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

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

June 4, 2016

Abstract

I examine treatment effect heterogeneity within an experiment to inform external validity. The local average treatment effect (LATE) gives an average treatment effect for compliers. I bound and estimate average treatment effects for always takers and never takers by extending marginal treatment effect methods. I use these methods to separate selection from treatment effect *heterogeneity*, generalizing the comparison of OLS to LATE. Applying these methods to the Oregon Health Insurance Experiment, I find that the treatment effect of insurance on emergency room utilization decreases from always takers to compliers to never takers. Previous utilization explains a large share of the treatment effect heterogeneity. Extrapolations show that other expansions could increase or decrease utilization.

*amanda.kowalski@yale.edu. I thank Saumya Chatrath, Aigerim Kabdiyeva, Samuel Moy, and Ljubica Ristovska for excellent research assistance. Joseph Altonji, John Asker, Steve Berry, Christian Brinch, Lasse Brune, Pedro Carneiro, Raj Chetty, Mark Duggan, Caroline Hoxby, Liran Einav, Amy Finkelstein, Matthew Gentzkow, Jonathan Gruber, John Ham, Guido Imbens, Dean Karlan, Larry Katz, Jonathan Levin, Rebecca McKibbin, 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, Stockholm University, UC Davis, UCLA, 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:** compliers, difference-in-difference test, sample local average treatment effect (SLATE), marginal untreated outcome test, Massachusetts health reform, program evaluation, selection.

1 Introduction

A researcher runs an experiment to estimate a treatment effect that is internally valid. However, the local average treatment effect (LATE) obtained from an experiment is not globally externally valid if the treatment effect varies across individuals. I use information available *within* an experiment to determine whether the treatment effect is likely to vary *across* experiments.

The LATE gives the average treatment effect for compliers who select into treatment strictly according to random assignment (Imbens and Angrist [1994]). The LATE is equal to the difference in average outcomes between treated and untreated compliers. If the LATE is internally valid, then selection into treatment is random among compliers, but selection need not be random in the experiment as a whole. In many experiments, there are always takers who select into treatment and never takers who do not select into treatment regardless of random assignment.

The LATE does not depend on the treated outcome of always takers or the untreated outcome of never takers, but these observed outcomes can be informative about selection and treatment effect heterogeneity. A difference in the average untreated outcomes of compliers and never takers provides evidence of selection. A difference in the average treated outcomes of compliers and always takers provides evidence of selection, treatment effect heterogeneity, or both. Under the same assumptions required to identify the LATE, I use observed outcomes and assumptions about observed outcomes to separate selection from treatment effect heterogeneity.

Always takers are more likely to select into treatment than compliers, who are more likely to select into treatment than never takers. Assuming weak monotonicity or linearity in untreated outcomes from always takers to compliers to never takers, I bound or estimate the unobserved average untreated outcome of always takers. If the bound or estimate plus the LATE cannot equal the observed average treated outcome of always takers, then I conclude that the treatment effect cannot be the same for always takers and compliers.

I bound or estimate the unobserved average treated outcome for never takers by assuming weak monotonicity or linearity in treated outcomes from always takers to compliers to never takers. If the bound or estimate minus the LATE cannot equal the observed average untreated outcome of never takers, then I conclude that the treatment effect cannot be the same for never takers and compliers. If the treatment effect varies from compliers to other individuals within an experiment, then the LATE cannot be globally externally valid.

My methods build on marginal treatment effect (MTE) methods developed by Björklund and Moffitt [1987], Heckman and Vytlacil [1999, 2005, 2007], Carneiro et al. [2011], and Brinch et al. [forthcoming]. 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 provide approaches to identify the MTE in settings with discrete instruments, thus allowing for the application of MTE methods to experiments. Experiments make the separation of selection and treatment effect heterogeneity clear.

The weak monotonicity assumptions that I impose to bound treatment effects on always takers and never takers are weaker than the linearity assumptions imposed in the literature. In some

cases, the bounds that I obtain do not include the LATE. Therefore, the bounds provide new tests that can reject global external validity under weaker assumptions than existing tests.

Imposing the linearity assumptions made in the literature, I estimate average treatment effects for always takers, never takers, and other groups of interest. The literature identifies an MTE with a discrete instrument, but it recovers inframarginal treatment effects using weights developed for a continuous instrument. For example, the literature recovers a single "treatment on the treated" estimate, but I recover separate average treatment effects for always takers and compliers using weights that I develop for a discrete instrument.

Using the weights, I decompose average treated outcomes into heterogeneous selection and treatment effects. I also decompose OLS estimates into heterogeneous selection and treatment effects. This decomposition generalizes the Hausman [1978] test that compares OLS to LATE to account for selection and treatment effect heterogeneity.

Finally, imposing additive separability between observables and unobservables as in the literature, I incorporate observables to estimate more general MTE functions. I show that these MTE functions can be used to estimate different treatment effects for individuals with different observable characteristics. I develop an approach to quantify how much treatment effect heterogeneity is explained by observables. I also develop approaches to extrapolate treatment effects to individuals in other potential experiments using variation in observables and unobservables.

I apply these methods to examine the impact of insurance on emergency room (ER) utilization using data from the Oregon Health Insurance Experiment (OHIE). 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. 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 ER utilization will change in response to other health insurance expansions such as the Affordable Care Act.

The OHIE is arguably the "gold standard" for evidence on the impact of insurance on ER utilization because it is a recent randomized experiment, but there is reason to question the external validity of the Oregon LATE, which indicates that health insurance increased ER utilization (Taubman et al. [2014]). LATEs from a credible natural experiment that increased health insurance coverage, the Massachusetts reform of 2006, show that ER utilization decreased or stayed the same immediately after the reform (Miller [2012], Smulowitz et al. [2011], Chen et al. [2011], Kolstad and Kowalski [2012]). Evidence on the ER utilization of other populations of newly insured individuals also yields varying results (Currie and Gruber [1996], Anderson et al. [2012, 2014], Newhouse and Rand Corporation Insurance Experiment Group [1993]).

I find that the treatment effect of insurance on ER utilization decreases from always takers to compliers to never takers in the OHIE. Previous ER utilization explains a large share of the treatment effect heterogeneity. The treatment effect heterogeneity that I find in Oregon indicates that a different policy 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 application of the MTE without observable heterogeneity in Section 3, and I discuss application of the MTE with observable heterogeneity in Section 4. In Section 5, I discuss extrapolation. In Section 6, I apply MTE methods to the OHIE, and I extrapolate the results to other contexts. In Section 7, 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

Let D represent a binary treatment such as health insurance coverage, and let Y represent an observed outcome such as emergency room utilization. Y_T is the potential outcome of an individual in the treated state ($D = 1$), and Y_U is 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, and Y_U represents potential emergency room utilization without health insurance. 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 if the net benefit of treatment, I_D, is greater than or equal to zero. I_D consists of an observed component p and an unobserved component U_D as follows:

$$I_D = p - U_D. \tag{1}$$

Since U_D enters (1) negatively, I refer to it as the unobserved net *cost* of treatment. U_D 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 an equal or 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.

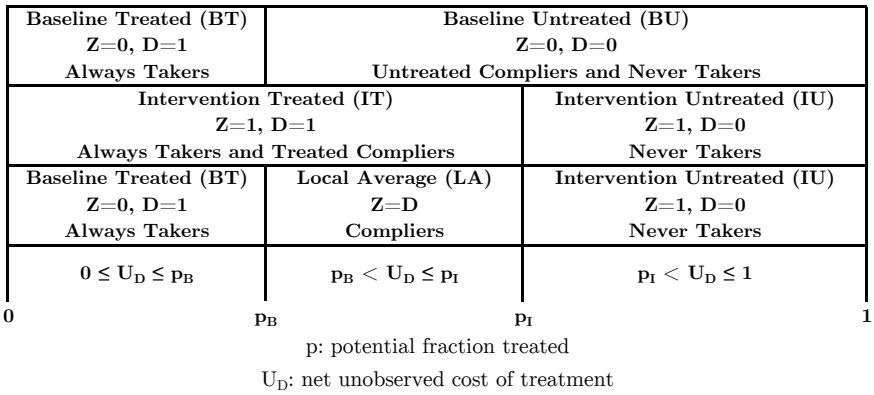
Since p enters (1) positively, I refer to it as the observed net *benefit* of treatment. In any experiment with a binary instrument Z that indicates winning the lottery, there are two observed values of p. The **baseline treatment probability** $\mathbf{p_B} \equiv P(D = 1|Z = 0)$ gives the potential fraction treated if the entire population were to remain in the baseline world without an experimental intervention. The **intervention treatment probability** $\mathbf{p_I} \equiv P(D = 1|Z = 1)$ gives the potential fraction treated if the entire population were to be eligible for the intervention. Because we can express $p \equiv P(D = 1|Z)$, we can refer to p as the potential fraction treated. The observed fraction treated in the full sample is a weighted average $P(D = 1) = s(p_B)p_B + s(p_I)p_I$, where $s(p_B) \equiv P(Z = 0)$ represents the share that loses the lottery and $s(p_I) \equiv P(Z = 1) = 1 - s(p_B)$

[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 standardized notation for concepts that I introduce later.

represents the share that wins the lottery.

Individuals select into treatment if their unobserved net cost of treatment U_D is less than or equal to the potential fraction treated p. As summarized in Figure 1, individuals with low net unobserved costs of treatment, $0 \leq U_D \leq p_B$, the **baseline treated (BT)**, select into treatment even if they lose the lottery ($D = 1$ and $Z = 0$), so they are **always takers**. Individuals with high net unobserved costs of treatment, $p_I < U_D \leq 1$, the **intervention untreated (IU)**, do not select into treatment even if they win the lottery ($D = 0$ and $Z = 1$), so they are **never takers**. The remaining individuals with intermediate net costs of treatment, $p_B < U_D \leq p_I$, the **compliers**, select into treatment strictly according to random assignment and determine the **local average (LA)**. Compliers who lose the lottery are **untreated compliers**, and compliers who win the lottery are **treated compliers**.[2]

Figure 1: Groups of Individuals in an Experiment



Always takers, compliers, and never takers often cannot be identified at the individual level because they are observed in mixed groups. Lottery losers who do not receive treatment ($D = 0$ and $Z = 0$), the **baseline untreated (BU)**, include untreated compliers and never takers ($p_B < U_D \leq 1$). Similarly, lottery winners who receive treatment ($D = 1$ and $Z = 1$), the **intervention treated (IT)**, include always takers and treated compliers ($0 \leq U_D \leq p_I$). In the full experimental sample of size N, the **randomized intervention sample treated (RIST)** includes all individuals with $D = 1$, all baseline and intervention treated. Similarly, the **randomized intervention sample untreated (RISU)** include all individuals with $D = 0$, all baseline and intervention untreated.

The depiction of p_B and p_I in Figure 1 provides more information than the first stage. By definition, $p_I - p_B$ is equal to the first stage, $P(D = 1|Z = 1) - P(D = 1|Z = 0)$. Therefore, the first stage gives the share of compliers, but 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

[2]The primitives of the model incorporate the assumptions required to identify a LATE discussed by Imbens and Angrist [1994]. The lottery affects the outcome through takeup (the instrument is relevant) and only through takeup (the exclusion restriction holds). Furthermore, the impact of the lottery on takeup is monotonic: there are no "defiers" that would have received the treatment at baseline but do not receive it given the intervention.

experimental intervention induces treatment of compliers with high or low net unobserved costs of treatment U_D relative to the entire sample.

An individual who is an always taker or never taker in a given experiment could be a complier in another experiment that induces different selection into treatment with a different instrument Z. Imagine that the distance between the baseline and intervention treatment probabilities becomes infinitesimal such that we can refer to a single intervention treatment probability. When the intervention treatment probability is zero, only the individual with the lowest net unobserved cost of treatment ($U_D = 0$) is a complier. As the intervention treatment probability increases from 0 to 1, the marginal individual at each value of $U_D = p$ is a complier.

2.2 The Marginal Treatment Effect MTE(p)

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:

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

$MTE(p)$ 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 some dimension of the net gain from treatment in dollars, then $MTE(p)$ can be interpreted as the willingness to pay for treatment with respect to Y for an individual at the margin of selecting into treatment, so $MTE(p)$ is a demand function. If the outcome Y instead represents some dimension of the cost of treatment in dollars, then $MTE(p)$ is a marginal cost function. In general, 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 also refer to the marginal untreated outcome $MUO(p)$ as the **marginal selection effect $MSE(p)$** because it identifies selection. Because untreated individuals do not receive treatment, any change in the untreated outcome as the fraction treated p increases reflects only selection. In the OHIE context, the difference in ER utilization between the uninsured lottery losers and the uninsured lottery winners identifies selection. $MUO(p)$ describes how the ER utilization of the marginal uninsured individual changes as coverage increases. 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 insurance markets, a downward-sloping $MUO(p)$ indicates **adverse selection**, and an upward-sloping $MUO(p)$ indicates **advantageous selection**. 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 basic plan decreases as enrollment in the generous plan p increases, then higher-cost individuals have adversely selected into the generous plan.

The marginal treated outcome $MTO(p)$ 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 treatment effects (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 potential 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 from $MUO(p)$. $MTO(p)$ includes the treatment effect and selection, and $MUO(p)$ includes only selection. Therefore, $MTE(p)$, the difference between $MTO(p)$ and $MUO(p)$, includes only the treatment effect.

In insurance markets, when 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 refer to it simply as moral hazard to avoid confusing it with selection.[3]

[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 heterogeneous moral hazard or selection or both. Under the Finkelstein and Poterba [2014] "unused observables" test, a correlation between a covariate and insurance coverage and a second correlation between the same covariate and insured spending could indicate heterogeneous moral hazard or selection or both. Under the Einav et al. [2013] cost curve test, an insured marginal cost curve $MTO(p)$ that is not constant could indicate heterogeneous moral hazard or selection or both. In Hackmann et al. [2015], my coauthors and I do not allow for heterogeneous moral hazard, so we refer to all variation in $MTO(p)$ as "selection." However, the cost curve test isolates selection when applied to the uninsured cost curve $MUO(p)$, and it isolates heterogeneous moral hazard when applied to the difference between the insured and uninsured cost curves, the $MTE(p)$.

2.3 Inframarginal Outcomes and Treatment Effects with MTE(p)

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

$$gTO = \int_0^1 \omega_g(p) MTO(p) dp \tag{2}$$

$$gUO = gSE = \int_0^1 \omega_g(p) MUO(p) dp \tag{3}$$

$$gTE = \int_0^1 \omega_g(p) MTE(p) dp \tag{4}$$

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 any 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: $gTE = gTO - gUO$. In Table 1, I introduce weights for the discrete groups of individuals introduced in Figure 1, as well as combinations of the groups.

In Column 1, which includes always takers, the **baseline treated treated outcome (BTTO)**: $E(Y_T|0 \leq U_D \leq p_B) = E(Y|Z = 0, D = 1)$ is observed, so it is reported in bold, along with all other quantities in Table 1 that do not require linearity of $MTO(p)$ or $MUO(p)$. The **baseline treated untreated outcome (BTUO)**: $E(Y_U|0 \leq U_D \leq p_B)$, is not observed because always takers always receive treatment, but it can be calculated with (3). The average treatment effect for always takers, the **baseline treated treatment effect (BTTE)**: $E(Y_T - Y_U|0 \leq U_D \leq p_B)$, can be calculated as the difference between the BTTO and the BTUO.

Column 2 includes the baseline untreated, which consists of untreated compliers and never takers. This group is policy-relevant because it includes all of the potential individuals to which health insurance could be expanded before the intervention. Column 3 includes the intervention treated, which consists of always takers and treated compliers.

As shown in Column 4, which includes never takers, the **intervention untreated untreated outcome (IUUO)**: $E(Y_U|p_I < U_D \leq 1) = E(Y|Z = 1, D = 0)$ is observed. The **intervention untreated treated outcome (IUTO)**: $E(Y_T|p_I < U_D \leq 1)$ is not observed, but it can be calculated with (2). The average treatment effect for never takers, the **intervention untreated treatment effect (IUTE)**: $E(Y_T - Y_U|p_I < U_D \leq 1)$, can be calculated as the difference between the IUTO and the IUUO.

Column 5 includes all treated individuals in the experiment, the randomized intervention sample treated, $RIST = s(p_B)BT + s(p_I)IT$. Column 6 includes all untreated individuals in the experiment, the randomized intervention sample untreated, $RISU = s(p_B)BU + s(p_I)IU$. Unlike the previous groups, these groups reflect the experimental design through the shares of individuals that win and lose the lottery.

Column 7 gives the local average weights, which represent the compliers. The **local average**

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

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

Calculation of the bold quantities does not rely on linearity of MTO(p) or MUO(p).

treatment effect (LATE): $E(Y_T - Y_U | p_B < U_D \leq p_I)$ gives the average treatment effect for compliers, which is equal to the difference between the **local average treated outcome (LATO)**: $E(Y_T | p_B < U_D \leq p_I)$ and the **local average untreated outcome (LAUO)**: $E(Y_U | p_B < U_D \leq p_I)$. 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 in experiments with always takers, always takers are also treated. The weights that I have introduced in Table 1 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. [forthcoming] 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 all treated individuals in the randomized intervention sample. The terms LATE, BTTE, ITTE, and RISTTE convey which groups of treated individuals are included, while "treatment on the treated" does not.

Column 8 reports the average weights $\omega_A(p) = 1$. The average weights represent all always takers, compliers, and never takers. The ATO, AUO, and ATE are not observed, but they can be calculated with $MTO(p)$, $MUO(p)$, and $MTE(p)$. In the OHIE context, 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 are the only weights in Table 1 that do not reflect a specific experimental intervention.

2.4 The External Validity of an Experiment

A general treatment effect gTE recovered from an experiment is **globally externally valid** if $MTE(p)$ is constant for all p. One treatment effect is **locally externally valid** for another if both treatment effects are equal. Empirically, the local average treatment effect from the OHIE might not be globally externally valid, but it could be locally externally valid for other treatment effects of interest.

3 Applying MTE(p) without Observables

3.1 Using Observables to Motivate Identification without Observables

Identification of $MTE(p)$ with an experiment relies on the same information that Katz et al. [2001] and Abadie [2003] use to identify the average observable characteristics of always takers, never takers, and compliers. Recall from the model 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 $p_I < U_D \leq 1$. The primitives of the model require that the experiment is **internally valid**: the distribution of

the unobserved net cost of treatment U_D is the same among lottery winners and losers.[4] Therefore, the shares of always takers, compliers, and never takers are the same among lottery winners and losers. The share of always takers is p_B; the share of compliers is $(p_I - p_B)$; and the share of never takers is $(1 - p_I)$. Although compliers cannot be observed directly, these shares help to identify their average characteristics.

Individuals who go untreated despite winning the lottery identify the average characteristics of never takers: $E(X|D = 0, Z = 1)$ 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. Using the shares of never takers and compliers, the average characteristics of untreated compliers are identified via

$$\frac{1}{p_I - p_B}\left[(1 - p_B)E(X|D = 0, Z = 0) - (1 - p_I)E(X|D = 0, Z = 1)\right]. \tag{5}$$

Similarly, 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 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. The average characteristics of treated compliers are identified via

$$\frac{1}{p_I - p_B}\left[p_I E(X|D = 1, Z = 1) - p_B E(X|D = 1, Z = 0)\right]. \tag{6}$$

Using the untreated and treated compliers, we can obtain an estimate of the weighted average characteristics of all compliers.[5]

In practice, few experimenters report the average characteristics of compliers, never takers, and always takers. If average characteristics are statistically the same across all groups, then they assert that the experimental LATE will be valid in other contexts. If characteristics are not the same, then experimenters still estimate the LATE. With some assumptions required for 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] The weighted average is equal to the average characteristics of untreated compliers from (5) weighted by $s(p_B)$ plus the average characteristics of treated compliers from (6) weighted by $s(p_I)$. We can also compare the untreated and treated compliers to test internal validity.

[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 by two-sample instrumental variable estimation. It is not surprising, then, that additional information from cross-tabulations can yield heterogeneous treatment effects.

