HEDG Working Paper 07/07

Use of instrumental variables in the presence of heterogeneity and self-selection: An application in breast cancer patients

Anirban Basu
James J. Heckman
Salvador Navarro-Lozano
Sergio Urzua

June 2007

ISSN 1751-1976

york.ac.uk/res/herc/hedgwp

Use of instrumental variables in the presence of heterogeneity and self-selection: An application in breast cancer patients

Anirban Basu[*,1], James J. Heckman[1,2],
Salvador Navarro-Lozano[3], Sergio Urzua[4]

[1]University of Chicago, [2]The American Bar Foundation,
[3]University of Wisconsin-Madison, [4]Northwestern University

June 26, 2007

ABSTRACT

Instrumental variables methods (IV) are widely used in the health economics literature to adjust for hidden selection biases in observational studies when estimating treatment effects. Less attention has been paid in the applied literature to the proper use of instrumental variables if treatment effects are heterogeneous across subjects and individuals select treatments based on expected idiosyncratic gains or losses from treatments. In this paper, we analyze the role of conventional instrumental variable analysis and alternative approaches using instrumental variables for estimating treatment effects for models with treatment heterogeneity and self-selection. Instead of interpreting IV estimates as the effect of treatment at an unknown margin of patients, we identify the marginal patients and we apply the method of local instrumental variables to estimate the Average Treatment Effect (ATE) and the Effect on the Treated (TT) on 5-year direct costs of breast conserving surgery and radiation therapy compared to mastectomy in breast cancer patients. We use a sample from the Outcomes and Preferences in Older Women, Nationwide Survey (OPTIONS) which is designed to be representative of all female Medicare beneficiaries (aged 67 or older) with newly diagnosed breast cancer between 1992 and 1994. Our results reveal some of the advantages and limitations of conventional and alternative IV methods in estimating mean treatment effect parameters.

Keywords: Self-selection, essential heterogeneity, instrumental variables, breast cancer, local instrumental variable method.

JEL Classification: C01, C21, C31

We are grateful to Daniel Polsky at the University of Pennsylvania and the OPTIONS team for allowing us to use their breast cancer dataset. We thank Willard Manning, Partha Deb and the seminar participants at the University of Chicago, University of Illinois at Chicago and at the American Society for Health Economists Inaugural Meeting for their helpful comments.

* Corresponding author:
Anirban Basu, PhD, Section of General Internal Medicine, Department of Medicine, University of Chicago
5841 S. Maryland Ave, MC 2007, Chicago IL 60637. abasu@medicine.bsd.uchicago.edu, Tel: +1 312-834-1796, Fax: +1 773-834-2238

1. Introduction

In many situations, people respond differently to the same treatment.[1] This is called response heterogeneity. In particular, when this differential response is based on characteristics not observed by the analyst, it is called unobserved heterogeneity. When individuals select into specific treatments based on characteristics that determine this heterogeneity self-selection arises.[2] The concepts of heterogeneity and self-selection of agents have become fundamental for the development of modern microeconometric tools for the analysis of individual choices and their consequences (Heckman, 2001).

Medical care is usually characterized by heterogeneity and self-selection and therefore, microeconometric methods that account for these are of great value for analyzing the effects of alternative medical technologies on costs, effectiveness and cost-effectiveness in health and medicine. This paper analyzes the role of instrumental variables in the construction of counterfactual outcomes related to health care policy making in the presence of both heterogeneity and self-selection.

The classic framework for evaluation in clinical outcomes research is the randomized experiment.[3] Randomization is the preferred methodology for the estimation of mean treatment effect in health economics and health services research too (Vanness and Mullahy, 2006). When randomization is not feasible or does not provide population coverage, researchers have relied on observational studies to estimate treatment effects. Instrumental variable (IV) analysis has been one of the cornerstone methods to address the issue of selection bias in observational studies. Recently, some researchers (Earle, 2001; Hadley, 2003; Brooks, 2003) have used IV analysis to evaluate alternative treatments in cancer for example – the types of evaluations that were by and large restricted to clinical trials.

Traditional IV methods assume that the only form of heterogeneity is that observed by the analyst. More recent methods relax this assumption and allow for unobserved heterogeneity in outcomes when applying IV. These methods, however, implicitly assume that the self-selection behavior is not influenced by the unobserved determinants of heterogeneity in outcomes.[4] In the medical care context, this assumption implies that either a) treatment effects are constant for every

[1] Here the term *treatment* is used in the generic sense. It can stand for medical treatments, social interventions, public policies, etc.
[2] Accordingly we can have selection based on observables and selection based on unobservables. See Heckman and Vytlacil (1999) and Heckman and Navarro (2004).
[3] Ideas on randomized experiments dating back to Neyman (1923, 1935) and Fisher (1935) have been instrumental in controlling for self-selection.
[4] This is the traditional assumption of most random effects and random-coefficient models in statistics and economics.

one in the population with the same observed characteristics or b) even if treatment effects are heterogeneous, patients (or their physicians) cannot anticipate these effects and use this information to select into the treatment that would potentially give them the largest benefits. That is, they have no information beyond what the analyst of an observational data possesses (see Heckman and Robb, 1985, and Heckman, 1996, 1997).[5] Either (a) or (b) represent a stretch for modeling treatment choices in health care, especially under the practical limitations of observational data to collect all relevant information pertaining to treatment choices.

Imbens and Angrist (1994) show that when these idiosyncratic gains are correlated with treatment receipt, standard IV methods (under additional monotonicity assumptions[6]) can identify the mean response of outcomes for persons induced to participate in treatment by a change in the value of the instrument. This treatment effect parameter is called the Local Average Treatment Effect (LATE). However, Imbens and Angrist (1994) do not identify who these patients are. Specifically, they do not identify who would change their treatment status in response to changes in instrument values.

In this context, the health economics literature correctly interprets LATE as the effect on some marginal population (McClellan, McNeil, and Newhouse, 1994; Brooks et al, 2003), admitting that the method does not identify the margin. This important limitation of LATE (and IV in general) is discussed in two health economics papers (Newhouse and McClellan, 1998; Harris and Remler, 1998) both concentrating on one of the first applications of instrumental variables on clinical outcomes (McClellan et al., 1994). As Newhouse and McClellan (1998) point out

".. unlike the population in a clinical trial, this marginal population who changes treatment because of location (instrumental variable in this case) is not straightforwardly identifiable by the clinician, that is, whether the patient is part of that population is not immediately obvious to the physician."

This paper has two main goals. First, we review in detail the theory behind essential heterogeneity and marginal treatment effects as these ideas are most likely to be extremely relevant to health economics and applied health services research. Moreover, we illustrate, using a step-by-step guide, how the local instrumental variable (LIV) methods could be applied in practice to estimate the marginal treatment effects and other treatment effect parameters.

Second, we make novel contributions to the LIV methods in developing methods to calculate LIV weights that simultaneously vary over both the unobserved characteristics and also the multiple

[5] Or if they have it, they do not use it when selecting treatment.
[6] This assumption is explained in detail in section 5. In summary, it states that all patients who are induced to change their treatment status in response to changes in the value of an instrument do so in the same "direction".

observed covariates in the outcome regression. These methods are fundamental to any practical application of LIV methods. Our empirical example is one of the first illustrations of how negative IV weights can occur when multiple instruments are used in the choice equation along with other covariates.

We draw on Heckman and Vytlacil (1999, 2001a, 2007) to present new methods that extend the IV-LATE approach to estimate mean responses to treatment but identifying the exact population affected by the treatment. We follow the literature in health economics and focus on identifying the average treatment affect (ATE) of a treatment compared to no treatment (or any other standard treatment). This is a difficult parameter to identify. ATE estimates the average gain if everyone undergoes treatment as compared to an alternative treatment or no treatment at all. This has been one of the most popular parameters of interest for health economists and policy analysts when making inference about health care policies (Vanness and Mullahy, 2006). We also discuss other mean treatment parameters such as the Treatment Effect on the Treated (TT), which estimates the average gain to those who actually select into treatment.

This paper proceeds as follows. In section 2, we set up a model for potential outcomes and formally define ATE and TT. We then illustrate how selection bias arises when traditional regression techniques are used to try and recover causal parameters (section 3). In section 4, we show how conventional linear instrumental variables econometric analysis, which has been widely used in health economics to make causal inferences, attempts to overcome these biases, and how it fails to identify ATE when selection of treatment depends on the heterogeneous outcome. In section 5, we introduce the concept of the Local Average Treatment Effect (LATE) and Marginal Treatment Effect (MTE) and its implementation via the local instrumental variable (LIV) method that can overcome the limitations of the traditional IV approach. In section 6, we illustrate these methods in an evaluation of the costs of localized breast cancer treatments in Medicare population. Discussion follows.

2. A Model of Heterogeneous Outcomes and the Average Treatment Effects

We distinguish between the underlying model generating potential outcomes as a result of receipt of a particular treatment, and the limited information available to the analyst who, retrospectively, aims to identify ATE by modeling the observed outcomes as a function of treatment choices. To simplify, we restrict our discussion to the case of two states – the *treated* state denoted

by j = 1 and the *untreated* state denoted by j = 0, and their corresponding potential outcomes represented by

$$Y_1 = \mu_1(X) + U_1, \qquad \textbf{(1a)}$$

$$Y_0 = \mu_0(X) + U_0, \qquad \textbf{(1b)}$$

where $\mu_j(X)$ is a function of characteristics (X) that are observed by the analyst and the agent, and U_j represents characteristics unobserved to the analyst but potentially known by the agent. In what follows, we use the terms *observed* and *unobserved* in reference to the analyst's perspective. If we assume that $E(U_j \mid X = x) = 0$, j = 0, 1, then we can interpret $\mu_j(X)$ as the conditional expectation of the outcome, that is $E\left(Y_j \mid X = x\right) = \mu_j(x)$.

The individual gain from treatment, $\Delta = Y_1 - Y_0$, contains two components: 1) the Average Treatment Effect (ATE(X))= $(\mu_1(X) - \mu_0(X))$,[7] which is the gain of receiving treatment for the average person with characteristics X, and *2)* $(U_1 - U_0)$, which is the idiosyncratic gain for a particular person undergoing treatment (Heckman, 1997). The second component constitutes ***unobserved heterogeneity*** and signifies that the incremental effect of treatment over no treatment may vary across individuals even after controlling for observable heterogeneity using covariates X.

The first component (ATE(X)) has received the most attention in many economic evaluations (Card, 2001; Imbens, 2004) and in the health economics literature (see Claxton, 1999; and Vanness and Mullahy, 2006 for an overview of its use in cost-effectiveness analysis). ATE conditional on X = x, $E(\Delta \mid X = x) = (\mu_1(x) - \mu_0(x))$ estimates the average gain if someone selected randomly from the general population with characteristics X = x undergoes treatment as compared to remaining untreated. An ideal experiment administered to persons selected at random from the general population with no non-compliance or non-response issues would estimate ATE.

A parameter that has significant policy relevance in health care but has received less attention is the effect of Treatment on the Treated (TT). This parameter evaluates the average gain from treatment among those who select into receiving treatment. Despite the disproportionate emphasis on ATE for policy making in health care, one can easily appreciate the importance of the TT parameter for health policy.[8] TT is one ingredient for determining whether a given treatment should be shut down or retained as a medical practice or in the formularies. It informs on whether the persons choosing the treatment benefit or not from it in gross terms. If we let D be an indicator that

[7] All treatment effect parameters can be defined as conditional on X. The unconditional version of a parameter can be obtained by integrating the corresponding conditional parameter using the observed distribution of X.

[8] See Auld (2005) for an application in health economics looking at TT. He looks at the causal effect of early initiation on adolescent smoking patterns among those initiating early smoking.

takes the value *1* if an individual selects into treatment and *0* if he does not, then TT conditional on $X = x$ is formally defined as

$$TT(x) = E(\Delta \mid X = x, D = 1) = (\mu_1(x) - \mu_0(x)) + E(U_1 - U_0 \mid X = x, D = 1). \qquad \textbf{(2)}$$

Notice that, conditional on $X = x$, TT is different from ATE only when the gains on unobservables are heterogeneous, i.e., when they are different for those who select into treatment. If the term $E(U_1 - U_0 \mid X = x, D = 1)$ in (2) is constant in the population, then the treatment effect is the same for everyone, and TT and ATE are identical (conditional on X). Thus, only when unobserved gains vary with treatment selection is the distinction between ATE and TT relevant.

