Distribution Agreement

Signature:

Yisroel Cahn Date

Essays in Labor Economics and Econometrics

By

Yisroel Cahn
Doctor of Philosophy

Economics

Esfandiar Maasoumi, Ph.D.
Advisor

Ruixuan Liu, Ph.D.
Committee Member

Evan Saltzman, Ph.D.
Committee Member

Accepted:

Kimberly J. Arriola, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Essays in Labor Economics and Econometrics

By

Yisroel Cahn
B.A & B.S., University of Maryland, Baltimore County, MD, 2016
M.A., Emory University, GA, 2020

Advisor: Esfandiar Maasoumi, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Economics
2022

Abstract

Essays in Labor Economics and Econometrics
By Yisroel Cahn

This dissertation explores methods of estimating and evaluating policy interventions that have heterogeneous effects on the outcome of interest.

Policy evaluation generally compares an observed outcome of interest with the estimated counterfactual outcome had the policy not been implemented. Chiefly, this is done by comparing means or means conditional on a subgroup of the population. However, if the policy has heterogeneous effects on the distribution of outcomes — for example, if 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 inequality or poverty. The first chapter surveys the literature on interpreting heterogeneous outcomes of a policy intervention.

The second chapter looks at an example — minimum wage policy. Recent proposals to raise the U.S. Federal minimum wage to $15 an hour are designed to improve the welfare of low-wage workers, but may involve important economic trade-offs. Although the effects of minimum wage on employment and wages have been well studied, little is known about its effects on hours worked which may be responsive to minimum wage changes. One reason for the lack of research is that hours are only observed for those who are employed, but workers could be exiting or entering the market and biasing the results. I fill this gap in the literature by employing a Heckman-type selection model to estimate the effects of minimum wage on hours worked, accounting for possible employment effects. Using U.S. Current Population Survey data, I found that increases to the minimum wage increased the hours worked of low-wage workers. However, I also found that the effects varied by industry—fast food and accommodation service workers saw their hours decrease while most other industries' minimum wage workers saw their hours increase, suggesting that market structure could be causing these findings. Additionally, I propose a new method to estimate jointly determined outcomes, showing that hours and wages both increased for low-wage workers.

The final chapter uses machine learning methods to predict intergenerational income mobility in the United States. The machine learning methods are (1) non-parametric and are not sensitive to functional form misspecification, (2) give an out-of-sample performance indicator, and (3) allow for predictors to be ranked by how importance they are to the overall prediction. I find that family wealth, and not parent income, is the most important predictor of child income for large increases in the income distribution.

Essays in Labor Economics and Econometrics

By

Yisroel Cahn
B.A & B.S., University of Maryland, Baltimore County, MD, 2016
M.A., Emory University, GA, 2020

Advisor: Esfandiar Maasoumi, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Economics
2022

Contents

List of Figures

List of Tables

Chapter 1

Inequality and Policy Evaluation

1.1 Introduction

Economic inequality and its causes are gaining renewed interest in both academic and political circles.[1] However, the standard methods of policy evaluation (average treatment effect, local average treatment effect, etc.)[2] focus on the mean outcome rather than the distribution of outcomes, and therefore do not allow for inequality or poverty comparisons. This chapter argues that inequality should be taken into account in policy decisions and offers possible ways of doing so.

In a 2011 issue of AER P&P, Anthony B. Atkinson asserts “[e]conomists need to be more explicit about the relation between welfare criteria and the objectives of governments, policymakers and individual citizens.” Indeed, such normative discussions are sidestepped when only means are considered.

For example, in evaluating “right to work” laws which forbid unions from interfering with the employment of nonunion workers, some workers may be hurt while

[1] For example, Blume and Durlauf (2015) discuss the merits and shortcomings of the popular book “Capital in the Twenty-First Century” by Thomas Piketty, and predict an increase in inequality research by academics.

[2] Heckman, 2010 defines three broad classes of policy evaluation problems that arise in economics. This chapter is concerned with the first class, i.e. evaluating the impacts of implemented interventions.

others benefit. Looking at the mean alone might indicate that such laws are neutral or favorable while ignoring ethical questions surrounding such laws.

Another example is forming combined classes for students with high test scores and students with low test scores. Examining the mean test scores of the resulting combined class might not give an accurate picture of who, if anyone, benefited and by how much. Reasonable evaluation criteria for such a policy might entail determining whether the gap in test scores between the two groups of students has been reduced or whether test scores are above a given threshold.

Furthermore, if there are multiple outcomes of interest, individuals might be affected positively in one outcome but negatively in another. In many cases, welfare might not only be measured in inequality or poverty in one dimension, but rather by some functional of the joint distribution of outcomes. In Chapter 2, I examine whether increases in minimum wage reduced the hours worked of those individuals whose wages were increased. Simply looking at average treatment effects of hours and wages separately would not accurately reflect such an outcome.

This also highlights an important aspect of welfare that has been overlooked by much of the policy evaluation literature — the effect of the policy in the short- medium- and long-term. Multidimensional inequality or poverty comparisons can address such issues by using outcomes of individuals in different periods as dimensions.

Methods for estimating counterfactual distributions to compare the distribution of outcomes with and without a policy change generally use decomposition methods. For an excellent survey on distribution decomposition methods, see Fortin et al. (2011).

1.2 Facts about Inequality and Poverty in the US

Figure 1.1 shows Gini indexes of weekly earnings in the US for employed individuals ages 18 through 64 from 1979 to 2019.[3] When the population is split by gender, inequality is lower in either group than when the two are pooled together, showing that much of the inequality displayed in the overall graph results from gaps between the genders. This highlights a shortcoming of the commonly used Gini index. The Gini index is not "subgroup decomposable," meaning that it does not permit a breakdown of the overall inequality in the population into subgroups. In the next section, subgroup decomposable inequality indexes are discussed.

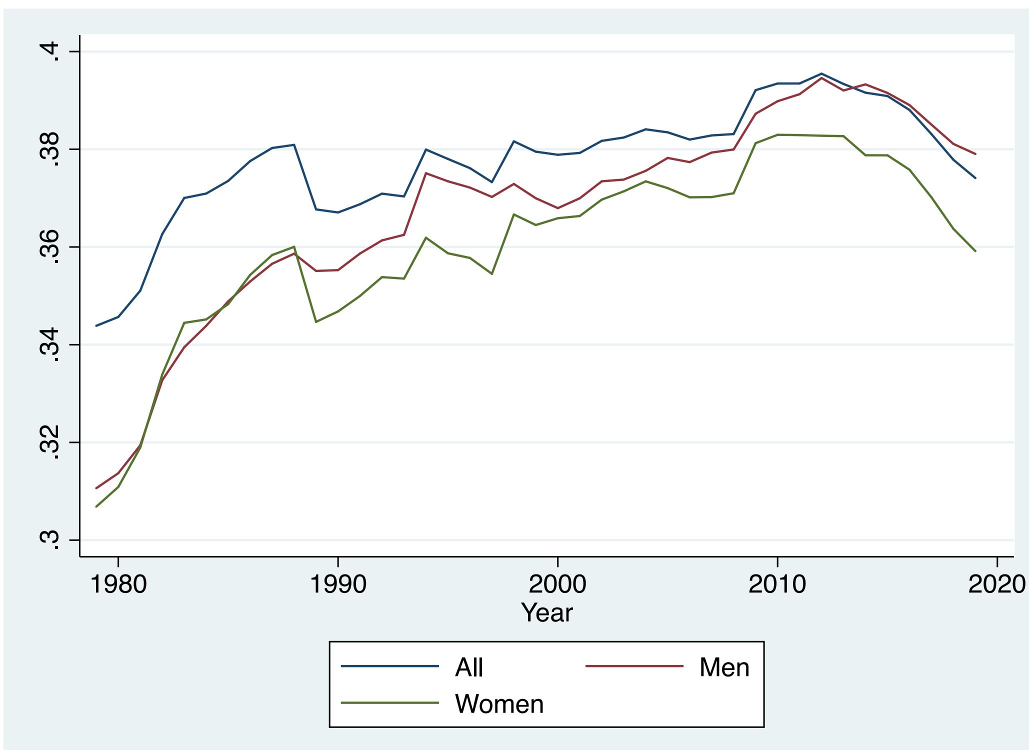


Figure 1.1: Gini Index

Inequality increased since the 1970's, but has decreased over the last few years.

Figure 2.1 shows mean weekly earnings, the 90th percentile of weekly earnings

[3]Data was collected from US Current Population Survey Outgoing Rotation Group. See Chapter 2 Section 5 for details on data cleaning.

minus the 10th, and the 90th percentile of weekly earnings minus the 50th, all in 2019 dollars. Mean weekly earnings increased since the 1970's and seem to be driven by weekly earnings increasing at the top of the distribution; the difference between the 90th percentile and the 10th percentile is proportional to the difference between the 90th percentile and the 50th percentile.

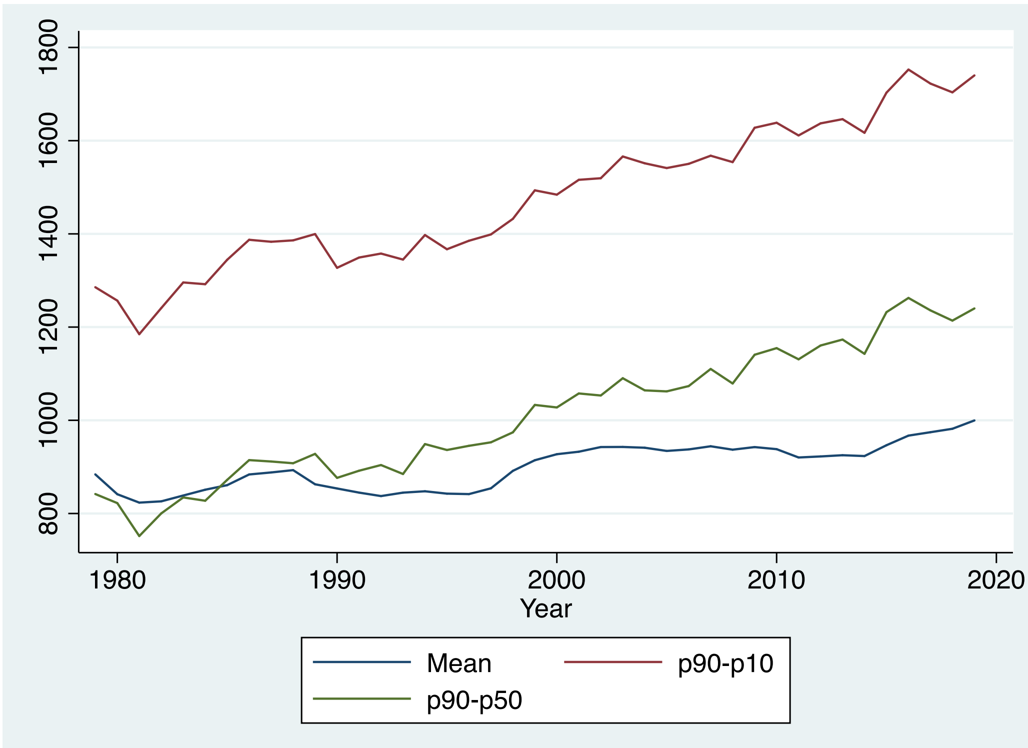


Figure 1.2: Weekly Earnings (2019 Dollars)

Figure 2.7 shows the Foster et al. (1984) poverty index of weekly earnings with different parameter values and a poverty threshold of $400 per week (half the median weekly earnings of individuals in 2019). As opposed to inequality, poverty seems to have remained constant and then decreased since the 1970's. Extreme poverty (FGT(2) which weighs individuals falling far below the poverty threshold more heavily) was always low.

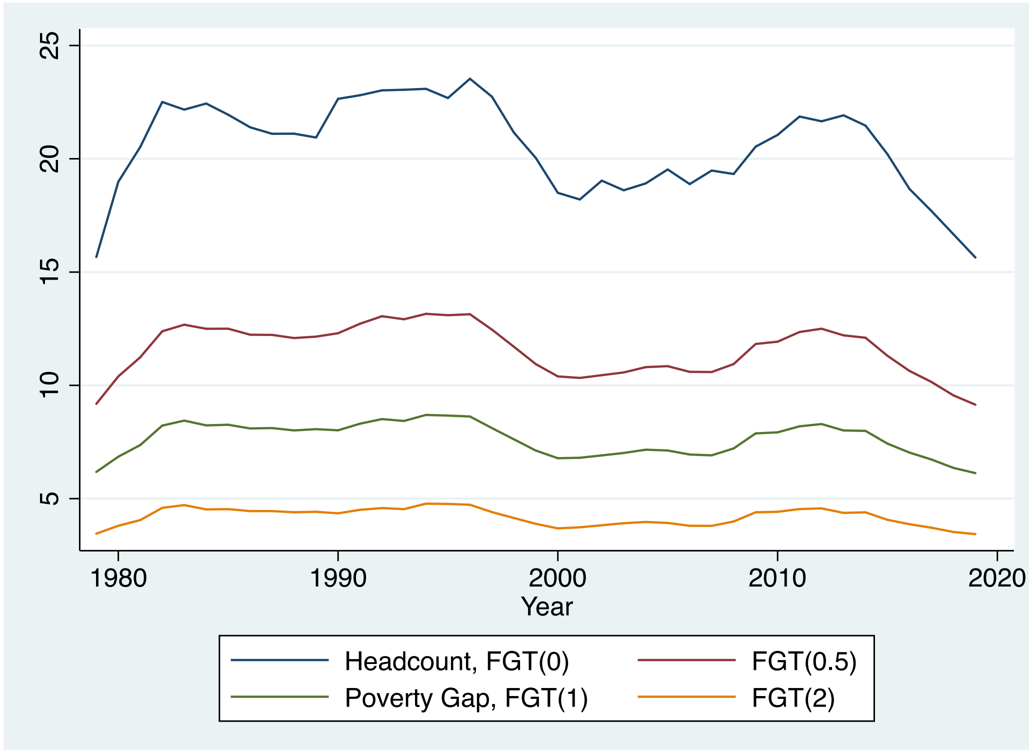


Figure 1.3: Poverty

1.3 Comparing Distributions

1.3.1 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. 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

$$\text{SWF} = W(u_1(\mathbf{x}_1), ..., u_n(\mathbf{x}_n)), \tag{1.1}$$

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 the SWF (1.1). First, using dominance criteria/stochastic dominance[4] that place restrictions on the utility functions of individuals in the population. Second, using an index that satisfies certain properties thought to define inequality.

Dominance criteria is generally less controversial than using an index because it uses fewer and less restrictive assumptions. However, if the dominance criteria are not present in a given population, this method does not produce a complete ordering of distributions. Although showing "first-order stochastic dominance" of one distribution over another is considered the gold standard in inequality comparisons, it cannot always be achieved. [5]

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 proposed policy would give all poor individuals a ten percent increase in income while giving wealthy individuals a twenty percent increase, most inequality indexes would assign such a distribution a higher level inequality even though every individual is made better off. Furthermore, many do not find the goal of reducing inequality for its own sake 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 which refers to some defined threshold of poverty) more reasonable than using an inequality index which is defined in terms of relationships between individuals.

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

[4]Dominance criteria are assumptions about which SWF is "preferable."

[5]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.

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, the SWF (1.1) becomes

$$\text{SWF} = \sum_{i=1}^{n} u(\mathbf{x}_i). \tag{1.2}$$

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 (1.2) is (strictly) larger than the other's.

It is 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.[6] With the class of utility functions $\mathcal{U}_1 \equiv \{u : u' \geq 0\}$, x first-order (strictly) stochastically dominates x' if the CDFs $F_x(w) < F_{x'}(w)$ for all $w \in \mathbb{R}$.[7] 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" (both distributions have the same mean), 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.

For the class of utility functions $\mathcal{U}_t \equiv \{u : u' \geq 0, u'' \leq 0, ..., (-1)^{t+1}u''^{\cdots\prime} \geq$

[6]This also allows us to compare distributions that do not have the same number of individuals.
[7]See Whang (2019) for details.

$0\}$, the distribution x is said to (strictly) t-order dominate x' if $\int_{-\infty}^{x} F_x^{\prime\cdots\prime}(w)dw < \int_{-\infty}^{x'} F_{x'}^{\prime\cdots\prime}(w)dw$ 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, an increase in a unit of education might change the marginal utility of an increase in income because those with more education might view an increase in income differently. 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 $w_1, w_2 \in \mathbb{R}$.[8]

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 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).

Inequality

There are two approaches to constructing inequality indexes — 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

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

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 to the information theoretic 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 indexes 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 indexes. 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 indexes.

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 to any subjectivity.

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.[9] 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

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

where $\delta_j = \alpha_j / \sum_j \alpha_j$. For the second step, Maasoumi (1986) uses the univariate generalized entropy inequality measure[10] over S.[11]

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

[10] See Shorrocks (1980).

[11] See Lin and Maasoumi (2019) for possible ways of calibrating first-step aggregation procedure.

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).[12]

Poverty

Foster et al. (1984) proposed a commonly used univariate class of poverty indexes that is subgroup decomposable. 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 has less poverty than another regardless of the poverty thresholds. In this way, their method is robust to the âunionâ (an individual is considered poor if she is below the poverty threshold in any dimension), âintersectionâ (an individual is considered poor if she is below the poverty threshold in all dimensions) or some âintermediateâ (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

[12] 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. An additional advantage of Maasoumi (1986) is that inequality need not be the focus of the aggregation (i.e. after the first step, the univariate distribution S can be used to construct other measures besides inequality such as the mean).

obtaining intermediate poverty indices, including the fuzzy set approach and the latent variables approach.

1.4 Conclusion

Atkinson (2011) lists several possible ways applied economists have rationalized neglecting welfare economics. These include assuming away differences in outcomes, assuming agreement on the welfare criteria, or even that welfare discussions are better suited to other disciplines. Atkinson argues against such stances and urges applied economists to renew their focus on welfare discussions. Indeed, a policy's distributional effects have important, non-trivial implications that are contingent on subjective values. This chapter offers a clear approach to explicitly accommodate different opinions regarding inequality and poverty, and heeds Atkinson's call to put welfare economics back in the spotlight.

