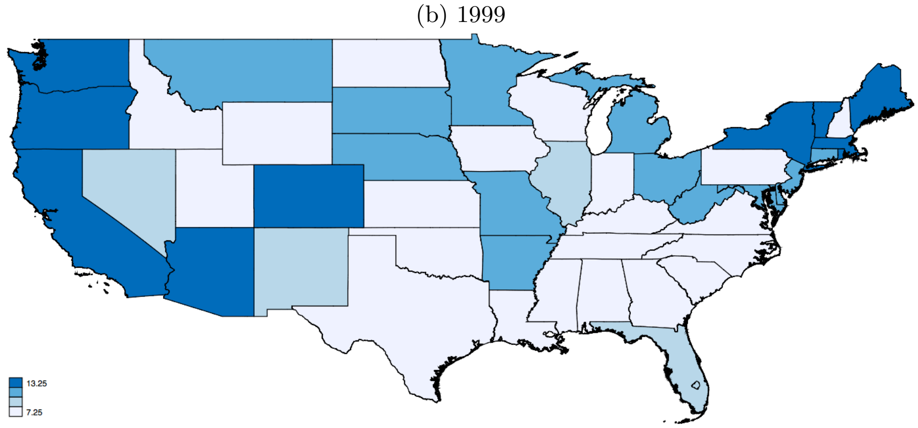
Figure 2: US Minimum Wage 1979-2019 in 2019 dollars