3.2 Identifying Bounds on Outcomes and Treatment Effects

Applying (5) and (6) to an outcome Y in lieu of a characteristic X identifies the local average untreated outcome (LAUO) and the local average treated outcome (LATO):

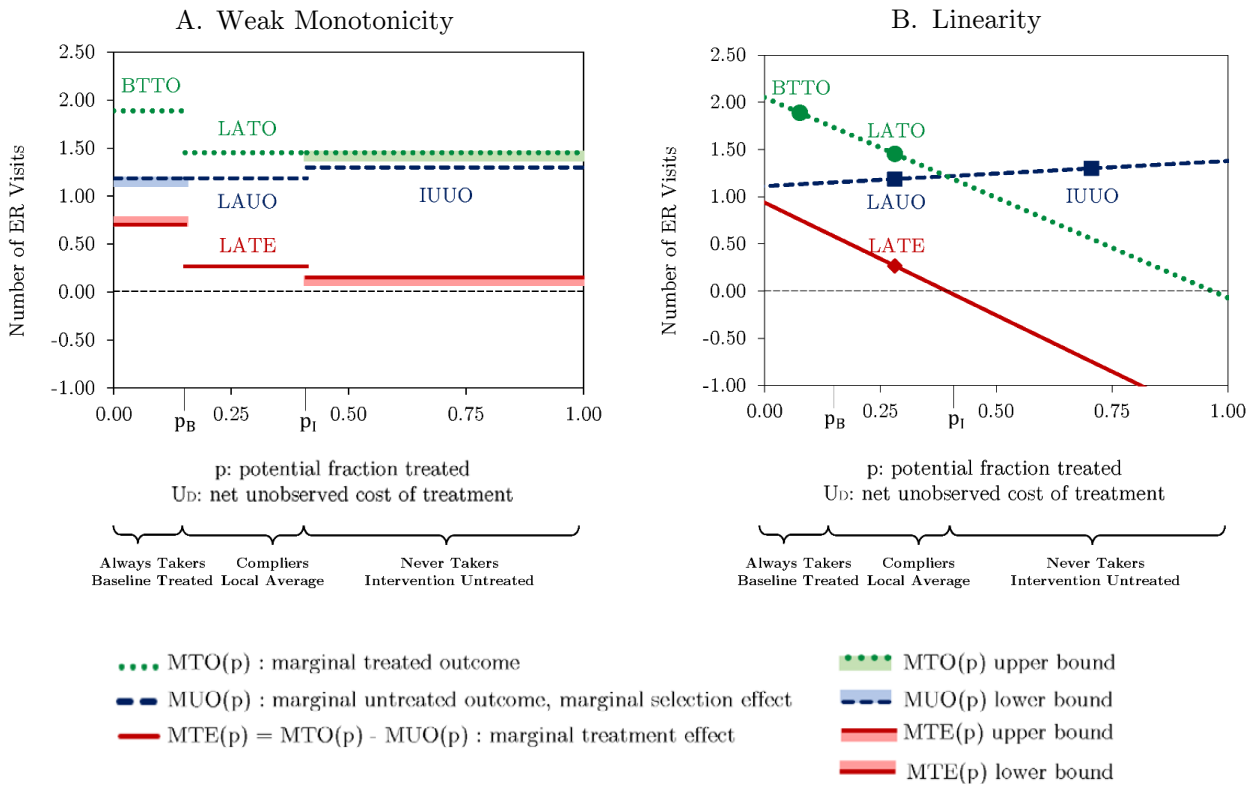
$$LAUO = \frac{1}{p_I - p_B}\left[(1-p_B)BUUO - (1-p_I)IUUO\right] \tag{7}$$

$$LATO = \frac{1}{p_I - p_B}\left[p_I ITTO - p_B BTTO\right], \tag{8}$$

where $BUUO = E(Y|D=0, Z=0)$, $IUUO = E(Y|D=0, Z=1)$, $BTTO = E(Y|D=1, Z=0)$, and $ITTO = E(Y|D=1, Z=1)$, following the notation introduced in Table 1.

Along the horizontal axis of Figure 2, the potential fraction treated p increases from no treatment to full treatment. As p increases, individuals with successively higher net unobserved costs of treatment U_D select into treatment. Always takers have lower net unobserved costs of treatment ($0 \leq U_D \leq p_B$) than treated compliers ($p_B < U_D \leq p_I$), so all of the always takers select into treatment before all of the treated compliers. 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.

Figure 2: Bounds and Estimates of $MTE(p)$ - Hypothetical example



Along the vertical axis of Figure 2, I depict treated outcomes, untreated outcomes, and treatment effects. In the left subfigures, I depict hypothetical observed values for the LAUO, the average untreated outcome of untreated compliers, and the IUUO, the average untreated outcome of never takers, with dashed lines over the relevant ranges of U_D. The hypothetical LAUO is less than the hypothetical IUUO, so the marginal selection function $MUO(p)$ slopes upward on average from p_B to 1. In the OHIE context, if the untreated compliers have lower average uninsured ER utilization than the never takers, then there is advantageous selection on average from the baseline level of coverage to full coverage. The average slope of $MUO(p)$ from 0 to p_B is not identified without further assumptions because the average untreated outcome for always takers, the baseline treated untreated outcome (BTUO), is not observed.

Comparison of the observed treated outcomes of always takers and treated compliers identifies the slope of the marginal treated outcome function $MTO(p)$ from 0 to p_I. In the hypothetical example depicted, the always takers have a higher average treated outcome than compliers ($BTTO > LATO$), so $MTO(p)$ slopes downward on average from 0 to p_I. In the OHIE context, if always takers have higher average insured ER utilization than treated compliers, then there could be adverse selection or a decreasing treatment effect on average as coverage increases from zero to the intervention level. The average slope of $MTO(p)$ from p_B to 1 is not identified without further assumptions because the average treated outcome for never takers, the intervention untreated treated outcome (IUTO), is not observed.

If we are willing to assume that the selection effect is weakly monotonic in p, implying that $MUO(p)$ is weakly monotonic in p, then we can obtain a bound on the average untreated outcome of always takers (the BTUO). The average outcome of untreated compliers provides an upper bound ($BTUO \leq LAUO < IUUO$) or a lower bound ($BTUO \geq LAUO > IUUO$), depending on the observed relationship between the average untreated outcomes of compliers and never takers. Alternatively, or in addition, if we are willing to assume that the treated outcome is weakly monotonic in p, implying that $MTO(p)$ is weakly monotonic in p, then we can obtain a bound on the average treated outcome of never takers (the IUTO). The average outcome of treated compliers provides an upper bound ($IUTO \leq LATO < BTTO$) or a lower bound ($IUTO \geq LATO > BTTO$). In the OHIE context, these bounds could be useful for forecasting the ER utilization associated with alternative policies.

Unlike alternative assumptions, weak monotonicity cannot be violated by observed outcomes[7] Weak monotonicity of $MUO(p)$ and $MTO(p)$ is a reasonable assumption if there is natural ordering of always takers, compliers, and never takers. Monotonic covariate values from always takers to compliers to never takers provide reassurance that weak monotonicity is reasonable.

The two separate assumptions of weak monotonicity of $MUO(p)$ and $MTO(p)$ form the basis for two separate tests of global external validity. As depicted in Figure 2, $LAUO < IUUO$, so we assume that $MUO(p)$ is weakly upward-sloping, implying that $BTUO \leq LAUO$. If we observe

[7]Huber et al. [2015] impose alternative assumptions to bound the BTUO and the IUTO. They assume that compliers have weakly larger treated and untreated outcomes than always takers and never takers. However, it is possible to observe violations of their assumptions, given that $LATO$, $LAUO$, $BTTO$, and $IUUO$ are observed.

that $BTTO > LATO$, so $MTO(p)$ is strictly upward-sloping in the same range, then the implied lower bound on the average treatment effect for always takers (BTTE) is strictly greater than the LATE, as depicted.[8] If we make the alternative or additional assumption that $MTO(p)$ is weakly upward-sloping, then the implied upper bound on the average treatment effect for never takers (IUTE) is strictly less than the LATE. Either bound can reject the global external validity of the LATE. The combination of both bounds implies $BTTE > LATE > IUTE$. Similarly, if $MUO(p)$ is weakly downward-sloping and MTO(p) is weakly upward-sloping, then the bounds imply $BTTE < LATE < IUTE$. In summary, if $MUO(p)$ and $MTO(p)$ have slopes of opposite sign, then the difference between them cannot be constant for all p, so both bounds reject global external validity.

If $MUO(p)$ and $MTO(p)$ have slopes of the same sign, then implied bounds on treatment effects are not informative about 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. It can also yield point estimates in lieu of bounds.

3.3 Identifying MTE(p)

Brinch et al. [forthcoming] impose linearity of $MUO(p)$ and $MTO(p)$ to identify a linear $MTE(p)$ with a binary instrument. Linearity is a stronger assumption than weak monotonicity. However, the traditional implicit assumption of global external validity of the LATE assumes that $MTE(p)$ is linear with a zero slope. Under linearity of $MUO(p)$ and $MTO(p)$, $MTE(p)$ is linear, but it can have a nonzero slope.

Linearity of $MUO(p)$ and $MTO(p)$ does not require that the individuals most likely to select into treatment have the largest treatment effects. For example, in the OHIE context, if the most risk averse individuals are the most likely to select into treatment, then $MUO(p)$ can slope downward. If the most risk averse individuals increase their ER utilization the least upon gaining insurance, then $MTO(p)$ and the resulting $MTE(p)$ can slope upward. In this example, the individuals most likely to select into treatment have the smallest treatment effects. The slopes of $MTO(p)$, $MUO(p)$, and $MTE(p)$ are determined empirically subject to the linearity assumptions.

To impose the linearity assumptions, Brinch et al. [forthcoming] 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. As depicted in the right subfigure of Figure 2, these assumptions introduce heterogeneity in outcomes within always takers, compliers, and never takers, while preserving

[8] The proof proceeds as follows:

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

the mean outcome within each group. Furthermore, they impose that the last always taker and the first complier have outcome $MTO(p_B)$, and the last complier and the first never taker have outcome $MUO(p_I)$.

Under these assumptions, the two points $(\frac{p_B+p_I}{2}, LAUO)$[9] and $(\frac{p_I+1}{2}, IUUO)$, labeled with square markers, identify the linear $MUO(p)$ and the two points $(\frac{p_B}{2}, BTTO)$ and $(\frac{p_B+p_I}{2}, LATO)$, labeled with circular markers, identify the linear $MTO(p)$:

$$MUO(p) = \frac{(1+p_I)BUUO - (1+p_B)IUUO}{p_I - p_B} + \frac{2(IUUO - BUUO)}{p_I - p_B}p \tag{9}$$

$$MTO(p) = BTTO - \frac{p_B}{p_I - p_B}(ITTO - BTTO) + \frac{2(ITTO - BTTO)}{p_I - p_B}p. \tag{10}$$

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

$$\begin{aligned} MTE(p) = &\frac{1}{p_I - p_B}(p_I(BTTO - BUUO) + p_B(IUUO - ITTO) + (IUUO - BUUO)) \\ &+ \frac{2}{p_I - p_B}((ITTO - IUUO) - (BTTO - BUUO))p. \end{aligned} \tag{11}$$

Given these functional forms, every element of Table 1 can be expressed in closed form.[10]

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

$$AUO(p) = BUUO - \frac{p_B}{p_I - p_B}(IUUO - BUUO) + \frac{IUUO - BUUO}{p_I - p_B}p \tag{12}$$

$$ATO(p) = BTTO - \frac{p_B}{p_I - p_B}(ITTO - BTTO) + \frac{ITTO - BTTO}{p_I - p_B}p, \tag{13}$$

from which they derive the marginal untreated outcome function $MUO(p)$[11] given by (9) and the marginal treated outcome function $MTO(p)$[12] given by (10).

[9] The combination of the linearity of $MUO(p)$ and the uniformity of U_D implies that the median complier ($p = \frac{p_B+p_I}{2}$) has the average treated outcome of all compliers.

[10] For example,

$$BTUO = \frac{(1+p_I-p_B)BUUO - IUUO}{p_I - p_B} \qquad BUTO = BTTO - \frac{BTTE - ITTO}{p_I - p_B}$$

$$ITUO = \frac{BUUO - (1+p_B-p_I)IUUO}{p_I - p_B} \qquad IUTO = \frac{ITTO(1-p_B+p_I) - BTTO}{p_I - p_B}.$$

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

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

3.4 Unexplained Heterogeneity with MTE(p)

If $MTE(p)$ has a nonzero slope, then there is unexplained treatment effect heterogeneity that can be quantified as follows:

$$RMSD = \sqrt{\int_0^1 (MTE(p) - ATE)^2 dp} \tag{14}$$

where the root mean squared deviation (RMSD) can be interpreted as the standard deviation of the unexplained variance in the outcome Y in the experimental sample, in the same units as Y. Under global external validity, $MTE(p) = ATE$ and $RMSD = 0$. The RMSD is more informative than the comparison of LATE to ATE. In the absence of global external validity, it can be possible that $LATE = ATE$, but it cannot be possible that $RMSD = 0$. Furthermore, when $LATE \neq ATE$, RMSD quantifies how much the treatment effect varies across individuals.

3.5 Identifying Optimal Treatment Probabilities with MTE(p)

$MTE(p)$ allows for positive treatment effects for some individuals and negative treatment effects for others. Suppose that $MTE(p)$ is downward-sloping. Define $\mathbf{p}^*$ as the **potential fraction treated p at which** $MTE(p)$ **is zero**:

$$p^* = -\frac{p_I(BTTO - BUUO) + p_B(IUUO - ITTO) + (IUUO - BUUO)}{2((ITTO - IUUO) - (BTTO - BUUO))}, \tag{15}$$

which gives the potential share of individuals in the experimental sample with a positive treatment effect when $MTE(p)$ is downward-sloping. 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. 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 $p_I \neq p^*$, then the optimal policy makes treatment more or less attractive to bring p_I closer to p^*. If $MTE(p)$ is always positive or negative in the range $0 \leq p \leq 1$, then it is optimal to treat everyone or no one, respectively.

Suppose that $MTE(p)$ is upward-sloping. 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 use blunt policy levers to target treatment optimally 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.

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] context, if Y measures the cost to the insurer in the generous plan relative to the basic plan, and p_B is optimal, then the LATE gives the deadweight loss due to moral hazard.[13]

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

3.6 Decomposing Treated Outcomes 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 weights ω_g, we can decompose the treated outcome into shares due to selection and treatment effects as follows:

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

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.

3.7 Decomposing OLS Estimates 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, the BTTO, to the observed outcome for a different group untreated individuals, the BUUO. To eliminate selection, we must compare the treated and untreated outcomes for the same individuals. We can compare the $BTTO = E(Y_T|0 \leq U_D \leq p_B)$ to the unobserved untreated outcome for the same group of individuals, the $BTUO = E(Y_U|0 \leq U_D \leq p_B)$, to obtain the BTTE. Alternatively, we can compare the $BUUO = E(Y_U|p_B < U_D \leq 1)$ to the unobserved treated outcome for the same group of individuals, the $BUTO = E(Y_T|p_B < U_D \leq 1)$, to obtain the BUTE. We can therefore decompose BOLS into shares due to selection and treatment effects in two ways as follows:

$$\frac{BOLS - BTTE}{BOLS} + \frac{BTTE}{BOLS} = 1$$

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

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

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 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. Decompositions of BOLS and IOLS are both of interest because they reflect selection and treatment effects for different groups of individuals.

Rather than reporting BOLS and IOLS separately, experimenters often report **randomized intervention sample OLS (RISOLS)**, the OLS estimate on the full sample. RISOLS is 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 shown in 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 $LATE = RISTTE = RISUTE$, and $LATE/RISOLS$ gives the share of the OLS estimate due to the treatment effect. However, if there is treatment effect heterogeneity, then RISTTE and RISUTE are directly comparable to RISOLS because they reflect the same group 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 $MTE(p)$, the comparison of BOLS to IOLS provides a test of global external validity.

3.8 Difference-in-Difference Test

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

$$Y = \lambda_{DZ} DZ + \lambda_D D + \lambda_Z Z + \lambda, \tag{16}$$

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 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 $BOLS = BTTO - BUUO$. 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 $IOLS - BOLS = ((ITTO - IUUO) - (BTTO - BUUO))$. If and only if IOLS is equal to BOLS, then $\lambda_{DZ} = 0$, and any treatment effect derived from $MTE(p)$ from an experiment is globally externally valid. When this condition holds, $MTE(p)$ has zero slope, per (11), so there is no treatment effect heterogeneity.

The regression in (16) 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.

Regardless of the outcome of the bounds tests introduced in Section 3.2, researchers will likely want to impose the linearity of $MTO(p)$ and $MUO(p)$ required for the difference-in-difference test. If the bounds reject global external validity, then linearity of $MTO(p)$ and $MUO(p)$ will allow researchers to recover a heterogeneous treatment effect with $MTE(p)$. If the bounds do not reject external validity, then linearity of $MTO(p)$ and $MUO(p)$ will allow researchers to run the difference-in-difference test. Researchers willing to run the difference-in-difference test can obtain $MTE(p)$ and all of the quantities derived from it without imposing further assumptions. Therefore, researchers willing to run tests from the literature that I implement with the difference-in-difference test should also be willing to report $MTE(p)$ and all of the quantities derived from it.

3.9 Difference-in-Difference Test Using Observables

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 (16) 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 coefficient 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.

3.10 Using Observables for Subgroup Analysis with MTE(p)

Researchers often perform subgroup analysis by estimating a LATE in each subgroup with covariate vector x. Researchers can also estimate a marginal treatment effect in each subgroup as follows:

$$MTE_x(p) = \frac{1}{p_{Ix} - p_{Bx}}(p_{Ix}(BTTO_x - BUUO_x) + p_{Bx}(IUUO_x - ITTO_x) + (IUUO_x - BUUO_x)) + \frac{2}{p_{Ix} - p_{Bx}}((ITTO_x - IUUO_x) - (BTTO_x - BUUO_x))p. \quad (17)$$

where (17) replaces all terms in (11) with their subgroup-specific values. For example, $BTTO_x = E(Y|D = 1, Z = 0, X = x)$. The comparison of $MTE(p)$ across subgroups informs whether the treatment effect varies in the same way with the unobserved cost of treatment U_D in each subgroup.

Even if the MTE is the same in each subgroup, then it can be misleading to compare LATEs across subgroups. The LATEs are only equal across all subgroups if the LATE in the full sample is globally externally valid. If the LATE in the full sample is not globally externally valid, then the LATEs are not necessarily comparable across subgroups because the probabilities of treatment conditional on losing the lottery $p_{Bx} \equiv P(D = 1|Z = 0, X = x)$ and winning the lottery $p_{Ix} \equiv P(D = 1|Z = 1, X = x)$ can differ across subgroups. The MTEs are comparable across subgroups even if p_{Bx} and p_{Ix} differ across subgroups.

4 Applying MTE(x,p) with Observables

4.1 Identifying MTE(x,p) with Observables

We can combine information across subgroups to estimate richer MTE functions. Within each subgroup with covariate vector x, the two points $(\frac{p_{Bx}+p_{Ix}}{2}, LAUO_x)$ and $(\frac{p_{Ix}+1}{2}, IUUO_x)$ identify a linear $MTO_x(p)$, and the two points $(\frac{p_{Bx}}{2}, BTTO)$ and $(\frac{p_{Bx}+p_{Ix}}{2}, LATO_x)$ identify a linear $MUO_x(p)$. However, if we assume that $MTE_x(p)$ is the same across subgroups, then we have more than four points to identify linear or *nonlinear* marginal treatment effects. By further subdividing the sample into finer subgroups, we can achieve nonparametric identification. Furthermore, if we are willing to impose some structure on how covariates enter an $MTE(x, p)$ function, then we can relax the assumption that $MTE_x(p)$ is the same across subgroups. If the structure holds, then it is more efficient to estimate a single $MTE(x, p)$ than it is to estimate a separate $MTE_x(p)$ within each subgroup.

Brinch et al. [forthcoming] specify the following functional forms that impose additive separability of observables and unobservables:

$$MTE(x, p) = E(Y_T - Y_U|X = x, U_D = p) = (\beta_T - \beta_U)'x + mte(p) \quad (18)$$

$$MTO(x,p) = E(Y_T|X=x, U_D=p) \qquad = \beta_T' x + mto(p) \tag{19}$$

$$MUO(x,p) = E(Y_U|X=x, U_D=p) \qquad = \beta_U' x + muo(p), \tag{20}$$

where $mto(p)$ and $muo(p)$ are general functions of p such as global or local polynomials, and $mte(p) = mto(p) - muo(p)$.[14] The first component of each function depends on a given observed vector of characteristics x, and the second component depends on the unobserved net cost of treatment $U_D = p$. Variation across subgroups in the observed outcomes identifies the additive shift terms $\beta_T' x$ and $\beta_U' x$, which determine the intercepts of each function. Variation across subgroups in the unobserved net cost of treatment through p_{Bx} and p_{Ix} identifies the parameters of the functions $mto(p)$ and $muo(p)$, which determine the slopes of each function. In these functional forms, the intercepts differ across subgroups, but the slopes do not.

It is tempting to think of $MTE(p)$ as an approximation to $MTE(x,p)$. However, the inclusion of covariates changes 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 $MTE(x,p)$ becomes a horizontal line.

To examine how the treatment effect varies with unobserved heterogeneity that still remains after taking all included covariates into account, the experimenter can graph the **sample marginal treatment effect (SMTE)**, the average $MTE(x,p)$ over all N individuals i in the experiment: $SMTE(p) = \frac{1}{N}\sum_i (\beta_T - \beta_U)' x_i + mte(p)$. To examine the how the treatment effect varies with observed heterogeneity, the experimenter can compare $MTE(x,p)$ for two different values of x. For example, the experimenter can examine the maximum amount of variation in the treatment effect that can be explained by observed heterogeneity by comparing the $MTE(x,p)$ with the smallest observable component, $\min MTE(x,p) = \min_i (\beta_T - \beta_U)' x_i + mte(p)$, to the $MTE(x,p)$ with the largest observable component, $\max MTE(x,p) = \max_i (\beta_T - \beta_U)' x_i + mte(p)$.

The $MTE(x,p)$ model generalizes the instrumental variable (IV) regression model. The traditional IV model imposes that the treatment effect is the same for all individuals, regardless of whether the model includes an additively-separable vector of observables. The IV model that interacts a vector of observables with a treatment indicator D allows the treatment effect to vary with observed heterogeneity. In contrast, the $MTE(x,p)$ model allows the treatment effect to vary with observed and unobserved heterogeneity.

4.2 Estimating MTE(x,p)

I detail a global polynomial algorithm for estimation of $MTE(x,p)$ in Section OA.1.[15] Estimation via low order global polynomials allows for extrapolation beyond the experimental support. Higher order global polynomials offer greater flexibility, but they rapidly approach positive or negative

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

[15] For all estimates, I bootstrap by household ID for 200 replications, and I report the standard deviation as the standard error or the 2.5 and 97.5 percentiles as the 95% confidence interval. I construct analogous intervals to obtain significance stars.

infinity just outside of experimental support. Local polynomials also offer greater flexibility, but they cannot be extrapolated without ad hoc assumptions. Furthermore, functions estimated via local polynomial estimation are not often smooth. Small jumps in the estimated the average treated and untreated outcome functions $ATO(x,p)$ and $AUO(x,p)$ can lead to wild fluctuations in the $MTE(x,p)$ functions derived from their slopes.