This chapter serves as a literature review of methods that compare heterogeneous outcomes that are referenced in the next chapters. Methods of estimating counterfactual distributions are laid out in Chapter 2, and mobility measures are discussed in Chapter 3.

Chapter 2

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

2.1 Introduction

2.1.1 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 several studies conclude that increased minimum wage reduces wage inequality in the United States[1] and has mild effects on employment,[2] little is known about its effects on hours worked,[3] which might vary for individuals earning

[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 2.2, there is a growing consensus that these increases had little or no effect on employment, and future minimum wage research should instead 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

different wages and working different hours. By analyzing the joint distribution of wages and hours worked, this chapter offers insight into whether increased minimum wage reduces poverty and inequality and what mechanisms cause any such change.

Minimum wage increases have theoretically ambiguous effects on hours worked. In a competitive market, minimum wage causes an excess supply of hours. Assuming there are no general equilibrium effects, this leads to a reduction in hours. However, in a monopsonistic market, increased minimum wage might still be below the marginal revenue brought in by a low-wage worker's hour of work, and lead to an increase in equilibrium hours worked.[4] The effect minimum wage has on hours is an empirical question that has received little attention in the literature.

Furthermore, estimating the effects of minimum wage on hours is difficult because hours are only observed for those who are employed, but workers could be exiting or entering the market and biasing the results. I fill the gap in the literature by employing a Heckman-type selection model to estimate the effects of minimum wage on hours worked, accounting for possible employment effects. However, unlike the traditional Heckman selection model which models a worker's selection decision into the labor force, I models employment selection as a market outcome (i.e., the employer is allowed to select the employee's hours).

The main identification assumption for these selection models is an exclusion restriction — a variable is included in the labor market participation equation that does not affect the outcome variable. As is common in the literature, number of children under age five is used as an exclusion restriction for wages. Education variables are used as exclusion restrictions for hours worked.[5]

---

paper, the same researchers found little to no impact (Jardim et al., 2017; Jardim et al., 2018).

[4]See Section 2.2 for further discussion and other possible market frictions determining the theoretical effects of minimum wage.

[5]Having an exclusion restriction is not necessary for identification in a type II Tobit model, see Wooldridge (2001). However, identification would be due to the non-linearity of the inverse Mills ratio, in which case functional form misspecification in the population model could be the main determinant of the inverse Mills ratio.

Using U.S. Current Population Survey data, I found that over the 2003-2019 period,[6] a one dollar increase in minimum wage lead to a 9% increase in wages for both men and women, a small increase in employment, and an increase in 1.43 hours worked per week for men and 0.64 hours worked per week for women. The effect on hours became negative when minimum wage was raised past $13.77 for men and $12.81 for women.[7]

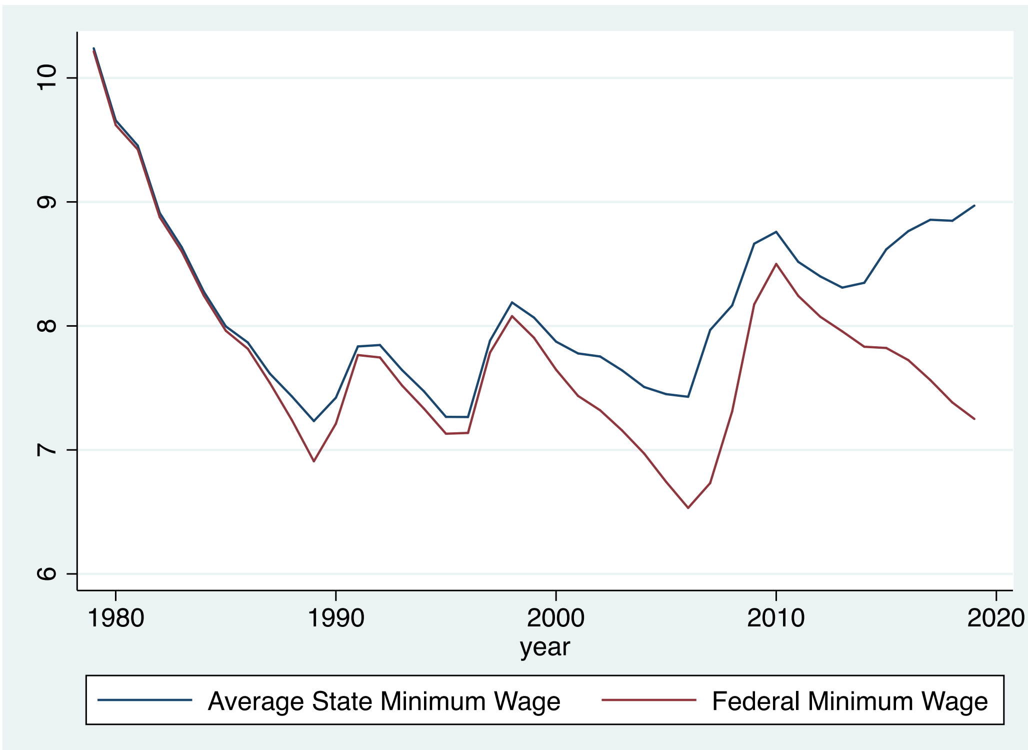
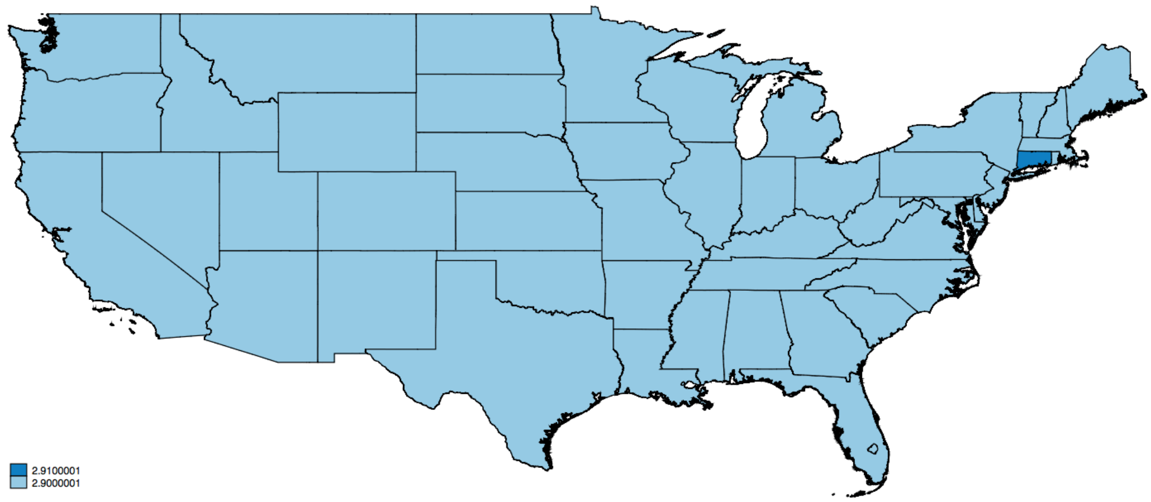
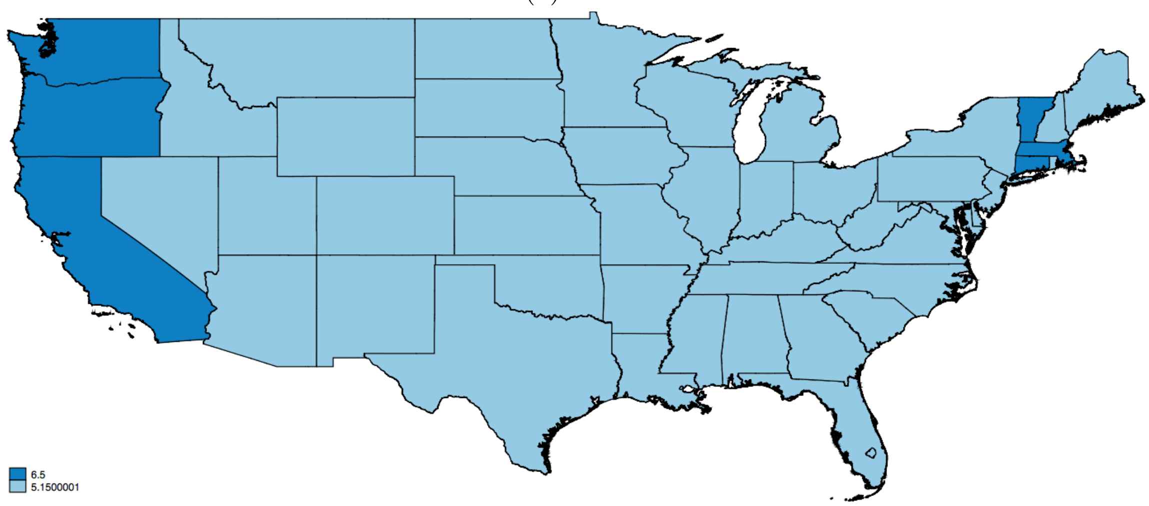


Figure 2.1: US Minimum Wage 1979-2019 in 2019 dollars

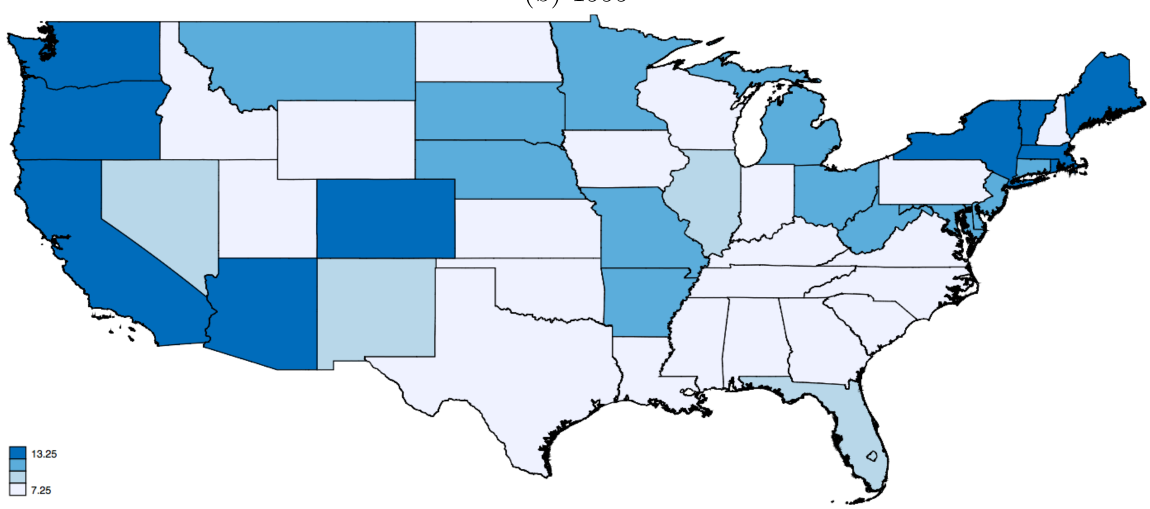
[6]This period had significant variation in state minimum wage by year — see Figures 2.1 and 2.2.
[7]See Table 2.5.



(a) 1979



(b) 1999



(c) 2019

Figure 2.2: Nominal State Minimum Wages in the US

When the effects of minimum wage are separated by industry, minimum wage workers in some industries saw their hours reduced, while the opposite was true in other industries. For example, men working minimum wage jobs in the accommodation and food services industry saw their weekly hours reduced by 1.18 hours per week when minimum wage was increased by one dollar. On the other hand, a one dollar increase in minimum wage caused men working in the manufacturing industry to work an additional 2.56 hours per week.[8]

It is not clear whether the industries with positive hour effects are highly concentrated. Figure 2.3 shows the Herfindahl-Hirschman Index (HHI) by industry in 2017.[9] Since industries are classified so broadly, none of the markets have a high HHI. Firms' direct competition is likely regional and the data available in this chapter are insufficient to determine which industry is concentrated. However, intuitively, it makes sense that accommodation and food services is competitive in local markets, while manufacturing is not. If so, a simple explanation for the reason a minimum wage increase had near-zero or positive employment and hour effects is that some industries are concentrated while others are competitive.

[8]See Tables 2.9 and 2.10.

[9]Industries are defined by North American Industry Classification System (NAICS) 2-digit codes. Information on HHI for Agriculture, Forestry, Fishing, and Hunting; Public Administration; and Retail Trade are not available.

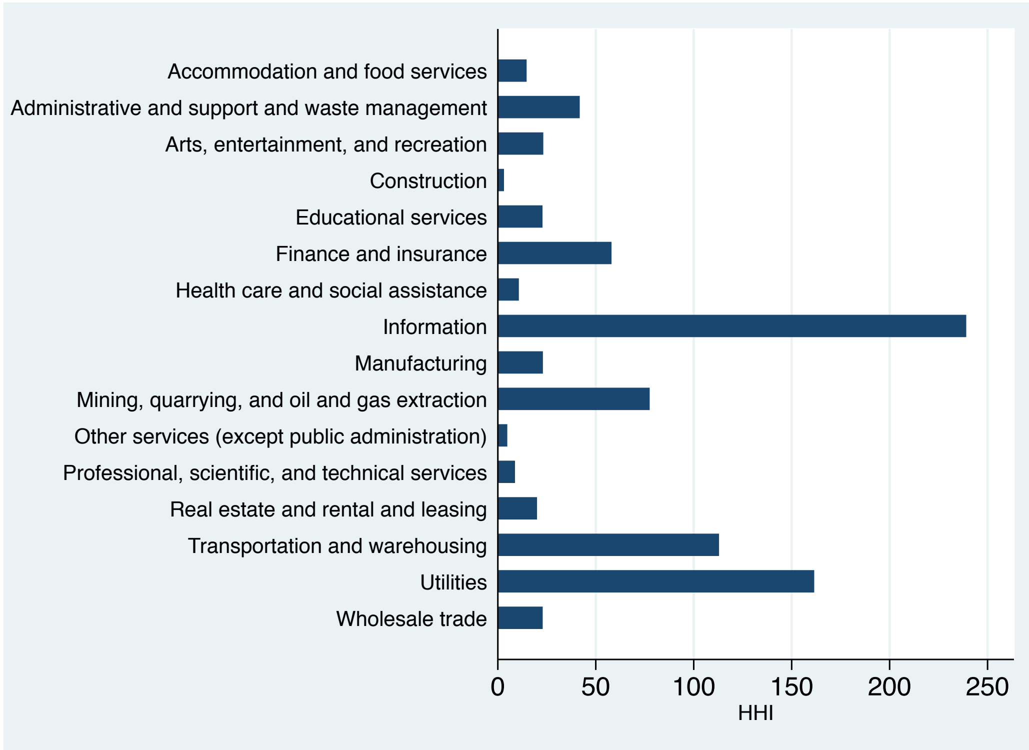


Figure 2.3: HHI by Industry, 2017

Additionally, it is possible minimum wage workers are able to work increased hours at their secondary job or find other part-time work. Since a large portion of minimum wage workers in the United States are paid by the hour and are part-time workers,[10] low-wage workers' increased desire to work more hours is met by work at a secondary job. In Table 2.1, after accounting for employment effects, the probability of a worker being part-time decreases with increased minimum wage. This is consistent with the two recent papers on Seattle's increased minimum wage (Jardim et al., 2017; Jardim et al., 2018) which suggests that workers were working additional hours at their secondary jobs.

[10]See Figures 2.9 and 2.10.

Table 2.1: Effect of Minimum Wage on Part-time Worker Status, 2003-2019

	(1) LPM	(2) Logit	(3) Probit	(4) Heckprobit	(5) LPM	(6) Logit	(7) Probit	(8) Heckprobit
Min. Wage	-0.02^{***}	-0.08^{***}	-0.05^{***}	-0.13^{***}	0.002	0.01	0.01	-0.03^{***}
	(0.01)	(0.03)	(0.02)	(0.01)	(0.004)	(0.02)	(0.01)	(0.01)
$\hat{\rho}$	-	-	-	0.08	-	-	-	0.21^{***}
	-	-	-	(0.05)	-	-		(0.04)
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Add. Controls	Y	Y	Y	Y	Y	Y	Y	Y
R^2/pseudo-R^2	0.22	0.18	0.18	-	0.13	0.1	0.1	-
N	40157	40157	40157	120184	64289	64289	64289	150552

Note: The columns show mean regression coeffiecents of regressing minimum wage on part-time worker status for workers earning at or below the minimum wage. Columns 1-4 are for men and Columns 5-8 are for women. LPM stands for linear probability model.$\hat{\rho}$ is an estimate of the employment effect in a Heckman probit selection model. N is the number of observations. Robust standard errors are in parentheses.

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Furthermore, it might not be market concentration that is causing the different effects of minimum wage on hours by industry, but rather the elasticity of demand for labor in those industries. If the labor costs in an industry are a relatively small percentage of overall expenses and the costs associated with reducing the labor force are high, then an increase in minimum wage is unlikely to result in a reduction to the labor force in that industry.

If the normative assumptions underpinning minimum wage policy are accepted, this chapter's results suggest minimum wage should be set by industry with competitive industries having lower minimum wages.