(a) 1979



(b) 1999



(c) 2019

Figure 3: Nominal State Minimum Wages in the US

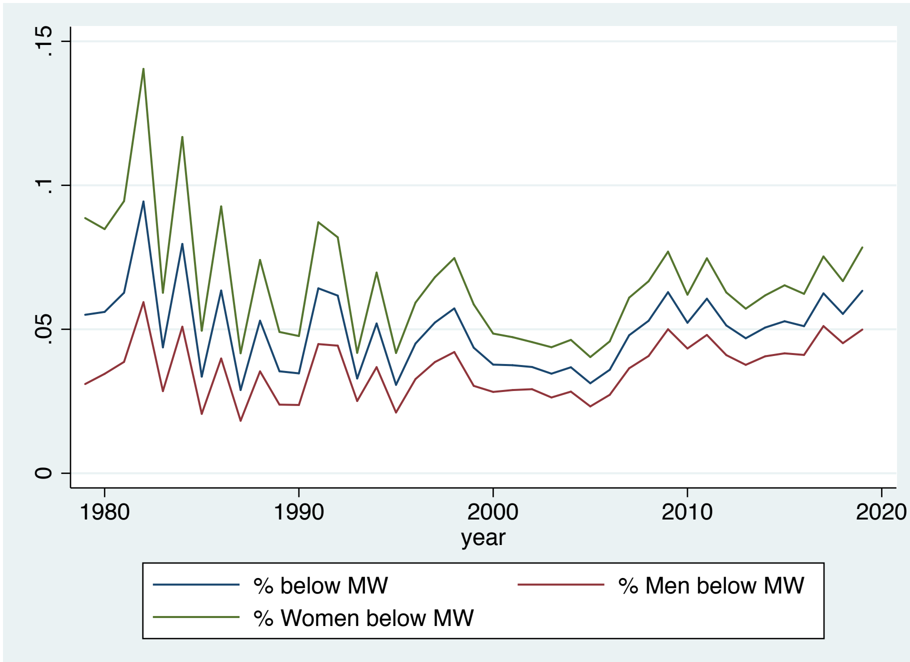


Figure 4: Percent of Worker earning at or Below Minimum Wage 1979-2019

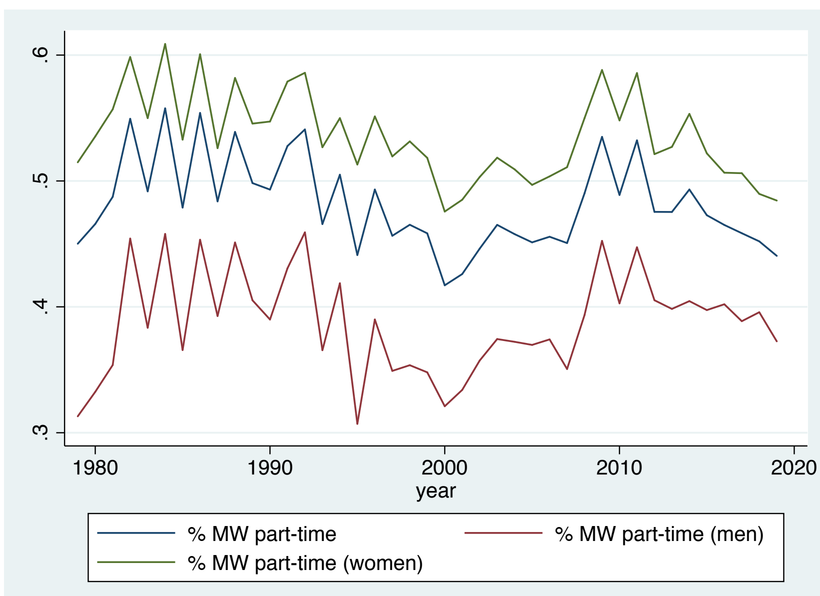


Figure 5: Percent of Minimum Wage Workers who are Part-time

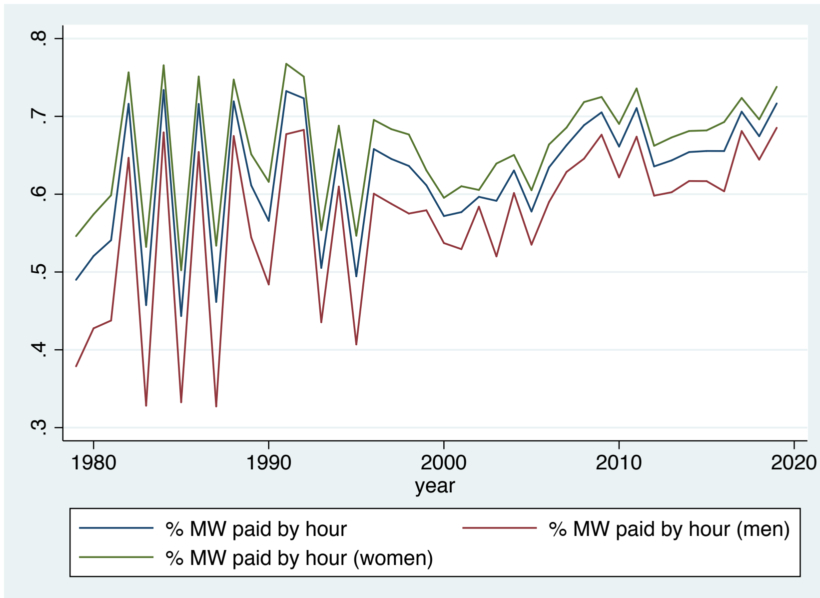
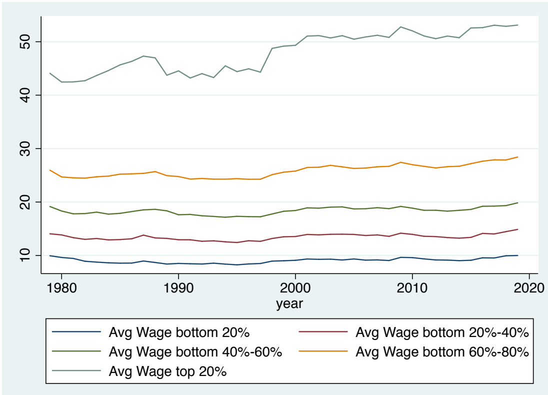
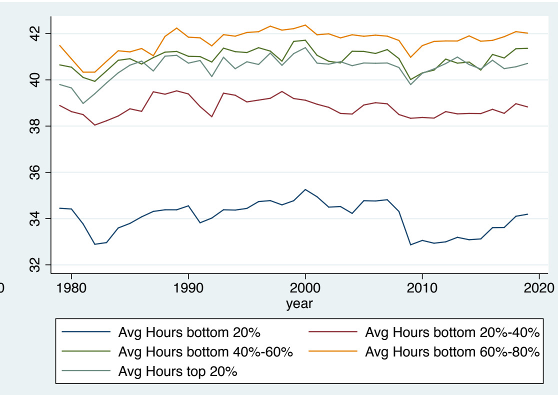


