6-1-2018

Extrapolation using Selection and Moral Hazard Heterogeneity from within the Oregon Health Insurance Experiment

Amanda E. Kowalski

EXTRAPOLATION USING SELECTION AND MORAL HAZARD HETEROGENEITY FROM WITHIN THE OREGON HEALTH INSURANCE EXPERIMENT

By

Amanda E. Kowalski

May 2018

COWLES FOUNDATION DISCUSSION PAPER NO. 2135

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Extrapolation using Selection and Moral Hazard Heterogeneity from within the Oregon Health Insurance Experiment*

Amanda E. Kowalski
Associate Professor, Department of Economics, Yale University
Faculty Research Fellow, NBER

May 17, 2018

Abstract

I aim to shed light on why emergency room (ER) utilization increased following the Oregon Health Insurance Experiment but decreased following a Massachusetts policy. To do so, I unite the literatures on insurance and treatment effects. Under an MTE model that assumes no more than the LATE assumptions, comparisons across always takers, compliers, and never takers can inform the impact of polices that expand and contract coverage. Starting from the Oregon experiment as the "gold standard," I make comparisons within Oregon and extrapolate my findings to Massachusetts. Within Oregon, I find adverse selection and heterogeneous moral hazard. Although previous enrollees increased their ER utilization, evidence suggests that subsequent enrollees will be healthier, and they will decrease their ER utilization. Accordingly, I can reconcile the Oregon and Massachusetts results because the Massachusetts policy expanded coverage from a higher baseline, and new enrollees reported better health.

*amanda.kowalski@yale.edu. This paper includes new material as well as material from NBER Working Paper 22363, "Doing More When You're Running LATE: Applying Marginal Treatment Effect Methods to Examine Treatment Effect Heterogeneity in Experiments" (Kowalski, 2016). I thank Saumya Chatrath, Aigerim Kabdiyeva, Samuel Moy, Ljubica Ristovska, and Matthew Tauzer for excellent research assistance. Joseph Altonji, John Asker, Steve Berry, Christian Brinch, Lasse Brune, Pedro Carneiro, Raj Chetty, Joseph Doyle, Mark Duggan, Caroline Hoxby, Liran Einav, Amy Finkelstein, Matthew Gentzkow, Jonathan Gruber, John Ham, Guido Imbens, Dean Karlan, Larry Katz, Michal Kolesar, Jonathan Levin, Rebecca McKibbin, Sarah Miller, Costas Meghir, Magne Mogstad, Mark Rosenzweig, Joseph Shapiro, Orie Shelef, Ashley Swanson, Ed Vytlacil, David Wilson, and seminar participants at Academia Sinica, AEA Annual Meeting, Annual Health Econometrics Workshop, Berkeley, BU/MIT/Harvard Health Economics, CHES, Chicago Harris, Dartmouth, Duke Fuqua, IFS, LSE, Michigan, NBER Summer Institute, Northwestern, Ohio State, Princeton, Rand, Santa Clara, SMU, Stanford, Stanford GSB, Stanford SITE, Stockholm, UBC, UC Davis, UC Irvine, UConn Development Conference, UCLA Anderson, USC, UT Austin, Yale, Wharton, Wisconsin, and WEAI provided helpful comments. NSF CAREER Award 1350132 and the Stanford Institute for Economic Policy Research (SIEPR) provided support. **Keywords:** compliers, marginal treatment effect, Massachusetts health reform, program evaluation, treated outcome test, untreated outcome test.

1 Introduction

Findings from the Oregon Health Insurance Experiment are considered the "gold standard" for evidence in health economics because they are based on a randomized lottery. The state of Oregon conducted the lottery in 2008 as a fair way to expand eligibility for its Medicaid health insurance program to a limited number of uninsured individuals. The lottery also effectively created a randomized experiment that facilitated evaluation of the impact of Medicaid. It is challenging to evaluate the impact of Medicaid because there could be "adverse selection" into Medicaid such that Medicaid enrollees would have worse outcomes than other individuals, even if they were not enrolled. Concerns about selection heterogeneity motivate the popularity of results from the Oregon experiment.

A headline finding from the Oregon experiment is that Medicaid increased emergency room (ER) utilization (Taubman et al., 2014). Legislation requires that emergency rooms see all patients regardless of insurance coverage, so the uninsured often access the healthcare system through the ER. There was hope that Medicaid would *decrease* ER utilization, either because of substitution toward primary care or because of improved health. However, it is plausible that Medicaid increased ER utilization because formerly uninsured individuals could visit the ER at lower personal cost when enrolled in Medicaid. Any treatment effect of Medicaid on ER utilization can be considered "moral hazard," which arises when individuals who gain insurance change their consumption of the insured good. The sign and magnitude of this type of moral hazard are important for policy evaluation because care provided in the ER is expensive, but the insured do not necessarily value additional ER care at its cost, leading to inefficiency.

Meta-analysis and comparisons made in the popular press suggest that moral hazard could vary across policies, implying that the headline finding from Oregon need not apply directly to all policies (Rovner, 2014; Sommers and Simon, 2017; Tavernise, 2014). Importantly, several studies on the Massachusetts health reform of 2006 show that ER utilization decreased or stayed the same (Chen et al., 2011; Kolstad and Kowalski, 2012; Miller, 2012; Smulowitz et al., 2011). Studies on other policies yield varying results (Anderson et al., 2012, 2014; Currie and Gruber, 1996; Newhouse and Rand Corporation Insurance Experiment Group, 1993).

To extrapolate from the Oregon experiment to other policies, including the Massachusetts reform, I unite the literatures on insurance and treatment effects. I use a generalized Roy (1951) model of the marginal treatment effect (MTE) introduced by Björklund and Moffitt (1987), in the tradition of Heckman and Vytlacil (1999, 2001, 2005), Carneiro et al. (2011), and Brinch et al. (2017), to characterize how selection and moral hazard vary independently as Medicaid enrollment expands and contracts. Although the MTE literature generally focuses on variation within a single context, I use variation from within the Oregon experiment to inform extrapolation to Massachusetts. I begin with a version of the MTE model shown by Vytlacil (2002) to assume no more than the local average treatment effect (LATE) assumptions of independence and monotonicity proposed by Imbens and Angrist (1994). My exposition of the model emphasizes the link between the MTE and always takers, compliers, and never takers, using the terminology of Angrist et al.

(1996) from the LATE literature.

Under the model, variation across always takers, compliers, and never takers is variation across an important margin: the margin of enrollment in Medicaid. Individuals are only always takers, compilers, or never takers with respect to a given policy. If Medicaid enrollment were to expand to more lottery entrants, then never takers would enroll, especially if a mandate required them to do so. If Medicaid enrollment were to contract, then compliers would disenroll, followed by always takers. Therefore, comparison of ER utilization across always takers, compliers, and never takers is central to identification of how selection and moral hazard vary as Medicaid enrollment expands and contracts. I use the model to calculate the average ER utilization of always takers, compliers with Medicaid, compliers without Medicaid, and never takers, following Imbens and Rubin (1997), Katz et al. (2001), Abadie (2002), and Abadie (2003).[1] I emphasize that calculation of the LATE does not require calculation of these four averages or even the ability to calculate them.[2] Therefore, these averages provide information that is not contained in the LATE that I can use for identification of selection and moral hazard heterogeneity.

I define selection and moral hazard heterogeneity using general functions from the model. A function that I refer to as the marginal untreated outcome (MUO) function defines selection heterogeneity along the entire enrollment margin, and the marginal treatment effect (MTE) function defines moral hazard heterogeneity along the entire enrollment margin. I weight these functions along ranges of the enrollment margin to construct the special cases of selection and moral hazard heterogeneity defined in the econometric literature (Angrist (1998) and Heckman et al. (1998) use a common set of definitions). I demonstrate that the definitions from the literature depend on the fraction of lottery winners in the Oregon experiment, which is not desirable if the goal is to characterize underlying behavior that is invariant to a parameter of the experimental design used to study it. Furthermore, under the definitions from the literature, it is not possible to identify selection or moral hazard heterogeneity without ancillary assumptions.

Using the general functions, I can identify a special case of selection heterogeneity without

[1] I illustrate the calculations graphically in Appendix C to build intuition. The calculations rely on average ER utilization for four observed groups: lottery losers with Medicaid (always takers only), lottery winners with Medicaid (always takers and compliers with Medicaid), lottery losers without Medicaid (never takers and compliers without Medicaid), and lottery winners without Medicaid (never takers only). Because of randomization, average ER utilization of lottery losers with Medicaid identifies average ER utilization with Medicaid for all always takers, even the lottery winners. Similarly, average ER utilization of lottery winners without Medicaid identifies average ER utilization without Medicaid for all never takers. Furthermore, the fraction of always takers among lottery losers and the fraction of never takers among lottery winners identify the respective fractions in the full sample. Using these fractions and average ER utilization for always takers with Medicaid and never takers without Medicaid, it is straightforward to back out average ER utilization for compliers with and without Medicaid from the average ER utilization for lottery winners with Medicaid and lottery losers without Medicaid. (It is not possible to calculate average ER utilization for always takers *without* Medicaid or never takers *with* Medicaid without ancillary assumptions because these groups do not change their enrollment based on the lottery.)

[2] The Wald (1940) approach to calculate the LATE divides the "reduced form," the average difference in ER utilization between lottery winners and losers, by the "first stage," the average difference in Medicaid enrollment between lottery winners and losers. As shown by Angrist (1990) and Angrist and Krueger (1992), it is possible to calculate the LATE using two separate datasets: one that includes ER utilization by lottery status (for the reduced form), and another that includes Medicaid enrollment by lottery status (for the first stage). However, it is not possible to observe ER utilization by lottery status *and* Medicaid enrollment in either dataset, so it is not possible to calculate any of the four averages.

further assumptions. I do so using a test that I refer to as the "untreated outcome test," which compares outcomes of compliers and and never takers. Similarly, I use a test that I refer to as the "treated outcome test," which compares outcomes of compliers and always takers, to identify the combined impact of selection and moral hazard heterogeneity. These tests are equivalent to tests proposed in the econometric literature by Bertanha and Imbens (2014); Guo et al. (2014); Black et al. (2015), and Mogstad et al. (2017), who also propose testing both jointly. The tests of Hausman (1978); Heckman (1979); Willis and Rosen (1979); Angrist (2004); Huber (2013) and Brinch et al. (2017) can also be expressed in terms of functions of the treated and untreated outcome tests (see Bertanha and Imbens, 2014). My innovation is in the interpretation. I emphasize that the distinction between the untreated and the treated is an important one, especially in insurance contexts where the distinction separates the insured from the uninsured. By relating these tests to general functions from the MTE model, I show that the untreated outcome tests isolates selection heterogeneity, and the treated outcome test identifies the combined influence of selection and treatment effect heterogeneity. Without ancillary assumptions, the joint implementation of the treated and untreated outcome tests is no more informative about treatment effect heterogeneity than the treated outcome test is on its own.

Tests from the insurance literature also cannot separately identify moral hazard heterogeneity within the Oregon experiment without ancillary assumptions. Uniting the insurance and treatment effects literatures, I demonstrate that Einav et al. (2010) cost curve tests from the insurance literature can be applied within the Oregon experiment. I emphasize that the Einav et al. (2010) cost curve test applied to uninsured costs is equivalent to the untreated outcome test. Therefore, it identifies selection heterogeneity. Furthermore, the Einav et al. (2010) cost curve test applied to insured costs is equivalent to the treated outcome test. Therefore, it identifies heterogeneous moral hazard, or heterogeneous selection, or both. The Einav et al. (2010) cost curve test applied to the difference between insured and uninsured costs would identify heterogeneous moral hazard, but it cannot be run within the Oregon experiment without ancillary assumptions because only one point on the curve is identified. Other tests from the insurance literature also cannot separately identify moral hazard heterogeneity within the Oregon experiment without ancillary assumptions. Under the Chiappori and Salanie (2000) positive correlation test, a correlation between insurance coverage and insured spending could indicate heterogeneous moral hazard, or heterogeneous selection, or both. Under the Finkelstein and Poterba (2014) unused observables test, a correlation between a covariate and insurance coverage and a second correlation between the same covariate and insured spending could also indicate heterogeneous moral hazard, or heterogeneous selection, or both.

To separately identify moral hazard heterogeneity, I impose two ancillary assumptions following Brinch et al. (2017), who impose them to study the impact of family size on child outcomes. In my context, the first ancillary assumption requires that ER utilization without Medicaid varies linearly with the fraction of the sample enrolled in Medicaid conditional on the lottery outcome. Under this assumption, the linear marginal untreated outcome (MUO) function identifies selection heterogeneity along the entire enrollment margin. The second ancillary assumption, which is equivalent to an assumption made by Olsen (1980) to study the impact of family size on maternal outcomes, requires linearity of a function that I refer to as the marginal treated outcome (MTO) function. In my context, the MTO function characterizes how ER utilization for Medicaid enrollees changes as Medicaid enrollment expands and contracts. I emphasize that the MTO function characterizes the combined impact of selection and moral hazard heterogeneity.

Under both ancillary assumptions, the difference between the MTO and MUO functions yields a linear marginal treatment effect (MTE) function, which I use to characterize how moral hazard varies as Medicaid enrollment expands and contracts. Linearity of the MTE function has precedent as a direct assumption in applied work (see Moffitt, 2008; French and Song, 2014). Furthermore, applied work that extrapolates as if the LATE applies everywhere makes an implicit assumption that the MTE function is linear and has zero slope.

To extrapolate from the Oregon experiment to other policies, including the Massachusetts reform, I reweight the Oregon MTE using weights that are special cases of general weights for MTE reweighting given by Heckman and Vytlacil (2007). Unlike the weights used by Brinch et al. (2017), these weights allow me to recover and predict outcomes and treatment effects for always takers, compliers, and never takers, which are of interest for extrapolation to specific policies that induce full enrollment or full dis-enrollment. I emphasize that using these weights, I only need one of the two ancillary assumptions to predict the impact of policies that enroll all never takers or dis-enroll all always takers.

To extrapolate using observables, I incorporate a shape restriction into both ancillary assumptions. This shape restriction is common in the MTE literature (Carneiro and Lee, 2009; Carneiro et al., 2011; Maestas et al., 2013; Kline and Walters, 2016; Brinch et al., 2017). In my context, it requires that included observables and the residual unobserved net cost of Medicaid enrollment have additively-separable impacts on potential ER utilization with and without Medicaid. Departing from the literature, which places little emphasis on the choice of included observables, I vary the included observables to determine which ones explain the moral hazard heterogeneity I find.

I begin my empirical analysis by taking the model without any ancillary assumptions to the data. By using publicly available data previously used to evaluate the Oregon Health Insurance Experiment, I encourage further replication and future work.[3] I replicate the headline finding that moral hazard is positive: Medicaid increases ER utilization. Per Imbens and Angrist (1994) and Angrist et al. (1996), the headline finding is interpretable as a LATE that only applies to compliers.

Within my analysis sample, only 26% of individuals are compliers, while 15% are always takers and 59% are never takers. Always and never takers are possible because individuals entered the experiment by joining a waitlist for Medicaid, but they were only required to provide eligibility documentation if they won. Therefore, always takers, who might not have been aware of their prior eligibility when they entered the waitlist, could still enroll even if they lost. Medicaid allows hospitals to facilitate retroactive enrollment of eligibles, so it is possible that some always takers enrolled after visiting the ER. On the other side of the spectrum, never takers did not enroll even if

[3]Publicly available data are rare in a landscape of proprietary and confidential data in health economics. I am grateful to the investigators of the Oregon Health Insurance Experiment for making their data available. I co-developed the Stata command *mtebinary* to aid replication of my results (Kowalski et al., 2016).

they won, either because they were ineligible or because they did not submit eligibility information in the required timeframe.

