Doing More When You're Running LATE: Applying Marginal Treatment Effect Methods to Experiments*

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

March 25, 2016

Abstract

Researchers run experiments to obtain a treatment effect estimate that is internally valid. However, the local average treatment effect (LATE) estimated by an experiment is not globally externally valid if the treatment effect varies across individuals. The LATE gives the average treatment effect for compliers who receive the treatment if and only if they win the experimental lottery. In many experiments, there are also always takers who always receive the treatment and never takers who never receive the treatment regardless of the experimental lottery. I show that it is possible to use such experiments to recover bounds on average treatment effects for always takers and never takers. These bounds can reject global external validity of the LATE in some cases, and they depend on weaker assumptions than existing tests of global external validity. Building on existing methods to recover a marginal treatment effect (MTE) with a discrete instrument, I develop weights that allow me to recover average treatment effects for discrete groups of individuals created by a discrete instrument, including always takers and never takers. I use the recovered treatment effects to decompose group average treated outcomes into selection and treatment effects. I also decompose the sample OLS estimate into a selection effect and a treatment effect. This decomposition generalizes the comparison of the OLS estimate to the LATE when the treatment effect is heterogeneous. I apply these methods to the Oregon Health Insurance Experiment. The Oregon LATE indicates that obtaining insurance increases emergency room (ER) utilization for compliers. I find that the treatment effect of insurance on ER utilization decreases from always takers to compliers to never takers. I also find that potential uninsured ER utilization decreases from always takers to compliers to never takers. Therefore, the selection effect and the treatment effect of insurance on insured ER utilization decrease as a larger fraction of individuals gain insurance. The heterogeneous selection and treatment effects that I recover from the OHIE indicate that a different policy experiment could increase or decrease ER utilization, depending on which individuals it induces to gain coverage.

*I thank Aigerim Kabdiyeva, Samuel Moy, and Ljubica Ristovska for excellent research assistance. Joseph Altonji, Steve Berry, Lasse Brune, Pedro Carneiro, Mark Duggan, Caroline Hoxby, Liran Einav, Amy Finkelstein, Matthew Gentzkow, Jonathan Gruber, Guido Imbens, Dean Karlan, Jonathan Levin, Sarah Miller, Magne Mogstad, Costas Meghir, Mark Rosenzweig, Joseph Shapiro, Orie Shelef, Ashley Swanson, Ed Vytlacil, David Wilson, and seminar participants at Academia Sinica, Berkeley, Princeton, Santa Clara, Singapore Management University, Stanford, Stanford GSB, USC, Yale, CHES and the WEAI provided helpful comments. NSF CAREER Award 1350132 and the Stanford Institute for Economic Policy Research (SIEPR) provided support. **Keywords:** average treatment effect (ATE), compliers, difference-in-difference test, local average treatment effect (LATE), Massachusetts health reform, meta-analysis, natural experiment, program evaluation, selection.

1 Introduction

Researchers run experiments to obtain a treatment effect estimate that is internally valid. However, the local average treatment effect (LATE) estimated by an experiment is not globally externally valid if the treatment effect varies across individuals. The LATE gives the average treatment effect for compliers who receive the treatment if and only if they win the experimental lottery. In many experiments, there are also always takers who always receive the treatment and never takers who never receive the treatment regardless of the experimental lottery. If the treatment effect varies from always takers to compliers to never takers, then the LATE is not globally valid. I focus on examining the external validity of experiments by using information available within experiments to identify average treatment effects for always takers, compliers, and never takers.

I build on marginal treatment effect (MTE) methods developed by Björklund and Moffitt [1987] and Heckman and Vytlacil [1999, 2005, 2007]. Traditionally, the MTE could only be identified in settings with continuous instruments. Therefore, MTE methods could not be applied to experiments with discrete or binary interventions. However, recent extensions developed by Brinch et al. [2012] provide approaches to identify the MTE in settings with discrete instruments, thus allowing for the application of MTE methods to experiments.

Under weaker assumptions than the assumptions that Brinch et al. [2012] impose to identify the MTE and to test for external validity, I show that it is possible to use experiments to recover bounds on average treatment effects for always takers and never takers. In some cases, these bounds do not include the LATE. Therefore, these bounds provide a new test that can reject global external validity under weaker assumptions than existing tests.

Next, building on the Brinch et al. [2012] approach to recover a marginal treatment effect with a discrete instrument, I develop weights that allow me to recover average treatment effects for discrete groups of individuals created by a discrete instrument. These weights differ from the weights for continuous instruments used by Brinch et al. [2012]. They allow me to recover average treatment effects for always takers, never takers, and other groups of interest from the MTE.

Using the treatment effects that I recover, I decompose group average treated outcomes into selection and treatment effects. I also decompose the sample OLS estimate into a selection effect and a treatment effect. This decomposition generalizes the comparison of the OLS estimate to the LATE when the treatment effect is heterogeneous.

I apply these methods to the Oregon Health Insurance Experiment (OHIE), an important randomized experiment for which the external validity of the findings is of particular policy-relevance. Through my

analysis of the OHIE, I contribute to the understanding of the impact of health insurance expansions on emergency room (ER) utilization. Legislation requires that emergency rooms see all patients, regardless of whether they have health insurance, making the ER the main portal through which the uninsured enter the healthcare system. ER utilization of the uninsured places a burden on other players in the healthcare system who pay for their uncompensated care. Furthermore, the uninsured themselves could potentially get higher quality, less expensive, and more coordinated care through other outlets. For these reasons, policymakers are particularly interested in how emergency room utilization will change in response to other health insurance expansions, especially those induced by the Affordable Care Act.

The OHIE is arguably the "gold standard" for evidence on the impact of health insurance expansions because it is a recent randomized experiment, but there is reason to question the external validity of the treatment effects derived from it. A central finding from the OHIE is that health insurance increased ER utilization for individuals who gained health insurance coverage through the OHIE lottery (Taubman et al. [2014]). However, results from a credible natural experiment that increased health insurance coverage, the Massachusetts reform of 2006, show that ER utilization decreased (Miller [2012], Smulowitz et al. [2011]) or stayed the same (Chen et al. [2011]), and admissions to the hospital from the emergency room (a proxy for emergency room visits) decreased (Kolstad and Kowalski [2012]). Related evidence on the ER utilization of other populations of newly insured individuals also yields varying results (see Currie and Gruber [1996], Anderson et al. [2012, 2014], Newhouse and Rand Corporation Insurance Experiment Group [1993]).

Using data from the OHIE, I find that the treatment effect of insurance on ER utilization decreases from always takers to compliers to never takers. I also find that potential uninsured ER utilization decreases from always takers to compliers to never takers. Therefore, the selection effect and the treatment effect of insurance on insured ER utilization decrease as a larger fraction of individuals gain insurance. The heterogeneous selection and treatment effects that I recover from the OHIE indicate that a different policy experiment could increase or decrease ER utilization, depending on which individuals it induces to gain coverage.

In the next section, I present a model that I use to define global external validity in terms of the MTE. I discuss identification and estimation of the MTE in Section 3. In Section 4, I apply MTE methods to the OHIE and I extrapolate the results to other contexts, including the Massachusetts health reform. In Section 5, I provide lessons for the design of future experiments, and I conclude.

2 The MTE and the External Validity of Experiments

2.1 Model of Selection into Treatment within an Experiment

Selection into treatment is the key model element. Suppose D represents a binary treatment such as health insurance coverage and Y represents an observed outcome such as emergency room utilization. Define Y_T as the potential outcome of an individual in the treated state ($D = 1$), and define Y_U as the potential outcome of an individual in the untreated state ($D = 0$).[1] In the OHIE context, Y_T represents potential emergency room utilization with health insurance coverage, and Y_U represents potential emergency room utilization without health insurance coverage. The following model relates the potential outcomes to the observed outcome:

$$Y = (1 - D)Y_U + DY_T.$$

In this model, an individual selects into treatment D based on the net benefit of treatment, I_D, which consists of an observed component p and an unobserved component U_D as follows:

$$I_D = p - U_D.$$

Since U_D enters (2.1) negatively, I refer to it as the unobserved net *cost* of treatment. The unobserved net cost of treatment can have any distribution, but the quantiles of any distribution are distributed uniformly between 0 and 1. I therefore normalize $U_D \sim U(0, 1)$ so that U_D represents the fraction of the population with a lower unobserved net cost of treatment. In the OHIE context, U_D could include pent-up demand for ER utilization, hypochondria, income, health, and any observable factor that is not specified in the model.

Since p enters (2.1) positively, I refer to it as the observed net *benefit* of treatment. I specify p such that it represents the potential fraction treated: $p \equiv P(D = 1|Z[, X])$ is the probability of treatment given the observed value of the instrument Z and an optional vector of covariates X. In OHIE context, Z is a binary indicator for winning the Oregon lottery. For any binary instrument, two values of p are observed. The first is the rate of treatment among lottery losers $p_B \equiv P(D = 1|Z = 0)$, which gives the potential fraction treated if the entire population were to remain in the baseline world without an experimental intervention. The second is the rate of treatment among lottery winners $p_I \equiv P(D = 1|Z = 1)$, which gives the potential fraction treated if the entire population were to be eligible for the intervention. The observed probability of treatment in the full experimental population is a weighted average $P(D = 1) = s(pB)p_B + s(pI)p_I$, where $s(p_B) \equiv P(Z = 0)$ represents the share that loses the lottery and $s(p_I) \equiv P(Z = 1) = 1 - s(p_B)$ represents

[1] Rubin [1974], Rubin [1977], and Holland [1986] developed the idea of potential outcomes. I have changed the traditional notation from Y_1 to Y_T and Y_0 to Y_U to facilitate simpler notation for concepts that I introduce later.

the share that wins the lottery.

In the baseline world without an experimental intervention, individuals with unobserved cost of treatment less than or equal to the observed net benefit of treatment are **baseline treated (BT)**: $0 \leq U_D \leq p_B$. All others are **baseline untreated (BU)**: $p_B < U_D \leq 1$. In a world with the experimental intervention, individuals with unobserved cost of treatment less than or equal to the observed net benefit of treatment are **intervention treated (IT)**: $0 \leq U_D \leq p_I$. All others are **intervention untreated (IU)**: $p_I < U_D \leq 1$.

Suppose that the experimental intervention increases the observed fraction of individuals to select into treatment, implying $p_B < p_I$. All of the baseline treated individuals continue to receive treatment: these are **always takers**: $0 \leq U_D \leq p_B$. The baseline untreated individuals with unobserved cost of treatment less than the intervention treatment probability select into treatment: these are **compliers**: $p_B < U_D \leq p_I$. The remaining baseline untreated individuals with the highest unobserved net cost of treatment, also known as the intervention untreated, remain untreated: these are **never takers**: $p_I < U_D \leq 1$.[2]

Compliers are individuals who change their takeup behavior from a specific baseline to a specific intervention; therefore, other individuals could be "compliers" for different interventions and different baselines. Always takers and never takers from one experiment can shift to being compliers in another experiment. Therefore, the terms "compliers," "always takers," and "never takers" are specific to a given baseline and intervention, and therefore specific to a given experiment. Imagine that the distance between the baseline and intervention becomes infinitesimal. Further imagine a continuum of potential interventions such that the individual with the lowest unobserved net cost of treatment ($U_D = 0$) is the complier for the first and the individual with the highest unobserved net cost of treatment ($U_D = 1$) is the complier for the last. For each potential intervention in the continuum, the marginal individual is the one for whom the unobserved net cost of treatment U_D is equal to the observed net benefit of treatment p.

2.2 The Marginal Treatment Effect

The **marginal treatment effect (MTE)**, as popularized by Heckman and Vytlacil [1999], is the difference between the treated potential outcome and the untreated potential outcome for an individual marginal to selecting into treatment – an individual for whom the unobserved net cost of treatment U_D is equal to the observed net benefit of treatment p:

$$MTE(p) = E(Y_T - Y_U | U_D = p).$$

[2]The primitives of the model guarantee that the intervention has a monotonic impact on takeup: there are no "defiers" that would have received the treatment at baseline but do not receive it given the intervention.

The MTE is defined for a particular value of p, but it can be informative to plot the function $MTE(p)$ as the potential fraction treated p increases from 0 to 1. If the outcome Y represents the gain from treatment in dollars, then the MTE can be interpreted as the willingness to pay for treatment for an individual at the margin of selecting into treatment. In this special case, the function $MTE(p)$ is a demand function. If the outcome Y instead represents the cost of treatment in dollars, then the function $MTE(p)$ is a marginal cost function. More generally, Y can represent any outcome that could be affected by treatment, in dollars or any other units. In the OHIE context, Y is a measure of emergency room utilization.

The marginal treatment effect is the difference between the **marginal treated outcome (MTO)**, and the **marginal untreated outcome (MUO)**:

$$MTO(p) = E(Y_T|U_D = p)$$
$$MUO(p) = E(Y_U|U_D = p).$$

I sometimes refer to the marginal untreated outcome $MUO(p)$ as the **marginal selection effect $MSE(p)$** because it identifies selection. A treatment effect should only be observed for individuals who change treatment status. Therefore, any change in the untreated outcome as the fraction treated increases reflects only selection. In the OHIE context, the difference in ER utilization between the uninsured lottery losers and the uninsured lottery winners identifies the slope of $MUO(p)$. Under the **marginal untreated outcome test for selection**, if $MUO(p)$ is not constant in any range of p, then there must be selection in that range.

The marginal untreated outcome test for selection generalizes the Einav et al. [2010] cost curve test for selection in insurance markets because it can be applied to any outcome Y and any treatment D. In the Einav et al. [2010] special case, Y is insurer cost and D is an indicator for enrollment in a generous insurance plan relative to a basic plan. If marginal insurer cost in the ungenerous plan decreases as enrollment in the generous plan p increases, then higher-cost individuals have selected into the generous plan. In this special case, a downward-sloping $MUO(p)$ indicates **adverse selection** into the generous plan, and an upward-sloping $MUO(p)$ indicates **advantageous selection** into the generous plan.

The slope of the marginal treated outcome function $MTO(p)$ reflects treatment effect heterogeneity as well as selection. In the OHIE context, $MTO(p)$ describes how the ER utilization of the marginal insured individual changes as coverage increases. If there is no treatment effect heterogeneity, then $MTO(p)$ reflects selection in the same way that $MUO(p)$ reflects selection: a downward slope indicates that individuals with higher values of the outcome have selected into treatment. If there is no selection, then $MTO(p)$ reflects how the treatment effect changes as the fraction treated increases: a downward slope indicates that individuals with bigger differences between the treated and untreated potential outcomes (and hence more to gain

from treatment) have selected into treatment. In the general case with treatment effect heterogeneity and selection, the slope of $MTO(p)$ at each fraction treated p depends on the sign and magnitude of the selection and treatment effects.

The marginal treatment effect $MTE(p)$ isolates the treatment effect from $MTO(p)$ by purging out selection identified by $MUO(p)$. In the special case where the treatment D represents insurance and the outcome Y represents insurer cost, the treatment effect identified by $MTE(p)$ is known as **moral hazard**. Moral hazard need not be the same across all individuals: $MTE(p)$ identifies how moral hazard varies with selection. Previous research has referred to the way that moral hazard varies with selection as "selection on moral hazard" (Einav et al. [2010]), but I prefer to refer to it simply as moral hazard to avoid confusing it with selection.[3]

2.3 The External Validity of An Experiment

A treatment effect recovered from an experiment is **globally externally valid** if the MTE is constant for all p, including $p < 0$ and $p > 1$. I define one treatment effect to be **locally externally valid** for another if both treatment effects are equal. Empirically, the local average treatment effect (LATE) from the OHIE might not be globally externally valid, but it could be locally externally valid for other treatment effects of interest.

3 Identifying and Estimating the MTE with an Experiment

3.1 Identifying Average Characteristics of Always Takers, Compliers, and Never Takers from an Experiment

Identification of the MTE with an experiment relies on the same information that Abadie [2003] uses to identify the average characteristics of always takers, never takers, and compliers. Recall that always takers are individuals with $0 \leq U_D \leq p_B$; compliers are individuals with $p_B < U_D \leq p_I$; and never takers are individuals with $p_I < U_D \leq 1$. Because U_D is distributed uniformly, the share of always takers is p_B; the share of compliers is $(p_I - p_B)$; and the share of never takers is $(1 - p_I)$. An experiment is **internally valid** if the distribution of the unobserved net cost of treatment U_D is the same among lottery winners and losers.

[3] Previous attempts to separate selection from moral hazard in insurance markets often conflate the two, especially if moral hazard varies. For example, under the Chiappori and Salanie [2000] "positive correlation" test, a correlation between insurance coverage and insured spending could indicate moral hazard as well as selection. Under the Finkelstein and Poterba [2014] "unused observables" test, a correlation between a covariate and insurance coverage a second correlation between the same covariate and insured spending could indicate moral hazard as well as selection. Under the Einav et al. [2013] cost curve test, an insured marginal cost curve $MTO(p)$ that is not constant could indicate moral hazard as well as selection. However, the cost curve test isolates the selection when applied to $MUO(p)$, and it isolates moral hazard when applied to $MTE(p)$.

4

The key to identification of average characteristics is that internal validity implies that the shares of always takers, compliers, and never takers are the same among lottery winners and losers. The individuals who go untreated despite winning the lottery identify the average characteristics of never takers: $E(X|D=0,Z=1)$, and the individuals who gain treatment despite losing the lottery identify the average characteristics of always takers: $E(X|D=1,Z=0)$. The average characteristics of the individuals who lose the lottery and go untreated $E(X|D=0,Z=0)$ are a weighted average of the average characteristics of never takers and **untreated compliers**: compliers who lose the lottery. Because we know the share of compliers and the average characteristics of never takers, we can calculate the average characteristics of untreated compliers via

$$\frac{1}{p_I - p_B}\left[(1-p_B)E(X|D=0,Z=0) - (1-p_I)E(X|D=0,Z=1)\right]. \tag{1}$$

Similarly, the average characteristics of the treated individuals who win the lottery $E(X|D=1,Z=1)$ are a weighted average of the average characteristics of always takers and **treated compliers**: compliers who win the lottery. We can calculate the average characteristics of treated compliers via

$$\frac{1}{p_I - p_B}\left[p_I E(X|D=1,Z=1) - p_B E(X|D=1,Z=0)\right]. \tag{2}$$

Because the untreated and treated compliers should have the same characteristics, we can take a weighted average of both groups using the sample weights $s(p_B)$ and $s(p_I)$ to obtain an estimate of the average characteristics of all compliers.[5]

In practice, few experimenters report the average characteristics of compliers, never takers, and always takers. Those who do hope that the compliers will have similar characteristics to the always takers and never takers. If average characteristics are statistically the same across all groups, then they have more confidence that the LATE that they estimate from the experiment will be valid in other contexts. However, if the average characteristics of compliers are different from the average characteristics of other groups, they simply acknowledge the difference and proceed to estimate the LATE. With MTE methods, experimenters can use the information embodied in the comparison of compliers to always takers and never takers to bound or estimate a marginal treatment effect *function* that generalizes the LATE.[6]

[4] Covariates can be used to test internal validity. If the lottery winners do not have the same same average characteristics as the losers, then it is unlikely that the unobserved net cost of treatment U_D is the same among lottery winners and losers.

[5] We can also compare the treated and untreated compliers to test internal validity because both groups should have the same characteristics if the randomization was conducted correctly.

[6] Identification of the average characteristics of always takers, never takers, and compliers requires cross-tabulations of the data by the treatment D as well as the instrument Z. In contrast, identification of the LATE only requires a tabulation of the outcome Y by the instrument Z and a separate tabulation of the treatment D by the instrument Z. In fact, even if the outcome Y and the treatment D are only available in separate datasets, then the LATE can still be obtained. It is not surprising, then, that additional information from cross-tabulations can yield a richer understanding of the treatment effect.

3.2 Identifying Bounds on Outcomes and Treatment Effects with an Experiment

Applying (1) and (2) to an outcome Y in lieu of a characteristic X identifies the **local average untreated outcome (LAUO)**: the average outcome of untreated compliers, and the **local average treated outcome (LATO)**: the average outcome of treated compliers:

$$LAUO = \frac{1}{p_I - p_B}\left[(1-p_B)BUUO - (1-p_I)IUUO\right] \tag{3}$$

$$LATO = \frac{1}{p_I - p_B}\left[p_I ITTO - p_B BTTO\right], \tag{4}$$

in terms of the **intervention untreated untreated outcome (IUUO)**: $E(Y|D=0,Z=1)$; the **baseline untreated untreated outcome (BUUO)**: $E(Y|D=0,Z=0)$; the **baseline treated treated outcome (BTTO)**: $E(Y|D=1,Z=0)$, and the **intervention treated treated outcome (ITTO)**: $E(Y|D=1,Z=1)$. The **local average treatment effect (LATE)** is the difference between the LATO and the LAUO.

The average outcome of untreated compliers (LAUO) relative to the average outcome of never takers (IUUO) identifies the average slope of the marginal selection function $MUO(p)$ from p_B to 1. Untreated compliers have lower net unobserved costs of treatment ($p_B < U_D \leq p_I$) than never takers ($p_I < U_D \leq 1$), so all of the untreated compliers select into treatment before all of the never takers. If the untreated compliers have a higher average outcome than the never takers ($LAUO > IUUO$), then the marginal selection function $MUO(p)$ slopes downward on average from p_B to 1. In the OHIE context, if the untreated compliers have higher average uninsured ER utilization than the never takers, then there is adverse selection on average from the baseline level of coverage to full coverage.

Similarly, the average outcome of treated compliers (LATO) relative to the average outcome of always takers (BTTO) identifies the slope of the marginal treated outcome function $MTO(p)$ from 0 to p_I. Always takers have lower net unobserved costs of treatment ($0 < U_D < p_B$) than treated compliers ($p_B < U_D < p_I$), so all of the always takers select in to treatment before all of the treated compliers. If the always takers have a higher average outcome than the treated compliers ($BTTO > LATO$), then $MTO(p)$ slopes downward on average from 0 to p_I. In the OHIE context, if the always takers have higher average insured ER utilization than the treated compliers, then there could be adverse selection or a decreasing treatment effect as coverage increases from zero to the level of coverage under the intervention.

The assumption that the marginal treated and untreated outcome functions $MTO(p)$ and $MUO(p)$ are weakly monotonic in p yields bounds on the average treatment effects for always takers and never takers.