3. OLS under Selection on Unobserved and Heterogeneous Gains

Since each subject either receives treatment or not, the observed outcome, *Y,* becomes

$$Y = DY_1 + (1-D)Y_0. \qquad \textbf{(3)}$$

This representation for the observed outcome is widely used in the literature.[9] The analyst does not observe both potential outcomes (treated and untreated state outcomes) for each subject. That is, he does not observe the pair (Y_1, Y_0) for anybody. If he did, he could simply form Y_1-Y_0 at the individual level and recover any treatment parameter.

The model for potential outcomes in equation (1a) and (1b) can be substituted into equation (2) to obtain the following model for the observed outcome (Quandt, 1972):

$$\begin{aligned} Y &= \mu_0(X) + D(\mu_1(X) - \mu_0(X)) + \{D(U_1 - U_0) + U_0\} \\ &= \mu_0(X) + DE(\Delta \mid X) + \{D(U_1 - U_0) + U_0\}. \end{aligned} \qquad \textbf{(4)}$$

Equation (4) can be interpreted as a standard regression model where the observed outcome Y is regressed on the covariates X and on an indicator for treatment D (interacted with the covariates X). However, the fact that the error term (given in curly brackets) depends on the treatment variable makes the analysis of this model non-standard. In order to see this, notice that from (4), the OLS estimator of the "effect of the treatment" is simply the difference in adjusted mean outcomes for treated and untreated individuals:

$$E(Y \mid X = x, D = 1) - E(Y \mid X = x, D = 0)$$

[9] See Neyman (1923), Fisher (1935), Roy (1951), Cox (1958), Quandt (1972, 1988), Rubin (1978).

$$= [\mu_1(x) + E(U_1 \mid X = x, D = 1)] - [\mu_0(x) + E(U_0 \mid X = x, D = 0)]$$

$$= E(\Delta \mid X) + [E(U_1 - U_0 \mid X = x, D = 1) + \{E(U_0 \mid X = x, D = 1) - E(U_0 \mid X = x, D = 0)\}]. \quad (5)$$

Equation (5) shows that the OLS treatment effect estimator is not necessarily an unbiased estimator for ATE, even asymptotically (Heckman, 1997). The term in square brackets in (5) represents the bias produced by the selection into treatment based on the potential outcomes. The bias consists of two parts – 1) the mean idiosyncratic gain from treatment for people who receive treatment (first term), and 2) the difference in the untreated outcomes between treated and untreated individuals (second term in curly brackets).

4. Role of Standard Instrumental Variables and its Limitations

Instrumental variable (IV) analysis is one of the most popular methods used to deal with selection biases. This method postulates the existence of a random variable (or variables) *Z* called instruments such that, conditional on the observable covariates *X*; these instruments are independent of potential outcomes but can predict treatment choice. Formally, an instrument *Z* must meet the following assumptions:

Assumption 1. The probability of treatment choice is a non-trivial function of the instrument *Z* conditional on *X*.

Assumption 2. The instrument *Z* is mean independent of the error terms in (1a) and (1b) conditional on *X*, i.e, $E(U_0 \mid X = x, Z = z) = 0$ and $E(U_1 \mid X = x, Z = z) = 0$.[10]

Notice that no formal model for the choice process (*D*) is required when implementing the IV approach. This, in principle, represents an attractive feature. However, as we show below, absence of an explicit choice model is also a drawback of the IV approach.

Under these assumptions, and considering the model of potential outcomes introduced in Section 2, we can consider three distinct scenarios.

(a) *Absence of unobserved heterogeneity*: This implies that $U_1 = U_0$ and all individuals get the same benefit from treatment (conditional on *X*). This model is called the "Dummy endogenous regression" model or the common coefficient model (Heckman, 1978; Heckman and Robb, 1986; LaLonde, 1986). In this situation, individual level treatment effects are homogenous and equal to

[10] Assumption 2 is usually stated as having a *Z* that is independent of the error term. The weaker version we state is enough to obtain identification of ATE in the absence of essential heterogeneity.

ATE for everyone. Consequently TT = ATE. Furthermore, the error term in (4) reduces to U_0. So selection bias occurs only through the dependence of D with U_0 and an instrument satisfying Assumption 1 and the mean independence assumptions with respect to U_0, (*i.e.,* $E(U_0 \mid X = x, Z = z) = 0$) would identify this common treatment effect parameter.

(b) *Presence of unobserved heterogeneity but selection into treatment does not depend on unobserved gains (non-essential heterogeneity)*: This implies that $U_1 \neq U_0$, but D is independent of $(U_1 - U_0)$ given X. Individuals with the same X respond differently to treatment but do not select into treatments based on the idiosyncratic difference between treatment outcomes. For example, these conditions would arise when heterogeneity in health outcomes is entirely *ex post* and patients either cannot predict the differences in *ex post* outcomes or believe that these differences have mean zero and hence do not base their treatment decisions on them. This model is a version of the ``uncorrelated'' random coefficients model in traditional econometrics. Since the distribution of (U_1 - U_0) does not vary by treatment status of individuals, E(U_1 - U_0|X=x, D=1) = E(U_1 - U_0|X=x) =0; this implies TT = ATE. As before, the selection bias in estimating ATE arises due to dependence of D with U_0 only, and under the conventional IV assumptions, ATE is identified.

(c) *Presence of unobserved heterogeneity and selection into treatment is based on unobserved gains (essential heterogeneity)*: This implies that $U_1 \neq U_0$, and D and (U_1 - U_0) are not statistically independent even conditional on X. Not only do individuals with the same X respond differently to treatment, their treatment choices are influenced by their knowledge of their idiosyncratic gains: $(U_1 - U_0)$. This situation, designated as ***essential heterogeneity*** (Heckman et al., 2006), is perhaps the most relevant case for analyzing treatment choices in health economics. In this more general model, the robustness features of standard IV methods disappear. Specifically, even if a proposed instrument Z satisfies the mean independence assumption (Assumption 2), the IV method does not, in general, estimate ATE (or TT). The assumption required for IV to estimate ATE in the context of (4) is

$$E(U_0 + D(U_1 - U_0) \mid X = x, Z = z) = 0 .$$

However, when D and $(U_1 - U_0)$ are correlated, $E(U_0 + D(U_1 - U_0) \mid X = x, Z = z) \neq 0$ in general, even if Z is mean independent of the outcomes because $E(U_1 - U_0 \mid X = x, Z = z, D = 1) \neq 0$. That is, the instrumental variables may be independent of the idiosyncratic gains in the overall population, but conditional on those who select to receive treatment, they may no longer be independent of the idiosyncratic gains in this subgroup.[11]

[11] In general, TT is not identified either, unless further assumptions are imposed. See Heckman (1997).

An additional consideration in models with *essential heterogeneity* is that different instruments identify different parameters. This is a direct consequence of the underlying dependency of the IV estimates on the model for treatment choice *D* (see Heckman et. al., 2006). In this context, an *a priori* specification of the choice model for *D* becomes necessary for the interpretation of IV estimators. We illustrate and develop this point below.

Essential heterogeneity is likely to occur in the analysis of health care decisions since the choice of medical treatment is likely guided by individual idiosyncratic gains from alternative treatment. Unfortunately, this problem has received relatively little or no attention in the health economics and health services literature. In Section 6 we study the case of breast cancer treatment. Our results illustrate the problems associated with the IV approach and the advantages of using a more general framework allowing for essential heterogeneity when studying health care applications.

5. Model for Treatment Choice, Local Average Treatment Effect (LATE) and the Marginal Treatment Effect (MTE)

When responses to treatment are heterogeneous and individuals select into treatment anticipating these heterogeneous effects, a clear specification of the choice model for *D* becomes essential for interpreting what IV estimates. The model can be specified using standard (microeconomic) choice theory where a subject's decision to receive treatment is the result of the maximization of his expected benefit (or utility). In such a model, if the expected benefit of being treated is larger than the benefit the individual expects to get if he remains untreated then the patient will choose to undergo treatment. Let *V* denote the difference in benefits of being treated versus remaining untreated net of costs. Then, the model of treatment choice can be denoted as follows:

$$V = \mu_V(Z, X) + U_V \qquad E(U_V)=0 \qquad D = 1(V > 0), \quad (6)$$

where 1(.) is an indicator function such that *D* = 1 if the patient chooses treatment (so his net benefit of treatment, *V*, is positive) and *D* = 0 if the patient chooses to remain untreated (so his net benefit of treatment *V* is negative). (*Z*, *X*) and U_V are, respectively, observed and unobserved factors

[12] The alternative formulation of the choice model is

$$V = \mu_V(Z, X) - U'_V \qquad E(U'_V)=0 \qquad D = 1(V > 0),$$

where the subtraction of the error term U'_V makes *P*(*Z*) enter as an upper limit of the CDF for U_V in (6) and (7). However, most traditional econometric software packages fit (6) and not the model with a negative error term. This leads to a disjoint between theory and the application we will pursue later. Therefore, we describe the theory in terms of the traditional choice model.

determining choice of treatment. We assume that U_V is independent of Z and X, and $E(V \mid Z, X) = \mu_V(Z, X)$.[13,14]

The propensity score is the probability of selecting treatment:

$$P(z,x) = \Pr(D = 1 \mid Z = z, X = x) = \Pr(U_V > -\mu_V(z,x)) = 1 - F_{U_V}(-\mu_V(z,x)) \quad \textbf{(7)}$$

where F_{U_V} is the cumulative distribution of U_V . Therefore,

$$D = 1(V > 0) = 1(U_V > -\mu_V(z,x)) \Leftrightarrow 1(F_{U_V}(U_V) > F_{U_V}(-\mu_V(z,x))) \Leftrightarrow 1(F_{U_V}(U_V) > 1 - P(z,x))$$

(8)

where $F_{U_V}(U_V) = U_D \sim Uniform(0,1)$ by construction.[15] From hereon, we denote $S(z,x) = 1 - P(z,x)$.

Consider for simplicity the single instrument case, i.e. Z is a scalar rather than a vector of instruments. Given model (6) and the assumed independence of Z and U_V, changing Z externally from U_V , shifts all people in the same direction (towards or against D =1). This produces "monotonicity" in the sense of Imbens and Angrist (1994).

Using this framework, we can define LATE as the average treatment effect for individuals who would change their treatment choice when Z moves from z to z' . That is for two values of the instrument Z, z and z' , LATE is defined as the difference in observed outcomes between those with $Z = z$ and $Z = z'$, divided by the difference in the propensity to select treatment for $Z = z$ and $Z = z'$. Formally, assuming monotonicity, and defining D_z as the random variable corresponding to choices when Z is set to z, we have that

$$\text{LATE}(x,z,z') = \frac{E(Y \mid X = x, Z = z') - E(Y \mid X = x, Z = z)}{\Pr(D = 1 \mid Z = z', X = x) - \Pr(D = 1 \mid Z = z, X = x)}$$

$$= \frac{E(Y_1 - Y_0 \mid X = x, D_{z'} = 1, D_z = 0)}{\Pr(D = 1 \mid Z = z', X = x) - \Pr(D = 1 \mid Z = z, X = x)}, \qquad \textbf{(9)}$$

and thus, LATE computes the mean gain to those induced to switch from no treatment to treatment by a change in Z form z to z' . However, in (9), the subpopulation induced to change treatment is

[13] Note that Z and X may share common covariates. We can relax this assumption to make U_V independent of Z given X, but we can allow the dependence between U_V and X to be general.

[14] The additive and separable representation of V is a common assumption used in the literature. Heckman and Vytlacil (2007) and Heckman et al (2006) analyze the consequences of weakening it.