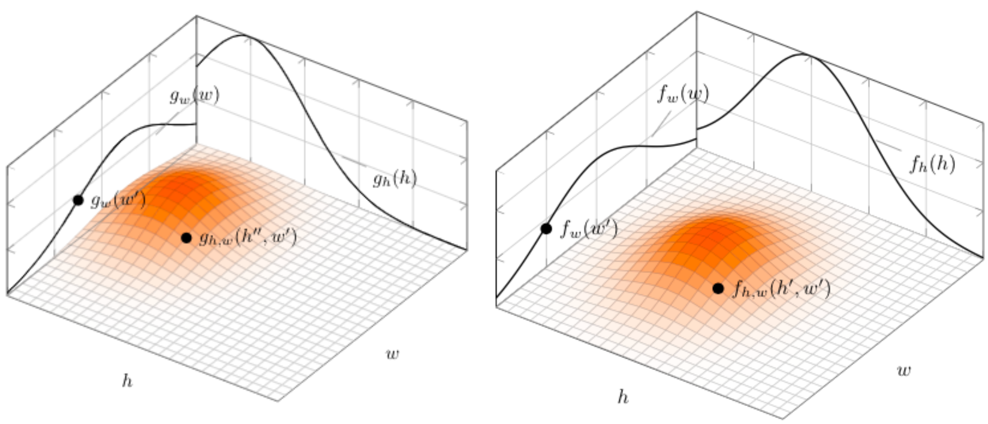
2.1.2 Wages and Hours as a Joint Outcome

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. It is possible for every individual's wage to be increased but for some individuals to experience a decrease in hours worked (assuming employment remains constant).[11] Perhaps employers do not adjust for increases in minimum wage by discharging workers or re-

[11] See Figure 2.4 for a graphical depiction.

ducing 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 are discussed in Chapter 1.



(a) With Minimum Wage Change (b) Without Minimum Wage Change

Figure 2.4: Comparing Joint Distributions

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 — the CDFs $G_w(w) < F_w(w)$ 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 (i.e. $g(w') < f(w')$), those same individuals are working fewer hours (i.e. $f_{h,w}(h', w') > g_{h,w}(h'', w')$).

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 a functional of the joint distribution of outcomes of interest with a policy and a functional of the counterfactual joint distribution without the policy.[12] Appendix C. provides more discussion on the importance of accounting for jointly determined outcomes.

Extending the method of Chernozhukov et al. (2013) for estimating counterfactual distributions using distribution regression, I employ 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 2.4. Additional robustness checks can be found in section 2.6.4.

I find that minimum wage increases caused both wages and hours worked to increase across the joint distribution for both men and women except at the very top of the wage distribution where minimum wage increases induced fewer individuals to work longer hours. This suggests minimum wage increases have large spillover effects and affect individuals working well above the minimum wage.

I find increased real Federal minimum wage from \$6.39 in 1989 to \$7.75 in 1992 (in 2019 dollars) caused the percentage of men earning around the median wage and

[12]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 the class of problems (P1) should also consider joint distributions.

working up to 36 hours per week to increase by 3% while it caused those working up to 40 hours to increase by 11% and those working up to 56 hours per week to increase by 17% (see Table 2.2). Surprisingly, the effect of increased minimum wage was much weaker for men in the bottom 10th percentile of wages. For those individuals, the minimum wage increase produced only a 5% increase in employees working up to 40 hours per week and a 6% increase in those working up to 56 hours per week. Men earning up to the 90th percentile and working 40 hours or more decreased by 3%. The effects on women, who tend to earn less and work fewer hours, was comparable. Increased real Federal minimum wage from \$6.53 in 2006 to \$8.07 in 2012, and decreased real minimum wage from \$7.63 in 1984 to \$6.39 in 1989 (in 2019 dollars) also had similar effects.[13]

Table 2.2: Minimum Wage Effect, Men 89-92

Wages/Hours	30	36	40	45	50	56
9.11	0.02*	0.03*	0.05*	0.05*	0.05*	0.06*
11.39	0.03*	0.04*	0.09*	0.09*	0.1*	0.11*
14.12	0.03*	0.03*	0.1*	0.11*	0.13*	0.13*
16.60	0.03*	0.03*	0.11*	0.12*	0.15*	0.15*
19.74	0.03*	0.03*	0.12*	0.14*	0.16*	0.17*
22.78	0.03*	0.03*	0.11*	0.12*	0.15*	0.16*
27.34	0.02*	0.02*	0.06*	0.07*	0.09*	0.1*
32.75	0.02	0.01	0	0.01	0.01	0.01
42.72	0.02	0	−0.03*	−0.03*	−0.03*	−0.03*

Note: Effect of having 1992's minimum wage on 1989's joint CDF of wages and hours for men.

$*$ = 95% confidence level using 50 bootstrap samples.

[13]Some of this could also be due to increased state minimum wages, which is accounted for in how this chapter defines the minimum wage variable. However, very few states had minimum wages different than the Federal minimum wage until 1990.

These results 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 increase, 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 work fewer hours as her wage increases.[14] In this way, minimum wage's heterogeneous effects are simple — consider Figure 2.6. Excluding general equilibrium effects, if the markets for hours of minimum wage, medium-wage, and high-wage labor are competitive, all workers should experience an increase in wages. However, medium-wage workers should see an increase in their hours worked, while minimum wage and high-wage workers should see a decrease.

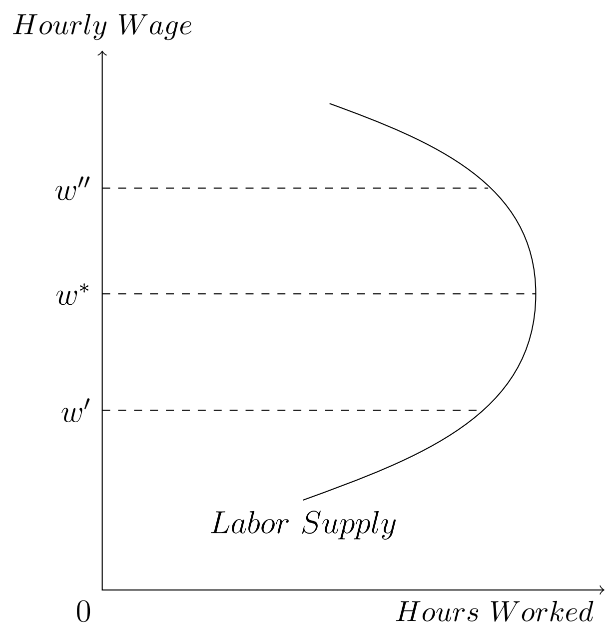
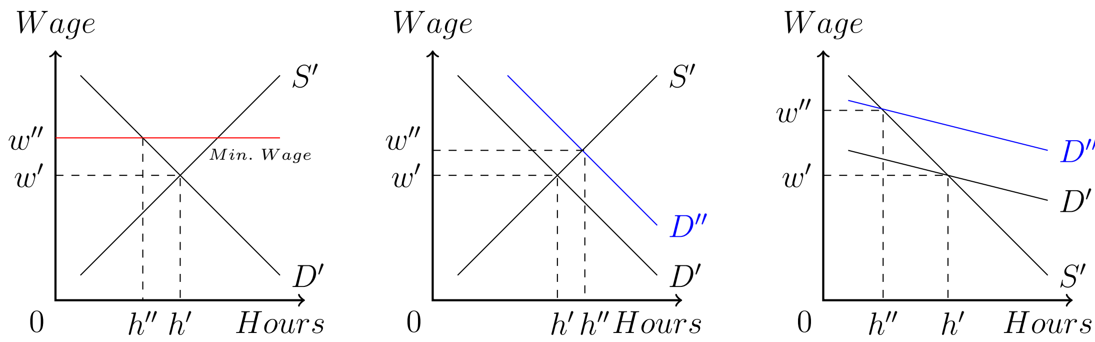


Figure 2.5: Backwards bending wage curve.

As hourly wage increases from w' to w^* to w'', a worker's hours increase and then decrease.

[14]See Figure 2.5.



(a) Min. Wage Workers. (b) Medium-Wage Workers. (c) High-Wage Workers.

Figure 2.6: Supply and Demand Illustration

Let S' and S'' represent a worker's supply of hours before and after a minimum wage increase, respectively. Let D' and D'' represent an employer's demand for hours of work before and after a minimum wage increase, respectively. In panel (a), when a minimum wage increase is binding, wages increase from w' to w'', however hours worked decreases from h' to h''. In panel (b), an increase in the price of minimum wage labor increases the demand for medium-wage laborers, thereby increasing both their wages and hours worked (assuming the two types of labor are substitutes). In panel (c), similarly, demand increases for high-wage workers, however since their supply is downward-sloping, wages increase but hours decrease.

This model of the wage and hour distributional effects due to a minimum wage increase might be an overly simplistic. Indeed, the decrease in minimum wage workers' hours does not seem present in the cases considered. Theoretical reasons why minimum wage workers might not see their hours reduced are the same as the reasons why they might not suffer employment effects laid out in Section 2.2.1.

2.1.3 The Remainder of the Chapter

Section 2.2 reviews the literature on the predictions and empirical effects of minimum wage policy on low-wage workers as well as its distributional effects. Section 2.3 describes the selection model used to account for employment effects. Section 2.4

presents the new methodology this chapter introduces on estimating counterfactual joint distributions. Section 3.3 discusses the data used in this chapter. Section 3.5 presents the results. Section 3.6 concludes.

2.2 Literature Review

2.2.1 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, 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.

2.2.2 Empirical Results on Low-wage Workers

Traditional empirical work using observational data found an increase in minimum wage led to decreases in employment.[15] 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

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

the reduction in hours worked was not considered.[16] 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.

2.2.3 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 ben-

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

efit 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.[17]

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.

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 I do.

[17] 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.

2.3 Employment Selection Model

Consider Heckman (1979) selection model (Type II Tobit model) for $i = 1, .., n$ observations

$$y_i = X_i'\beta + \epsilon_i,$$
$$D_i = \mathbb{1}\{Z_i'\gamma + \nu_i > 0\},$$

with $\begin{pmatrix} \epsilon_i \\ \nu_i \end{pmatrix} \Big| X_i, Z_i \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right)$, so that the variance of ϵ_i is σ^2 and the variance of ν_i is normalized to 1. Hence, for a random sample of the population $(X_i, Z_i, D_i, D_i y_i)$ with the outcome of interest y_i only observed when $D_i = 1$

$$\begin{aligned} E[\epsilon_i|X_i, \text{sample selection rule}] &= E[\epsilon_i|X_i, D_i = 1] \\ &= E[\epsilon_i|X_i, Z_i'\gamma + \nu_i > 0] \\ &= E[\epsilon_i|X_i, \nu_i > -Z_i'\gamma]. \end{aligned}$$

Therefore,

$$E[y_i|X_i, \text{sample selection rule}] = X_i'\beta + E[\epsilon_i|\nu_i > -Z_i'\gamma].$$

Since ϵ_i and ν_i are assumed to be bivariate normal,

$$E[\epsilon_i|\nu_i > -Z_i'\gamma] = \sigma\lambda(Z_i'\gamma),$$

where $\lambda(Z_i'\gamma)$ is the inverse Mills ratio $\frac{\phi(Z_i'\gamma)}{\Phi(Z_i'\gamma)}$, $\phi(\cdot)$ is the PDF of a normal distribution, and $\Phi(\cdot)$ is the CDF of a normal distribution.[18] In that way, the inverse Mills ratio

[18] See Heckman (1979) for details.

is a monotone decreasing function of the probability that the observation is selected into the sample and including it in the regression of X_i on y_i corrects for the omitted variable bias due to selection (Heckman's two-step procedure).

2.4 Joint Distribution Model

2.4.1 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

$$F_Y(y) = \int_{\mathcal{X}} F_{Y|X}(y|x) dF_X(x), \tag{2.1}$$

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 are 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 (2.1) can be rewritten as

$$F_{Y\langle t,v \rangle}(y) = \int_{\mathcal{X}_v} F_{Y_t|X_t}(y|x) dF_{X_v}(x), \tag{2.2}$$

$t, v \in \{0, 1\}$ 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 made.

To further demonstrate the usefulness of the counterfactual distributions, consider the case in which X is partitioned into two random variables, $X = (X_a, X_b)$.[19] Then by the law of iterated probability

$$F_Y(y) = \int_{\mathcal{X}_b} \int_{\mathcal{X}_a} F_{Y|X_a,X_b}(y|x_a, x_b) dF_{X_a|X_b}(x_a|x_b) dF_{X_b}(x_b). \tag{2.3}$$

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

$$F_{Y\langle t|s,v\rangle}(y) = \int_{\mathcal{X}_{b,s}} \int_{\mathcal{X}_{a,v}} F_{Y|X_{a,t},X_{b,t}}(y|x_a, x_b) dF_{X_{a,s}|X_{b,s}}(x_a|x_b) dF_{X_{b,v}}(x_b), \tag{2.4}$$

$t, s, v \in \{0, 1\}$. 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.[20][21]

There are several proposed methods for estimating 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

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

[20]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 adhere to the interpretation of a change from $X_{a,1}$ to $X_{a,0}$ on the distribution of Y_0).

[21]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 similarly decomposed.

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.[22]

Suppose now that there are two outcomes of interest, Y^1 and Y^2.[23] 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

$$F_{Y^1,Y^2}(y_1, y_2) = C(F_{Y^1}(y_1), F_{Y^2}(y_2)). \tag{2.5}$$

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

2.4.2 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.[24] 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

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

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

[24] 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.

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 $(j, k) \in \mathcal{JK}$.

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 2.4.2. 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.

Theorem 2.4.2 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.

2.4.3 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 $\mathbf{X}_v$ denote 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

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

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:

$$F_{Y^i\langle t|s,v\rangle}(y) := \int_{\mathcal{C}_v} \int_{\mathcal{M}_s} F_{Y_t^i|m_t,\mathbf{c}_t}(y|m, \mathbf{c}) dF_{m_s|\mathbf{c}_s}(m|\mathbf{c}) dF_{\mathbf{c}_v}(\mathbf{c}), \tag{2.6}$$

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 $t = v = s$, (2.6) becomes the unconditional distribution $F_{Y^i\langle t|t,t\rangle} = F_{Y_t^i}$—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

$$F_{\mathbf{Y}\langle t|s,v\rangle}(y_1, y_2) = C(F_{Y^1\langle t|s,v\rangle}(y_1), F_{Y^2\langle t|s,v\rangle}(y_2)). \tag{2.7}$$

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

2.4.4 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 $F_{Y_1^i|m_1,\mathbf{c}_1}(y|m,\mathbf{c})$ by distribution regression (discussed below) to obtain $\hat{F}_{Y_1^i|m_1,\mathbf{c}_1}(y|m,\mathbf{c})$

4. Obtain $\hat{F}_{Y^i\langle 1|0,1\rangle}(y) = \int_{\mathcal{C}_1}\int_{\mathcal{M}_0} \hat{F}_{Y_t^i|m_1,\mathbf{c}_1}(y|m,\mathbf{c})d\hat{F}_{m_0|\mathbf{c}_0}(m|\mathbf{c})d\hat{F}_{\mathbf{c}_1}(\mathbf{c})$

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 or 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

$$\hat{F}_{Y_1^i|\mathbf{X}_1}(y|\mathbf{X}) = \Lambda(P(\mathbf{X})\hat{\beta}(y)) \quad y \in \mathcal{Y}_i,$$

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

$$\hat{\beta}(y) = \arg\max_b \sum_{j=1}^{n}\{\mathbb{1}\{Y_j^i \leq y\}\ln(P(\mathbf{X}_j)'b) + \mathbb{1}\{Y_j^i \geq y\}\ln((1-P(\mathbf{X}_j))'b)\}.$$

I use a logit link function.

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

$$\hat{F}_{\mathbf{Y}\langle t|s,v\rangle}(y_1,y_2) = \hat{C}(\hat{F}_{Y^1\langle t|s,v\rangle}(y_1), \hat{F}_{Y^2\langle t|s,v\rangle}(y_2)), \tag{2.8}$$

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

This chapter 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 $(\hat{U}_i^1, \hat{U}_i^2) = (\hat{F}_{Y^1\langle t|s,v\rangle}(Y_i^1), \hat{F}_{Y^2\langle t|s,v\rangle}($ 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.

2.4.5 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 $\hat{s}^2(y_1, y_2) = B^{-1}\sum_{j=1}^{B}\left(\left(\hat{F}^{*(j)}_{\mathbf{Y}\langle 1|0,1\rangle}(y_1, y_2)\right.\right.$

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 $[L(y_1, y_2), U(y_1, y_2)] = \left[\hat{\Delta}_{y_1,y_2} \pm c(1-\alpha)\hat{s}\right.$

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.

2.5 Data

2.5.1 Data Cleaning

I use the Current Population Survey Merged Outgoing Rotation Group (CPS MORG) for the years 1979 to 2019 and was obtained through the NBER website.[25] CPS MORG is different than the March CPS[26] 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 for that week.[27]

I clean the data using the specifications of Autor et al. (2016) and state minimum wage data was obtained through Vaghul and Zipperer (2016).[28] Wages are in 2019 dollars and Consumer Price Indexes were obtained from Federal Reserve Bank of Minneapolis.[29] 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).[30]

However, unlike Autor et al. (2016), I use 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

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

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

[27] See Lemieux (2006).

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

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

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

of who is supplying this hour". While this approach may have originally deviated from using weekly earnings as the dependent variable to account for labor market participation decisions, it makes welfare comparisons more difficult since there is no way of knowing whether individuals actually have higher wage and hours worked bundles (incomes). Additionally, it treats hours as exogenously given. My 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, and Standard Metropolitan Statistical Area (SMSA).

Two exclusion restrictions are needed for identification in the selection models. As is common in the literature, I use number of children under age five as an exclusion restriction that potentially affects labor force participation but is uncorrelated with hourly wage. Education variables are used as the exclusion restriction that potentially affects labor force participation but is uncorrelated with hours worked.

HHI data by industry in 2017 was obtained through the US Census Bureau.[31]

2.5.2 Data Visualization

Figure 2.1 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 rates across the country. Figure 2.2 shows the states' minimum wages in 1979, 1999, and 2019. Table 2.3 shows the timeline of nominal Federal minimum wage increases.

[31] Available at https://data.census.gov/.

Table 2.3: 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

Figure 2.7 shows the percentage of workers earning at or below the minimum wage. The percentage of men earning at or below the minimum wage is lower than that of women. Figure 2.8 shows the evolution of hourly wage and weekly hours worked by five quantile ranges for hourly wage. The gap between the top and bottom quantile ranges of hourly wages in the pooled sample widens over time but seems to be mainly driven by the gap between the top and bottom quantile ranges of hourly wages for women widening. Additionally, Figure 2.8 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 seem to be more volatile for workers at the bottom of the wage distribution.