Examining ER utilization, I find that never takers have lower average ER utilization than compliers without Medicaid, which identifies adverse selection from compliers to never takers via the untreated outcome test. Applying the treated outcome test, I find that always takers have higher average ER utilization than compliers with Medicaid. This result is consistent with decreasing moral hazard or adverse selection from always takers to compliers. To separately identify moral hazard heterogeneity, I impose the ancillary assumptions, and I find that moral hazard decreases from always takers to compliers to never takers, so much so that average moral hazard is *negative* for never takers. Therefore, subsequent expansions of Medicaid in Oregon could result in decreased use of the ER.

Before extrapolating to Massachusetts, I assess whether my findings from within Oregon apply within Massachusetts. To do so, I recast my previous work on the Massachusetts reform from Hackmann et al. (2015), which builds on the Einav et al. (2010) model from the insurance literature, in terms of the MTE model with ancillary assumptions. I show that the marginal cost curve for the insured minus the uninsured is an MTE function. Therefore, Hackmann et al. (2015) shows a decreasing MTE function for total health care costs, which is consistent with the decreasing MTE function for ER utilization that I find in Oregon.

I acknowledge that there are several factors that could have differed between the Massachusetts and Oregon contexts. The Oregon policy was a randomized lottery conducted only for individuals who entered it, and the Massachusetts reform was a state-wide policy. Furthermore, institutional features of the health care environment could have differed across states. As discussed by Miller (2012), Massachusetts had an uncompensated care pool that might have encouraged excess emergency care before its dissolution and replacement under the Massachusetts reform. Also, health insurance terms could also have differed, especially since Oregon expanded Medicaid alone and Massachusetts also expanded other types of coverage. Despite these differences, the MTE in Oregon and the MTE in Massachusetts both slope downward, suggesting that in both contexts, those who increase their utilization the most in response to health insurance coverage enroll first.

Given decreasing MTE functions in Oregon and Massachusetts, I rely on the Oregon experiment as the "gold standard," and I characterize the Massachusetts reform as an expansion of Medicaid along the Oregon MTE. Because enrollment levels were high in Massachusetts before and after the reform, MTE-reweighting predicts that Massachusetts compliers respond to insurance like a subset of Oregon never takers. With MTE-reweighting, I predict a decrease in ER utilization in Massachusetts of the same order of magnitude as the decrease found by Miller (2012). MTE-reweighting thus offers a plausible potential pathway to reconcile the increase in ER utilization found in Oregon with the decrease in ER utilization found in Massachusetts.

To explore potential mechanisms for why the impact of coverage on ER utilization is positive for some groups but negative for others, I examine observables. I begin by examining self-reported health. Finkelstein et al. (2012) shows that Medicaid improved self-reported health within the Oregon experiment, so I only compare the self-reported health of groups without Medicaid: compliers who lost the lottery and never takers. I find that 55% of Oregon compliers report fair or poor health, while only 34% of Oregon never takers report fair or poor health. The difference is statistically different from zero, indicating adverse selection on self-reported health via the untreated outcome test. I also find suggestive evidence of adverse selection on self-reported health within Massachusetts. However, the difference between Massachusetts and Oregon is even more striking than the difference within Massachusetts: only 21% of Massachusetts compliers report fair or poor health. These comparisons suggest an important potential mechanism for my findings—upon gaining coverage, individuals in worse health increase their ER utilization, while individuals in better health decrease their ER utilization.

Examination of ER utilization from before the Oregon lottery took place corroborates the role of health as a potential mechanism for why ER utilization increased in Oregon but decreased in Massachusetts. Before the lottery took place, always takers visited the ER more than compliers, who visited the ER more than never takers, indicating adverse selection. Furthermore, when I include previous ER utilization in the Oregon MTE, I can explain the entire decrease in moral hazard from always takers to compliers to never takers. This evidence suggests that differences in previous ER utilization between Oregon compliers and Massachusetts compliers could explain the entire difference between the positive Oregon LATE and the negative Massachusetts LATE. Unfortunately, I do not observe previous ER utilization in my Massachusetts data, so I cannot extrapolate with it directly, but MTE-reweighting effectively allows me to extrapolate based on an unobservable that captures previous ER utilization and health.

When I use available observables for extrapolation, I can still reconcile the Oregon and Massachusetts results with MTE-reweighting, even though I cannot with LATE-reweighting following Angrist and Fernandez-Val (2013) in the tradition of Hotz et al. (2005). I examine the three common observables available for all individuals in the Oregon and Massachusetts data: age, gender, and an indicator for communications in English. These observables are similar across both contexts. When I incorporate them into the Oregon MTE, they do not explain differences in ER utilization across always takers, compliers, and never takers. Furthermore, consistent with evidence from Taubman et al. (2014), I find that treatment effects on compliers are positive within the vast majority of subgroups determined by the common observables. Accordingly, LATE-reweighting would yield a Massachusetts LATE even more positive than the Oregon LATE. In contrast, MTE-reweighting with an MTE that incorporates the common observables can still reconcile the positive LATE in Oregon with the negative LATE in Massachusetts.

2 Model of Selection and Moral Hazard Heterogeneity

2.1 First Stage: Enrollment in Medicaid

Let D represent enrollment in Medicaid, which can be thought of as the "treatment" offered by the Oregon Health Insurance Experiment. Let V_T represent potential utility in the treated state (enrolled in Medicaid, $D = 1$), and let V_U represent potential utility in the untreated state (not enrolled in Medicaid, $D = 0$).[4] The following definition relates realized utility V to the potential utilities:

$$V = V_U + (V_T - V_U)D. \tag{1}$$

I specify the potential utilities as follows:

$$V_T = \mu_T(Z, X) + \nu_T \tag{2}$$
$$V_U = \mu_U(Z, X) + \nu_U, \tag{3}$$

where ν_T and ν_U are unobserved terms with unspecified distributions, X is an optional vector of observed covariates, and Z is an observed binary instrument.[5] The unspecified functions $\mu_T(\cdot)$ and $\mu_U(\cdot)$ translate the covariates and instrument into units of utility. In the Oregon context, the instrument represents the outcome of the randomized lottery. I begin by considering only individuals who participated in the lottery, so $Z \in \{0, 1\}$ for all individuals. Individuals with $Z = 0$ are lottery losers, so I refer to them as the "control group." Individuals with $Z = 1$ are lottery winners that receive the lotteried intervention, an opportunity to be eligible for a lotteried Medicaid program, so I refer to them as the "intervention group." I need different terminology for the intervention group ($Z = 1$) and the treatment group ($D = 1$) because not all Oregon lottery winners enrolled in Medicaid. I assume

A.1. (First Stage Independence) The random variable $\nu_U - \nu_T$ is independent of Z conditional on X, which implies that $F(\nu_U - \nu_T \mid X)$, denoted as U_D, is independent of Z conditional on X.

A.2. (First Stage Technical Assumption) The cumulative distribution function of $\nu_U - \nu_T$ conditional on X, which I denote with F, is continuous and strictly increasing.

These assumptions allow me to derive the following equation for enrollment in Medicaid conditional on the lottery outcome:

$$D = 1\{U_D \leq \mathrm{P}(D = 1 \mid Z = z, X)\}, \tag{4}$$

where $U_D = F(\nu_U - \nu_T \mid X)$. The enrollment equation (4) follows from the statement that individuals enroll in Medicaid if their potential treated utility V_T exceeds their untreated potential utility V_U. To make my exposition of the model self-contained, I present the proof in Appendix A.

[4] Rubin (1974), Rubin (1977), and Holland (1986) are credited with developing the idea of potential outcomes. Here, I use this idea for an unobserved utility V rather than an observed outcome Y. Also, I use subscripts T and U rather than 1 and 0 to emphasize that they represent the treated and untreated states and avoid confusion with the states of winning and losing a lottery.

[5] For expositional clarity, I specify (2) and (3) separately, but for the derivation of the enrollment equation, I only need to specify the difference between them:

$$V_T - V_U = \widetilde{\mu}(Z, X) + \widetilde{\nu}$$

where $\widetilde{\mu}(Z, X) = \mu_T(Z, X) - \mu_U(Z, X)$ and $\widetilde{\nu} = \nu_T - \nu_U$. The terms $\mu_T(Z, X)$ and $\mu_U(Z, X)$ need not be additively separable from each other, and the terms ν_T and ν_U need not be additively separable from each other. Vytlacil (2002) shows that the additive separability of $\widetilde{\mu}(Z, X)$ from $\widetilde{\nu}$ is equivalent to the LATE monotonicity assumption of Imbens and Angrist (1994). Intuitively, an interaction between the two terms could allow winning the lottery to increase enrollment for some individuals and decrease enrollment for others.

The enrollment equation (4) implies that individuals enroll in Medicaid if their value of U_D is less than the threshold $\mathrm{P}(D = 1 \mid Z = z, X)$. The model implies that U_D is distributed uniformly between 0 and 1. For completeness, I present the proof in Appendix B. I interpret U_D as the "(unobserved net) cost of enrollment," such that individuals with the lowest cost of enrollment enroll in Medicaid first. The term "net" emphasizes that U_D represents enrollment costs net of enrollment benefits and therefore that it does not make a material distinction between costs and benefits.

There are two special cases of the enrollment equation (4) for the lottery losers and winners:

$$D = 1\{U_D \leq p_{CX}\} \qquad \text{where } p_{CX} = P\big(D = 1 \mid Z = 0, X\big), \tag{5}$$

$$D = 1\{U_D \leq p_{IX}\} \qquad \text{where } p_{IX} = P\big(D = 1 \mid Z = 1, X\big), \tag{6}$$

where the probabilities p_{CX} and p_{IX} can be estimated in the control group ($Z = 0$) and the intervention group ($Z = 1$), respectively. Without loss of generality, I proceed with $p_{CX} \leq p_{IX}$.[6] To ensure that the lottery is relevant for enrollment, but not necessarily coincident with the lottery outcome, I assume:

A.3. (First Stage Relevance) $\mu_T(Z, X) - \mu_U(Z, X)$ is a nondegenerate random variable conditional on X.

A.4. (First Stage Enrollment Differs from Lottery Outcome with Positive Probability) $0 < \mathrm{P}(D = 1 \mid Z = z, X) < 1$.

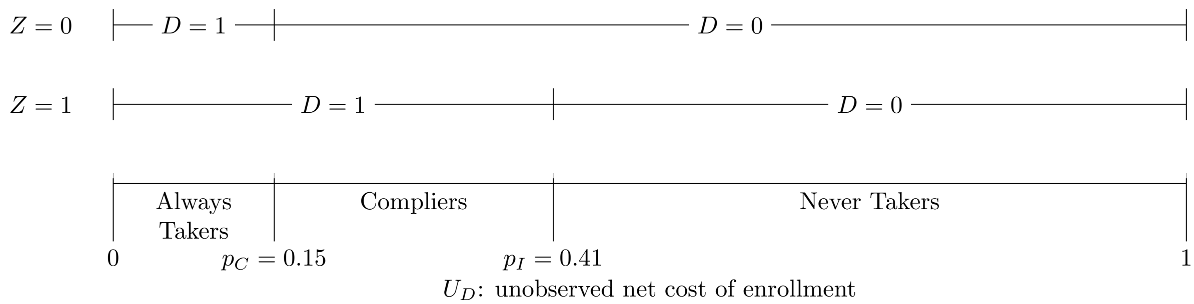
In practice, assumptions A.3 and A.4 are verifiable.

As I show in Figure 1 these assumptions allow me to identify three distinct ranges of the enrollment margin, U_D. As originally shown by Vytlacil (2002), the three ranges of U_D correspond to ranges for always takers, compliers, and never takers (the model excludes defiers). Within my analysis sample, 15% of lottery losers enroll and 41% of lottery winners enroll. Accordingly, in Figure 1, I depict $p_C = 0.15$ and $p_I = 0.41$, suppressing X to emphasize that these quantities are averages in the full analysis sample, not in a sample conditional on X. In the top line of Figure 1, I depict the lottery losers. By (5), I infer that enrolled lottery losers, who must be always takers, have $0 \leq U_D \leq 0.15$. In the middle line of Figure 1, I depict the lottery winners. By (6), I infer that the unenrolled lottery winners, who must be never takers, have $0.41 < U_D \leq 1$. In the bottom line of Figure 1, I depict U_D for lottery losers and winners on the same axis, and I label the implied ranges of U_D for always and never takers. Individuals with values of U_D in the middle range, $0.15 < U_D \leq 0.41$, enroll in Medicaid if they win the lottery, but they do not enroll if they lose the lottery, so these individuals must be compliers.

The depiction of U_D that I show in the bottom line of Figure 1 provides more information than the "first stage" typically reported in the LATE literature, which is equal to $P(D = 1 \mid Z = 1) - P(D = 1 \mid Z = 0) = p_I - p_C$. The first stage conveys the share of compliers, but it does not

[6] If $p_{CX} \geq p_{IX}$, then it is always possible to rename the intervention group as the control group and vice versa so that this condition is satisfied.

Figure 1: Ranges of U_D for Always Takers, Compliers, and Never Takers



convey the separate shares of always takers and never takers. I emphasize that the ordering from always takers to compliers to never takers along U_D is an ordering across an important margin: the margin of enrollment in Medicaid. As Medicaid enrollment expands along the enrollment margin, always takers enroll first, followed by compliers, followed by never takers.

2.2 Second Stage: Relating Enrollment in Medicaid to ER Utilization

I relate Medicaid enrollment D to realized ER utilization Y as follows:

$$Y = Y_U + (Y_T - Y_U)D, \tag{7}$$

where I specify potential ER utilization with Medicaid Y_T and without Medicaid Y_U as follows:

$$Y_T = g_T(X, U_D, \gamma_T) \tag{8}$$

$$Y_U = g_U(X, U_D, \gamma_U), \tag{9}$$

where $g_T(\cdot)$ and $g_U(\cdot)$ are unspecified functions that need not be additively separable in their observed an unobserved components, unlike the potential utility functions in (2) and (3). X is the same optional vector of observed covariates from the first stage of the model, U_D is the unobserved net cost of enrollment from the first stage of the model, and γ_T and γ_U represent additional unobserved heterogeneity in the second stage of the model. I assume:

A.5. (Second Stage Independence) The random vector (U_D, γ_T) and the random vector (U_D, γ_U) are independent of Z conditional on X.

In Appendix C, I use this assumption to calculate the expected values of Y_T for always takers and compliers and the expected values of Y_U for compliers and never takers from the model, consistent with the approaches of Imbens and Rubin (1997), Katz et al. (2001), Abadie (2002), and Abadie (2003), which rely on the LATE assumptions of independence and monotonicity. As a whole, the model, given by the utility equations (1)–(3), the selection equations (4)–(6), the potential outcome equations (7)–(9), and assumptions (A.1)–(A.5) assumes no more than the LATE assumptions, as shown by Vytlacil (2002) and applied by Heckman and Vytlacil (2005).[7]

[7]To ensure that the model assumes no more than the LATE assumptions, I only make the following stylistic

2.3 Selection and Moral Hazard Heterogeneity in the Model

I characterize selection and moral hazard heterogeneity on Y along the entire enrollment margin using functions from the MTE literature (see Carneiro and Lee, 2009; Brinch et al., 2017):

$$\text{Selection and Moral Hazard Heterogeneity:} \quad \text{MTO}(x,p) = \mathrm{E}\left[Y_T \mid X = x, U_D = p\right] \tag{10}$$

$$\text{Selection Heterogeneity:} \quad \text{MUO}(x,p) = \mathrm{E}\left[Y_U \mid X = x, U_D = p\right] \tag{11}$$

$$\text{Moral Hazard Heterogeneity:} \quad \text{MTE}(x,p) = \mathrm{E}\left[Y_T - Y_U \mid X = x, U_D = p\right], \tag{12}$$

where x is a realization of the covariate vector X and p is a realization of the unobserved net cost of enrollment U_D. I refer to the first two functions as the "marginal treated outcome (MTO)" and "marginal untreated outcome (MUO)" functions. Their difference yields the "marginal treatment effect (MTE)" function of Heckman and Vytlacil (1999, 2001, 2005).