The average treatment effect for always takers, **the baseline treated treatment effect (BTTE)** is the difference between the $BTTO$ and the **baseline treated untreated outcome (BTUO)**. The BTUO is not observed. However, under weak monotonicity of $MUO(p)$ the average untreated outcome of compliers provides a bound: $(BTUO \leq LAUO < IUTO)$ or $(BTUO \geq LAUO > IUTO)$. Similarly, the average treatment effect for never takers, **the intervention untreated treatment effect (IUTE)** is equal to the difference between the **intervention untreated treated outcome (IUTO)** and the $IUUO$. Under weak monotonicity of $MTO(p)$, the average treated outcome of compliers provides a bound on the $IUTO$: $(IUTO \leq LATO < BTTO)$ or $(IUTO \geq LATO > BTTO)$. In the OHIE context, these bounds could be useful to ER providers who want to bound ER utilization in the full lottery sample if Oregon were to change its intervention to provide health insurance to no one or to require everyone to have coverage.

In the case where $LAUO > IUUO$ and $BTTO < LATO$, the implied upper bound on the average treatment effect for always takers (BTTE) that is informative about the global external validity of the LATE because it is strictly less than the LATE.[7] In the same case, the implied lower bound on the average treatment effect for never takers ($IUTE$) is informative because it is strictly greater than the LATE. The combination of both bounds implies $BTTE < LATE < IUTE$. Similarly, in the case where $LAUO < IUUO$ and $BTTO > LATO$, the bounds are also informative, implying $BTTE > LATE > IUTE$. Global external validity requires that all treatment effects are equal, so global external validity can be rejected in both cases. In these cases, the marginal untreated outcome and the marginal untreated outcome functions have slopes of the opposite sign, so the difference between them cannot be constant for all p.

In the two remaining cases in which the slopes of the marginal untreated outcome and marginal treated outcome curves have slopes of the same sign, the bounds on outcomes could be useful for bounding outcomes if never takers gained treatment or if always takers lost treatment. However, the implied bounds on treatment effects in these cases are not informative about global external validity. In these cases, it is possible that the treatment effect varies, but weak monotonicity of $MTO(p)$ and $MUO(p)$ alone cannot reject global external validity. Additional structure on $MTO(p)$ and $MUO(p)$ can yield a test of global external validity that is informative in all cases, and it can yield point estimates in lieu of bounds.

[7] The proof proceeds as follows:

$$
\begin{aligned}
LAUO > IUUO \implies & BTUO && \geq LAUO && \text{(by weak monotonicity of MUO(p))} \\
\implies & BTTO - BTTE - BTTO && \geq LAUO && \text{(by } BTTE = BTTO - BTUO - BUUO\text{))} \\
\implies & -BTTE && \geq LAUO - BTTO && \\
\implies & BTTE && \leq BTTO - LAUO && \\
\implies & BTTE && \leq LATE + BTTO - LATO && \text{(by } LATE = LATO - LAUO\text{)} \\
\implies & BTTE && \leq LATE + BTTO - LATO < LATE && \text{(if } BTTO < LATO\text{).}
\end{aligned}
$$

3.3 Identifying the Linear MTE with an Experiment

Brinch et al. [2012] impose structure on $MUO(p)$ and $MTO(p)$ to identify a linear MTE with a binary instrument. They assume that the slope of $MUO(p)$ at every point from 0 to 1 is equal to the average slope of $MUO(p)$ from p_B to 1. They also assume that the slope of $MTO(p)$ at every point from 0 to 1 is equal to the average slope of $MTO(p)$ from 0 to p_I. These assumptions introduce heterogeneity in outcomes within always takers, compliers, and never takers, while preserving the mean outcomes. Furthermore, $MUO(p_I)$ is the untreated outcome of the last complier and the first never taker to select into treatment, and $MTO(p_B)$ is the treated outcome of the last always taker and the first complier to select into treatment.

Under these assumptions, the two points $(\frac{p_B+p_I}{2}, LAUO)$ and $(\frac{p_I+1}{2}, IUUO)$ identify the linear $MUO(p)$ and the two points $(\frac{p_B}{2}, BTTO)$ and $(\frac{p_B+p_I}{2}, LATO)$ identify the linear $MTO(p)$:

$$MUO(p) = \frac{(1+p_I)BUUO - (1+p_B)IUUO}{p_I - p_B} + \frac{2(IUUO - BUUO)}{p_I - p_B}p \tag{5}$$

$$MTO(p) = BTTO - \frac{p_B}{p_I - p_B}(ITTO - BTTO) + \frac{2(ITTO - BTTO)}{p_I - p_B}p. \tag{6}$$

The linear MTE is the difference between the linear marginal treated outcome function $MTO(p)$ and the linear marginal untreated outcome function $MUO(p)$:

$$\begin{aligned} MTE(p) = &\frac{1}{p_I - p_B}(p_I(BTTO - BUUO) + p_B(IUUO - ITTO) + (IUUO - BUUO)) \\ &+ \frac{2}{p_I - p_B}((ITTO - IUUO) - (BTTO - BUUO))p. \end{aligned} \tag{7}$$

Brinch et al. [2012] derive $MTE(p)$ without constructing the LATO and the LAUO using the **average untreated outcome (AUO)**: $\boldsymbol{AUO(p)}$= $E(Y_T|X = x, U_D > p) = \int_p^1 MUO(p)dp$, and the **average treated outcome (ATO)**: $\boldsymbol{ATO(p)}$= $E(Y_T|X = x, U_D \leq p) = \int_0^p MTO(p)dp$. Linearity of $MTO(p)$ and $MUO(p)$ implies linearity of $AUO(p)$ and $ATO(p)$. The two points $(p_B, BUUO)$ and $(p_I, ITTO)$ identify the linear $AUO(p)$, and the two points $(p_B, BTTO)$ and $(p_I, ITTO)$ identify the linear $ATO(p)$:

$$AUO(p) = BUUO - \frac{p_B}{p_I - p_B}(IUUO - BUUO) + \frac{IUUO - BUUO}{p_I - p_B}p \tag{8}$$

$$ATO(p) = BTTO - \frac{p_B}{p_I - p_B}(ITTO - BTTO) + \frac{ITTO - BTTO}{p_I - p_B}p, \tag{9}$$

from which they derive the linear marginal untreated outcome function $MUO(p)$[8] and the linear marginal treated outcome function $MUO(p)$.[9]

[8] $MUO(p) = \frac{d[(1-p)AUO(p)]}{d(1-p)} = -\frac{d[(1-p)AUO(p)]}{dp} = -(1-p)\frac{dAUO(p)}{dp} - AUO(p)$.

[9] $MTO(p) = \frac{d[pATO(p)]}{dp} = p\frac{dATO(p)}{dp} + ATO(p)$

3.4 Identifying Treated Outcomes, and Untreated Outcomes, and Treatment Effects from the Linear MTE from an Experiment

We can construct inframarginal treated outcomes, untreated outcomes, and treatment effects from the linear $MTO(p)$, $MUO(p)$ and $MTE(p)$ by applying general weights $\omega_g(p)$ to all three functions:

$$gTO = \int_0^1 \omega_g(p) MTO(p) dp \tag{10}$$

$$gUO = gSE = \int_0^1 \omega_g(p) MUO(p) dp = \int_0^1 \omega_g(p) MSE(p) dp \tag{11}$$

$$gTE = \int_0^1 \omega_g(p) MTE(p) dp \tag{12}$$

where gTO is the general weighted average **treated outcome (TO)**, $gUO = gSE$ is the general weighted average **untreated outcome (UO)** or **selection effect (SE)**, and gTE is the general weighted average **treatment effect (TE)**. For general weights ω_g, the weighted average treatment effect is equal to the difference between the weighted average treated outcome and the weighted average untreated outcome: $gTE = gTO - gUO$.

Experiments involve several interesting groups of individuals g, each represented by a different set of general weights $\omega_g(p)$. First consider baseline treated individuals, also known as always takers, for whom $0 \leq U_D \leq p_B$. For baseline treated individuals, the weights are

$$\omega_{BT}(p) = \begin{cases} \frac{1}{p_B} & \text{if } 0 \leq p \leq p_B \\ 0 & \text{if } p_B < p \leq 1, \end{cases}$$

as shown in Column 1 of Table 1. We observe the potential treated outcome for these individuals: $E(Y_T | 0 \leq U_D \leq p_B) = E(Y | D = 1, Z = 0) = BTTO$. The $BTTO$ and all other quantities that do not require the linear MTE for identification are depicted in bold in Table 1. We do not observe the baseline treated untreated outcome $BTUO = E(Y_U | 0 \leq U_D \leq p_B)$ because always takers always receive treatment. However, we can calculate it with (11). We can also calculate the average effect of treatment on always takers, the baseline treated treatment effect $BTTE = E(Y_T - Y_U | 0 \leq U_D \leq p_B)$.

Similarly, Column 2 gives weights for the baseline untreated individuals (compliers and never takers). For these individuals, we observe the baseline untreated untreated outcome (BUUO). The linear MTE identifies the baseline untreated treated outcome (BUTO) and the baseline untreated treatment effect (BTTE). Columns 3 and 4 give weights for the intervention treated (IT) individuals (always takers and compliers) and the intervention untreated (IU) individuals (never takers).

Table 1: Treated Outcomes, Untreated Outcome, and Treatment Effects

			(1) Baseline Treated (Always Takers) BT	(2) Baseline Untreated BU	(3) Intervention Treated IT	(4) Intervention Untreated (Never Takers) IU	(5) Randomized Intervention Sample Treated RIST	(6) Randomized Intervention Sample Untreated RISU	(7) Local Average (Compliers) LA	(8) Average A
(a)	MTO(p)	Treated Outcome TO	**BTTO**	BUTO	**ITTO**	IUTO	**RISTTO**	RISUTO	**LATO**	ATO
(b)	MUO(p)	Untreated Outcome UO	BTUO	**BUUO**	ITUO	**IUUO**	RISTUO	**RISUUO**	**LAUO**	AUO
(c)	MTE(p) = MTO(p) - MUO(p)	Treatment Effect TE = TO - UO	BTTE	BUTE	ITTE	IUTE	RISTTE	RISUTE	**LATE**	ATE
(d)	$\omega_g(p) =$	if $0 \leq p \leq p_B$:	$\mathbf{\frac{1}{p_B}}$	**0**	$\mathbf{\frac{1}{p_I}}$	**0**	$\mathbf{\frac{1}{p_B + s(p_I)(p_I - p_B)}}$	**0**	**0**	**1**
		if $p_B < p \leq p_I$:	**0**	$\mathbf{\frac{1}{(1-p_B)}}$	$\mathbf{\frac{1}{p_I}}$	**0**	$\mathbf{\frac{s(p_I)}{p_B + s(p_I)(p_I - p_B)}}$	$\mathbf{\frac{s(p_B)}{1 - s(p_I)p_I - s(p_B)p_B}}$	$\mathbf{\frac{1}{(p_I - p_B)}}$	**1**
		if $p_I < p \leq 1$:	**0**	$\mathbf{\frac{1}{(1-p_B)}}$	**0**	$\mathbf{\frac{1}{(1-p_I)}}$	**0**	$\mathbf{\frac{1}{1 - s(p_I)p_I - s(p_B)p_B}}$	**0**	**1**
(e)	Decomposition	Selection UO/TO	BTUO/**BTTO**	**BUUO**/BUTO	ITUO/**ITTO**	**IUUO**/IUTO	RISTUO/**RISTTO**	**RISUUO**/RISUTO	**LAUO**/**LATO**	AUO/ATO
		Treatment Effect TE/TO	BTTE/**BTTO**	BUTE/BUTO	ITTE/**ITTO**	IUTE/IUTO	RISTTE/**RISTTO**	RISUTE/RISUTO	**LATE**/**LATO**	ATE/ATO
(f)	OLS Estimates	OLS = TTO - UUO	**BOLS = BTTO - BUUO**		**IOLS = ITTO - IUUO**		**RISOLS = RISTTO - RISUUO**		-	-
(g)	OLS Decomposition	Selection (OLS - TE)/OLS	(**BOLS** - BTTE)/**BOLS**	(**BOLS** - BUTE)/**BOLS**	(**IOLS** - ITTE)/**IOLS**	(**IOLS** - IUTE)/**IOLS**	(**RISOLS** - RISTTE)/**RISOLS**	(**RISOLS** - RISUTE)/**RISOLS**	-	-
		Treatment Effect TE/OLS	BTTE/**BOLS**	BUTE/**BOLS**	ITTE/**IOLS**	IUTE/**IOLS**	RISTTE/**RISOLS**	RISUTE/**RISOLS**		

Calculation of the bold quantities does not rely on the linear MTE.

In contrast to the baseline and intervention weights, the randomized intervention sample (RIS) weights that I introduce in Columns 5 and 6 reflect the fraction of the sample that loses the lottery $s(p_B)$. The randomized intervention sample treated treated outcome (RISTTO) and the randomized intervention sample untreated untreated outcome (RISTUUO) are both observed. In the context of the OHIE, they give the average ER utilization among the insured and the uninsured individuals in the experiment, respectively.

Column 7 reports the local average weights, which yield the LATE. Experimenters often refer to the LATE as the "treatment on the treated" estimate, which can be misleading. The LATE gives the treatment effect on compliers, but always takers are also treated. The weights that I have introduced allow me to calculate treatment effects for various treated groups, while the traditional Heckman and Vytlacil [2007] weights for a continuous instrument used by Brinch et al. [2012] only yield one "treatment on the treated" estimate. Using my weights, the baseline treated treatment effect BTTE gives a "treatment on the treated" estimate for always takers; the intervention treated treatment effect ITTE gives a "treatment on the treated" estimate for always takers and compliers; and the randomized intervention sample treated treatment effect RISTTE gives a "treatment on the treated" estimate for always takers and compliers, weighted by the share that loses the lottery $s(p_B)$. The terms LATE, BTTE, ITTE, and RISTTE convey which groups of treated individuals are included, while "treatment on the treated" does not.

Column 6 reports the average weights $\omega_A(p) = 1$. The average weights are equal to the sum of the weights for always takers, compliers and never takers $\omega_A(p) = \omega_{BT}(p) + \omega_{LA}(p) + \omega_{IU}(p)$. The ATO, AUO, and ATE are not observed, but they can be calculated with the linear MTE. In the context of the OHIE, the ATO gives the average ER utilization if all individuals were insured, and the AUO gives the average ER utilization if all individuals were uninsured.

The average weights $\omega_A(p) = 1$ are the only weights in Table 1 that do not expect a specific experimental intervention. Any given experimental intervention shifts the treatment probability from a specific p_B to a specific p_I. By substituting hypothetical values of p_B and p_I in the formulas in Table 1, experimenters can forecast treated outcomes, untreated outcomes, and treatment effects from hypothetical new experimental intervention. The randomized intervention sample weights $\omega_R IS(p)$ reflect the experimental design as well as the experimental intervention because they reflect the share of individuals who win the lottery $s(p_I)$. By substituting hypothetical values of $s(p_I)$ and $s(p_B) = 1 - s(p_I)$ into $\omega_R IS(p)$, experimenters can forecast treated outcomes and untreated outcomes from hypothetical new experiments that alter the share of individuals who lose the lottery.

Estimates of treated and untreated outcomes can be useful for budgeting hypothetical new experiments. In the OHIE context, suppose that Oregon policymakers have the results from the OHIE, and they are contemplating making more coverage available via a new experimental intervention in which lottery winners

who sign up for health insurance will also receive an additional incentive, such as $200 in cash. Suppose that pilot tests indicate that the new intervention treatment probability p_I' will exceed p_I from the OHIE. By computing the RISTTO with cost as an outcome for hypothetical values of p_I' and $s(p_I)$, policymakers can determine how many individuals they can classify as lottery winners in the new experiment while still meeting their budget goals.

3.5 Identifying Optimal Treatment Probabilities with the Linear MTE from an Experiment

The linear MTE allows for positive treatment effects in some groups ($gTE > 0$) and negative treatment effects in others ($gTE < 0$). Consider the case where the MTE is downward-sloping[10]. Define $\mathbf{p}^*$ as the **potential fraction treated p at which the linear MTE is zero**:

$$p^* = -\frac{p_I(BTTO - BUUO) + p_B(IUUO - ITTO) + (IUUO - BUUO)}{2((ITTO - IUUO) - (BTTO - BUUO))}. \tag{13}$$

If $0 < p^* < 1$, then p^* gives the potential share of individuals in the full sample with a positive treatment effect. The downward-sloping MTE indicates that individuals with positive treatment effects select into coverage first, so the first p^* of individuals to select into treatment have a positive treatment effect, and the remaining individuals have a negative treatment effect.

In this case, if a policymaker wants all individuals with positive treatment effects to receive treatment, then p^* gives the optimal value of the intervention treatment probability p_I. If the intervention treatment probability from an experiment is not equal to the optimal probability, then the experimenter can recommend policies that make treatment more or less attractive to bring p_I closer to p^*. The experimenter can then test whether $p_I = p^*$ with a new experiment.

Next consider the case where the MTE is upward-sloping[11] with $0 < p^* < 1$. If a policymaker wants all individuals with positive treatment effects to receive treatment, then the optimal fraction of individuals to treat is $(1 - p^*)$. Unfortunately, it is harder to target those individuals because the first p^* individuals to select into treatment should *not* receive it. In this case, the optimal policy does not simply involve making the treatment more or less attractive for all individuals. Rather, it involves targeting the treatment to the individuals who should receive it.

Instead, it involves it also involves manipulating the net unobserved cost of treatment U_D such that the individuals with positive treatment effects select in to treatment before the individuals with negative treatment effects.

[10] $((ITTO - IUUO) - (BTTO - BUUO)) < 0$
[11] $((ITTO - IUUO) - (BTTO - BUUO)) > 0$

The remaining cases are straightforward. If the linear MTE is always positive in the range $0 < p < 1$, then it is optimal to treat everyone. If the linear MTE is always negative in the range $0 < p < 1$, then it is optimal to treat no one.

If the outcome Y measures the benefit of treatment in dollars, then $MTE(p)$ can be used to calculate the deadweight loss that results from treating a suboptimal fraction of individuals. If the baseline treatment probability is optimal ($p_B = p^*$), then the deadweight loss is equal to the integral of $MTE(p)$ from p_B to p_I, which is also equal to the LATE. Under this interpretation, the LATE is the distortion associated with shifting the treatment probability from the baseline probability p_B to the intervention probability p_I with the intervention. In the Einav et al. [2010] OHIE context, if Y measures the cost to the insurer in the generous plan relative to the ungenerous plan, and p_B is optimal, then the LATE gives the deadweight loss due to moral hazard.[12]

The optimal treatment threshold need not be zero. Suppose that there are two different linear MTE curves: one measures benefit in dollars, and the other measures cost in dollars. Given these two curves, the optimal treatment threshold does not occur at p^* where the benefit MTE intersects zero; it occurs where the benefit MTE intersects the cost MTE (marginal benefit equals marginal cost).

3.6 Decomposing Treated Outcomes from an Experiment into Selection and Treatment Effects

All of the treatment effects in Table 1 have been purged of selection. However, all of the treated outcomes do reflect selection. For any group g of individuals represented by general weights ω_g, we can decompose the treated outcome into shares due to selection and treatment effects as follows:

$$\underbrace{\frac{gSE}{gT0}}_{\text{selection}} + \underbrace{\frac{gTE}{gT0}}_{\text{treatment}} = 1,$$

because $gTO = gSE + gTE$. In the OHIE context, this decomposition tells us what share of insured ER utilization in any group g is due to the composition of the group as opposed to moral hazard in that group.

We can also decompose a *change* in treated outcomes across groups into selection and treatment effects. It is particularly useful to understand the change in the treated outcome induced by an experiment: $ITTO -$

[12]The LATE does not give the deadweight loss due to selection, which has been purged from the MTE.

$BTTO$. We can decompose the difference into shares due to selection and treatment as follows[13]:

$$\underbrace{\frac{IUUO - BUUO}{ITTO - BTTO}}_{\text{selection}} + \underbrace{\frac{(ITTO - IUUO) - (BTTO - BUUO)}{ITTO - BTTO}}_{\text{treatment}} = 1.$$

This decomposition requires the same assumptions as the linear MTE, but it can be calculated directly from observable quantities even if the linear MTE has not been calculated. Different experiments start from different baseline probabilities of treatment p_B and induce different intervention probabilities of treatment p_I. However, this decomposition should give the same outcome regardless of the experiment.[14] In the OHIE context, this decomposition tells us what share of the change in insured ER utilization induced by the experiment is due to selection relative to moral hazard. I also implement a similar decomposition to assess what share of the difference in utilization between always takers and compliers is due to selection relative to moral hazard.

3.7 Decomposing OLS Estimates from an Experiment into Selection and Treatment Effects

Consider an OLS regression run on the sample of individuals that lose the lottery, the baseline individuals. The **baseline OLS (BOLS)** estimate is the difference between the baseline treated treated outcome (BTTO) and the baseline untreated untreated outcome (BUUO). BOLS can be affected by selection because it compares the observed outcome for a group of treated individuals, $BTTO = E(Y_T|0 \leq U_D \leq p_B)$, to the observed outcome for a different group untreated individuals $BUUO = E(Y_U|p_B \leq U_D \leq 1)$, instead of comparing it to the potential untreated outcome for the treated individuals $BTUO = E(Y_U|0 \leq U_D \leq p_B)$. We can decompose BOLS into shares due to selection and treatment effects as follows:

$$\underbrace{\frac{BTTE - BOLS}{BOLS}}_{\text{selection}} + \underbrace{\frac{BTTE}{BOLS}}_{\text{treatment}} = 1.$$

An alternative reason for why BOLS can be affected by selection is that it compares the observed outcome for a group of untreated individuals $BUUO = E(Y_U|p_B \leq U_D \leq 1)$ to the observed outcome for a different treated individuals $BTTO = E(Y_T|0 \leq U_D \leq p_B)$, instead of comparing it to the potential treated outcome for the untreated individuals $BUTO = E(Y_T|p_B \leq U_D \leq 1)$. Therefore, we can also decompose BOLS into

[13] This follows because $ITTO = ITUO + ITTE$ and $BTTO = BTUO + BTTE$. Furthermore, $ITUO - BTUO = IUUO - BUUO$.