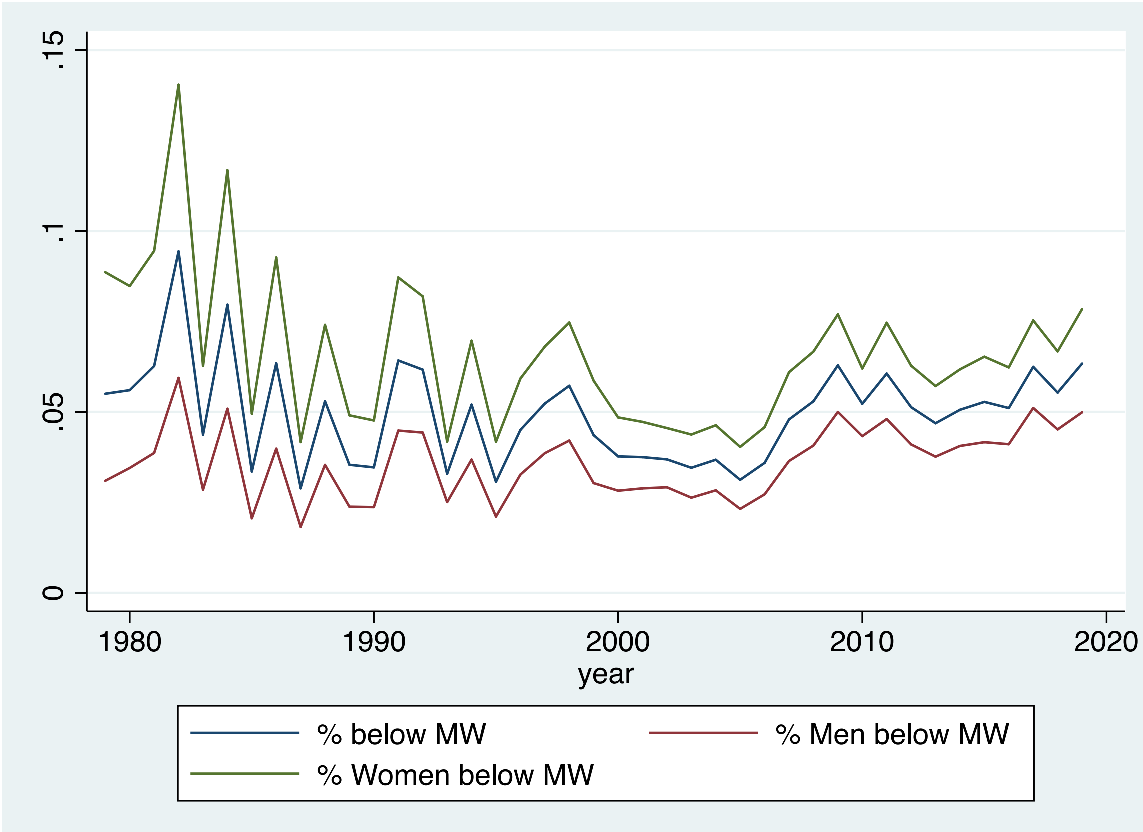
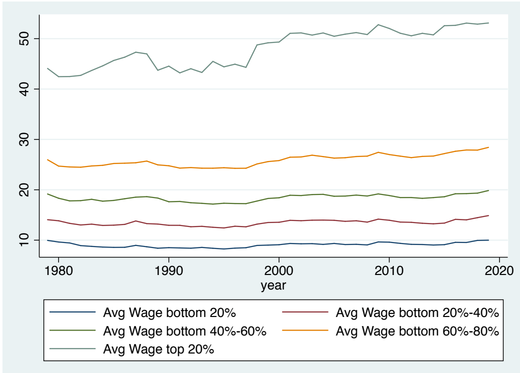
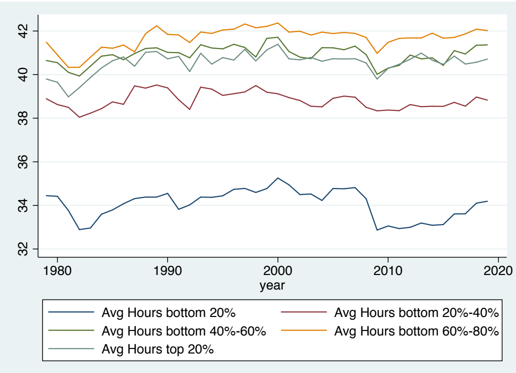


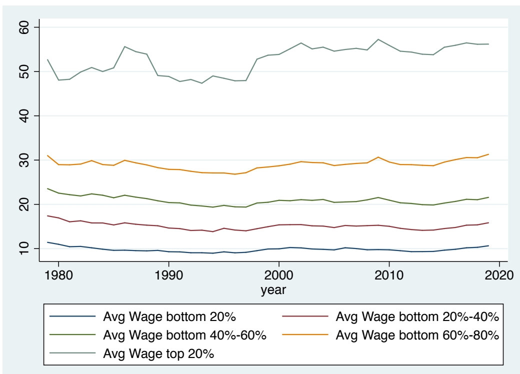
Figure 2.7: Percent of Worker earning at or Below Minimum Wage 1979-2019



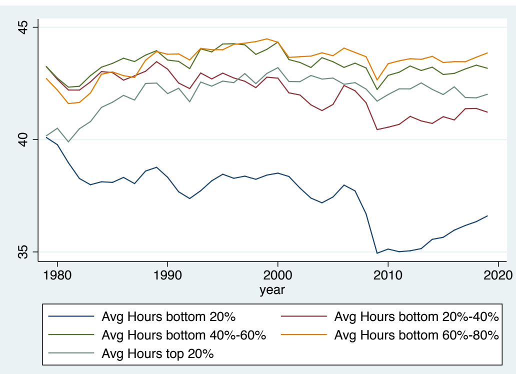
(a) US Wages in 2019 dollars (Pooled)



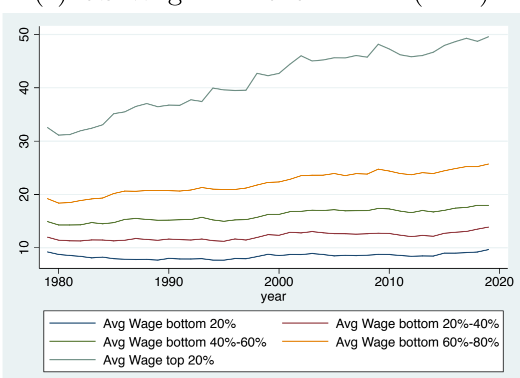
(b) US Weekly Hours Worked (Pooled)



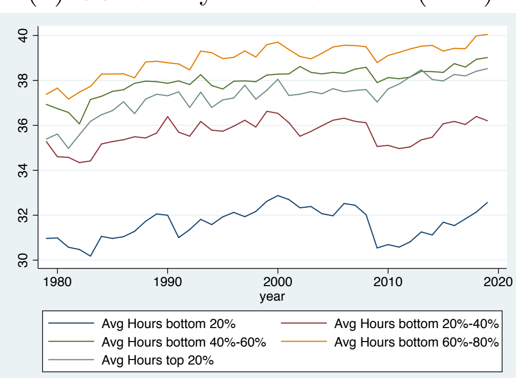
(c) US Wages in 2019 dollars (Men)



(d) US Weekly Hours Worked (Men)



(e) US Wages in 2019 dollars (Women)



(f) US Weekly Hours Worked (Women)

Figure 2.8: Average Wages and Hours of 5 Quantiles Ranges of Wages 1979-2019

Figure 2.9 shows the percentage of minimum wage workers who are part-time. Around half of minimum wage workers are part-time. Figure 2.10 shows the percentage of minimum wage workers who are paid by the hours. After 2000, between 60 and 70 percent of minimum wage workers were paid by the hour.

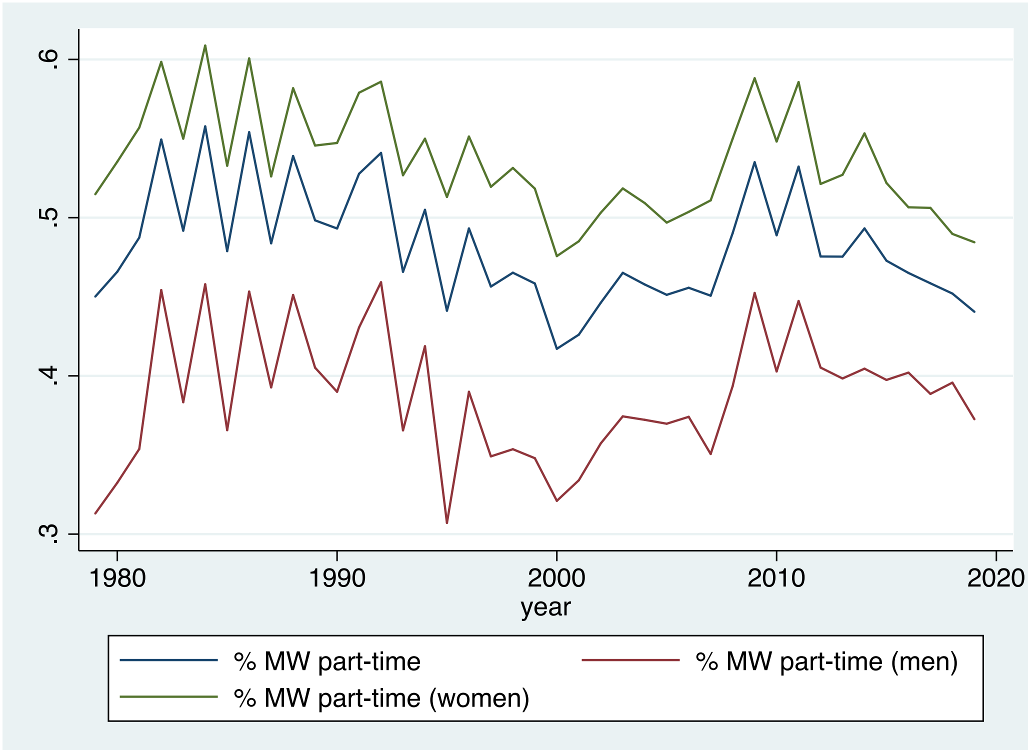


Figure 2.9: Percent of Minimum Wage Workers who are Part-time

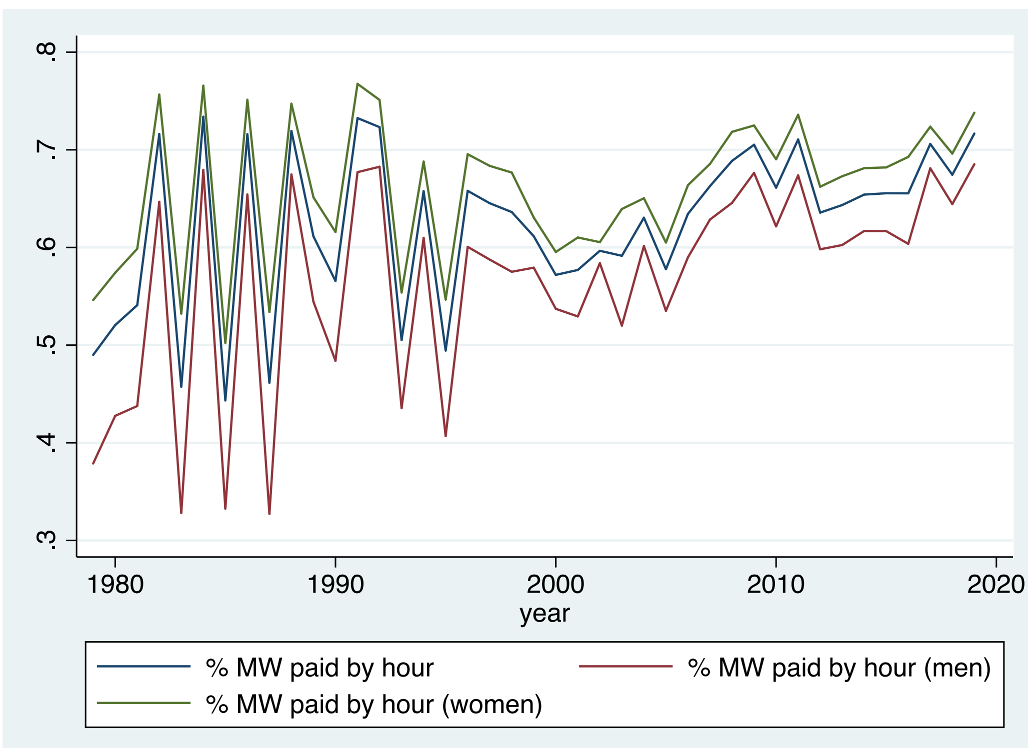
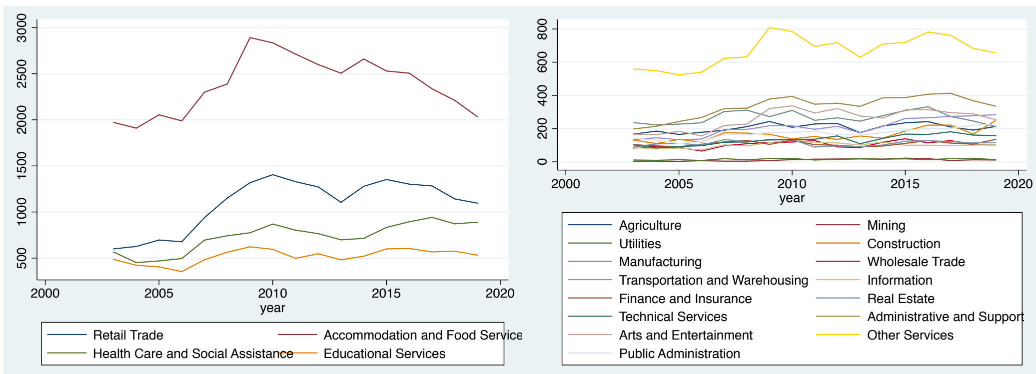


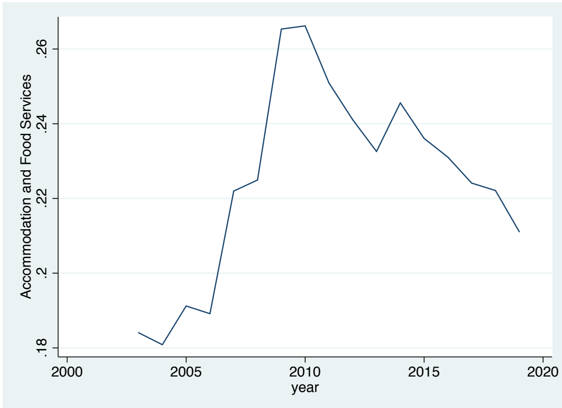
Figure 2.10: Percent of Minimum Wage Workers paid by the Hour

Figure 2.3 shows the HHI by industry defined by NAICS 2-digit industry codes. In formation and Utilities are more concentrated industries, while Construction and Health Care and Social Assistance are less concentrated. Figure 2.11 shows the number of minimum wage workers per industry in the sample, and Figure 2.12 shows the percentage of workers in the several industries. Figure 2.13 show the unemployment rate by industry.

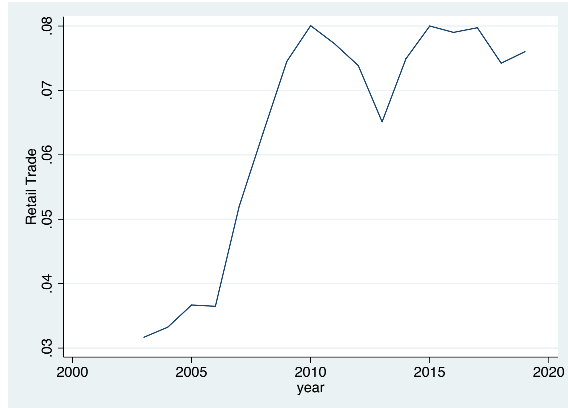


(a) (b)

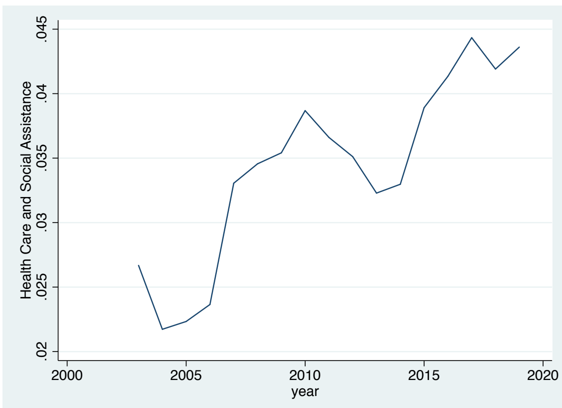
Figure 2.11: Number of Workers Paid Min. Wage by Industry, 2003-2019