[15] Since $E(D \mid X - x, Z = z) = P(x,z)$

not clearly identified since, in the absence on an underlying choice model, the relevant margin at which this change in behavior is taking place is not specified.[16]

Progress on this problem was made by Heckman and Vytlacil (1999, 2001a, 2007) and Vytlacil (2002), who develop the interpretation of LATE in terms of the choice model in (6). Their interpretation is illustrated in Figure 1, which presents a stylized profile of mean conditional treatment effects over U_D. Recall that U_D represents the unobserved characteristics that determine treatment. Once we condition on the observed factors X and the unobserved U_D, the conditional mean treatment effects $E(\Delta|X=x,\ U_D=u_D)$ are exactly the same for each individual with the same value of $U_D = u_D$, despite having different values of Z (or $P(Z,X)$). For any value of the instrument $Z = z$ (and $X = x$), the patients for whom $U_D > S(z,x)$ receive treatment while patients with $U_D \leq S(z,x)$ remain untreated (see Figure 1a). In addition, notice that the expected value of the observed outcome for this group of patients can be written as the weighted average of those who receive treatment and those who do not:

$$E(Y \mid Z = z, X = x) =$$

$$\Pr(D = 1 \mid Z = z, X = x) \cdot E(Y_1 \mid D = 1, Z = z, X = x) + \Pr(D = 0 \mid Z = z, X = x) \cdot E(Y_0 \mid D = 0, Z = z, X = x) =$$
$$\Pr(U_D > S(z,x)) \cdot E(Y_1 \mid U_D > S(z,x), x) + \Pr(U_D \leq S(z,x)) \cdot E(Y_0 \mid U_D \leq S(z,x), x) \qquad \textbf{(10)}$$

By the definition of an instrument (Assumption 1), we can vary the value of $Z = z$ (given $X=x$), and therefore $P(Z = z, X = x) = P(z,x)$ and hence, $S(Z = z, X = x) = S(z,x)$, non-trivially with respect to the distribution of U_D. Thus, consider two groups of patients, one with $Z = z$ and the other with $Z = z'$ from the same distribution of U_D. Let $S(z,x) \geq S(z',x)$ for every patient. Using expression (10), we have that the difference in the observed outcomes between these two groups of patients is then,

$$E(Y \mid z,x) - E(Y \mid z',x)$$

$$= \left[\Pr(U_D > S(z,x)) \cdot E(Y_1 \mid U_D > S(z,x), x) + \Pr(U_D \leq S(z,x)) \cdot E(Y_0 \mid U_D \leq S(z,x), x)\right]$$

$$-\left[\Pr(U_D > S(z',x)) \cdot E(Y_1 \mid U_D > S(z',x), x) + \Pr(U_D \leq S(z',x)) \cdot E(Y_0 \mid U_D \leq S(z',x), x)\right]$$

$$= \Pr(S(z',x) < U_D < S(z,x)) \cdot \left[E(Y_1 \mid S(z',x) < U_D < S(z,x), x) - E(Y_0 \mid S(z',x) < U_D < S(z,x), x)\right]$$

[16] The interpretation of LATE is even more limited in the case of multiple instruments. When Z is a vector of instruments, the "monotonicity" assumption does not necessarily hold and the second equality in (9) can break down (Heckman et al. 2006).

$$= \Pr(P(z,x) < U_D < P(z',x)) \cdot \left[E(Y_1 \mid S(z',x) < U_D < S(z,x), x) - E(Y_0 \mid S(z',x) < U_D < S(z,x), x) \right], \quad (11)$$

where the last two equalities follow from the fact that U_D ~*Uniform(0,1)*. This is illustrated in Figure 1b, where the mean potential outcomes outside the limits of the margin cancel out. Thus, combining (9) and (11), we can conclude that LATE identifies the average effect for a group of patients who are within the margin defined by $S(z,x)$ and $S(z',x)$ (Heckman and Vytlacil, 1999, 2007):

$$\text{LATE}(x,z,z') = E(Y_1 - Y_0 \mid X = x, S(z',x) < U_D < S(z,x)). \quad (12)$$

LATE is often referred in the health literature as the treatment effect for the marginal patients (McClellan et al, 1994, Brooks et al, 2003). The marginal patients are defined as the subset of patients whose treatment choices varies with the instrument. Imbens and Angrist (1994) define the LATE parameter from hypothetical manipulation of the choice probability or values for the instrument. Heckman and Vytlacil (1999, 2007) draw on choice theory and derive LATE (and also other treatment effect parameters, as explained below) in the context of the generalized Roy Model (Roy, 1951; Heckman and Sedlacek, 1985). Relating IV to choice models helps to identify the margin of U_D selected by instruments. IV, working through $S(Z,X)$, selects different slices of U_D and defines mean treatment effects for those slices.

In a model with a scalar and binary instrument with only two points in the support of $P(Z,X)$, the IV estimate and the overall LATE estimate are the same. When there are more than two distinct values of Z, an overall LATE (the standard IV estimator) can be estimated by a weighted average of the pairwise LATE parameters based on ordered values of the scalar instrument Z (Yitzhaki, 1989, Imbens and Angrist, 1994). However, Heckman et al. (2006) showed that when vector instruments enter the choice model, the traditional IV method may produce misleading inferences since the IV estimate can be negative even if all the pairwise LATE estimates are positive. This is because the weights used to compute the overall LATE can be negative if the choice model is determined by a vector of instruments and the analyst uses only some of those instruments in the calculations. This point is illustrated in our empirical example.

LATE is an interpretable parameter when the observed variation in the instrument defines the question for which the analyst seeks an answer, *e.g.,* if the analyst has access to an instrument, Z, that takes two values (z_1 and z_2) and the question he seeks to answer is precisely what happens when the instrument is changed from z_1 to z_2. However, when the policy being analyzed does not conform closely to the instrument used, it is not always clear who the marginal patients associated

with the policy are, and consequently, whether or not the marginal patients defined by LATE are those on which the clinical decision making should rely.

In order to address some of these limitations and to better understand the distribution of treatment effects in the population, we can use the Marginal Treatment Effect (MTE) first introduced by Björklund and Moffitt, 1987 (see also Heckman, 1997; Heckman and Smith, 1998; Heckman and Vytlacil, 1999, 2000, 2001a). The MTE is the average gain to patients who are indifferent between receiving *treatment 1* versus *treatment 0* given *X* and *Z*. These are the patients at the margin as defined by *X* and *Z*. Formally, MTE can be defined as:

$$\begin{aligned}\text{MTE}(x,z) &= E(\Delta \mid X=x, Z=z, V=0) = E(\Delta \mid X=x, U_V = -\mu_V(z,x)) \\ &= \mu_1(x) - \mu_0(x) + E(U_1 - U_0 \mid U_V = -\mu_V(z,x)) \\ &= \mu_1(x) - \mu_0(x) + E(U_1 - U_0 \mid U_D = S(z,x)), \qquad \textbf{(13)}\end{aligned}$$

where the last equality follows from the fact that $S(Z,X)$ is a monotonic transformation of the mean utility $\mu_V(Z,X)$ while U_D is a monotonic function of U_V. In Figure 1, the mean conditional treatment effect at each level of U_D is the value of the MTE at that level of U_D. Evaluation of the MTE parameter at low values of U_D averages the outcome gain for those individuals whose unobservable characteristics make them less likely to undergo treatment, while evaluation of MTE parameter at high values of U_D gives the gain for those patients with unobservable characteristics which make them more likely to undergo treatment. For example, LATE is a weighted sum of all MTE within the margin at which LATE is identified (Figure 1 (b)). In the limit, as $\mu_V(z',x) \to \mu_V(z,x)$, LATE converges to MTE under standard regularity conditions.

Treatment Parameters and IV Effects as Weighted Averages of MTE

An additional feature of MTE is that all mean treatment effects parameters, including the ATE, TT, OLS, and the IV effect, can be calculated from weighted averages of MTE. These weights can be obtained from the data (Heckman and Vytlacil, 2007; and Heckman et al., 2006).[17] For example, the ATE is the sum of all MTE across all distinct values of U_D, weighted equally (conditional on X). A more formal description of these weights is given below.

Equation (13) shows that the MTE is identified on the support of $S(Z,X)$. For notational convenience, from hereon we define $u_D \equiv S(z,x)$, i.e., u_D is the value of $S(Z,X)$ used to define the

[17] A software designed to calculate these weights can be found at http://jenni.uchicago.edu/underiv

margin of indifference in (13). Using this notation, an average value for the MTE at each level of U_D can be obtained by integrating over the distribution of X conditional on U_D=u_D That is,

$$\begin{aligned} MTE(u_D) &= E_{X|U_D=u_D}(MTE(x,u_D)) \\ &= E_{X|U_D=u_D}\{(\mu_1(x)-\mu_0(x))+E(U_1-U_0 \mid U_D=u_D)\}. \end{aligned} \qquad \textbf{(14)}$$

Additionally, by integrating these average MTEs over the distribution of U_D (which by construction is *Uniform(0,1)*) we can obtain the (unconditional) Average Treatment Effect:

$$\begin{aligned} ATE &= E_{U_D}(MTE(u_D)) \\ &= E_{U_D}E_{X|U_D=u_D}(\mu_1(x)-\mu_0(x)+E(U_1-U_0 \mid U_D=u_D). \end{aligned} \qquad \textbf{(15)}$$

Here, the last term in (15) drops out because $E_{U_D}E(U_1-U_0 \mid U_D=u_D)=E\left(U_1-U_0\right)=0$. Equation (15) suggests that the *weights* for the MTE(x, u_D) that yield the ATE can be constructed from the empirical joint distribution of (*X, S(Z,X)*) directly. Alternatively, since U_D is distributed as *Uniform(0,1)*, simply integrating MTE(u_D) over the full support of U_D yields ATE.

Obtaining the weights to estimate TT and the IV estimator is a bit more complicated than determining the weights for ATE, but they can be computed readily using the data at hand. Intuitively, for TT, the weights for MTE evaluated at high values of U_D are relatively larger than those evaluated at low values of U_D. This is because, by definition, larger values of U_D represent greater propensity to select treatment based on unobserved characteristics. The TT weights can be written as (Heckman et al., 2006):

$$\varpi_{TT}(x,u_D)=\frac{\Pr(S(Z)\le u_D \mid X=x,U_D=u_D)}{\int\int \Pr(D=1 \mid X=x,U_D=u_D)du_D dF(X)}. \qquad \textbf{(16)}$$

The weights for an IV estimator relate the IV estimate to the distribution of MTE. Consider the case of a general scalar function of instrumental variables, $J(Z)$, that is used to identify the IV effect.[18] The IV estimator, conditional on X, is given by:

$$\frac{Cov(J(Z),Y \mid X=x)}{Cov(J(Z),D \mid X=x)}. \qquad \textbf{(17)}$$

Without loss of generality, we center J around its mean so that $E(\tilde{J}(Z) \mid X=x)=0$, where $\tilde{J}(Z)=J(Z)-E(J(Z) \mid X)$. Substituting the model for Y from (3) we have,

[18] That is $Cov(J(Z), D \mid X = x) \neq 0$. The propensity score, $P(Z)$ is one such function, but not necessarily the only one available. In fact if $J(Z)$ is monotonically related to $P(Z)$, then the joint distribution of ($J(Z)$, $P(Z)$) collapses to a single distribution: that of $P(Z)$ (See Heckman et al, 2006).

$$Cov(\tilde{J}(Z), Y \mid X = x) = E(\tilde{J}(Z) \cdot [Y_0 + D \cdot (Y_1 - Y_0)] \mid X = x)$$

$$= \int_{U_D} E(\tilde{J}(Z) \cdot D \mid X = x, u_D) \cdot E(Y_1 - Y_0 \mid u_D, X = x) du_D$$

$$= \int_{U_D} \left\{ E(\tilde{J}(Z) \mid X = x, S(Z, X) \le u_D) \cdot \Pr(S(Z, x) \le u_D) \mid X = x \right\} \cdot MTE(x, u_D) du_D \,, \qquad \textbf{(18)}$$

where we use the fact that, conditional on X=x and U_D= u_D, $J(Z) \cdot D$ is independent of Y_0 and $D \cdot (Y_1 - Y_0)$ (second equality), and the definition of MTE(x, u_D) in (13) (third equality). Therefore, the weights that relate the MTE(x, u_D) to the IV estimate are:[19]

$$\varpi_{IV}(x, u_D) = \frac{\left\{ E(\tilde{J}(Z) \mid X = x, S(Z, x) \le u_D) \cdot \Pr(S(Z, x) \le u_D \mid X = x) \right\}}{Cov(\tilde{J}(Z), D \mid X = x)} . \qquad \textbf{(19)}$$