[14] To see this, note that the slope of the linear marginal untreated outcome function MUO relative to the slope of the linear marginal treated outcome function MTO determines the share due to selection.

shares due to selection and treatment effects as follows:

$$\underbrace{\frac{BUTE - BOLS}{BOLS}}_{\text{selection}} + \underbrace{\frac{BUTE}{BOLS}}_{\text{treatment}} = 1.$$

Similarly, the **intervention OLS (IOLS)** estimate is the difference between the intervention treated treated outcome (ITTO) and the intervention untreated untreated outcome (IUUO). We can decompose IOLS into shares due to selection and treatment effects in two ways, as shown in Row e of Table 1. Selection and treatment effects can vary from BOLS to IOLS, so there is no reason to expect that their respective decompositions will yield the same answers.

Rather than estimating BOLS and IOLS separately, experimenters often report **randomized intervention sample OLS (RISOLS)**, the OLS estimate on the full sample. RISOLS is a weighted average of BOLS and IOLS, weighted by the share of individuals lotteried out $s(p_B)$. RISOLS is also equal to the difference between the randomized intervention sample treated treated outcome (RISTTO) and the randomized intervention sample untreated untreated outcome (RISUUO). We can decompose RISOLS into the shares due to selection and treatment effects in two ways as as shown in Row e of Table 1.

Experimenters often compare LATE to RISOLS with the intent of obtaining the share of RISOLS due to the treatment effect. If there is no treatment effect heterogeneity, then $(LATE/RISOLS) = (RISTE/RISOLS) = (RISUE/RISOLS)$, and all three fractions give the share of the OLS estimate due to the treatment effect. However, if there is treatment effect heterogeneity, then RISTE and RISUE are directly comparable to RISOLS because they reflect the same mix of individuals, but LATE is not directly comparable to RISOLS because it only reflects compliers.

Although it is common to report RISOLS, it is not a very informative statistic for two reasons. First, unlike BOLS, it reflects the impact of the experimental intervention. Second, unlike BOLS and IOLS, RISOLS reflects the share of the sample that loses the lottery $s(p_B)$, so it changes with the experimental design. I recommend reporting BOLS and IOLS in addition to RISOLS. Under the assumptions required to identify the linear MTE, the comparison of BOLS to IOLS provides a test for global external validity.

3.8 Difference-in-Difference Test

Angrist [2004], Brinch et al. [2012], and Bertanha and Imbens [2014] propose tests for global external validity that I implement using the following difference-in-difference regression:

$$Y = \lambda_{DZ} DZ + \lambda_D D + \lambda_Z Z + \lambda, \tag{14}$$

where Y is the outcome, λ_D is the coefficient on the binary indicator for selecting into the treatment D, λ_Z is the coefficient on the binary indicator for winning the lottery Z, λ_{DZ} is the coefficient on the interaction of selecting into treatment and winning the lottery, and λ is the coefficient on the constant term. This regression compares four observable experimental average outcomes: the baseline treated treated outcome $BTTO = E(Y|D = 1, Z = 0)$; the baseline untreated untreated outcome $BUUO = E(Y|D = 0, Z = 0)$; the intervention treated treated outcome $ITTO = E(Y|D = 1, Z = 1)$; and the intervention untreated untreated outcome $IUUO = E(Y|D = 0, Z = 1)$.

The coefficient λ_D is equal to $BTTO - BUUO$, which is equal to the baseline OLS estimate $BOLS$. On its own, λ_D does not inform the presence or absence of selection or a heterogeneous treatment effect. Even if $\lambda_D = 0$, there could be selection and a heterogeneous treatment effect that balances it.

The coefficient λ_Z is equal to $IUUO - BUUO$. If $\lambda_Z = 0$, then there is no selection. However, the absence of selection does not imply global external validity because the treatment effect can still be heterogeneous. In general, even if there is no selection, $BOLS \neq IOLS \neq RISOLS \neq LATE$.

The coefficient λ_{DZ} is equal to $((ITTO - IUUO) - (BTTO - BUUO))$, which is equal to $IOLS - BOLS$. If and only if IOLS is equal to BOLS, then $\lambda_{DZ} = 0$, and any treatment effect derived from the linear MTE from an experiment is globally externally valid. When this condition holds, the linear MTE has zero slope, per (7), so there is no treatment effect heterogeneity.

The regression in (14) makes these tests simple to implement. The asymptotic or bootstrapped standard errors from the regression provide direct tests for whether each coefficient is equal to zero. The joint test of $\lambda_{DZ} = \lambda_D = 0$, which tests whether the treatment effect is globally externally valid *and* equal to zero can be implemented as a post-estimation t-test.

The difference-in-difference test can reject external validity even if $MTO(p)$ and $MUO(p)$ have slopes of the same sign. In contrast, the bounds that I introduced in Section 3.2 can only reject global external validity if $MTO(p)$ and $MUO(p)$ have slopes of the opposite sign because they impose less structure. The bounds impose weak monotonicity of $MTO(p)$ and $MUO(p)$, but the difference-in-difference test imposes linearity of $MTO(p)$ and $MUO(p)$. An experimenter willing to impose linearity of $MTO(p)$ and $MUO(p)$ for the difference-in-difference test can recover the linear MTE and all of the quantities derived from it without imposing further assumptions.

3.9 Difference-in-Difference Test Using Covariates

We can incorporate covariates into the difference-in-difference test to formalize the comparison of the characteristics of always takers, never takers, and compliers discussed in Section 3.1. Suppose that we implement (14) using a single covariate from the vector X as the dependent variable in lieu of the outcome Y. In this

implementation, the coefficient λ_D tests whether the observable characteristic is related to baseline takeup; the coefficient λ_Z tests whether the experiment induces selection on that observable characteristic; and the coefficient λ_{DZ} tests whether the observable characteristic has a different relationship to intervention takeup than it does to baseline takeup.

We can obtain further insight by regressing the outcome Y on the same covariate in the sample of lottery losers. Using the estimated coefficients, we can obtain a predicted outcome for all lottery losers and winners, and we can use that predicted outcome as the dependent variable in a new difference-in-difference test. If we find a nonzero coefficient using the actual outcome, but we do not reject that the coefficients is equal to zero using the predicted outcome, then we have found an observable basis for baseline takeup, selection, or selection on the treatment effect, respectively.

We can also implement a more powerful test by predicting the outcome Y using the entire vector of covariates X in the sample of lottery losers. If we cannot reject zero for all coefficients in the resulting difference-in-difference test, then we can be more confident that all selection has an observable basis. In insurance markets, if there is an observable basis for selection, then pricing or risk adjustment on that observable basis could alleviate or eliminate welfare losses (Bundorf et al. [2012].

3.10 Using Covariates for Subgroup Analysis and Sample Re-weighting with Linear MTE Methods

Subgroup analysis divides the sample into subgroups, generally one covariate at a time, and compares the LATEs estimated within each subgroup. If the LATE on the full sample is globally externally valid, then the LATE should be the same in each subgroup. The linear MTE should also be the same within each subgroup. However, if the LATE on the full sample is not globally externally valid, then the LATE and MTE need not be the same in each subgroup. Even if the linear MTE is the same within each subgroup, each subgroup s can have a different support from p_{Bs} to p_{Is} and hence a different LATE.

If subgroup analysis yields different LATEs in each subgroup, experimenters often re-weight the experimental sample to extrapolate the results to a second sample. Sample re-weighting should produce relevant results in the second sample if the LATE in each subgroup is globally externally valid. However, if the LATE in each subgroup is not globally externally valid, then the experimenter should re-weight using the MTE in each subgroup, taking p_{Bs} and p_{Is} in the second sample into account.

3.11 Identifying a Linear or Nonlinear MTE with an Experiment and Covariates

We can combine information across subgroups to estimate richer MTE functions. Within each subgroup s, the two points $(p_{Bs}, BUUO_s)$ and $(p_{Is}, IUUO_s)$ identify a linear average cost curve for the untreated, and the two points $(p_{Bs}, BTTO_s)$ and $(p_{Is}, ITTO_s)$ identify a linear average cost curve for the treated. However, if we assume that the MTE is the same across subgroups, then we have four points to identify linear or *nonlinear* average cost curves for the treated and untreated. By further subdividing the sample into finer subgroups, we can achieve nonparametric identification of the MTE. Furthermore, if we are willing to impose some structure on how covariates enter the MTE, then we need not assume that the MTE is the same across subgroups.

Brinch et al. [2012] specify the following functional forms that impose some structure on how covariates enter the MTE, MTO, and MUO while allowing them to differ across subgroups with the same vector of characteristics $X = x$:

$$MTE(x,p) = E(Y_T - Y_U | X = x, U_D = p) = (\beta_T - \beta_U)'x + (mto(p) - muo(p)). \tag{15}$$

$$MTO(x,p) = E(Y_T | X = x, U_D = p) = \beta_T' x + mto(p) \tag{16}$$

$$MUO(x,p) = E(Y_U | X = x, U_D = p) = \beta_U' x + muo(p). \tag{17}$$

where the first component of each function depends on the observed vector of characteristics x, and the second component depends on the unobserved net cost of treatment U_D. Variation across subgroups in the observed outcomes identifies the additive shifts $\beta_T' x$ and $\beta_U' x$. Variation across subgroups in the unobserved net cost of treatment p_{Bx} and p_{Ix} identifies the slopes $mto(p)$ and $muo(p)$.[15]

The additive separability imposed by (15)-(17) is weaker than additive separability often imposed by experimenters. For example, by including covariates additively in an instrument variable regression model, experimenters assume that covariates induce the same additive shift in the observed outcome Y regardless of treatment status. In contrast, (15)-(17) allow covariates to shift the treated potential outcome Y_T and the untreated potential outcome Y_U by different amounts.

The experimenter can summarize heterogeneity in the treatment effect across subgroups by graphing the average MTE over all individuals i in the experiment: $\overline{MTE(x,p)} = \sum_i (\beta_T - \beta_U)' x_i + (mto(p) - muo(p))$. To demonstrate the maximum range of treatment effects, the experimenter can also graph the MTEs with the smallest and largest observable components: $\min MTE(x,p) = \min_i (\beta_T - \beta_U)' x_i + (mto(p) - muo(p))$ and $\max MTE(x,p) = \max_i (\beta_T - \beta_U)' x_i + (mto(p) - muo(p))$. A table that reports $(\beta_T - \beta_U)' x$ for each

[15] I specify the slopes in lowercase to avoid confusion with $MTO(p)$ and $MUO(p)$, the linear marginal treated and untreated outcome functions that do not depend on x.

subgroup and $(mto(p) - muo(p))$ for the reference subgroup gives all information necessary to construct $MTE(x,p)$ for each subgroup.

3.12 Identifying Treated Outcomes, Untreated Outcomes, and Treatment Effects from the Linear or Nonlinear MTEs with Covariates from an Experiment

We can construct inframarginal treated outcomes, untreated outcomes, and treatment effects from the linear or nonlinear $MTO(x,p)$, $MUO(x,p)$ and $MTE(x,p)$ by applying general weights $\omega_g(x,p)$ to all three functions:

$$gTO(x) = \int_0^1 \omega_g(x,p) MTO(x,p) dp \tag{18}$$

$$gUO(x) = gSE(x) = \int_0^1 \omega_g(x,p) MUO(x,p) dp = \int_0^1 \omega_g(x,p) MSE(x,p) dp \tag{19}$$

$$gTE(x) = \int_0^1 \omega_g(x,p) MTE(x,p) dp \tag{20}$$

the weights $\omega_g(x,p)$ are the same as those given in Row D of Table 1, except that they are a function of the predicted propensity score p_{Bx} or p_{Ix} for individuals with covariate vector x in lieu of the full sample p_B or p_I. These weights are very different from the weights graphed by Brinch et al. [2012]. The weights that I have developed here explicitly account for the binary instrument. Rather than reporting (18)-(20) at a single value of x, I evaluate them for every x in the sample and report the averages $\overline{gTO(x)}$, $\overline{gUO(x)}$, and $\overline{gTE(x)}$.

3.13 Identifying Optimal Treatment Probabilities with the MTE from an Experiment and Covariates

Brinch et al. [2012] do not discuss the possibility of using the MTE model that incorporates covariates to identify heterogeneous treatment effects across subgroups. Such an exercise could be useful to an experimenter who wants to predict which groups of individuals are likely to react positively or negatively to an intervention. Identification of positive and negative treatment effects from an MTE with covariates is similar to identification in the case without covariates discussed in Section 3.5, but the predictions are richer.

Suppose that the linear MTE or the nonlinear MTE with covariates is weakly downward-sloping. In some subgroups of individuals with covariate vector $X = x$, the observable component of the MTE, $(\beta_T - \beta_U)'x$, might be large enough that the MTE is always above zero for all potential treated fractions $0 \leq p \leq 1$. In those subgroups, all members have a positive treatment effect. Similarly, in other subgroups, $(\beta_T - \beta_U)'x$

might be negative enough that the MTE is always below zero for all potential treated fractions $0 \leq p \leq 1$. In those subgroups, all members have a negative treatment effect.

In any remaining subgroups, the MTE crosses zero at $0 < p_x^* < 1$. Individuals in those subgroups will have a positive treatment effect if they have unobserved net costs of treatment U_D less than the threshold p_x^* in their subgroup; they will have a negative treatment effect otherwise. Equivalently, the first p_x^* of individuals in the subgroup to sign up for treatment will have a positive treatment effect, and the remaining individuals will have a negative treatment effect. Thus, the sign of the treatment effect depends on the potential fraction treated. If a policymaker wants all individuals with positive treatment effects to receive treatment, then p_x^* gives the optimal fraction of the subgroup to treat.

Suppose that the linear MTE is upward-sloping or that a nonlinear MTE is upward-sloping for some $0 < p_x^* < 1$. In these cases, analysis follows as discussed in Section 3.5. the optimal policy will shift the order that individuals select into treatment by altering U_D.

3.14 Extrapolating an MTE to Another Experiment on the Same Sample

Suppose that the only difference between two different experiments on the same sample is that they involve different baseline and intervention treatment probabilities p_I and p_B. In this case, both experiments should recover the same MTE functions. In practice, if the MTE varies across experiments, then we can conclude, in the spirit of a Hausman [1978] test, that one of the experiments is invalid or that an assumption required for MTE estimation is violated.

If anything other than the treatment probabilities varies across experiments, then the MTE functions might not be comparable. For example, Y must be measured in the same way across experiments for the MTEs to be the same. Even within a single experiment, the MTE for one outcome can be upward-sloping while the MTE for another outcome can be downward-sloping. For example, in the OHIE context, if primary care utilization and ER utilization are substitutes for all individuals, then the MTE for primary care can be upward-sloping even as the MTE for ER utilization is downward-sloping.

The treatment D must also be measured in the same way across experiments the MTEs to be the same. Different measures of treatment result in different intervention and treatment probabilities p_I and p_B. They also result in different marginal treated and untreated outcome curves because the individuals used to identify those curves vary as the definition of treatment varies.

For the MTE with covariates, the entire vector of covariates must be available and measured in a standardized way for the MTEs across experiments to be the same. It is tempting to think of the linear MTE is an approximation to the MTE with covariates. However, the covariates change the interpretation of the unobserved net cost of treatment U_D. As more covariates are included in model, they are purged from the

residual unobserved net cost of treatment U_D in the spirit of Altonji et al. [2005]. In the limit, if every element of the unobservable becomes observed, then the MTE becomes a horizontal line.

Identification of the MTE with an experiment requires that the experiment changes the threshold for selection into treatment, but it does not change the order in which individuals select into treatment given by U_D. If one experiment does not change U_D but a second experiment does, then the MTE functions identified by both experiments need not be the same. The MTE could be upward-sloping in one experiment and downward-sloping in another if one experiment changes the ordering of selection into treatment. However, the ATEs recovered from the MTEs from both experiments will be the same.

Suppose that the only difference between a proposed experiment and an existing experiment on the same sample is that one experiment offers a larger treatment incentive than the other, resulting in a larger intervention treated fraction p_I. In this case, it should be reasonable for the experimenter to extrapolate from the MTE function from one experiment to the next. However, even in this case, the experimenter should exercise caution if the intervention treated fraction p_I induced by the new experiment is far outside of the support of the existing experiment.

3.15 Extrapolating to the Experiment Interpreted as a Natural Experiment

Any experiment can be interpreted as a natural experiment that took place in the post-period but not in the pre-period for lottery winners. If some individuals receive the treatment in the pre-period, ensuring that there are some always-takers, then we can estimate an MTE with the natural experiment and compare it to the MTE estimated with the randomized experiment. Assumign that the conditions require to compare two different experiments on the same sample from Section 3.14 are met, the MTEs from both experiments should equal, so we can compare them with a Hausman [1978] test. In the OHIE context, no individuals receive insurance in the pre-period because they must be uninsured to enter the lottery, so we cannot estimate a separate MTE using the natural experiment. However, we can use the observed change in outcomes from the pre-period to the experimental period ($Y - Y_{pre}$) to validate the predictions from the MTE estimated with the randomized experiment.

3.16 Extrapolating the MTE to an Experiment on a Different Sample

I provide several cases under which it is possible to compare MTEs from experiments on different samples. Each case imposes some structure on how the samples are drawn from a broader pool. In each case, assume that the conditions required to compare two different experiments on the same sample from Section 3.14 are met.

In the simplest case, suppose that both samples are drawn at random from a broader pool. This case

differs from the comparison of different experiments on the same sample because the same individuals need not be in the experimental data for both experiments. In this case, the MTEs should be the same for both experiments.

Next, suppose that both samples are not strictly drawn at random from the broader pool because the second experiment over-samples individuals with a characteristic that is observable in both samples. Further suppose that the observable characteristic is independent of the unobserved net cost of treatment U_D in the experiment that does not over-sample. In this case, the potential fraction treated p has the same meaning in both samples, and the MTEs should be the same for both experiments.

In a related case, suppose that the second experiment over-samples individuals with a characteristic that is observable in both samples, but that characteristic is *not* independent of the unobserved net cost of treatment U_D in the first experiment. In this case, the linear MTEs need not be equal because the potential fraction treated p does not have the same meaning in both examples. However, by specifying the correct model of over-sampling, it should be possible to construct MTEs with covariates that are the same for both experiments.

In a different case, suppose that the first experimental sample is drawn at random from the broader pool and the second experimental sample is drawn at random from a set of individuals who sign up for a lottery from the broader pool. To compare the MTEs, the experimenter must take a stand on the range of unobserved cost of treatment U_D from the broader pool that is represented by the individuals who sign up for the lottery. Suppose that the distribution of observable characteristics does not differ across samples. One natural assumption is that the fraction f of individuals who sign up for the lottery would be the first to select into treatment in the broader sample. If this assumption holds, then the MTE from $0 \leq p \leq 1$ on the individuals who sign up for the lottery should be equal to the MTE from $0 \leq p \leq f < 1$ on the broader sample. In this case, extrapolation from the sample of individuals who sign up for the lottery to the broader pool requires extrapolation to potential treated fractions that exceed full treatment: $p > 1$.

3.17 Estimating the MTE from an Experiment

The linear MTE is exactly identified, so estimation is straightforward. Estimation of the quantities derived from it is also straightforward because they can be expressed in closed form. Confidence intervals for the linear MTE function and all quantities derived from it can be obtained via bootstrapping.[16]

Estimation of the MTE with covariates requires more assumptions. I detail my preferred algorithm for estimation of the MTE with covariates via a global polynomial in Appendix A.[17] After choosing the

[16] I bootstrap by household ID for 200 replications, and I report the 2.5 and 97.5 percentiles as the 95% confidence interval. I obtain significance stars for other intervals by constructing analogous confidence intervals.

[17] I intend to make Stata code available.

order of the global polynomial, I estimate propensity scores and the average treated and untreated outcome functions $ATO(x,p)$ and $AUO(x,p)$. From those estimates, I construct estimates of the marginal treated and untreated outcome functions $MTO(x,p)$ and $MUO(x,p)$ and the marginal treatment effect. I obtain confidence intervals for the MTE with covariates and all quantities derived from it via bootstrapping.

In theory, higher order global polynomials offer greater flexibility. In practice, if the common support of the treated and untreated propensity scores from the experiment do not span potential treated fractions from $0 \leq p \leq 1$, then it is desirable to impose more structure on the MTE to extrapolate beyond the experimental support. Other than the LATE, all treatment effects in Table 1 require extrapolation beyond the experimental support. Extrapolation is less reliable with higher order polynomials because the estimated function decreases to negative infinity or increases to infinity just outside of experimental support.

In theory, local polynomials also offer greater flexibility. However, local polynomials can only be estimated within the support, and extrapolation requires ad hoc assumptions. Furthermore, local polynomial estimation often results in functions that are not smooth. If the average treated and untreated outcome functions $ATO(x,p)$ and $AUO(x,p)$ are not smooth, then the functions derived from their slopes, the $MTO(x,p)$, $MUO(x,p)$, and $MTE(x,p)$, are subject to large fluctuations that might not be merited by the underlying data.

4 Application: The Oregon Health Insurance Experiment

4.1 OHIE Replication Results

As a first step toward estimating marginal treatment effects in the OHIE, I replicate the main LATE estimates reported in Taubman et al. [2014] using publicly-available Oregon administrative data. Following Taubman et al. [2014], I specify the endogenous variable D as an indicator for any Medicaid coverage, which includes Medicaid coverage obtained via the lottery or the existing Medicaid eligibility guidelines. I refer to individuals with $D = 1$ as "treated" or "insured," and individuals with $D = 0$ as "untreated" or "uninsured."[18] I examine three measures of emergency room utilization Y: an indicator for any ER visit, a count of the number of ER visits, and dollar amount of ER total charges.[19] All three measures include individuals with zero visits.

The first column of Table 2 replicates the main results from Taubman et al. [2014] using the full sample of

[18] Several individuals with $D = 0$ gained health insurance coverage through other means, but they were still "untreated" and "uninsured" by Medicaid.

[19] Even though they body of Taubman et al. [2014] does not report results using ER total charges, I examine it because it is more continuous than the other two measures of ER uilization. ER total charges (reported in the data as "total charges") is the sum of the list prices of all care provided during the ER visit and any associated hospitalization. The amounts actually paid, which are not observed, are generally much lower than total charges because of discounts. However, because the insured and uninsured receive different discounts, the comparison of total charges is more informative for this exercise than the comparison of actual payments would be.

administrative data.[20] The coefficient in the top panel, which I replicate exactly, indicates that individuals who receive Medicaid coverage increase the probability that they visit the ER by 6.97 percentage points on a base of 34 percentage points among the lottery losers (a 21% increase). The coefficient in the middle panel indicates that individuals who receive Medicaid coverage increase their visits to the ER by 0.388 visits on a base of 1.00 visits among the lottery losers (a 39% increase).[21] The coefficient in the bottom panel indicates that individuals who receive Medicaid coverage increase their total charges by $847 on a base of $3,620 among lottery losers (a 23% increase.).