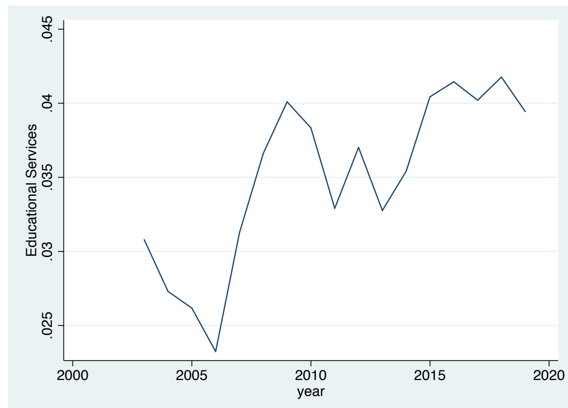
(a) Accommodation and Food Services



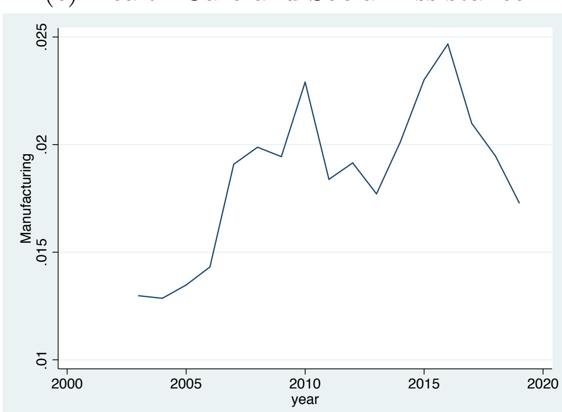
(b) Retail Trade



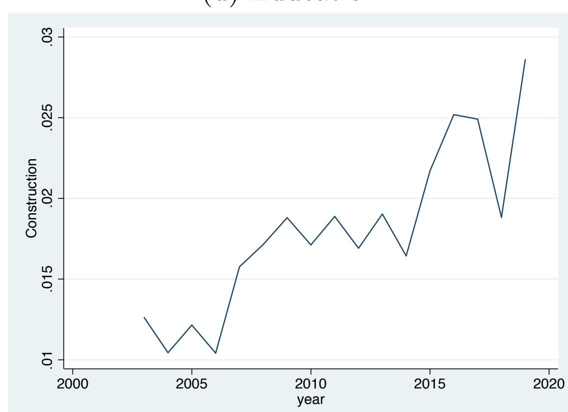
(c) Health Care and Social Assistance



(d) Education



(e) Manufacturing



(f) Construction

Figure 2.12: Percent of Industry Paid Min. Wage, 2003-2019

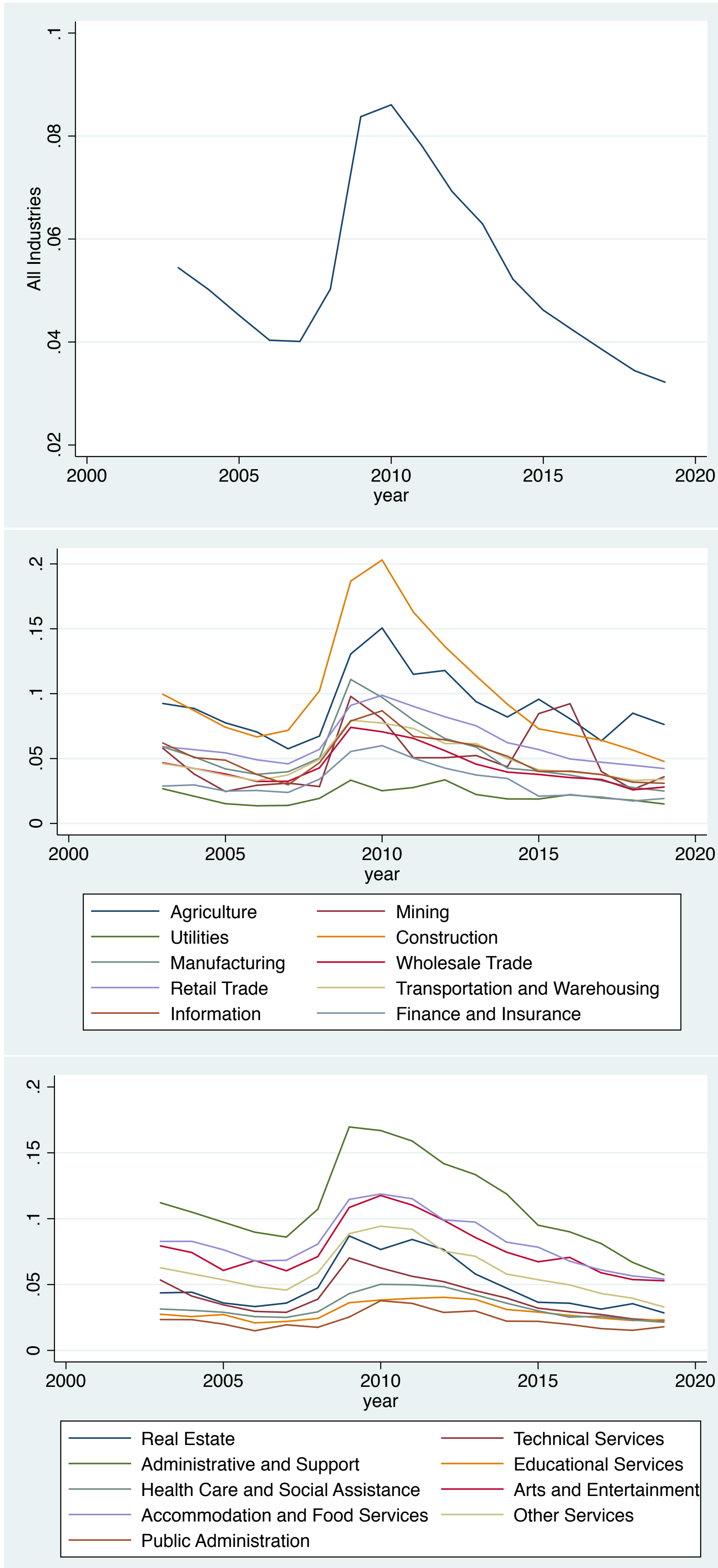


Figure 2.13: Unemployment Rate, 2003-2019

2.6 Results

2.6.1 Effects on Minimum Wage Workers

In this section, the minimum wage variable (MW) is defined as the minimum wage in the region in which an individual works. Hence, MW varies by state and time, and therefore the years 2003 through 2019 with more variation in state minimum wage are used.

The fixed effects models for hours and log(wages) are

$$\text{hours}_{its} = \alpha_s + \gamma_t + \delta^H X_i^H + \beta^H \text{MW}_{ts} + \epsilon_{its}^H, \tag{2.9}$$

and

$$\log(\text{wages})_{its} = \alpha_s + \gamma_t + \delta^W X_i^W + \beta^W S\text{MW}_{ts} + \epsilon_{its}^W, \tag{2.10}$$

where α_s and γ_t are state and year fixed effects, X_i^H and X_i^W are vectors of control variables for hours worked and wages, and ϵ_{its}^H and ϵ_{its}^W are error terms.

Equations 2.9 and 2.10 can be altered to have a quadratic MW variable as

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

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 $-\frac{\beta_1}{2\beta_2}$.

Table 2.4 shows the fixed effects models of the effects of minimum wage (and minimum wage squared) on workers earning at or below the minimum wage on hours or log(wages). Table 2.5 includes a Heckman correction. Table 2.6 shows the effect of minimum wages on employment status for workers earning at or below the minimum wage.

Table 2.4: Minimum Wage Workers, 2003-2019

	Hours				log(Wages)	
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\beta}_1$	1.04**	-2.07	-0.18	-4.62^{***}	0.2***	0.19***
	(0.43)	(5.79)	(0.21)	(1.72)	(0.04)	(0.02)
$\hat{\beta}_2$	-	0.18	-	0.26***	-	-
	-	(0.32)	-	(0.09)	-	-
$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$	-	5.90	-	8.88	-	-
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y
R^2	0.58	0.58	0.62	0.62	0.09	0.07
N	40157	40157	64289	64289	51502	77464

Note: The columns show mean regression coefficients of regressing minimum wage (and minimum wage squared) on hours or log(wage) with state and year fixed effects for worker earning at or below the minimum wage. Columns 1, 2, and 5 are regressions coefficients when for men and the remaining columns are regressions coefficients for women. N is the number of observations in per state-year group. Robust standard errors are in parentheses.

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 2.5: Minimum Wage Workers with Heckman Correction, 2003-2019

	Hours				log(Wages)	
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\beta}_1$	1.43^{***}	4.13^{***}	0.64^{***}	2.05^{**}	0.09^{***}	0.09^{***}
	(0.28)	(1.47)	(0.19)	(0.94)	(0.02)	(0.01)
$\hat{\beta}_2$	-	-0.15^{*}	-	-0.08	-	-
	-	(0.08)	-	(0.05)	-	-
$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$	-	13.77	-	12.81	-	-
Inverse Mills Ratio	-4.85^{**}	-4.6^{**}	-4.53^{***}	-4.4^{***}	-0.59^{***}	-0.56^{***}
	(1.97)	(1.97)	(1.46)	(1.46)	(0.09)	(0.06)
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y
N	120184	120184	150552	150552	104336	126276

Note: The columns show mean regression coefficients of regressing minimum wage (and minimum wage squared) on hours or log(wage) with state and year fixed effects for worker earning at or below the minimum wage with a two-step Heckman correction. Columns 1, 2, and 5 are regressions coefficients for men and the remaining columns are regressions coefficients for women. N is the number of observations. Standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 2.6: Minimum Wage Workers LFP, 2003-2019

	(1)	(2)	(3)	(4)	(5)	(6)
	LPM	Logit	Probit	LPM	Logit	Probit
Min. Wage	0.077***	0.376***	0.225***	0.071***	0.367***	0.216***
	(0.003)	(0.016)	(0.009)	(0.003)	(0.014)	(0.008)
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y
R^2	0.18	0.14	0.14	0.14	0.11	0.11
N	120184	120184	120184	150552	150552	150552

Note: The columns show mean regression coefficients of regressing minimum wage on labor force participation for workers earning at or below the minimum wage. Columns 1-3 are for men and Columns 4-6 are for women. LPM stands for linear probability model. N is the number of observations. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

2.6.2 Effects on Minimum Wage Workers by industry

Tables 2.7 and 2.8 show the effects of minimum wage on minimum wage workers with state and year fixed effects by industry without correcting for employment effects. Minimum wage has no statistically significant effects on the hours of minimum wage workers but has a statistically significant positive effect on wages.

Tables 2.7 and 2.8 show the effects of minimum wage on hours or log(wages) of workers earning at or below the minimum wage without accounting for employment effects. Tables 2.9 and 2.10 show the same thing, but with a Heckman correction. Tables 2.11 and 2.12 show the effect of minimum wages on employment status for workers earning at or below the minimum wage in different industries.

2.6.3 Joint Distribution Results

Table 2.2 shows the effect of the 1992 increase in minimum wage on the distribution of hours and wages for men, when compared to the 1989 minimum wage ($\hat{F}_{\mathbf{Y}_{\langle 1|1,1\rangle}} -$

Table 2.7: Effects of Min. Wage on Min. Wage Workers by Industry, 2003-2019

	Accommodation and Food Services									Retail Trade		
	Hours				log(Wages)			Hours			log(Wages)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\hat{\beta}_1$	0.17	2.39	0.5	0.29	0.14^{***}	0.15^{***}	2.26^{***}	-15.01	0.26	14.33	0.08^{**}	0.14^{***}
	(0.56)	(7.1)	(0.45)	(3.8)	(0.05)	(0.03)	(0.72)	(19.8)	(0.69)	(11.84)	(0.04)	(0.04)
$\hat{\beta}_2$	-	-0.12	-	0.01	-	-	-	0.95	-	-0.79	-	-
	-	(0.38)	-	(0.21)	-	-	-	(1.07)	-	(0.65)	-	-
$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$	-	9.63	-	-11.42	-	-	-	7.86	-	9.08	-	-
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.61	0.61	0.61	0.61	0.05	0.08	0.62	0.62	0.65	0.65	0.05	0.04
N	10354	10354	20536	20536	10354	20536	6032	6032	9005	9005	6032	9005
	Health Care and Social Assistance									Education		
	Hours				log(Wages)			Hours			log(Wages)	
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
$\hat{\beta}_1$	2.98	99.17^{**}	-0.58	-31.03^{***}	-0.11^{**}	0.32^{***}	1.74	-10.42	0.03	-8.44	-0.01	-0.03
	(2.91)	(49.04)	(1.02)	(9.42)	(0.11)	(0.06)	(1.53)	(14.59)	(1.04)	(6.21)	(0.09)	(0.1)
$\hat{\beta}_2$	-	-5.26^{**}	-	1.74^{***}	-	-	-	0.66	-	0.5	-	-
	-	(2.66)	-	(0.52)	-	-	-	(0.77)	-	(0.36)	-	-
$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$	-	9.42	-	8.91	-	-	-	7.85	-	8.41	-	-
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.57	0.57	0.55	0.55	0.11	0.07	0.67	0.67	0.65	0.65	0.1	0.08
N	1762	1762	8798	8798	1762	8798	1990	1990	5302	5302	1990	5302

Note: The columns show mean regression coefficients of regressing minimum wage (and minimum wage squared) on hours or log(wage) for workers earning at or below the minimum wage in different industries. Columns 1, 2, 5, 7, 8, 11, 13, 14, 17, 19, 20, and 23 are regressions coefficients when only male observations are used and the remaining columns are are regressions coefficients when only female observations are used. N is the number of observations in per state-year group. Robust standard errors are in parentheses. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 2.8: Effects of Min. Wage on Min. Wage Workers by Industry, 2003-2019

	Manufacturing									Construction		
		Hours			log(Wages)			Hours				log(Wages)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\hat{\beta}_1$	0.7	-8.05	-0.14	-2.53	0.59**	0.42***	3.34	89.22***	4.38	-22.25	0.03	1.14*
	(0.79)	(20.2)	(0.63)	(10)	(0.26)	(0.13)	(4.78)	(35.36)	(6.17)	(28.24)	(0.16)	(0.66)
$\hat{\beta}_2$		0.49		0.14*				−4.89***		1.81		
	-	(1.11)	-	(0.55)	-	-	-	(1.95)	-	(1.9)	-	-
$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$	-	8.18	-	9.35	-	-	-	9.12	-	6.13	-	-
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.54	0.54	0.62	0.62	0.11	0.15	0.55	0.55	0.83	0.83	0.07	0.37
N	1951	1951	1904	1904	1951	1904	1908	1908	377	377	1908	377

Note: The columns show mean regression coefficients of regressing minimum wage (and minimum wage squared) on hours or log(wage) for workers earning at or below the minimum wage in different industries. Columns 1, 2, 5, 7, 8, and 11 are regressions coefficients when only male observations are used and the remaining columns are are regressions coefficients when only female observations are used. N is the number of observations in per state-year group. Robust standard errors are in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.9: Effects of Min. Wage on Min. Wage Workers by Industry with Employment Effects, 2003-2019

	Accommodation and Food Services						Retail Trade					
	Hours				log(Wages)		Hours				log(Wages)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\hat{\beta}_1$	-1.18^{*}	-5.7^{**}	-0.08	0.06	0.18^{***}	0.15^{***}	1.57	2.18	1.36	1.84	-0.06	0.05
	(0.64)	(2.79)	(0.32)	(1.67)	(0.02)	(0.01)	(1.23)	(4.29)	(0.88)	(3.07)	(0.06)	(0.04)
$\hat{\beta}_2$	-	0.24^{*}	-	-0.01	-	-	-	-0.03	-	-0.02	-	-
	-	(0.14)	-	(0.09)	-	-	-	(0.22)	-	(0.15)	-	-
$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$	-	11.95	-	4.07	-	-	-	33.89	-	38.27	-	-
Inv. Mills Ratio	-17.44^{***}	-18.64^{***}	-2.37	-2.35	0.25	0.43^{***}	-6.77	-6.61	2.23	2.43	-1.04^{***}	-0.53^{***}
	(5.27)	(5.34)	(3.38)	(3.39)	(0.16)	(0.12)	(6.98)	(7.06)	(4.13)	(4.3)	(0.27)	(0.14)
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Add. Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	15920	15920	29505	29505	15571	28982	14185	14185	19441	19441	13110	18327
N-Censored	5023	5023	7748	7748	5023	7748	6981	6981	9061	9061	6981	9061

	Health Care and Social Assistance						Education					
	Hours				log(Wages)		Hours				log(Wages)	
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
$\hat{\beta}_1$	3.09^{**}	-6.8	2.9^{***}	2.2	0.27^{**}	-0.03	2.33^{***}	13.26^{***}	0.24	5.12^{**}	0.12	-0.06
	(1.6)	(7.86)	(0.95)	(3.16)	(0.11)	(0.06)	(0.67)	(4.25)	(0.41)	(2.14)	(0.11)	(0.08)
$\hat{\beta}_2$	-	0.54	-	0.04	-	-	-	-0.62^{***}	-	-0.28^{**}	-	-
	-	(0.42)	-	(0.16)	-	-	-	(0.24)	-	(0.12)	-	-
$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$	-	6.25	-	-29.83	-	-	-	10.67	-	9.15	-	-
Inv. Mills Ratio	11.99	11.95	7.3	7.07	0.13	-1.14^{***}	-5.79	-7.81	-20.07^{***}	-20.59^{***}	-0.74	-1.19^{**}
	(10.47)	(10.47)	(6)	(6.08)	(0.49)	(0.25)	(7.64)	(7.69)	(5.09)	(5.15)	(0.54)	(0.47)
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Add. Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	4484	4484	21113	21113	3643	18437	7508	7508	22438	22438	4074	10656
N-Censored	1853	1853	9453	9453	1853	9453	2017	2017	5120	5120	2017	5120

Note: The columns show mean regression coefficient of regressing minimum wage (and minimum wage squared) on hours or log(wage) on workers earning at or below the minimum wage with Heckman correction for employment effects by industry. Columns 1, 2, 5, 7, 8, 11, 13, 14, 17, 19, 20, and 23 for men and the remaining columns are for women. N is the number of observations, and $N - N$-Censored is the number of uncensored observations. Standard errors are in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.10: Effects of Min. Wage on Min. Wage Workers by Industry with Employment Effects, 2003-2019

	Manufacturing							Construction				
	Hours				log(Wages)			Hours			log(Wages)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\hat{\beta}_1$	2.56*	8.19	4.06***	17.97***	0.03	0.13	1.88	25.69***	3.51	-4.69	0.07	0.4
	(1.37)	(6.79)	(1.2)	(6.2)	(0.2)	(0.11)	(1.99)	(7.4)	(2.54)	(12.84)	(0.14)	(0.25)
$\hat{\beta}_2$	-	-0.31^{**}	-	-0.77^{**}	-	-	-	-1.26^{***}	-	0.46	-	-
	-	(0.37)	-	(0.34)	-	-	-	(0.38)	-	(0.71)	-	
$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$	-	13.13	-	11.71	-	-	-	10.21	-	5.09	-	-
Inverse Mills Ratio	-3.87	-3.27	5.84	7.85	-1.59^{*}	-0.8^{*}	-4.42	0.22	9.09	8.95	-0.06	0.31
	(10.05)	(10.07)	(8.87)	(8.96)	(0.87)	(0.46)	(9.6)	(9.63)	(11.61)	(11.6)	(0.59)	(0.84)
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	10834	10834	6937	6937	9367	6217	15690	15690	1847	1847	14791	1590
N-Censored	7388	7388	4278	4278	7388	4278	12858	12858	1208	1208	12858	1208