Like the TT weights, these weights also integrate to 1 over the support of {X, U_D}. Note however, that $E(\tilde{J}(Z) \mid X = x, S(Z, x) \le u_D)$ can be negative even if the weights add up to 1 over the support of {X, U_D}. Generally, if $E(\tilde{J}(Z) \mid X = x, S(Z, x) \le u_D)$ is weakly monotonic in u_D, then the weights are always positive.[20]

Method of Local Instrumental Variables

The method of local instrumental variables (LIV) can be used to estimate the MTE (Heckman, 1997; Heckman and Vytlacil, 1999, 2000; Heckman, 2001, Vytlacil, 2002; Heckman et al., 2006). LIV identifies the MTE over the support of the propensity score. To see why, consider that

$$E(Y \mid Z = z, X = x) = E(DY_1 + (1 - D)Y_0 \mid Z = z, X = x)$$

$$= E(Y_0 \mid X = x) + E(D(Y_1 - Y_0) \mid Z = z, X = x)$$

$$= E(Y_0 \mid X = x) + E((Y_1 - Y_0) \mid D = 1, X = x) \Pr(D = 1 \mid Z = z, X = x)$$

$$= E(Y_0 \mid X = x) + \int_{S(z,x)=1-P(z,x)}^{1} E((Y_1 - Y_0) \mid U_D = u, X = x) du \,.$$

Now, if we take the rate of change of the mean outcome with respect to the probability of receiving treatment evaluated at a particular value of $S(z,x)$ =1 -$P(z,x)$:

[19] See Heckman et al (2006) for a detailed derivation of these weights.
[20] This is violated when $J(Z)$ and $P(Z)$ are not perfectly dependent so that increase in $J(Z)$ may lead to decrease in $P(Z)$ over some values of Z while the opposite is true over other values of Z.

$$\frac{\partial}{\partial P(z,x)} E(Y \mid Z=z, X=x)\Big|_{1-P(z,x)=u_D} = E((Y_1 - Y_0) \mid X=x, u_D = 1-P(z,x)) = \text{MTE}(x, u_D)\,. \quad \textbf{(20)}$$

Expression (20) shows that the LIV estimand, (the derivative of $E(Y \mid Z=z, X=x)$), identifies the marginal treatment effect given $U_D=u_D$ (≡ 1 - *P(z,x)*) (Heckman and Vytlacil, 1999). Once MTE is estimated via LIV, the other mean treatment effect parameters can also be estimated using different weighted averages of the estimated MTE.

Estimation of MTE and ATE using the method of LIV

In practice, there are different strategies for estimating the MTE using the logic developed in the previous section. In this paper, we consider the following simple and intuitive strategy. Since, $Y=DY_1 + (1-D)Y_0$, we have

$$\begin{aligned} &E(Y \mid X=x, P(Z,X)=P(z,x)) = E(DY_1 + (1-D)Y_0 \mid X=x, P(Z,X)=P(z,x)) \\ &= \mu_0(x) + P(z,x)\cdot(\mu_1(x)-\mu_0(x)) + E(U_0 \mid P(Z,X)=P(z,x)) \\ &\qquad + P(z,x)\cdot E(U_1 - U_0 \mid P(Z,X)=P(z,x), D=1) \\ &= \mu_0(x) + P(z,x)\cdot(\mu_1(x)-\mu_0(x)) + K(P(z,x))\,, \end{aligned} \quad \textbf{(21)}$$

where $K(P(z,x))$ is a general function of the propensity score *P(z,x)*. Then, following (20), the LIV estimator is given by

$$\frac{\partial E(Y \mid X=x, P(z,x))}{\partial P(z,x)}\Big|_{1-P(z,x)=u_D} = (\mu_1(x)-\mu_0(x)) + \frac{\partial K(P(z,x))}{\partial P(z,x)} = \text{MTE}(x, u_D). \quad \textbf{(22)}$$

From this last expression we observe that the key element for the estimation of MTE is the function $K(P(z,x))$. This function can be estimated using different econometric techniques. In our empirical analysis we use a flexible approximation to $K(P(z,x))$ based on a polynomial on the propensity score. [21] Specifically, equation (21) is implemented by regressing the outcome *Y* on all covariates, the propensity score, the interaction of the propensity score with all covariates, and a polynomial on the propensity score. The MTE is then directly computed following equation (22).

Notice also that equation (22) provides the intuition for the simple test that Heckman et al (2006) propose to determine whether an analyst can safely ignore the complications induced by heterogeneity and self-selection. Specifically, if $\partial K(P(z,x))/\partial P(z,x)$ does not vary with $P(z,x)$ (i.e. if

[21] The polynomial in the propensity score can be interpreted as a semi-parametric method for approximating the *K(P(z,x))* function (see, e.g., Powell, 1994). Alternatively, this regression, and in particular *K(P(z,x))* can be estimated non-parametrically using, for example, local linear regression or splines.

linearity of outcomes in the probability of selection is accepted) the MTE is constant across all margins of selection and the constant treatment effect model suffices to explain the data.

If the test of linearity fails, then one needs to account for the fact that the MTE is not constant in the population. One can then weight the MTE differentially to obtain various treatment effect parameters as explained in the previous section and by Heckman and Vytlacil (2005, 2007b).

6. A Health-Care Application

We demonstrate the importance of essential heterogeneity in a health-care application. Specifically, we analyze the effect on 5-year medical costs of breast-conserving surgery with radiation (BCSRT) compared to mastectomy (MST) in patients with breast cancer. In this sample, all patients undergo either BCSRT or MST. Throughout this section, we refer to BCSRT as the treated state (corresponding to Y_1) and MST as the untreated state (corresponding to Y_0) and all treatment effect parameters are defined as the differences in outcomes between BCSRT and MST ($Y_1 - Y_0$).

Our data come from the OPTIONS (Outcomes and Preferences in Older Women, Nationwide Survey) project (Hadley et al, 1992). The OPTIONS sample was designed to be representative of all female elderly Medicare beneficiaries (aged 67 or older) with newly diagnosed, early-stage breast cancer in Medicare's-fee-for service program between 1992 and 1994. The dataset was constructed in four steps: 1) Medicare claims for persons with a breast cancer diagnosis or relevant surgery procedure codes for calendar years 1992 to 1994 were obtained. 2) Additional exclusions were applied so that the sample was limited to women for whom BCSRT and MST would be considered equivalent from the clinical point of view. Cases for which breast cancer was not the primary diagnosis were deleted. 3) Surgeons identified in the dataset were surveyed to verify study eligibility of the patients based on the presence of primary stage I and II invasive disease and the absence of the preceding exclusion criteria (as in (2)). 4) Additional conditions were applied to exclude patients who were in a Medicare health maintenance organization in the month of the survey because their cost data were not available in the claims file. Finally, patients who had breast-conservation surgery but did not receive radiation due to unidentified reasons were excluded (17%). Further details of the specific exclusion criteria used can be found elsewhere (Hadley et al., 2003; Polsky et al., 2003). This data provides a unique opportunity to analyze a large national sample of Medicare beneficiaries with confirmed local stage of breast cancer. Moreover, we choose this data for comparability to results published in the literature based on this dataset (Hadley et al., 2003; Polsky et al., 2003).

All 5-year Medicare payments from inpatient, outpatient, and physician Part-B claims were used to estimate direct medical costs, including costs related to breast cancer treatment and all other medial costs covered by Medicare. Total costs were calculated using an annual 3% discount rate. The final sample consists of 2,517 patients of whom 1,813 patients had a MST and the remaining had BCSRT. The distribution of patient characteristics by treatment type is published elsewhere (Polsky et al., 2003).

The covariates that we control for are variables that are both measurable and theoretically predictive of costs. In addition to the treatment indicator, we include age at the time of surgery, cancer stage, Charlson co-morbidity index, patient-specific Medicare payments in the year before surgery categorized into 5 groups, and race. Because claims do not contain socioeconomic data, we used percentage of college graduates, median household income and percentage below poverty level by 5-digit zip-code level of the women's residence as proxies for socioeconomic status. Additionally, we adjusted for county-level data on health system characteristics, such as hospital admissions, number of nursing homes and an indicator for urban area.

The primary goal of the analysis is to estimate the average treatment effect parameter (ATE) and the treatment effect on the treated (TT) parameter on total costs associated with BCSRT as compared to MST. Additionally, we compare these parameters with what IV estimates.[22] The variables used as valid instruments include a regional dummy variable (NORTH) to represent regional variations in practice patterns, and a continuous variable that represents the Medicare physician fee differential (FEEDIF) between mastectomy and breast conserving surgery calculated at the 3-digit zip-code level of the treating physician. NORTH represented a geographical area plausibly independent of underlying health but appeared to influence treatment, perhaps through a historical practice style effect. In particular, women residing in the Northeast, Midwest and Pacific census divisions (represented by indicator NORTH) were more likely to receive BCSRT compared to MST. Medicare fees are assumed to be exogenous and independent of unobservable health of patients and preferences of patients and physicians because they were determined by a combination of the resource-based fee specified by the Medicare Fee Schedule, which is independent of any particular physician's or patient's characteristics, and the average historical Medicare payment in the geographic area. Further details and justification for these instruments are available in Hadley et al. (2002).

[22] In this empirical application, we use a linear and additive separable specification to model cost which is highly skewed. This is a limitation of the present analysis and we discuss this issue further below. The primary reason for using this specification is the comparability of our results to previously published work with this data (Hadley et al, 2003; Polsky and Basu, 2006). Moreover, this functional form passes all the goodness of fit tests that are used to identify systematic biases in modeling the mean function of costs data (Basu et al., 2004; Manning et al., 2006).

We first present empirical evidence on the presence of essential heterogeneity in our application using the test proposed by Heckman et al. (2006). We then use the local instrumental variable estimator to compute the marginal treatment effect (MTE) as well as ATE and TT. Finally, we compare these treatment parameters with the IV *effects* estimated using NORTH and FEEDIF as instruments.[23]

Testing for Essential Heterogeneity and the Local Instrumental Variable Estimator

First, we estimate the propensity score of choosing treatment as a function of all covariates and also the instruments NORTH and FEEDIF. Both instrumental variables are significant predictors of treatment choice ($p < 0.0001$ for each).[24]

Figure 2(a) illustrates the distribution of the predicted propensity score for choosing BCSRT separately for patients who chose BCSRT and those who chose MST. It also depicts the identified support where we find positive density of the propensity score for both treatment sub-samples. We cannot identify MTE over the entire (0,1) support.[25] Although we do find people near 0, there is essentially no mass close to 1. This means (unconditional) ATE is not identified in the sample without further assumptions (Heckman and Vytlacil, 1999, 2007).

Using the test of linearity proposed by Heckman et al. (2006), we explore the assumption of a constant treatment effect (i.e. the absence of essential heterogeneity). In this test (following equation (22)), we regress 5-year costs on all covariates, the propensity score, the interaction of propensity score with all covariates, and a polynomial on the propensity score. The joint test of significance for the polynomial coefficients reveals whether there is essential heterogeneity in treatment effects. Table 1 presents these results and shows that we find that the cubic terms of the polynomial are jointly significant.[26] This indicates that there is strong evidence of self-selection based on heterogeneous and unobserved gains in the population and, consequently, that standard IV estimates are not necessarily informative on ATE or TT.

---

[23] All analyses are done in Stata® 9.0 and Gauss®. Estimates of MTE and weights are validated against the MTE-software developed in Heckman et al. (2006).

[24] From our choice model we obtain $\hat{\alpha}_{NORTH}$ =0.279 (0.065) with associated p-value< 0.001 and $\hat{\alpha}_{FEEDIF}$ = 0.002 (0.0005) with associated p-value < 0.001.