4.3 Identifying Optimal Treatment Probabilities with MTE(x,p)

Predictions from $MTE(x,p)$ allow the experimenter to assess which observable subgroups x are likely to react positively or negatively to an intervention. For some subgroups, the observable component of $MTE(x,p)$, $(\beta_T - \beta_U)'x$, might be large enough that the treatment effect is always positive. Similarly, there could be other subgroups in which the treatment effect is always negative. In any remaining subgroups, if $MTE(x,p)$ is linear, then analysis of positive and negative treatment effects follows from Section 3.5. If $MTE(x,p)$ is nonlinear, then it can cross zero at more than one point. Under the optimal policy, $MTE(x,p)$ should be weakly decreasing, and it should cross the optimal treatment threshold only once.

4.4 Inframarginal Outcomes and Treatment Effects with MTE(x,p)

We can construct inframarginal treated outcomes, untreated outcomes, and treatment effects for each subgroup with covariate vector x as follows:

$$gTO(x) = \int_0^1 \omega_g(x,p_x) MTO(x,p_x) dp_x \tag{21}$$

$$gUO(x) = gSE(x) = \int_0^1 \omega_g(x,p_x) MUO(x,p_x) dp_x \tag{22}$$

$$gTE(x) = \int_0^1 \omega_g(x,p_x) MTE(x,p_x) dp_x \tag{23}$$

where $p_x = p(D=1|Z, X=x)$ and the general weights are the same as those given in the bottom row of Table 1 with p_{Bx} and p_{Ix} in lieu of p_B and p_I.[16] To summarize (21)-(23) across subgroups in the experimental sample, using each individual i in the experimental sample, we can construct:

$$SgTO = \frac{1}{\sum_i P(i \in g)} \sum_i P(i \in g) gTO(x_i) \tag{24}$$

$$SgUO = SgSE = \frac{1}{\sum_i P(i \in g)} \sum_i P(i \in g) gUO(x_i) \tag{25}$$

$$SgTE = \frac{1}{\sum_i P(i \in g)} \sum_i P(i \in g) gTE(x_i), \tag{26}$$

[16] I estimate these probabilities using predicted probabilities from the estimation algorithm in Section OA.1.

where $P(i \in g)$ is the probability that each individual i in the sample is a member of group g.[17] The prefix S indicates the sample average over individuals with observables x_i in group g. For example, the **sample local average treatment effect (SLATE)** is the average treatment effect for the compilers, and the **sample average treatment effect (SATE)** is the average treatment effect for everyone in the experimental sample. The SLATE and the SATE average $MTE(x,p)$ over p and x, while the LATE and the ATE average $MTE(p)$ over p.

It is more informative to compare $SgTE$ across samples S with different observable characteristics than it is to compare $SgTE$ across groups g with different unobservable characteristics. The comparison of gTE across g is informative about global external validity. However, individuals in different groups can have different observable characteristics, so $SgTE$ need not equal for all g, even if there is no unexplained heterogeneity.

4.5 Unexplained Heterogeneity with MTE(x,p)

We can quantify unexplained treatment effect heterogeneity in $MTE(x,p)$ by generalizing (14) as follows:

$$RMSD(X_c) = \sqrt{\int_0^1 (SMTE(p) - SATE)^2 dp} \tag{27}$$

where $RMSD(X_c)$ can be interpreted as the standard deviation of the unexplained variance in the experimental sample, after taking a subset of the vector X_c of the available the vector of observables X into account with $MTE(x_c,p)$. Using this expression, we can decompose the unexplained treatment effect heterogeneity in $MTE(p)$ into the portion is explained by observables and the portion that remains unexplained:

$$\underbrace{\frac{RMSD(X_0) - RMSD(X_c)}{RMSD(X_0)}}_{\text{explained}} + \underbrace{\frac{RMSD(X_c)}{RMSD(X_0)}}_{\text{unexplained}} = 1,$$

where X_0 represents the null set of covariates. If $RMSD(X_c) = 0$, then all treatment effect heterogeneity is explained by covariates X_c incorporated into $MTE(x_c,p)$.

5 Extrapolation

5.1 Extrapolation Considerations

In theory, two experiments that are exactly the same should recover the same MTE. Therefore, a Hausman [1978] test should not reject the null hypothesis that both MTEs are the same, and it

[17] Table OA9 gives the expression for $P(i \in g)$ for the group g in each column. The probability that each individual is a complier, $P(i \in LA)$, incorporates the observed characteristics of compliers through p_{Bx} and p_{Ix}, in a way that is consistent with the observed characteristics of compliers derived in Section 3.1. I incorporate characteristics in this way so that the Oregon SLATE only reflects the observed characteristics of the Oregon compliers. Therefore, it is informative to compare the Oregon SLATE to the Oregon LATE, which reflects only Oregon compliers. For all g, the SgTE only reflects the observed characteristics of individuals in the group in the sample.

should be possible to extrapolate from one experiment to the other using either MTE. In practice, if anything varies across experiments, then experimenters should exercise caution in extrapolation.

For example, the outcome 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. In the OHIE context, if ER utilization and primary care utilization are substitutes, 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 for 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 functions because the individuals used to identify those functions vary as the definition of treatment varies.

Unobserved heterogeneity U_D must also be the same across experiments for the MTEs to be the same. If two experiments are drawn at random from a broader pool, then $MTE(p)$ and $MTE(x,p)$ should be the same. However, in $MTE(x,p)$, unobserved heterogeneity U_D is a function of observed heterogeneity captured in covariate vector X. Therefore, X must be measured in the same way in both experiments for U_D in $MTE(x,p)$ to be the same. If X is measured in the same way in two experiments, but one experiment over-samples on some element of X, then it should be possible to re-weight $MTE(x,p)$ so that it is comparable across experiments, as I discuss in Section 5.3.

Unobserved heterogeneity U_D is not likely to be the same in one experiment that samples at random from a broad pool as it is in another experiment that samples at random from lottery entrants from the same pool. However, the experimenter can compare MTEs from both experiments by taking a stand on the range of U_D from the broad pool that is represented by the individuals who enter the lottery. One natural assumption is that the fraction f of individuals who enter the lottery would be the individuals with the lowest net unobserved costs of treatment in the broad pool. If this assumption holds, then the MTE from $0 \leq p \leq 1$ on the sample individuals who enter the lottery should be equal to the MTE from $0 \leq p \leq f < 1$ on the broader sample. Extrapolation from the sample of individuals who enter the lottery to the broader pool requires extrapolation to potential treated fractions that exceed full treatment: $p > 1$.

If the instrument Z is the same across experiments, then the MTEs should be the same. However, there are also cases in which the instrument can differ and the MTEs can still be the same. For example, suppose that the one instrument has a strictly larger treatment incentive than another, such as free treatment vs. discounted treatment. If U_D is the same in both experiments, then the baseline treatment probability should be the same in both experiments, but the intervention treatment probability should be greater in the experiment with the larger incentive. The LATEs from both experiments need not be the same, but the MTEs should be the same. One caveat is that the estimated $MTE(p)$ might not be the same across experiments if the true $MTE(p)$ is nonlinear and the two intervention probabilities are very different, requiring more reliance on linear extrapolation.

If two instruments offer very different incentives, then the MTEs might not be the same, but

the ATEs recovered from the MTEs should still be the same. Just as U_D is a function of observed heterogeneity through X, it is also a function of the instrument through Z. Suppose that men have larger treatment effects than women and the instrument in one experiment incentivizes men to take up treatment but the instrument in another experiment incentivizes women to take up treatment. The MTE from the first experiment will be downward-sloping because as the fraction treated increases from 0 to 1, men are treated first and then women. However, the MTE from the second experiment will be upward-sloping because women select into treatment first. In this case, $MTE(p)$ will be different in both experiments. However, the amount of unexplained treatment effect heterogeneity as calculated with RMSD will be the same, and the ATEs will be the same.

5.2 Extrapolating with MTE(p)

If extrapolation is merited, then gTO, gUO, and gTE can be extrapolated to reflect different treatment probabilities using (2)-(4) by substituting hypothetical values of p_B, p_I, $s(p_B)$, and $s(p_I)$ into the formulas for the general weights $w_g(p)$ in Table 1. For example, suppose that Oregon policymakers are contemplating making more coverage available via a new intervention in which lottery winners who sign up for health insurance receive discounted, as opposed to free, coverage. Pilot tests indicate that the new intervention treatment probability will be p'_I. By extrapolating $RISTTO$ with cost as an outcome, policymakers can determine what share of individuals $s(p'_I)$ they can declare as lottery winners given available funds.

5.3 Extrapolating with MTE(x,p)

$SgTO$, $SgUO$, and $SgTE$ can be extrapolated to reflect different treatment probabilities and characteristics as follows:

$$SgTO(\mathbb{A},\mathbb{B}) = \frac{1}{\sum_{i\in\mathbb{A}} P(i\in g)} \sum_{i\in\mathbb{A}} P(i\in g) \int_0^1 \omega_g(x_{i\mathbb{A}}, p_{xi\mathbb{B}}) MTO(x_{i\mathbb{A}}, p_{xi\mathbb{B}}) dp_{xi\mathbb{B}}, \tag{28}$$

$$SgUO(\mathbb{A},\mathbb{B}) = SgSE(\mathbb{A},\mathbb{B}) = \frac{1}{\sum_{i\in\mathbb{A}} P(i\in g)} \sum_{i\in\mathbb{A}} P(i\in g) \int_0^1 \omega_g(x_{i\mathbb{A}}, p_{xi\mathbb{B}}) MUO(x_{i\mathbb{A}}, p_{xi\mathbb{B}}) dp_{xi\mathbb{B}}, \tag{29}$$

$$SgTE(\mathbb{A},\mathbb{B}) = \frac{1}{\sum_{i\in\mathbb{A}} P(i\in g)} \sum_{i\in\mathbb{A}} P(i\in g) \int_0^1 \omega_g(x_{i\mathbb{A}}, p_{xi\mathbb{B}}) MTE(x_{i\mathbb{A}}, p_{xi\mathbb{B}}) dp_{xi\mathbb{B}}, \tag{30}$$

where each individual i in sample $\mathbb{A}$ has observed covariate vector $x_{i\mathbb{A}}$. In sample $\mathbb{B}$, individuals with the same x as individual i have treatment probability $p_{xi\mathbb{B}}$. $\mathbb{A}$ and $\mathbb{B}$ can be the same sample, and either can be the actual experimental sample. If both are the experimental sample, conditioning on $\mathbb{A}$ and $\mathbb{B}$ can be suppressed, and the equations simplify to (24)-(26).

For example, the SLATE can be extrapolated to a hypothetical sample $\mathbb{A}$ that has more men

if $MTE(x,p)$ includes a covariate x for men. Other approaches have been proposed to re-weight local average treatment effects based on observed characteristics. My approach also allows for re-weighting based on unobserved characteristics that manifest themselves as different treatment probabilities for the same observed characteristics.

5.4 Decomposing Differences in Extrapolations with MTE(x,p)

Using (28)-(30), we can decompose differences between two samples $\mathbb{A}$ and $\mathbb{B}$ into differences in observables (explained) vs. unobservables (unexplained). For example, we can decompose the difference between two SLATEs in two ways as follows:

$$\frac{SLATE(\mathbb{A},\mathbb{A}) - SLATE(\mathbb{B},\mathbb{A})}{SLATE(\mathbb{A},\mathbb{A}) - SLATE(\mathbb{B},\mathbb{B})} + \frac{SLATE(\mathbb{B},\mathbb{A}) - SLATE(\mathbb{B},\mathbb{B})}{SLATE(\mathbb{A},\mathbb{A}) - SLATE(\mathbb{B},\mathbb{B})} = 1. \tag{31}$$

$$\underbrace{\frac{SLATE(\mathbb{A},\mathbb{B}) - SLATE(\mathbb{B},\mathbb{B})}{SLATE(\mathbb{A},\mathbb{A}) - SLATE(\mathbb{B},\mathbb{B})}}_{\text{explained}} + \underbrace{\frac{SLATE(\mathbb{A},\mathbb{A}) - SLATE(\mathbb{A},\mathbb{B})}{SLATE(\mathbb{A},\mathbb{A}) - SLATE(\mathbb{B},\mathbb{B})}}_{\text{unexplained}} = 1. \tag{32}$$

5.5 Extrapolation to 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. Therefore, if pre-period data are available, it is possible to estimate the MTE using the natural experiment. Equality of the MTE from the randomized and natural experiments validates the results. In the OHIE context, no individuals receive insurance in the pre-period because they must be uninsured to enter the lottery, so there are no always takers, and I cannot estimate a separate MTE using the natural experiment. However, I can use the observed change in outcomes from the pre-period to the experimental period, $Y - Y_{pre}$, to validate extrapolations from the MTE estimated with the randomized experiment. To assess the performance of extrapolations from the MTE, I can compare them to extrapolations from the LATE and the RISOLS. I can also compare them to extrapolations of the MTE from Monte Carlo simulations in which the estimated MTE is the true MTE.

6 Application: The Oregon Health Insurance Experiment

6.1 Replication

I replicate the main LATEs reported in Taubman et al. [2014] using publicly-available Oregon administrative data. I examine three measures of emergency room utilization Y: an indicator for any ER visit, a count of the number of ER visits, and a dollar amount of ER total charges.[18] All

[18]For each outcome, I run regressions on the largest set of observations for which all variables are available. Even though Taubman et al. [2014] does not include ER total charges in the main results, I examine it because it is more continuous than the other two measures of ER utilization. ER total charges (reported in the data as "total charges")

measures include individuals with zero visits. The endogenous variable D is an indicator for any Medicaid coverage, which includes Medicaid coverage obtained via the lotteried program or the main program. I refer to individuals with $D = 1$ as "treated" or "insured," and individuals with $D = 0$ as "untreated" or "uninsured."[19]

The coefficient in the first column of the top panel of Table 2, 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).[20] 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. 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.

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. 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, 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 a 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%. 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 in the full experimental sample is the same for lottery winners and losers.[21] Therefore, it is unlikely that OHIE results that do not control for the number of lottery entrants are internally valid. As noted in Taubman et al. [2014], the LATEs for

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 as a measure of resource utilization than the comparison of actual payments would be.

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

[20] I cannot replicate the result 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].

[21] Indeed, the comparison of the characteristics of lottery winners and losers yields several statistically significant differences in the full experimental sample.

Table 2: OHIE Replication and Extension

Any ER Visits					
	(1)	(2)	(3)	(4)	(5)
Medicaid	0.0697	0.0763	-0.0146	0.1816†	0.0531†
	(0.0251)**	(0.0266)***	(0.0271)	(0.0684)**	(0.0286)*
	[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
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
	(0.121)***	(0.152)***	(0.156)	(0.248)***	(0.175)
	[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
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
	($767)	($785)	($788)	($1,432)	($927)
	[$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
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 coefficients in Columns (4) and (5): ††† p<0.01, †† p<0.05, † p<0.1.

ER utilization are not robust to the removal of the control for the number of lottery entrants. As reported in Column 3, the coefficients for all three specifications of ER utilization are negative, and none are statistically different from zero.

Columns 4 and 5 report results from separate regressions for individuals in households with two lottery entrants or a single lottery entrant, respectively.[22] The results within these subsamples should be internally valid because randomization within each subsample should result in the same distribution of U_D among lottery winners and losers. The Taubman et al. [2014] approach of controlling for the number of lottery entrants could also produce internally valid results if the LATE in the full sample is globally externally valid, and thus the treatment effect is the same regardless of the number of lottery entrants. However, if the treatment effect varies across subsamples with different numbers of lottery entrants, then a linear control for the number of lottery entrants does not guarantee internal validity.

Comparison of Columns 4 and 5 provides the first evidence that the LATEs from Oregon are not globally externally valid and that the treatment effect could vary with selection. Across all measures of ER utilization, the LATE is larger for individuals in households with two lottery entrants: 18 vs. 5 percentage points, 0.7 vs. 0.3 visits, and \$878 vs. \$428. The coefficients for any visits are statistically different from each other at the 10% level. If individuals in who entered the lottery with household members had a stronger desire to gain coverage than individuals who entered alone, then this comparison provides preliminary evidence that the treatment varies with selection: the impact of insurance on ER utilization is larger for individuals more likely to gain coverage.

The sample with one lottery entrant is my preferred replication sample. Because it includes the vast majority of the full sample, it is likely to be more representative of other samples of interest. One difficulty in extrapolating 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 administrative data, so it is not possible to further restrict the sample with one lottery entrant to households with only one member.

6.2 Average Characteristics of Always Takers, Compliers, and Never Takers

The first column of Table 3 provides summary statistics on my replication sample. I report summary statistics on the observables available in the administrative data, including gender, age, selection of written materials in English, measures of pre-period utilization, enrollment in SNAP (food stamps) and TANF (welfare), and whether the individual signed up for the lottery on the first day. Columns 2 and 3 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 raise concerns about internal validity.[23] In contrast, as shown in Columns 4 and 5, the treated and untreated individuals do not

[22] I omit the small number of individuals with three lottery entrants.

[23] The coefficients on female and SNAP reject internal validity at the 10% level, but the coefficients on the predicted outcomes, which provide a higher-powered test less subject to multiple hypothesis testing, do not.

have the same average values of these covariates. The individuals who take up treatment are more likely to be female, on SNAP or TANF, and they have higher ER utilization in the pre-period. Thus, there seems to be some observable basis for selection into insurance.

Columns 6 through 9 report cross-tabulations of the data based on whether each individual won the lottery, as well as whether each individual gained Medicaid coverage. Some individuals who lost the lottery still gained coverage through the main Medicaid program, so they must be always takers. They make up 15.2% of the lottery losers, so we can infer that they make up the same percentage of the overall sample given internal validity. Therefore, $p_B = 0.152$. As shown in Column 6, always takers have a 72% probability of being female, which is much larger than the probabilities in other columns. It is likely that the main Medicaid program had more generous eligibility thresholds for women. Some women might not have been eligible for Medicaid when they signed up for the lottery, but they might have become eligible upon becoming pregnant. Other individuals might have been eligible through the main program when they entered the lottery, but they did not know about their eligibility or they had not taken steps to enroll. Emergency rooms often sign eligible individuals up for Medicaid after they incur ER charges, which could partially explain why always takers have higher pre-period utilization than other groups.

Column 9 reports statistics on never takers who lost the lottery but did not enroll in Medicaid. Entrants were not required to submit proof of eligibility to enter the lottery. However, winners were required to submit eligibility information and to meet the eligibility requirements to enroll in Medicaid. Therefore, never takers did not gain coverage, either because they did not submit their information in time or because they were not eligible. As shown, never takers had much lower pre-period ER utilization than always takers, amassing less than half of the total charges. Never takers make up 58.9% of the lottery winners, and thus the same fraction of the full sample by internal validity. Therefore, $p_I = 1 - 0.589 = 0.411$. The always takers plus the compliers make up almost three quarters of the sample of individuals who entered the lottery, so the treatment effects on them should be of interest to policymakers, but the LATE only gives the average treatment effect on the 25.9% of the sample that are compliers.

Columns 10 and 11 report the average characteristics of treated and untreated compliers, calculated from comparison of Columns 6 and 7 and the shares of always takers and never takers via (5) and (6).[24] Although the treated and untreated compliers appear to have slightly different characteristics, the t-test results for predicted outcomes reported in the second panel show that the characteristics are not statistically different, providing further evidence of internal validity.

Column 12 reports the combined characteristics of the treated and untreated compliers. 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. In the OHIE context, the comparison of covariates across Columns 1 and 12 raises minimal concerns about global external validity. However, compliers are included in 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 comparison of the characteristics of the compliers to the characteris-

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

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 Compliers)	Local Average (Untreated Compliers)	Local Average (All Compliers)
	RIS			RIST	RISU	BT	BU	IT	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
Any ER visits, pre-period	0.34	0.34	0.34	0.42	0.32	0.45	0.32	0.39	0.31	0.36	0.35	0.35
Number of ER visits, pre-period	0.87	0.86	0.87	1.18	0.77	1.36	0.78	1.05	0.73	0.87	0.88	0.88
ER total charges, pre-period	$2,440	$2,387	$2,468	$3,514	$2,099	$4,210	$2,156	$3,024	$1,942	$2,328	$2,642	$2,534
On SNAP, pre-period	0.57	0.58	0.57	0.75	0.51	0.77	0.53	0.74	0.47	0.72	0.67	0.69
On TANF, pre-period	0.02	0.03	0.02	0.07	0.01	0.09	0.01	0.05	0.01	0.02	0.01	0.01
Signed up for lottery on first day	0.09	0.10	0.09	0.12	0.09	0.10	0.09	0.13	0.07	0.15	0.13	0.13
Predicted outcomes												
Any ER visits	0.37	0.37	0.37	0.42	0.35	0.44	0.35	0.41	0.34	0.39	0.38	0.39
Number of ER visits	1.09	1.09	1.09	1.40	0.99	1.54	1.01	1.30	0.95	1.15	1.15	1.15
ER total charges	$3,935	$3,915	$3,945	$4,826	$3,652	$5,222	$3,716	$4,546	$3,475	$4,150	$4,264	$4,225
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 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

	(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	*	*	*	*
Any ER visits, pre-period	0.002	0.009	-0.045***	0.132***	-0.013	***	***	***	***
Number of ER visits, pre-period	-0.002	-0.007	-0.259***	0.579***	-0.045	***	***	***	***
ER total charges, pre-period	-$81	-$314	-$973***	$2,054***	-$214	***	***	***	***
On SNAP, pre-period	0.012*	0.045*	0.033**	0.236***	-0.063***	***	***	***	***
On TANF, pre-period	0.003	0.012	-0.049***	0.084***	0.001	***	***	***	***
Signed up for lottery on first day	0.006	0.023	0.051***	0.005	-0.016***	***	***	***	***
Predicted outcomes									
Any ER visits	0.002	0.007	-0.020**	0.089***	-0.013***	***	***	***	***
Number of ER visits	0.002	0.007	-0.186***	0.532***	-0.060**	***	***	***	***
ER total charges	-$29	-$113	-$435**	$1,506***	-$241**	***	**	***	***
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; Statistical significance was assessed using bootstrapping.

tics of always and never takers casts more doubt on global external validity than the comparison of the compliers to the full sample. At the same time, monotonicity in characteristics from always takers to compliers to never takers provides reassurance that it is reasonable to extrapolate across the groups to recover heterogeneous treatment effects.