For comparison to Taubman et al. [2014], I report standard errors clustered by household ID in brackets. For comparison to MTE results, I also report standard errors block bootstrapped by household ID in parentheses. Both standard errors are similar. The estimates for any visits and the number of visits are statistically different from zero at the 1% level, and the total charges estimate is not statistically different from zero at conventional levels.

Because the linear MTE does not incorporate covariates, I examine robustness of the OHIE LATEs to the exclusion of covariates. Following Taubman et al. [2014], the results in the first column include two covariates. The first is a measure of ER utilization before the experimental period, specified in the same way as the outcome Y.[22] When I omit this covariate in Column 2, the point estimates remain almost unchanged for the first two measures of ER utilization. The point estimate for charges decreases substantially, but it remains positive.

The second covariate is a count of the number of lottery entrants in the household. Multiple individuals in the same household could enter the OHIE lottery by signing up for a waitlist for Medicaid coverage. However, if any individual in the household won the lottery, then all household members were treated as winners. About 20% of entrants had another entrant in their household, and very small fraction had two other entrants in their household. Because of the lottery design, individuals in households with more than one entrant won the lottery at a much higher rate: 57% vs. 34%.

It is unlikely that OHIE results that do not control for the number of lottery entrants are internally valid. Because the indicator for winning the lottery Z is not balanced on the number of lottery entrants, it is unlikely that the distribution of U_D is the same for lottery winners and losers. For example, individuals in households whose members communicated about how to sign up for the waitlist might have been more likely to sign up for the waitlist and to use the ER upon winning the lottery. Indeed, the comparison of the characteristics of lottery winners and losers yields several statistically significant differences in the full

[20] For each outcome, I run regressions on the largest set of observations for which all variables are available.

[21] I cannot replicate the result in the bottom panel exactly because of censoring and truncation performed to limit the identification of human subjects in the publicly-available data, but my estimate is very similar to the coefficient of 0.41 on a base of 1.02 visits reported in Taubman et al. [2014].

[22] The period from before the experimental period is January 1, 2007 to March 9, 2009, and the experimental period is from March 10, 2008 to September 2009.

Table 2: OHIE Replication and Extension

Any ER Visits					
	(1)	(2)	(3)	(4)	(5)
Medicaid	0.0697	0.0763	-0.0146	0.182†	0.0531†
	(0.0251)***	(0.0268)***	(0.0282)	(0.0720)**	(0.0276)**
	[0.0239]***	[0.0257]***	[0.0266]	[0.0661]***	[0.0279]*
Covariates	Any pre-visits, Lottery Entrants	Lottery Entrants	No Covariates	No Covariates	No Covariates
Regression sample	Full sample	Full sample	Full sample	2 Lottery Entrants	1 Lottery Entrant
Observations	24,646	24,646	24,646	4,951	19,643
R-squared	0.151	0.030	-	0.015	0.013
E[Y|Z=0]	0.34	0.34	0.34	0.21	0.37
Number of ER Visits					
	(1)	(2)	(3)	(4)	(5)
Medicaid	0.388	0.344	-0.048	0.700†	0.267†
	(0.106)***	(0.124)**	(0.128)	(0.237)***	(0.142)**
	[0.107]***	[0.131]***	[0.134]	[0.237]***	[0.151]*
Covariates	Pre-visits, Lottery Entrants	Lottery Entrants	No Covariates	No Covariates	No Covariates
Regression sample	Full sample	Full sample	Full sample	2 Lottery Entrants	1 Lottery Entrant
Observations	24,615	24,622	24,622	4,948	19,622
R-squared	0.339	0.022	-	0.013	0.010
E[Y|Z=0]	1.00	1.00	1.00	0.45	1.09
ER Total Charges					
	(1)	(2)	(3)	(4)	(5)
Medicaid	$847	$509	-$990	$878	$428
	($796)	($838)	($844)	($1,408)	($958)
	[$769]	[$807]	[$805]	[$1,361]	[$935]
Covariates	Pre-charges, Lottery Entrants	Lottery Entrants	No Covariates	No Covariates	No Covariates
Regression sample	Full sample	Full sample	Full sample	2 Lottery Entrants	1 Lottery Entrant
Observations	24,621	24,630	24,630	4,950	19,628
R-squared	0.088	0.006	-	0.002	0.002
E[Y|Z=0]	$3,620	$3,639	$3,639	$1,639	$3,971

*** p<0.01, ** p<0.05, * p<0.1; Bootstrapped standard errors in parentheses, asymptotic standard errors in square brackets. Standard errors are clustered at the household level.
Test of equality of coefficiens in Columns (4) and (5): ††† p<0.01, †† p<0.05, † p<0.1.
Source: Oregon Administrative Data
To obtain the bootstrapped standard errors, we block bootstrap for 200 replications, and we report the standard deviation of the replications as an estimate of the standard error.

sample. As noted in Taubman et al. [2014], the LATEs for ER utilization are not robust to the removal of the control for the number of lottery entrants. In Column 3, the coefficients for all three specifications of ER utilization are negative, and none are statistically different from zero.

The next two columns report results from separate regressions for individuals with two lottery entrants and one lottery entrant. The results within these subsamples should be internally valid.[23] However, the comparison across these subsamples suggests that the results are not globally externally valid. Across all measures of ER utilization, the LATE of insurance on ER utilization is substantially larger in the subsample with two entrants: 18 vs. 5 percentage points, 0.7 vs. 0.3 visits, and \$830 vs. \$430.[24] The coefficients are statistically different from each other at the 10% level for first two measures.

The subsample with one lottery entrant is my preferred replication sample. Because it includes the vast majority of the full OHIE sample, it is likely to be more representative of other samples of interest. One difficulty in extrapolating from either OHIE subsample to any other sample of interest is that a variable that captures the same information as the number of lottery entrants is unlikely to be available. Household size is a potential candidate, but it is distinct from the number of lottery entrants because not all members of a household entered the lottery. Household size is not available in the OHIE administrative data, so it is not possible to further restrict the sample with one lottery entrant to households with only one member.

4.2 Estimated Average Characteristics of Always Takers, Never Takers, and Compliers

The first column of Table 3 provides summary statistics on the OHIE sample with one lottery entrant. The sample is 56% female, the average age is 41, and 91% selected materials written in English.[25] The next two columns show that the lottery winners and losers have the same average values of these covariates, and the corresponding t-tests reported in the bottom panel do not reject the null of internal validity. In contrast, as shown in Columns 4 and 5, the treated and untreated individuals do not have the same average values of these covariates. The individuals who take up treatment are 64% female, while the individuals who do not are only 53% female. Thus, there seems to be some observable basis for selection into insurance.

Columns 6 through 9 report cross-tabulations of the data based on lottery status and treatment. Always takers have a 72% probability of being female, but never takers have only a 53% probability of being female.

[23] The Taubman et al. [2014] approach of controlling for the number of lottery entrants could also produce internally valid results, depending on whether a linear control effectively balances the unobserved net cost of treatment U_D. More recent analysis by Finkelstein et al. [2015] divides the sample as I do.

[24] The average ER utilization of lottery losers is much higher among the subsample with multiple lottery entrants, suggesting that individuals in households with more ER utilization did not enter the lottery at a higher rate. However, individuals in household with more pent-up demand for ER utilization might have entered the lottery at a higher rate. Although ER care is available to the uninsured, it is not necessarily free.

[25] I report statistics on these covariates because they are all defined in a comparable way in the Behavioral Risk Factor Surveillance System (BRFSS), so they can be used for extrapolation to other samples.

Table 3: Average Characteristics and Outcomes of Always Takers, Never Takers, and Compliers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Randomized Intervention Sample Average	Intervention	Baseline	Randomized Intervention Sample Treated	Randomized Intervention Sample Untreated	Baseline Treated (Always Takers)	Baseline Untreated (Never Takers and Untreated Compliers)	Intervention Treated (Always Takers and Treated Compliers)	Intervention Untreated (Never Takers)	Local Average Treated (Treated Compliers)	Local Average Untreated (Untreated Compliers)	Local Average (All Compliers)
	RIS	(Z=1)	(Z=0)	(D=1) RIST	(D=0) RISU	(D=1, Z=0) BT	(D=0, Z=0) BU	(D=1, Z=1) IT	(D=0, Z=1) IU	LAT	LAU	LA
Covariates												
Female	0.56	0.55	0.56	0.64	0.53	0.72	0.53	0.58	0.53	0.50	0.55	0.53
Age in 2009	40.7	40.7	40.7	40.5	40.7	39.4	40.9	41.3	40.3	42.4	42.4	42.4
English	0.91	0.91	0.91	0.91	0.91	0.90	0.91	0.92	0.91	0.93	0.92	0.92
Predicted outcomes												
Any ER visits	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.38	0.38	0.38
Number of ER visits	1.12	1.12	1.12	1.13	1.11	1.12	1.12	1.13	1.11	1.14	1.14	1.14
ER total charges	$4,009	$4,003	$4,013	$4,016	$4,007	$3,899	$4,033	$4,098	$3,937	$4,215	$4,251	$4,238
Outcomes												
Any ER visits	0.37	0.38	0.37	0.51	0.33	0.55	0.33	0.48	0.31	0.44	0.39	0.40
Number of ER visits	1.12	1.16	1.09	1.73	0.92	1.89	0.95	1.62	0.85	1.45	1.19	1.28
ER total charges	$4,009	$4,082	$3,971	$6,996	$3,061	$8,794	$3,109	$5,732	$2,930	$3,944	$3,516	$3,664
N for outcomes												
Any ER visits	19,643	6,755	12,888	4,737	14,906	1,959	10,929	2,778	3,977	1,751	3,341	5,092
Number of ER visits	19,622	6,743	12,879	4,725	14,897	1,956	10,923	2,769	3,974	1,745	3,333	5,078
ER total charges	19,628	6,752	12,876	4,726	14,902	1,951	10,925	2,775	3,977	1,752	3,341	5,093

	T-tests		***Difference-in-difference Tests***							
	(2) - (3)	*(10) - (11)*	λ_{DZ} *[(8) - (6)] - [(9) - (7)]*	λ_D *(6) - (7)*	λ_Z *(9) - (7)*	$\lambda_{DZ} = 0$ $\lambda_D = 0$	$\lambda_{DZ} = 0$ $\lambda_Z = 0$	$\lambda_D = 0$ $\lambda_Z = 0$	$\lambda_{DZ} = 0$ $\lambda_Z = 0$ $\lambda_D = 0$	
Covariates										
Female	-0.012	-0.045	-0.133***	0.188***	-0.006	***	***	***	***	
Age in 2009	-0.019	-0.075	2.504***	-1.470***	-0.668***	***	***	***	***	
English	0.002	0.007	0.018*	-0.008	-0.003					
Predicted outcomes										
Any ER visits	-0.001	-0.003	0.004	0.002	-0.003	***		**	***	
Number of ER visits	-0.005	-0.019	0.005	0.020	-0.012	***		***	***	
ER total charges	-$6	-$21	$266**	-$98	-$89	***	***	**	***	
Outcomes										
Any ER visits	0.014*	0.053**	-0.045***	0.213***	-0.023***	***	***	***	***	
Number of ER visits	0.069*	0.267*	-0.171	0.939***	-0.104**	***	***	***	***	
ER total charges	$111	$428	-$2,882***	$5,685***	-$179	***	***	***	***	

*** p<0.01, ** p<0.05, * p<0.1; With the exception of joint tests, where asymptotic F-Tests were used, p-values were computed based on bootstrapped standard errors.
Source: Oregon Administrative Data, 1 Lottery Entrant in Household

Predicted outcomes were obtained using a linear model among lottery losers controlling for common controls only, which includes binary variables for female, age in 2009, English, and all two-way interaction terms between these covariates. The reported statistics reflect predictions obtained for both lottery losers and lottery winners using the coefficients from this linear probability model.

Columns 10 and 11 report the average characteristics of treated and untreated compliers, calculated via (2) and (1).[26] Although the treated and untreated compliers appear to have slightly different characteristics, the t-test results reported in the second panel show that the characteristics are not statistically different, providing further evidence of internal validity.

The combined characteristics of the treated and untreated compliers are reported in Column 12. Existing studies that report average characteristics of compliers often compare the average characteristics of compliers to the average characteristics of the full sample to informally assess external validity, and they do not necessarily even report the characteristics of other groups. In the OHIE, compliers are more male (47% vs. 44%), slightly older (42.4 vs. 40.7), and more likely to request materials in English (92% vs. 91%) than

[26]Previous research on the OHIE has reported average characteristics of compliers (Finkelstein et al. [2015]).

individuals in the full OHIE sample. Though the differences raise some some concerns about global external validity, the magnitues of the differences are small relative to the levels for the full sample.

However, compliers are included in the full sample (they make up 25% of the full sample), so it is more informative to compare the compliers to the always takers and never takers than it is to compare them to the full sample. The compliers and the never takers are 53% female, but the always takers are 72% female. The always takers are also younger on average than the compliers (39.4 vs. 42.4), and the never takers are even younger (40.3). The always takers are less likely to request materials in English (90%) than the compliers (92%) and the never takers (91%). The differences in age and English status across the three groups are relatively small, but the difference in age between the always takers and the compliers is large. Because of the large age difference, the comparison of the compliers to the always takers casts more doubt on global external validity than the comparison of the compliers to the full sample.

4.3 Difference-in-Difference Test Results

The difference-in-difference test using covariates formalizes the comparison of the compliers to the rest of the sample. Results in the second panel show that some covariates are related to selection ($\lambda_Z \neq 0$); some covariates are related to baseline takeup ($\lambda_D \neq 0$); and some covariates have different relationships to baseline takeup than they have to intervention takeup ($\lambda_{DZ} \neq 0$). When we use these covariates to predict the outcomes Y among the lottery losers, we still see some evidence that casts doubt on global external validity.

Results from the difference-in-difference test using the three measures of ER utilization show a statistically significant rejection of global external validity ($\lambda_{DZ} \neq 0$) for two measures of ER utilization. The results also show statistically significant evidence of selection ($\lambda_Z \neq 0$) for two measures of ER utilization. The rejection of the null of no selection indicates that RISOLS is biased. The rejection of global external validity indicates that the LATE does not apply to all individuals.

4.4 Estimates of the Linear MTE

Figure 1 plots the estimated linear $MTO(p)$, $MUO(p)$, and $MTE(p)$ for all three measures of ER utilization. The potential fraction treated p increases along the horizontal axis from no treatment to full treatment. As p increases, individuals with successively higher net unobserved costs of treatment U_D select into treatment. The baseline treatment probability $p_B = 0.152$ indicates that always takers make up the first 15.2% of the sample to select into treatment. The always takers select into treatment even if they lose the lottery, so they have the lowest net unobserved costs of treatment ($0 \leq U_D \leq p_B$). The intervention treatment probability $p_I = 0.411$ indicates that the next 25.9% (=0.411-0.152) of individuals to select into treatment

are compliers. The compliers have net unobserved costs of treatment ($p_B < U_D \leq p_I$) that are higher than the net unobserved cost of the last always taker to select into treatment ($U_D = p_B$). The compliers that win the lottery select into treatment (they are the treated compliers), and the compliers that lose the lottery that do not select into treatment (they are the untreated compliers). The remaining 40.1% of the sample with the highest net unobserved costs of treatment ($p_I < U_D \leq 1$) are never takers who do not select into treatment even if they win the lottery. The labels on the bottom axis show how p_B and p_I divide the baseline and intervention treated and untreated.

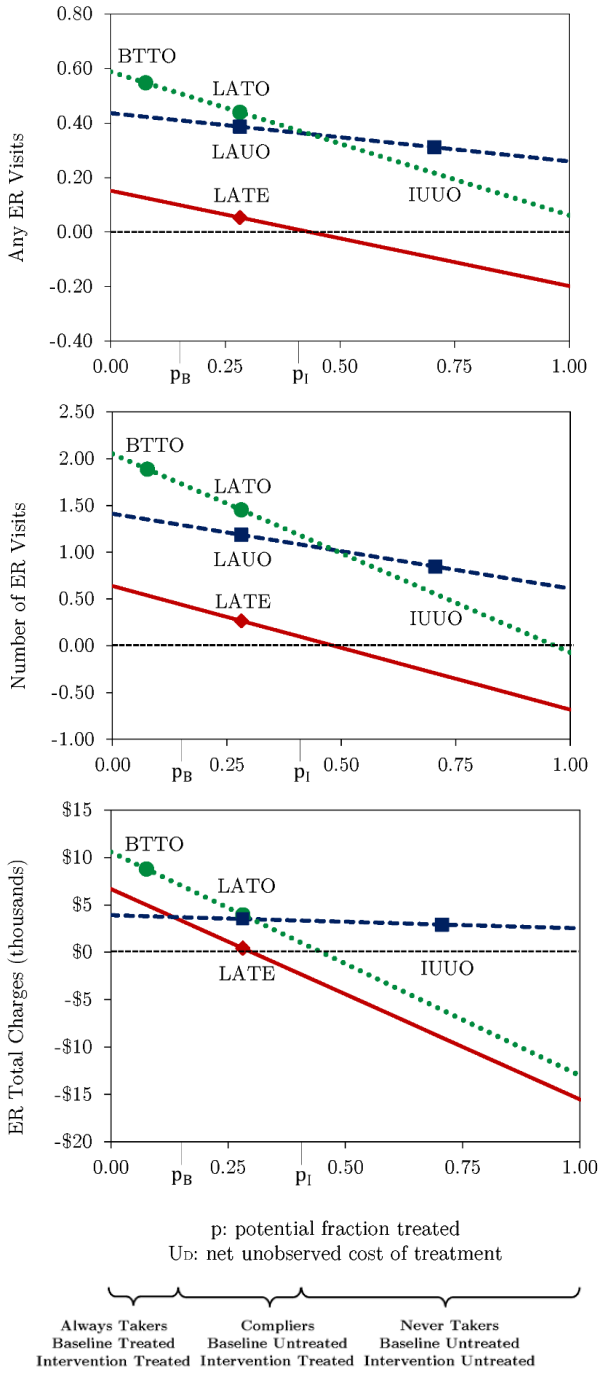
The depiction of p_B and p_I along the horizontal axis of Figure 1 provides more information than the first stage estimate. By definition, $(p_I - p_B)$ is equal to the first stage estimate ($P(D = 1|Z = 1) - P(D = 1|Z = 0)$). Therefore, the first stage estimate gives the share of compliers, but it provides less information than the horizontal axis of Figure 1 because it does not convey the shares of always takers and never takers separately. The reporting of p_B or p_I in addition to the first stage informs whether the experimental intervention changes selection into treatment for individuals with high or low net unobserved costs of treatment relative to the entire sample.

The dashed line depicts the linear marginal untreated outcome function $MUO(p)$, which is identified by the two points depicted with square markers. One point, $(\frac{p_B+p_I}{2}, LAUO)$, shows that the median complier has the average untreated outcome for compliers, which is implied by the linearity of $MUO(p)$ and the uniformity of U_D. The other point, $(\frac{p_I+1}{2}, IUUO)$, shows that the median never taker has the average untreated outcome for all never takers. For all three measures of ER utilization, the line that connects both points, slopes downward, indicating adverse selection. As reported in the first columns of Table 4, the downward slope of $MUO(p)$ is statistically different from zero for any visits and the number of visits.

The dotted line depicts the linear marginal treated outcome function $MTO(p)$, which is identified by the two points, $(\frac{p_B}{2}, BTTO)$ and $(\frac{p_B+p_I}{2}, LATO)$, depicted with circular markers. For all three measures of ER utilization, $BTTO > LATO$, so the linear $MTO(p)$ slopes downward. The downward slope indicates either selection or moral hazard that changes with selection, or both.

The solid line depicts the marginal treatment effect $MTE(p)$. The LATE, which gives the average treatment effect for compliers, is the single point on $MTE(p)$ with the diamond marker. Because $MTE(p)$ is not equal to the LATE for all p, it is clear from the figure that the LATE is not globally externally valid. In this application, because the marginal treated and untreated outcome functions have slopes of the same sign, we cannot reject global external validity using the bounds introduced in Section 3.2. However, we have rejected global external validity in Section 4.3 using the difference-in-difference test. In this application, the linear MTE relies on the same assumptions as the difference-in-difference test, but it provides more information.

Figure 1: Linear MTE



For all measures of ER utilization, the MTE is downward-sloping, indicating that moral hazard is largest for the first individuals to select into treatment, and it decreases as subsequent individuals select into treatment. Across all measures of ER utilization, the marginal treatment effects for always takers are positive and larger than the marginal treatment effects for compliers. This pattern could arise if the individuals with the most pent-up demand for ER utilization select into coverage regardless of the lottery outcome, and individuals with lower levels of pent-up demand only select into coverage if they win the lottery.

For all measures of ER utilization, the MTE is positive for some individuals and negative for others. The marginal treatment effect changes from positive to negative when the fraction treated increases to $p^* = 0.43$ for any visits, $p^* = 0.48$ for the number of visits, and $p^* = 0.30$ for total charges. For total charges, $p* < p_I$, which indicates that even though OHIE compliers have positive treatment effects on average, some compliers decrease their total charges when they select into insurance. All never takers have negative treatment effects for total charges, and most never takers have negative treatment effects for the other two measures. There are some never takers with positive treatment effects for any visits and number of visits because $p^* > p_I$.

In health economics, there is a long-standing question about whether there is heterogeneity in moral hazard across individuals who use different amounts of care. If moral hazard is the same in levels across all individuals, as would be the case if the LATE from Oregon were globally externally valid, then efforts to reduce moral hazard among high users would be just as effective as efforts to reduce moral hazard among low users. However, if moral hazard is greatest among the high users, then efforts that focus on curtailing their moral hazard will have the greatest impact. The slope of the estimated marginal selection effect shows that the individuals most likely to sign up for coverage are the individuals that would have the most utilization if they were uninsured, and the slope of the estimated marginal treatment effect shows that the individuals most likely to sign up for coverage increase their utilization the most upon gaining coverage. Therefore, moral hazard is greatest among the individuals who consume the most care.