I emphasize that the MUO function characterizes selection heterogeneity on Y along the entire enrollment margin and that the MTE function characterizes moral hazard heterogeneity on Y along the entire enrollment margin. The MTO function is the sum of the MUO and MTE functions, so it captures selection plus moral hazard heterogeneity on Y along the entire enrollment margin. At each point p along the enrollment margin, the slope of the MUO function determines whether selection is adverse or advantageous. Similarly, the slope of the MTE function determines whether moral hazard is decreasing or increasing. Uniting insurance terminology with econometric terminology, I emphasize that the MTE function characterizes "moral hazard heterogeneity" on Y along the enrollment margin, which is equivalent to "selection on moral hazard," "selection on the treatment effect," "selection on the gain," and "treatment effect heterogeneity" on Y along the enrollment margin.

Definitions of selection and moral hazard hetergeneity widely used in the econometrics literature, for example, by Angrist (1998) and Heckman et al. (1998), imply the following equations:[8]

$$\text{Selection Heterogeneity from Literature:} \quad \mathrm{E}[Y_U \mid D = 1] - \mathrm{E}[Y_U \mid D = 0] \tag{13}$$

$$\text{Moral Hazard Heterogeneity from Literature:} \quad \mathrm{E}[Y_T - Y_U \mid D = 1] - \mathrm{E}[Y_T - Y_U \mid D = 0]. \tag{14}$$

I can weight the MUO and MTE functions to obtain the definitions from the literature as special

changes to the model presented in Heckman and Vytlacil (2005): (i) I use slightly different notation; (ii) I add equations (2) and (3) to provide an intuitive derivation of the selection equation (4), explaining in footnote 5 that only the *difference* between (2) and (3) matters; (iii) I present A.1 and A.5 as two different assumptions to emphasize that the model relies on independence assumptions in the first and second stages; (iv) I assume A.2 which implies that the distribution of U_D is absolutely continuous with respect to Lebesgue measure, but all proofs hold with the weaker assumption; (v) I assume A.4 which adds a Z term that I believe was omitted from $1 > \mathrm{P}(D = 1 \mid X) > 0$ in Heckman and Vytlacil (2005); (vi) I make U_D explicit in the potential outcome equations (8)-(9); and (vii) I omit the assumption that $\mathrm{E}[Y_T]$ and $\mathrm{E}[Y_U]$ are finite, which will be implied by my ancillary assumptions.

[8] The literature refers to (13) as the "selection bias term" that results from using the observed outcome difference $\mathrm{E}[Y \mid D = 1] - \mathrm{E}[Y \mid D = 0]$ to estimate $\mathrm{E}[Y_T - Y_U \mid D = 1]$, the effect of treatment on the treated of Rubin (1977).

cases. To demonstrate, I express (13) as the following weighted integral of the MUO function:

$$= \int_0^1 \Big[\mathrm{P}(Z=0 \mid D=1)\,\omega(p,0,p_c) + \mathrm{P}(Z=1 \mid D=1)\,\omega(p,0,p_I) \\ - \mathrm{P}(Z=0 \mid D=0)\,\omega(p,p_c,1) - \mathrm{P}(Z=1 \mid D=0)\,\omega(p,p_I,1) \Big] \,\mathrm{MUO}(p)\,\mathrm{d}p \qquad (15)$$

with weights $\omega(p,p_L,p_H) = 1\{p_L \leq p < p_H\}/(p_H - p_L)$. In doing so, I make it clear that selection heterogeneity from the literature depends on the fraction of lottery losers and the fraction of lottery winners, which is not desirable if the goal is to characterize underlying behavior that is invariant to parameters of the experimental design used to study it. Furthermore, the special case of selection heterogeneity from the literature is not identified without ancillary assumptions because the average untreated outcome for always takers is not observed. In contrast, I identify a different policy-relevant special case of selection heterogeneity without ancillary assumptions.

3 Identifying Selection and Moral Hazard Heterogeneity

3.1 The Untreated Outcome Test

I identify a special case of selection heterogeneity along one range of the enrollment margin, U_D, without ancillary assumptions by comparing average ER utilization without Medicaid between compliers ($p_C < U_D \leq p_I$) and never takers ($p_I < U_D \leq 1$):

$$\mathrm{E}[Y_U \mid p_C < U_D \leq p_I] - \mathrm{E}[Y_U \mid p_I < U_D \leq 1] = \int_0^1 (\omega(p,p_C,p_I) - \omega(p,p_I,1))\,\mathrm{MUO}(p)\,\mathrm{d}p, \qquad (16)$$

where $\omega(p,p_L,p_H) = 1\{p_L \leq p < p_H\}/(p_H - p_L)$. I refer to the test of the null hypothesis that (16) is equal to zero as the "untreated outcome test," which is refutable and verifiable with a large enough sample. This test is equivalent to tests proposed in the econometric literature by Bertanha and Imbens (2014), Guo et al. (2014), Black et al. (2015), and Mogstad et al. (2017). My innovation is in interpretation. By relating the test to the marginal untreated outcome function from the MTE model, I show that that a rejection of the untreated outcome test identifies selection heterogeneity. My innovation relative to the insurance literature is that by modeling the Oregon experiment with the MTE model, I show that it is possible to test for selection heterogeneity within the Oregon experiment. I emphasize that the untreated outcome test is equivalent to the Einav et al. (2010) cost curve test for selection heterogeneity applied to uninsured costs.

3.2 The Treated Outcome Test

Similarly, I can also compare average ER utilization with Medicaid between always takers ($0 \leq U_D \leq p_C$) and compliers ($p_C < U_D \leq p_I$) without ancillary assumptions:

$$\mathrm{E}[Y_T \mid 0 \leq U_D \leq p_C] - \mathrm{E}[Y_T \mid p_C < U_D \leq p_I] = \int_0^1 (\omega(p,0,p_C) - \omega(p,p_C,p_I))\,\mathrm{MTO}(p)\,\mathrm{d}p, \qquad (17)$$

where $\omega(p, p_L, p_H) = 1\{p_L \leq p < p_H\}/(p_H - p_L)$. I refer to the test of the null hypothesis that (17) is equal to zero as the "treated outcome test." This test is equivalent to tests proposed in the econometric literature by Bertanha and Imbens (2014), Guo et al. (2014), Black et al. (2015), and Mogstad et al. (2017). It is also equivalent to the Einav et al. (2010) cost curve test from the insurance literature when applied to insured costs. By relating the treated outcome test to the marginal treated outcome function from the MTE model, I show that a rejection of the treated outcome test implies moral hazard heterogeneity, or selection heterogeneity, or both types of heterogeneity, between always takers and compliers.

I emphasize that the distinction between the treated and the untreated is a meaningful one, especially in insurance contexts where the distinction separates the insured and the uninsured. It is tempting to assert that the treated and untreated outcome tests must be equally informative because it is possible to rename the treated group as the untreated group and vice versa. However, changing the definition of the treatment also changes the definition of the treatment effect. The treatment effect is $Y_T - Y_U$, not $|Y_T - Y_U|$, so the distinction between treated and untreated matters. The untreated outcome test cannot identify treatment effect heterogeneity. The treated outcome test can reflect treatment effect heterogeneity, but it cannot identify treatment effect heterogeneity without ancillary assumptions.

Other tests from the econometric literature also cannot identify treatment effect heterogeneity without ancillary assumptions. Tests from the econometric literature can often be expressed in terms of functions of the treated *and* untreated outcome tests (see the discussion in Bertanha and Imbens, 2014). I emphasize that without ancillary assumptions, the joint treated and untreated outcome test is no more informative about treatment effect heterogeneity than the treated outcome test is on its own.

Tests from the insurance literature also cannot separately identify moral hazard heterogeneity in the context of the Oregon experiment without ancillary assumptions. The cost curve test of Einav et al. (2010) applied to insured costs is equivalent to the treated outcome test. Therefore, it identifies heterogeneous moral hazard, or heterogeneous selection, or both. The Einav et al. (2010) cost curve test applied to the difference between insured and uninsured costs would identify heterogeneous moral hazard, but it cannot be run within the Oregon experiment without ancillary assumptions because only one point on the curve is identified. Other tests from the insurance literature also cannot separately identify moral hazard heterogeneity within the Oregon experiment without ancillary assumptions. Under the Chiappori and Salanie (2000) positive correlation test, a correlation between insurance coverage and insured spending could indicate heterogeneous moral hazard, or heterogeneous selection, or both. Under the Finkelstein and Poterba (2014) unused observables test, a correlation between a covariate and insurance coverage and a second correlation between the same covariate and insured spending could also indicate heterogeneous moral hazard, or heterogeneous selection, or both.

4 Extrapolation Using Selection and Moral Hazard Heterogeneity

Using the untreated outcome test, I can identify selection heterogeneity along a range of the Medicaid enrollment margin. Using the treated outcome test, I can identify possible moral hazard heterogeneity along another range of the enrollment margin, but I cannot separate it from selection heterogeneity. To predict how moral hazard would change under a wide range of policies, I impose ancillary assumptions to identify how moral hazard changes along the entire enrollment margin.

4.1 Under Ancillary Assumptions

I begin by imposing the following ancillary assumptions following Brinch et al. (2017), who impose AA.1 and AA.2 to examine the impact of family size on child outcomes, and Olsen (1980), who imposes an assumption equivalent to AA.1 to examine the impact of family size on maternal outcomes:

AA.1. (Linearity of $\mathrm{MTO}(p)$ in p) From (8), let $Y_T = \alpha_T + \theta_T U_D + \gamma_T$, where $\mathrm{E}[\gamma_T \mid U_D = p] = 0$. Therefore, $\mathrm{MTO}(p) = \mathrm{E}[Y_T \mid U_D = p] = \alpha_T + \theta_T p$.

AA.2. (Linearity of $\mathrm{MUO}(p)$ in p) From (9), let $Y_U = \alpha_U + \theta_U U_D + \gamma_U$, where $\mathrm{E}[\gamma_U \mid U_D = p] = 0$. Therefore, $\mathrm{MUO}(p) = \mathrm{E}[Y_U \mid U_D = p] = \alpha_U + \theta_U p$.

In my context, these assumptions require that potential ER utilization varies linearly with the fraction of the sample enrolled in Medicaid conditional on the lottery outcome. These assumptions are informed by the model because they relate the unobserved net cost of enrollment U_D in the first stage to potential outcomes in the second stage. Under these assumptions, the variation that identifies the untreated outcome test identifies the intercept and slope of the MUO function, and the variation that identifies the treated outcome test identifies the intercept and slope of the MTO function.

The combination of AA.1 and AA.2 implies the following linear MTE function:

$$\mathrm{MTE}(p) = \mathrm{E}[Y_T - Y_U \mid U_D = p] = (\alpha_T - \alpha_U) + (\theta_T - \theta_U)p.$$

In my context, linearity of the MTE function implies that moral hazard varies linearly with the fraction of the sample enrolled in Medicaid conditional on the lottery outcome. Linearity of the MTE function has precedent as a direct assumption in applied work (see Moffitt, 2008; French and Song, 2014). Applied work that extrapolates to other policies using the LATE also makes an implicit assumption that the MTE function is linear and has zero slope.

Under AA.1 and AA.2, the slope of the linear MTE function identifies moral hazard heterogeneity on Y along the entire enrollment margin. Under AA.2, the slope of the linear MUO function identifies selection heterogeneity on Y along the entire enrollment margin. Therefore, under AA.1 and AA.2, selection and moral hazard heterogeneity are separately identified along the entire enrollment margin.

Using selection and moral hazard heterogeneity along the enrollment margin, I can extrapolate to predict the impacts of a wide range of policies. To do so, I integrate the weighted MTE function and its component MTO and MUO functions over a general range of the enrollment margin $p_L < U_D \leq p_H$ as follows:

$$\mathrm{E}\left[Y_T \mid p_L < U_D \leq p_H\right] = \int_0^1 \omega(p, p_L, p_H)\mathrm{MTO}(p)\,\mathrm{d}p \tag{18}$$

$$\mathrm{E}\left[Y_U \mid p_L < U_D \leq p_H\right] = \int_0^1 \omega(p, p_L, p_H)\mathrm{MUO}(p)\,\mathrm{d}p \tag{19}$$

$$\mathrm{E}\left[Y_T - Y_U \mid p_L < U_D \leq p_H\right] = \int_0^1 \omega(p, p_L, p_H)\mathrm{MTE}(p)\,\mathrm{d}p, \tag{20}$$

using weights $\omega(p, p_L, p_H) = 1\{p_L < p \leq p_H\}/(p_H - p_L)$. These weights are special cases of general weights for MTE-reweighting given by Heckman and Vytlacil (2007). Unlike the weights used by Brinch et al. (2017), these weights allow me to recover exact values of observed average outcomes for always takers ($0 \leq U_D \leq p_C$), compliers ($p_C < U_D \leq p_I$), and never takers ($p_I < U_D \leq 1$), which are of interest for extrapolation to specific policies. To predict the impact of specific policies that enroll all never takers or dis-enroll all always takers, I only need one of the two ancillary assumptions.

4.2 Incorporating Observables

When extrapolating from the Oregon experiment to the Massachusetts reform, it might be desirable to assess the impact of adjusting for observable differences between the two populations. To do so, I incorporate observables directly into the linear MUO and MTO functions via a shape restriction commonly used in the MTE literature (see Brinch et al., 2017; Carneiro and Lee, 2009; Carneiro et al., 2011; Maestas et al., 2013). In my context, the shape restriction requires that included observables and the remaining unobserved net cost of Medicaid enrollment have additively-separable impacts on potential ER utilization with and without Medicaid:

AA.3. (Linearity of $\mathrm{MTO}(x, p)$ in p) From (8), let $Y_T = \beta_T' X + \delta_T U_D + \xi_T$ where $\mathrm{E}\left[\xi_T \mid X = x, U_D = p\right] = 0$. Therefore, $\mathrm{MTO}(x, p) = \mathrm{E}\left[Y_T \mid X = x, U_D = p\right] = \beta_T' x + \delta_T p$.

AA.4. (Linearity of $\mathrm{MUO}(x, p)$ in p) From (9), let $Y_U = \beta_U' X + \delta_U U_D + \xi_U$ where $\mathrm{E}\left[\xi_U \mid X = x, U_D = p\right] = 0$. Therefore, $\mathrm{MUO}(x, p) = \mathrm{E}\left[Y_U \mid X = x, U_D = p\right] = \beta_U' x + \delta_U p$.

Assumptions AA.3 and AA.4 imply the following MTE function:

$$\mathrm{MTE}(x, p) = \mathrm{E}\left[Y_T - Y_U \mid X = x, U_D = p\right] = (\beta_T - \beta_U)' x + (\delta_T - \delta_U) p.$$

I present an algorithm for estimation of these functions that simplifies the Heckman et al. (2006) algorithm in Appendix D.[9] I reweight these functions using the same approach that I use in (19)–

[9]For inference, I bootstrap using 200 replications, and I report the standard deviation as the standard error or the 2.5 and 97.5 percentiles as the 95% confidence interval.