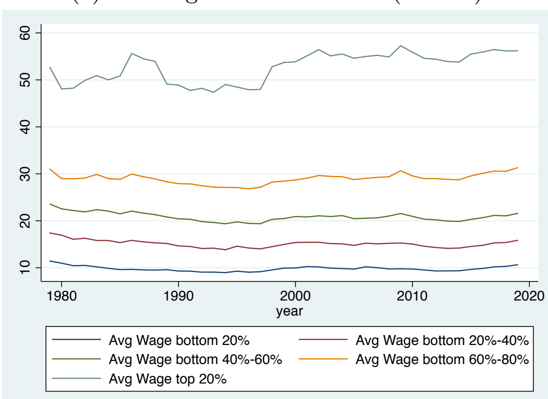
Figure 6: Percent of Minimum Wage Workers paid by the Hour



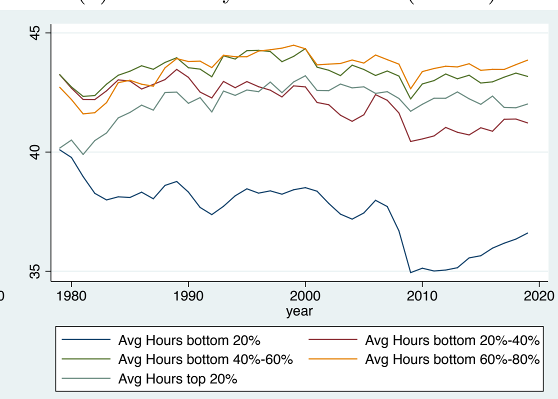
(a) US Wages in 2019 dollars (Pooled)



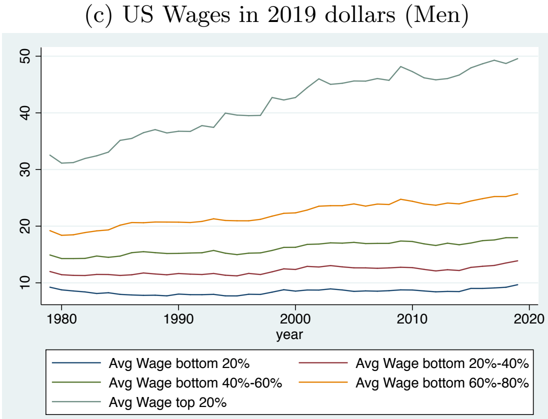
(b) US Weekly Hours Worked (Pooled)



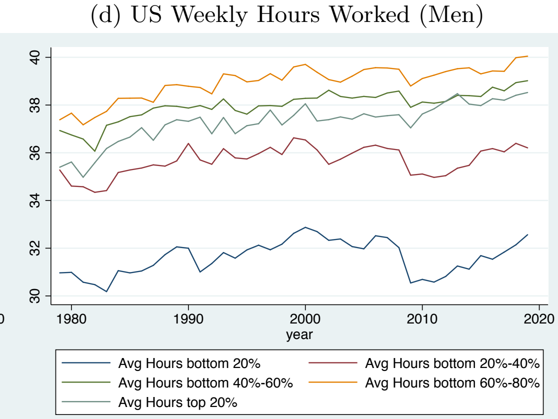
(c) US Wages in 2019 dollars (Men)



(d) US Weekly Hours Worked (Men)



(e) US Wages in 2019 dollars (Women)



(f) US Weekly Hours Worked (Women)