6.3 Difference-in-Difference Test

The difference-in-difference test using covariates formalizes the comparison of the compliers to the rest of the sample. Results in the second panel of Table 3 show that some covariates are related to baseline takeup ($\lambda_D \neq 0$); some covariates are related to selection ($\lambda_Z \neq 0$); 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 statistically-significant evidence that casts doubt on global external validity. Results from the difference-in-difference test using the outcomes in lieu of the covariates reject global external validity ($\lambda_{DZ} \neq 0$) at the 10% level or better. 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 includes a nonzero selection effect. The rejection of global external validity indicates that the LATE does not apply to all individuals under the assumption that $MUO(p)$ and $MTO(p)$ are linear in p.

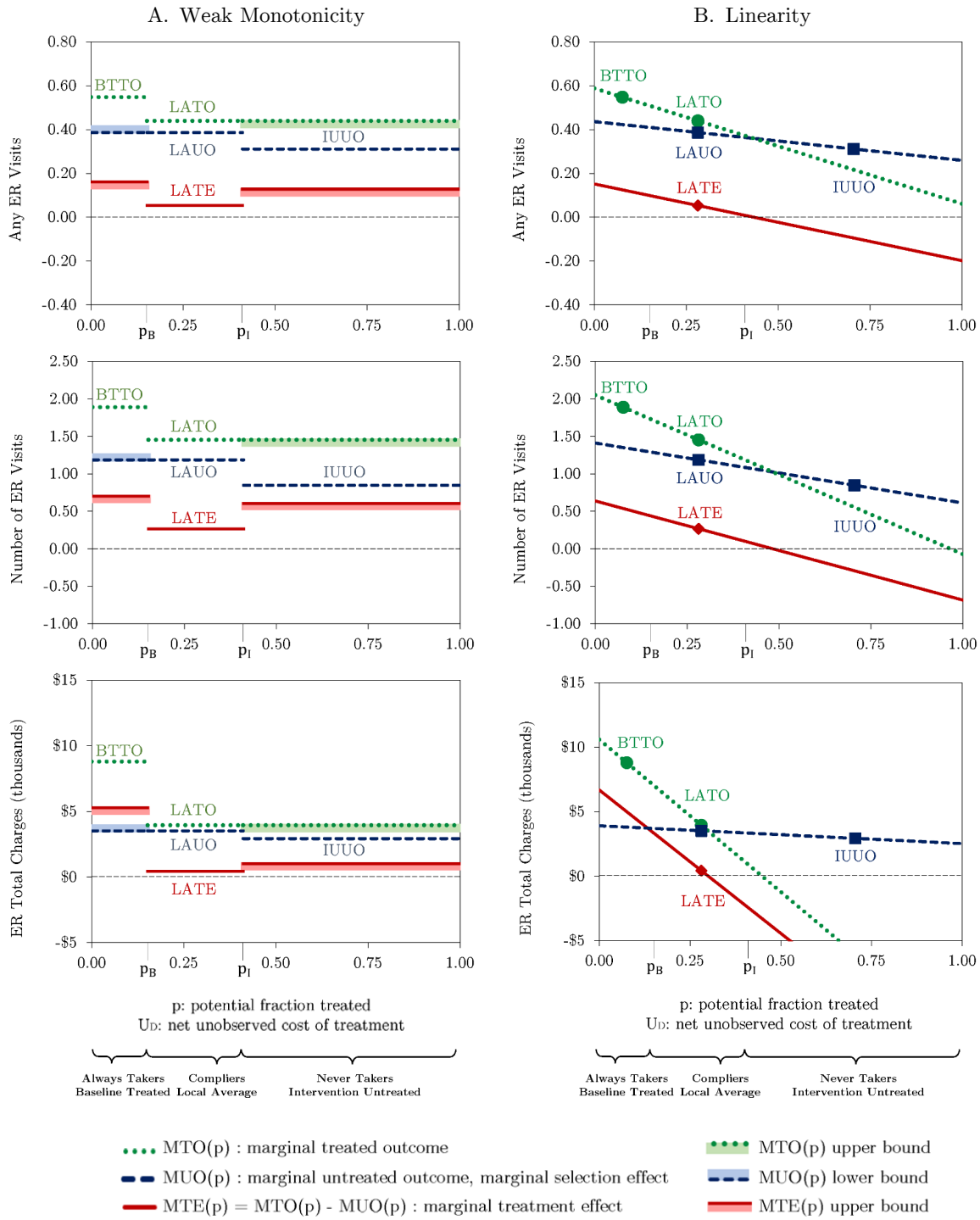
6.4 Bounds on Outcomes and Treatment Effects

The left subfigures in Figure 3 show that we cannot reject global external validity of the Oregon LATE under weak monotonicity of $MTO(p)$ and $MUO(p)$ alone. Each subfigure includes a different measure of the ER utilization outcome Y. Each plots the BTTO, LATO, LAUO, and IUUO, as reported in Table 3 over the relevant range of U_D. For number of visits, the only difference between the actual outcomes plotted in Figure 3 and the hypothetical outcomes plotted in Figure 2 is the value of the IUUO (the hypothetical value is 1.3, and the actual value is 0.85).

For all three measures of ER utilization, the untreated outcome for never takers (IUUO) is smaller than the untreated outcome for compliers (LAUO), indicating adverse selection into health insurance, instead of advantageous selection, as depicted by the hypothetical example. For all three measures of ER utilization, the treated outcome for always takers (BTTO) exceeds the treated outcome compliers (LATO), also indicating adverse selection, or a decreasing treatment effect, or both. Because the $MUO(p)$ and $MTO(p)$, both slope downward, the bounds do not reject global external validity of the LATE in this application.

However, the bounds could still be informative for ER providers in Oregon. The bounds on the BTTE imply that upon gaining insurance, always takers increase the 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, and their total charges by no more than \$5,638. The bounds on the IUTE imply that upon gaining insurance, never takers increase their average visit probability by no more than 13 percentage points ($IUTE \leq LATO - IUUO = 0.44 - 0.31$), their number of visits by no more than

Figure 3: Bounds and Estimates of $MTE(p)$



0.6, and their total charges by no more than $1,014. If Oregon decided to require everyone in the lottery sample to have insurance, the bounds imply that their average visit probability would be no more than 0.27 ($ITTO p_I + (LATO - IUUO)(1 - p_I) = 0.48(0.411) + (0.44 - 0.31)(1 - 0.411)$), their average number of visits would be no more than 1.02, and their average charges would be no more than $2,953.

6.5 Estimates of MTE(p)

The solid lines in the right subfigures of Figure 3 depict the marginal treatment effect $MTE(p)$ for each measure of ER utilization. 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, as indicated by the difference-in-difference test.

For all measures of ER utilization, $MTE(p)$ 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. As reported in the first columns of Table 4, the slope of $MTE(p)$ is statistically different from zero at the 1% level for any visits and total charges.[25] The slope is not statistically different from zero for the number of visits, but the LATE is not statistically different from zero, either. 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.

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, in the OHIE, moral hazard is greatest among the individuals who consume the most care.

For all measures of ER utilization, $MTE(p)$ 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

[25] I do not plot confidence intervals for $MUO(p)$, $MTO(p)$, and $MTE(p)$ because they can be misleading. Unlike the confidence intervals obtained from integrating each function over a range of p, which are of interest and reported in Table 4, confidence intervals at a particular value of p are not of independent interest, and they are generally much wider.

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.[26] However, the negative treatment effects for never takers are not observed in the OHIE because the never takers do not gain insurance.

Although it is plausible for never takers to have negative treatment effects (for example, they could plausibly decrease their ER utilization by substituting to primary care), it is not plausible for never takers to have negative outcomes (for example, they could not plausibly have a negative visit probability, a negative number of visits, or negative charges). As depicted in Figure 3, $MTO(p)$ and $MUO(p)$ are almost always positive when ER utilization is measured in terms of any visits or the number of visits ($MTO(p)$ for number of visits is negative for 3% of the sample). However, when ER utilization is measured in terms of ER total charges, $MTO(p)$ is negative for 45% of the sample – all never takers and some compliers. Therefore, even though it could be desirable to specify the outcome in terms of ER total charges because it is more continuous than the other measures, the linear extrapolation of ER total charges is the least plausible. Accordingly, when I report inframarginal outcomes and treatment effects, I place the least emphasis on estimates that rely on $MTO(p)$ for ER total charges in the range of U_D in which it is negative.

6.6 Inframarginal Outcomes and Treatment Effects from MTE(p)

Table 4 reports average treated outcomes, untreated outcomes, and treatment effects recovered from $MTO(p)$, $MUO(p)$, and $MTE(p)$. 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 linearity of $MTO(p)$ and $MUO(p)$. On average, always takers visit the ER with probability 0.55, they make 1.89 visits, and they 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 $MUO(p)$. The estimated BTUO shows 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 estimated BTTE shows that upon gaining insurance, always takers increase their average probability of an ER visit by 0.12, their average number of visits by 0.54, and their average total charges by \$4,944. All of these estimates are much larger than the corresponding estimates for compliers reported in Column 7, and the crosses indicate that they are statistically different.

Column 2 gives treated outcomes, untreated outcomes, and treatment effects for the baseline untreated individuals, which include never takers and untreated compliers. This group is policy-relevant because it represents the potential pool of individuals to whom coverage could be expanded after the experiment. The average untreated outcome for these individuals, BUUO, is observed, but the average treated outcome is not. Weighting the marginal treated outcome function $MTO(p)$ gives

[26] There are some never takers with positive treatment effects for any visits and number of visits because $p^* > p_I$.

Table 4: Treated Outcomes, Untreated Outcomes, and Treatment Effects in Oregon: $MTE(p)$

			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Function	Group	Baseline Treated (Always Takers)	Baseline Untreated (Never Takers and Untreated Compliers)	Intervention Treated (Always Takers and Treated Compliers)	Intervention Untreated (Never Takers)	Randomized Intervention Sample Treated	Randomized Intervention Sample Untreated	Local Average (Treated and Untreated Compliers)	Average
Intercept	Slope	g	BT	BU	IT	IU	RIST	RISU	LA	A
Any ER Visits: : RMSD = 0.10* (0.03, 0.18)**										
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)†††**	(0.18, 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)†††
(0.55, 0.63)	(-0.76, -0.30)	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	(0.36, 0.49)†††	**(0.33, 0.34)†††**	(0.35, 0.45)†††	**(0.30, 0.33)†††**	(0.36, 0.47)†††	**(0.32, 0.34)†††**	**(0.35, 0.43)**	(0.33, 0.36)†††
(0.37, 0.52)	(-0.31, -0.06)	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	(0.05, 0.19)†††	(-0.15, 0.05)†††	(0.03, 0.13)†††	(-0.23, 0.03)†††	(0.04, 0.15)†††	(-0.17, 0.04)†††	**(0.00, 0.10)**	(-0.11, 0.06)†††
(0.06, 0.23)	(-0.62, -0.11)	TE	BTTE	BUTE	ITTE	IUTE	RISTTE	RISUTE	**LATE**	ATE
Number of ER Visits: RMSD = 0.38*(0.03, 0.85)**										
MTO(p)		Treated	**1.89*** **	0.83**	**1.62*** **	0.55	**1.73*** **	0.76**	**1.45*** **	0.99***
2.05***	-2.12***	Outcome	**(1.73, 2.03)†††**	(0.26, 1.49)†††	**(1.47, 1.72)†††**	(-0.18, 1.40)†††	**(1.61, 1.81)†††**	(0.14, 1.47)†††	**(1.25, 1.66)**	(0.53, 1.54)†††
(1.81, 2.28)	(-3.48, -0.61)	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	(1.04, 1.74)†††	**(0.91, 1.00)†††**	(1.01, 1.55)†††	**(0.78, 0.92)†††**	(1.03, 1.63)†††	**(0.89, 0.97)†††**	**(1.00, 1.43)**	**(0.94, 1.11)†††**
(1.06, 1.85)	(-1.50, -0.21)	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	(0.12, 0.88)	(-0.70, 0.54)	(0.01, 0.62)	(-1.08, 0.58)	(0.07, 0.70)	(-0.79, 0.55)	**(-0.09, 0.54)**	(-0.50, 0.51)
(0.14, 1.07)	(-2.94, 0.44)	TE	BTTE	BUTE	ITTE	IUTE	RISTTE	RISUTE	**LATE**	ATE
ER Total Charges: RMSD = $6,414*($3,197, $9,667)**										
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. Statistical significance (difference from the LATO, LAUO, or LATE): ††† p<0.01, †† p<0.05, † p<0.1.
Calculation of the bold quantities does not rely on linearity of MTO(p) or MUO(p).

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 a BUTO of 0.83. The BUTE in terms of ER total charges is also negative, but I interpret it with caution because the BUTO is also negative.

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 never takers 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.

Column 7 reports the local average treatment effect LATE for comparison to the other treatment effects. Even though the LATEs in terms of number of visits and total charges are not statistically different from zero at conventional levels,[27] many of the treatment effects on other groups are, indicating that even if there is no detectable treatment effect on the compliers in an experiment, there could be detectable treatment effects on other groups of interest. As shown with crosses, the LATE is often statistically different from other treatment effects of interest, indicating treatment effect heterogeneity.

Column 8 reports ATO and AUO estimates that indicate that the average number of visits among everyone that entered the lottery would be 0.99 with insurance but 1.01 without insurance, implying a very small ATE. The ATE is so small because there are roughly as many individuals with positive treatment effects as negative treatment effects. Because heterogeneous treatment effects negate each other, the simple comparison of ATE to LATE obscures a substantial amount of treatment effect heterogeneity. The RMSD estimates at the top of each panel show that the standard deviation of the unexplained variance in ER utilization across all individuals that entered the lottery is a 10% visit probability (27% of the average probability), or 0.38 visits (34% of the average number of visits).

The ATE among all entrants is policy-relevant because it gives the treatment effect on all individuals who entered the lottery. The ATE among all entrants extrapolates the treatment effect to individuals who were not eligible for coverage, since proof of eligibility was not required to enter the lottery. As I discuss in Section OA.2, I can use eligibility information in the administrative data to recover an $MTE(p)$ and treatment effects for eligibles. The estimates are very similar to the estimates that include ineligibles, so I include the ineligibles in my preferred sample.[28]

[27] The standard errors obtained by recovering the LATE from $MTE(p)$ are exactly the same as those obtained from the regressions in Table 2.

[28] I aim for the OHIE to a be a model for other applications, and I do not want to suggest that other applications require eligibility information.

6.7 Treated Outcome Decomposition Results

In Table 5, I decompose the treated outcomes from Table 4 into selection and treatment components. Column 1 shows that selection accounts 71% of the observed number of visits among always takers. 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. The 99% confidence intervals reject one and zero for all three measures of ER utilization, indicating that the treated outcome for always takers reflects a combination of selection and treatment effects.

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. Table 4 shows that untreated compliers and never takers would have a 28% probability of visiting the ER if they had insurance, but they have a 33% probability without insurance, so the BUTE is negative. The decomposition in Table 5 shows that the predicted probability of visiting the ER with insurance would be 1.17 times higher if the treatment effect were instead zero.

I can also decompose the *difference* in outcomes between insured lottery winners (the intervention treated) and the insured lottery losers (the baseline treated). 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 it is for 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 visit probability difference, 38% of the visit number difference, and 6% 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 upon gaining coverage.[29]

[29]The comparison of the intervention treated (always takers and compliers) to the baseline treated (always takers), though policy-relevant, is not as direct as the comparison of always takers to 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 shows that health insurance increased the ER utilization of always takers by an average of 0.54 visits, and the LATE 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

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

	(1) Baseline Treated BT	(2) Baseline Untreated BU	(3) Intervention Treated IT	(4) Intervention Untreated IU	(5) Randomized Intervention RIST	(6) Randomized Intervention RISU	(7) Local Average LA	(8) Average A
Any ER Visits								
Selection	0.77***	1.17***	0.83***	1.44***	0.81***	1.23***	0.88***	1.07***
UO/TO	(0.65, 0.92)†††	(0.87, 1.81)	(0.73, 0.95)†††	(0.91, 3.67)	(0.70, 0.93)†††	(0.89, 2.08)	(0.77, 1.01)†	(0.86, 1.44)
	BTUO/**BTTO**	**BUUO**/BUTO	ITUO/**ITTO**	**IUUO**/IUTO	RISTUO/**RISTTO**	**RISUUO**/RISUTO	**LAUO**/**LATO**	AUO/ATO
Treatment Effect	0.23***	-0.17	0.17***	-0.44	0.19***	-0.23	0.12*	-0.07
TE/TO	(0.08, 0.35)†††	(-0.81, 0.13)†††	(0.05, 0.27)†††	(-2.67, 0.09)†††	(0.07, 0.30)†††	(-1.08, 0.11)†††	(-0.01, 0.23)†††	(-0.44, 0.14)†††
	BTTE/**BTTO**	BUTE/BUTO	ITTE/**ITTO**	IUTE/IUTO	RISTTE/**RISTTO**	RISUTE/RISUTO	**LATE**/**LATO**	ATE/ATO
OLS =	**0.21*** **		**0.17*** **		**0.18*** **			
TTO - UUO	**(0.19, 0.23)**		**(0.15, 0.19)**		**(0.16, 0.20)**		-	-
	BOLS = BTTO - BUUO		**IOLS = ITTO - IUUO**		**RISOLS = RISTTO - RISUUO**			
Selection	0.41***	1.23***	0.53***	1.57***	0.45***	1.34***		
(OLS - TE)/OLS	(0.14, 0.76)†††	(0.76, 1.67)	(0.20, 0.86)†††	(0.82, 2.54)	(0.16, 0.79)†††	(0.77, 1.96)		
	(**BOLS** - BTTE)/**BOLS**	(**BOLS** - BUTE)/**BOLS**	(**IOLS** - ITTE)/**IOLS**	(**IOLS** - IUTE)/**IOLS**	(**RISOLS** - RISTTE)/**RISOLS**	(**RISOLS** - RISUTE)/RISOLS	-	-
Treatment Effect	0.59***	-0.23	0.47***	-0.57	0.55***	-0.34		
TE/OLS	(0.24, 0.86)†††	(-0.67, 0.24)†††	(0.14, 0.80)†††	(-1.54, 0.18)†††	(0.21, 0.84)†††	(-0.96, 0.23)†††		
	BTTE/**BOLS**	BUTE/**BOLS**	ITTE/**IOLS**	IUTE/**IOLS**	RISTTE/**RISOLS**	RISUTE/**RISOLS**		
Number of ER Visits								
Selection	0.71***	1.15**	0.77***	1.53	0.75***	1.22**	0.82***	1.02***
UO/TO	(0.55, 0.93)†††	(0.63, 3.22)	(0.62, 0.99)††	(-15.13, 14.18)	(0.60, 0.96)††	(0.61, 5.15)	(0.66, 1.07)	(0.66, 1.92)
	BTUO/**BTTO**	**BUUO**/BUTO	ITUO/**ITTO**	**IUUO**/IUTO	RISTUO/**RISTTO**	**RISUUO**/RISUTO	**LAUO**/**LATO**	AUO/ATO
Treatment Effect	0.29***	-0.15	0.23**	-0.53	0.25**	-0.22	0.18	-0.02
TE/TO	(0.07, 0.45)†††	(-2.22, 0.37)††	(0.01, 0.38)†††	(-13.18, 16.13)	(0.04, 0.40)†††	(-4.15, 0.39)††	(-0.07, 0.34)†††	(-0.92, 0.34)†††
	BTTE/**BTTO**	BUTE/BUTO	ITTE/**ITTO**	IUTE/IUTO	RISTTE/**RISTTO**	RISUTE/RISUTO	**LATE**/**LATO**	ATE/ATO
OLS =	**0.94*** **		**0.77*** **		**0.81*** **			
TTO - UUO	**(0.78, 1.07)**		**(0.62, 0.90)**		**(0.68, 0.89)**		-	-
	BOLS = BTTO - BUUO		**IOLS = ITTO - IUUO**		**RISOLS = RISTTO - RISUUO**			
Selection	0.43***	1.13***	0.52***	1.38***	0.46***	1.21***		
(OLS - TE)/OLS	(0.11, 0.85)†††	(0.37, 1.71)	(0.16, 0.98)††	(0.36, 2.63)	(0.13, 0.90)††	(0.35, 2.06)		
	(**BOLS** - BTTE)/**BOLS**	(**BOLS** - BUTE)/**BOLS**	(**IOLS** - ITTE)/**IOLS**	(**IOLS** - IUTE)/**IOLS**	(**RISOLS** - RISTTE)/**RISOLS**	(**RISOLS** - RISUTE)/RISOLS	-	-
Treatment Effect	0.57***	-0.13	0.48**	-0.38	0.54**	-0.21		
TE/OLS	(0.15, 0.89)†††	(-0.71, 0.63)†††	(0.02, 0.84)†††	(-1.63, 0.64)†††	(0.10, 0.87)†††	(-1.06, 0.65)†††		
	BTTE/**BOLS**	BUTE/**BOLS**	ITTE/**IOLS**	IUTE/**IOLS**	RISTTE/**RISOLS**	RISUTE/**RISOLS**		
ER Total Charges								
Selection	0.43***	-1.03	0.63***	-0.48**	0.53***	-0.80	0.89***	-2.64
UO/TO	(0.24, 0.70)†††	(-14.04, 5.94)	(0.39, 0.93)†††	(-3.27, -0.20)††	(0.31, 0.80)†††	(-5.61, 5.08)	(0.55, 1.50)	(-27.00, 20.16)
	BTUO/**BTTO**	**BUUO**/BUTO	ITUO/**ITTO**	**IUUO**/IUTO	RISTUO/**RISTTO**	**RISUUO**/RISUTO	**LAUO**/**LATO**	AUO/ATO
Treatment Effect	0.57***	2.03	0.37***	1.48**	0.47***	1.80	0.11	3.64
TE/TO	(0.30, 0.76)†††	(-4.94, 15.04)	(0.07, 0.61)†††	(1.20, 4.27)††	(0.20, 0.69)†††	(-4.08, 6.61)	(-0.50, 0.45)†††	(-19.16, 28.00)
	BTTE/**BTTO**	BUTE/BUTO	ITTE/**ITTO**	IUTE/IUTO	RISTTE/**RISTTO**	RISUTE/RISUTO	**LATE**/**LATO**	ATE/ATO
OLS =	**$5,685*** **		**$2,803*** **		**$3,935*** **			
TTO - UUO	**($4,475, $6,868)**		**($2,021, $3,602)**		**($3,182, $4,623)**		-	-
	BOLS = BTTO - BUUO		**IOLS = ITTO - IUUO**		**RISOLS = RISTTO - RISUUO**			
Selection	0.12	2.08***	0.25	4.21***	0.16	2.75***		
(OLS - TE)/OLS	(-0.16, 0.49)†††	(1.31, 2.66)†††	(-0.38, 0.87)†††	(1.89, 7.68)†††	(-0.21, 0.63)†††	(1.48, 3.99)†††		
	(**BOLS** - BTTE)/**BOLS**	(**BOLS** - BUTE)/**BOLS**	(**IOLS** - ITTE)/**IOLS**	(**IOLS** - IUTE)/**IOLS**	(**RISOLS** - RISTTE)/**RISOLS**	(**RISOLS** - RISUTE)/RISOLS	-	-
Treatment Effect	0.88***	-1.08***	0.75***	-3.21***	0.84***	-1.75***		
TE/OLS	(0.51, 1.16)	(-1.66, -0.31)†††	(0.13, 1.38)	(-6.68, -0.89)†††	(0.37, 1.21)	(-2.99, -0.48)†††		
	BTTE/**BOLS**	BUTE/**BOLS**	ITTE/**IOLS**	IUTE/**IOLS**	RISTTE/**RISOLS**	RISUTE/**RISOLS**		

*** p<0.01, ** p<0.05, * p<0.1. Bootstrapped 95% confidence intervals in parentheses. Statistical significance (difference from the LATE): ††† p<0.01, †† p<0.05, † p<0.1 (only indicated for the treatment effects).
Calculation of the bold quantities does not rely on linearity of MTO(p) or MUO(p).