(20). Departing from the literature, which places little emphasis on the choice of observables, I use MTE-reweighting to determine which observables explain moral hazard heterogeneity and are thus more important for extrapolation.

5 Results

Starting from the premise that the Oregon experiment is the "gold standard," my focus is to use selection and moral hazard heterogeneity from within the experiment to predict the impact to other policies. The focus of my work is not to evaluate the design or implementation of the lottery, which has been discussed in Baicker et al. (2013, 2014); Taubman et al. (2014), and Finkelstein et al. (2016). I am able to replicate results from Taubman et al. (2014), limited only by minor changes made to the publicly available data to limit identification of individuals with large and uncommon numbers of ER visits. The probability of winning the lottery was only random conditional on the number of entrants from the household, so I conduct my analysis using all individuals that were the only members of their household to enter, resulting in a sample size of 19,643 individuals with administrative data on their visits to the ER.

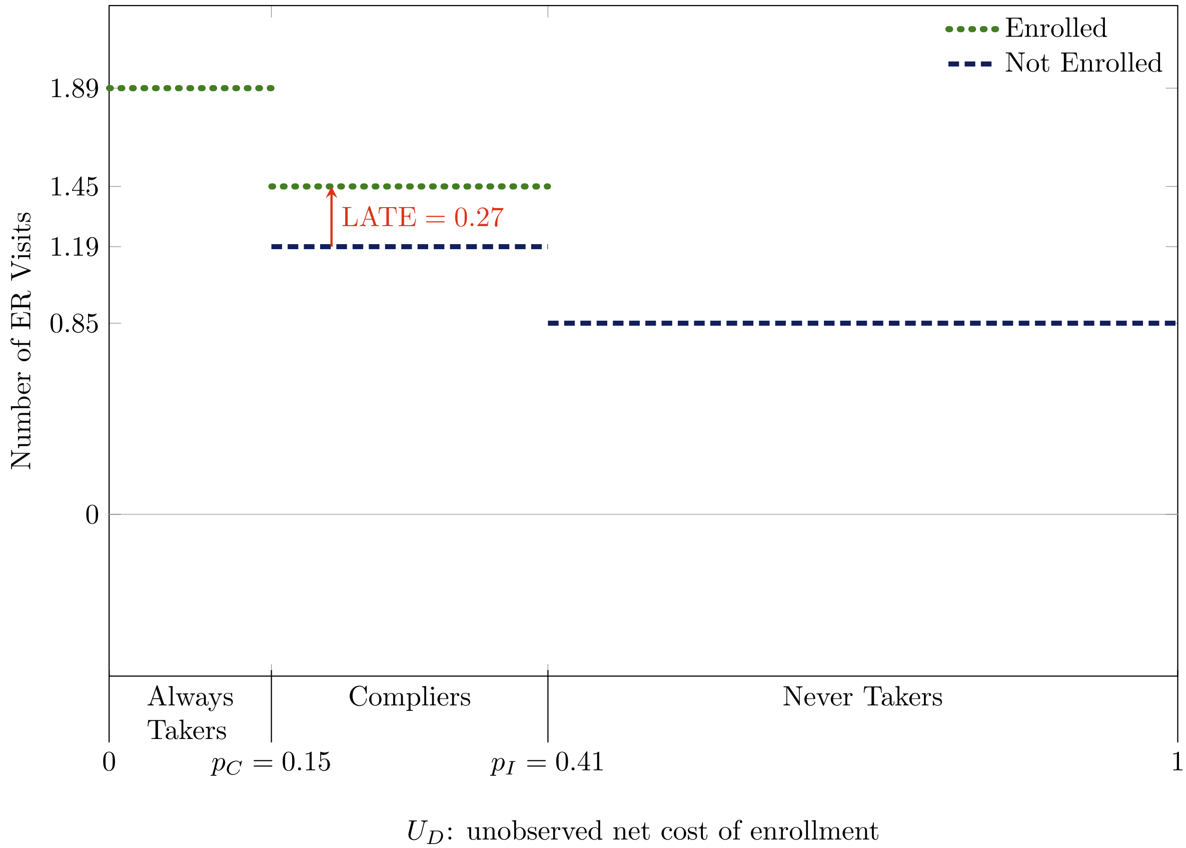
5.1 Identifying Selection and Moral Hazard Heterogeneity

As I depict along the horizontal axis of Figure 2, the sample consists of 15% always takers, 26% compliers, and 59% never takers. Along the vertical axis, I show that during the study period from March 10, 2008 to September 30, 2009, always takers visited the ER 1.89 times, compliers visited 1.45 times if enrolled and 1.19 times if not, and never takers visited 0.85 times. The difference in visits between enrolled and unenrolled compliers is equal to the LATE, as shown by Imbens and Rubin (1997). I depict the LATE with an arrow to indicate its magnitude and direction. The estimated LATE of 0.27 is consistent the headline finding of Taubman et al. (2014), who show that moral hazard is positive—Medicaid increases ER utilization—for compliers.

Figure 2 provides more information than the LATE. As originally shown by Angrist (1990); Angrist and Krueger (1992), the calculation of the LATE does not require the ability to calculate averages of ER utilization conditional on Medicaid enrollment. Using the Wald (1940) approach, the reduced form $\mathrm{E}[Y|Z=1]-\mathrm{E}[Y|Z=0]$ is equal to 0.07, and the first stage $\mathrm{E}[D|Z=1]-\mathrm{E}[D|Z=0]$ is equal to 0.26. Dividing the reduced form divided by the first stage yields a LATE of 0.27 visits, which is equal to the LATE reported in Figure 2.

The depiction in Figure 2 emphasizes that the LATE need not apply to always or never takers. There could be reason to question whether the LATE applies to always and never takers if their average outcomes are different from those of compliers. I compare average outcomes with the untreated and treated outcome tests.

Figure 2: Number of ER Visits for Always Takers, Compliers, and Never Takers



Note. The number of ER visits represents the total number of visits to the emergency department during the study period from March 10, 2008 to September 30, 2009. p_C is the probability of enrollment in the control group, and p_I is the probability of enrollment in the intervention group.

5.1.1 The Untreated Outcome Test

The untreated outcome test provides evidence of adverse selection across compliers and never takers in Oregon. When not enrolled in Medicaid, compliers visit the ER an average of 1.19 times, while never takers visit 0.85 times. The average difference of 0.34 visits, reported as the result of the untreated outcome test in Table 1, is statistically different from zero. The difference cannot reflect treatment from Medicaid since both groups are not enrolled. Under the model, compliers enroll in Medicaid before never takers, so the difference identifies adverse selection into Medicaid along this range of the enrollment margin. Given adverse selection, I predict that enrollees in subsequent expansions would use the ER less on average in the absence of Medicaid than current enrollees.

5.1.2 The Treated Outcome Test

When enrolled in Medicaid, always takers visit the ER an average of 1.89 times while compliers visit an average of 1.45 times. The average difference of 0.44 visits, reported as the result of the treated outcome test in Table 1, is statistically different from zero. Under the model, always takers enroll

Table 1: Number of ER Visits for Always Takers, Compliers, and Never Takers

	Mean				
	(1) Always Takers	(2) Compliers	(3) Never Takers	Untreated Outcome Test (2) - (3)	Treated Outcome Test (1) - (2)
Number of ER Visits					
Enrolled	1.89	1.45	0.55		0.44
	(0.08)	(0.11)	(0.45)		(0.17)
Not Enrolled	1.35	1.19	0.85	0.34	
	(0.17)	(0.11)	(0.03)	(0.13)	
Treatment Effect	0.54	0.27	-0.29		
	(0.19)	(0.15)	(0.45)		

Note. Bootstrapped standard errors are in parentheses. The shaded cells report extrapolated values from MTE-reweighting via (18)–(20) for enrolled individuals (N=4,725) and individuals who are not enrolled (N=14,897). The number of ER visits represents the total number of visits to the emergency department during the study period from March 10, 2008 to September 30, 2009.

in Medicaid before compliers, so their greater visits with Medicaid could reflect adverse selection, decreasing moral hazard, or both. Without ancillary assumptions, I cannot separate selection heterogeneity from moral hazard heterogeneity. However, based on the combined influence of selection and moral hazard heterogeneity, I predict that ER utilization per Medicaid enrollee would decrease if the Medicaid expansion implemented with the Oregon lottery were repealed.

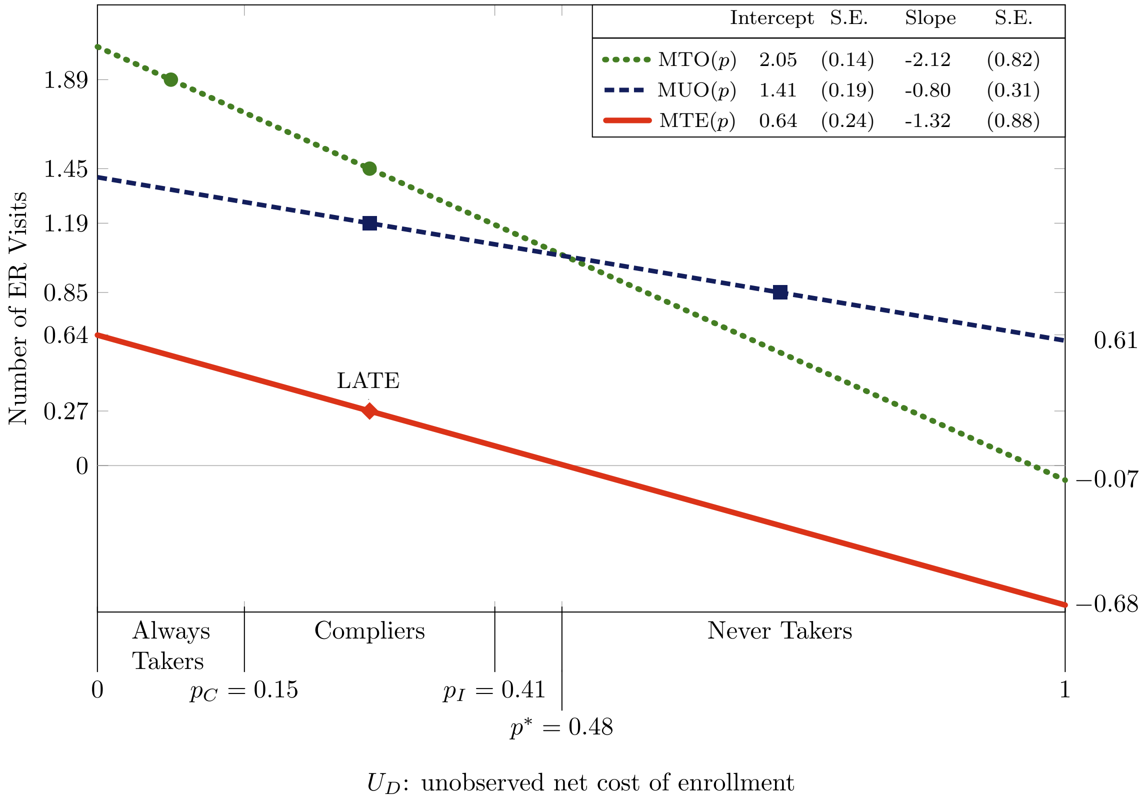
5.2 Extrapolation Under Ancillary Assumptions

5.2.1 Oregon MTE(p)

Although I can make some policy predictions using the results of the treated and untreated outcome tests, I impose ancillary assumptions to predict the moral hazard impact of a wider set of policies that induce changes in enrollment. Figure 3 depicts the MTO, MUO, and MTE functions under ancillary assumptions AA.1 and AA.2. On the vertical axis, the two points labeled with circular markers indicate the average outcomes of always takers and enrolled compliers, which fall at the median of the support for each group on the horizontal axis. These two points identify the intercept and slope of the MTO function, depicted with a dotted line. The two points labeled with square markers identify the intercept and slope of the MUO function, depicted with a dashed line. I depict the MTE function, the vertical difference between the MTO and MUO functions, with a solid line. As shown, the MTE function is positive for low levels of enrollment and negative for high levels of enrollment, even though the LATE is positive.

For extrapolation to specific policies that dis-enroll all always takers or enroll all never takers, I report values that I obtain via reweighting with (18)-(20) in the shaded cells of Table 1. I only observe always takers when enrolled in Medicaid, and they visit the ER 1.89 times. Extrapolation from observed ER utilization for unenrolled compliers and never takers indicates that if always

Figure 3: MTO(p), MUO(p), and MTE(p)



Note. Bootstrapped standard errors are in parentheses. The number of ER visits represents the total number of visits to the emergency department during the study period from March 10, 2008 to September 30, 2009. p_C is the probability of enrollment in the control group, and p_I is the probability of enrollment in the intervention group.

takers were not enrolled in Medicaid, they would visit the ER 1.35 times, such that the treatment effect of Medicaid for always takers is an increase of 0.54 visits. In contrast, extrapolation from enrolled compliers and always takers shows that Medicaid *decreases* the average average ER utilization of never takers by 2.9 visits. This finding suggests that if Medicaid were extended to all never takers, their ER utilization would decrease. Accordingly, a hypothetical policy that builds on the 2008 Oregon policy by enrolling all lottery entrants in Medicaid would decrease ER utilization by 0.16 per person—the weighted average treatment effect for all untreated compliers and never takers in the sample.[10]

5.2.2 Comparison to Massachusetts MTE(p)

Before extrapolating to Massachusetts, I assess whether the MTE within Oregon applies within Massachusetts. I begin by recasting my previous work on the Massachusetts reform from Hackmann et al. (2015), which builds on the Einav et al. (2010) model from the insurance literature, in

[10] The weighted average treatment effect for all untreated compliers and never takers in the sample is given by:

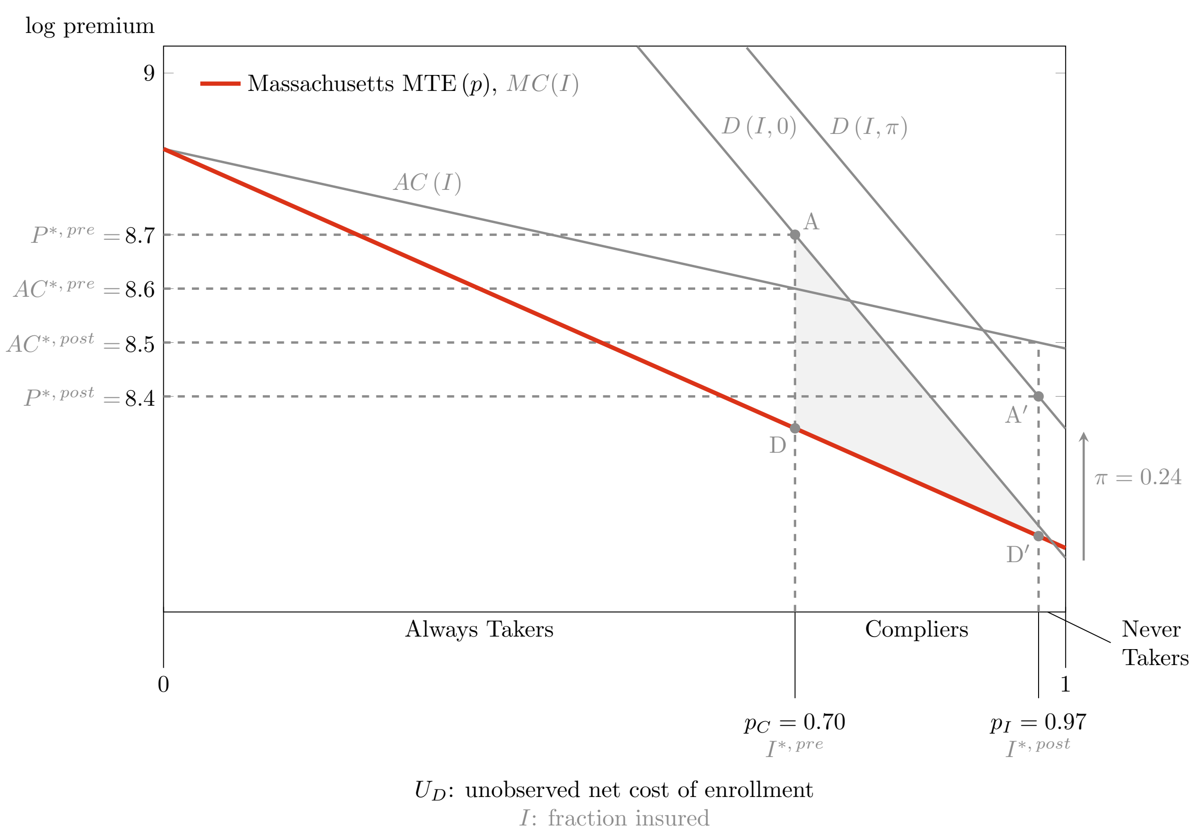
$$\frac{\mathrm{P}(Z=0)(p_I - p_C) * \mathrm{E}[Y_T - Y_U \mid p_C < U_D \leq p_I] + (1-p_I) * \mathrm{E}[Y_T - Y_U \mid p_I < U_D \leq 1]}{\mathrm{P}(Z=0)(p_I - p_C) + (1-p_I)}.$$