Figure 7: Average Wages and Hours of 5 Quantiles Ranges of Wages 1979-2019

Tables

Table 1: Nominal Federal Minimum Wage Timeline

Date	Minimum wage
January 1, 1979	$2.90
January 1, 1980	$3.10
April 1, 1990	$3.80
April 1, 1991	$4.25
October 1, 1996	$4.75
September 1, 1997	$5.15
July 24, 2007	$5.85
July 24, 2008	$6.55
July 24, 2009	$7.25

Note: Minimum wages are for for all covered nonexempt workers.
Source: https://www.dol.gov/agencies/whd/minimum-wage/history/chart

Table 2: Men 89-92

Wages \ Hours	24	35	40	42	45	48	52	60
6.89	0.04*	0.03*	0.02*	0.02*	0.01*	0.01*	0.01*	0.01*
9.35	0.02*	0.05*	0.07*	0.07*	0.06*	0.06*	0.05*	0.05*
11.73	0.02*	0.04*	0.10*	0.10*	0.08*	0.08*	0.07*	0.07*
14.25	0.01*	0.02*	0.09*	0.09*	0.07*	0.07*	0.03*	0.03*
17.05	0.01*	0.02*	0.13*	0.13*	0.09*	0.09*	0.04*	0.04*
20.21	0.01*	0.01*	0.14*	0.14*	0.09*	0.09*	0.02*	0.02*
23.87	0.01*	0.01*	0.4*	0.15*	0.1*	0.1*	0.01*	0.01*
28.45	0.01*	0	0.1*	0.11*	0.1*	0.12*	0.04*	0.04*
36.45	0.01*	0	0.02*	0.03*	0.02*	0.04*	0.01*	0.01*

Note: $*$ = 95% confidence level using 30 bootstrap samples.

Table 3: Women 89-92

Wages \ Hours	24	35	40	42	45	48	52	60
6.89	0.07*	0.05*	0	0	0	0	0	0
9.35	0.07*	0.16*	0.01*	0.01*	0.01*	0.01*	0	0
11.73	0.05*	0.13*	0.04*	0.04*	0.03*	0.03*	0.02*	0.02*
14.25	0.03*	0.09*	0.01*	0.01*	0	0	−0.02*	−0.02*
17.05	0.03*	0.08*	0.04*	0.04*	0.01*	0.01*	−0.01*	−0.01*
20.21	0.02*	0.07*	0.06*	0.06*	0.04*	0.03*	0.01*	0.01*
23.87	0.02*	0.06*	0.06*	0.06*	0.03*	0.02*	−0.004*	−0.004*
28.45	0.01*	0.05*	0.04*	0.04*	0.05*	0.04*	0.01*	0
36.45	0.01*	0.05*	0.01*	0.01*	0.01*	0.01*	0.02*	0.02*

Note: $*$ = 95% confidence level using 30 bootstrap samples.

Table 4: Men 89-84

Wages \ Hours	24	35	40	42	45	48	53	60
6.86	0.02*	0.03*	0	0	0	0	0	0
9.20	0.004*	0.05*	0.03*	0.03*	0.02*	0.02*	0.01*	0.01*
11.62	−0.005*	0.04*	0.05*	0.05*	0.04*	0.03*	0.01*	0.01*
14.09	−0.009*	0.03*	0.08*	0.08*	0.06*	0.05*	0.01*	0.01*
16.85	−0.01*	0.03*	0.09*	0.09*	0.06*	0.05*	0	0
19.86	−0.01*	0.03*	0.15*	0.15*	0.11*	0.1*	0.03*	0.03*
23.29	−0.01*	0.03*	0.13*	0.13*	0.09*	0.07*	0.01*	−0.01*
27.61	−0.01*	0.02*	0.1*	0.1*	0.12*	0.1*	0.005*	0.005*
34.85	−0.02*	0.02*	0.06*	0.06*	0.08*	0.09*	0.05*	0.05*

Note: $*$ = 95% confidence level using 30 bootstrap samples.