6.8 OLS Decomposition Results

As shown in Table 5, the BOLS, IOLS, RISOLS estimates are positive for all measures of ER utilization, indicating that insured individuals have higher ER utilization than uninsured individuals whether they won the lottery or not. For all three measures of ER utilization, BOLS exceeds IOLS, indicating that the treatment effect decreases with the potential treated fraction, 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 LATE by RISOLS, then we would conclude that the treatment effect is responsible for only 27% (0.05/0.18) of RISOLS for any visits, 33% of RISOLS for number of visits, and 11% of RISOLS for total charges.

If we allow for a heterogeneous treatment effect by dividing RISTTE by RISOLS, we see that the treatment effect actually has a greater role. The treatment effect is responsible for 55% (0.10/0.18) of RISOLS for any visits, 54% of RISOLS for number of visits, and 83% of RISOLS for total charges. The comparison of LATE to RISOLS understates the role of the treatment effect and overstates the role of the selection effect because it does not acknowledge that treatment effects for always takers (which is included in the RISTTE but not the LATE) are larger than the treatment effects for compliers.

6.9 Subgroup Analysis Results from LATE and MTE(p)

Given the large differences in pre-period ER utilization observed between always takers, compliers, and never takers in Table 3, Table 6 divides the sample into subgroups according to pre-period ER utilization. Tables OA3-OA8 divide the sample according to the other covariates. The first row reports the LATE in each subgroup. Across all tables, almost all of the LATEs are positive. Therefore, is unlikely that traditional LATE re-weighting methods that rely only on observed heterogeneity can explain why some health insurance expansions could decrease ER utilization.

However, it can be misleading to compare or re-weight LATEs across subgroups when those LATEs are not globally externally valid. The slope of $MTE(p)$ is often statistically different from zero, indicating that there is often unexplained treatment effect heterogeneity *within* a subgroup. Furthermore, as shown in the second row, the baseline and intervention treatment probabilities p_B and p_I vary across subgroups, so the LATEs will not be comparable.

Comparison of Columns 1 and 2 show that though there is substantial treatment effect heterogeneity in the full sample, there is very little treatment effect heterogeneity among individuals that visited the ER in the pre-period. In this subgroup, the LATE indicates that insurance increases the probability of a visit by 7 percentage points, and the RMSD estimate indicates that the standard deviation of the unexplained variance in the treatment effect is only 0.3 percentage points. In contrast, in the full sample, the RMSD is 10 percentage points.

When $MTE(p)$ slopes downward, p^* gives the share of the sample with a positive treatment effect. In most subgroups reported in Tables OA3-OA8, $MTE(p)$ predicts that less than half of

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.

Table 6: LATE and MTE(p) Subgroup Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Any ER Visits			**Number of ER Visits**			**ER Total Charges**	
	Full sample	Any ER visits in the pre-period	No ER visits in the pre-period	Full sample	>1 ER visits in the pre-period	≤ 1 ER visit in the pre-period	Full sample	Top 10% of ER total charges in the pre-period	Bottom 90% of ER total charges in the pre-period
LATE	0.05*	0.07*	0.04	0.27	0.29	0.32***	$428	$5,778	$27
	(0.00, 0.10)	(0.00, 0.16)	(-0.02, 0.10)	(-0.09, 0.54)	(-0.92, 1.43)	(0.08, 0.49)	(-$1,436, $2,142)	(-$9,128, $18,514)	(-$1,294, $1,302)
vs. full sample	-			-			-		
vs. complementary sample	-			-			-		
$\mathbf{p_B}$	0.15***	0.20***	0.13***	0.15***	0.22***	0.14***	0.15***	0.23***	0.14***
	(0.15, 0.16)	(0.19, 0.21)	(0.12, 0.13)	(0.15, 0.16)	(0.21, 0.24)	(0.13, 0.14)	(0.15, 0.16)	(0.21, 0.26)	(0.14, 0.15)
vs. full sample	-	***	***	-	***	***	-	***	***
vs. complementary sample	-	***		-	***		-	***	
$\mathbf{p_I}$	0.41***	0.47***	0.38***	0.41***	0.49***	0.39***	0.41***	0.51***	0.40***
	(0.40, 0.42)	(0.45, 0.49)	(0.36, 0.39)	(0.40, 0.42)	(0.47, 0.52)	(0.38, 0.41)	(0.40, 0.42)	(0.47, 0.55)	(0.39, 0.41)
vs. full sample	-	***	***	-	***	***	-	***	***
vs. complementary sample	-	***		-	***		-	***	
MTE(p) intercept	0.15***	0.07	0.13***	0.64***	0.11	0.45***	$6,677***	$25,628***	$3,353***
	(0.06, 0.23)	(-0.11, 0.21)	(0.04, 0.23)	(0.14, 1.07)	(-1.90, 2.33)	(0.15, 0.76)	($3,555, $9,326)	($3,776, $47,466)	($1,121, $5,175)
vs. full sample	-			-			-	*	***
vs. complementary sample	-			-			-	*	
MTE(p) slope	-0.35***	0.01	-0.38***	-1.32	0.51	-0.49	-$22,218***	-$53,606*	-$12,262***
	(-0.62, -0.11)	(-0.34, 0.42)	(-0.68, -0.08)	(-2.94, 0.44)	(-5.38, 5.76)	(-1.51, 0.63)	(-$33,486, -$11,076)	(-$105,584, $6,817)	(-$19,281, -$3,406)
vs. full sample	-	***		-			-		***
vs. complementary sample	-	*		-			-		
p*	0.43***	-6.13	0.35***	0.48	-0.21	0.92	0.30***	0.48*	0.27***
	(0.27, 0.97)	(-11.83, 5.97)	(0.19, 1.01)	(-0.92, 2.26)	(-4.10, 5.77)	(-5.09, 7.27)	(0.22, 0.45)	(-0.77, 1.71)	(0.16, 0.49)
vs. full sample	-			-			-		
vs. complementary sample	-			-			-		
RMSD	0.10***	0.003***	0.11***	0.38***	0.15***	0.14***	$6,414***	$15,475***	$3,540***
	(0.03, 0.18)	(0.001, 0.13)	(0.02, 0.19)	(0.03, 0.85)	(0.02, 1.81)	(0.01, 0.44)	($3,197, $9,667)	($1,359, $30,479)	($983, $5,566)
N	19,643	6,709	12,934	19,622	3,405	16,210	19,628	1,962	17,657

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

Individuals with missing values for the corresponding pre-utilization measure were not included in the subgroups.

individuals have a positive treatment effect. However, pre-period utilization is a strong predictor of positive treatment effects. The $MTE(p)$ reported in Column 2 indicates that all individuals with ER visits in the pre-period have positive treatment effects, but the $MTE(p)$ reported in Column 2 indicates that only 35% of the remaining individuals have positive treatment effects.

The traditional instrumental variable regression model with covariates imposes that the treatment effect is the same regardless of observed heterogeneity. As reported in Table OA2, the instrumental variable estimates that incorporate covariates yield treatment effect estimates that are the same for all subgroups. The instrumental variable regression model can allow the treatment effect to vary with observed heterogeneity if it includes interactions between the covariates and the endogenous variable. $MTE(x,p)$ allows the treatment effect to vary with observed and unobserved heterogeneity.

6.10 Estimates of MTE(x,p)

Figure 4 depicts the linear[30] $SMTE(p)$ estimated with all covariates.[31] For all three measures of ER utilization, the $SMTE(p)$ estimated with all covariates is less steep than the $MTE(p)$ reported with a solid line, indicating that covariates have a role in explaining why the treatment effect varies with the quantile of the sample that selects into treatment.[32] The $SMTE(p)$ for the number of ER visits is almost horizontal, indicating that almost all unobserved heterogeneity is explained.

Table 7 shows that when ER utilization is measured in terms of the number of ER visits, the inclusion of all covariates decreases the standard deviation of unexplained heterogeneity from $RMSD(X_0) = 0.38$ visits to $RMSD(X_4) = 0.07$ visits, so 83% of $RMSD(X_0)$ is explained. For the other measures of ER utilization, covariates have less predictive power, as demonstrated by the comparison of $MTE(p)$ and $SMTE(p)$. The inclusion of all covariates explains 28% of the variation in the treatment effect in terms of any visits but only 8% in terms of total charges.

Next, I attempt to understand which covariates are most important for explaining heterogeneity in the treatment effect. Using a dotted line, in the right panel of Figure 4, I report $SMTE(p)$ obtained from only the "common covariates" that are also available in the Behavioral Risk Factor Surveillance System (BRFSS), on the grounds that these covariates are most likely to be available

[30] In Figure OA3, I report robustness of $MTE(x,p)$ to the order of the global polynomial by plotting the estimated quadratic and cubic $SMTE(p)$ using all covariates. The linear, quadratic, and cubic polynomials all depict meaningful treatment effect heterogeneity. All three polynomials generally decrease as the fraction treated increases, but cubic and higher order polynomials vary widely, especially at high ranges of U_D, because extrapolations rapidly approach positive or negative infinity in regions where there is no data.

[31] "All covariates" include: "common covariates:" female, English, binary variables for each year of age, as well as all two-way interaction terms; "pre-period utilization:" a binary variable for any ER visits in the pre-period and a continuous variable for ER total charges in the pre-period (included in all specifications, regardless of outcome); binary variables for SNAP and TANF enrollment in the pre-period, and a binary variable for whether the individual signed up for the lottery on the first day. Section OA.3 discusses the propensity scores predicted with these covariates.

[32] It is readily apparent from these results that MTE methods provide an alternative approach to conditional quantile IV estimation. Victor Chernozhukov, Ivan Fernandez-Val, and I developed a censored quantile instrumental variable (CQIV) estimator (Chernozhukov et al. [2015]) to examine variation in moral hazard across individuals. The results showed limited variation in moral hazard across the conditional quantiles of the expenditure distribution (Kowalski [2016]), which is consistent with the results from $SMTE(p)$. However, unlike the results from $MTE(p)$, they offered limited information about variation across the unconditional quantiles of the expenditure distribution, especially since the CQIV algorithm requires covariates.

Figure 4: $MTE(x, p)$

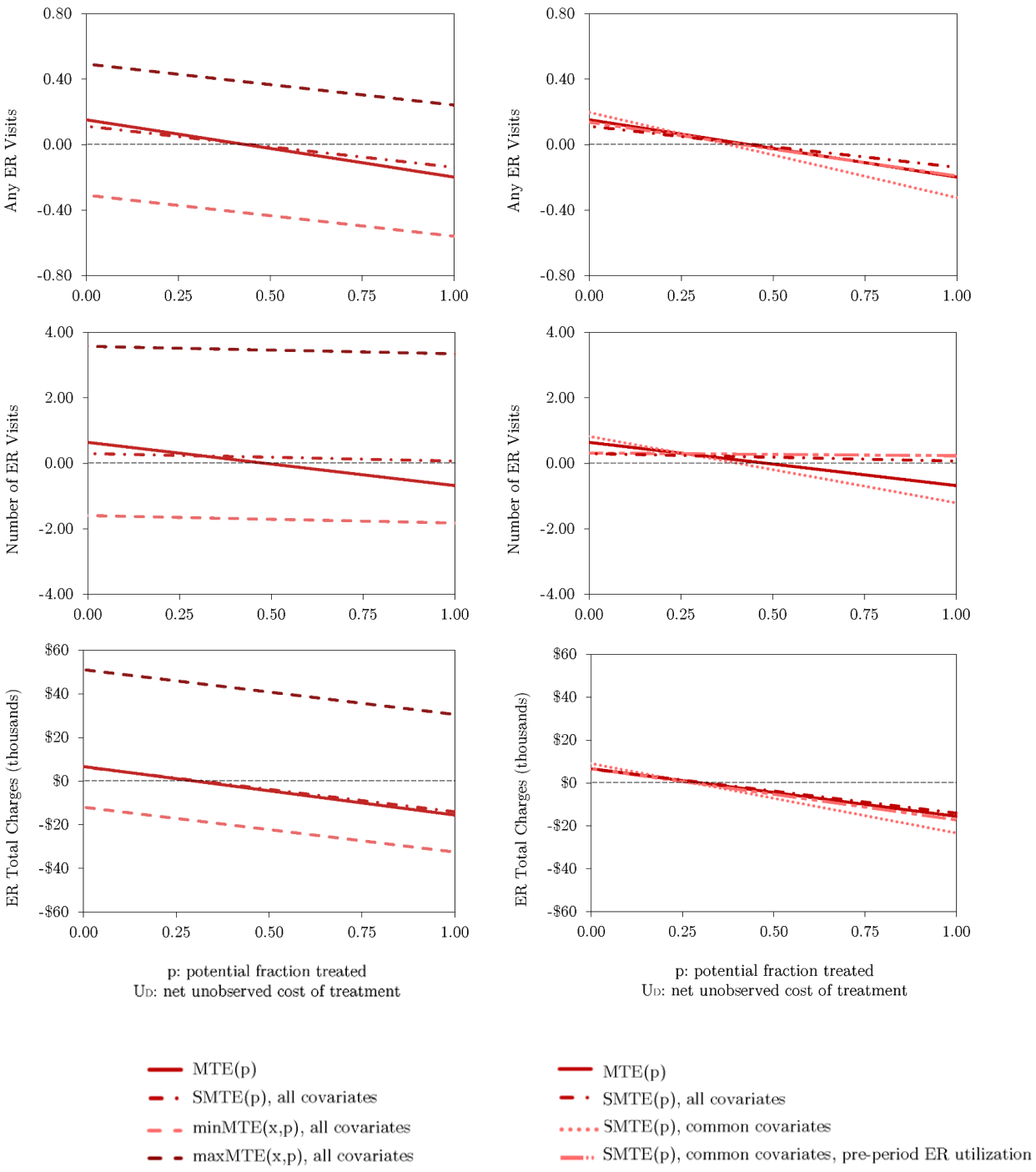


Table 7: Explained and Unexplained Treatment Effect Heterogeneity

	RMSD(X_c)	Explained $\frac{RMSD(X_0)-RMSD(X_c)}{RMSD(X_0)}$	Unexplained $\frac{RMSD(X_c)}{RMSD(X_0)}$	X_c Common covariates	X_c Pre-period ER utilization	X_c All covariates
			Any ER Visits			
X_0	0.10*** (0.03, 0.18)	0.00 (0.00, 0.00)†††	1.00*** (1.00, 1.00)			
X_1	0.15*** (0.07, 0.23)	-0.49*** (-1.76, -0.06)†††	1.49*** (1.06, 2.76)†††	X		
X_2	0.09*** (0.02, 0.17)	0.06 (-0.52, 0.53)†††	0.94*** (0.47, 1.52)	X	X	
X_3	0.07*** (0.01, 0.15)	0.28 (-0.18, 0.77)†††	0.72*** (0.23, 1.18)	X	X	X
			Number of ER Visits			
X_0	0.38*** (0.03, 0.85)	0.00 (0.00, 0.00)†††	1.00*** (1.00, 1.00)			
X_1	0.59*** (0.08, 1.00)	-0.54 (-7.79, 0.49)†††	1.54*** (0.51, 8.79)	X		
X_2	0.02*** (0.01, 0.49)	0.94 (-8.99, 0.98)†††	0.06*** (0.02, 9.99)	X	X	
X_3	0.07*** (0.01, 0.43)	0.83 (-6.96, 0.99)†††	0.17*** (0.01, 7.96)	X	X	X
			ER Total Charges			
X_0	\$6,414*** (\$3,197, \$9,667)	0.00 (0.00, 0.00)†††	1.00*** (1.00, 1.00)			
X_1	\$9,351*** (\$5,325, \$12,821)	-0.46*** (-0.87, -0.12)†††	1.46*** (1.12, 1.87)†††	X		
X_2	\$6,884*** (\$3,045, \$9,784)	-0.07 (-0.42, 0.36)†††	1.07*** (0.64, 1.42)	X	X	
X_3	\$5,930*** (\$2,676, \$8,805)	0.08 (-0.26, 0.42)†††	0.92*** (0.58, 1.26)	X	X	X

Statistical significance (difference from 0): *** p<0.01, ** p<0.05, * p<0.1. Statistical significance (difference from 1): ††† p<0.01, †† p<0.05, † p<0.1 (only indicated for the decompositions).

for other samples of interest. For all three measures of ER utilization, the inclusion of these covariates increases unexplained variation in the treatment effect, making $SMTE(p)$ steeper, and increasing $RMSD(X_1)$ by 50% relative to $RMSD(X_0)$, as reported in Table 7.[33]

The inclusion of measures of pre-period utilization substantially decreases the unexplained variation in the treatment effect, as shown in Figure 4 and quantified with $RMSD(X_2)$. The inclusion of other covariates less likely to be available in other contexts, SNAP enrollment, TANF enrollment, and whether the individual signed up on the first day, explain only slightly more variation in the treatment effect for any visits and total charges, and they reduce the explained variation for the number of ER visits.

I next explore which observable groups have the smallest and largest treatment effects. As shown in the left of Figure 4, for the number of visits, $\min MTE(x,p)$ represents $MTE(x,p)$ for 54 year-old males who do not request materials in English, had no ER visits in the pre-period, were not on SNAP or TANF, and did not sign up on the first day of the lottery. For this group, $MTE(x,p)$ is negative for all $0 \leq p \leq 1$, indicating negative treatment effects for all individuals in this group that entered the lottery. The corresponding $\max MTE(x,p)$ indicates positive treatment effects for 42 year-old males who requested materials in English, had 17 ER visits and \$15,759 in charges in the pre-period, were on SNAP but not on TANF, and did not sign up on the first day of the lottery. The predicted visit difference between individuals with the two covariate vectors is almost six visits, which is very large relative to the predicted visit difference from the individual with the lowest to the individual with the highest unobserved net cost of treatment U_D within any covariate vector.