[25] Note that although one of the instruments employed is a binary indicator (NORTH), we are able to estimate a (almost) continuous support of the propensity score using the other X's in the choice equation.

[26] The statistical test on essential heterogeneity is carried out using bootstrapping techniques.

In light of the evidence on essential heterogeneity, we apply the LIV estimator to recover the MTE over the distribution of U_D. We use the derivative of the cubic formulation as our LIV estimand. [27] The predicted values of the propensity score allow us to define the values of u_D over which MTE can be identified (Heckman and Vytlacil, 2001a). The larger the support of the propensity score, the bigger the set over which MTE can be recovered. [28]

First we obtain estimates of MTE(x,u_D). Since X represents a vector of covariates in this application, we reduce its dimensionality by using demi-deciles (or twentiles) of the linear predictor $X'\beta$ in the equation (22). [29] We denote these demi-deciles as η_q hereon, where q = 1,2,…,20. Thus, using our coefficient estimates from the above regression (equation (22) under a cubic polynomial formulation for *K(P(z,x))*) we estimate MTE(η_q,u_D) by using average $x'\hat{\beta}$ for each η_q and varying u_D between 0 and 1. Note that MTE estimates using a value of $P(Z,X)$ = p are associated with u_D = (1 – p). The MTE(η_q,u_D) is shown in Figure 2 (b). The empirical joint density of (η_q,u_D) is shown in Figure 2 (c) and it also represents the weights for MTE(η_q,u_D) required to calculate the *empirical* ATE (estimated over the observed common support).

To obtain the unconditional MTE(u_D), we integrate MTE(η_q,u_D) over the empirical distribution of η_q at each value of u_D. The MTE(u_D) of BCSRT over MST is displayed in Figure 2(d). Standard errors for MTE(u_D) at each of these points are estimated via 500 bootstrap replicates. Figure 2(d) shows that for high values of u_D (representing patients with latent characteristics that make then most likely to choose BCSRT compared to MST), the MTE is significantly positive. Patients most likely to choose BCSRT incur significantly higher costs if they are given BCSRT rather than MST. We discuss the implication of this result later in this section. This effect disappears for the middle values of u_D, where MTE is not significantly different from zero and then becomes significant again at lower values of u_D. Since we could not identify the higher end of the support for P(Z,X), we could not estimate MTE for the patients least likely to choose BCSRT (i.e. low values of u_D) based on their unobserved characteristics. [30]

The Treatment Parameters

[27] We also considered a quartic specification for our LIV estimand. Although the quartic specification was significant with the F-test (p=0.018), it did not pass the likelihood-ratio test (p=0.095).

[28] With parametric approaches, assumptions about functional form can estimate MTE over ranges of u_D that are not identified with our choice model and sample. This is not the case when non-parametric techniques are used instead.

[29] Note that β corresponds to the coefficients on the interaction term of X and $P(Z,X)$ in the LIV estimand.

[30] The results are similar when only NORTH is included in the choice model. On the contrary, when only FEEDIF is used as an instrument we do not find evidence of essential heterogeneity. However, it is misleading to infer from this result that essential heterogeneity is absent based on this result. This is because Heckman et al.'s test relies on the correct specification of the choice model; that is it relies on the analyst incorporating all factors that necessarily affect choice of treatment. Thus, by omitting the IV NORTH, we are eliminating a major source of variation in choice. Consequently, it may be always appropriate to use the LIV methods with multiple instruments unless there is a strong theoretical evidence for one factor affecting choice.

Next, following equations (15) and (16), we calculate the weights associated with ATE and TT. These are shown in Figure 2 (c) and 2 (e). As is intuitively reasonable, the TT weights are larger for higher values of u_D than for lower values. The ATE weights are simply the empirical joint density of (η_q, u_D).

Estimates of the mean treatment effects parameters, ATE and TT, are reported in Table 2. Even though we do not recover the whole support of propensity score, the TT parameter is still interpretable as an *empirical* TT estimate that is conditional on those individuals we observe selecting treatment. The TT (standard error in parenthesis) is estimated to be \$52,329 (15,496).

Likewise, the *empirical* ATE parameter is $\$41{,}493\ (14{,}594)$. However, in order to interpret the ATE parameter as the average effect of BCSRT compared to MST we require the full support of the propensity score. Therefore, in the absence of full support, estimation entails some form of extrapolation. We extrapolate MTE(u_D) over the missing range of u_D, to obtain an ATE estimate of \$88,975 (35,221), that is statistically significant.[31] Note that ATE is estimated to be larger than TT primarily due to the extrapolated portion of the MTE at low values of u_D, and therefore should be interpreted cautiously. In fact, for this application it is important to consider whether the ATE provides any policy relevant estimate. Notice that the ATE estimate is significantly different from the TT estimate at the 5% level.

Interpreting the IV Estimator under Heterogeneous Treatment Effects

As shown above, the IV estimator can be represented as a weighted average of the marginal treatment effect where the weighting scheme depends on the instrument considered in the analysis as well as on the specification of the choice model. In particular, depending on the relationship between the instrument and the rest of the variables included in the choice model, the IV weights can be positive, negative or zero (regardless of the shape of the MTE). This affects the way the IV estimate is interpreted.

Our analysis considers only two potential instruments: NORTH and FEEDIF. In our sample, these two variables are negatively correlated (correlation coefficient = -0.120; $p < 0.001$). This means that if we compare patients with NORTH = 0 to NORTH =1, average FEEDIF will be lower for the later group. Consequently, moving from NORTH = 0 to NORTH =1 may increase the average propensity to choose BCSRT for some patients while *it may decrease* it for others if the decrease in FEEDIF more than compensates for the effect of NORTH.[32] Thus, if we are looking at

[31] Since u_D is uniformly distributed the computation of ATE from the extrapolated MTE(u_D) is valid. The same logic cannot be applied to the case of TT.
[32] This is because the coefficients on both NORTH and FEEDIF are positive in the choice model. See footnote 24.

the IV estimate of NORTH or FEEDIF separately, where the correct choice model contains both, the IV weights (following (19)) may be negative over some ranges of (η_q, u_D).

This is illustrated in Figures 3, where MTE(u_D) along with the IV weights corresponding to different specifications of $J(Z)$ are presented for two selected demi-deciles of $X'\hat{\beta}$. Figure 3 reveals the negative IV weights over some ranges of u_D where either NORTH or FEEDIF is used as the instrumental variable but both enter the choice model. This exercise shows that the individual IV estimates arising from such an analysis should be cautiously interpreted, since they put negative weight over some ranges of (η_q, u_D) where MTE(η_q, u_D) is positive. On the contrary, and as predicted, the overall IV weights, calculated using the entire linear index of the choice model, are always non-negative (see Figure 3). In fact, these weights were non-negative for every combination of (η_q, u_D).

The weighted averages of MTE, weighted by the corresponding IV weights are shown in Table 2. In order to illustrate our main points, we focus on the NORTH variable. The IV effect of NORTH, using the MTE approach, is estimated to be \$29,580 (15,997) when the choice model contains both NORTH and FEEDIF. Notice that this IV estimate of NORTH is a Local Average Treatment effect (LATE) since the regional variation is an indicator variable. In comparison, the traditional linear IV estimator using NORTH estimates the effect of BCSRT compared to MST at \$44,941 (21,790). [33] These estimates do not tally with each other because our MTE based estimate depends on the specification of the choice model which contains both NORTH and FEEDIF as instruments, and due to the covariance between these two instruments, the MTE-based IV estimate is calculated using negative weights over some regions. Our analysis illustrates that caution should be exercised when interpreting LATE estimates of individual IV variables when the underlying choice model contains multiple instruments.

In contrast, when we use the entire linear index of the choice model as the instrument, the overall IV effect is estimated to be \$33,708 (13,213), which lies close to the corresponding linear IV estimate using both NORTH and FEEDIF. [34] Had we specified our choice model using only NORTH as the instrument, IV weights would have been non-negative across all ranges of (η_q, u_D), and the MTE approach estimate of the IV effect of NORTH would be \$43,997 (22,437), an estimate that is

[33] The linear IV estimates tally with those reported by Hadley et al. (2003) and by Polsky and Basu (2006).

[34] Notice that the IV estimator computed using the weighted average of the MTE differs slightly from the linear IV estimand. This is due to the parametric approximation used in the estimation of the MTE and the associated IV weights. Since the main point of our analysis is to illustrate the importance of accounting for the presence of essential heterogeneity, we consider these differences negligible. Furthermore, given our sample size, the implementation of non-parametric techniques appears as an extremely costly alternative from an efficiency point of view. Another source of difference is that we have used a probit model to generate the propensity score, whereas the linear IV method follows a two-stage least squares (TSLS) approach and uses a linear probability model in the first stage.

again very similar to the linear IV estimate of $44,941 (21,790) based on NORTH only (Table 2). [35] Table 2 also describes the identifying formula for each estimate in order to clarify the differences between the estimates.

Margin of choice affected by an instrument

Following the discussion in Section 5, we interpret the IV estimand for NORTH as the average treatment effect for individuals who would change their treatment choice when moving from a non-North region to the North region. To further understand the margin at which this effect is estimated, we use our estimated choice model on treatment choices as a function of all observed covariates in the outcomes equation and the NORTH and FEEDIF variable. First we compute the propensity score when we "turn" the NORTH variable off and on for each individual, say, $\hat{p}(0)$ and $\hat{p}(1)$, respectively. Next, within each demi-decile, η_q, and at a specific value of U_D = u_D we find whether any subject would change their treatment status as a response to a change in the value of the instrument (NORTH). That is, we check whether there are subjects for whom $\hat{p}(0) \leq 1 - u_D$, but $\hat{p}(1) > 1 - u_D$. If there are such subjects, then the IV NORTH is able to identify a treatment effect defined by that (η_q, u_D) margin. Figure 4 illustrates that margins identified by the IV NORTH and these are the same margins over which we estimate non-zero IV weights (Figure 3(a)) for NORTH.[36]

Discussion of results

Applying the method of LIV to estimate the causal treatment effect on 5-year medical costs for breast-conserving surgery versus mastectomy, we find significant evidence of essential heterogeneity in the data. The main empirical finding of our analysis is that the treatment effect on 5-year medical costs for patients whose unobserved characteristics make them most likely to receive BCSRT is significantly positive. The MTE evaluated at large values of u_D rises to about $200,000, and it represents the marginal effect for the group of patients whose unobserved characteristics make them most likely to select BCSRT. One could imagine that these are patients who may have strong preference for the aesthetics of surgery outcomes (Nold et al., 2000). However, the MTE estimated is for the marginal patients who share similarly strong preferences for the aesthetics but are still indifferent to selecting between BCSRT and MST given U_D = u_D (=1 –

[35] Similarly, the estimated IV *effect* of FEEDIF when both NORTH and FEEDIF enter the choice model is $25,381 (14,653) (see Table 2). This is very different from the linear IV estimate of $12,884 (32,081) using only FEEDIF. When we specify our choice model to depend on FEEDIF only, then the MTE based approach produces an IV estimate of $13,170 (26,495), again very close to the linear IV estimate.

36 The margin identified by FEEDIF can be constructed using the same logic. However, this entails a cumbersome analysis since FEEDIF is a continuous variable.

$p(z)$) and η_q. This indicates that the observed characteristics for these patients, such as clinical stage of cancer for example, may be such that they would discourage them to choosing BCSRT. This is further supported in Figure 5 where we plot the proportion of patients with advanced (stages B and C) cancer. We find that for large values of u_D almost all marginal patients have advanced cancer, where BCSRT is clinically less attractive (Veronesi et al. 2002). Therefore, our results indicates that treating these patients with BCSRT would lead to much higher costs than using MST, possibly due to complication and recurrent cancer. In fact, we confirm that this large treatment effect is driven mostly by costs incurred beyond the first year of receiving BCSRT or MST.