Using this equation, I obtain (0.27*0.66*(0.41-0.15)+(-0.29)*(1.00-0.41))/(0.66*(0.41-0.15)+(1.00-0.41))= -0.16.

terms of the MTE model with ancillary assumptions AA.1 and AA.2. To do so, in Figure 4, I reproduce Figure 8 from Hackmann et al. (2015) using notation consistent with the MTE model while presenting notation from the original figure in a lighter typeface. In Hackmann et al. (2015), the Massachusetts reform shifts the demand curve upward by the individual mandate penalty π and shifts the average cost curve downward by the change in markups. Equilibrium occurs at the intersection of the demand curve and the average cost curve net of markups, which occurs at $I^{*,pre} = p_C$ before the reform and $I^{*,post} = p_I$ after the reform. Log premiums before and after the reform, depicted by A and A', identify the slope of the linear demand curve, given the penalty π. Marginal costs before and after the reform, depicted by D and D', identify the slope of the linear marginal cost curve.

In Hackmann et al. (2015), as in Einav et al. (2010), the marginal cost curve is obtained from the difference between the marginal insurer spending on the insured and marginal insurer spending on the uninsured. Therefore, the marginal cost curve represents a Massachusetts MTE function in terms of the log premium. This function, like the Oregon MTE function in terms of the number of ER visits, is downward sloping, indicating that in both contexts, moral hazard decreases as the fraction enrolled increases. I do not observe ER costs or visits in the Hackmann et al. (2015) data, but evidence from the Oregon experiment shows that ER costs and total health care costs are

Figure 4: Figure 8 from Hackmann et al. (2015) Recast as Massachusetts MTE(p)

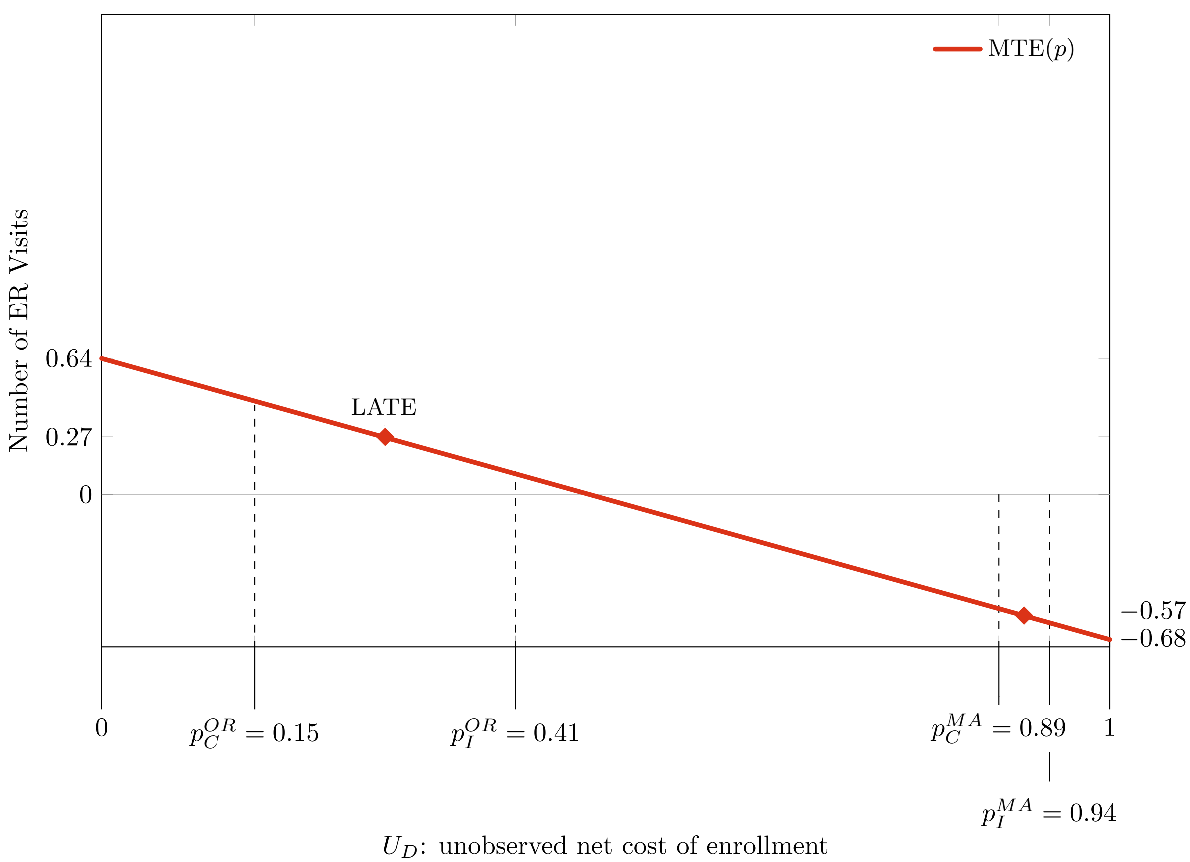


complements (Taubman et al., 2014).

5.2.3 MTE-Reweighting from Oregon MTE(p) to Massachusetts

Returning to the Oregon data because the Oregon experiment is the "gold standard," I characterize the Massachusetts reform as a subsequent expansion of Medicaid along the Oregon MTE. To do so, I reproduce the Oregon MTE in Figure 5. I label the probability of health insurance coverage in Massachusetts before the reform as $p_C^{MA} = 0.89$ and after the reform as $p_I^{MA} = 0.94$. I obtain these values from the Behavioral Risk Factor Surveillance System (BRFSS) data that I used to study the Massachusetts reform in Kolstad and Kowalski (2012). Unlike the Hackmann et al. (2015) data, which only capture enrollment in the individual health insurance market, the BRFSS data capture enrollment in the entire state. It is important to capture enrollment in the entire state for comparison to the literature on the impact of the Massachusetts reform on ER utilization (Chen et al., 2011; Kolstad and Kowalski, 2012; Miller, 2012; Smulowitz et al., 2011).

Figure 5: Extrapolation of MTE(p) to Massachusetts



Note. The number of ER visits represents the total number of visits to the emergency department during the study period from March 10, 2008 to September 30, 2009. p_C^{OR} is the probability of enrollment in the control group in Oregon, p_I^{OR} the probability of enrollment in the intervention group in Oregon, p_C^{MA} the probability of enrollment in the control group in the Massachusetts reform, and p_I^{MA} the probability of enrollment in the intervention group in the Massachusetts reform.

As shown, enrollment levels before and after the Massachusetts reform would entail enrollment of a subset of never takers in Oregon. Therefore, application of the Oregon MTE to Massachusetts implies that Massachusetts compliers are comparable to a subset of Oregon never takers in terms of their unobserved net cost of enrollment U_D. There is a case to be made that the Oregon sample is actually a subset of the Massachusetts sample along the lower range of U_D because all individuals in the Oregon sample entered a lottery for Medicaid. Therefore, it is likely conservative to compare Massachusetts compliers to this particular subset of Oregon never takers.

MTE-reweighting the Oregon MTE via (20) over the range from $p_C^{MA} = 0.89$ to $p_I^{MA} = 0.94$, I predict that the Massachusetts reform should have decreased ER visits by an average of 0.57 visits among Massachusetts compliers. Miller (2012) finds that insurance enrollment induced by the Massachusetts reform decreased ER visits by 0.67 to 1.28 visits per person per year, depending on the empirical strategy. The decrease that I find over the 19 months from March 10, 2008 to September 30, 2009 translates into a decrease of 0.36 visits per person per year (=(0.57/19)*12), which is smaller than her estimates but of the same order of magnitude. Therefore, my extrapolations can reconcile the increase in ER utilization in Oregon with the decrease in ER utilization in Massachusetts using only variation in the unobserved net cost of enrollment U_D.

5.3 Extrapolation Incorporating Observables

5.3.1 Self-Reported Health

To explore potential mechanisms for why the impact of coverage on ER utilization is positive for some groups but negative for others, I examine observables. I begin by examining self-reported health. I observe self-reported health for almost all individuals in the Massachusetts BRFSS data, and I observe self-reported health for a subset of individuals in the Oregon administrative data who were surveyed. Using the Oregon data, Finkelstein et al. (2012) shows that Medicaid improved self-reported health, so I only compare the self-reported health of groups without Medicaid: compliers who lost the lottery and never takers. As I describe in Appendix C, I obtain the average probability that individuals in these groups reported fair or poor health.

As shown in Table 2, within Oregon and Massachusetts, I find that never takers are less likely to be in fair or poor health than compliers who are not enrolled in Medicaid, consistent with adverse selection via the untreated outcome test. However, differences in self-reported health are more striking across both contexts than they are within each context. As I show in Table 2, 55% of Oregon compliers report fair or poor health, while only 34% of Oregon never takers report fair or poor health. In stark contrast, only 21% of Massachusetts compliers report fair or poor health. These comparisons suggest an important potential mechanism for my findings. Upon gaining coverage, individuals in worse health (Oregon compliers) increase their ER utilization, while individuals in better health (Oregon never takers and Massachusetts compliers) decrease their ER utilization.

Table 2: Always Takers, Compliers, and Never Takers: Oregon vs. Massachusetts

	Means				Difference in Means	
		(1)	(2)	(3)		
	All	Always Takers	Compliers	Never Takers	(1) - (2)	(2) - (3)
Oregon Health Insurance Experiment of 2008						
Fair or Poor Health, Not Enrolled[a]	0.42	-	0.55	0.34	-	0.20
	(0.01)		(0.03)	(0.01)		(0.04)
Number of Pre-period ER Visits	0.87	1.36	0.88	0.73	0.48	0.15
	(0.01)	(0.05)	(0.07)	(0.03)	(0.09)	(0.09)
Common Observables						
Age	40.69	39.45	42.41	40.25	-2.96	2.16
	(0.09)	(0.29)	(0.41)	(0.19)	(0.53)	(0.57)
Female	0.56	0.72	0.53	0.53	0.19	0.003
	(0.003)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
English	0.91	0.90	0.92	0.91	-0.02	0.01
	(0.002)	(0.01)	(0.01)	(0.004)	(0.01)	(0.01)
N	19,643	2,986	5,092	11,565		
Massachusetts Health Reform of 2006						
Fair or Poor Health, Not Enrolled[a]	0.19	-	0.21	0.18	-	0.03
	(0.02)		(0.03)	(0.01)		(0.04)
Common Observables						
Age	42.00	42.15	42.42	38.98	-0.26	3.43
	(0.086)	(0.12)	(1.41)	(0.49)	(1.49)	(1.57)
Female	0.51	0.52	0.43	0.38	0.10	0.04
	(0.003)	(0.004)	(0.05)	(0.02)	(0.05)	(0.06)
English	0.96	0.98	0.86	0.81	0.12	0.05
	(0.001)	(0.001)	(0.02)	(0.02)	(0.02)	(0.04)
N	62,456	55,966	3,175	3,314		

Note. Bootstrapped standard errors are in parentheses. Data for the Massachusetts health reform are taken from pooled annual samples of the Behavioral Risk Factor Surveillance System (BRFSS) from years 2004–2009 and restricted to ages 21–64 (the age range of the Oregon sample). For the Massachusetts health reform, treatment is an indicator that equals one for individuals with any form of health insurance ("Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare?"). The instrument is an indicator that equals one in the post-period of the expansion on and after July 2007. "Age" is measured in year 2008 for the Oregon Health Insurance Experiment and in year 2006 for the Massachusetts health reform. "Female" is a binary indicator for the gender of the respondent. "English" is a binary indicator that equals one for individuals in the Oregon Health Insurance Experiment who requested materials in English and that equals one for individuals in the BRFSS who completed the interview in English. The number of pre-period visits is measured before the study period from January 1, 2007 to March 9, 2008. "Fair or Poor Health" equals one when individuals self-report having fair or poor health on a 5-point scale. [a]Number of observations in the Oregon Health Insurance Experiment with nonmissing self-reported health: 5,833. Number of observations in the BRFSS with nonmissing self-reported health: 62,161.

5.3.2 Previous ER Utilization

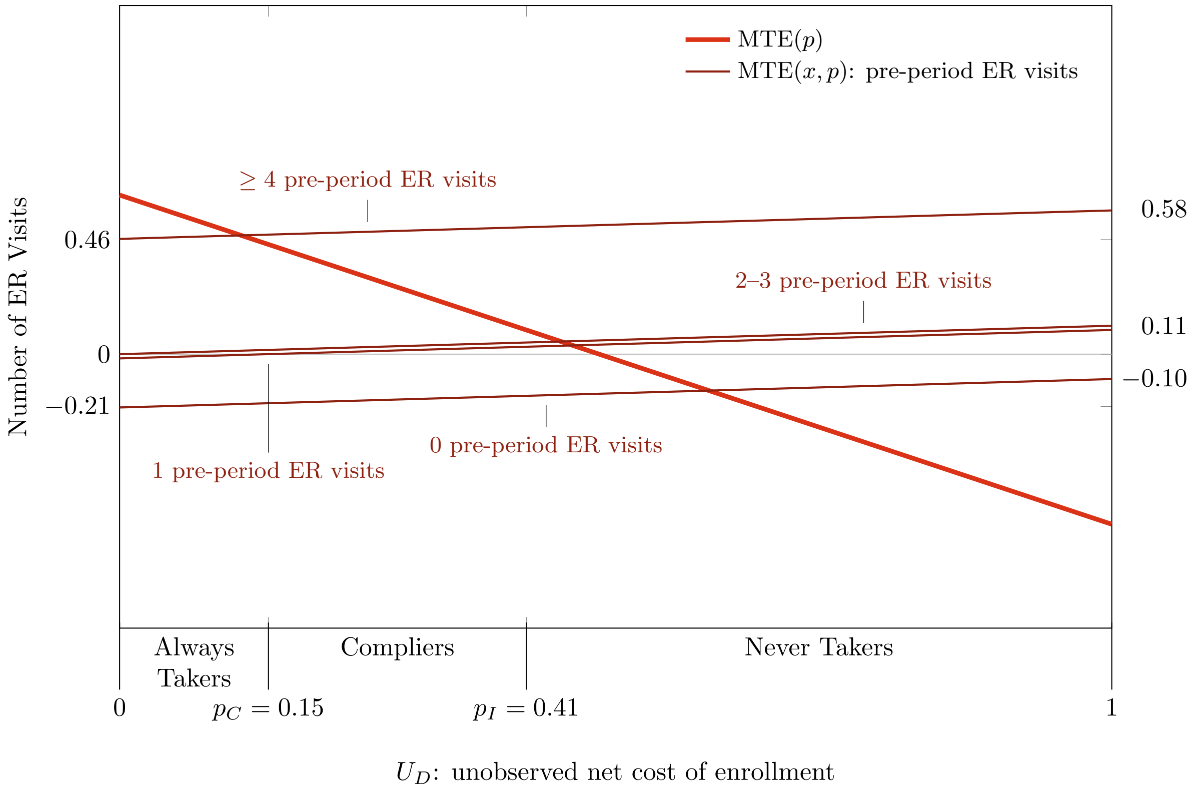
Examination of ER utilization from before the Oregon lottery took place corroborates the role of health as a potential mechanism for why ER utilization increased in Oregon but decreased in Massachusetts. For each individual in the Oregon administrative data, I observe the total number of pre-period ER visits from January 1, 2007, to March 9, 2008. I report the average number of per-period ER visits for always takers, compliers, and never takers in Table 2, calculated as described in Appendix C. Always takers visited the ER an average of 1.36 times, while compliers visited an average of 0.88 times, and never takers visited an average of 0.73 times. The monotonic relationship in previous ER utilization across these groups indicates adverse selection on previous ER utilization: individuals with larger previous ER utilization are more likely to enroll in Medicaid.