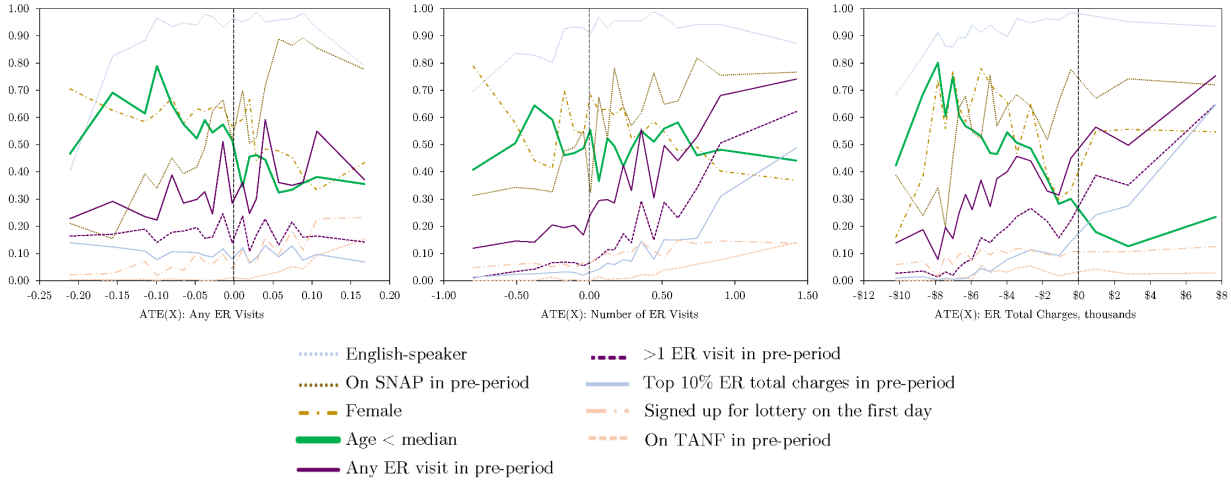
The estimated $MTE(x,p)$ can be thought of as a calculator that produces an estimate of the treatment effect for an individual with covariate vector x who signs up for treatment when the fraction treated within the group is p. In practice, it might be difficult to determine what value of p to input into the calculator for a given individual, so it might be preferable to develop a calculator from $ATE(x)$, which integrates over all p, to provide general guidelines on the impact of treatment across different demographic groups. If covariates explain all of the unobserved heterogeneity, then $MTE(x,p)$=$ATE(x)$, so no information is lost by integrating over p.

Figure 5 presents statistics on $ATE(x)$ predicted for every individual in the sample. The horizontal axis groups individuals into 20 bins with the same number of individuals in them, so that the bins represent the ventiles of $ATE(x)$. The vertical line indicates where $ATE(x)$ goes from negative to positive. The other lines give average covariate values for individuals at each ventile of $ATE(x)$. The thickest solid line shows that individuals with ages below the median age are more prevalent among the groups with higher values of $ATE(x)$. The solid line of medium thickness shows that individuals with a pre-period visit to the ER in increase their ER utilization more than other individuals when they gain insurance.

Table OA9 reports the SATE, the sample average of $ATE(x)$, as well as all of the other sam-

[33]To understand why the inclusion of covariates can increase unexplained variation, consider a simple example in which the young have larger treatment effects than the old and always takers have larger treatment effects than never takers. However, always takers are more likely to be old. In this example, inclusion of a covariate for age can increase unexplained variation in the treatment effect.

Figure 5: Average Observables Across Ventiles of $ATE(x)$



ple treatment effects. $SgTE$ need not equal the corresponding gTE. For example, SLATE is an estimate of the treatment effect for compliers that incorporates observables, and LATE is an estimate of the treatment effect for compliers that does not incorporate observables. If compliers are observably different from other groups, then the SLATE need not equal the LATE, even if there is no remaining unobserved heterogeneity.[34]

6.11 Extrapolating to the Natural Experiment

Before validating $MTE(p)$ using the natural experiment that took place from the pre-period to the experimental period, I perform a Monte Carlo exercise to benchmark how well it should perform in absolute terms and relative to the LATE and the RISOLS in my application. I discuss the implementation of the Monte Carlo exercise in Section OA.4. The first two columns of Table 8 report how well each estimator performs in the Monte Carlo designed to simulate the randomized experiment. In Column 1, the true treatment effect θ is equal to the estimated LATE from the OHIE. The LATE estimator generally has the smallest mean bias and RMSE, but the $MTE(p)$ estimator performs very similarly, especially as compared to the RISOLS estimator, which always performs the worst. For example, the LATE estimator under-predicts total charges by \$1.10, the $MTE(p)$ estimator under-predicts total charges by \$1.30, and the RISOLS estimator over-predicts total charges by \$634. In contrast, in Column 2, when the MTE(p) from the OHIE is the true θ, the $MTE(p)$ estimator substantially out-performs the other estimators on both dimensions.

[34]Even the gTO and gUO that are observed and reported in bold in Table 1 need not be equal to the corresponding $SgTO$ and $SgUO$. As discussed in Section OA.1, the global polynomial estimation algorithm for $MTE(x,p)$ estimates two separate regressions: one for the randomized intervention sample treated and another for the randomized intervention sample untreated. Therefore, the observed BTTO can be different from the predicted SBTTO because the prediction includes all always takers, including those that win the lottery, but the observed BTTO includes only the always takers that lose the lottery.

Table 8: Validation Exercises

	(1)		(2)		(3)		(4)		(5)	
			Simulated data						Observed data	
	θ = LATE		**θ = MTE(p)**		**D*θ = D*LATE**		**D*θ = D*MTE(p)**		**D*θ = D*(Y-Y_{pre})**	
	Randomized experiment		Randomized experiment		Natural experiment		Natural experiment		Natural experiment	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Any ER Visits										
MTE(p)	-0.00016 [1]	0.003 [2]	-0.00039 [1]	0.072 [1]	-0.00004 [2]	0.001 [2]	-0.00004 [1]	0.016 [1]	0.00214 [1]	0.274 [1]
LATE	-0.00016 [2]	0.002 [1]	0.07611 [2]	0.126 [2]	-0.00004 [1]	0.001 [1]	-0.01096 [2]	0.03 [2]	-0.00878 [2]	0.274 [2]
RISOLS	0.08122 [3]	0.081 [3]	0.20282 [3]	0.227 [3]	0.01958 [3]	0.04 [3]	0.01958 [3]	0.044 [3]	0.02172 [3]	0.277 [3]
Number of ER Visits										
MTE(p)	-0.00051 [2]	0.012 [2]	-0.00131 [1]	0.271 [1]	-0.00003 [1]	0.006 [2]	-0.00003 [1]	0.062 [1]	-0.02925 [1]	1.29 [1]
LATE	-0.00029 [1]	0.011 [1]	0.28857 [2]	0.479 [2]	-0.00006 [2]	0.005 [1]	-0.04135 [2]	0.112 [2]	-0.07185 [3]	1.296 [3]
RISOLS	0.36696 [3]	0.367 [3]	0.82727 [3]	0.911 [3]	0.08837 [3]	0.18 [3]	0.08837 [3]	0.195 [3]	0.05499 [2]	1.293 [2]
ER Total Charges										
MTE(p)	-$1.3 [2]	$20.7 [2]	-$17.9 [1]	$4,534.7 [1]	-$0.2 [1]	$10.6 [2]	-$0.2 [1]	$1,037.4 [1]	-$34.6 [1]	$10,984.8 [1]
LATE	-$1.1 [1]	$18.8 [1]	$4,853.8 [2]	$8,042.9 [2]	-$0.2 [2]	$9.2 [1]	-$692.9 [3]	$1,886.1 [3]	-$725.8 [3]	$11,091.9 [3]
RISOLS	$634.0 [3]	$634.0 [3]	$8,366.0 [3]	$10,540.9 [3]	$152.6 [3]	$311.0 [3]	$152.6 [2]	$1,287.8 [2]	$99.5 [2]	$10,994.6 [2]

Rankings for bias, in brackets, are based on absolute value.

Columns 3 and 4 report how well each estimator performs in the Monte Carlo designed to simulate the natural experiment. In the natural experiment, the observed change in outcomes, $Y - Y_{pre}$, should only reflect the treatment effect for individuals who gain coverage (the always takers and compliers); it should be zero otherwise. Therefore, I examine the ability of each estimator to recover the average *observed* treatment effect across all observations, $E[D\theta]$. The performance of the estimators is similar. On the whole, the Monte Carlo results suggest that extrapolating based on $MTE(p)$ sacrifices a small amount of efficiency when the true treatment effect is equal to the LATE, but it has huge gains when the true treatment effect is equal to $MTE(p)$.

Next, I turn to validating the MTE results using the observed data from the natural experiment. 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.[35]. Therefore, I continue with the validation exercise to demonstrate its application, but I interpret the findings with caution.

In Column 5, the bias and RMSE should be directly comparable to the bias and RMSE from the Monte Carlo exercises in Columns 3 or 4 if either LATE or $MTE(p)$ give the true treatment effect. In practice, the bias and RMSE in the observed data are much larger.[36] While I interpret the results with caution because of measurement of Y_{pre}, the $MTE(p)$ estimator out-performs the other estimators.

[35]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.

[36]One explanation is that Y_{pre} is not directly comparable to Y for the reasons discussed above; another is that the true treatment effect is not equal to the LATE or $MTE(p)$; a third is that something changed from before to after the experiment such that even individuals who did not gain coverage changed their ER utilization.

6.12 Extrapolating to Massachusetts

Before extrapolating from the Oregon Health Insurance Experiment to the Massachusetts health reform, I acknowledge that there are several factors that could have differed between the 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 individuals and the Massachusetts reform was a state-wide policy. Therefore, Oregon impacts likely occurred through the demand-side, but Massachusetts impacts could also have occurred through the supply-side.

Furthermore, institutional features of the health care environment could have differed across states. As discussed by Miller [2012], Massachusetts had an uncompensated care pool that might have encouraged excess emergency care before its dissolution and replacement under the Massachusetts reform. Also, both states could have had different social norms regarding emergency room vs. primary care utilization.[37] Health insurance terms could also have differed, especially since Oregon expanded Medicaid alone and Massachusetts also expanded other types of coverage.

For extrapolation, I make the important assumption that D, which represents Medicaid in Oregon, can be extrapolated to all types of health insurance in Massachusetts. I define the Massachusetts instrument Z to indicate individuals in Massachusetts after the reform. The resulting Massachusetts baseline and intervention treatment probabilities, $p_B = 0.905$ and $p_I = 0.956$, are both very high relative to the Oregon $p_B = 0.152$ and $p_I = 0.411$.[38] However, the relevant treatment probabilities for extrapolation to Massachusetts could be even higher.

It is likely that individuals who entered the Oregon lottery for health insurance had a stronger desire to obtain coverage than individuals who obtained coverage after the Massachusetts health reform to avoid paying a penalty. Therefore, the relevant p_B and p_I for extrapolation to Massachusetts from Oregon could exceed 1. I proceed under the conservative assumption that the distribution of unobserved heterogeneity U_D is the same in the Oregon and Massachusetts samples so that p_B and p_I from Oregon and Massachusetts are comparable.

In Column 1 of the first row of the left panels of Table 9, I report the Oregon LATE for each measure of ER utilization for reference. All LATEs are positive. Applying Massachusetts weights to $MTE(p)$ from Oregon in Column 2, I find negative LATEs for all three measures of ER utilization. These extrapolations to Massachusetts imply that insurance should decrease the visit probability by 0.17, the number of visits by 0.58, and charges by \$13,797.[39] The Miller [2012] examination of the Massachusetts reform finds that insurance decreases the number of visits by 0.5 visits per year, which is squarely in the range of my extrapolations. Therefore, my extrapolations potentially reconcile the Oregon and Massachusetts results using only variation in unobservables.

Next, I examine the impact of observables on the extrapolations. I obtain observable characteristics of the Massachusetts population with and without insurance before and after the reform using

[37] Those social norms could have differed between urban and rural areas. Though Massachusetts is more urban than Oregon, the Oregon administrative data on ER utilization are only from the Portland area.

[38] I obtain these probabilities from the BRFSS, as summarized in Table OA1.

[39] The Massachusetts LATEs for any visits and total charges are statistically different from zero and from the Oregon LATEs at the 10% level or better. The Massachusetts LATE for the number of visits is not statistically significant, but neither is the underlying Oregon LATE.

Table 9: Extrapolation from Oregon to Massachusetts

A. Comparison of LATE and SLATE	(1)	(2)	(1) - (2)
Any ER Visits			
LATE(·)	0.05*	-0.17*	0.22***
	(0.00, 0.10)	(-0.36, 0.00)	(0.07, 0.40)
	(OR)	(MA)	(OR) - (MA)
SLATE(· , ·)	0.05	-0.28***	0.33***
	(-0.01, 0.11)	(-0.47, -0.09)	(0.16, 0.50)
	(OR, OR)	(MA, MA)	(OR, OR) - (MA, MA)
Number of ER Visits			
LATE(·)	0.27	-0.58	0.85
	(-0.09, 0.54)	(-1.71, 0.66)	(-0.28, 1.88)
	(OR)	(MA)	(OR) - (MA)
SLATE(· , ·)	0.26	-1.03	1.28**
	(-0.08, 0.54)	(-2.03, 0.27)	(0.09, 2.24)
	(OR, OR)	(MA, MA)	(OR, OR) - (MA, MA)
ER Total Charges			
LATE(·)	$428	-$13,797***	$14,225***
	(-$1,436, $2,142)	(-$22,256, -$5,843)	($7,070, $21,525)
	(OR)	(MA)	(OR) - (MA)
SLATE(· , ·)	$40	-$20,832***	$20,872***
	(-$2,506, $2,612)	(-$29,948, -$10,359)	($11,537, $28,828)
	(OR, OR)	(MA, MA)	(OR, OR) - (MA, MA)

B. SLATE Decompositions	(3) SLATE(· , OR)	(4) SLATE(· , MA)	(5) Unexplained
Any ER Visits			
SLATE(OR, ·)	0.05	-0.29***	1.04***
	(-0.01, 0.11)	(-0.48, -0.09)	(0.99, 1.09)
	(OR, OR)	(OR, MA)	$\frac{\text{(OR, OR)} - \text{(OR, MA)}}{\text{(OR, OR)} - \text{(MA, MA)}}$
SLATE(MA, ·)	0.05	-0.28***	1.00***
	(-0.01, 0.12)	(-0.47, -0.09)	(0.94, 1.06)
	(MA, OR)	(MA, MA)	$\frac{\text{(MA, OR)} - \text{(MA, MA)}}{\text{(OR, OR)} - \text{(MA, MA)}}$
Explained	-0.005	-0.04*	
	(-0.06, 0.06)†††	(-0.09, 0.01)†††	
	$\frac{\text{(OR, OR)} - \text{(MA, OR)}}{\text{(OR, OR)} - \text{(MA, MA)}}$	$\frac{\text{(OR, OR)} - \text{(MA, MA)}}{\text{(OR, OR)} - \text{(MA, MA)}}$	
Number of ER Visits			
SLATE(OR, ·)	0.26	-1.07	1.04***
	(-0.08, 0.54)	(-2.13, 0.26)	(0.90, 1.23)
	(OR, OR)	(OR, MA)	$\frac{\text{(OR, OR)} - \text{(OR, MA)}}{\text{(OR, OR)} - \text{(MA, MA)}}$
SLATE(MA, ·)	0.25	-1.03	0.99***
	(-0.09, 0.55)	(-2.03, 0.27)	(0.76, 1.11)
	(MA, OR)	(MA, MA)	$\frac{\text{(MA, OR)} - \text{(MA, MA)}}{\text{(OR, OR)} - \text{(MA, MA)}}$
Explained	0.01	-0.04	
	(-0.11, 0.24)†††	(-0.23, 0.10)†††	
	$\frac{\text{(OR, OR)} - \text{(MA, OR)}}{\text{(OR, OR)} - \text{(MA, MA)}}$	$\frac{\text{(OR, OR)} - \text{(MA, MA)}}{\text{(OR, OR)} - \text{(MA, MA)}}$	
ER Total Charges			
SLATE(OR, ·)	$40	-$21,324***	1.02***
	(-$2,506, $2,612)	(-$30,698, -$10,789)	(1.00, 1.06)
	(OR, OR)	(OR, MA)	$\frac{\text{(OR, OR)} - \text{(OR, MA)}}{\text{(OR, OR)} - \text{(MA, MA)}}$
SLATE(MA, ·)	-$193	-$20,832***	0.99***
	(-$2,770, $2,143)	(-$29,948, -$10,359)	(0.95, 1.01)
	(MA, OR)	(MA, MA)	$\frac{\text{(MA, OR)} - \text{(MA, MA)}}{\text{(OR, OR)} - \text{(MA, MA)}}$
Explained	0.01	-0.02**	
	(-0.01, 0.05)†††	(-0.06, -0.0001)†††	
	$\frac{\text{(OR, OR)} - \text{(MA, OR)}}{\text{(OR, OR)} - \text{(MA, MA)}}$	$\frac{\text{(OR, OR)} - \text{(MA, MA)}}{\text{(OR, OR)} - \text{(MA, MA)}}$	

Sources: Oregon Administrative Data, 1 lottery entrant in household and Behavioral Risk Factor Surveillance System 2004-2009, Massachusetts data
Statistical significance (difference from 0): *** p<0.01, ** p<0.05, * p<0.1. Statistical significance (difference from 1): ††† p<0.01, †† p<0.05, † p<0.1 (only indicated for the decompositions).

the BRFSS data that I used in Kolstad and Kowalski [2012].[40] The only three covariates that are defined consistently in the Massachusetts and Oregon data are gender, age, and whether written materials were requested in English. I compare common covariates from the Massachusetts and Oregon data in Table OA1. The Oregon never takers and the Massachusetts compliers do not necessarily have similar observables, even though the extrapolation of the LATE suggested that their unobservables should be similar. However, as discussed in Section 6.10, these covariates increase unexplained treatment effect heterogeneity in Oregon, so it is not surprising that these covariates do not explain differences between the Massachusetts and Oregon samples. Indeed, quantitative

[40]Data from the other published studies that examine the impact of the Massachusetts health reform are not available at the individual level, or they only include individuals who visit a hospital or emergency room, making them unsuitable for this exercise. The BRFSS data do not include any measures of emergency room utilization, so I cannot estimate a separate $MTE(p)$ or $MTE(x, p)$ in Massachusetts.

evidence in Table 9 confirms that these observables do not explain the differences between the samples.[41]

My methods allow for extrapolation based on unobservables, which is empirically important in extrapolations from Oregon to Massachusetts. The SLATEs that extrapolate using only Massachusetts observables, reported in Column 3, are almost always positive. Therefore, existing sample re-weighting methods that incorporate only available observables would not likely yield negative treatment effects in Massachusetts. However, the SLATEs that extrapolate using Massachusetts unobservables, reported in Column 4, are all negative, regardless of whether they also extrapolate using Massachusetts observables. Though based on many restrictive assumptions, my extrapolations can potentially reconcile the positive effect of insurance on ER utilization estimated using the Oregon Health Insurance Experiment with the positive effect of insurance on ER utilization estimated using the Massachusetts health reform.

7 Conclusion

7.1 Considerations for Experimental Design

The exercise of applying MTE methods to the OHIE brings to light several issues that should be considered in the design of future experiments. The first issue is that it is easier to compare interventions, outcomes, treatments, covariates, and samples that are consistently defined within and across experiments. To facilitate comparison, data must be collected on all of these dimensions. It is especially important to collect data on treatment. "Intent to treat" estimates produced without treatment data reflect selection and treatment effect heterogeneity, and the two cannot be separated without data on treatment. It is also important to collect data on always takers and never takers. Many experiments only collect data only on individuals that remain in the experiment, potentially excluding always takers and never takers.

A subtler issue is that if a proposed policy would have always takers and never takers, then experiments to study the proposed policy should allow for always takers and never takers. Without them, the MTE cannot be identified. In the absence of always and never takers, the LATE will recover the ATE, but there will be no way to recover other policy-relevant LATEs of interest. For example, suppose that researchers want to know the impact of a policy that would make free health insurance available but not require eligibles to gain coverage. By forcing all lottery winners to gain coverage and forcing all lottery losers to go uninsured, experimenters recover an ATE that could differ from the policy-relevant LATE, and they hinder their ability to learn about selection. In clinical trials, if the patients with the most to gain select into treatment first, then designing a trial

[41] The second row of each left panel in Table 9 compares the SLATE estimated using Oregon observables and unobservables in Column 1 to the SLATE estimated using Massachusetts observables and unobservables in Column 2. All of the Oregon SLATEs are positive, and all of the Massachusetts SLATEs are negative. As shown in the next column, the difference between the Oregon and Massachusetts SLATEs is larger than the difference between the Oregon and Massachusetts LATEs. The decompositions reported in the right panels of Table 9 show that covariates explain from negative 4% to 1% of the difference in the SLATEs, depending on the decomposition and the measure of ER utilization.

that employs strong encouragement techniques to entice all patients who win the lottery to select into treatment could inadvertently dilute the policy-relevant treatment effect that would occur in the absence of strong incentives to select into treatment.

Given these issues, 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. "Selective trials" proposed by Chassang et al. [2012] have promise. With a continuous intervention, or even several different discrete interventions, the assumptions required to identify treatment effect heterogeneity are weaker.

7.2 Summary

Meta-analysis investigates external validity by examining treatment effect heterogeneity across experiments. I investigate external validity by examining treatment effect heterogeneity *within* an experiment. The understanding of treatment effect heterogeneity within an experiment also informs the optimal targeting of treatment based on observables and unobservables.

I examine treatment effect heterogeneity within the Oregon Health Insurance Experiment. I find that the treatment effect and the selection effect of insurance on ER utilization decreases from always takers to compliers to never takers. This finding informs a long-standing question by showing that insurance increased ER utilization the most for the highest users. It also informs a current policy-relevant question by showing that the impact of an insurance expansion on ER utilization depends on the individuals covered by the expansion.

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. *Journal of Political Economy*, forthcoming.

Pedro Carneiro, James J. Heckman, and Edward J. Vytlacil. Estimating marginal returns to education. *American Economic Review*, 101(6):2754–81, October 2011. doi: 10.1257/aer.101.6.2754. URL `http://www.aeaweb.org/articles/?doi=10.1257/aer.101.6.2754`.

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.

Victor Chernozhukov, Ivn Fernndez-Val, and Amanda E. Kowalski. Quantile regression with censoring and endogeneity. *Journal of Econometrics*, 186(1):201 – 221, 2015. ISSN 0304-4076. doi: http://dx.doi.org/10.1016/j.jeconom.2014.06.017. URL `http://www.sciencedirect.com/science/article/pii/S0304407614001717`.

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`.