4.5 Estimated Treated Outcomes, Untreated Outcomes, and Treatment Effects from the Linear MTE

Table 4 reports average treated outcomes, untreated outcomes, and treatment effects recovered from the linear MTO, MUO, and MTE functions as discussed in Section 3.4. Column 1 reports estimates for always takers, the baseline treated. The baseline treated treated outcome (BTTO) is observed, so it is reported in bold, along with all other quantities that do not require the assumptions required to identify linear MTE. On average, always takers visit the ER with probability 0.55, they make 1.89 visits, and the incur $8,794 in total charges. The baseline treated untreated outcome (BTUO) is not observed because all always takers receive coverage, but it can be estimated by weighting the linear marginal untreated outcome function $MUO(p)$.

Table 4: Treated Outcome, Untreated Outcome, and Treatment Effects in Oregon

	Function			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Baseline Treated	Baseline Untreated	Intervention Treated	Intervention Untreated	Randomized Intervention Sample Treated	Randomized Intervention Sample Untreated	Local Average	Average
	Intercept	Slope		(Always Takers)			(Never Takers)			(Compliers)	
				BT	BU	IT	IU	RIST	RISU	LA	A
Any ER Visits	*MTO(p)*		Treated	**0.55*** **	0.28***	**0.48*** **	0.22***	**0.51*** **	0.27***	**0.44*** **	0.32***
	0.59***	-0.53***	Outcome	**(0.53, 0.57)**	(0.19, 0.38)	**(0.46, 0.50)**	(0.09, 0.34)	**(0.49, 0.52)**	(0.16, 0.37)	**(0.41, 0.47)**	(0.24, 0.40)
	(0.56, 0.62)	(-0.75, -0.31)	TO	**BTTO**	BUTO	**ITTO**	IUTO	**RISTTO**	RISUTO	**LATO**	ATO
	MUO(p)		Untreated	0.42***	**0.33*** **	0.40***	**0.31*** **	0.41***	**0.33*** **	**0.39*** **	0.35***
	0.44***	-0.18***	Outcome	(0.35, 0.50)	**(0.33, 0.34)**	(0.34, 0.46)	**(0.30, 0.33)**	(0.34, 0.48)	**(0.32, 0.33)**	**(0.34, 0.44)**	(0.33, 0.37)
	(0.35, 0.53)	(-0.33, -0.03)	UO	BTUO	**BUUO**	ITUO	**IUUO**	RISTUO	**RISUUO**	**LAUO**	AUO
	MTE(p)		Treatment	0.12***	-0.05	0.08***	-0.10*	0.10***	-0.06*	**0.05** **	-0.02
	0.15***	-0.35***	Effect	(0.04, 0.20)	(-0.15, 0.04)	(0.02, 0.14)	(-0.22, 0.03)	(0.03, 0.16)	(-0.16, 0.04)	**(0.00, 0.11)**	(-0.10, 0.05)
	(0.06, 0.24)	(-0.61, -0.12)	TE	BTTE	BUTE	ITTE	IUTE	RISTTE	RISUTE	**LATE**	ATE
Number of ER Visits	*MTO(p)*		Treated	**1.89*** **	0.83**	**1.62*** **	0.55	**1.73*** **	0.76*	**1.45*** **	0.99***
	2.05***	-2.12***	Outcome	**(1.74, 2.05)**	(0.02, 1.48)	**(1.50, 1.73)**	(-0.49, 1.41)	**(1.63, 1.82)**	(-0.11, 1.46)	**(1.22, 1.66)**	(0.35, 1.52)
	(1.80, 2.38)	(-4.02, -0.60)	TO	**BTTO**	BUTO	**ITTO**	IUTO	**RISTTO**	RISUTO	**LATO**	ATO
	MUO(p)		Untreated	1.35***	**0.95*** **	1.25***	**0.85*** **	1.29***	**0.92*** **	**1.19*** **	**1.01*** **
	1.41***	-0.80***	Outcome	(1.06, 1.72)	**(0.91, 1.00)**	(1.03, 1.53)	**(0.79, 0.91)**	(1.04, 1.61)	**(0.89, 0.96)**	**(1.01, 1.41)**	**(0.94, 1.11)**
	(1.08, 1.82)	(-1.43, -0.26)	UO	BTUO	**BUUO**	ITUO	**IUUO**	RISTUO	**RISUUO**	**LAUO**	**AUO**
	MTE(p)		Treatment	0.54***	-0.12	0.37**	-0.29	0.44***	-0.17	**0.27**	-0.02
	0.64***	-1.32*	Effect	(0.15, 0.91)	(-0.93, 0.54)	(0.05, 0.65)	(-1.28, 0.58)	(0.12, 0.73)	(-1.02, 0.54)	**(-0.08, 0.58)**	(-0.67, 0.53)
	(0.16, 1.09)	(-3.23, 0.19)	TE	BTTE	BUTE	ITTE	IUTE	RISTTE	RISUTE	**LATE**	ATE
ER Total Charges	*MTO(p)*		Treated	**$8,794*** **	-$3,006*	**$5,732*** **	-$6,068**	**$6,996*** **	-$3,824*	**$3,944*** **	-$1,218
	$10,582***	-$23,601***	Outcome	**($7,626, $9,902)**	(-$7,423, $1,380)	**($4,987, $6,547)**	(-$11,844, -$420)	**($6,356, $7,591)**	(-$8,617, $903)	**($2,557, $5,436)**	(-$4,858, $2,409)
	($8,828, $12,393)	(-$34,299, -$12,906)	TO	**BTTO**	BUTO	**ITTO**	IUTO	**RISTTO**	RISUTO	**LATO**	ATO
	MUO(p)		Untreated	$3,801***	**$3,109*** **	$3,621***	**$2,930*** **	$3,695***	**$3,061*** **	**$3,516*** **	$3,214***
	$3,905***	-$1,383	Outcome	($2,034, $5,809)	**($2,906, $3,345)**	($2,284, $5,145)	**($2,545, $3,341)**	($2,180, $5,423)	**($2,899, $3,276)**	**($2,445, $4,744)**	($2,831, $3,697)
	($1,892, $6,208)	(-$5,200, $1,878)	UO	BTUO	**BUUO**	ITUO	**IUUO**	RISTUO	**RISUUO**	**LAUO**	AUO
	MTE(p)		Treatment	$4,994***	-$6,115***	$2,111***	-$8,998***	$3,301***	-$6,885***	**$428**	-$4,432**
	$6,677***	-$22,218***	Effect	($2,587, $6,998)	(-$10,552, -$1,638)	($387, $3,584)	(-$14,857, -$3,206)	($1,440, $4,873)	(-$11,686, -$2,053)	**(-$1,436, $2,142)**	(-$8,056, -$723)
	($3,555, $9,326)	(-$33,486, -$11,076)	TE	BTTE	BUTE	ITTE	IUTE	RISTTE	RISUTE	**LATE**	ATE

*** p<0.01, ** p<0.05, * p<0.1. Bootstrapped 95% confidence intervals in parentheses.

Source: Oregon Administrative Data, 1 Lottery Entrant in Household.

Calculation of the bold quantities does not rely on the linear MTE.

To obtain the bootstrapped 95% confidence intervals, we block bootstrap by household ID for 200 replications, and we report the 2.5 and 97.5 percentiles as the 95% confidence interval.

The BTUO results show that on average, if the always takers were uninsured, they would visit the ER with probability 0.42, they would make 1.35 visits, and their ER total charges would be $3,801. The baseline treated treatment effect BTTE results show that upon gaining insurance, on average, always takers increase their probability of an ER visit by 0.12, they increase their number of visits by 0.54, and they increase their ER total charges by $4,944. All of these estimates are much larger than the corresponding LATE estimates, indicating that there is meaningful variation in moral hazard between the always takers and the compliers.

The bounds on the BTTE introduced in Section 3.2, which assume weak monotonicity of $MTO(p)$ and $MUO(p)$, are much larger than the BTTE estimates obtained assuming linearity of $MTO(p)$ and $MUO(p)$. The bounds on the BTTE imply that upon gaining insurance, always takers increase their average visit probability by no more than 0.16 ($BTTE \leq BTTO - LAUO = 0.55 - 0.39$), their number of visits by no more than 0.7 (1.89-1.19), and their total charges by no more than $5,638 ($8,794-$3,156). These bounds are not informative about the global external validity of the LATE from the OHIE because they do not rule out the LATE. It is very plausible that other applications will yield linear $MTO(p)$ and $MUO(p)$ estimates with slopes of opposite sign and thus bounds that will be informative about the global external validity of other LATEs.

Even though the bounds on the BTTE are not informative about global external validity, the bounds on the BTUO could be informative about what ER utilization would have been had Oregon decided not to provide coverage to anyone who entered the lottery. The bounds on the BTUO imply that if always takers lost their coverage, then their average ER utilization would be no less than the average ER utilization for untreated compliers (LATO): 0.39 visit probability, 1.19 visits, and $3,516 in charges. The BUUOs in Column 2 show that all of the non-always takers without coverage average a 33% visit probability, 0.95 visits, and $3,109 total charges. Therefore, the average visit probability of everyone in the lottery sample would be no less than 0.25 ($= 0.39p_B + 0.33(1 - p_B)$), the average number of visits would be no less than 0.74, and average ER charges would be no less than $2,365.

Column 4 gives results for the never takers, the intervention untreated. Never takers visited the ER an average of 0.85 times, much less frequently than the always takers, who visited 1.89 times, and the compliers, who visited 1.19 times. However, the IUTE estimates imply that they would visit the ER even less, an average of 0.29 fewer times, if they had health insurance. The estimates also imply that they would have a 10 percentage point lower probability of visiting the ER, which is about half of their observed probability of visiting the ER. The bounds developed in Section 3.2 imply that upon gaining insurance, never takers would increase their probability of visiting the ER by no more than 13 percentage points ($BTTE \leq LATO - IUUO = 0.44 - 0.31$), and they would increase their number of visits by no more than 0.6 (=1.45-0.85). Therefore, if Oregon decided to require everyone in the lottery sample to have

insurance, then the average visit probability of everyone in the lottery sample would be no more than 0.27 ($= ITTOp_I + 0.13(1 - p_I)$), and the average number of visits would be no more than 1.02.

Column 2 gives treated outcomes, untreated outcomes, and treatment effects for the baseline untreated (BU) individuals, which includes all individuals except the always takers. This group is policy-relevant because it represents the potential pool of individuals to whom coverage could be expanded. The average untreated outcome for these individuals, the BUUO, is observed, but the average treated outcome is not. Weighting the marginal treated outcome function $MTO(p)$ gives an estimate of the insured ER utilization of these uninsured individuals. The baseline untreated individuals would visit the ER with 28% probability when insured, but we only observe them visiting the ER with 33% probability when uninsured, so the BUTE implies that insurance decreases the probability of an ER visit by 5 percentage points for all individuals who were uninsured at baseline. The number of visit results tell a similar story. The results show that insurance decreases the number of visits by 0.12, from a BUUO of 0.95 to at BUTO of 0.83. The BUTE in terms of ER total charges is also negative: baseline untreated individuals decrease their ER utilization by \$6,115 upon gaining coverage.

On its own, the BUTE for total charges seems plausible. However, because we have plotted $MUO(p)$ and calculated the average untreated outcome derived from it, we see that the linear extrapolation for uninsured charges does not appear plausible at high values of p because it predicts negative ER utilization. Total charges could be the preferred measure of ER utilization on theoretical grounds because it is the most continuous. However, on empirical grounds, the extrapolation of the MTE for ER total charges seems the least plausible at high values of p. Although the linear extrapolation of the MTO and the MTE for total charges appears to be reasonable within the experimental support from p_B to p_I, I prefer to extrapolate to all untreated groups using the any visit and number of visits results. I focus on those results for the other untreated groups.

Column 7 reports the local average treatment effect LATE for comparison to the other treatment effects. Although the local average treated outcome LATO, the local average untreated outcome LAUO, and the LATE can be calculated without the MTE, the exact same values can be recovered from the MTE. Furthermore, the reported bootstrapped confidence intervals on the LATE are exactly the same as those that result from obtaining the LATE via an instrumental variable regression. Even though the LATE in terms of ER utilization is not statistically different from zero at conventional levels, all of the other treatment effects for ER total charges are statistically different from zero, indicating that even if there is no detectable impact on the compliers in an experiment, there could be detectable impacts on other groups of interest. The confidence intervals for the LATE often do not even include other treatment effects of interest, so reporting the LATE alone would have offered a very limited picture.

Column 8 reports the ATO that would result if the entire sample had insurance, the AUO that would result if the entire sample did not have insurance, and the implied average treatment effect of gaining insurance. The total charges results should be interpreted with caution because they involve extrapolation to individuals with high values of U_D. The any visits and number of visit results show that the average treatment effect is nearly zero because there are roughly as many individuals with positive treatment effects as negative treatment effects.

4.6 Treated Outcome Decomposition Results

In Table 5, I decompose the treated outcomes from Table 4 into selection and treatment components following the approach that I introduced in Section 3.6. For the always takers, selection accounts for 77% of the probability of observing any visit, 71% of the observed number of visits, and 43% of ER total charges. The 99% confidence intervals reject one and zero, indicating that the treated outcome for always takers reflects a combination of selection and treatment effects. In other words, 71% of the visits that the always takers make to the emergency room would still take place were they to lose coverage. However, always takers also increase their utilization when they gain coverage, and that moral hazard is responsible for 29% of the visits that they make to the ER.

As shown in Column 7, the average utilization of compliers shows a greater role for selection. For compliers who gain insurance, selection explains 88% of the probability of any visit, 82% of the number of visits, and 89% of total charges. The decomposition rejects full selection at the 90% level or higher for the first two measures of ER utilization, as shown by the significance crosses. However, when ER utilization is measured in terms of total charges, some compliers, (those with $p^* = 0.30 \leq U_D \leq 0.41 = p_I$) have negative treatment effects. The combination of negative treatment effects and positive selection effects results in a decomposition that cannot reject full selection.

The decompositions of the treated outcomes for all of the untreated groups also reflect negative treatment effects. For always takers, the BUTO reported in Table 4 shows that they would have a 28% probability of visiting the ER if they had insurance. The decomposition in Table 5 shows that predicted probability of visiting the ER with insurance would be 1.17 times higher, but for the negative treatment effect.

I can also decompose the *difference* in treated outcomes induced by the OHIE as discussed in Section 3.6. The results of this decomposition should be of interest to insurers because they explain why average ER utilization is lower for insured lottery winners than insured lottery losers. Relative to the insured lottery losers, the insured lottery winners are 7 percentage points less likely to visit the ER, they visit the ER 0.26 fewer times, and their total charges are $3,062 lower. The slope of the marginal untreated outcome function relative to the marginal treated outcome function indicates that selection explains 33% (-.18/-.53) of the

Table 5: Decompositions of Treated Outcomes and OLS Estimates into Selection and Treatment Effects

			(1) Baseline Treated (Always Takers) BT	(2) Baseline Untreated BU	(3) Intervention Treated IT	(4) Intervention Untreated (Never Takers) IU	(5) Randomized Intervention Sample Treated RIST	(6) Randomized Intervention Sample Untreated RISU	(7) Local Average (Compliers) LA	(8) Average A
Any ER Visits	Decomposition	Selection UO/TO	0.77*** (0.64, 0.92)††† BTUO/**BTTO**	1.17*** (0.89, 1.78) **BUUO**/BUTO	0.83*** (0.71, 0.96)††† ITUO/**ITTO**	1.44*** (0.92, 3.33) **IUUO**/IUTO	0.81*** (0.69, 0.94)††† RISTUO/**RISTTO**	1.23*** (0.89, 2.01) **RISUUO**/RISUTO	0.88*** (0.76, 1.00)†† **LAUO**/**LATO**	1.07*** (0.87, 1.42) AUO/ATO
		Treatment Effect TE/TO	0.23*** (0.08, 0.36)††† BTTE/**BTTO**	-0.17 (-0.78, 0.11)††† BUTE/BUTO	0.17*** (0.04, 0.29)††† ITTE/**ITTO**	-0.44* (-2.33, 0.08)††† IUTE/IUTO	0.19*** (0.06, 0.31)††† RISTTE/**RISTTO**	-0.23* (-1.01, 0.11)††† RISUTE/RISUTO	0.12** (0.00, 0.24)††† **LATE**/**LATO**	-0.07 (-0.42, 0.13)††† ATE/ATO
	OLS Estimates	OLS = TTO - UUO	**0.21*** (0.19, 0.24) BOLS = BTTO - BUUO**		**0.17*** (0.14, 0.19) IOLS = ITTO - IUUO**		**0.18*** (0.16, 0.20) RISOLS = RISTTO - RISUUO**		-	-
	OLS Decomposition	Selection (OLS - TE)/OLS	0.41*** (0.08, 0.80)††† (**BOLS** - BTTE)/**BOLS**	1.23*** (0.77, 1.65) (**BOLS** - BUTE)/**BOLS**	0.53*** (0.11, 0.88)††† (**IOLS** - ITTE)/**IOLS**	1.57*** (0.86, 2.42) (**IOLS** - IUTE)/**IOLS**	0.45*** (0.09, 0.84)††† (**RISOLS** - RISTTE)/**RISOLS**	1.344*** (0.780, 1.943) (**RISOLS** - RISUTE)/**RISOLS**	-	-
		Treatment Effect TE/OLS	0.59*** (0.20, 0.92)††† BTTE/**BOLS**	-0.23 (-0.65, 0.23)††† BUTE/**BOLS**	0.47*** (0.12, 0.89)††† ITTE/**IOLS**	-0.57* (-1.42, 0.14)††† IUTE/**IOLS**	0.55*** (0.16, 0.91)††† RISTTE/**RISOLS**	-0.344* (-0.943, 0.220)††† RISUTE/**RISOLS**		
Number of ER Visits	Decomposition	Selection UO/TO	0.71*** (0.54, 0.92)††† BTUO/**BTTO**	1.15** (0.62, 5.66) **BUUO**/BUTO	0.77*** (0.61, 0.97)†† ITUO/**ITTO**	1.53 (-22.34, 14.80) **IUUO**/IUTO	0.75*** (0.58, 0.93)††† RISTUO/**RISTTO**	1.22* (-2.60, 6.31) **RISUUO**/RISUTO	0.82*** (0.64, 1.06)† **LAUO**/**LATO**	1.02*** (0.66, 2.95) AUO/ATO
		Treatment Effect TE/TO	0.29*** (0.08, 0.46)††† BTTE/**BTTO**	-0.15 (-4.66, 0.38)†† BUTE/BUTO	0.23** (0.03, 0.39)††† ITTE/**ITTO**	-0.53 (-13.80, 23.34) IUTE/IUTO	0.25*** (0.07, 0.42)††† RISTTE/**RISTTO**	-0.22 (-5.31, 3.60)† RISUTE/RISUTO	0.18 (-0.06, 0.36)††† **LATE**/**LATO**	-0.02 (-1.95, 0.34)††† ATE/ATO
	OLS Estimates	OLS = TTO - UUO	**0.94*** (0.78, 1.10) BOLS = BTTO - BUUO**		**0.77*** (0.65, 0.88) IOLS = ITTO - IUUO**		**0.81*** (0.71, 0.90) RISOLS = RISTTO - RISUUO**		-	-
	OLS Decomposition	Selection (OLS - TE)/OLS	0.43*** (0.14, 0.82)††† (**BOLS** - BTTE)/**BOLS**	1.13*** (0.35, 1.83) (**BOLS** - BUTE)/**BOLS**	0.52*** (0.18, 0.93)†† (**IOLS** - ITTE)/**IOLS**	1.38*** (0.34, 2.83) (**IOLS** - IUTE)/**IOLS**	0.46*** (0.15, 0.85)††† (**RISOLS** - RISTTE)/**RISOLS**	1.209*** (0.329, 2.150) (**RISOLS** - RISUTE)/**RISOLS**	-	-
		Treatment Effect TE/OLS	0.57*** (0.18, 0.86)††† BTTE/**BOLS**	-0.13 (-0.83, 0.65)††† BUTE/**BOLS**	0.48** (0.07, 0.82)††† ITTE/**IOLS**	-0.38 (-1.83, 0.66)††† IUTE/**IOLS**	0.54*** (0.15, 0.85)††† RISTTE/**RISOLS**	-0.209 (-1.150, 0.671)††† RISUTE/**RISOLS**		
ER Total Charges	Decomposition	Selection UO/TO	0.43*** (0.24, 0.70)††† BTUO/**BTTO**	-1.03* (-14.04, 5.94)† **BUUO**/BUTO	0.63*** (0.39, 0.93)††† ITUO/**ITTO**	-0.48** (-3.27, -0.20)†† **IUUO**/IUTO	0.53*** (0.31, 0.80)††† RISTUO/**RISTTO**	-0.80* (-5.61, 5.08)† **RISUUO**/RISUTO	0.89*** (0.55, 1.50) **LAUO**/**LATO**	-2.64 (-27.00, 20.16) AUO/ATO
		Treatment Effect TE/TO	0.57*** (0.30, 0.76)††† BTTE/**BTTO**	2.03 (-4.94, 15.04) BUTE/BUTO	0.37*** (0.07, 0.61)††† ITTE/**ITTO**	1.48** (1.20, 4.27)†† IUTE/IUTO	0.47*** (0.20, 0.69)††† RISTTE/**RISTTO**	1.80 (-4.08, 6.61) RISUTE/RISUTO	0.11 (-0.50, 0.45)††† **LATE**/**LATO**	3.64 (-19.16, 28.00) ATE/ATO
	OLS Estimates	OLS = TTO - UUO	**$5,685*** ($4,475, $6,868) BOLS = BTTO - BUUO**		**$2,803*** ($2,021, $3,602) IOLS = ITTO - IUUO**		**$3,935*** ($3,182, $4,623) RISOLS = RISTTO - RISUUO**		-	-
	OLS Decomposition	Selection (OLS - TE)/OLS	0.12 (-0.16, 0.49)††† (**BOLS** - BTTE)/**BOLS**	2.08*** (1.31, 2.66) (**BOLS** - BUTE)/**BOLS**	0.25 (-0.38, 0.87)††† (**IOLS** - ITTE)/**IOLS**	4.21*** (1.89, 7.68)††† (**IOLS** - IUTE)/**IOLS**	0.16 (-0.21, 0.63)††† (**RISOLS** - RISTTE)/**RISOLS**	2.75*** (1.48, 3.99)††† (**RISOLS** - RISUTE)/**RISOLS**	-	-
		Treatment Effect TE/OLS	0.88*** (0.51, 1.16) BTTE/**BOLS**	-1.08 (-1.66, -0.31)††† BUTE/**BOLS**	0.75*** (0.13, 1.38) ITTE/**IOLS**	-3.21*** (-6.68, -0.89)††† IUTE/**IOLS**	0.84*** (0.37, 1.21) RISTTE/**RISOLS**	-1.75*** (-2.99, -0.48)††† RISUTE/**RISOLS**		

Bootstrapped 95% confidence intervals in parentheses. Statistical significance (difference from 0): *** p<0.01, ** p<0.05, * p<0. Statistical significance (difference from 1): ††† p<0.01, †† p<0.05, † p<0.1 (only indicated for the decompositions).
Source: Oregon Administrative Data, 1 Lottery Entrant in Household.
Calculation of the bold quantities does not rely on the linear MTE.
To obtain the bootstrapped 95% confidence intervals, we block bootstrap by household ID for 200 replications, and we report the 2.5 and 97.5 percentiles as the 95% confidence interval.

visit probability difference, 38% (-.80/-2.12) of the visit number difference, and 6% (-1,383/-23,601) of the total charge difference. In other words, some of the difference in ER utilization between insured lottery winners and insured lottery losers reflects adverse selection – the lottery losers that took up coverage had a higher propensity to consume ER care even when uninsured. However, the main reason for the difference is moral hazard – the lottery losers that took up coverage increased their utilization by more when they gained health insurance.