Note: The columns show mean regression coefficient of regressing minimum wage (and minimum wage squared) on hours or log(wage) on workers earning at or below the minimum wage with Heckman correction for employment effects by industry. Columns 1, 2, 5, 7, 8, 11, 13, 14, 17, 19, 20, and 23 for men and the remaining columns are for women. N is the number of observations, and $N - N$-Censored is the number of uncensored observations. Standard errors are in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.11: Min. Wage Workers LFP by Industry, 2003-2019

	Accommodation and Food Services								Retail Trade			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	LPM	Logit	Probit	LPM	Logit	Probit	LPM	Logit	Probit	LPM	Logit	Probit
Min Wage	0.07***	0.4***	0.24***	0.05***	0.32***	0.18***	0.12***	0.56***	0.34***	0.12***	0.58***	0.35***
	(0.01)	(0.05)	(0.03)	(0.01)	(0.04)	(0.02)	(0.01)	(0.05)	(0.03)	(0.01)	(0.04)	(0.02)
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2/pseudo-R^2	0.09	0.07	0.07	0.1	0.08	0.08	0.13	0.1	0.1	0.13	0.1	0.1
N	15920	15910	15910	29505	29496	29496	14185	14183	14183	19441	19441	19441
	Health Care and Social Assistance								Education			
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
	LPM	Logit	Probit	LPM	Logit	Probit	LPM	Logit	Probit	LPM	Logit	Probit
Min Wage	0.07***	0.37***	0.22***	0.1***	0.49***	0.29***	0.05***	0.26***	0.15***	0.04***	0.19***	0.11***
	(0.02)	(0.08)	(0.05)	(0.01)	(0.04)	(0.02)	(0.01)	(0.07)	(0.04)	(0.01)	(0.04)	(0.02)
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2/pseudo-R^2	0.15	0.12	0.12	0.11	0.09	0.09	0.11	0.09	0.09	0.08	0.08	0.08
N	4484	4483	4483	21113	21113	21113	7508	7506	7506	22438	22438	22438

Note: The columns show regression coefficients of regressing minimum wage on labor force participation for workers earning at or below minimum wage in different industries. Columns 1-3, 8-10, 13-15, and 18-21 are for men and the remaining columns are for women. LPM stands for linear probability model. N is the number of observations. Robust standard errors are in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.12: Min. Wage Workers LFP by Industry, 2003-2019

			Manufacturing						Construction			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	LPM	Logit	Probit	LPM	Logit	Probit	LPM	Logit	Probit	LPM	Logit	Probit
Min Wage	0.07***	0.32***	0.2***	0.07***	0.33***	0.2***	0.08***	0.44***	0.26***	0.09***	0.47***	0.28***
	(0.01)	(0.05)	(0.03)	(0.01)	(0.06)	(0.04)	(0.01)	(0.05)	(0.03)	(0.03)	(0.14)	(0.08)
State and Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2/pseudo-R^2	0.17	0.14	0.14	0.18	0.14	0.14	0.15	0.15	0.15	0.22	0.2	0.2
N	10834	10831	10831	6937	6932	6932	15690	15684	15684	1847	1845	1845

Note: The columns show regression coefficients of regressing minimum wage on labor force participation for workers earning at or below minimum wage in different industries. Columns 1-3 and 8-10 are for men and the remaining columns are for women. LPM stands for linear probability model. N is the number of observations. Robust standard errors are in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

$\hat{F}_{\mathbf{Y}_{\langle 1|0,1\rangle}}$). Tables 2.13 through 2.17 show similar counterfactual effects for both men and women and for different comparison years. Since these tables are differences in CDFs, a way to interpret the tables is, for example in Table 2.2, there was a 10% increase in men earning up to $14.12 per hour and working up to 40 hours per week in 1992 assuming a counterfactual minimum wage from 1989.

Table 2.13: Minimum Wage Effect, Women 89-92

Wages/Hours	20	28	32	37	40	48
8.20	0.02*	0.01*	0.02*	0.02*	0.03*	0.03*
9.57	0.03*	0.02*	0.04*	0.04*	0.06*	0.06*
11.39	0.03*	0.03*	0.05*	0.04*	0.08*	0.09*
13.21	0.04*	0.04*	0.06*	0.06*	0.13*	0.13*
15.16	0.04*	0.03*	0.06*	0.05*	0.11*	0.12*
17.84	0.04*	0.02*	0.05*	0.03*	0.08*	0.08*
20.50	0.04*	0.01*	0.05*	0.01*	0.03*	0.03*
25.06	0.04*	0.01*	0.05*	0.01	0.01	0.01
32.57	0.04*	0.01*	0.06*	-0.01	−0.02*	-0.03*

Note: Effect of having 1992's minimum wage on 1989's joint CDF of wages and hours for women.

$*$ = 95% confidence level using 50 bootstrap samples.

Table 2.14: Minimum Wage Effect, Men 89-84

Wages/Hours	30	40	44	49	55
9.84	0.02*	0.05*	0.05*	0.05*	0.06*
12.31	0.03*	0.09*	0.09*	0.1*	0.11*
15.38	0.03*	0.12*	0.13*	0.14*	0.16*
18.46	0.03*	0.16*	0.17*	0.2*	0.22*
21.53	0.03*	0.15*	0.16*	0.19*	0.21*
24.61	0.02*	0.14*	0.15*	0.18*	0.19*
28.62	0.02*	0.04*	0.05*	0.06*	0.05
33.20	0.02	0.01	0.01	0.02	0
42.82	0.02	−0.03*	−0.03*	−0.03*	−0.05*

Note: Effect of having 1984's minimum wage on 1989's joint CDF of wages and hours for men.

$*$ = 95% confidence level using 50 bootstrap samples.

Table 2.15: Minimum Wage Effect, Women 89-84

Wages/Hours	20	26	32	38	40	45
8.24	0.02*	0.03	0.03	0.04	0.05	0.05
9.84	0.03*	0.02*	0.03*	0.05*	0.07*	0.07*
11.07	0.04*	0.03*	0.04*	0.06*	0.1*	0.1*
12.31	0.04*	0.03*	0.04*	0.06*	0.11*	0.11*
14.77	0.05*	0.03*	0.05*	0.08*	0.15*	0.17*
16.55	0.04*	0.02*	0.03*	0.06*	0.12*	0.13*
19.10	0.04*	0.01*	0.01	0.04*	0.07*	0.08*
22.76	0.04*	0	0	0.03*	0.03*	0.04*
28.61	0.04	−0.004*	-0.01	0.02*	-0.01	-0.01

Note: Effect of having 1984's minimum wage on 1989's joint CDF of wages and hours for women.

$*$ = 95% confidence level using 50 bootstrap samples.

Table 2.16: Minimum Wage Effect, Men 06-12

Wages/Hours	28	37	40	48	55
9.58	0.02*	0.03*	0.05*	0.05*	0.06*
11.30	0.04*	0.06*	0.11*	0.12*	0.13*
13.92	0.03*	0.06*	0.15*	0.16*	0.17*
16.71	0.03*	0.06*	0.19*	0.21*	0.22*
20.05	0.02*	0.05*	0.2*	0.23*	0.25*
23.56	0.02*	0.04*	0.19*	0.22*	0.24*
28.33	0.01*	0.03*	0.16*	0.19*	0.2*
35.69	0.01	0.02*	0.09*	0.11*	0.11*
49.42	0	0.01	0	0	-0.02

Note: Effect of having 2012's minimum wage on 2006's joint CDF of wages and hours for men.

$*$ = 95% confidence level using 50 bootstrap samples.

Table 2.17: Minimum Wage Effect, Women 06-12

Wages/Hours	20	29	35	40	48
8.91	0.02*	0.02*	0.04*	0.05*	0.05*
10.30	0.04*	0.04*	0.07*	0.1*	0.1*
11.97	0.04*	0.04*	0.07*	0.12*	0.12*
13.95	0.05*	0.04*	0.08*	0.15*	0.15*
16.71	0.05*	0.04*	0.08*	0.19*	0.2*
19.36	0.05*	0.03*	0.07*	0.17*	0.18*
23.36	0.04*	0.02	0.06*	0.13*	0.14*
28.96	0.04*	0.01	0.05*	0.07*	0.07*
39.64	0.04*	0	0.04*	0.01	0

Note: Effect of having 2012's minimum wage on 2006's joint CDF of wages and hours for women.

$*$ = 95% confidence level using 50 bootstrap samples.

These tables suggest that, for both men and women, there are considerable spillover effects. Although minimum wage should only directly affect workers at the bottom of the wage distribution, the majority of the effects were on workers working 40 or more hours per week (the bulk of the labor force) and earning wages close to the median of the wage distribution. This would suggest a strong compensating wage differential; when wages of the lowest paid workers increase, employers have to pay all workers more to entice them to remain at their jobs because outside work options have improved.

Additionally, the increase in wages of all workers due to minimum wage increases means hours of medium-wage workers should increase and hours of high-wage workers should decrease.[32] This is present in Figures 2.2-2.17 as higher minimum wage caused

[32] See discussion of the backward bending nature of the wage curve in Section 3.1.

workers at the top of the wage distribution to work fewer hours. Indeed, for women who generally work lower-paid jobs, there is a smaller negative effect of increased minimum wage on hours worked at the top of the wage distribution.

2.6.4 Distribution Effects with Conditional Mean Methods

As stated in the introduction, the conditional mean methods in this section take either hours or wages as exogenously given. While their results cannot capture effects that vary by both wages and hours, they serve as a benchmark.

The results from Table 2.18 and 2.19 show fixed effect regression of minimum wage (and minimum wage squared) on hours and log(wages) when the data is split into ten buckets of wage earners.

Table 2.18: Effect of Minimum Wage with State and Year Fixed Effects, 2003-2019

	Hours Worked				log(Wages)			
	Men		Women		Men		Women	
Total	0.041	(0.118)	0.178^{**}	(0.076)	0.021^{***}	(0.007)	0.011^{***}	(0.003)
%0-%10	0.406	(0.508)	-0.097	(0.199)	0.021	(0.018)	-0.014	(0.017)
%10-%20	-0.475^{*}	(0.291)	0.385^{***}	(0.143)	-0.003	(0.002)	-0.001	(0.001)
%20-%30	0.502	(0.365)	0.107	(0.189)	0	(0.001)	-0.002	(0.002)
%30-%40	-0.073	(0.352)	0.315^{**}	(0.155)	0.001	(0.001)	0	(0.001)
%40-%50	0.25	(0.17)	0.304^{*}	(0.172)	0.002	(0.001)	0.002^{***}	(0.001)
%50-%60	-0.346	(0.22)	-0.125	(0.158)	0	(0.001)	0.002^{*}	(0.001)
%60-%70	-0.096	(0.373)	0.291^{*}	(0.172)	0.003^{***}	(0.001)	-0.001	(0.001)
%70-%80	0.244	(0.202)	0.022	(0.211)	0.002	(0.001)	0	(0.001)
%80-%90	0.43	(0.291)	0.608^{***}	(0.136)	-0.001	(0.002)	0.001	(0.004)
%90-%100	0.015	(0.187)	0.129	(0.358)	0.002	(0.006)	-0.006	(0.006)

Note: The columns show mean regression coefficients of regressing minimum wage on hours or log(wage) with state and year fixed effects and additional controls for different buckets of wage earners. Robust standard errors in parenthesis. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 2.19: Effect of Minimum Wage on Hours Worked, 2003-2019

			Men		
	$\hat{\beta}_1$		$\hat{\beta}_2$		$-\frac{\hat{\beta}_1}{2\hat{\beta}_2}$
Total	1.41^{**}	(0.67)	-0.08^{**}	(0.04)	8.56
%0-%10	-6.17^{**}	(3.03)	0.39^{**}	(0.17)	7.83
%10-%20	4.94^{***}	(1.67)	-0.32^{***}	(0.09)	7.79
%20-%30	-5.1^{***}	(1.97)	0.34^{***}	(0.11)	7.61
%30-%40	4.83^{*}	(2.59)	-0.3^{**}	(0.15)	8.11
%40-%50	-1.45	(1.2)	0.1	(0.07)	7.08
%50-%60	0.82	(1.27)	-0.07	(0.08)	5.84
%60-%70	5.27^{**}	(2.23)	-0.33^{**}	(0.14)	8.03
%70-%80	2.53^{**}	(1.09)	-0.14^{**}	(0.06)	9.16
%80-%90	5.22^{***}	(1.29)	-0.29^{***}	(0.08)	9.03
%90-%100	0.74	(1.02)	-0.04	(0.06)	8.52
			Women		
Total	0.35	(0.55)	-0.01^{**}	(0.03)	16.73
%0-%10	-3.25^{**}	(1.44)	0.19^{*}	(0.09)	8.37
%10-%20	0.14	(1.38)	0.01^{*}	(0.08)	-4.55
%20-%30	-0.14	(1.45)	0.02^{*}	(0.09)	4.73
%30-%40	-2.63^{**}	(1.16)	0.18^{*}	(0.07)	7.25
%40-%50	1.91^{*}	(1.14)	-0.1^{*}	(0.07)	9.84
%50-%60	1.17	(1.05)	-0.08^{*}	(0.06)	7.49
%60-%70	-0.61	(1.41)	0.05^{*}	(0.08)	5.55
%70-%80	0.11	(1.6)	-0.005	(0.1)	10.53
%80-%90	0.97	(1.18)	-0.02^{*}	(0.07)	22.33
%90-%100	3.34^{**}	(1.71)	-0.19	(0.11)	8.87

Note: The columns show mean regression coefficients of regressing of minimum wage and minimum wage squared on hours worked with state and year fixed effects for different buckets of wage earners. Robust standard errors in parenthesis. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Roth and Sant'Anna (2021) formally state criteria for when partitioning the data into buckets and running fixed effects regressions is valid.[33]

2.7 Conclusion

Minimum wage policy's effect on hours worked was not well known because hours are only observed whose who are employed. Using a Heckman-type selection model to account for possible employment effects, I found that small increases to the minimum wage increased hours worked.

Additionally, considering jointly determined outcomes 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, I find that increased minimum wage had positive effects on both hourly wage and hours worked for all individuals except at the very top of the wage distribution, where hours were reduced. A possible explanation for these findings is that not all industries where minimum wage workers are employed are competitive. This chapter finds evidence for such a theory, since the effects of minimum wage policy on the hours of minimum wage workers differs by industry.

With richer data on firms, understanding the effects of minimum wage on hours worked by different market concentrations would corroborate my findings. Also, the method for estimating jointly determined outcomes presented in this chapter can help shed light on many controversial policy proposals that cause "winners" and "losers" at different parts of the joint distribution of outcomes.

[33]They require a "parallel trends" condition on the CDF of the untreated potential outcomes.

Chapter 3

Who Needs Help: Off-the-shelf Machine Learning Methods to Predict Income Mobility

3.1 Introduction

Determining the degree of intergenerational income mobility (or lack thereof) in American society has important policy implications. Furthermore, determining which subgroups of the population have the lowest intergenerational income mobility is crucial for fashioning effective programs to assist those most in need of help.

This chapter answers the above questions by using Random Forest and Gradient Boosting machine learning methods to determine which characteristics are the most important predictors of intergenerational mobility. I find that family wealth is the most important predictor of large increases and decreases in intergenerational mobility, suggesting that family wealth, not just income, should be an important factor in determining welfare assistance.

Interestingly, the importance of distinguishing between income and wealth was argued by Thomas Piketty in his popular book *Capital in the Twenty-first Century* (2017). In it, he asserts that in the long term the rate of return on capital (r) is greater than economic growth (g) and causes wealth to be concentrated among the rich. While Blume and Durlauf (2015) question the theoretical foundations of Piketty's claim, they contend that further theoretical and empirical analysis is warranted.

Pfeffer and Killewald (2017) find that parent and grandparent wealth are important predictors of wealth mobility. I find that parent wealth is also an important predictor of intergenerational income mobility. I further find that other influential factors include whether a parent's highest level of education included some college, whether the parent was African-American, and whether the parent was married.

Although I do not find geographical location important, I only considered four regions of the United States. With more data, geographical location might play a larger role, as found by Chetty, Hendren, Kline, and Saez (2014) who use commuting zones as regions.

Many machine learning and algorithmic methods are not widely used in economics because their goal is to make predictions about certain variables given other variables, whereas econometrics is typically concerned with the estimation and inference of a particular functional of a joint distribution of the data (see Athey and Imbens, 2019).[1] Accordingly, econometric methods can determine the causal effect of an intervention whereas machine learning methods cannot.

However, if the goal is to determine who is in most need of an intervention, machine learning methods are best for identifying such groups, and traditional econometric methods can determine the most beneficial forms of intervention.

There are several advantages to machine learning methods: (1) They are nonparametric and do not make strong functional form assumptions, (2) The models are validated and not subject to over-fitting concerns, and (3) Their out-of-sample perfor-

[1] Athey and Imbens, 2019 review and suggest new avenues for research that use machine learning methods and are relevant to economics problems.

mance is known. Additionally, some machine learning methods like Random Forest and Gradient Boosting use regression trees and give easily interpretable importance charts using an impurity function.[2]

Section 3.2 lists some of the terminology commonly used in the machine learning literature and throughout this chapter. Section 3.3 describes the data used in this chapter. Section 3.4 describes how to measure intergenerational income mobility and how the Random Forest and Gradient Boosting algorithms work. It also discusses how to measure predictor importance. Section 3.5 analyzes the results. Section 3.6 concludes.