Table 5: Women 89-84

Wages \ Hours	24	35	40	42	45	48	53	60
6.86	0.08*	0.05*	0	0	0	0	0	0
9.20	0.08*	0.18*	0.04*	0.04*	0.03*	0.03*	0.03*	0.03*
11.62	0.05*	0.13*	0.05*	0.04*	0.03*	0.03*	0.03*	0.03*
14.09	0.04*	0.1*	0.06*	0.05*	0.04*	0.03*	0.03*	0.03*
16.85	0.03*	0.09*	0.07*	0.06*	0.04*	0.03*	0.03*	0.03*
19.86	0.02*	0.08*	0.08*	0.07*	0.05*	0.04*	0.04*	0.04*
23.29	0.01*	0.07*	0.1*	0.09*	0.06*	0.06*	0.06*	0.06*
27.61	0.01*	0.06*	0.06*	0.05*	0.05*	0.05*	0.05*	0.05*
34.85	0.01*	0.06*	0.04*	0.03*	0.03*	0.03*	0.06*	0.06*

Note: $*$ = 95% confidence level using 30 bootstrap samples.

Table 6: Men 06-12

Wages \ Hours	21	32	37	40	41	45	55	55	65
8.33	0.05*	0.06*	0.05*	0.04*	0.04*	0.04*	0.03*	0.03*	0.03*
11.13	0.03*	0.09*	0.1*	0.08*	0.08*	0.07*	0.06*	0.06*	0.05*
13.88	0.02*	0.07*	0.1*	0.09*	0.09*	0.07*	0.05*	0.05*	0.05*
16.51	0.02*	0.06*	0.09*	0.14*	0.14*	0.12*	0.09*	0.08*	0.05*
19.47	0.02*	0.05*	0.07*	0.15*	0.15*	0.12*	0.07*	0.06*	0.02*
22.91	0.01*	0.04*	0.06*	0.18*	0.18*	0.13*	0.08*	0.06*	0.01*
27.31	0.01*	0.03*	0.05*	0.18*	0.18*	0.15*	0.08*	0.06*	−0.01*
33.57	0.01*	0.02*	0.04*	0.13*	0.13*	0.12*	0.08*	0.05*	−0.02*
44.76	0.01*	0.02*	0.03*	0.09*	0.09*	0.08*	0.1*	0.09*	0.01*

Note: $*$ = 95% confidence level using 30 bootstrap samples.

Table 7: Women 06-12

Wages \ Hours	21	32	37	40	41	45	55	55	65
8.33	0.1*	0.07*	0.06*	0.03*	0.03*	0.03*	0.02*	0.02*	0.02*
11.13	0.08*	0.16*	0.17*	0.07*	0.07*	0.06*	0.05*	0.05*	0.05*
13.88	0.07*	0.12*	0.16*	0.07*	0.06*	0.05*	0.03*	0.03*	0.02*
16.51	0.06*	0.09*	0.13*	0.09*	0.08*	0.07*	0.04*	0.03*	0.02*
19.47	0.06*	0.07*	0.10*	0.1*	0.08*	0.06*	0.03*	0.02*	0
22.91	0.06*	0.06*	0.09*	0.11*	0.1*	0.07*	0.03*	0.02*	0
27.31	0.05*	0.05*	0.07*	0.13*	0.11*	0.08*	0.03*	0.02*	0
33.57	0.05*	0.03*	0.05*	0.1*	0.09*	0.09*	0.03*	0.02*	0
44.76	0.05*	0.02*	0.03*	0.05*	0.04*	0.05*	0.04*	0.02*	0

Note: $*$ = 95% confidence level using 30 bootstrap samples.

Table 8: Effect of Minimum Wage with State and Year Fixed Effects, 2003-2019

	Hours Worked				log(Wages)			
	Men		Women		Men		Women	
Total	0.168***	(0.059)	0.343***	(0.047)	0.018***	(0.003)	0.015***	(0.003)
Bottom %20	-0.044	(0.186)	0.228**	(0.112)	0.01	(0.007)	0.008	(0.005)
%20-%40	0.264*	(0.154)	0.325***	(0.103)	0.003**	(0.001)	0.001	(0.001)
%40-%60	0.14	(0.132)	0.336***	(0.098)	0.002**	(0.001)	0.001	(0.001)
%60-%80	0.208*	(0.119)	0.282***	(0.099)	0.001	(0.001)	0	(0.002)
Top %20	0.258**	(0.101)	0.524***	(0.113)	0.001	(0.003)	0.005	(0.004)

Note: Standard error in parenthesis.
$*$ = significant at a 90% confidence level.
$**$ = significant at a 95% confidence level.
$***$ = significant at a 99% confidence level.