I also find heterogeneous moral hazard by previous ER utilization: individuals with larger previous ER utilization are more likely to increase their ER utilization upon Medicaid enrollment. Indeed, I find that previous ER utilization can explain the entire decrease in moral hazard from always takers to compliers to never takers. There is substantial variation in pre-period ER utilization: 66% of individuals have zero visits, 17% have one visit, 11% have 2 to 3 visits, and 6% have 4 or more visits in the pre-period. By incorporating controls for each of these visit ranges into the MTE, I obtain a separate MTE(x, p) for each range. As depicted in Figure 6, the MTE(p) function, which does not incorporate observables, has a pronounced downward slope, indicating substantial unexplained heterogeneity in moral hazard. However, when I incorporate controls for previous ER utilization into the MTE(x, p) function, I eliminate the negative slope, indicating that the unexplained heterogeneity in moral hazard can be explained by previous ER utilization.

Looking beyond the slope of the MTE(x, p) function to its level at various values of pre-period ER visits reveals a clear monotonic relationship between pre-period ER visits and the treatment effect of Medicaid enrollment on subsequent ER visits. As depicted in Figure 6, the MTE(x, p) for individuals with 4 or more pre-period visits is always positive, and the MTE(x, p) for individuals with zero pre-period visits is always negative. Intuitively, individuals with high numbers of ER visits in the pre-period increase their ER utilization upon gaining coverage, while individuals with zero ER visits in the pre-period decrease their ER utilization upon gaining coverage.

Individuals induced to gain coverage through the Massachusetts reform had better health and presumably lower ER utilization than individuals induced to gain coverage through the Oregon experiment, which could explain why ER utilization decreased in Massachusetts but decreased in Oregon. Unfortunately, I do not observe previous ER utilization in my Massachusetts data. Furthermore, data from published studies that examine the impact of the Massachusetts health reform on ER utilization are either not available at the individual level or they only include individuals who visit a hospital or emergency room, making them unsuitable for comparison to data on pre-period ER visits in Oregon (Chen et al., 2011; Kolstad and Kowalski, 2012; Miller, 2012; Smulowitz et al., 2011). Therefore, I cannot weight the Oregon MTE(x, p) to reflect previous ER utilization in Massachusetts, however MTE-reweighting with with MTE(p) effectively allows me to extrapolate based on an unobservable that captures previous ER utilization and health.

Figure 6: MTE(x,p) with Previous ER Utilization



Note. The number of ER visits represents the total number of visits to the emergency department during the study period from March 10, 2008 to September 30, 2009. Pre-period ER visits refers to a group of indicators for visiting the ER 0 times, 1 time, 2–3 times, and 4 or more times during the pre-period from January 1, 2007 to March 9, 2008. p_C is the probability of enrollment in the control group, and p_I is the probability of enrollment in the intervention group. In this figure, the function for 1 pre-period ER visits has been shifted downward slightly to make it easier to discern from the function for 2–3 pre-period ER visits.

5.3.3 LATE-Reweighting with Common Observables

To explore the potential for extrapolation with observables alone to reconcile the increase in ER utilzation in Oregon with the decrease in Massachusetts, I begin by considering LATE-reweighting, which relies only on observables. The three common observables available for all individuals in the Massachusetts BRFSS data and the Oregon administrative data are age, gender, and an indicator for communications in English. As I show in Table 2, the common observables are similar in both samples. Such similarity is plausible even though Massachusetts is more urban than Oregon because the Oregon administrative ER data are only drawn from the Portland area. Common observables are also generally similar for Oregon and Massachusetts compliers, suggesting that LATE-reweighting will not yield a Massachusetts LATE that differs dramatically from the Oregon LATE.

To examine variation in the common observables available for LATE-reweighting, I use each common observable to divide the sample into two subgroups, and I report LATEs within each subgroup in Table 3. As shown, the LATEs within each subgroup are all positive, with the exception

Table 3: Subgroup Analysis of Common Observables with LATE and MTE(p)

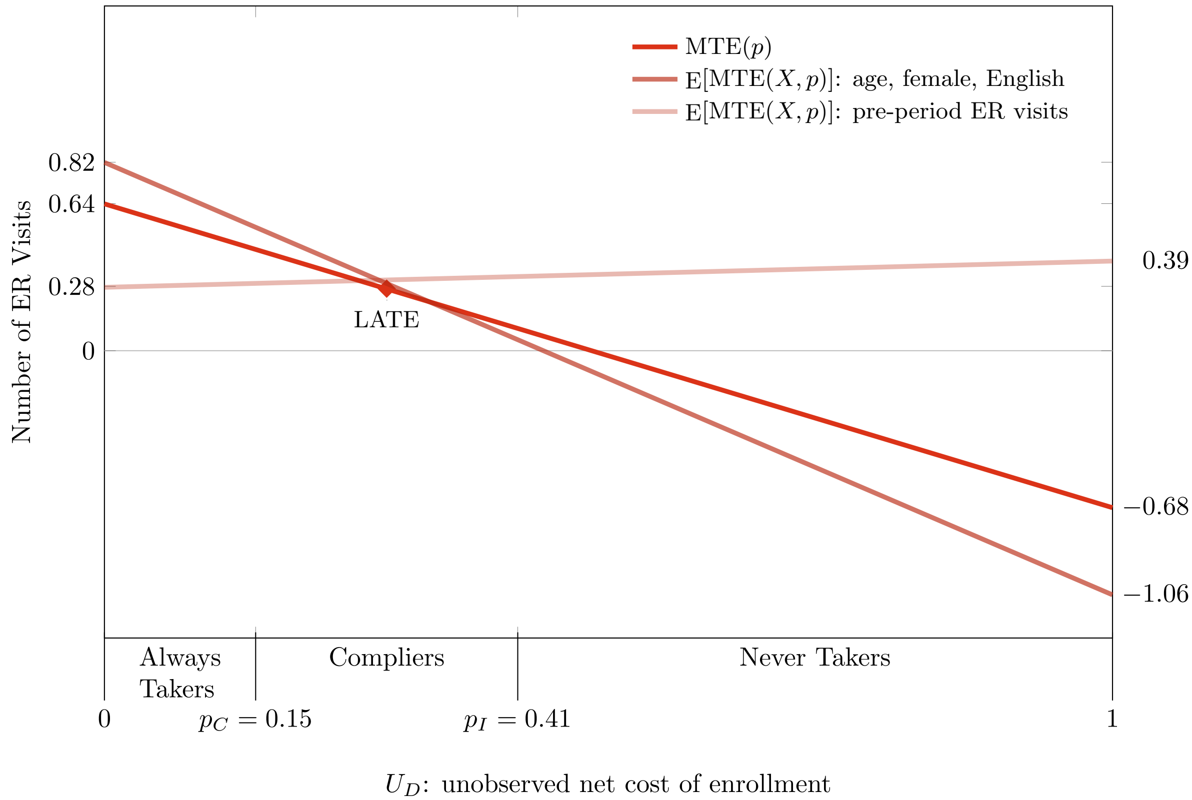
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Age ≥ median[a]	Age < median[a]	Female	Male	English	Non-English
Oregon Health Insurance Experiment of 2008							
LATE	0.27	0.14	0.44	0.14	0.39	0.30	-0.15
	(0.15)	(0.18)	(0.25)	(0.21)	(0.21)	(0.16)	(0.34)
p_C	0.15	0.13	0.17	0.20	0.10	0.15	0.16
	(0.003)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.01)
p_I	0.41	0.43	0.39	0.43	0.38	0.41	0.38
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
MTE intercept	0.64	0.98	0.31	0.48	0.92	0.72	0.14
	(0.24)	(0.28)	(0.39)	(0.32)	(0.33)	(0.25)	(0.47)
MTE slope	-1.32	-3.01	0.48	-1.06	-2.20	-1.51	-1.07
	(0.88)	(1.04)	(1.49)	(1.08)	(1.40)	(0.92)	(2.07)
p^*	0.48	0.33	-0.63	0.45	0.42	0.48	0.13
	(2.84)	(0.85)	(10.37)	(1.49)	(3.47)	(4.53)	(11.99)
N	19,622	9,816	9,806	10,932	8,690	17,871	1,751
Massachusetts Health Reform of 2006							
p_C	0.90	0.93	0.87	0.92	0.87	0.91	0.55
	(0.003)	(0.003)	(0.005)	(0.003)	(0.005)	(0.003)	(0.02)
p_I	0.95	0.96	0.93	0.96	0.93	0.96	0.74
	(0.002)	(0.002)	(0.004)	(0.002)	(0.004)	(0.002)	(0.02)
N	62,456	40,492	21,964	38,808	23,648	59,233	3,223

Note. Bootstrapped standard errors are in parentheses. The number of ER visits represents the total number of visits to the emergency department during the study period from March 10, 2008 to September 30, 2009. The value p^* indicates the share of the sample with positive treatment effects when the MTE(p) curve slopes downward and the share of the sample with negative treatment effects when the MTE(p) curve slopes upward. "Age" is measured in year 2008 for the Oregon Health Insurance Experiment and in year 2006 for the Massachusetts health reform. "English" is an indicator variable for individuals in the Oregon Health Insurance Experiment who requested materials in English and that equals one for individuals in the BRFSS who completed the interview in English. "Non-English" is the complement of "English." [a]The median age in the Oregon Health Insurance Experiment is 41. I use the same age to construct the Massachusetts subgroups.

of the LATE within the group that requested communication in a language other than English. Taubman et al. (2014) report LATEs within a wide variety of observable subgroups and also find that almost all are positive. Because LATEs within each subgroup are almost all positive, LATE-reweighting based on any of the common observables yields a positive treatment effect for almost any weights. Therefore, LATE-reweighting based only on the common observables cannot reconcile the positive treatment effect in Oregon with the negative treatment effect in Oregon.

It is not surprising that LATE-reweighting with common observables cannot explain treatment effect heterogeneity across Oregon and Massachusetts because observables cannot explain treatment

Figure 7: MTE(x, p) with Previous ER Utilization vs. MTE(x, p) with Common Observables



Note. The number of ER visits represents the total number of visits to the emergency department during the study period from March 10, 2008 to September 30, 2009. Pre-period ER visits refers to a group of indicators for visiting the ER 0 times, 1 time, 2–3 times, and 4 or more times during the pre-period from January 1, 2007 to March 9, 2008. "Age" is measured in year 2008. "Female" is a binary indicator for the gender of the respondent. "English" is a binary indicator that equals one for individuals who requested materials in English. The specification with common covariates (age, female, English) includes all two-way interactions. p_C is the probability of enrollment in the control group, and p_I is the probability of enrollment in the intervention group.

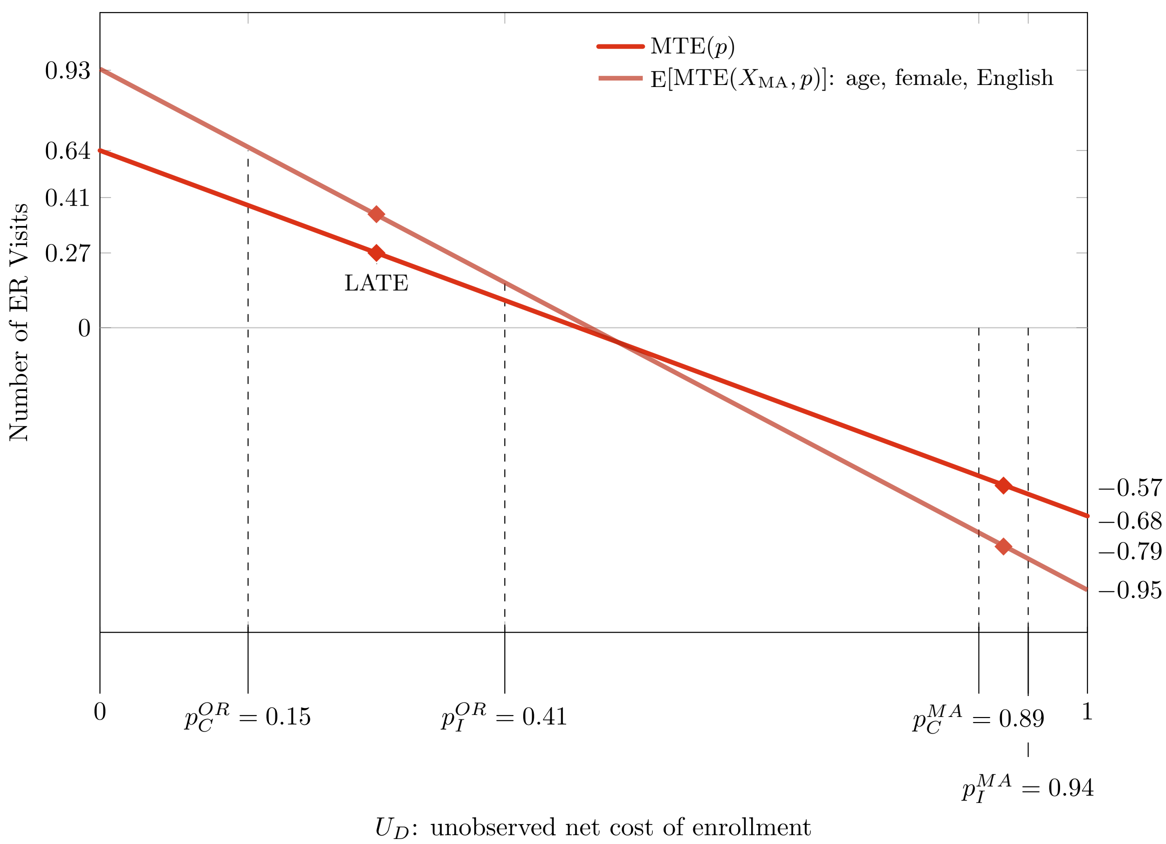
effect heterogeneity within Oregon. To demonstrate, I estimate an MTE within each subgroup, and I report the slope and intercept in Table 3. In almost all subgroups, the MTE slopes downward. When the MTE slopes downward, the horizontal intercept p^* gives the fraction of individuals predicted to have positive treatment effects. In all but one subgroup, even though the LATEs are positive, the MTEs predict that the majority of individuals have *negative* treatment effects, indicating that the common observables leave substantial heterogeneity unexplained.