3.2 Terminology

This section is meant to establish terms used in the chapter and to avoid confusion between machine learning and econometric terminology. Covariates or regressors are also called *features*. Data used in estimation is called the *training sample*, whereas data used to determine out-of-sample performance is called the *testing sample*. Estimating an unordered discrete dependent variable is known as *classification*, and when there are only two possible outcomes, it is called *binary classification*. *Hyperparameters* or *tuning parameters* are parameters whose values are set to control the learning algorithm, whereas endogenously determined parameters are usually called *weights*.

3.3 Data

This chapter uses data from the Panel Study of Income Dynamics (PSID).[3] The PSID is a longitudinal data set of households living in the United States starting from 1968 of over 18,000 individuals living in 5,000 families. It uses March Current Population

[2]See Section 3.4.
[3]Available at https://psidonline.isr.umich.edu/.

Survey (CPS) weights to make the sample nationally representative and contains income, wealth, occupation, health, and demographic information.

I compare head of households in 1991 with their offspring in 2017 between the ages of 25 and 64. The sample is made up of 2017 parent-child pairs. Only children present in the household in 1991 are used.

Table 3.1 shows some descriptive statistics about the sample. The descriptive statistics of the parents and children are similar with a few notable exceptions. First, because the children were surveyed and present in the household in 1991, their ages in 2017 vary less than their parents, with the majority between 25 and 34. And second, the children have higher levels of education than their parents.

Table 3.1: Descriptive Statistics

	N=2017			
	Parent (1991)		Child (2017)	
	Mean or %	(Std. Dev.)	Mean or %	(Std. Dev.)
Demographics				
Age	39.1	(10.3)	35.2	(6.5)
Age group 25-34	39 %		52.4 %	
Age group 35-44	34.4 %		37.8 %	
Age group 45-54	15.3 %		9.2 %	
Age group 55-64	11.26 %		0.6 %	
White	82.3 %		80.2 %	
African American	16.5 %		17.4 %	
Other	1.2 %		2.4 %	
Male	83.4 %		80.5 %	
Female	16.6 %		19.5 %	
Education				
Less than high school	19.2 %		7.7 %	
High school	38.6 %		22.3 %	
Some college	13.5 %		27.4 %	
Batchlors	16.7 %		22.2 %	
Postgraduate	12.1 %		20.4 %	
Region				
Northeast	23.1 %		18.4 %	
North Central	31.1 %		26.7 %	
South	30.22 %		34.7 %	
West	15.07 %		19.33 %	
Other	0.5 %		0.9 %	
Income (2017 dollars)				
Quantile 1 (lowest)	12,542.2	(6,508.6)	15,179	(7,711)
Quantile 2	31,026	(4,285.8)	33,701	(4,260.9)
Quantile 3	46,486	(4,298.3)	49,305.31	(4,623.3)
Quantile 4	66,148.3	(7,289.9)	69,403.3	(7,887)
Quantile 5 (highest)	123,872.1	(7,5031.4)	140,874.9	(8,1834.2)

Figure 3.1 shows the correlation between family wealth in 1989 and a child's wage in 2017.

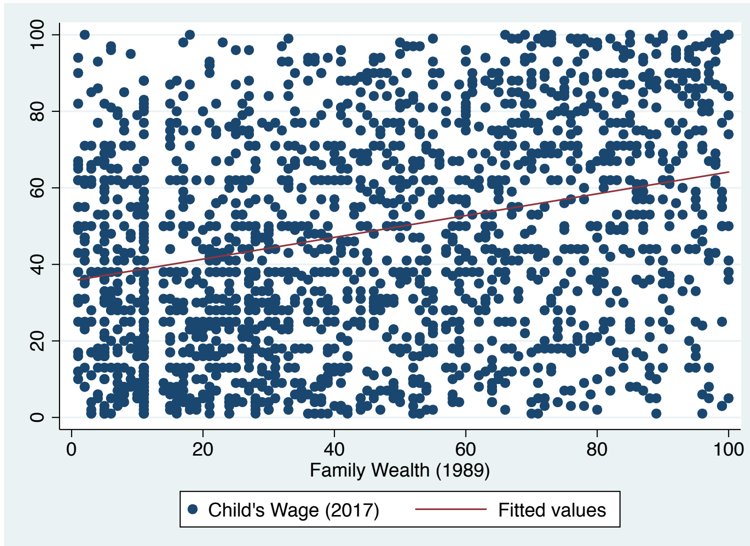


Figure 3.1: Correlation between Family Wealth and Child Income

The features used in the Random Forest and Gradient Boosting algorithms are: parent age, parent income percentile, family wealth percentile, and dummy variables for parent gender, marital status, geographic region,[4] industry, occupation,[5] race (White, African-American, and Other), and highest level of education completed (None, Some High School, High School, Some College, College, and Postgraduate).

[4]The four regions used are

NORTHEAST: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont

NORTH CENTRAL: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin

SOUTH: Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, Washington DC, West Virginia

WEST: Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming.

[5]Occupation and industry dummy variables are from from 1970 Census of Population. Please refer to Appendix 2, Wave XIV documentation, for complete listings.

3.4 Methods

3.4.1 Measuring Mobility

As pointed out by Chetty, Hendren, Kline, Saez, and Turner (2014), measuring intergenerational income mobility amounts to characterizing the joint distribution of parent and child income, which is comprised of (1) the joint distribution of parent and child ranks, i.e. the *copula* of the distribution, and (2) the *marginal* distributions of parent and child income. The marginal distributions determine the inequality within a generation, while the copula determines inequality across generations.

Common statistics that summarize the joint distribution of parent and child income are (i) intergenerational elasticities (IGE) of child income with respect to parent income, (ii) rank correlation between children and parents, and (iii) rank transition probability matrices.[6] Only methods (ii) and (iii) depend solely on the copula, while IGE combines features of both the copula and marginal distributions.

Methods that only summarize the copula and not the marginal distributions are useful for distinguishing changes in inequality from intergenerational mobility. However, it is important to note that doing so creates relative measures of intergenerational income mobility, i.e. measures that only depend on intergenerational ranks and may be misleading depending on the normative question at hand.[7]

The age-adjusted rank-rank slope (i.e. (ii)) is obtained from the regression:

$$RankI_c = \gamma_0 + \gamma_1 RankI_p + \gamma_2 Age_c + \gamma_3 Age_c^2 + \gamma_2 Age_p + \gamma_3 Age_p^2 + \epsilon_c, \qquad (3.1)$$

where $RankI_c$ is the rank of the child in the child's income distribution, $RankI_p$ is the rank of the parent in the parent's income distribution, Age_c is the child's age,

[6] See, for example, Black and Devereux (2011) for a review.

[7] An absolute measure of income mobility that measures mobility relative to some income level might be, for example, the percent of children earning more than their parents or the average difference between child income and mean parent income.

and Age_p is the parent's age. The coefficient γ_1 is the rank-rank slope coefficient and unaffected by inequality differences in the marginal distributions.

Similarly, the age-adjusted IGE (i.e. (i)) is obtained by the slope on the following regression:

$$\log(I_c) = \lambda_0 + \lambda_1 \log(I_p) + \lambda_2 Age_c + \lambda_3 Age_c^2 + \lambda_2 Age_p + \lambda_3 Age_p^2 + \epsilon_c, \qquad (3.2)$$

where $\log(I_c)$ is the natural logarithm of the child's income and $\log(I_p)$ is the natural logarithm of the parent's income. The coefficient λ_1 is the IGE and can be interpreted as approximately the percent change in income of a child that is caused by a percent change in income of the parent, all else being equal.

Methods (i) and (ii) depend strongly on the functional form of the relationship between parent and child income. Additionally, it does not characterize the heterogeneity in mobility by subgroup, e.g. mobility might be different for African-American parents and children. While this might be remedied by computing (i) or (ii) by subgroup, that would still not be informative as to which subgroups are most relevant in predicting mobility differently from the general population.

I offer a new method for characterizing relative intergenerational mobility that allows for ranking of the most important subgroups for predicting mobility. Let the new measure of mobility be called $M_{increase}(p)$ and be equal to 1 if the child has moved up p percentiles from the parent, and 0 otherwise. Similarly, let $M_{decrease}(p)$ and be equal to 1 if the child has moved down p percentiles from the parent, and 0 otherwise. Since a child whose parent is already in the top p-percentile cannot move above that p-percentile, the highest p-percentile should be eliminated from the sample to avoid ambiguity. The same holds true for children with parents in the lowest p-percentile. These new measures depend only on the copula and can easily be predicted with machine learning methods that allow for the most important subgroups that predict mobility to be ranked.

3.4.2 Random Forest

Consider the *Random Forest* model for classification (Breiman et al., 1984, Breiman, 2001). The data consists of w inputs and a response, for each of N observations: (x_i, y_i) for $i = 1, 2, \ldots, N$, with $x_i = (x_{i1}, x_{i2}, \ldots, x_{iw})$. The idea is to partition the sample into M regions $R_1, R_2, \ldots, R_M$ and estimate the regression function within each region as the average outcome. The partitioning is done sequentially based on the covariates x_i passing a threshold. In that way, the space of all joint predictor variable values is partitioned into disjoint regions as represented by terminal nodes of a decision tree. The algorithm is displayed below.

Algorithm 1 Random Forest for Classification

1. For $b = 1$ to B:
   (a) Draw a bootstrap sample, $\mathbf{X}^*$ and Y^*, of size N from the training data.
   (b) Make a random forest decision tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
      (i) Select m variables at random from the w variables.
      (ii) Pick the best variable/split-point among the m.
      (iii) Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_b\}_1^B$. Let $\hat{C}_b(x)$ be the class prediction of the bth random-forest tree. Then $\hat{C}_{rf}^B(x) = majority\ vote\{\hat{C}_b(x)\}_1^B$.

For step (ii): for covariate k and threshold c the sum of in-sample squared errors is

$$Q(c,k) = \sum_{i:X_{ik} \leq c} (Y_i - \bar{Y}_{k,c,l})^2 + \sum_{i:X_{ik} > c} (Y_i - \bar{Y}_{k,c,r})^2$$

where l and r denote left and right and

$$\bar{Y}_{k,c,l} = \sum_{i:X_{ik} \leq c} Y_i / \sum_{i:X_{ik} \leq c} 1 \text{ and } \bar{Y}_{k,c,r} = \sum_{i:X_{ik} > c} Y_i / \sum_{i:X_{ik} > c} 1.$$

To split the sample, minimize $Q(c, k)$ over all covariates $k = 1, ...m$ and all thresholds

$c \in (-\infty, \infty)$. Alternatively, an entropy-based criterion for sample splitting can be used.[8]

The hyperparameters of the random forest are (i) `criterion`: splitting criterion (entropy or gini), (ii) `max_depth`: (n_{min}) is maximum depth of the tree, (iii) `max_features`: (m) is maximum number of features random forest considers to split a node (either the square root of the total number of features or the logarithm base 2 of the total number of features.), and (iv) `n_estimators`: (B) is the number of bootstraps.

The hyperparameters are selected by k-fold cross validation.

3.4.3 Gradient Boosting

A *tree* can formally be expressed as

$$T(x;\Theta) = \sum_{j=1}^{J} \gamma_j \mathbb{1}(x \in R_j),$$

with parameters $\Theta = \{R_J, \gamma_j\}_1^J$ and $j = 1, 2, \ldots, J$ disjoint partitions of the parameter space (terminal nodes).

Using some loss function to measure closeness, the *Gradient Boosting* induces a tree at the mth iteration to be as close as possible to the negative gradient. The algorithm is displayed below.

Tuning parameters include the number of iterations M and the sizes of each of the constituent trees J_m, $m = 1, 2, \ldots, M$ and are selected by k-fold cross validation.

[8] See Hastie et al. (2001) for details.

Algorithm 2 Gradient Boosting for Classification

1. Initialize $f_0(x) = \arg\min_\gamma \sum_{i=1}^{N} L(y_i, \gamma)$,
   where $L(y_i, \gamma)$ is a logit loss function.
2. For $m = 1$ to M:
   (a) For $i = 1, 2, \ldots, N$ compute:
   "pseudo redisual" $r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}}$
   (b) Fit a regression tree to the targets r_{im} giving terminal regions R_{jm},
   $j = 1, 2, \ldots, J_m$.
   (c) For $j = 1, 2, \ldots, J_m$ compute
   $\gamma_{jm} = \arg\min_\gamma \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma)$.
   (d) Update $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} \mathbb{1}(x \in R_{jm})$.
3. Output $\hat{f}(x) = f_M(x)$.

3.4.4 Variable Importance

In a node m, representing a region R_m with N_m observations, let

$$\hat{p}_m = \frac{1}{N_m} \sum \mathbb{1}(y_i = 1),$$

be the proportion of class 1 observations in node m. Two measures of node impurity are

- Gini Impurity: $2\hat{p}_m(1 - \hat{p}_m)$
- Entropy Impurity: $-\hat{p}_m \log \hat{p}_m - (1 - \hat{p}_m) \log(1 - \hat{p}_m)$.

Impurity is used to measure the predictive strength of the variables. Gini impurity measures how often a randomly chosen element from the sample would be misclassified if it were randomly assigned values of the m variables in the region. Entropy impurity measures the information loss by randomly assigning values to the m variables in the region.

3.5 Results

Table 3.2 shows the age-adjusted intergenerational elasticities. There do not seem to be significant differences in IGE at different ages of the parent.

Table 3.2: Inter-generational Elasticities

	IGE	(SE)	N
Overall	0.14^{***}	(0.04)	2017
By age (4 groups)			
(1) Age 25-34	0.13^{***}	(0.04)	938
(2) Age 35-44	0.15^{***}	(0.05)	668
(3) Age 45-54	0.41^{***}	(0.16)	282
(4) Age 55-64	0.05	(0.15)	129
By age (2 groups)			
(5) Age 25-44	0.14^{***}	(0.11)	1606
(6) Age 45-64	0.15^{***}	(0.11)	411
Test of differences (p-values)			
(2) vs. (1)	0.77		
(3) vs. (1)	0.09^{*}		
(4) vs. (1)	0.59		
(5) vs. (6)	0.96		

Note: $p^{*} < .10$, $p^{**} < 0.05$, and $p^{***} < 0.01$ based on two-tailed tests. Grouped by parent's age.

Overall, a income IGE is 0.14, which is much lower than the wealth IGE which

Pfeffer and Killewald (2017) find, meaning there is more mobility in income than in wealth.

Table 3.3 shows age-adjusted rank slope coefficients. Like IGE, there do not seem to be a significant differences corresponding to different parent ages. Overall, a 10-percentile increase in the parent's income distribution is associated with a 1.4 percentile increase in the child's distribution. Income exhibits much more mobility than what Pfeffer and Killewald (2017) find for wealth. They find the overall age-adjusted rank-rank slope coefficient for wealth is 0.39.

Table 3.3: Inter-generational Rank Correlations

	Rank Slope	(SE)	N
Overall	0.18^{***}	(0.04)	2017
By age (4 groups)			
(1) Age 25-34	0.16^{***}	(0.05)	938
(2) Age 35-44	0.17^{***}	(0.06)	668
(3) Age 45-54	0.29^{***}	(0.09)	282
(4) Age 55-64	0.18^{**}	(0.17)	129
By age (2 groups)			
(5) Age 25-44	0.17^{***}	(0.04)	1606
(6) Age 45-64	0.21^{***}	(0.09)	411
Test of differences (p-values)			
(2) vs. (1)	0.93		
(3) vs. (1)	0.23		
(4) vs. (1)	0.95		
(5) vs. (6)	0.71		

Note: $p^{*} < .10$, $p^{**} < 0.05$, and $p^{***} < 0.01$ based on two-tailed tests. Grouped by parent's age.

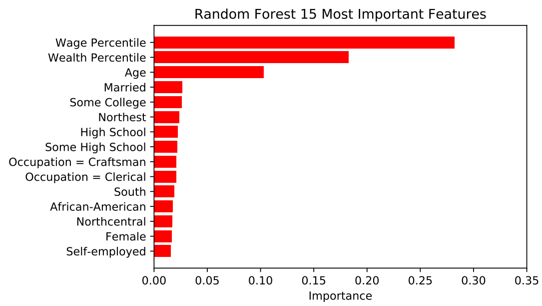
Table 3.4 shows the income quintile transition matrix.

Table 3.4: Income Quintile Transition Matrix

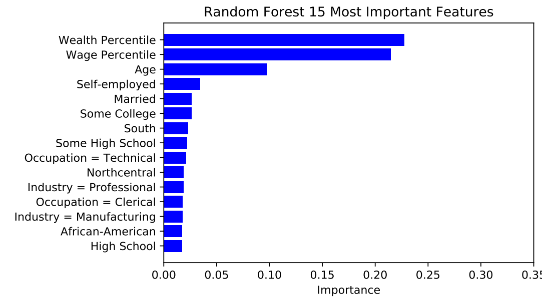
Parent (1991)		Child (2017)					
		Lowest [<= $26k]	Quintile 2 [$26k-$40k]	Quintile 3 [$40k-$57k]	Quintile 4 [$58k-$85k]	Highest [>=$85K]	Total
Lowest	[<= $22k]	28	21	17	19	16	100
Quintile 2	[$23k-$37k]	26	24	18	16	16	100
Quintile 3	[$37k-$54k]	18	20	23	25	14	100
Quintile 4	[$55k-$78k]	15	18	20	23	25	100
Highest	[>=$78K]	15	14	23	20	28	100

Note: In 2017 dollars. N=2017.

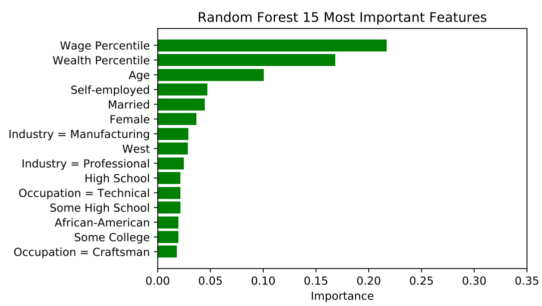
Figure 3.2 shows the most important features (based on Gini impurity) for moving up in the income distribution using Random Forest classification. For smaller movements, parent income percentile is the most important predictor. For large movements, parent wealth percentile is the most important predictor. Figure 3.3 shows the same thing but with Gradient Boosting classification. The results are similar.