Similarly, patients who react with great concern to the presence of cancer might wish to undergo MST even when their clinical conditions make them candidates for BCSRT. For such patients, whose unobserved preferences make them more likely to select MST, getting BCSRT treatment may lead to additional healthcare utilization generated by their anxiousness. This may be one of the reasons we see that the MTE is also increasing for lower values of u_D. In fact, we do not observe anyone getting BCSRT if their unobserved characteristic make them strongly likely to get MST. This conforms to the observation by Nold et al. (2000) that, "if a women wants to have MRM-NR (a form of MST) even when she is a candidate for BCS (same as BCSRT), a surgeon's input is overshadowed by the fear of cancer."

Our results have strong implications for the use of the appropriate mean treatment effect parameter for policy decisions. Even though ATE, estimated over the whole support of U_D, would traditionally be used to evaluate MST versus BCSRT in cost-effectiveness analysis, the fact that no patients choose BCSRT if they have strong preference for MST implies that efficient policy decisions regarding coverage on BCSRT should focus on TT instead of ATE. In our example, the TT is lower than the ATE, thereby relatively favoring the use of BCSRT compared to MST.

7. Conclusion

Instrumental variables (IV) analysis is the most commonly used econometric method for addressing selection biases. However, the recent econometric literature has demonstrated several limitations of the traditional IV approach. For one, an IV estimate of ATE can be interpreted as the parameter that would be produced from a well-designed randomized clinical trial only under the assumption that treatment effects are the same for every one in the population or that, even if treatment effects are heterogeneous, patients or their physicians do not have (or use) any additional information to select into treatment beyond what the analyst of an observational data

possesses. Such assumptions are clearly a stretch for modeling treatment choices in health care, especially under the practical limitations of observational data to collect all relevant information pertaining to treatment choices.

When these assumptions are relaxed and subjects are allowed to self-select into treatment based on their idiosyncratic gains, the standard IV methods identify parameters that only reflect the treatment effects for a group of patients who would change their treatment status in responses to changes in the levels of the instrumental variable (Imbens and Angrist, 1994; and Heckman et al, 2006). This also implies that different instruments identify different treatment effects simply because they represent the effect of the treatment over different groups of patients. Thus it is really difficult to interpret IV results in general, and for clinical practices in particular, where patients are often believed to select treatment based on their idiosyncratic (unobserved) gains.

In this paper we implement the method of local instrumental variable (LIV) to address the limitations of the IV approach in the context of a health care application in which agents decide treatment status based on their unobserved gains (essential heterogeneity). Specifically, we analyze the treatment effects (measure as a 5-year medical cost) associated with a breast-conserving surgery (BCSRT) versus an alternative (mastectomy or MST). We analyze not only the effect of the treatment as measured by the IV estimates, but also by the marginal treatment effects (MTE), the average treatment effect (ATE), and the effect of treatment on those treated (TT). Unlike the IV estimates, each of these parameters answers a well-defined economic question under essential heterogeneity.

We first test for the presence of essential heterogeneity in our sample using a version of the test proposed by Heckman et al, 2006. Our results indicate strong evidence for the presence of essential heterogeneity. We then use the LIV estimates to recover MTE, ATE, and TT. We estimate an average cost (effect) associated with BCSRT versus MST (ATE) of $41,493, whereas the estimated average cost (effect) on those treated (TT) is $52,329. Our estimated IV effects on the other hand, are found in the range $12,884 – $44,941. This confirms that the IV approach, in general, does not identify ATE or TT.

Finally, we illustrate the problems of interpretation associated with IV by empirically analyzing its decomposition as a weighted average of the MTE. We present evidence of the presence of negative IV weights in regions of the support of u_D in which the marginal effects (MTE) are positive. This means that estimated positive marginal effects enters with negative signs in the IV calculation. Thus, in principle, IV estimates can be negative even though the marginal effects are

positive, simply because of the IV weighting scheme. This demonstrates the limitations of the IV approach.

Although the conventional IV method may seem more robust and less demanding, it does not properly account for the presence of essential heterogeneity and is difficult to interpret in terms of a particular treatment question. LIV on the other hand does not suffer from these problems, but imposes stronger requirements on the data[37] and requires specification of a choice model. Alternative methods (see for example the factor based approach of Carneiro, Hansen and Heckman 2003; and Heckman, Stixrud and Urzua, 2006) that allow for essential heterogeneity can also be implemented, albeit under stronger assumptions.

On a final note, we would like to point out that there is a large literature on non-linear models that are more robust for modeling cost data due to the various characteristics of such data (see Manning (2006) for a review). Although, in our application, we use a linear-in-parameter and additive separable specification to model potential outcomes (costs) and apply the LIV method, this choice is primarily driven by the need for clarity of explaining the methods and for comparability of IV results with previously published work analyzing our application (Hadley et al., 2003; Polsky and Basu , 2006). More flexible specifications for the outcome equations are technically allowed when implementing the LIV method (Heckman et al. 2006). Extending this method to non-linear models, such as log-link generalized linear models, is likely to be an interesting area of future work.

[37] This is particularly important in the semi-parametric versions of the LIV estimator.

References:

Angrist JK, Krueger AB. *Empirical strategies in labor economics*. In O. Ashenfelter and D. Card (Eds.), Handbook of Labor Economics, New York: North Holland, 1999; Vol. 3A, 1277-1366.

Angrist J, Imbens G, Rubin G. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 1996; 91:444-455.

Basu A. Manning WG, Mullahy J. Comparing alternative models: Log vs Proportional hazard? *Health Economics* 2004; 13(8): 749-765.

Björklund A, Moffitt R. The estimation of wage gains and welfare gains in self-selection. *Review of Economics and Statistics* 1987; 69(1) 42-49.

Brooks JM, Chrischilles E, Scott S, Chen-Hardee S. Was Lumpectomy Underutilized for Early Stage Breast Cancer? – Instrumental Variables Evidence for Stage II Patients from Iowa. *Health Services Research* 2003; 38(6), Part I: 1385-1402.

Card D. Estimating the return to schooling. Progress on some persistent econometric problems. *Econometrica* 2001; 69(5): 1127-60.

Carneiro P, Hansen K, Heckman JJ. Estimating distribution of treatment effects with an application to returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* 2003; 44(2): 361-422.

Chamberlain G, Imbens G. Random effect estimator with many instrumental variables. *Econometrica* 2004; 71(1): 295 – 306.

Claxton K. The Irrelevance of Inference: A Decision-Making Approach to the Stochastic Evaluation of Health Care Technologies. *Journal of Health Economics* 1999; 18: 341-364.

Cox DR. The planning of Experiments. Wiley: New York, 1958.

Earle CE, Tsai JS, Gelber RD, Weinstein MC, Neumann PJ, Weeks JC. Effectiveness of chemotherapy for advanced lung cancer in the elderly: Instrument variable and propensity analysis. *Journal of Clinical Oncology* 2001; 19(4): 1064-1070.

Fisher RA. Design of Experiments. Oliver and Boyd: Edinburgh, 1935.

Hadley J, Mitchell JM, Mandelblatt J. Medicare fees and small area variations in the treatment of localized breast cancer. *New England Journal of Medicine* 1992; 52:334-360.

Hadley J, Polsky D, Mandelblatt JS, Mitchell JM, Weeks JC, Wang Q, Hwang YT, OPTIONS Research Team. An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a Medicare population. *Health Economics* 2003; 12: 171-186.

Harris KM, Remler DK. Who is the marginal patient? Understanding instrumental variable estimates of treatment effects. *Health Services Research* 1998; 33(5): 1337-1360.

Heckman JJ. Simultaneous equations model with both continuous and discrete endogenous variables with and without structural shift in the equations. In Studies in Nonlinear Estimation, eds. Steven Goldfield and Richard Quandt, Ballinger, 1976.

Heckman JJ. Dummy endogenous variables in a simultaneous equations system. *Econometrica* 1978; 46(4): 931-959.

Heckman JJ. Addendum to sample selection bias as a specification error. In Evaluation Studies Review Annual, vol 5 eds. Ernst Stromsdorfer and George Farkas, Sage Publications: Beverly Hills, CA, 1980.

Heckman JJ. Comments on Angrist, Imbens, and Rubin: Identification of causal effects using instrumental variables. *Journal of American Statistical Association* 1996; 91: 434.

Heckman JJ. Instrumental Variables: A study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources* 1997; 32 (3): 441-462.

Heckman JJ. Micro data, heterogeneity, and the evaluation of public policy: Nobel Lecture. *Journal of Political Economy* 2001; 109(4): 673-748.

Heckman JJ, Robb R. *Alternative methods for evaluating the impact of interventions*. In Longitudinal Analysis of Labor Market Data, Heckman JJ, Singer B (eds), Wiley: New York, 1985; 156-245.

Heckman, J.J., Sedlacek, G., 1985. Heterogeneity, aggregation and market wage functions: An empirical model of self selection in the labor market. Journal of Political Economy 93(6): 1077-25.

Heckman JJ, Honore B. The Empirical Content of the Roy Model. *Econometrica* 1990; 58 (5): 1121-49.

Heckman JJ, Smith J. Evaluating the welfare state. In Econometrics and Economic Theory in the 20th Century: The Ragnar Firsch Centennial eds. Strom S., Econometric Society Monograph Series, Cambridge University Press: Cambridge, 1998.

Heckman JJ, Stixrud J, Urzua S. The Effect of Cognitive and Noncognitive Skills on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics* 2006; v24(3):411-481.

Heckman JJ, Navarro S. Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics* 2004; 86 (1): 30-57.

Heckman JJ, Vytlacil E. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences* 1999; 96: 4730-4734.

Heckman JJ, Vytlacil E. 2000. The relationship between treatment parameters within a latent variable framework. *Economic Letters* 2000; 66(1): 33-39.

Heckman JJ, Vytlacil E. *Local instrumental variables*. In C. Hsiao, K. Morimue, and J.L. Powell (Eds.) Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in the Honor of Takeshi Amemiya, Cambridge University Press: New York, 2001a; pp 1-46.

Heckman JJ, Vytlacil E. Policy relevant treatment effects. *American Economic Review* 2001b; 91(2): 107-111.

Heckman JJ, Vytlacil E. Structural equations, treatment effects and econometric policy evaluation. *Econometrica* 2005; 73(3): 669-738.

Heckman JJ, Urzua S, Vytlacil E. Understanding instrumental variables in models with essential heterogeneity. Review of Economics and Statistics 2006; 88(3): 389-432.

Heckman JJ, Vytlacil E. Econometric evaluation of social programs. In J. Heckman and E. Leamer (Eds.) Handbook of Econometrics vol 6, Elsevier: Amsterdam, 2007.

Imbens GW. 2004. Non-parametric estimation of average treatment effects under exogeneity. A review. Review of Economics and Statistics 19(3): 358-64.

Imbens G, Angrist J. Identification and estimation of local average treatment effects. *Econometrica* 1994; 62(2): 467-475.

LaLonde R. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 1986; 76(4): 604-620.

Manning WG, Basu A, Mullahy J. Generalized modeling approaches on risk adjustment to skewed outcomes. *Journal of Health Economics* 2006; 24(3): 465 - 488.

Manning WG. *Dealing with skewed data on costs and expenditures*. In Elgar Companion to Health Economics. Ed. Jones A., Cheltenham: UK, 2006.

McClellan MB, Newhouse JB, McNeil BJ. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? *Journal of the American Medical Association* 1994; 272(11): 859-66

Newhouse J, McClellan MB. Econometrics in Outcomes Research: The Use of Instrumental Variables. *Annual Review of Public Health* 1998; 19:17-34.

Neyman J. Sur les applications de la thar des probabilities aux experiences Agaricales: *Essay des principle* 1923. [English translation of excerpts by D. Dabrowska and T. Speed, in *Statistical Sciences* 1990; 5: 463-472.]

Neyman J. Statistical problems in agricultural experimentation (with discussions). Supplement to the *Journal of the Royal Statistical Society, Series B* 1935; 2:107-180.

Nold JR, Beamer L, Helmer SD, McBoyle MF. Factors influencing a women's choice to undergo breast-conserving surgery versus modified radical mastectomy. *The American Journal of Surgery* 2000; 180(6): 413-418.

Polsky D, Basu A. *Selection bias in observational data.* In Elgar Companion to Health Economics. Ed. Jones A., Cheltenham: UK, 2006.