I can also decompose difference in treated outcomes between the always takers and treated compliers. The always takers visited the ER an average of 1.89 times, while the compliers with insurance visited the ER an average of 1.45 times. The BTTE estimate shows that health insurance increased the ER utilization of always takers by an average of 0.54 visits, and the LATE estimates shows that health insurance increased emergency room (ER) utilization for compliers by an average of 0.26 visits. The comparison of the decompositions in Columns 1 and 7 shows that moral hazard is responsible for a larger share of utilization for always takers than it is for compliers. Furthermore, differences in moral hazard between two groups explain 62% =(0.54-0.26)/(1.89-1.45) of the difference in visits between the two groups.

4.7 OLS Decomposition Results

As shown in Table 5, the baseline OLS (BOLS), intervention OLS (IOLS), and randomized intervention sample OLS (RISOLS) estimates are positive for all measures of ER utilization, indicating that insured individuals have higher ER utilization than uninsured individuals. Unlike the treatment effects estimated via the MTE, the OLS estimates can reflect selection in addition to a heterogeneous treatment effect. For all three measures of ER utilization, $IOLS < BOLS$, indicating that there is heterogeneity in the treatment effect, as previously formalized with the difference-in-difference test.

If, as in standard practice, we were to assume that the LATE is globally externally valid and divide the LATE by the RISOLS, we would conclude that the treatment effect is responsible for only 27% (=0.05/0.18) of the RISOLS estimate for any visits, 33% (=0.27/0.81) of the RISOLS estimate for number of visits, and 11% (=428/3,935) of the RISOLS for total charges. Instead, if we allow for a heterogeneous treatment effect estimated by the MTE and divide the RISTTE by the RISOLS, we see that the treatment effect actually has a much greater role. The treatment effect is responsible for 55% (0.10/0/18) of RISOLS for any visits, 54% (0.44/0.81) of RISOLS for number of visits, and 83%(=3,301/3,935) of RISOLS for total charges. The comparison of LATE to RISOLS understates the role of the treatment effect in this application because it does not acknowledge that treatment effects for always takers are larger than the treatment effects for compliers.

As discussed in Section 3.7, the RISOLS estimate is not very informative about the baseline world before

the experiment because it reflects the effect of the experimental intervention, and it reflects the experimental design through the shares that received the experimental intervention by winning the lottery. If we instead decompose the baseline OLS estimate, as shown in Table 5, we see that baseline OLS is less biased by selection than we would have concluded based on the results of the other decompositions: the treatment effect was responsible for 59% of the OLS estimate of the visit probability, 57% of the number of visits, and 88% of total charges.

4.8 Subgroup Analysis Estimates from LATE and Linear MTE

The columns of Tables 6-8 divide the OHIE sample into subgroups based on gender, English status, and age. For all measures of ER utilization, the LATE is larger for men, individuals who request materials in English, and individuals younger than median age. LATEs in some subroups are statistically different from the LATEs in the complementary subgroup and the full sample. However, only the subgroup that requested materials in a language other than English has negative LATEs for all three measures of utilization, but those LATEs are not statistically different from zero. The subgroup analysis reported by Taubman et al. [2014] in Table S14 yields similar results. On the whole, LATE subgroup analysis suggests that health insurance increases emergency room utilization for all or almost all individuals. Therefore, sample-re-weighting based only on observed hetergeneity in LATEs is unlikely to explain why some health insurance expansions decrease ER utilization.

However, the baseline and intervention treatment probabilities p_B and p_I vary quite a bit across subgroups. Therefore, if the MTE varies with p, even if it is exactly the same in all subgroups, then the LATEs in each subgroup will not be the same. The most noticeable difference in the baseline treatment probability is between women and men. The values of p_B in each subgroup indicate that 20% of women but only 10% of men gain coverage if they lose the lottery. Stated another way, 20% of women are always takers, but only 10% of men are always takers. This difference is another manifestation of the finding reported in Section 3.1 that compliers and the never takers are 53% female, but the always takers are 72% female.

The next rows report the intercepts and slopes of the linear MTE in each subgroup. For almost all measures of ER utilization and almost all subgroups, the linear MTE has a statistically-significant downward slope, indicating that there is treatment effect heterogeneity *within* each subgroup. As discussed in Section 3.5, when the linear MTE slopes downward, p^* gives the share of the sample with a positive treatment effect. In most subgroups, the MTE predicts that less than half of individuals have a positive treatment effect. When ER utilization is measured in terms of total charges, the linear MTE predicts that a sizeable share of Oregon compliers in each subgroup have negative treatment effects.

Heterogeneity in the MTE within each subgroup can potentially explain why some health insurance

Table 6: LATE and Linear MTE Subgroup Analysis: Any ER visits

	(1) Full sample	(2) Female	(3) Male	(4) English-speaker	(5) Non-English speaker	(6) Older than or at median age	(7) Younger than median age
LATE	0.05**	0.01	0.10***	0.06**	-0.11	0.05	0.06
	(0.00, 0.11)	(-0.07, 0.07)	(0.02, 0.17)	(0.01, 0.12)	(-0.26, 0.06)	(-0.02, 0.11)	(-0.02, 0.15)
Comparison with complementary sample	-	*	*	*	*		
Comparison with full sample	-	*	*		**		
$\mathbf{p_B}$	0.15***	0.20***	0.10***	0.15***	0.16***	0.13***	0.17***
	(0.15, 0.16)	(0.19, 0.20)	(0.09, 0.10)	(0.14, 0.16)	(0.14, 0.18)	(0.12, 0.14)	(0.16, 0.18)
Comparison with complementary sample	-	***	***			***	***
Comparison with full sample	-	***	***			***	***
$\mathbf{p_I}$	0.41***	0.44***	0.38***	0.41***	0.38***	0.43***	0.39***
	(0.40, 0.42)	(0.42, 0.46)	(0.36, 0.40)	(0.40, 0.43)	(0.35, 0.43)	(0.41, 0.45)	(0.38, 0.41)
Comparison with complementary sample	-	***	***			***	***
Comparison with full sample	-	***	***		*	***	***
Linear MTE intercept	0.15***	0.12*	0.21***	0.18***	-0.07	0.25***	0.07
	(0.06, 0.24)	(-0.01, 0.26)	(0.09, 0.33)	(0.08, 0.28)	(-0.31, 0.15)	(0.14, 0.36)	(-0.07, 0.21)
Comparison with complementary sample	-			*	*	*	*
Comparison with full sample	-			**	**	**	*
Linear MTE slope	-0.35***	-0.35**	-0.45**	-0.41***	-0.12	-0.71***	-0.03
	(-0.61, -0.12)	(-0.76, -0.02)	(-0.84, -0.09)	(-0.73, -0.16)	(-0.81, 0.69)	(-1.06, -0.37)	(-0.43, 0.35)
Comparison with complementary sample	-					**	**
Comparison with full sample	-			*		***	**
p*	0.43***	0.34*	0.46**	0.44***	-0.63	0.35***	2.54
	(0.28, 0.98)	(-0.03, 0.85)	(0.27, 1.01)	(0.29, 0.74)	(-2.02, 2.51)	(0.26, 0.49)	(-6.55, 4.38)
Comparison with complementary sample	-						
Comparison with full sample	-					*	
N	19,643	10,943	8,700	17,892	1,751	9,827	9,816

*** p<0.01, ** p<0.05, * p<0.1; Bootstrapped 95% confidence interval in parentheses.
Source: Oregon Administrative Data, 1 Lottery Entrant in Household

Table 7: LATE and Linear MTE Subgroup Analysis: Number of ER visits

	(1) Full sample	(2) Female	(3) Male	(4) English-speaker	(5) Non-English speaker	(6) Older than or at median age	(7) Younger than median age
LATE	0.27	0.14	0.39*	0.30*	-0.15	0.14	0.44
	(-0.08, 0.58)	(-0.27, 0.56)	(-0.04, 0.83)	(-0.06, 0.64)	(-0.76, 0.40)	(-0.29, 0.51)	(-0.07, 1.02)
Comparison with complementary sample	-						
Comparison with full sample	-						
$\mathbf{p_B}$	0.15***	0.20***	0.10***	0.15***	0.16***	0.13***	0.17***
	(0.15, 0.16)	(0.19, 0.20)	(0.09, 0.10)	(0.14, 0.16)	(0.14, 0.18)	(0.12, 0.14)	(0.16, 0.18)
Comparison with complementary sample	-	***	***			***	***
Comparison with full sample	-	***	***			***	***
$\mathbf{p_I}$	0.41***	0.43***	0.38***	0.41***	0.38***	0.43***	0.39***
	(0.40, 0.42)	(0.42, 0.45)	(0.37, 0.40)	(0.40, 0.43)	(0.35, 0.43)	(0.41, 0.45)	(0.38, 0.41)
Comparison with complementary sample	-	***	***			***	***
Comparison with full sample	-	***	***		*	***	***
Linear MTE intercept	0.64***	0.48	0.92***	0.72***	0.14	0.98***	0.31
	(0.16, 1.09)	(-0.14, 1.05)	(0.23, 1.49)	(0.24, 1.20)	(-0.64, 1.05)	(0.47, 1.58)	(-0.47, 1.19)
Comparison with complementary sample	-						
Comparison with full sample	-			*			*
Linear MTE slope	-1.32*	-1.06**	-2.20*	-1.51*	-1.07	-3.01***	0.48
	(-3.23, 0.19)	(-3.16, 1.00)	(-4.60, 0.32)	(-3.39, 0.08)	(-5.39, 2.30)	(-4.83, -1.17)	(-2.98, 3.37)
Comparison with complementary sample	-					**	**
Comparison with full sample	-					**	*
p*	0.48*	0.45	0.42*	0.48*	0.13	0.33**	-0.63
	(-1.75, 1.96)	(-3.22, 2.16)	(-1.36, 1.86)	(-0.94, 2.01)	(-0.92, 2.10)	(0.19, 0.56)	(-2.56, 2.50)
Comparison with complementary sample	-						
Comparison with full sample	-						
N	19,622	10,932	8,690	17,871	1,751	9,816	9,806

*** p<0.01, ** p<0.05, * p<0.1; Bootstrapped 95% confidence interval in parentheses.
Source: Oregon Administrative Data, 1 Lottery Entrant in Household

Table 8: LATE and Linear MTE Subgroup Analysis: ER Total Charges

	(1) Full sample	(2) Female	(3) Male	(4) English-speaker	(5) Non-English speaker	(6) Older than or at median age	(7) Younger than median age
LATE	$428	$358	$458	$579	-$1,698	-$977	$2,431
	(-$1,436, $2,142)	(-$1,786, $3,052)	(-$2,279, $2,486)	(-$1,498, $2,543)	(-$5,325, $2,757)	(-$3,285, $1,376)	(-$231, $5,095)
Comparison with complementary sample	-					*	*
Comparison with full sample	-					*	*
$\mathbf{p_B}$	0.15***	0.20***	0.10***	0.15***	0.16***	0.13***	0.17***
	(0.15, 0.16)	(0.19, 0.20)	(0.09, 0.10)	(0.14, 0.16)	(0.14, 0.19)	(0.12, 0.14)	(0.16, 0.18)
Comparison with complementary sample	-	***	***			***	***
Comparison with full sample	-	***	***			***	***
$\mathbf{p_I}$	0.41***	0.44***	0.38***	0.41***	0.38***	0.43***	0.39***
	(0.40, 0.42)	(0.42, 0.45)	(0.36, 0.40)	(0.40, 0.42)	(0.34, 0.42)	(0.41, 0.45)	(0.38, 0.41)
Comparison with complementary sample	-	***	***			***	***
Comparison with full sample	-	***	***		*	***	***
Linear MTE intercept	$6,677***	$3,526*	$12,621***	$7,141***	$3,273	$12,943***	$1,962
	($3,555, $9,326)	(-$536, $6,938)	($7,705, $17,980)	($3,848, $9,828)	(-$2,518, $12,965)	($7,459, $17,572)	(-$1,807, $5,848)
Comparison with complementary sample	-	***	***			***	***
Comparison with full sample	-	**	***			***	***
Linear MTE slope	-$22,218***	-$10,050*	-$51,011***	-$23,270***	-$18,170*	-$49,711***	$1,660
	(-$33,486, -$11,076)	(-$22,997, $2,825)	(-$76,451, -$30,594)	(-$35,548, -$10,572)	(-$45,109, $1,932)	(-$67,343, -$31,371)	(-$13,008, $13,983)
Comparison with complementary sample	-	***	***			***	***
Comparison with full sample	-	***	***			***	***
p*	0.30***	0.35	0.25***	0.31***	0.18	0.26***	-1.18
	(0.22, 0.45)	(-1.08, 1.24)	(0.20, 0.30)	(0.23, 0.46)	(-0.76, 0.46)	(0.21, 0.31)	(-7.68, 8.82)
Comparison with complementary sample	-						
Comparison with full sample	-		**			*	
N	19,628	10,939	8,689	17,877	1,751	9,813	9,815

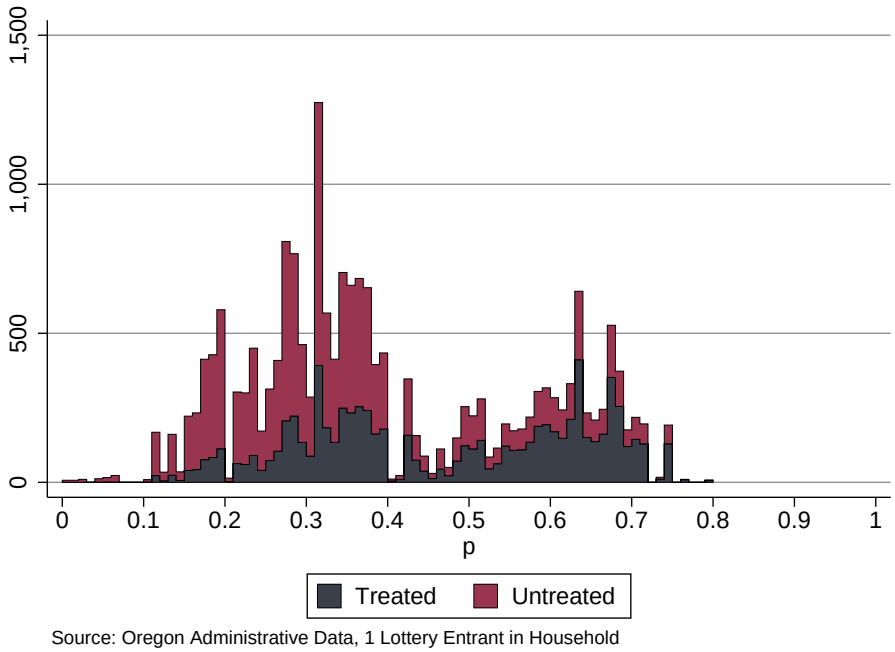
*** p<0.01, ** p<0.05, * p<0.1; Bootstrapped 95% confidence interval in parentheses.
Source: Oregon Administrative Data, 1 Lottery Entrant in Household

expansions decrease ER utilization. Although some variation is visible across subgroups, the linear MTE in each subgroup is broadly similar to the linear MTE in the full sample. The treatment effect varies with unobserved heterogeneity U_D in the full sample, and it also varies with unobserved heterogeneity in a similar way in most subgroups. Because the MTE is downward-sloping within each subgroup, it might be more efficient to combine all subgroups and estimate a single MTE with covariates.

4.9 Estimates of the MTE with Covariates

To estimate MTEs with covariates following the algorithm in Appendix A, I estimate a health insurance coverage propensity score for each individual in the sample based on covariates and whether the individual won the lottery. In Figure 2, I report a histogram of the estimated propensity scores in increments of 0.01. I shade the histogram to reflect the shares of treated and untreated individuals in each bin. Because the MTE is the difference between the marginal treated outcome and the marginal untreated outcome, the MTE is only nonparametrically identified in the common support of the treated and untreated. The range of the common support indicates that based on observable characteristics, few individuals have very low or high predicted probabilities of obtaining health insurance coverage, even conditional on winning the lottery. Therefore, I estimate the MTE with a global polynomial so that I can extrapolate outside of the common support.

Figure 2: Distribution of Estimated Propensity Scores in the OHIE



Source: Oregon Administrative Data, 1 Lottery Entrant in Household

For all three measures of ER utilization, Figure 3 depicts the estimated average linear MTE with covariates $\overline{MTE(x,p)}$ with a solid line. For comparison, Figure 3 depicts the linear MTE without covariates with a dot dash line. The magnitudes of the estimated MTEs are very similar. $\overline{MTE(x,p)}$ and $MTE(p)$ both indicate a substantial amount of treatment effect heterogeneity across the unobserved cost of treatment U_D.

The difference between the two dashed lines indicates the widest estimated range of treatment effect heterogeneity across any two groups with different vectors of covariates x. The vector of covariates the yields the minimum and maximum shift in the MTE differs across the three measures of ER utilization. For ER total charges, $\min MTE(x,p)$ depicts the MTE for 54 year old males who request materials in a language other than English. As discussed in Section 3.13, the MTE for this group is always below zero, implying that all members of this group decrease their ER utilization when they gain coverage. For ER total charges, $\max MTE(x,p)$ depicts the MTE for 62 year old males who request materials in English. The MTE for this group is always above zero, implying that all members of this group increase their ER utilization when they gain coverage.

Table 9 reports average treated outcomes $\overline{gTO(x)}$, $\overline{gUO(x)}$, and average treatment effects $\overline{gTE(x)}$ from the linear MTE with covariates $MTE(x,p)$ for comparison to the corresponding results from the linear MTE without covariates $MTE(p)$ reported in Table 4. As discussed in Appendix A, the global polynomial estimation algorithm for $MTE(x,p)$ estimates two separate regressions: one for the treated and another for the untreated members of the randomized intervention sample. Therefore the average predicted outcomes for the treated and untreated randomized intervention samples, $\overline{RISTTO(x)}$ and $\overline{RISUUO(x)}$ are equal to the observed average outcomes for the treated and untreated randomized intervention samples, and they are depicted in bold because they can be calculated without $MTE(x,p)$. The other columns of Table 9 report average predicted outcomes for various subsets of the treated and untreated randomized intervention samples. Because the functional form of the estimated global polynomial is not fully nonparametric, the average predicted outcomes should not exactly equal the average observed outcomes in subsets of the regression sample. However, the global polynomial is fairly flexible, so the average predicted outcomes from $MTE(x,p)$ are very similar to the observed outcomes depicted in bold in Table 4.