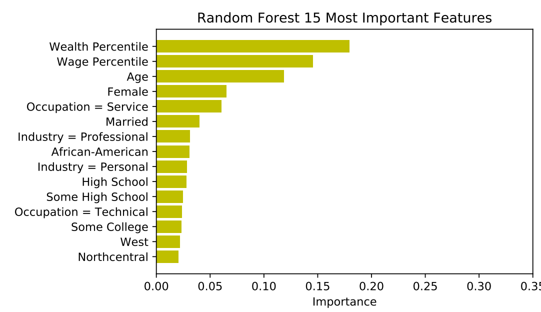
(a) Increase of 20 percentile or more. Testing Error: 0.26



(b) Increase of 40 percentile or more. Testing Error: 0.19

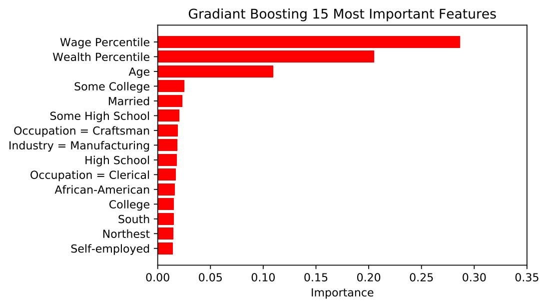


(c) Increase of 60 percentile or more. Testing Error: 0.11

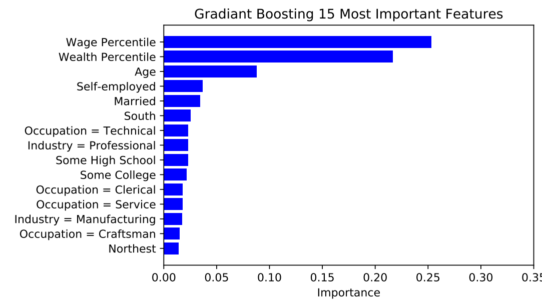


(d) Increase of 80 percentile or more. Testing Error: 0.03

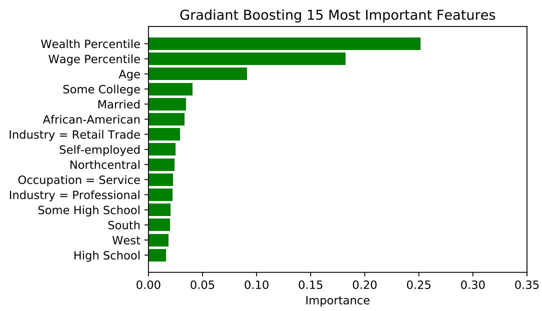
Figure 3.2: Most Important Features for Moving Up in the Income Distribution, Random Forest



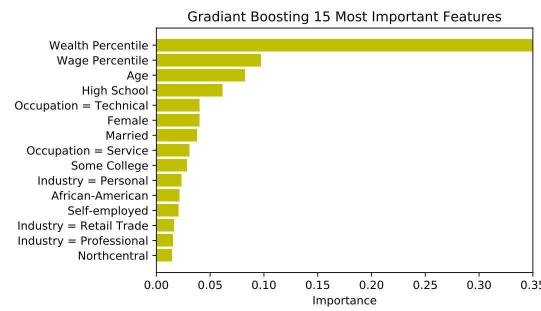
(a) Increase of 20 percentile or more. Testing Error: 0.30



(b) Increase of 40 percentile or more. Testing Error: 0.19



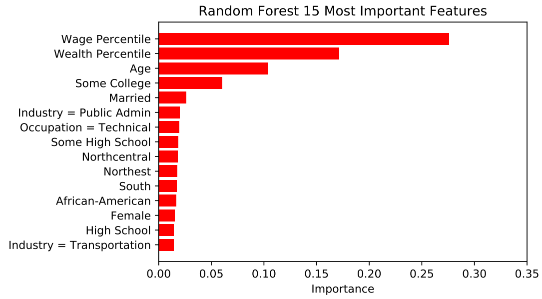
(c) Increase of 60 percentile or more. Testing Error: 0.13



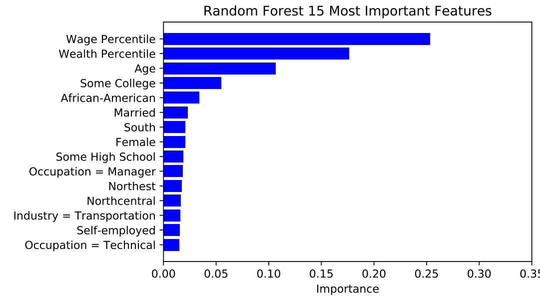
(d) Increase of 80 percentile or more. Testing Error: 0.06

Figure 3.3: Most Important Features for Moving Up in the Income Distribution, Gradient Boosting

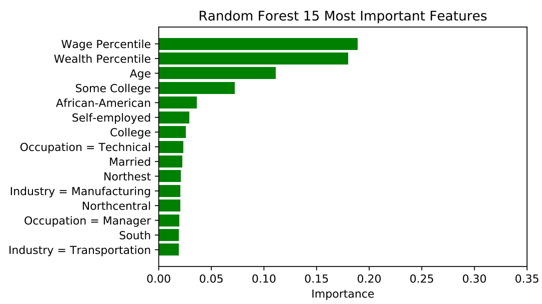
Figure 3.4 shows the most important features (based on Gini impurity) for moving down in the income distribution using Random Forest classification. Just as with increases in income, for smaller movements, parent income percentile is the most important predictor. For large movements, parent wealth percentile is the most important predictor. Figure 3.5 shows the same thing but with Gradient Boosting classification. The results are similar.



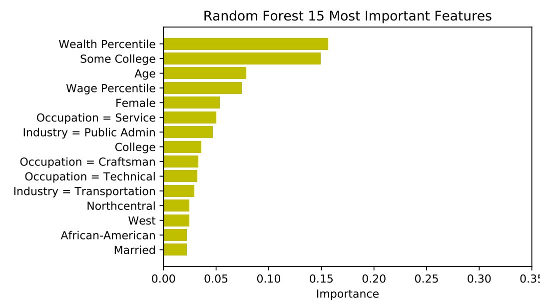
(a) Decrease of 20 percentile or more.
Testing Error: 0.32



(b) Decrease of 40 percentile or more.
Testing Error: 0.28

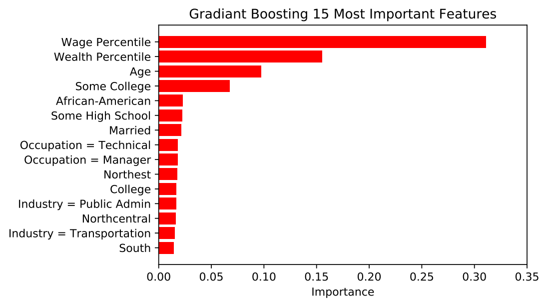


(c) Decrease of 60 percentile or more.
Testing Error: 0.24

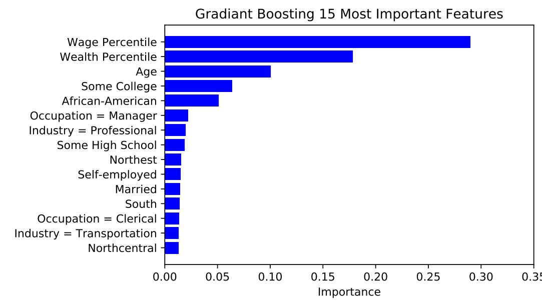


(d) Decrease of 80 percentile or more.
Testing Error: 0.09

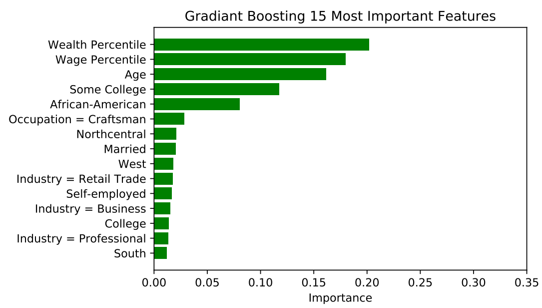
Figure 3.4: Most Important Features for Moving Down in the Income Distribution, Random Forest



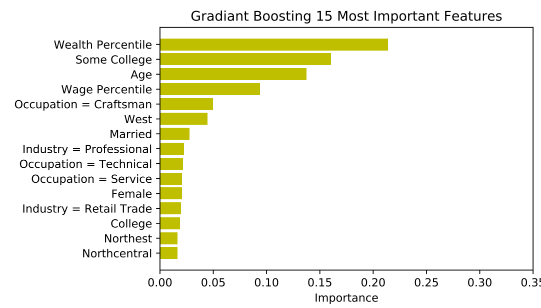
(a) Decrease of 20 percentile or more.
Testing Error: 0.26



(b) Decrease of 40 percentile or more.
Testing Error: 0.29



(c) Decrease of 60 percentile or more.
Testing Error: 0.25



(d) Decrease of 80 percentile or more.
Testing Error: 0.06

Figure 3.5: Most Important Features for Moving Down in the Income Distribution, Gradient Boosting

3.6 Conclusion

Predictors of income intergenerational mobility have important policy implications. Machine learning methods such as Random Forest and Gradient Boosting are non-parametric, give a performance indicator based on out-of-sample prediction, and allow predictors to be ranked. Using such methods, this chapter found family wealth percentile is the most important predictor of large increases and decreases in intergenerational income mobility.

Appendix A

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

$\hat{F}_{y^1,y^2}$, obtained by empirical copula and the empirical CDF. Table A.2 shows $c(0.75) - c(0.25)$, i.e. 0.75-quantile-0.25-quantile of

$$\left\{ \max_{y_1 \in T_1, y_2 \in T_2} \left| \left(\hat{F}^{*(j)}_{\mathbf{Y}}(y_1, y_2) - \hat{F}^{*(j)}_{\mathbf{Y}}(y_1, y_2) \right) - \left(\hat{F}^{(j)}_{\mathbf{Y}}(y_1, y_2) - \hat{F}^{(j)}_{\mathbf{Y}}(y_1, y_2) \right) \right| \Big/ \hat{s}(y_1, y_2) \right\}_{j=1}^{B},$$

which should roughly be a table of 2's if the t-ratio is approximately normally distributed.

Both the point estimates and confidence bands seem to be working correctly.

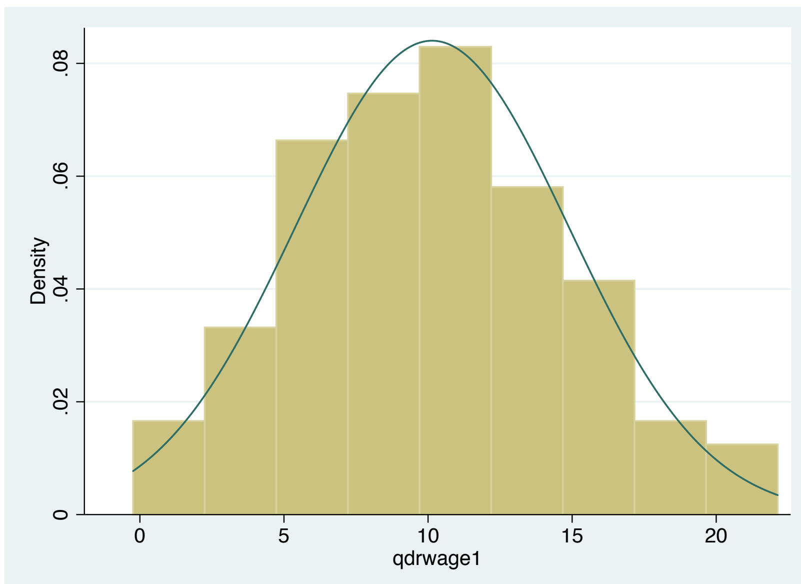


Figure A.1: Simulation of y_1

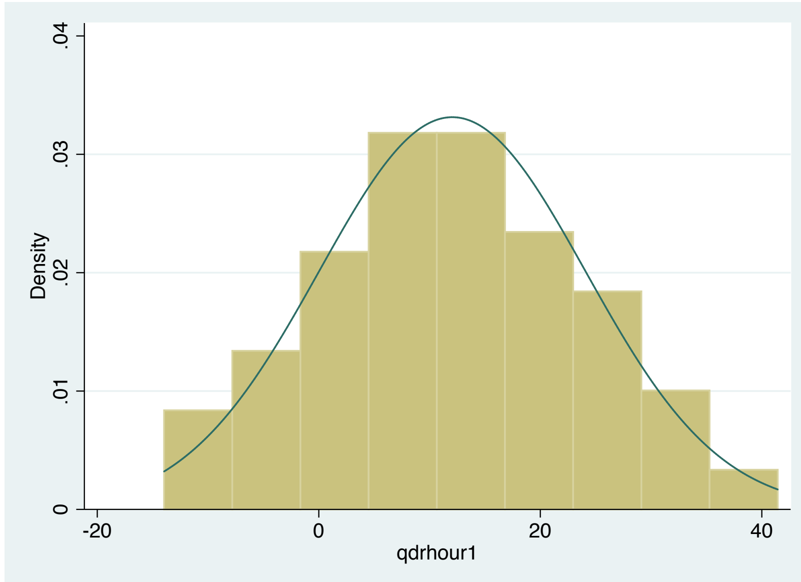


Figure A.2: Simulation of y_2

Table A.1: Simulated Difference

y^1/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	0	0	0	0
5.93	0	0	0	0	0	0	0	0	0
7.31	0	0	0	0	-0.01	-0.01	-0.01	0	0
8.62	0	0	0	0	-0.01	-0.02	-0.01	-0.01	-0.01
9.95	0	-0.01	-0.01	-0.01	-0.02	-0.03	-0.03	-0.03	-0.03
11.27	0	0	0	0	-0.02	-0.02	-0.02	-0.02	-0.02
12.53	0	0	-0.01	-0.01	-0.02	-0.03	-0.03	-0.03	-0.03
14.08	0	0	0	-0.01	-0.02	-0.03	-0.03	-0.03	-0.03
16.39	0	0	-0.01	-0.01	-0.03	-0.03	-0.03	-0.03	-0.03

Note: $*$ = 95% confidence level using 50 bootstrap samples.

Table A.2: $c(0.75) - c(0.25)$

y^1/y^2	-4.65	1.07	5.26	8.52	11.39	14.91	18.33	22.64	28.38
3.72	2.04	1.99	1.99	2.05	1.89	1.7	1.82	1.71	1.75
5.93	1.95	2.29	1.63	1.95	1.84	1.75	2	2.02	2.03
7.31	2.21	1.97	1.99	1.84	2.19	1.9	2.04	2.13	2.11
8.62	2.61	1.84	1.88	2.26	2.14	1.8	1.81	1.83	2.14
9.95	2.06	2.15	1.91	2.56	2.32	2.14	2.11	2.01	1.89
11.27	2.06	2.26	2.04	2.43	2.35	2	2.17	1.8	2.01
12.53	2.2	2.04	2.13	2.06	1.96	1.79	2.08	2.17	2.2
14.08	2.02	1.99	1.84	2.09	2.01	2.31	2.03	1.97	2.06
16.39	1.93	1.93	2.15	1.99	2.09	2.49	1.98	1.84	2.23

Appendix B

Proofs

Proof of Theorem 2.4.2. 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,

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

and

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

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. ■

Appendix C

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.[1] 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[2] and quantiles[3] 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

[1] 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.

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

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

data splitting to account for outcome heterogeneity is that these groupings can be arbitrary, with results 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.[4] Although these two methods estimate different "objects," they are in essence capturing the same idea of heterogeneous outcomes.

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 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. Moreover, interpreting the results would quickly become infeasible as the number of outcomes of interest increases. Alternatively, comparing multidimensional distributions, discussed in Cahn and Maasoumi (2022), 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.[5] It might be tempting, for example, to 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 in situations where the policy affects both outcomes of interest.

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

[5]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.

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 ease of interpretability of the findings and the fact that this method allows for outcomes to be affected heterogeneously while not relying on exogeneity of any of the outcomes of interest.

Bibliography

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:

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

Athey, S., & Imbens, G. (2019). Machine learning methods economists should know about.

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

Atkinson, A. B. (2011). The restoration of welfare economics. *American Economic Review*, *101*(3), 157–61.

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.

Black, S., & Devereux, P. (2011). Recent developments in intergenerational mobility. In O. Ashenfelter & D. Card (Eds.). Elsevier.

Blume, L., & Durlauf, S. (2015). Capital in the twenty-first century: A review essay. *Journal of Political Economy*, *123*(4), 749–777.

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

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth; Brooks.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Cahn, Y., & Maasoumi, E. (2022). *Modeling heterogeneous outcomes* (tech. rep.). Emory University.

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.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *. *The Quarterly Journal of Economics*, *129*(4), 1553–1623.

Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the united states still a land of opportunity? recent trends in intergenerational mobility. *American Economic Review*, *104*(5), 141–47.

Deheuvels, P. (1979). La fonction de dã©pendance empirique et ses propriã©tã©s. *Acad. Royl. Belg., Bull. C1 Sci. 5iÃ¨me sÃ©r*, *65*, 274â292.

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.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. Springer New York Inc.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.

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

Lin, Z., & Maasoumi, E. (2019). Elongated pseudo-panels from time series of cross sections *.

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.

Pfeffer, F. T., & Killewald, A. (2017). Generations of Advantage. Multigenerational Correlations in Family Wealth. *Social Forces*, *96*(4), 1411–1442.

Piketty, T. (2017). *Capital in the twenty-first century* (A. Goldhammer, Trans.). Belknap Press.

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?

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.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data* (Vol. 1). The MIT Press.