Polsky D, Mandelblatt JS, Weeks J, Venditti L, Hwang YT, Glick HA, Hadley J, Schulman KA. Economic evaluation of breast cancer treatment: Considering the value of patient choice. *Journal of Clinical Oncology* 2003; 21:1139-1146.

Powell JL. *Estimation of Semiparametric Models*. In Handbook of Econometrics, Volume 4, edited by R. Engle and D. McFadden. 2443–2521, Elsevier: Amsterdam, 1994.

Quandt RE. A new approach to estimating switching regressions. *Journal of the American Statistical Association* 1972; 67: 306-310.

Quandt RE. *The econometrics of disequilibrium*. Oxford: Blackwell, 1988.

Roy AD, Some thoughts on the distribution of earnings. *Oxford Economic Papers* 1951; 3: 135-146.

Rubin D. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 1978; 6: 34-58.

Staiger D. Stock JH. Instrumental Variables Regression with Weak Instruments. *Econometrica* 1997; 65: 557-586.

Vanness DJ, Mullahy J. Perspectives on Mean-based Evaluation of Health Care. In Jones A. (Eds) The Elgar Companion to Health Economics, Edward Elgar Publishing: Cheltenham, 2006.

Veronesi U, Cascinelli N, Mariani L, Greco M, Saccozzi R, Luini A, Aguilar M, Marubini E. Twenty-year follow-up of a randomized study comparing breast–conserving surgery with radical mastectomy for early breast cancer. *New England Journal of Medicine* 2002; 347(16): 1127-32.

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

Yitzhaki S. On using linear regression in welfare economics. Working paper 217, Department of Economics, Hebrew University. (available at http://jenni.uchicago.edu/underiv), 1989.

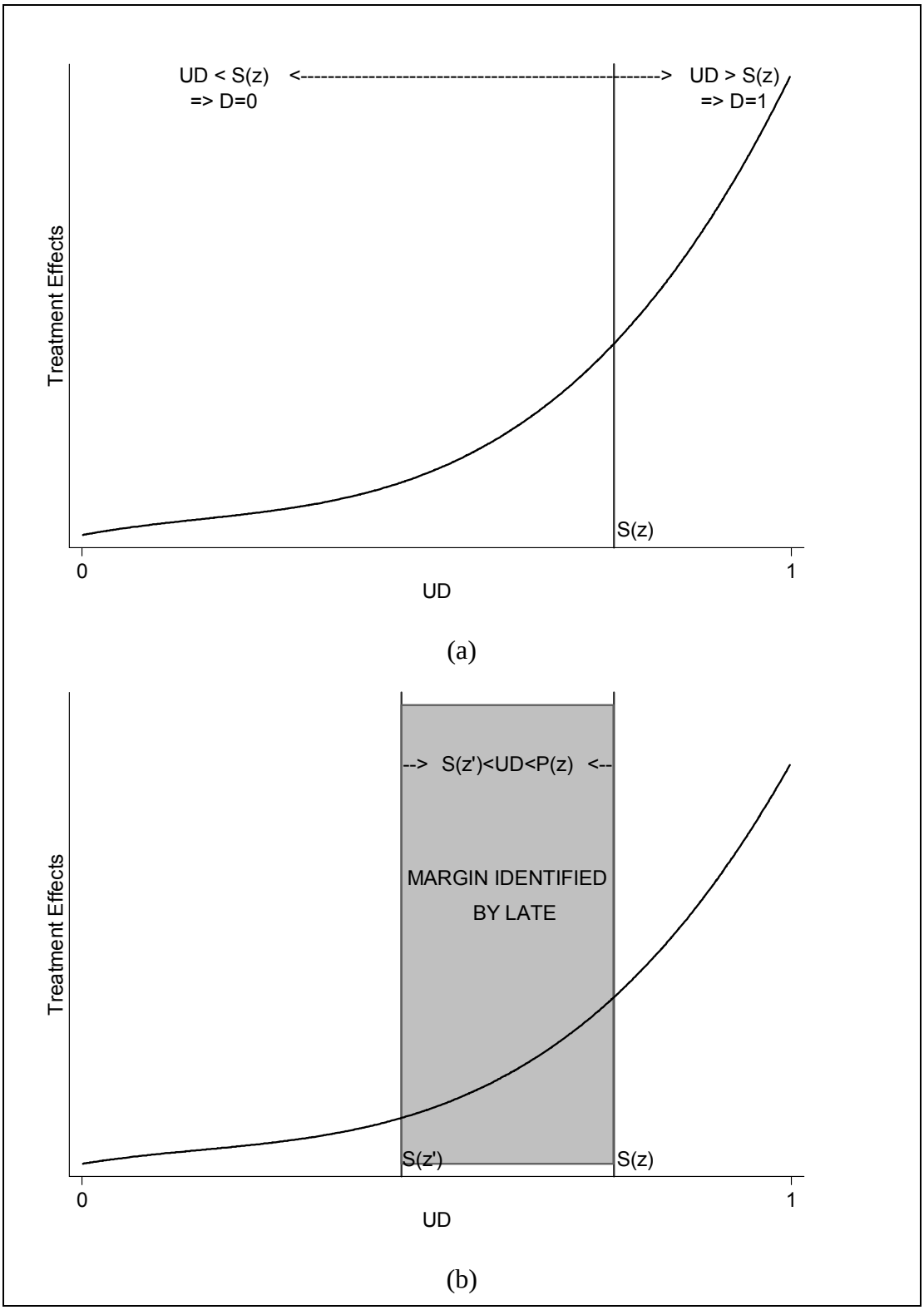
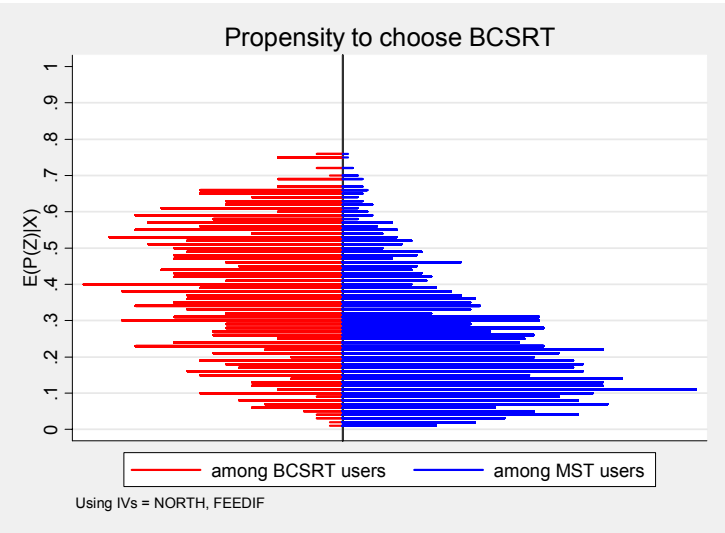
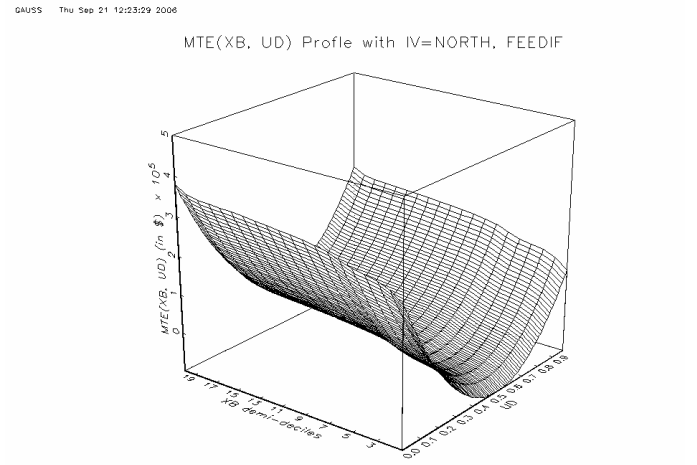


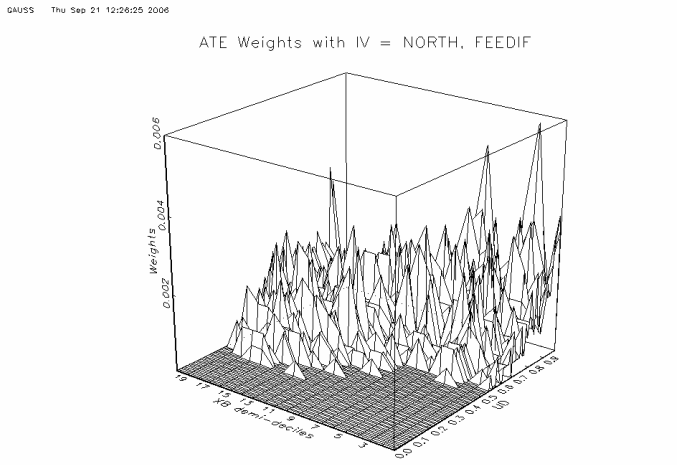
Figure 1: Stylized mean conditional treatment effects at each level of U_D. (a) Treatment choice based on $S(z)$; (b) Margin identified by LATE with two distinct values of an instrumental variable. Here dependence on X is suppressed for clarity.



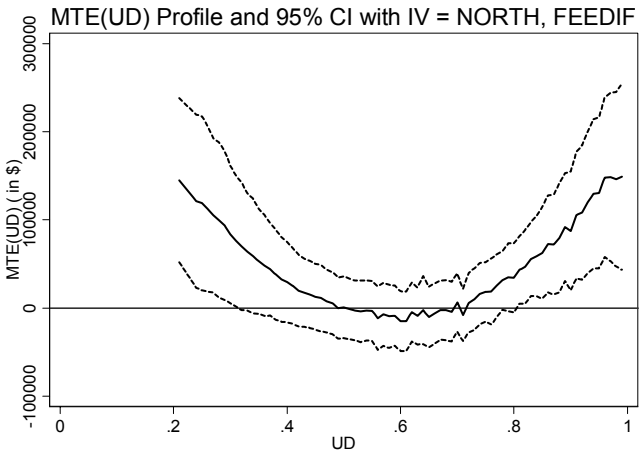
(a)



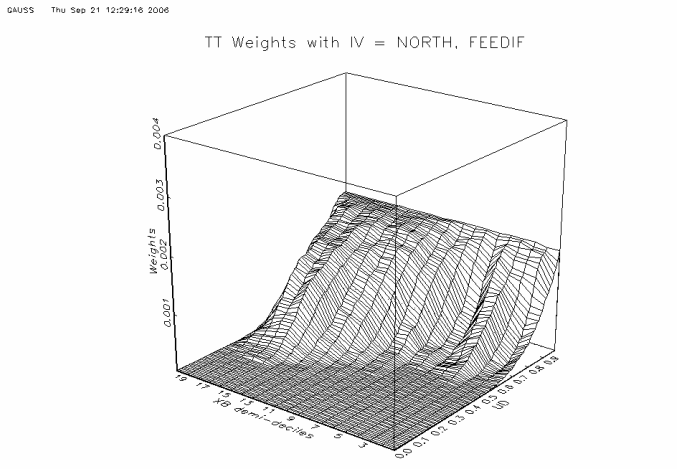
(b)



(c)



(d)



(e)

Figure 2: Propensity scores, treatment effects and IV weights using both instruments, NORTH and FEEDIF, as choice predictors. (a) Estimated propensity score for BCSRT among BCSRT and MST users; (b) $MTE(\eta_q, u_D)$ of BCSRT over MST; (c) $\varpi_{ATE}(\eta_q, u_D)$; (d) $MTE(u_D)$ of BCSRT over MST and 95% CI; and (e) $\varpi_{TT}(\eta_q, u_D)$;