In theory, even though the linear $MTE(p)$ shows a substantial amount of heterogeneity across the unobservable U_D, the inclusion of covariates in $MTE(x,p)$ could make all unobserved heterogeneity observed, resulting in a flat $MTE(x,p)$. In that case, all of the average treatment effects derived from $MTE(x,p)$ should be equal to each other, and thus they should not be equal to the corresponding heterogeneous treatment effects derived from $MTE(x,p)$. In the case in which the inclusion of covariates $MTE(x,p)$ removes some unobserved heterogeneity, but some remains, the differences between $gTE(x)$ and the corresponding $\overline{gTE(x)}$ should reflect the influence of the included covariates on the treatment effect in group g. For ex-

Figure 3: Linear MTE with Covariates

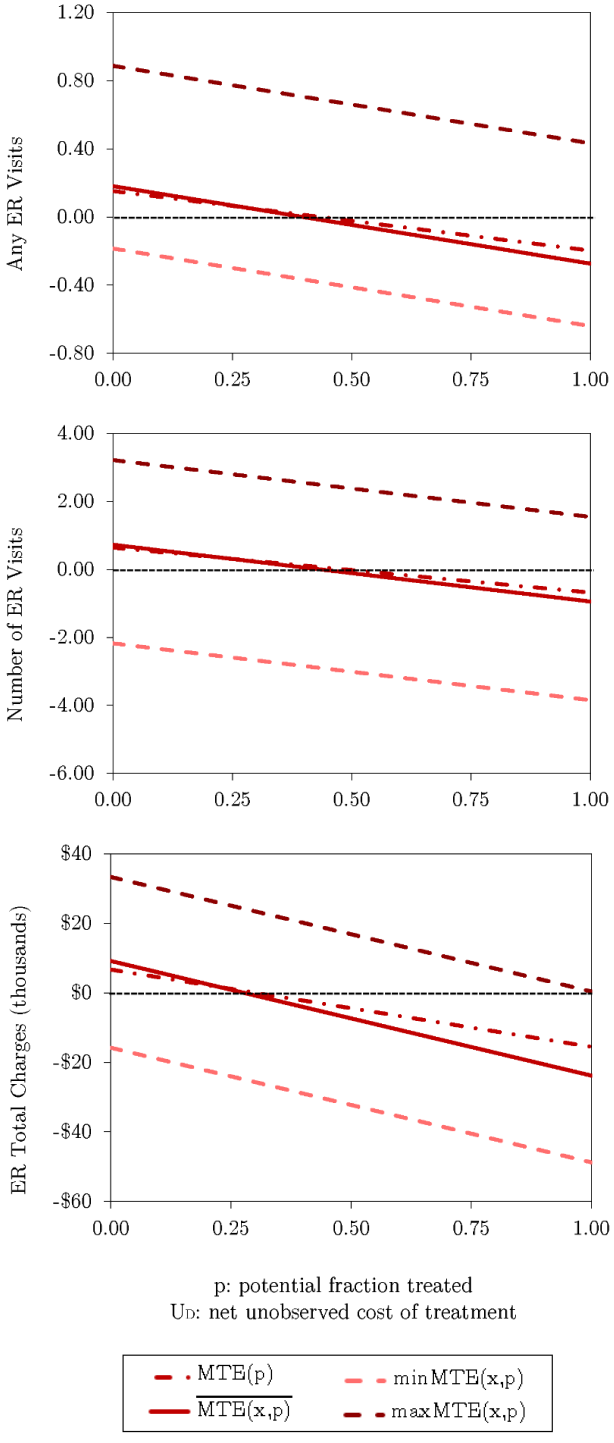


Table 9: Treated Outcome, Untreated Outcome, and Treatment Effects in Oregon: Linear MTE with Covariates

		(1) Baseline Treated (Always Takers) $\overline{BT(x)}$	(2) Baseline Untreated $\overline{BU(x)}$	(3) Intervention Treated $\overline{IT(x)}$	(4) Intervention Untreated (Never Takers) $\overline{IU(x)}$	(5) Randomized Intervention Sample Treated $\overline{RIST(x)}$	(6) Randomized Intervention Sample Untreated $\overline{RISU(x)}$	(7) Local Average (Compliers) $\overline{LA(x)}$	(8) Average $\overline{A(x)}$
Any ER Visits	Treated	0.55***	0.26***	0.48***	0.16*	**0.51*** **	0.24***	0.43***	0.30***
	Outcome	(0.53, 0.57)	(0.11, 0.36)	(0.46, 0.51)	(-0.02, 0.29)	**(0.50, 0.53)**	(0.08, 0.34)	(0.38, 0.47)	(0.18, 0.38)
	TO	$\overline{BTTO(x)}$	$\overline{BUTO(x)}$	$\overline{ITTO(x)}$	$\overline{IUTO(x)}$	$\overline{RISTTO(x)}$	$\overline{RISUTO(x)}$	$\overline{LATO(x)}$	$\overline{ATO(x)}$
	Untreated	0.41***	0.34***	0.40***	0.31***	0.40***	**0.33*** **	0.38***	0.35***
	Outcome	(0.31, 0.50)	(0.33, 0.35)	(0.32, 0.47)	(0.29, 0.33)	(0.32, 0.49)	**(0.32, 0.34)**	(0.32, 0.43)	(0.33, 0.37)
	UO	$\overline{BTUO(x)}$	$\overline{BUUO(x)}$	$\overline{ITUO(x)}$	$\overline{IUUO(x)}$	$\overline{RISTUO(x)}$	$\overline{RISUUO(x)}$	$\overline{LAUO(x)}$	$\overline{AUO(x)}$
	Treatment	0.14***	-0.08*	0.08**	-0.15**	0.11**	-0.09*	0.05	-0.05
	Effect	(0.05, 0.24)	(-0.22, 0.02)	(0.01, 0.16)	(-0.33, -0.02)	(0.03, 0.20)	(-0.25, 0.01)	(-0.02, 0.11)	(-0.17, 0.04)
	TE	$\overline{BTTE(x)}$	$\overline{BUTE(x)}$	$\overline{ITTE(x)}$	$\overline{IUTE(x)}$	$\overline{RISTTE(x)}$	$\overline{RISUTE(x)}$	$\overline{LATE(x)}$	$\overline{ATE(x)}$
Number of ER Visits	Treated	1.91***	0.74	1.61***	0.35	**1.75*** **	0.64	1.39***	0.90***
	Outcome	(1.75, 2.06)	(-0.16, 1.56)	(1.50, 1.78)	(-0.78, 1.39)	**(1.67, 1.87)**	(-0.32, 1.51)	(1.12, 1.67)	(0.16, 1.57)
	TO	$\overline{BTTO(x)}$	$\overline{BUTO(x)}$	$\overline{ITTO(x)}$	$\overline{IUTO(x)}$	$\overline{RISTTO(x)}$	$\overline{RISUTO(x)}$	$\overline{LATO(x)}$	$\overline{ATO(x)}$
	Untreated	1.34***	0.95***	1.26***	0.84***	1.30***	**0.92*** **	1.16***	1.01***
	Outcome	(0.90, 1.69)	(0.92, 1.01)	(0.93, 1.54)	(0.78, 0.92)	(0.92, 1.60)	**(0.89, 0.97)**	(0.93, 1.38)	(0.93, 1.11)
	UO	$\overline{BTUO(x)}$	$\overline{BUUO(x)}$	$\overline{ITUO(x)}$	$\overline{IUUO(x)}$	$\overline{RISTUO(x)}$	$\overline{RISUUO(x)}$	$\overline{LAUO(x)}$	$\overline{AUO(x)}$
	Treatment	0.57***	-0.21	0.34**	-0.49	0.45**	-0.29	0.23	-0.11
	Effect	(0.17, 1.02)	(-1.12, 0.62)	(0.02, 0.77)	(-1.63, 0.53)	(0.12, 0.88)	(-1.26, 0.60)	(-0.15, 0.62)	(-0.85, 0.61)
	TE	$\overline{BTTE(x)}$	$\overline{BUTE(x)}$	$\overline{ITTE(x)}$	$\overline{IUTE(x)}$	$\overline{RISTTE(x)}$	$\overline{RISUTE(x)}$	$\overline{LATE(x)}$	$\overline{ATE(x)}$
ER Total Charges	Treated	\$9,118***	-\$6,565*	\$5,648***	-\$11,195**	**\$7,232*** **	-\$7,776*	\$2,679***	-\$4,186*
	Outcome	(\$7,909, \$10,115)	(-\$13,263, \$1,279)	(\$4,852, \$6,571)	(-\$18,629, -\$1,657)	**(\$6,576, \$7,893)**	(-\$14,647, \$525)	(\$816, \$5,096)	(-\$9,532, \$2,078)
	TO	$\overline{BTTO(x)}$	$\overline{BUTO(x)}$	$\overline{ITTO(x)}$	$\overline{IUTO(x)}$	$\overline{RISTTO(x)}$	$\overline{RISUTO(x)}$	$\overline{LATO(x)}$	$\overline{ATO(x)}$
	Untreated	\$3,253***	\$3,180***	\$3,369***	\$2,968***	\$3,316***	**\$3,124*** **	\$3,287***	\$3,170***
	Outcome	(\$1,039, \$5,720)	(\$2,939, \$3,440)	(\$1,678, \$5,322)	(\$2,532, \$3,379)	(\$1,377, \$5,485)	**(\$2,909, \$3,326)**	(\$2,107, \$4,677)	(\$2,702, \$3,702)
	UO	$\overline{BTUO(x)}$	$\overline{BUUO(x)}$	$\overline{ITUO(x)}$	$\overline{IUUO(x)}$	$\overline{RISTUO(x)}$	$\overline{RISUUO(x)}$	$\overline{LAUO(x)}$	$\overline{AUO(x)}$
	Treatment	\$5,865***	-\$9,744**	\$2,280**	-\$14,163***	\$3,916***	-\$10,900***	-\$608	-\$7,357**
	Effect	(\$2,913, \$7,977)	(-\$16,446, -\$1,975)	(\$339, \$4,208)	(-\$21,742, -\$4,501)	(\$1,636, \$5,890)	(-\$17,769, -\$2,629)	(-\$2,703, \$2,100)	(-\$12,857, -\$1,330)
	TE	$\overline{BTTE(x)}$	$\overline{BUTE(x)}$	$\overline{ITTE(x)}$	$\overline{IUTE(x)}$	$\overline{RISTTE(x)}$	$\overline{RISUTE(x)}$	$\overline{LATE(x)}$	$\overline{ATE(x)}$

*** p<0.01, ** p<0.05, * p<0.1. Bootstrapped 95% confidence intervals in parentheses.

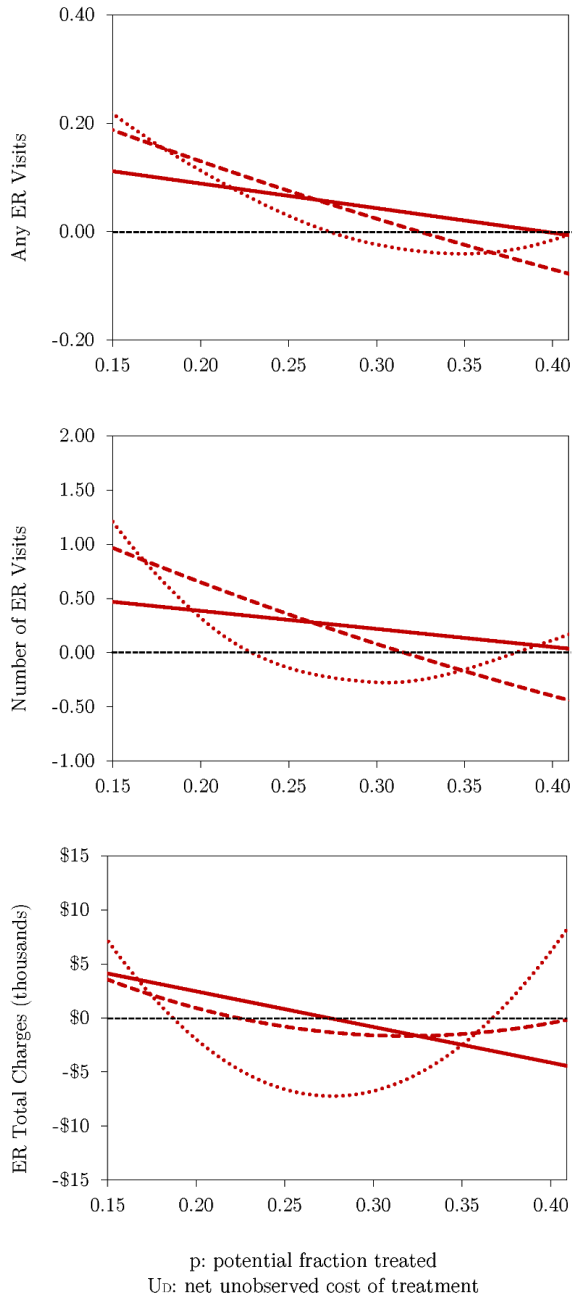
Source: Oregon Administrative Data, 1 Lottery Entrant in Household.

To obtain the bootstrapped 95% confidence intervals, we block bootstrap by household ID for 200 replications, and we report the 2.5 and 97.5 percentiles as the 95% confidence interval.

ample, unless the LATE without covariates is globally externally valid, $\overline{LATE(x)}$ need not be equal to the LATE without covariates. However, since $\overline{MTE(x,p)}$ is so similar to $MTE(p)$, as depicted in Figure 3, the all cells in Table 9 do not differ much from the corresponding cells in Table 4. Therefore, relative to the estimated linear MTE and the estimated linear MTE in each subgroup, the estimated linear MTE with covariates does not add much to our understanding of treatment effect heterogeneity in the OHIE.

In Figure 4, I report robustness of $MTE(x,p)$ to the order of the global polynomial by plotting the estimated quadratic $\overline{gTE(x)}$ and cubic $\overline{gTE(x)}$. I horizontal axis of these plots only includes the common support of the estimated treated and untreated propensity scores. As depicted, the linear, quadratic, and cubic polynomials all depict meaningful variations in treatment effect heterogeneity. Furthermore, with the exception of the cubic polynomial for ER total charges, they show that the treatment effect decreases as the fraction treated increases. Higher order global polynomials vary widely, especially outside of the common support because flexible extrapolation results in predictions that tend very quickly to positive or negative infinity in regions where there is no data.

Figure 4: Linear and Nonlinear MTEs with Covariates



4.10 Extrapolating the MTE to the Experiment Interpreted as a Natural Experiment

As discussed in Section 3.15, any experiment can be interpreted as a natural experiment that took place in the post-period but not in the pre-period for lottery winners. In the OHIE natural experiment, there are no always takers who had coverage in both periods because only uninsured individuals could enter the lottery. Therefore, we cannot estimate a separate MTE using the natural experiment, but we can use the observed change in outcomes from the pre-period to the experimental period ($Y - Y_{pre}$) to validate the predictions from the MTE.

Before extrapolating the MTE to the natural experiment, I run a Monte Carlo exercise to benchmark how well extrapolations based on the MTE should perform relative to extrapolations based on the LATE and the RISOLS in the OHIE randomized and natural experiments. I discuss the implementation of the Monte Carlo exercise in Appendix B.

Table 10 reports the results from the Monte Carlo exercise. Column 1 reports the mean bias and mean RMSE from 1,000 Monte Carlo simulations in which the true treatment effect θ is equal to the estimated LATE from the OHIE. In Column 2, the true treatment effect is equal to the estimated MTE from the OHIE.

For each measure of ER utilization, the first set of results report how well each estimator performs relative to the true treatment effect in the randomized experiment. I report the rank of each estimator in terms of absolute bias and RMSE in brackets, and I highlight the winning estimator. As reported in Column 1, when the LATE is the true value, the LATE has the smallest absolute mean bias for any ER visits, but the linear MTE has the smallest absolute mean bias for the number of visits and total charges. The magnitude of the mean bias for the linear MTE for the number of visits shows that the linear MTE under-predicts the true treatment effect for total charges by only \$1.23. The LATE has the lowest RMSE for all three measures of utilization, but the linear MTE comes in a close second. As reported in Column 2, when the MTE is the true value, the linear MTE substantially out-performs the LATE in terms of bias and RMSE. The RISOLS always performs the worst in terms of bias and RMSE.

For each measure of ER utilization, the second set of results report how well each estimator performs relative to the true treatment effect in the natural experiment. To mimic the natural experiment, for each observation, I set the true treatment effect equal to D^*TE, and I set the estimated treatment effect equal to $\widehat{D^*TE}$, since only the individuals who gain coverage should change their utilization from the pre-period to the experimental period. The performance of the linear MTE relative to the LATE and the RISOLS is similar for the natural experiment and the randomized experiment. These results suggest that in a simulation

Table 10: Monte Carlo Exercise

	(1)				(2)			
	θ = LATE				θ = MTE			
	Bias		RMSE		Bias		RMSE	
Any ER Visits								
TE								
RISOLS	0.08124	[3]	0.081	[3]	0.20284	[3]	0.227	[3]
LATE	0.00006	[1]	0.002	[1]	0.07642	[2]	0.127	[2]
Linear MTE	0.00008	[2]	0.003	[2]	0.00019	[1]	0.071	[1]
D*TE								
RISOLS	0.01959	[3]	0.040	[3]	0.01959	[3]	0.071	[3]
LATE	0.00002	[2]	0.001	[1]	-0.01089	[2]	0.030	[2]
Linear MTE	0.00001	[1]	0.001	[2]	0.00001	[1]	0.016	[1]
Number of ER Visits								
TE								
RISOLS	0.36703	[3]	0.367	[3]	0.82740	[3]	0.911	[3]
LATE	-0.00005	[2]	0.011	[1]	0.28907	[2]	0.479	[1]
Linear MTE	0.000003	[1]	0.012	[2]	0.00003	[1]	0.479	[1]
D*TE								
RISOLS	0.08842	[3]	0.180	[3]	0.08842	[3]	0.195	[3]
LATE	-0.00001	[1]	0.005	[1]	-0.04126	[2]	0.112	[2]
Linear MTE	-0.00002	[2]	0.006	[2]	-0.00002	[1]	0.062	[1]
ER Total Charges								
TE								
RISOLS	634.24	[3]	634.236	[3]	8,369.13	[3]	10544.320	[3]
LATE	-1.33	[2]	18.475	[1]	4,854.97	[2]	8044.900	[2]
Linear MTE	-1.23	[1]	20.172	[2]	-16.04	[1]	4534.230	[1]
D*TE								
RISOLS	152.71	[3]	311.207	[3]	152.71	[2]	1287.023	[2]
LATE	-0.3	[1]	9.064	[1]	-693.46	[3]	1886.286	[3]
Linear MTE	-0.31	[2]	10.386	[2]	-0.31	[1]	1036.856	[1]

The presented results are means obtained from 1,000 samples of size corresponding to the sample siz
Rankings for bias are based on absolute value.

designed to mimic the OHIE, there is not much to lose by running the linear MTE if the true treatment effect is equal to the LATE, but there is much to gain by running the linear MTE if the true treatment effect is equal to the linear MTE.

Next, we turn to validating our MTE results using the natural experiment. For the natural experiment, the true treatment effect is equal to the difference in outcomes between the experimental period and the pre-period: $D^*\theta = Y - Y_{pre}$. Unfortunately, the pre-period outcome Y_{pre} is not directly comparable to the experimental outcome Y. Individuals had to be uninsured for 6 months to enter the lottery, but the pre-period data aggregate ER utilization over a longer time period, and they do not include any information on pre-period insurance coverage.[27]. Therefore, I interpret the findings from the validation exercise with caution.

I did not include any estimators that require covariates in the Monte Carlo, but I do include estimators with covariates in the validation exercise. I perform the validation exercise in the full sample and in all subgroups reported in Table6. In the full sample, I assess the performance of the linear MTE, the linear MTE with covariates, the LATE, and the RISOLS. In the subgroups, I also assess the performance of the linear MTE, the LATE, and the RISOLS in each subgroup.

Column 1 of Table 11 reports results from the validation exercise in the full sample. The bias and RMSE from the validation exercise in the full sample should be directly comparable to the bias and MSE from the Monte Carlo exercise designed to mimic the natural experiment. In practice, the biases from the validation exercise are only somewhat larger, but the RMSEs are substantially larger. The linear MTE has the smallest absolute mean bias for every measure of ER utilization, and it also ranks first in terms of RMSE for any ER visits and ER total charges. For number of visits, the linear MTE with covariates ranks first in terms of RMSE, suggesting that the additional structure improves the efficiency of the MTE estimate for that outcome.

[27]The pre-period took place from January 1, 2007 to March 9, 2008, and the post-period took place from March 10, 2008 through September 30, 2009.

Table 11: Extrapolating to the Experiment Interpreted as a Natural Experiment, $D^*\theta = Y - Y_{pre}$

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Full sample		Female		Male		English speaker		Non-English speaker		Older than or at median age		Younger than median age	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Any ER Visits														
Linear MTE	0.002 [1]	0.274 [1]	0.007 [3]	0.290 [3]	-0.005 [2]	0.253 [2]	0.001 [1]	0.277 [1]	0.009 [3]	0.235 [3]	-0.005 [2]	0.273 [2]	0.010 [4]	0.275 [4]
Linear MTE in subgroup	-	-	-0.003 [1]	0.289 [1]	0.005 [3]	0.253 [3]	0.006 [2]	0.278 [3]	-0.036 [6]	0.246 [5]	0.003 [1]	0.274 [3]	0.002 [2]	0.274 [2]
Linear MTE with covariates	0.022 [4]	0.277 [3]	0.029 [7]	0.291 [4]	0.013 [6]	0.257 [7]	0.022 [6]	0.279 [5]	0.030 [5]	0.249 [7]	0.013 [3]	0.267 [1]	0.032 [7]	0.287 [7]
LATE	-0.009 [2]	0.274 [2]	-0.006 [2]	0.290 [2]	-0.013 [5]	0.254 [5]	-0.009 [4]	0.278 [4]	-0.002 [1]	0.235 [1]	-0.016 [5]	0.275 [5]	-0.002 [3]	0.274 [3]
LATE in subgroup	-	-	-0.018 [4]	0.292 [5]	-0.003 [1]	0.252 [1]	-0.007 [3]	0.278 [2]	-0.040 [7]	0.248 [6]	-0.017 [6]	0.275 [6]	0.001 [1]	0.274 [1]
RISOLS	0.022 [3]	0.277 [4]	0.029 [6]	0.295 [7]	0.012 [4]	0.254 [4]	0.021 [5]	0.281 [6]	0.028 [4]	0.241 [4]	0.014 [4]	0.274 [4]	0.030 [6]	0.280 [6]
RISOLS in subgroup	-	-	0.027 [5]	0.294 [6]	0.015 [7]	0.254 [6]	0.023 [7]	0.281 [7]	0.007 [2]	0.235 [2]	0.019 [7]	0.276 [7]	0.025 [5]	0.278 [5]
Number of ER Visits														
Linear MTE	-0.029 [1]	1.290 [2]	-0.014 [1]	1.373 [2]	-0.049 [4]	1.179 [4]	-0.036 [2]	1.325 [3]	0.039 [4]	0.858 [3]	-0.042 [2]	1.241 [3]	-0.017 [1]	1.338 [3]
Linear MTE in subgroup	-	-	-0.047 [3]	1.375 [3]	-0.018 [1]	1.176 [3]	-0.023 [1]	1.324 [2]	-0.072 [5]	0.866 [4]	-0.025 [1]	1.242 [5]	-0.034 [3]	1.338 [2]
Linear MTE with covariates	-0.047 [2]	1.283 [1]	-0.035 [2]	1.331 [1]	-0.063 [6]	1.212 [7]	-0.053 [4]	1.313 [1]	0.017 [3]	0.914 [7]	-0.053 [5]	1.204 [1]	-0.040 [4]	1.362 [7]
LATE	-0.072 [4]	1.296 [4]	-0.065 [4]	1.377 [4]	-0.080 [7]	1.187 [6]	-0.078 [7]	1.332 [7]	-0.005 [1]	0.854 [1]	-0.081 [6]	1.249 [6]	-0.063 [6]	1.342 [5]
LATE in subgroup	-	-	-0.093 [7]	1.382 [7]	-0.061 [5]	1.181 [5]	-0.071 [6]	1.330 [6]	-0.106 [6]	0.881 [5]	-0.106 [7]	1.257 [7]	-0.026 [2]	1.337 [1]
RISOLS	0.055 [3]	1.293 [3]	0.081 [6]	1.380 [6]	0.023 [2]	1.174 [1]	0.049 [3]	1.326 [4]	0.120 [7]	0.889 [6]	0.042 [3]	1.240 [2]	0.068 [7]	1.343 [6]
RISOLS in subgroup	-	-	0.076 [5]	1.379 [5]	0.032 [3]	1.175 [2]	0.059 [5]	1.327 [5]	0.013 [2]	0.854 [2]	0.049 [4]	1.241 [4]	0.061 [5]	1.342 [4]
ER Total Charges														
Linear MTE	-35 [1]	10,985 [1]	61 [2]	10,336 [3]	145 [2]	11,750 [1]	-93 [3]	11,370 [2]	564 [4]	5,731 [4]	-441 [4]	12,950 [2]	371 [4]	8,582 [6]
Linear MTE in subgroup	-	-	-342 [4]	10,344 [4]	256 [4]	11,769 [2]	-21 [1]	11,369 [1]	-108 [2]	5,612 [1]	-34 [1]	12,921 [1]	104 [1]	8,519 [1]
Linear MTE with covariates	1,396 [4]	11,646 [4]	1,688 [7]	10,770 [7]	992 [7]	12,762 [7]	1,331 [7]	11,968 [7]	2,098 [7]	7,326 [7]	846 [5]	13,403 [7]	1,969 [7]	9,478 [7]
LATE	-726 [3]	11,092 [3]	-776 [5]	10,431 [5]	-663 [5]	11,871 [5]	-782 [6]	11,488 [6]	-148 [3]	5,632 [3]	-1,080 [6]	13,152 [5]	-372 [5]	8,549 [4]
LATE in subgroup	-	-	-778 [6]	10,432 [6]	-670 [6]	11,873 [6]	-744 [5]	11,477 [5]	-678 [6]	5,793 [6]	-1,422 [7]	13,291 [6]	148 [3]	8,522 [3]
RISOLS	100 [2]	10,995 [2]	173 [3]	10,332 [2]	7 [1]	11,776 [3]	44 [2]	11,377 [3]	669 [5]	5,789 [5]	-282 [3]	12,973 [4]	481 [6]	8,571 [5]
RISOLS in subgroup	-	-	4 [1]	10,326 [1]	243 [3]	11,789 [4]	110 [4]	11,379 [4]	-12 [1]	5,624 [2]	101 [2]	12,962 [3]	116 [2]	8,520 [2]

Source: Oregon Administrative Data, 1 Lottery Entrant in Household

The pre-period was defined from January 1, 2007 to March 9, 2008. The post-period was defined from March 10, 2008 and September 30, 2009 (inclusive).