5.3.4 MTE-Reweighting with Common Observables

When I include all of the common observables as well as their two-way interactions in the MTE, substantial heterogeneity remains unexplained. I emphasize the comparison of unexplained heteregeneity across various MTE functions in Figure 7. To do so, I present E[MTE(x, p)] functions, which average included observed heterogeneity across all individuals. Consistent with the depiction in Figure 6, the inclusion of pre-period ER visits in MTE(x, p) results in a function that is flatter than MTE(p). Therefore, the inclusion of pre-period ER visits decreases unexplained heterogeneity in the treatment effect. In contrast, the inclusion of the common observables in MTE(x, p) results in a function that is steeper than MTE(p). Therefore, the inclusion of pre-period ER visits *increases* unexplained heterogeneity in the treatment effect.

MTE-reweighting with observables can still proceed if there is unexplained heterogeneity in the treatment effect. To reweight the Oregon MTE with common observables for extrapolation to Massachusetts, I estimate the average MTE(x, p) for compliers in Massachusetts to construct E[MTE(X_{MA}, p)]. In Figure 8, I plot E[MTE(X_{MA}, p)]. Reweighting the Oregon MTE to predict the impact of the Massachusetts reform on ER utilization, I apply (20) using the pre-reform level of coverage in Massachusetts p_C^{MA} and the post-reform level of coverage in Massachusetts p_I^{MA}. I predict that the Massachusetts reform will decrease emergency room utilization by 0.79 visits over

Figure 8: Extrapolation of MTE(x, p) to Massachusetts



Note. The number of ER visits represents the total number of visits to the emergency department during the study period from March 10, 2008 to September 30, 2009. "Age" is measured in year 2008 for the Oregon Health Insurance Experiment and in year 2006 for the Massachusetts health reform."English" is an indicator variable for individuals in the Oregon Health Insurance Experiment who requested materials in English and that equals one for individuals in the BRFSS who completed the interview in English. The specification with common covariates (age, female, English) includes all two-way interactions. p_C^{OR} is the probability of enrollment in the control group in Oregon, p_I^{OR} the probability of enrollment in the intervention group in Oregon, p_C^{MA} the probability of enrollment in the control group in the Massachusetts reform, and p_I^{MA} the probability of enrollment in the intervention group in the Massachusetts reform.

an approximately 19 month period. Translating this decrease into an annual decrease, I predict a decrease of 0.50 visits per person per year (=(0.79/19)*12), which is even closer to the Miller (2012) estimates of 0.67 to 1.28 than the decrease that I predict with the Oregon MTE(p), also plotted for comparison.

Figure 8 illustrates that accounting for differences in U_D between Oregon and Massachusetts has a much larger impact than accounting for differences in common observables between Oregon and Massachusetts. If I account for the observables of Massachusetts compliers with E[MTE(X_{MA}, p)], but do not account for range of U_D for Massachusetts compliers, then I predict a Massachusetts LATE of 0.41, which is even more positive than the LATE of 0.27 estimated in Oregon. Such an approach, which can be considered a form of LATE-reweighting, does not reconcile the positive LATE in Oregon with the negative LATE in Massachusetts, given that common observables do not explain treatment effect heterogeneneity across U_D in Oregon. This finding demonstrates that the power of LATE-weighting to reconcile results across contexts is limited by the common observables available for reweighting.

It should come as no surprise that the common observables that happen to be available in two contexts might not explain treatment effect heterogeneity across the two contexts. The MTE offers guidance on *which* observables can potentially explain treatment effect heterogeneity. Within Oregon and across Oregon and Massachusetts, it appears that variables that capture heterogeneity in self-reported health can potentially explain treatment effect heterogeneity. Previous ER utilization can potentially explain treatment effect heterogeneity, but common observables cannot. Regardless, MTE-reweighting with the common observables can still reconcile the positive treatment effect induced by the Oregon experiment with the negative treatment effect induced by the Massachusetts reform.

6 Conclusion

I aim to shed light on why emergency room (ER) utilization increased following the Oregon Health Insurance Experiment but decreased following the Massachusetts reform. To do so, I combine insights from the literatures on insurance and treatment effects. Starting from the Oregon Health Insurance Experiment as the "gold standard," I find heterogeneous moral hazard: although Oregon compliers increase their ER utilization upon gaining coverage, Oregon never takers would decrease their ER utilization upon gaining coverage. I also find heterogeneous selection: Oregon never takers report better health than Oregon compliers.

Next, I extrapolate my findings from within the Oregon experiment to Massachusetts. Given higher levels of coverage in Massachusetts, Massachusetts compliers are comparable to a subset of Oregon never takers. Like Oregon never takers, Massachusetts compliers report better health than Oregon compliers. Therefore, I can reconcile the increase in ER utilization induced by the Oregon Health Insurance Experiment with the decrease in ER utilization induced by the Massachusetts while. Upon gaining coverage, individuals who report worse health increase their ER utilization, while individuals who report better health decrease their ER utilization.

Appendix A Proof of the Enrollment Equation ($D = 1\{U_D \le \mathrm{P}(D = 1 \mid Z = z, X)\}$)

Medicaid enrollment D is given by

$$\begin{aligned}
D &= 1\{0 \le V_T - V_U\} \\
&= 1\{0 \le \mu_T(z, X) + \nu_T - \mu_U(z, X) - \nu_U\} \\
&= 1\{\nu_U - \nu_T \le \mu_T(z, X) - \mu_U(z, X)\} \\
&= 1\{F(\nu_U - \nu_T) \le F(\mu_T(z, X) - \mu_U(z, X))\} && (F \text{ increasing under A.2}) \\
&= 1\{U_D \le F(\mu_T(z, X) - \mu_U(z, X))\} && (U_D = F(\nu_U - \nu_T \mid X) \text{ by definition}) \\
&= 1\{U_D \le \mathrm{P}(D = 1 \mid Z = z, X)\},
\end{aligned}$$

where the last equality follows from

$$\begin{aligned}
F(\mu_T(z, X) - \mu_U(z, X) \mid X) &= \mathrm{P}(\nu_U - \nu_T \le \mu_T(z, X) - \mu_U(z, X) \mid X) \\
&= \mathrm{P}(\nu_U - \nu_T \le \mu_T(Z, X) - \mu_U(Z, X) \mid Z = z, X) \\
&\qquad ((\nu_U - \nu_T) \perp Z \mid X \text{ by A.1}) \\
&= \mathrm{P}(0 \le \mu_T(Z, X) + \nu_T - \mu_U(Z, X) - \nu_U \mid Z = z, X) \\
&= \mathrm{P}(0 \le V_T - V_U \mid Z = z, X) \\
&= \mathrm{P}(D = 1 \mid Z = z, X). \qquad \blacksquare
\end{aligned}$$

Appendix B Proof that $U_D = F(\nu_U - \nu_T)$ is uniformly distributed between 0 and 1

The result that U_D is distributed uniformly between 0 and 1 is not a separate assumption of the model. It is due to the "probability integral transformation," which shows that the cumulative distribution function of any random variable $\widetilde{\nu} = \nu_T - \nu_U$ applied to itself must be distributed uniformly between 0 and 1 (for example, see Casella and Berger (2002, page 54)):

A random variable Y is distributed uniformly between 0 and 1 if and only if $F_Y(x) = x$ for $0 \le x \le 1$. Therefore, U_D is distributed uniformly between 0 and 1:

$$\begin{aligned}
F_{U_D}(u) &= P(U_D \le u) \\
&= P(F(\nu_U - \nu_T) \le u) \\
&= P(\nu_U - \nu_T \le F^{-1}(u)) && (F \text{ increasing under A.2}) \\
&= F(F^{-1}(u)) = u. && (F \text{ continuous under A.2})
\end{aligned}$$

$\blacksquare$

Appendix C Averages for Always Takers, Compliers, and Never Takers

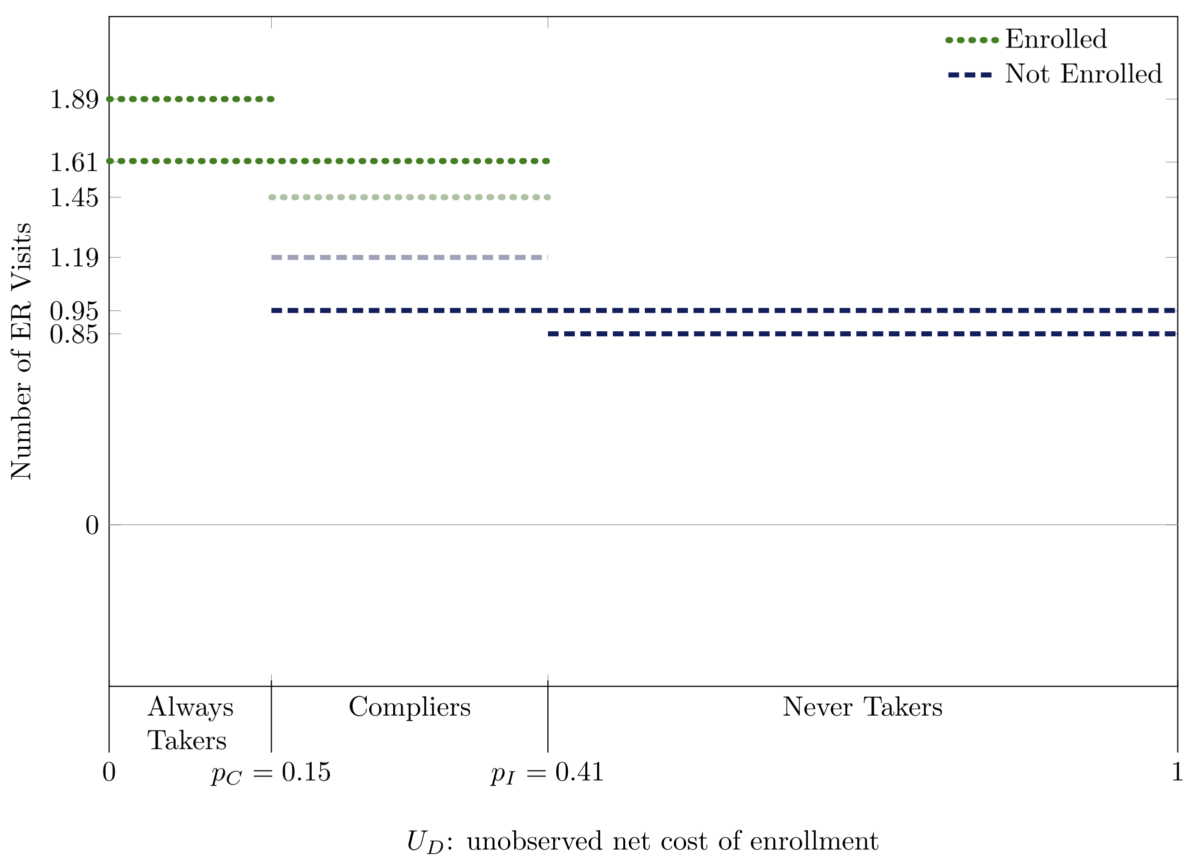
Imbens and Rubin (1997), Katz et al. (2001), Abadie (2002), and Abadie (2003) rely on the LATE assumptions to calculate average outcomes and observables of always takers, compliers, and never takers. For consistency with my exposition, I perform the same calculations using the MTE model that assumes no more than the LATE assumptions. I build intuition with a graphical illustration that follows from the model.

I identify the expected value of Y_T for always takers as follows, supressing X for simplicity:

$$\begin{aligned}
\mathrm{E}[Y \mid D=1, Z=0] &= \mathrm{E}[Y_U + D(Y_T - Y_U) \mid D=1, Z=0] && \text{(by (7))} \\
&= \mathrm{E}[Y_T \mid D=1, Z=0] \\
&= \mathrm{E}[Y_T \mid 0 \leq U_D \leq p_C, Z=0] && \text{(by (6), where } p_C = P(D=1|Z=0)) \\
&= \mathrm{E}[g_T(U_D, \gamma_T) \mid 0 \leq U_D \leq p_C, Z=0] && \text{(by (8))} \\
&= \mathrm{E}[g_T(U_D, \gamma_T) \mid 0 \leq U_D \leq p_C] && (Z \perp (U_D, \gamma_T) \text{ by (A.5)}) \\
&= \mathrm{E}[Y_T \mid 0 \leq U_D \leq p_C].
\end{aligned}$$

I use similar steps to calculate the expected value of Y_T for lottery winners enrolled in Medicaid $\mathrm{E}[Y_T \mid 0 \leq U_D \leq p_I] = \mathrm{E}[Y \mid D=1, Z=1]$, the expected value of Y_U for never takers $\mathrm{E}[Y_U \mid p_I <$

Figure C1: Graphical Illustration of Average Number of ER Visits for Compliers



Note. The number of ER visits represents the total number of visits to the emergency department during the study period from March 10, 2008 to September 30, 2009. p_C is the probability of enrollment in the control group, and p_I is the probability of enrollment in the intervention group.

$U_D \leq 1] = \mathrm{E}[Y \mid D = 0, Z = 1]$, and the expected value of Y_U for lottery losers not enrolled in Medicaid $\mathrm{E}[Y_U \mid p_C < U_D \leq p_I] = \mathrm{E}[Y \mid D = 0, Z = 0]$. I then use the four resulting values to calculate the expected value of Y_T for compliers enrolled in Medicaid:

$$\mathrm{E}[Y_T \mid p_C < U_D \leq p_I] = \frac{p_I}{p_I - p_C}\mathrm{E}[Y_T \mid 0 \leq U_D \leq p_I] - \frac{p_C}{p_I - p_C}\mathrm{E}[Y_T \mid 0 \leq U_D \leq p_C],$$

and the expected value of Y_U for compliers not enrolled in Medicaid:

$$\mathrm{E}[Y_U \mid p_C < U_D \leq p_I] = \frac{1 - p_C}{p_I - p_C}\mathrm{E}[Y_U \mid p_C < U_D \leq 1] - \frac{1 - p_I}{p_I - p_C}\mathrm{E}[Y_U \mid p_I < U_D \leq 1].$$

I illustrate the calculations graphically using values from Oregon data in Figure C1. I use bolded dotted lines to depict average ER utilization when enrolled in Medicaid, Y_T, for two observed groups: lottery losers enrolled in Medicaid ($0 \leq U_D \leq p_C$) and lottery winners enrolled in Medicaid ($0 \leq U_D \leq p_I$). I use bolded dashed lines to depict average ER utilization when not enrolled in Medicaid, Y_U, for two observed groups: lottery losers not enrolled in Medicaid ($p_C < U_D \leq 1$) and lottery winners not enrolled in Medicaid ($p_I < U_D \leq 1$). I depict the calculated outcomes for compliers with lighter shading.

To calculate the average observable X for each group, I begin with the same approach. Even though average *outcomes* of compliers should depend on whether they win or lose the lottery, average *observables* of compliers should not. Therefore, I weight the average observables of compliers who win and lose the lottery by their respective probabilities:

$$\mathrm{E}[X \mid p_C < U_D \leq p_I] = \mathrm{P}(Z = 1)\mathrm{E}[X \mid p_C < U_D \leq p_I] + \mathrm{P}(Z = 0)\mathrm{E}[X \mid p_C < U_D \leq p_I].$$

Appendix D Estimating MTO(x, p), MUO(x, p), and MTE(x, p) of order $k \geq 1$ in p.