Table 9: Effect of Minimum Wage with State and Year Fixed Effects, 2003-2019

		Men			
	$\hat{\beta}_1$		$\hat{\beta}_2$		$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$
Total	0.35	(0.361)	-0.011	(0.021)	15.898
Bottom %20	0.316	(1.261)	-0.022	(0.075)	7.327
%20-%40	0.152	(0.999)	0.007	(0.06)	-11.038
%40-%60	-2.146^{**}	(0.907)	0.14^{**}	(0.055)	7.653
%60-%80	1.556^{**}	(0.744)	-0.082^{*}	(0.045)	9.469
Top %20	1.096^{*}	(0.565)	-0.049	(0.033)	11.085
		Women			
	$\hat{\beta}_1$		$\hat{\beta}_2$		$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$
Total	-0.317	(0.287)	0.04^{**}	(0.017)	3.98
Bottom %20	-0.361	(0.814)	0.036	(0.05)	4.947
%20-%40	-2.108^{***}	(0.681)	0.15^{***}	(0.042)	7.007
%40-%60	1.021	(0.649)	-0.042	(0.039)	12.208
%60-%80	-1.212^{**}	(0.576)	0.09^{***}	(0.034)	6.754
Top %20	0.946	(0.592)	-0.024	(0.033)	19.615

Note: Standard error in parenthesis.

$*$ = significant at a 90% confidence level.

$**$ = significant at a 95% confidence level.

$***$ = significant at a 99% confidence level.

Appendix

A.1. Simulation

With a sample size of 3000, let

$$
\begin{aligned}
x &\sim N(10,3),\\
z &\sim N(2,12),\\
\epsilon_1 &\sim N(0,4),\\
\epsilon_2 &\sim N(0,4),\\
y^1 &= x+\epsilon_1,\\
y^2 &= x+z+\epsilon_2.
\end{aligned}
$$

Hence

$$
\begin{aligned}
y^1 &\sim N(10,5)\\
y^2 &\sim N(12,13).
\end{aligned}
$$

Figures A.1.1 and A.1.2 show histograms of $\hat{F}_{y^1}$ and $\hat{F}_{y^1}$ by running distribution regressions with logit link functions of x on y^1 and x,z on y^2, respectively. Table A.1.1 shows the difference between the estimated joint distribution, $\hat{F}_{y^1,y^2}$, obtained by empirical copula and the empirical CDF.