Individuals with missing values for the outcome variable or the corresponding pre-period measure were excluded.

Rankings for bias are based on absolute value.

Columns 2-7 report results from the validation exercise in each subgroup. As the highlighting shows, the three MTE estimators tend to out-perform the two LATE estimators and the two RISOLS estimators. It is difficult to discern a clear ranking within the three MTE estimators, but the MTE estimator within the subgroup tends to the have the lowest RMSE for ER total charges. On the whole, the Monte Carlo exercise and the validation exercise on the natural experiment show that the MTE estimators out-perform the LATE and OLS estimators in simulated and actual data.

4.11 Extrapolating the MTE to an Experiment on a Different Sample: The Massachusetts Health Reform

The extrapolation for the results from the OHIE to the Massachusetts health reform is of particular policy-relevance. Though the LATE from the OHIE shows that insurance increases ER utilization, the LATE from the Massachusetts Health Reform shows that insurance decreases ER utilization or leaves it unchanged. The extrapolation of the LATE from Oregon to Massachusetts cannot explain the difference in findings, but perhaps the extrapolation of the MTE can. In Sections 3.14 and 3.16, I discussed the restrictive conditions required to extrapolate the results from one experiment to another experiment on a different sample. Under the strong assumption that all of those conditions are met, I extrapolate the results from the OHIE to the Massachusetts Health Reform.

Before undertaking this exercise, I acknowledge that there are several factors that could have differed between both empirical contexts that MTE methods will not address directly. At a fundamental level, the Oregon expansion was a randomized experiment open to a relatively small group of subjects and the Massachusetts reform was a state-wide policy. Therefore, Oregon impacts likely occurred through the demand-side, but Massachusetts impacts could also occur through the supply-side. Health insurance terms could also differ, especially since Oregon expanded Medicaid alone and Massachusetts also expanded other types of coverage.

Furthermore, institutional features of the health care environment could differ 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 health reform. Also, both states could have different social norms regarding emergency room vs. primary care usage. Those social norms could differ between urban areas and rural areas. Though Massachustts is more urban than Oregon, the Oregon administrative data on ER utilization are only from the Portland area.

Table 12 presents observable demographic characteristics from the OHIE and Massachusetts side-by-side. The data from the Massachusetts health reform are the same data from the Behavioral Risk Factor Surveillance System (BRFSS) that I used in Kolstad and Kowalski [2012], restricted to include only individuals from

Massachusetts. Overall, the Massachusetts sample from the BRFSS includes 62,541 individuals, making it much larger than our primary OHIE sample of 19,643 individuals.

As shown in the top row, these data do not include any measures of emergency room utilization, but they do allow me to compare individual-level characteristics from the Massachusetts health reform with individual-level characteristics from the OHIE. The data from the other published studies that examine the impact of the Massachusetts health reform on emergency room visits are not at the individual level, or they only include individuals who visit the emergency room, making them unsuitable for this exercise.

Table 12: Average Characteristics and Outcomes of Always Takers, Never Takers, and Compliers: Oregon vs. Massachusetts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Oregon Health Insurance Experiment				Massachusetts Health Reform			
	Sample Average	Always Takers	Never Takers	Compliers	Sample Average	Always Takers	Never Takers	Compliers
Y: Any ER visit	0.37	0.55	0.31	0.40	-	-	-	-
Y: Number of ER visits	1.12	1.89	0.85	1.28	-	-	-	-
Y: ER total charges	$4,009	$8,794	$2,930	$3,661	-	-	-	-
Z: Selected in Oregon lottery	0.34	0.00	1.00	0.34	-	-	-	-
Z: Massachusetts, Post-Reform	-	-	-	-	0.42	0.00	1.00	0.42
D: Medicaid	0.24	1.00	0.00	0.34	0.92	1.00	0.00	0.42
X: Lottery entrants in household	1.00	1.00	1.00	1.00	-	-	-	-
X: Number of adults in household	-	-	-	-	1.86	1.86	1.82	1.86
X: Age in 2009	40.7	39.4	40.3	42.4	42.0	42.2	39.0	42.4
X: Female	0.56	0.72	0.53	0.53	0.51	0.52	0.38	0.43
X: English	0.91	0.90	0.91	0.92	0.96	0.98	0.81	0.86
Number of Observations	19,643	1,959	3,977	5,092	62,456	25,918	1,856	3,175

Sources: Oregon Administrative Data, 1 lottery entrant in household and Behavioral Risk Factor Surveillance System 2004-2009, Massachusetts data
Note that for the Massachusetts sample, there are more people in the treatment group than in the control group because there are more years of data in the post-reform period than in the pre-reform period. The pre-reform period spans 2004 through March 2006. The post-reform period spans July 2007 through 2009. The during-reform period, which spans April 2006 through June 2007, has been excluded from the analysis.

The key variables for extrapolation are the instrument Z and the treatment D. I define the Massachusetts instrument Z so that indicates whether the individual was in the sample post-reform. As discussed in the table notes, the Massachusetts sample includes slightly less data in the post-reform period. I define the treatment D such that the treatment represents Medicaid in Oregon and all types of insurance coverage in Oregon. This is a large assumption.

The next three rows compare the common covariates available in the Oregon data and the Massachusetts data. As shown, the Oregon sample is younger, they are more likely to be female, and they are less likely to request materials in English. Furthermore, *compliers* in Oregon are even younger, more female, and less likely to request materials in English than *compliers* in Massachusetts. Column 6 shows that the vast majority of the Massachusetts sample consists of always takers whose coverage was not affected by the Massachusetts health reform.

In Massachusetts, the baseline treatment probability p_B and the intervention treatment probability p_I are both very high relative to the OHIE experimental support. Furthermore, as discussed in Section 3.16, since the Oregon sample consists of individuals who selected to enter a lottery for insurance, the relevant p_B and p_I for extrapolation from Oregon could actually exceed 1. However, I proceed under the very conservative assumption that the distribution of unobserved heterogeneity U_D is the same in Oregon and Massachusetts, so that p_B and p_I from Oregon and Massachusetts are comparable.

Applying Massachusetts weights to the linear MTE from Oregon, I find negative LATEs for all three measures of ER utilization. I hesitate to read too much into the magnitudes of the estimates since they are so far outside of the Oregon support. The Oregon MTE implies that insurance should decrease ER utilization in Massachusetts by decreasing the visit probability by 0.17, decreasing the number of ER visits by -0.58, and decreasing ER total charges by $13,797. The any ER visit and ER total charges results are statistically different from zero at at least the 10% level. Though based on many restrictive assumptions, these results could potentially reconcile the entire discrepancy in the LATEs for ER utilization between the Oregon Health Insurance Experiment and the Massachusetts Health Reform.

5 Experimental Design for External Validity

The exercise of applying MTE methods to the OHIE brings to light several issues that should be considered in the design of future experiments to maximize the usefulness of MTE methods. The first issue is that a wide array of covariates should be collected to aid in tracing out a nonlinear MTE function. Ideally, these covariates should be defined such that they can also be obtained from other experimental or non-experimental data. By defining covariates in a standardized fashion, the MTE function from one context such as Oregon can be applied to another such as Massachusetts. Covariates used for stratification should be defined such that they can also be obtained from other experimental or non-experimental data.

A second issue is that researchers should collect data on always takers, compliers, and never takers. If researchers do not collect follow-up data on experimental subjects who lose the lottery but attain the intervention by other means, then always takers cannot be identified. Because the OHIE researchers collected data on individuals who lost the lottery but gained health insurance through other means, the MTE function can be estimated. However, if they did not collect such data, as is common in clinical trials, then only one of the separate pieces of the MTE function could be estimated.

A subtler issue is that experiments should be designed such that always takers and never takers are possible. If there are no always takers or no never takers, the full MTE function cannot be estimated. For example, in their empirical application, Brinch et al. [2012] cannot estimate the full MTE because there are

no never takers who have twins but do not end up with an extra child.

A related issue is that going to great lengths to encourage all participants who win the lottery to receive the intervention could make the estimated LATE less locally externally valid for other LATEs of interest. It could also limit the ability of the researchers to use the estimates to produce externally relevant ones. In the extreme case, if there is no selection into or out of treatment, then the LATE will be equal to the ATE. However, if the policy intervention based on the experiment will allow individuals to select into or out of treatment, then the ATE might not be locally externally valid for the LATE of interest, but there will be no way to estimate the LATE of interest with the experiment.

Another manifestation of going to great lengths to encourage full takeup of the experimental intervention is that the estimated ATE could be *smaller* than the LATE that would result from a policy of interest. If the individuals with the largest treatment effects select into treatment first, then then going to great lengths to get all individuals to take up the treatment if they win the lottery could dilute the policy-relevant LATE of interest. This insight could be especially valuable in clinical trials.

Finally, perhaps the most productive way to improve the ability of experiments to recover treatment effect heterogeneity with MTE methods is to run experiments with continuous instruments. The idea of "selective trials" proposed by Chassang et al. [2012] seems consistent with this idea. With a continuous intervention, or even several different discrete interventions, the assumptions required to identify treatment effect heterogeneity with the MTE are weaker.

5.1 Conclusion

Researchers run experiments to obtain a treatment effect estimate that is internally valid. However, the local average treatment effect (LATE) estimated by an experiment is not globally externally valid if the treatment effect varies across individuals. The LATE gives the average treatment effect for compliers who receive the treatment if and only if they win the experimental lottery. In many experiments, there are also always takers who always receive the treatment and never takers who never receive the treatment regardless of the experimental lottery. I show that it is possible to use such experiments to recover bounds on average treatment effects for always takers and never takers. These bounds can reject global external validity of the LATE in some cases, and they depend on weaker assumptions than existing tests of global external validity.

Building on existing methods to recover a marginal treatment effect (MTE) with a discrete instrument, I develop weights that allow me to recover average treatment effects for discrete groups of individuals created by a discrete instrument, including always takers and never takers. I use the recovered treatment effects to decompose group average treated outcomes into selection and treatment effects. I also decompose the sample OLS estimate into a selection effect and a treatment effect. This decomposition generalizes the comparison

of the OLS estimate to the LATE when the treatment effect is heterogeneous.

I apply these methods to the Oregon Health Insurance Experiment. The Oregon LATE indicates that obtaining insurance increases emergency room (ER) utilization for compliers. I find that the treatment effect of insurance on ER utilization decreases from always takers to compliers to never takers. I also find that potential uninsured ER utilization decreases from always takers to compliers to never takers. Therefore, the selection effect and the treatment effect of insurance on insured ER utilization decrease as a larger fraction of individuals gain insurance. The heterogeneous selection and treatment effects that I recover from the OHIE indicate that a different policy experiment could increase or decrease ER utilization, depending on which individuals it induces to gain coverage.

References

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

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

Joseph G. Altonji, Todd E. Elder, and Christopher R. Taber. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113(1):151–184, February 2005. URL `https://ideas.repec.org/a/ucp/jpolec/v113y2005i1p151-184.html`.

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. doi: 10.1257/pol.4.1.1. URL `http://www.aeaweb.org/articles.php?doi=10.1257/pol.4.1.1`.

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.

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

Marinho Bertanha and Guido W. Imbens. External validity in fuzzy regression discontinuity designs. Working Paper 20773, National Bureau of Economic Research, December 2014. URL `http://www.nber.org/papers/w20773`.

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.

BRFSS. Behavioral risk factor surveillance system survey data. *Center for Disease Control*. URL http://www.cdc.gov/brfss/annual_data/annual_2013.html.

Christian N Brinch, Magne Mogstad, and Matthew James Wiswall. Beyond late with a discrete instrument. heterogeneity in the quantity-quality interaction of children. Technical report, Research Department of Statistics Norway, 2012.

M. Kate Bundorf, Jonathan Levin, and Neale Mahoney. Pricing and Welfare in Health Plan Choice. *American Economic Review*, 102(7):3214–48, December 2012. URL https://ideas.repec.org/a/aea/aecrev/v102y2012i7p3214-48.html.

Sylvain Chassang, Gerard Padro I Miquel, and Erik Snowberg. Selective trials: A principal-agent approach to randomized controlled experiments. *American Economic Review*, 102(4):1279–1309, 2012. doi: 10.1257/aer.102.4.1279. URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.102.4.1279.

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.

Liran Einav, Amy Finkelstein, Stephen Ryan, Paul Schrimpf, and Mark R Cullen. Selection on moral hazard in health insurance. *The American Economic Review*, 103(1):178, 2013.

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 Finkelstein, Nathaniel Hendren, and Erzo F.P. Luttmer. The value of medicaid: Interpreting results from the oregon health insurance experiment. Working Paper 21308, National Bureau of Economic Research, June 2015. URL http://www.nber.org/papers/w21308.

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

James J. Heckman and Edward Vytlacil. Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3):669–738, 05 2005. URL http://ideas.repec.org/a/ecm/emetrp/v73y2005i3p669-738.html.

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. 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.

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

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.

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.

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

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. ISSN 0196-0644. doi: http://dx.doi.org/10.1016/j.annemergmed.2011.02.020. URL http://www.sciencedirect.com/science/article/pii/S0196064411001338.

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. doi: 10.1126/science.1246183. URL http://www.sciencemag.org/content/343/6168/263.abstract.

Appendices

A Global Polynomial MTE Estimation of an MTE with Covariates

Our goal is to estimate the marginal treatment effect $MTE(x,p)$, the marginal treated outcome $MTO(x,p)$, and the marginal untreated outcome $MUO(x,p)$, which we have specified as

$$MTE(x,p) = E(Y_T - Y_U|X=x, U_D=p) = (\beta_T - \beta_U)'x + (mto(p) - muo(p)). \tag{21}$$

$$MTO(x,p) = E(Y_T|X=x, U_D=p) = \beta_T'x + mto(p) \tag{22}$$

$$MUO(x,p) = E(Y_U|X=x, U_D=p) = \beta_U'x + muo(p). \tag{23}$$

Step 1: Specify the order M of the global polynomial

We specify the order $M \geq 1$ of the global polynomial for the unobservable components of the average treated and untreated outcome functions $ATO(x,p)$ and $ATO(x,p)$ as follows:

$$ATO(x,p) = E(Y_T|X=x, U_D \leq p) = \beta_T'x + ATO(p) \quad = \beta_T'x + \sum_{m=0}^{M} \gamma_{Tm} p^m \tag{24}$$

$$AUO(x,p) = E(Y_U|X=x, U_D > p) = -\beta_U'x + AUO(p) = -\beta_U'x + \sum_{m=0}^{M} \gamma_{Um} p^m. \tag{25}$$

These specifications imply that $MTE(x,p)$, $MTO(x,p)$ and $MUO(x,p)$ have the functional forms specified in (21)-(23) with M^{th} order global polynomials for $MTO(p)$[28], $MUO(p)$[29], and $(MTO(p) - MUO(p))$:

$$MTO(x,p) = \beta_T'x + \gamma_{T0} + \sum_{m=1}^{M} (m+1)\gamma_{Tm} p^m \tag{26}$$

$$MUO(x,p) = \beta_U'x - \gamma_{U0} - \sum_{m=1}^{M} m\gamma_{Um} p^{m-1} + \sum_{m=1}^{M} (m-1)\gamma_{Um} p^m. \tag{27}$$

[28] $MTO(p) = \frac{d[pATO(p)]}{dp} = p\frac{dATO(p)}{dp} + ATO(p)$

[29] $MUO(p) = \frac{d[(1-p)AUO(p)]}{d(1-p)} = -\frac{d[(1-p)AUO(p)]}{dp} = -(1-p)\frac{dAUO(p)}{dp} - AUO(p)$.

Step 2: Estimate the propensity score p

After dropping individuals with missing values for the outcome Y, we regress treatment D on the instrument Z and the covariates X. In our baseline, X includes interaction terms between covariates. We also interact Z with X so that we can harness variation in $p_B x$ and $p_I x$ across subgroups $X = x$. Using the coefficient estimates, we predict a propensity score $p \equiv P(D = 1|Z, X)$ for each individual.

Because the MTE is the difference between the marginal treated outcome and the marginal untreated outcome, we only estimate $MTE(x, p)$ on a common support for $MTO(x, p)$ and $MUO(x, p)$. We drop observations outside of the maximum common support of the predicted propensity scores conditional on $D = 1$ and $D = 0$. In addition, we drop observations above the 5th percentile and below the 95th percentile of the maximum common support.

Step 2: Estimate $ATO(x, p)$ and $AUO(x, p)$

We estimate the average treated outcome function $ATO(x, p)$ using only the treated observations (the observations with $D = 1$) that were not dropped in Step 2. We regress the outcome Y on the covariates X and a global polynomial in the predicted propensity score as specified in (24). We save the predicted coefficients β_T and $\sum_{m=0}^{M} \gamma_{Tm}$.

Similarly, we estimate the average untreated outcome function $ATO(x, p)$ using only the untreated observations (the observations with $D = 0$) that were not dropped in Step 2. We regress the outcome Y on the covariates X and a global polynomial in the predicted propensity score as specified in (25). We save the predicted coefficients β_U and $\sum_{m=0}^{M} \gamma_{Um}$.

Step 3: Construct estimates of $MTO(x, p)$, $MUO(x, p)$, and $MTE(x, p)$

Using the predicted coefficients saved from Step 2, we construct estimates of the marginal treated and untreated outcome functions $MTO(x, p)$ following (26) and (27). We then construct an estimate of the marginal treatment effect function $MTE(x, p)$ by reporting the difference between $MTO(x, p)$ and $MUO(x, p)$. These estimates give a value of the MTE for each value of the covariate x and each propensity score p.

B Implementation of the Monte Carlo Exercise

I generate each Monte Carlo to mimic my OHIE replication sample as closely as possible. Each Monte Carlo has the same number of observations N as my OHIE replication sample. I draw U_D so that it is uniformly distributed from 0 to 1. (This is equivalent to drawing ν from any distribution and setting U_D equal to the quantiles of ν.) I generate the binary instrument Z such that $P(Z = 1) = s(p_B)$ where $s(p_B)$ is the share of

lottery winners in the OHIE. I generate the binary treatment D such that $D = 1$ for the $p_B N$ observations with the lowest values of U_D among the observations with $Z = 0$. I also set $D = 1$ for the $p_I N$ observations with the lowest values of U_D among the observations with $Z = 1$. I generate $Y_U = MUO(U_D)$ using the $MUO(p)$ that I estimate in the OHIE so that there is some selection.

Next, I simulate two different versions of the outcome Y for each of the three measures of ER utilization. The first version reflects a homogenous treatment effect ("$\theta = LATE$") and the second version reflects a heterogeneous treatment effect ("$\theta = MTE$"). I generate

$$Y_T(\theta) = \begin{cases} Y_U + LATE & \text{if } \theta = LATE \\ Y_U + MTE(U_D) & \text{if } \theta = MTE \end{cases}$$

using the $LATE$ and the $MTE(p)$ that I estimate in the OHIE. Finally, I generate two versions of the observed outcome:

$$Y(LATE) = (1 - D)Y_U + DY_T(LATE)$$
$$Y(MTE) = (1 - D)Y_U + DY_T(MTE).$$

I retain the simulated $Y(LATE)$, $Y(MTE)$, D, Z, and the true treatment effect θ for each observation.

In each Monte Carlo sample, for $Y(LATE)$, $Y(MTE)$, I obtain an estimate of the treatment effect $\hat{\theta}$ using three estimators: RISOLS, LATE, and RISTTE from the linear MTE. I calculate the bias and MSE as follows:

$$Bias(\hat{\theta}) = E[\hat{\theta} - \theta]$$
$$RMSE(\hat{\theta}) = \sqrt{E[(\hat{\theta} - \theta)^2]}$$

I repeat for 1,000 Monte Carlo samples, and I report the mean bias and MSE across all samples. This exercise validates how well each estimator performs in the simulated randomized experiment.

In the natural experiment as opposed to the randomized experiment, we are interested in the predicted observed change in outcomes $D^*\hat{\theta}$ relative to the observed change in outcomes $D^*\theta$ because the true change in outcomes should be zero for all individuals who do not take up coverage ($D = 0$), and it should be equal to θ for all individuals who do take up coverage. Therefore, I also calculate mean bias and MSE as follows:

$$Bias(\hat{D\theta}) = E[D\hat{\theta} - D\theta]$$
$$RMSE(\hat{D\theta}) = \sqrt{E[(\hat{D\theta} - D\theta)^2]}$$

This exercise validates how well each estimator performs in the simulated natural experiment.