Martin B. Hackmann, Jonathan T. Kolstad, and Amanda E. Kowalski. Adverse selection and an individual mandate: When theory meets practice. *American Economic Review*, 105(3):1030–66, 2015. doi: 10.1257/aer.20130758. URL `http://www.aeaweb.org/articles.php?doi=10.1257/aer.20130758`.

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.

Martin Huber, Lukas Laffers, and Giovanni Mellace. Sharp iv bounds on average treatment effects on the treated and other populations under endogeneity and noncompliance. *Journal of Applied Econometrics*, 2015.

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

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

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

Amanda Kowalski. Censored quantile instrumental variable estimates of the price elasticity of expenditure on medical care. *Journal of Business & Economic Statistics*, 34(1):107–117, 2016. doi: 10.1080/07350015.2015.1004072. URL http://dx.doi.org/10.1080/07350015.2015.1004072.

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.

OA Online Appendix

OA.1 Global Polynomial MTE Estimation of an MTE with Covariates

After choosing the 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.

Step 1: Specify the order M of the global polynomial

I 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:

$$ATO(x,p) = E(Y_T|X=x, U_D \leq p) = \beta_T' x + ATO(p) = \beta_T' x + \sum_{m=0}^{M} \gamma_{Tm} p^m \tag{33}$$

$$AUO(x,p) = E(Y_U|X=x, U_D > p) = \beta_U' x + AUO(p) = \beta_U' x + \sum_{m=0}^{M} \gamma_{Um} p^m. \tag{34}$$

These specifications imply that $MTE(x,p)$, $MTO(x,p)$[42] and $MUO(x,p)$[43] have the functional forms specified in (18)-(20) with M^{th} order global polynomials for $mto(p)$, $muo(p)$, and $mte(p)$:

$$MTO(x,p) = \beta_T' x + \gamma_{T0} + \sum_{m=1}^{M} (m+1)\gamma_{Tm} p^m \tag{35}$$

$$MUO(x,p) = \beta_U' x + \sum_{m=0}^{M} \gamma_{Um} p^m + \sum_{m=1}^{M} m\gamma_{Um}(p^m - p^{m-1}). \tag{36}$$

Step 2: Estimate the propensity score p

After dropping individuals with missing values for the outcome Y, I regress treatment D on the instrument Z and covariates X. I interact Z with X to harness variation in p_{Bx} and p_{Ix} across subgroups. I predict a propensity score $p_x \equiv P(D=1|Z,X)$ for each individual. The predicted propensity scores can sometimes be less than zero or greater than one.

Step 3: Estimate ATO(x,p) and AUO(x,p)

I estimate the average treated outcome function $ATO(x,p)$ using only the treated observations (the observations with $D=1$). I regress the outcome Y on the covariates X and a global polynomial in the predicted propensity score as specified in (33). I save the predicted coefficients. Similarly, I estimate the average untreated outcome function $ATO(x,p)$ using only the untreated observations (the observations with $D=0$).

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

[43] $MUO(x,p) = \frac{d[(1-p)AUO(x,p)]}{d(1-p)} = -\frac{d[(1-p)AUO(x,p)]}{dp} = -(1-p)\frac{dAUO(x,p)}{dp} + AUO(x,p)$.

Step 4: Construct estimates of MTO(x,p), MUO(x,p), and MTE(x,p)

Using the predicted coefficients saved from Step 1, I construct estimates of the marginal treated and untreated outcome functions $MTO(x,p)$ following (35) and (36). I construct $MTE(x,p)$ as the difference between $MTO(x,p)$ and $MUO(x,p)$.

Step 5: Construct estimates of SgTO, SgUO, and SgTE

The predicted propensity scores estimated in Step 2 are censored such that all negative propensity scores are censored at 0 and all propensity scores greater than 1 are censored at 1. For each individual i in the sample, we obtain a value of p_{Bx} and p_{Ix}. Using the $MTO(x,p)$, $MUO(x,p)$, and $MTE(x,p)$ estimated in Step 4, we calculate the treated outcomes, untreated outcomes, and treatment effects for each individual, which we average across all individuals by incorporating the indicator function $P(i \in g)$. $P(i \in g)$ ensures that each SgTO, SgUO, and SgTE average only incorporates the observable characteristics of the respective group g.

OA.2 Estimates Excluding Medicaid Ineligibles

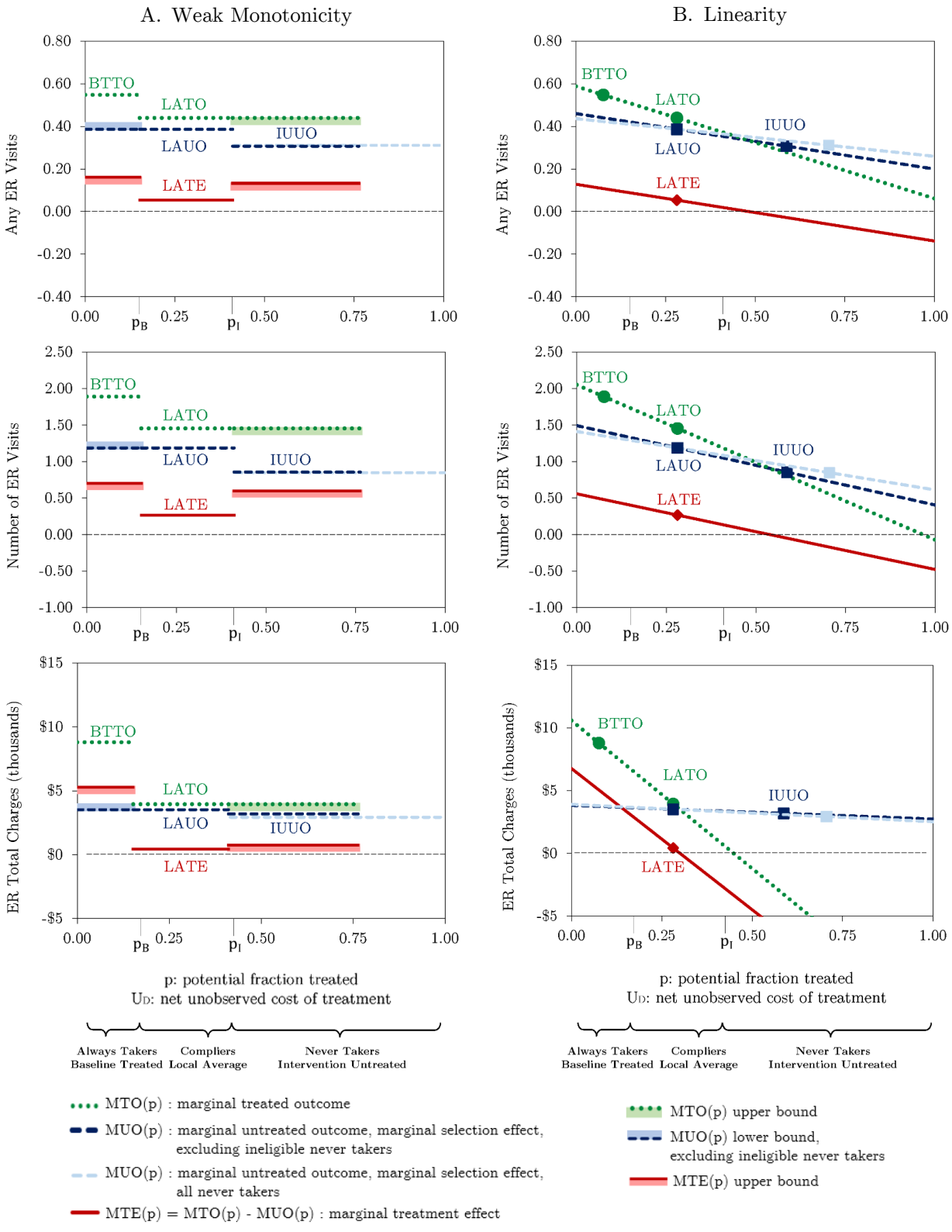
Eligibility information is not available for all individuals who lost the lottery, so it is not possible to restrict the sample to exclude ineligibles and estimate a LATE via an instrumental variable regression just on the remaining individuals. However, given available eligibility information, it is possible to estimate an $MTE(p)$ restricted to exclude ineligibles. The Oregon administrative data contain information on whether each individual who won the lottery had an approved application for the lotteried program. The only way to enroll in Medicaid without an approved application was to be eligible for the main program.

Among the intervention untreated (never takers who won the lottery), 40% submitted an application that was not approved, so we know that at least 40% of never takers were ineligible for the lotteried program, and we can identify those individuals in the data. Ineligibility dictates that these individuals have the highest unobserved net costs of treatment U_D in the full sample. Therefore, the support that excludes the ineligibles excludes the 40% of never takers with the highest values of U_D. The exclusion of ineligible never takers does not guarantee that the remaining never takers are eligible, but it gives an upper bound on the fraction of never takers that could be eligible.

Figure OA1 plots the IUUO of the remaining never takers over the remaining support. The $MUO(p)$ in sample that excludes the inelgibles is very similar to the IUUO from the full sample, as depicted by the lighter dashed line. Therefore, $MTE(p)$ and $MUO(p)$ are also similar. The ATE that excludes ineligibles is -0.005 for any ER visits, 0.0405 for the number of ER visits, and -$4,498 for total charges. These ATEs are all very similar to the ATEs that include inelgibiles.[44]

[44]The ATEs that include ineligibles are -0.02 for any ER visits, -0.02 for the number of ER visits, and -$4,432 for total charges.

Figure OA1: Bounds and Estimates of $MTE(p)$, Excluding Medicaid Ineligibles



OA.3 Distribution of Estimated Propensity Scores

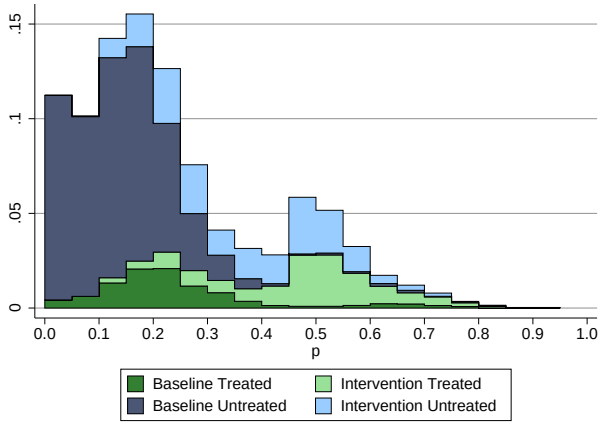
As part of the global polynomial estimation algorithm for $MTE(x,p)$, I estimate a propensity score that gives the predicted probability that each individual has insurance given observed characteristics X and lottery winning status Z. I report a histogram of the estimated propensity scores in increments of 0.05 in Figure OA2. I shade the histogram to reflect the shares of baseline and intervention treated and untreated individuals in each bin. Because $MTE(x,p)$ is the difference between the marginal treated outcome $MTO(x,p)$ and the marginal untreated outcome $MUO(x,p)$, $MTE(x,p)$ it is only nonparametrically identified in the common support of the treated and untreated. With this motivation, [Brinch et al., forthcoming] follow Carneiro et al. [2011] and identify a common support using the estimated propensity scores for the treated and untreated to estimate $MTE(x,p)$ via a local polynomial.

However, I estimate $MTE(x,p)$ via a global polynomial for extrapolation, so in the interest of using all the data, I do not define a common support for the treated and untreated. Furthermore, the estimated propensity scores reported in Figure OA2a give the support for the average treated and untreated outcome functions $ATO(x)$ and $AUO(x)$. However, the support for the marginal treated and untreated outcome functions $MTO(p)$ and $MUO(p)$, reported in Figure OA2b is wider. The average treated outcome function is estimated on a mix of baseline and intervention treated. Baseline treated have support from $0 \leq U_D \leq p_{Bx}$, and intervention treated have support from $0 \leq U_D \leq p_{Ix}$. Similarly, the average untreated outcome function is estimated on a mix of baseline untreated ($p_{Bx} < U_D \leq 1$) and intervention untreated ($p_{Bx} < U_D \leq 1$). In Figure OA2b, I draw uniformly from the relevant support for each individual to illustrate the support for the marginal functions.[45]

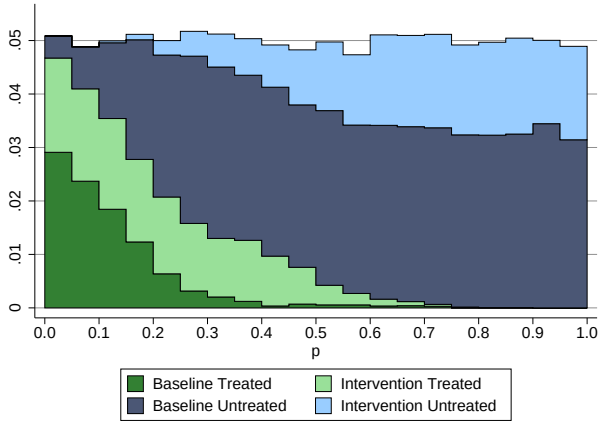
The support for the marginal functions is very different from the support for the average functions. Although there appears to be a common support for the treated and untreated at low values of U_D, there are almost no treated observations at values of U_D above 0.7. Per the discussion in Section OA.2, most never takers in that range were not eligible for Medicaid. Therefore, estimates of $MTE(x,p)$ are extrapolations for high values of U_D, but they are generally within the common support for low values of U_D.

[45] Because I estimate the propensity scores with a linear probability model, some estimated propensity scores can be less than zero or greater than one. In practice, when I incorporate all covariates, 964 observations have negative estimated propensity scores, and no observations have estimated propensity scores greater than 1. I use all of the estimated propensity scores for estimation of $MTE(x,p)$. However, to calculate treatment effects from $MTE(x,p)$ following the discussion in Section 4.4, I censor the estimated propensity scores so that they fall between 0 and 1. I report the censored propensity scores in Figure OA2. I draw from the support estimated for each individual with covariate vector x using the censored propensity scores.

Figure OA2: Distribution of Estimated Propensity Scores



(a) Support for Average Treated and Untreated Outcomes



(b) Support for Marginal Treated and Untreated Outcomes

OA.4 Implementation of the Monte Carlo Exercise

Each Monte Carlo sample has the same number of observations N as my OHIE replication sample. I generate the binary instrument Z such that $s(p_B)N$ individuals have $Z = 1$, where $s(p_B)$ is the share of lottery winners in the OHIE. I draw U_D so that it is uniformly distributed from 0 to 1 across all observations. (This is equivalent to drawing ν from any distribution and setting U_D equal to the quantiles of ν.) I generate the binary treatment D such that

$$D = \begin{cases} 1 \text{ if } 0 \leq U_D \leq pB & \& \ Z = 0 \\ 1 \text{ if } 0 \leq U_D \leq pI & \& \ Z = 1 \\ 0 \text{ otherwise.} \end{cases}$$

I generate $Y_U = MUO(U_D)$ using the $MUO(p)$ that I estimate in the OHIE for each measure of ER utilization so that there is some selection. Next, I simulate two different versions of the outcome Y. 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. I generate the observed outcome for each θ: $Y(\theta) = (1-D)Y_U + DY_T(\theta)$. I retain the simulated $Y(\theta)$, D, Z, and the true treatment effect θ for each observation.

In each Monte Carlo sample, for $Y(LATE)$ and $Y(MTE)$, I obtain an estimate of the treatment effect $\hat{\theta}$ using three estimators: $MTE(p)$, LATE, and RISOLS.[46] I calculate the bias and RMSE 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 RMSE across all samples. This exercise validates how well each estimator performs in the simulated randomized experiment.

In the natural experiment, the observed change in outcomes from before to after the experiment should only reflect the treatment effect for individuals who gain coverage (the always takers and compliers); it should be zero otherwise. Therefore, in the natural experiment, we are interested in how well each estimator recovers the observed treatment effect $D\theta = Y - Y_{pre}$ for each observation. I construct $D\hat{\theta}$ and $D\theta$ for each observation, I calculate E[$D\hat{\theta}$] and E[$D\theta$] across all observations, and I report the bias and RMSE.

OA.5 Additional Figures and Tables

[46] I do not use the $MTE(x,p)$ estimator because that estimator would require data on covariates. The most important covariates seem to be those that measure pre-period utilization, and those covariates are not available for the natural experiment (pre-period data are not available in the pre-period).

Figure OA3: $MTE(x, p)$ Robustness to Global Polynomial Order

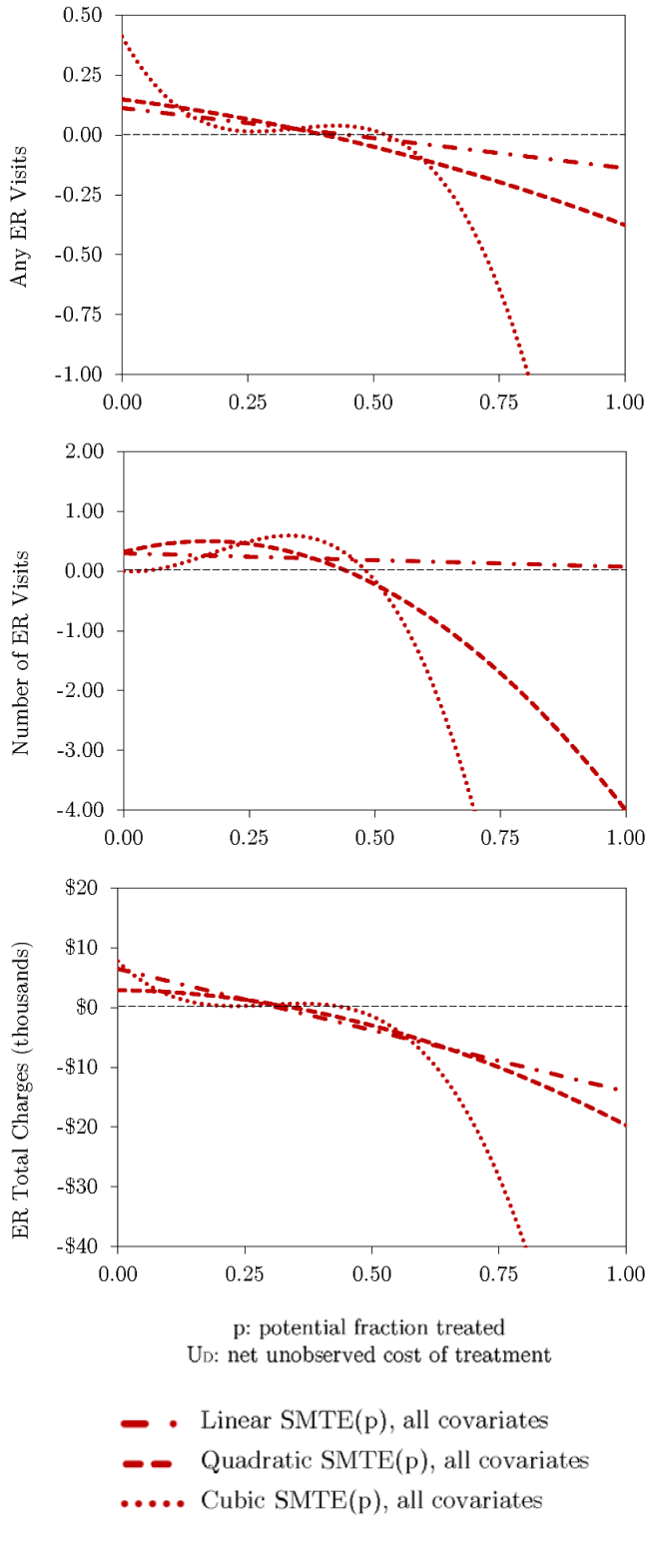


Table OA1: Average Characteristics of Always Takers, Never Takers, and Compliers: Oregon vs. Massachusetts

	(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
Age in 2009	40.7	39.4	40.3	42.4	42.0	42.2	39.0	42.4
Female	0.56	0.72	0.53	0.53	0.51	0.52	0.38	0.43
English	0.91	0.90	0.91	0.92	0.96	0.98	0.81	0.86
Number of Observations	19,643	2,986	11,565	5,092	62,456	56,548	2,733	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.

Summary statistics in the Massachusetts sample were calculated using frequency weights.

The number of observations reflects the sample counts for all always takers, never takers, and compliers.