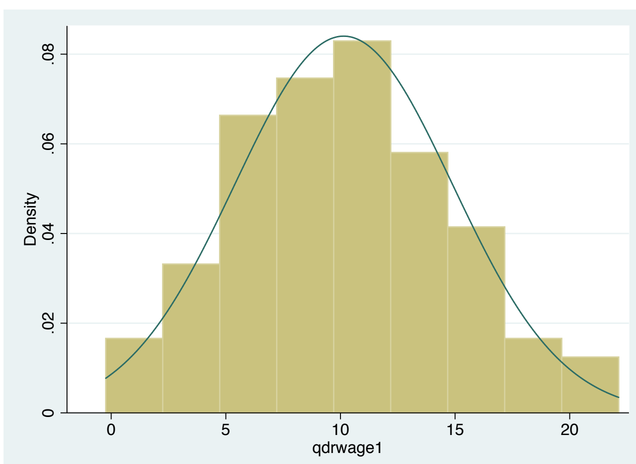


Figure A.1.1

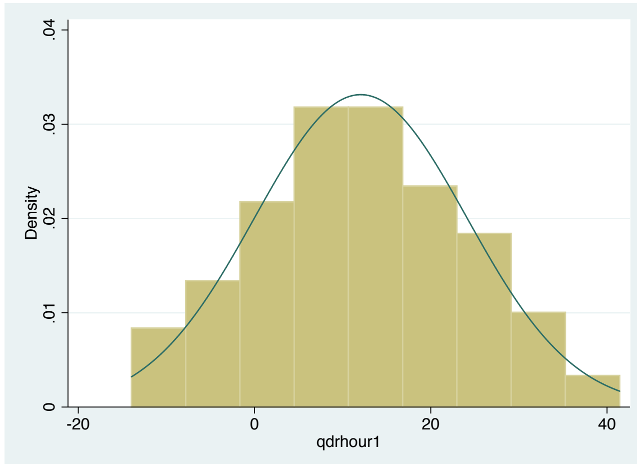


Figure A.1.2

Table 10: Table A.1.1

$y^1 \backslash y^2$	-4.65	1.07	5.26	8.52	11.39	14.91	18.33	22.64	28.38
3.72	0	0	0	0	-0.006	0	0	0	0
5.93	0	0	0	0	0	0	0	0	0
7.31	0	-0.005	0	0	0	-0.006	0	0	0
8.62	0	0	-0.005	0	-0.01	-0.007	0	-0.012	-0.013
9.95	0	0	0	0	-0.014	-0.01	0	0	0
11.28	0	0	0	0	0	0	0.005	0	0.01
12.53	0	0	0	0	0	0.006	0.014	0.009	0.017
14.08	0	0	0	0	-0.014	0	0.009	0	0
16.37	-0.006	-0.005	0	-0.006	-0.011	0	0	0	0

B.1. Proofs

Proof of Theorem 3.2. The first part of this proof is essentially identical to the proof of Lemma 2.1 of Chernozhukov et al. (2013). Note that $Y^1 = \mathbb{1}\{J = j\}Y^{1*}$ and $Y^2 = \mathbb{1}\{J = j\}Y^{2*}$. Also, by definition,

$$Y_j^1 := Y^1|J = j \text{ and } X_k^1 := X^1|J = k, \tag{$\star$}$$

and

$$Y_j^2 := Y^2|J = j \text{ and } X_k^2 := X^2|J = k, \tag{$\star\star$}$$

Then, by the law of iterated probability

$$\begin{aligned}
F_{Y_j^{1*}|J}(y|k) &= \int_{\mathcal{X}_k^1} F_{Y_j^{1*}|J,X^1}(y|k,x^1)dF_{X^1|J}(x|k) \\
&= \int_{\mathcal{X}_k^1} F_{Y_j^{1*}|J,X^1}(y|j,x^1)dF_{X^1|J}(x|k) \\
&= \int_{\mathcal{X}_k^1} F_{Y_j^{1*}|X_j^1}(y|x^1)dF_{X_k^1}(x).
\end{aligned}$$

The second equality follows from Assumption 2 and the last equality follows from ($\star$). Next, by the law of iterated probability

$$\begin{aligned}
F_{Y_j^{2*}|J}(y|k) &= \int_{\mathcal{X}_k^2} F_{Y_j^{2*}|J,X^2}(y|k,x^2)dF_{X^2|J}(x|k) \\
&= \int_{\mathcal{X}_k^1} F_{Y_j^{2*}|J,X^2}(y|j,x^2)dF_{X^2|J}(x|k) \\
&= \int_{\mathcal{X}_k^2} F_{Y_j^{2*}|X_j^2}(y|x^2)dF_{X_k^2}(x).
\end{aligned}$$

The second equality follows from Assumption 2, and the last equality follows from ($\star\star$).

By Sklar's Theorem, under Assumption 3,

$$F_{Y^{1*},Y^{2*}\langle j,k\rangle}(y^1,y^2) = C(F_{Y^{1*}_{\langle j,k\rangle}}(y^1), F_{Y^{2*}_{\langle j,k\rangle}}(y^2))$$

is unique. ■