While this paper considers the case where MTO(x, p) and MUO(x, p) are polynomials of order $k = 1$, the steps below present the general estimation strategy for polynomials of order $k \geq 1$:

1. Estimate propensity scores, $\widehat{p}$, for all individuals in the sample by fitting

$$D = \alpha_0 + \alpha_1 Z + \alpha_2' X + \alpha_3'(X'Z) + \varepsilon$$

and using $\widehat{\alpha}_0$, $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, and $\widehat{\alpha}_3$ to predict D conditional on Z and observables X.

2. For each $j \in \{0, \dots, k\}$, apply $h_T^{(j)}(\cdot)$ derived in Appendix E and $h_U^{(j)}(\cdot)$ derived in Appendix F to the propensity scores, $\widehat{p}$, estimated in step 1. These transformations will allow us to use the estimated propensity scores to obtain coefficients for the marginal outcome functions in the next steps:

$$h_T^{(j)}(\widehat{p}) = \frac{\widehat{p}^j}{1 + j},$$
$$h_U^{(j)}(\widehat{p}) = \frac{1 - \widehat{p}^{1+j}}{(1 + j)(1 - \widehat{p})}.$$

3. Condition the sample on treated individuals ($D = 1$), and use OLS to estimate:

$$Y = \beta_T' X + \sum_{j=0}^{k} \lambda_{Tj} h_T^{(j)}(\widehat{p}) + \gamma_T.$$

   Construct $\mathrm{MTO}(x,p) = \widehat{\beta}_T' x + \sum_{j=0}^{k} \widehat{\lambda}_{Tj} p^j$.

4. Condition the sample on untreated individuals ($D = 0$), and use OLS to estimate:

$$Y = \beta_U' X + \sum_{j=0}^{k} \lambda_{Uj} h_U^{(j)}(\widehat{p}) + \gamma_U.$$

   Construct $\mathrm{MUO}(x,p) = \widehat{\beta}_U' x + \sum_{j=0}^{k} \widehat{\lambda}_{Uj} p^j$.

5. Construct $\mathrm{MTE}(x,p) = \mathrm{MTO}(x,p) - \mathrm{MUO}(x,p) = (\widehat{\beta}_T - \widehat{\beta}_U)' x + \sum_{j=0}^{k} (\widehat{\lambda}_{Tj} - \widehat{\lambda}_{Uj}) p^j$.

Appendix E Derivation of $\mathrm{MTO}(x,p)$.

The objective of this appendix is to derive a transformation that can be applied to propensity scores to estimate the MTO function of the following form:

$$\mathrm{MTO}(x,p) = \beta_T' x + \sum_{j=0}^{k} \lambda_{Tj} p^j. \tag{21}$$

We derive the MTO function from the average treated outcome function, which can be estimated directly. Taking the expectation of Y conditional on observables x and conditional on enrollment in Medicaid given propensity score p gives the average treated outcome function:

$$\begin{aligned}
\mathrm{E}\left[Y_T \mid X = x, 0 \le U_D \le p\right] &= \mathrm{E}[\beta_T' X + \sum_{j=0}^{k} \lambda_{Tj} U_D^j + \gamma_T \mid X = x, 0 \le U_D \le p] \\
&= \mathrm{E}[\beta_T' X + \sum_{j=0}^{k} \lambda_{Tj} U_D^j \mid X = x, 0 \le U_D \le p] \qquad (\mathrm{E}[\gamma_T \mid X, U_D] = 0) \\
&= \beta_T' x + \sum_{j=0}^{k} \lambda_{Tj} \mathrm{E}[U_D^j \mid X = x, 0 \le U_D \le p] \\
&= \beta_T' x + \sum_{j=0}^{k} \lambda_{Tj} \int_0^p \frac{1}{p} u_D^j \, \mathrm{d}u_D \qquad (U_D \sim U[0,1]) \\
&= \beta_T' x + \sum_{j=0}^{k} \lambda_{Tj} \frac{p^j}{1+j}. \qquad (22)
\end{aligned}$$

Transforming the propensity scores using $h_T^{(j)}(p) := \frac{p^j}{1+j}$ gives the MTO parameters, λ_{Tj}, from (21).

Appendix F Derivation of MUO(x, p).

The objective of this appendix is to derive a transformation that can be applied to propensity scores to estimate the MUO function of the following form:

$$\mathrm{MUO}(x, p) = \beta_U' x + \sum_{j=0}^{k} \lambda_{Uj} p^j. \tag{23}$$

We derive the MUO function from the average untreated outcome function, which can be estimated directly. Taking the expectation of Y conditional on observables x and conditional on enrollment in Medicaid given propensity score p gives the average untreated outcome function:

$$\begin{aligned}
\mathrm{E}\left[Y_U \mid X = x, p \leq U_D \leq 1\right] &= \mathrm{E}[\beta_U' X + \sum_{j=0}^{k} \lambda_{Uj} U_D^j + \gamma_U \mid X = x, p < U_D \leq 1] \\
&= \mathrm{E}[\beta_U' X + \sum_{j=0}^{k} \lambda_{Uj} U_D^j \mid X = x, p < U_D \leq 1] \qquad (\mathrm{E}[\gamma_U \mid X, U_D] = 0) \\
&= \beta_U' x + \sum_{j=0}^{k} \lambda_{Uj} \mathrm{E}[U_D^j \mid X = x, p < U_D] \\
&= \beta_U' x + \sum_{j=0}^{k} \lambda_{Uj} \int_p^1 \frac{1}{1-p} u_D^j \, \mathrm{d}u_D \qquad (U_D \sim U[0,1]) \\
&= \beta_U' x + \sum_{j=0}^{k} \lambda_{Uj} \frac{1 - p^{1+j}}{(1+j)(1-p)}.
\end{aligned} \tag{24}$$

Transforming the propensity scores using $h_U^{(j)}(p) := \frac{1-p^{1+j}}{(1+j)(1-p)}$ gives the MUO parameters, λ_{Uj}, from (23).

References

Oregon health insurance experiment public use data. URL http://www.nber.org/oregon/data.html.

Alberto Abadie. Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American statistical Association*, 97(457):284–292, 2002.

Alberto Abadie. Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics*, 113(2):231–263, 2003.

Michael Anderson, Carlos Dobkin, and Tal Gross. The effect of health insurance coverage on the use of medical services. *American Economic Journal: Economic Policy*, 4(1):1–27, 2012.

Michael L Anderson, Carlos Dobkin, and Tal Gross. The effect of health insurance on emergency department visits: Evidence from an age-based eligibility threshold. *Review of Economics and Statistics*, 96(1):189–195, 2014.

JD Angrist. Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica*, 66(2):249–288, 1998.

Joshua D Angrist. Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records. *The American Economic Review*, pages 313–336, 1990.

Joshua D Angrist. Treatment effect heterogeneity in theory and practice*. *The Economic Journal*, 114(494):C52–C83, 2004.

Joshua D Angrist and Ivan Fernandez-Val. Extrapolate-ing: External validity and. In *Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress*, volume 51, page 401. Cambridge University Press, 2013.

Joshua D Angrist and Alan B Krueger. The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American statistical Association*, 87(418):328–336, 1992.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

Katherine Baicker, Sarah Taubman, Heidi L. Allen, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Eric C. Schneider, Bill J. Wright, Alan M. Zaslavsky, and Amy N. Finkelstein. The oregon experiment - effects of medicaid on clinical outcomes. *New England Journal of Medicine*, 368(18):1713–1722, 2013.

Katherine Baicker, Amy Finkelstein, Jae Song, and Sarah Taubman. The impact of medicaid on labor market activity and program participation: Evidence from the oregon health insurance experiment. *American Economic Review*, 104(5):322–28, 2014.

Marinho Bertanha and Guido W. Imbens. External validity in fuzzy regression discontinuity designs. Working Paper 20773, National Bureau of Economic Research, December 2014.

Anders Björklund and Robert Moffitt. The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics*, pages 42–49, 1987.

Dan A Black, Joonhwi Joo, Robert LaLonde, Jeffrey A Smith, and Evan J Taylor. Simple tests for selection bias: Learning more from instrumental variables. 2015.

Christian N Brinch, Magne Mogstad, and Matthew Wiswall. Beyond late with a discrete instrument. *Journal of Political Economy*, 125(4):985–1039, 2017.

Pedro Carneiro and Sokbae Lee. Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics*, 149(2):191–208, 2009.

Pedro Carneiro, James J. Heckman, and Edward J. Vytlacil. Estimating marginal returns to education. *American Economic Review*, 101(6):2754–81, October 2011.

George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Christopher Chen, Gabriel Scheffler, and Amitabh Chandra. Massachusetts' health care reform and emergency department utilization. *New England Journal of Medicine*, 365(12):e25, 2011.

Pierre-André Chiappori and Bernard Salanie. Testing for asymmetric information in insurance markets. *Journal of political Economy*, 108(1):56–78, 2000.

Janet Currie and Jonathan Gruber. Health insurance eligibility and child health: lessons from recent expansions of the medicaid program. *Quarterly Journal of Economics*, 111(2):431–466, 1996.

Liran Einav, Amy Finkelstein, and Mark R Cullen. Estimating welfare in insurance markets using variation in prices. *The Quarterly Journal of Economics*, 125(3):877, 2010.

Amy Finkelstein and James Poterba. Testing for asymmetric information using unused observables in insurance markets: evidence from the uk annuity market. *Journal of Risk and Insurance*, 81 (4):709–734, 2014.

Amy F. Finkelstein, Sarah Taubman, Bill J. Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi L. Allen, Katherine Baicker, and the Oregon Health Study Group. The oregon health insurance experiment: Evidence from the first year. *The Quarterly Journal of Economics*, 127(3):1057–1106, 2012.

Amy N. Finkelstein, Sarah L. Taubman, Heidi L. Allen, Bill J. Wright, and Katherine Baicker. Effect of medicaid coverage on ed use — further evidence from oregon's experiment. *New England Journal of Medicine*, 375(16):1505–1507, 2016. PMID: 27797307.

Eric French and Jae Song. The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy*, 6(2):291–337, 2014.

Zijian Guo, Jing Cheng, Scott A Lorch, and Dylan S Small. Using an instrumental variable to test for unmeasured confounding. *Statistics in medicine*, 33(20):3528–3546, 2014.

Martin B. Hackmann, Jonathan T. Kolstad, and Amanda E. Kowalski. Adverse selection and an individual mandate: When theory meets practice. *American Economic Review*, 105(3):1030–66, 2015.

Jerry A Hausman. Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, pages 1251–1271, 1978.

James Heckman, Sergio Urzua, and Edward Vytlacil. Estimation of treatment effects under essential heterogeneity. 2006.

James J. Heckman and Edward Vytlacil. Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3):669–738, 05 2005.

James J. Heckman and Edward J. Vytlacil. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences*, 96(8):4730–4734, 1999.

James J. Heckman and Edward J. Vytlacil. Local instrumental variables. In Cheng Hsiao, Kimio Morimune, and James L. Powell, editors, *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, pages 1–46. Cambridge University Press, 2001.

James J Heckman and Edward J Vytlacil. Econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics*, 6: 4875–5143, 2007.

James J. Heckman, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. Characterizing selection bias using experimental data. *Econometrica*, 66(5):1017–1098, 1998.

JJ Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153–162, 1979.

Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.

V Joseph Hotz, Guido W Imbens, and Julie H Mortimer. Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics*, 125(1):241–270, 2005.

Martin Huber. A simple test for the ignorability of non-compliance in experiments. *Economics Letters*, 120(3):389–391, 2013.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–75, 1994.

Guido W Imbens and Donald B Rubin. Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, 64(4):555–574, 1997.

Lawrence F Katz, Jeffrey R Kling, Jeffrey B Liebman, et al. Moving to opportunity in boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics*, 116 (2):607–654, 2001.

Patrick Kline and Christopher R Walters. Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics*, 131(4):1795–1848, 2016.

Jonathan T. Kolstad and Amanda E. Kowalski. The impact of health care reform on hospital and preventive care: Evidence from massachusetts. *Journal of Public Economics*, 96:909–929, December 2012.

Amanda Kowalski. Censored quantile instrumental variable estimates of the price elasticity of expenditure on medical care. *Journal of Business & Economic Statistics*, 34(1):107–117, 2016.

Amanda Kowalski, Yen Tran, and Ljubica Ristovska. MTEBINARY: Stata module to compute Marginal Treatment Effects (MTE) With a Binary Instrument. Statistical Software Components, Boston College Department of Economics, December 2016. URL https://ideas.repec.org/c/boc/bocode/s458285.html.

Nicole Maestas, Kathleen J Mullen, and Alexander Strand. Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *The American Economic Review*, 103(5):1797–1829, 2013.

Sarah Miller. The effect of insurance on emergency room visits: an analysis of the 2006 massachusetts health reform. *Journal of Public Economics*, 96(11):893–908, 2012.

Robert Moffitt. Estimating marginal treatment effects in heterogeneous populations. *Annales d'Economie et de Statistique*, pages 239–261, 2008.

Magne Mogstad, Andres Santos, and Alexander Torgovitsky. Using instrumental variables for inference about policy relevant treatment effects. Working Paper 23568, National Bureau of Economic Research, July 2017.

Joseph P. Newhouse and Rand Corporation Insurance Experiment Group. *Free for all?: lessons from the RAND health insurance experiment*. Harvard University Press, 1993.

Randall J Olsen. A least squares correction for selectivity bias. *Econometrica: Journal of the Econometric Society*, pages 1815–1820, 1980.

Julie Rovner. Medicaid expansion boosted emergency room visits in oregon, January 2014. URL http://www.npr.org/sections/health-shots/2014/01/02/259128081/medicaid-expansion-boosted-emergency-room-visits-in-oregon.

Andrew Donald Roy. Some thoughts on the distribution of earnings. *Oxford economic papers*, 3 (2):135–146, 1951.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Donald B Rubin. Assignment to treatment group on the basis of a covariate. *Journal of Educational and Behavioral statistics*, 2(1):1–26, 1977.

Peter B. Smulowitz, Robert Lipton, J. Frank Wharam, Leon Adelman, Scott G. Weiner, Laura Burke, Christopher W. Baugh, Jeremiah D. Schuur, Shan W. Liu, Meghan E. McGrath, Bella Liu, Assaad Sayah, Mary C. Burke, J. Hector Pope, and Bruce E. Landon. Emergency department utilization after the implementation of massachusetts health reform. *Annals of Emergency Medicine*, 58(3):225 – 234.e1, 2011.

Benjamin D Sommers and Kosali Simon. Health insurance and emergency department use-a complex relationship. *The New England journal of medicine*, 376(18):1708, 2017.

Sarah L. Taubman, Heidi L. Allen, Bill J. Wright, Katherine Baicker, and Amy N. Finkelstein. Medicaid increases emergency-department use: Evidence from oregon's health insurance experiment. *Science*, 343(6168):263–268, 2014.

Sarah Tavernise. Emergency visits seen increasing with health law. *New York Times*, 2014. URL http://www.nytimes.com/2014/01/03/health/access-to-health-care-may-increase-er-visits-study-suggests.html.

Edward Vytlacil. Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, 70(1):331–341, 2002.

Abraham Wald. The fitting of straight lines if both variables are subject to error. *Ann. Math. Statist.*, 11(3):284–300, 09 1940.

Robert J Willis and Sherwin Rosen. Education and self-selection. *Journal of political Economy*, 87 (5, Part 2):S7–S36, 1979.