Table OA2: OHIE Replication and Extension

Any ER Visits				
	(1)	(2)	(3)	(4)
Medicaid	0.0531	0.0551	0.0512	0.0456
	(0.0286)*	(0.0278)**	(0.0262)**	(0.0264)*
	[0.0279]*	[0.0274]**	[0.0254]**	[0.0256]*
Covariates	No covariates	Common covariates	Common covariates and pre-period ER utilization	All covariates
Regression sample	1 Lottery Entrant	1 Lottery Entrant	1 Lottery Entrant	1 Lottery Entrant
Observations	19,643	19,643	19,624	19,624
E[Y|Z=0]	0.37	0.37	0.37	0.37
Number of ER Visits				
	(1)	(2)	(3)	(4)
Medicaid	0.267	0.276	0.326	0.310
	(0.175)	(0.171)*	(0.132)**	(0.133)**
	[0.151]*	[0.149]*	[0.123]***	[0.124]**
Covariates	No covariates	Common covariates	Common covariates and pre-period ER utilization	All covariates
Regression sample	1 Lottery Entrant	1 Lottery Entrant	1 Lottery Entrant	1 Lottery Entrant
Observations	19,622	19,622	19,624	19,624
E[Y|Z=0]	1.09	1.09	1.09	1.09
ER Total Charges				
	(1)	(2)	(3)	(4)
Medicaid	$428	$458	$539	$530
	($927)	($944)	($877)	($886)
	[$935]	[$931]	[$865]	[$875]
Covariates	No covariates	Common covariates	Common covariates and pre-period ER utilization	All covariates
Regression sample	1 Lottery Entrant	1 Lottery Entrant	1 Lottery Entrant	1 Lottery Entrant
Observations	19,628	19,628	19,624	19,624
E[Y|Z=0]	$3,971	$3,971	$3,930	$3,930

*** 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.

Table OA3: LATE and MTE(p) Subgroup Analysis: Any ER visits, 1 of 2

	(1) Full sample	(2) Female		(3) Male		(4) English		(5) Non-English		(6) Age ≥ median		(7) Age < median
LATE	0.05*	0.01		0.10***		0.06**		-0.11		0.05		0.06
	(0.00, 0.10)	(-0.08, 0.08)		(0.03, 0.18)		(0.00, 0.12)		(-0.28, 0.05)		(-0.01, 0.12)		(-0.04, 0.14)
vs. full sample	-							**				
vs. complementary sample	-						**					
$\mathbf{p_B}$	0.15***	0.20***		0.10***		0.15***		0.16***		0.13***		0.17***
	(0.15, 0.16)	(0.19, 0.21)		(0.09, 0.10)		(0.14, 0.16)		(0.14, 0.19)		(0.12, 0.14)		(0.16, 0.18)
vs. full sample	-	***		***						***		***
vs. complementary 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, 0.43)		(0.34, 0.42)		(0.41, 0.44)		(0.38, 0.41)
vs. full sample	-	***		***						***		***
vs. complementary sample	-		***								***	
MTE(p) intercept	0.15***	0.12*		0.21***		0.18***		-0.07		0.25***		0.07
	(0.06, 0.23)	(-0.04, 0.25)		(0.08, 0.32)		(0.08, 0.27)		(-0.35, 0.20)		(0.15, 0.36)		(-0.08, 0.20)
vs. full sample	-					**				***		*
vs. complementary sample	-						*				**	
MTE(p) slope	-0.35***	-0.35*		-0.45**		-0.41***		-0.12		-0.71***		-0.03
	(-0.62, -0.11)	(-0.71, 0.03)		(-0.85, -0.02)		(-0.69, -0.18)		(-1.07, 0.83)		(-1.05, -0.41)		(-0.48, 0.38)
vs. full sample	-									***		**
vs. complementary sample	-										**	
p*	0.43***	0.34*		0.46**		0.44***		-0.63		0.35***		2.54
	(0.27, 0.97)	(-0.02, 1.34)		(0.23, 1.67)		(0.29, 0.81)		(-3.09, 7.03)		(0.27, 0.47)		(-7.66, 8.82)
vs. full sample	-											
vs. complementary sample	-											
RMSD	0.10***	0.10***		0.13***		0.12***		0.03***		0.21***		0.01***
	(0.03, 0.18)	(0.01, 0.21)		(0.02, 0.24)		(0.05, 0.20)		(0.00, 0.35)		(0.12, 0.30)		(0.00, 0.14)
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.

Individuals with missing values for the corresponding pre-utilization measure were not included in the subgroups.

Table OA4: LATE and MTE(p) Subgroup Analysis: Any ER visits, 2 of 2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full sample	On SNAP in the pre-period	Not on SNAP in the pre-period	On TANF in the pre-period	Not on TANF in the pre-period	Signed up for lottery on first day	Did not sign up for lottery on first day
LATE	0.05*	0.06**	0.00	0.07	0.05*	0.03	0.06*
	(0.00, 0.10)	(0.00, 0.12)	(-0.11, 0.11)	(-1.68, 1.12)	(0.00, 0.10)	(-0.09, 0.15)	(-0.01, 0.11)
vs. full sample	-						
vs. complementary sample	-						
$\mathbf{p_B}$	0.15***	0.21***	0.08***	0.63***	0.14***	0.16***	0.15***
	(0.15, 0.16)	(0.20, 0.21)	(0.07, 0.09)	(0.58, 0.69)	(0.13, 0.15)	(0.14, 0.18)	(0.14, 0.16)
vs. full sample	-	***	***	***	***		
vs. complementary sample	-		***		***		
$\mathbf{p_I}$	0.41***	0.52***	0.26***	0.75***	0.40***	0.55***	0.40***
	(0.40, 0.42)	(0.51, 0.54)	(0.24, 0.27)	(0.68, 0.81)	(0.39, 0.41)	(0.52, 0.59)	(0.38, 0.41)
vs. full sample	-	***	***	***	***	***	***
vs. complementary sample	-		***		***		***
MTE(p) intercept	0.15***	0.15***	0.14*	-0.69	0.14***	0.16	0.15***
	(0.06, 0.23)	(0.04, 0.26)	(0.00, 0.29)	(-5.48, 2.58)	(0.05, 0.23)	(-0.05, 0.42)	(0.05, 0.25)
vs. full sample	-						
vs. complementary sample	-						
MTE(p) slope	-0.35***	-0.23*	-0.81**	1.10	-0.35***	-0.38	-0.35***
	(-0.62, -0.11)	(-0.50, 0.02)	(-1.44, -0.19)	(-3.89, 7.88)	(-0.62, -0.10)	(-0.96, 0.18)	(-0.63, -0.10)
vs. full sample	-						
vs. complementary sample	-						
$\mathbf{p^*}$	0.43***	0.63*	0.17*	0.63	0.42***	0.43	0.43***
	(0.27, 0.97)	(-0.34, 2.01)	(-0.09, 0.41)	(-2.93, 2.16)	(0.27, 0.98)	(-0.62, 1.56)	(0.25, 1.03)
vs. full sample	-	*	**				
vs. complementary sample	-		*				
RMSD	0.10***	0.07***	0.23***	0.32***	0.10***	0.11***	0.10***
	(0.03, 0.18)	(0.01, 0.14)	(0.05, 0.41)	(0.02, 2.27)	(0.03, 0.18)	(0.01, 0.28)	(0.03, 0.18)
N	19,643	11,181	8,462	464	19,179	1,827	17,816

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

Individuals with missing values for the corresponding pre-utilization measure were not included in the subgroups.

Table OA5: LATE and MTE(p) Subgroup Analysis: Number of ER Visits 1 of 2

	(1)	(2)		(3)	(4)	(5)	(6)		(7)
	Full sample	Female		Male	English	Non-English	Age ≥ median		Age < median
LATE	0.27	0.14		0.39*	0.30	-0.15	0.14		0.44
	(-0.09, 0.54)	(-0.40, 0.61)		(-0.08, 0.79)	(-0.09, 0.58)	(-0.70, 0.40)	(-0.20, 0.51)		(-0.10, 0.91)
vs. full sample	-								
vs. complementary 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)
vs. full sample	-	***		***			***		***
vs. complementary 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)		(0.36, 0.40)	(0.40, 0.43)	(0.34, 0.42)	(0.41, 0.45)		(0.38, 0.41)
vs. full sample	-	***		***			**		**
vs. complementary sample	-		***					**	
MTE(p) intercept	0.64***	0.48		0.92***	0.72***	0.14	0.98***		0.31
	(0.14, 1.07)	(-0.40, 1.10)		(0.22, 1.55)	(0.17, 1.21)	(-0.81, 0.81)	(0.34, 1.60)		(-0.57, 1.08)
vs. full sample	-				*		*		
vs. complementary sample	-								
MTE(p) slope	-1.32	-1.06		-2.20	-1.51	-1.07	-3.01***		0.48
	(-2.94, 0.44)	(-3.17, 1.25)		(-4.96, 0.93)	(-3.24, 0.35)	(-4.46, 2.78)	(-4.94, -0.91)		(-2.18, 3.75)
vs. full sample	-						**		*
vs. complementary sample	-							**	
p*	0.48	0.45		0.42	0.48	0.13	0.33***		-0.63
	(-0.92, 2.26)	(-3.32, 3.83)		(-0.60, 1.36)	(-2.92, 1.43)	(-1.25, 2.73)	(0.20, 0.63)		(-3.13, 3.83)
vs. full sample	-								
vs. complementary sample	-								
RMSD	0.38***	0.31***		0.63***	0.44***	0.31***	0.87***		0.14***
	(0.03, 0.85)	(0.02, 0.92)		(0.06, 1.43)	(0.03, 0.93)	(0.02, 1.29)	(0.26, 1.43)		(0.02, 1.08)
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.

Individuals with missing values for the corresponding pre-utilization measure were not included in the subgroups.

Table OA6: LATE and MTE(p) Subgroup Analysis: Number of ER Visits, 2 of 2

	(1) Full sample	(2) On SNAP in the pre-period	(3) Not on SNAP in the pre-period	(4) On TANF in the pre-period	(5) Not on TANF in the pre-period	(6) Signed up for lottery on first day	(7) Did not sign up for lottery on first day
LATE	0.27	0.25	0.17	2.03	0.24	0.13	0.29
	(-0.09, 0.54)	(-0.22, 0.63)	(-0.24, 0.63)	(-3.46, 11.46)	(-0.14, 0.50)	(-0.54, 0.80)	(-0.10, 0.59)
vs. full sample	-						
vs. complementary sample	-						
$\mathbf{p_B}$	0.15***	0.21***	0.08***	0.63***	0.14***	0.16***	0.15***
	(0.15, 0.16)	(0.20, 0.21)	(0.07, 0.09)	(0.58, 0.69)	(0.13, 0.15)	(0.14, 0.18)	(0.14, 0.16)
vs. full sample	-	***	***	***	***		
vs. complementary sample	-		***		***		
$\mathbf{p_I}$	0.41***	0.52***	0.26***	0.75***	0.40***	0.55***	0.40***
	(0.40, 0.42)	(0.51, 0.54)	(0.24, 0.27)	(0.68, 0.81)	(0.39, 0.41)	(0.51, 0.59)	(0.38, 0.41)
vs. full sample	-	***	***	***	***	***	***
vs. complementary sample	-		***		***		***
MTE(p) intercept	0.64***	0.68*	0.31	-15.69**	0.67***	0.30	0.69***
	(0.14, 1.07)	(-0.06, 1.27)	(-0.36, 0.88)	(-64.75, -2.78)	(0.15, 1.10)	(-0.95, 1.45)	(0.13, 1.18)
vs. full sample	-			**			
vs. complementary sample	-				**		
MTE(p) slope	-1.32	-1.16	-0.81	25.71**	-1.58*	-0.47	-1.48
	(-2.94, 0.44)	(-2.64, 0.74)	(-3.67, 2.06)	(4.07, 97.03)	(-3.14, 0.34)	(-3.79, 2.99)	(-3.35, 0.39)
vs. full sample	-			**			
vs. complementary sample	-				**		
$\mathbf{p^*}$	0.48	0.58	0.38	0.61***	0.42*	0.63	0.47
	(-0.92, 2.26)	(-1.59, 3.82)	(-0.84, 3.39)	(0.39, 0.92)	(-0.60, 1.48)	(-3.67, 11.37)	(-1.56, 2.09)
vs. full sample	-						
vs. complementary sample	-						
RMSD	0.38***	0.34***	0.23***	7.42***	0.46***	0.14***	0.43***
	(0.03, 0.85)	(0.01, 0.76)	(0.02, 1.06)	(1.66, 33.17)	(0.04, 0.91)	(0.00, 1.13)	(0.04, 0.97)
N	19,622	11,163	8,459	464	19,158	1,825	17,797

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

Individuals with missing values for the corresponding pre-utilization measure were not included in the subgroups.

Table OA7: LATE and MTE(p) Subgroup Analysis: ER Total Charges, 1 of 2

	(1) Full sample	(2) Female		(3) Male	(4) English	(5) Non-English	(6) Age ≥ median		(7) Age < median
LATE	$428	$358		$458	$579	-$1,698	-$845		$2,447*
	(-$1,436, $2,142)	(-$1,786, $3,052)		(-$2,279, $2,486)	(-$1,498, $2,543)	(-$5,325, $2,757)	(-$3,285, $1,376)		(-$256, $5,163)
vs. full sample	-								*
vs. complementary 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)
vs. full sample	-	***		***			***		***
vs. complementary 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)
vs. full sample	-	***		***			***		***
vs. complementary sample	-		***					***	
MTE(p) intercept	$6,677***	$3,526*		$12,621***	$7,141***	$3,273	$12,978***		$1,043
	($3,555, $9,326)	(-$536, $6,938)		($7,705, $17,980)	($3,848, $9,828)	(-$2,518, $12,965)	($7,299, $17,478)		(-$2,109, $5,619)
vs. full sample	-	**		***			***		***
vs. complementary sample	-		***					***	
MTE(p) slope	-$22,218***	-$10,050		-$51,011***	-$23,270***	-$18,170*	-$49,187***		$4,986
	(-$33,486, -$11,076)	(-$22,997, $2,825)		(-$76,451, -$30,594)	(-$35,548, -$10,572)	(-$45,109, $1,932)	(-$66,927, -$31,935)		(-$13,008, $15,926)
vs. full sample	-	***		***			***		***
vs. complementary sample	-		***					***	
$\mathbf{p^*}$	0.30***	0.35		0.25***	0.31***	0.18	0.26***		-0.21
	(0.22, 0.45)	(-1.08, 1.24)		(0.20, 0.30)	(0.23, 0.46)	(-0.76, 0.46)	(0.21, 0.31)		(-3.40, 8.82)
vs. full sample	-			**					
vs. complementary sample	-								
RMSD	$6,414***	$2,901***		$14,726***	$6,717***	$5,245***	$14,199***		$1,439***
	($3,197, $9,667)	($317, $6,639)		($8,832, $22,070)	($3,052, $10,262)	($620, $13,022)	($9,219, $19,320)		($52, $4,933)
N	19,628	10,939		8,689	17,877	1,751	10,309		9,319

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

Individuals with missing values for the corresponding pre-utilization measure were not included in the subgroups.

Table OA8: LATE and MTE(p) Subgroup Analysis: Total Charges, 2 of 2

	(1) Full sample	(2) On SNAP in the pre-period	(3) Not on SNAP in the pre-period	(4) On TANF in the pre-period	(5) Not on TANF in the pre-period	(6) Signed up for lottery on first day	(7) Did not sign up for lottery on first day
LATE	$428	$838	-$964	$6,696	$359	$135	$460
	(-$1,436, $2,142)	(-$1,040, $2,744)	(-$4,415, $2,853)	(-$22,902, $61,820)	(-$1,490, $2,030)	(-$4,420, $4,184)	(-$1,646, $2,320)
vs. full sample	-						
vs. complementary sample	-						
$\mathbf{p_B}$	0.15***	0.21***	0.08***	0.63***	0.14***	0.16***	0.15***
	(0.15, 0.16)	(0.20, 0.21)	(0.07, 0.09)	(0.58, 0.69)	(0.13, 0.15)	(0.14, 0.18)	(0.14, 0.16)
vs. full sample	-	***	***	***	***		
vs. complementary sample	-		***		***		
$\mathbf{p_I}$	0.41***	0.52***	0.26***	0.75***	0.40***	0.55***	0.40***
	(0.40, 0.42)	(0.51, 0.54)	(0.24, 0.27)	(0.68, 0.81)	(0.39, 0.41)	(0.51, 0.59)	(0.38, 0.41)
vs. full sample	-	***	***	***	***	***	***
vs. complementary sample	-		***		***		***
MTE(p) intercept	$6,677***	$7,535***	$5,134**	-$2,132	$7,285***	$6,455	$6,742***
	($3,555, $9,326)	($3,795, $11,332)	($879, $10,134)	(-$116,349, $155,529)	($4,153, $9,843)	(-$1,474, $14,257)	($3,555, $9,629)
vs. full sample	-				***		
vs. complementary sample	-						
MTE(p) slope	-$22,218***	-$18,369***	-$36,019**	$12,824	-$25,522***	-$17,821*	-$22,977***
	(-$33,486, -$11,076)	(-$28,906, -$8,015)	(-$67,999, -$3,445)	(-$170,457, $141,769)	(-$37,015, -$13,320)	(-$39,988, $2,145)	(-$36,598, -$9,550)
vs. full sample	-				***		
vs. complementary sample	-						
$\mathbf{p^*}$	0.30***	0.41***	0.14*	0.17	0.29***	0.36	0.29***
	(0.22, 0.45)	(0.31, 0.59)	(0.00, 0.30)	(-3.51, 5.41)	(0.22, 0.40)	(-0.68, 0.74)	(0.20, 0.50)
vs. full sample	-	**	**		**		
vs. complementary sample	-		**				
RMSD	$6,414***	$5,303***	$10,398***	$3,702***	$7,368***	$5,145***	$6,633***
	($3,197, $9,667)	($2,314, $8,344)	($1,360, $19,630)	($546, $62,713)	($3,845, $10,685)	($507, $11,544)	($2,757, $10,565)
N	19,628	11,171	8,457	463	19,165	1,825	17,803

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

Individuals with missing values for the corresponding pre-utilization measure were not included in the subgroups.

Table OA9: Treated Outcomes, Untreated Outcomes, and Treatment Effects in Oregon: $MTE(x,p)$

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Group		Baseline Treated (Always Takers)	Baseline Untreated (Never Takers and Untreated Compliers)	Intervention Treated (Always Takers and Treated Compliers)	Intervention Untreated (Never Takers)	Randomized Intervention Sample Treated	Randomized Intervention Sample Untreated	Local Average (Treated and Untreated Compliers)	Average
P(i ∈ g)		1(Z=0, D=1)	1(Z=0, D=0)	1(Z=1, D=1)	1(Z=1, D=0)	1(D=1)	1(D=0)	$\frac{(p_{Ix}-p_{Bx})}{p_{Ix}}P(i \in IT) + \frac{(p_{Ix}-p_{Bx})}{(1-p_{Bx})}P(i \in BU)$	1
g		SBT	SBU	SIT	SIU	SRIST	SRISU	SLA	SA
Any ER Visits	Treated	0.54***	0.30***	0.48***	0.24***	0.49***	0.26***	0.41***	0.34***
	Outcome	(0.52, 0.56)	(0.22, 0.38)	(0.46, 0.50)	(0.14, 0.35)	(0.48, 0.56)	(0.18, 0.37)	(0.37, 0.45)	(0.27, 0.41)
	TO	SBTTO	SBUTO	SITTO	SIUTO	SRISTTO	SRISUTO	SLATO	SATO
	Untreated	0.45***	0.33***	0.41***	0.31***	0.42***	0.32***	0.37***	0.35***
	Outcome	(0.38, 0.51)	(0.33, 0.34)	(0.37, 0.46)	(0.30, 0.33)	(0.38, 0.49)	(0.31, 0.34)	(0.33, 0.42)	(0.34, 0.37)
	UO	SBTUO	SBUUO	SITUO	SIUUO	SRISTUO	SRISUUO	SLAUO	SAUO
	Treatment	0.09***	-0.03	0.07***	-0.07	0.07***	-0.06	0.04*	-0.01
	Effect	(0.02, 0.17)	(-0.12, 0.05)	(0.02, 0.12)	(-0.17, 0.04)	(0.02, 0.15)	(-0.15, 0.05)	(-0.01, 0.09)	(-0.08, 0.06)
	TE	SBTTE	SBUTE	SITTE	SIUTE	SRISTTE	SRISUTE	SLATE	SATE
Number of ER Visits	Treated	1.91***	1.09***	1.58***	0.95***	1.69***	1.01***	1.34***	1.21***
	Outcome	(1.74, 2.06)	(0.59, 1.64)	(1.45, 1.70)	(0.32, 1.63)	(1.58, 1.88)	(0.47, 1.67)	(1.14, 1.57)	(0.80, 1.67)
	TO	SBTTO	SBUTO	SITTO	SIUTO	SRISTTO	SRISUTO	SLATO	SATO
	Untreated	1.53***	0.94***	1.27***	0.85***	1.36***	0.89***	1.11***	1.03***
	Outcome	(1.20, 1.88)	(0.91, 0.99)	(1.01, 1.54)	(0.79, 0.93)	(1.10, 1.69)	(0.84, 0.97)	(0.89, 1.34)	(0.96, 1.12)
	UO	SBTUO	SBUUO	SITUO	SIUUO	SRISTUO	SRISUUO	SLAUO	SAUO
	Treatment	0.38**	0.15	0.31**	0.09	0.33**	0.11	0.24*	0.18
	Effect	(0.02, 0.72)	(-0.37, 0.69)	(0.04, 0.56)	(-0.56, 0.80)	(0.03, 0.60)	(-0.46, 0.75)	(-0.04, 0.49)	(-0.24, 0.61)
	TE	SBTTE	SBUTE	SITTE	SIUTE	SRISTTE	SRISUTE	SLATE	SATE
ER Total Charges	Treated	$8,787***	-$2,331	$5,700***	-$5,015*	$6,043***	-$4,690	$3,312***	-$626
	Outcome	($7,726, $9,774)	(-$6,677, $1,605)	($4,950, $6,421)	(-$10,371, $145)	($5,333, $9,195)	(-$9,183, $1,145)	($1,407, $4,870)	(-$4,136, $2,625)
	TO	SBTTO	SBUTO	SITTO	SIUTO	SRISTTO	SRISUTO	SLATO	SATO
	Untreated	$3,901***	$3,053***	$3,491***	$2,922***	$3,690***	$3,068***	$3,053***	$3,173***
	Outcome	($2,204, $5,814)	($2,860, $3,290)	($2,122, $4,880)	($2,532, $3,287)	($2,309, $5,474)	($2,771, $3,315)	($1,926, $4,294)	($2,787, $3,579)
	UO	SBTUO	SBUUO	SITUO	SIUUO	SRISTUO	SRISUUO	SLAUO	SAUO
	Treatment	$4,886***	-$5,385***	$2,208***	-$7,937***	$2,354***	-$7,758***	$1,012	-$3,798**
	Effect	($2,396, $6,791)	(-$9,704, -$1,461)	($727, $3,823)	(-$13,341, -$2,757)	($956, $6,056)	(-$12,181, -$1,814)	(-$573, $2,986)	(-$7,323, -$466)
	TE	SBTTE	SBUTE	SITTE	SIUTE	SRISTTE	SRISUTE	SLATE	SATE

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