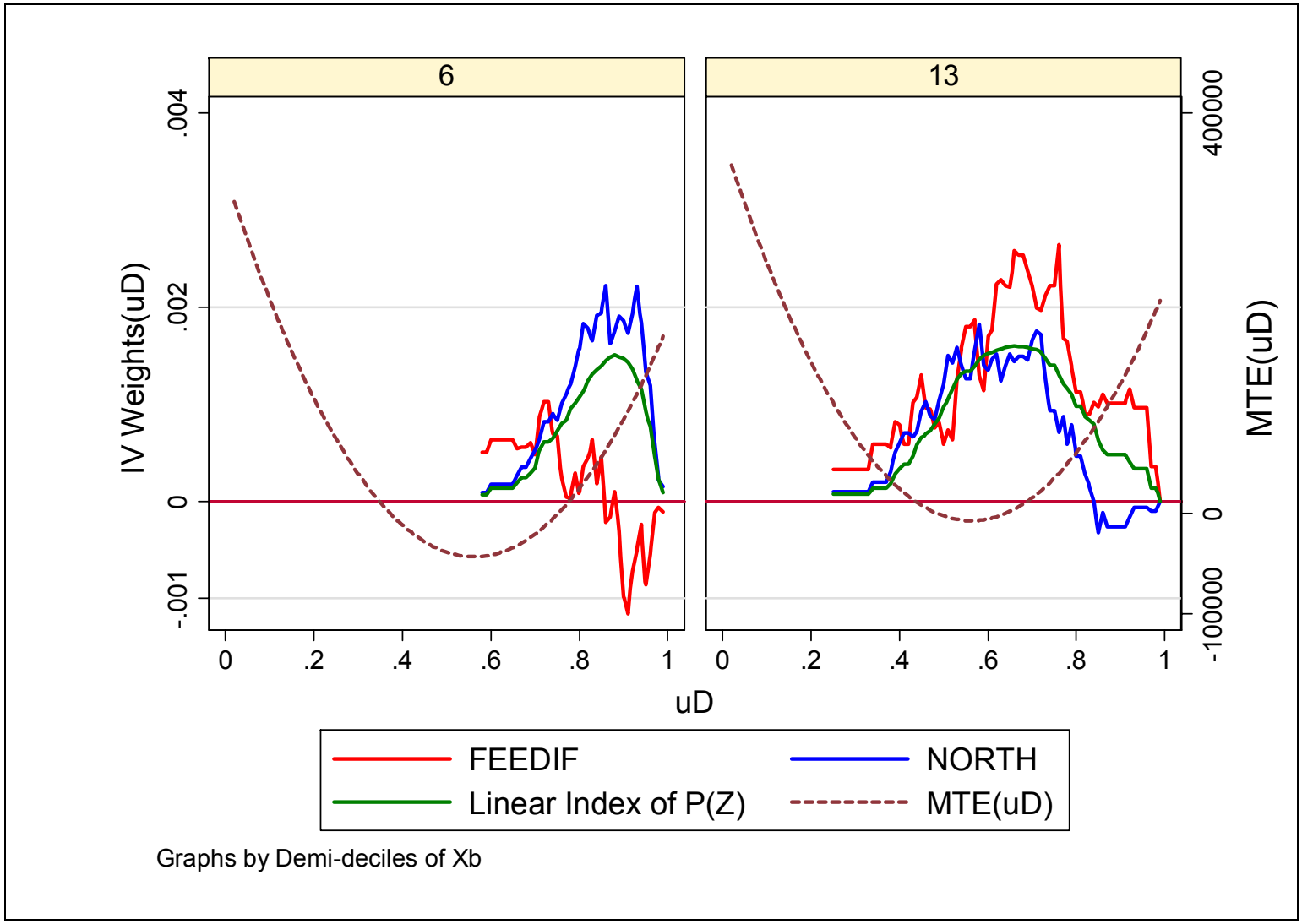


Figure 3: Based on a choice model that contains both NORTH and FEEDIF, IV weights, $\varpi_{IV}(\eta_q, u_D)$, for NORTH or FEEDIF as the sole instrument and for the entire linear index from choice model as the instrument along with MTE(U_D) at the 6th and 13th demi-decile of $X\hat{\beta}$.

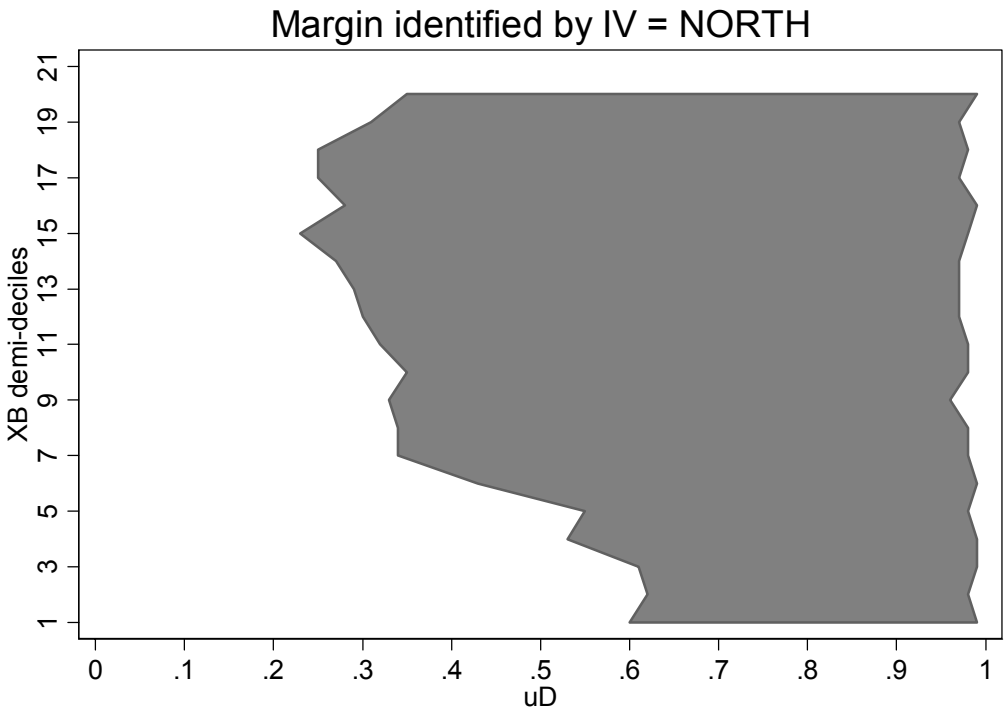


Figure 4: Margin identified by IV NORTH using the LATE approach

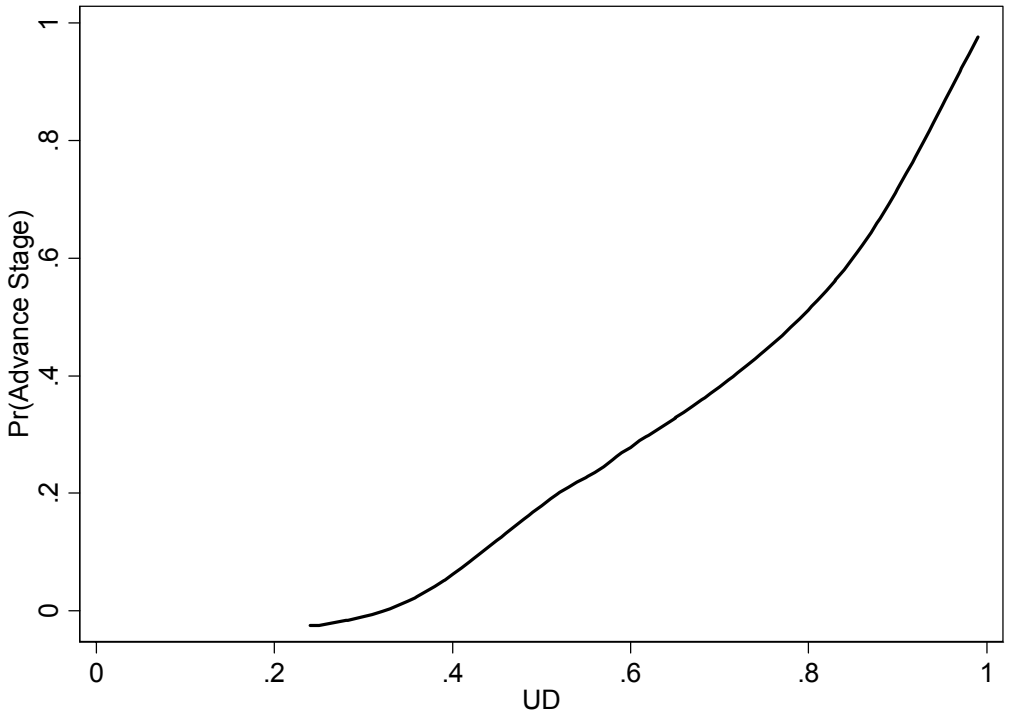


Figure 5: The proportion of marginal patients with advanced (stages B and C) cancer at each value of U_D.

Table 1: Tests of Linearity of conditional expectation of $E(Y|P(Z,X), X)$, where $P(Z,X)$ is a function of both NORTH and FEEDIF.

IV = NORTH, FEEDIF	Linear	Quadratic	Cubic	Quartic
p	24106 (30956)	22953 (56085)	235459^{+} (98431)	391171^{+} (136049)
p^2		1623 (65857)	-527099^{+} (211843)	-1314492^{+} (520164)
p^3			401761^{+} (153017)	1973011^{+} (960337)
p^4				-1072507 (647140)
F-statistic* (p-value)	-	0.00 (0.976)	4.35 (0.013)	3.35 (0.018)
L-R Chi-Square Statistic** (p-value)	-	0.00 (0.980)	6.99 (0.008)	2.79 (0.095)

+ p-value < 0.05. * Joint test of higher-order polynomials of propensity score p.
** Test for adding subsequent higher order polynomial of propensity score p.

Table 2: Treatment Effects under different estimators

Model	MTE approach		Linear IV	
	Treatment effects Mean (sd)	Identifying formula for IV effects†	Treatment effects Mean (sd)	Identifying formula for IV effects
Using IV=NORTH, FEEDIF in choice model			32,136 (15,005)+	$=\frac{Cov(Y,\hat{D}_{TSLS}(Z_1,Z_2))}{Var(\hat{D}_{TSLS}(Z_1,Z_2))}$
IV effect (for linear index from choice model)	33,708 (13,213) +	$=\frac{Cov(Y,\hat{D}_{LIV}(Z_1,Z_2))}{Cov(\hat{\mu}_V(Z_1,Z_2),\hat{D}_{LIV}(Z_1,Z_2))}$	-	
IV effect (for NORTH only)	29,580 (15,997)	$=\frac{Cov(Y,Z_1)}{Cov(Z_1,\hat{D}_{LIV}(Z_1,Z_2))}$	-	
IV effect (for FEEDIF only)	25,381 (14,653)	$=\frac{Cov(Y,Z_2)}{Cov(Z_2,\hat{D}_{LIV}(Z_1,Z_2))}$	-	
ATE (Empirical) *	41,493 (14,594) +		-	
ATE**	88,975 (35,221) +		-	
TT	52,329 (15,496) +		-	
Using IV=NORTH in choice model			44,941 (21,790)+	$=\frac{Cov(Y,\hat{D}_{TSLS}(Z_1))}{Var(\hat{D}_{TSLS}(Z_1))}$ ††
IV effect (for linear index from choice model)	43,997 (22,437)+	$=\frac{Cov(Y,Z_1)}{Cov(\hat{\mu}_V(Z_1),\hat{D}_{LIV}(Z_1))}$	-	
Using IV=FEEDIF in choice model			12,884 (32,081)	$=\frac{Cov(Y,\hat{D}_{TSLS}(Z_2))}{Var(\hat{D}_{TSLS}(Z_2))}$ ††
IV effect (for linear index from choice model)	13,170 (26,495)	$=\frac{Cov(Y,Z_2)}{Cov(\hat{\mu}_V(Z_2),\hat{D}_{LIV}(Z_2))}$	-	

+ $p < 0.05$; * estimated based on identified support for U_D; ** estimated after extrapolating MTE(u_D) over the missing support of U_D.

† Dependence on X suppressed for clarity. Z_1 = NORTH; Z_2 = FEEDIF; $\hat{D}$ = Pr(BCSRT)

††Since in linear IV, $\hat{D}_{TSLS}$ is estimated via a linear probability model, $\frac{Cov(Y,\hat{D}_{TSLS}(Z_1))}{Var(\hat{D}_{TSLS}(Z_1))}=\frac{Cov(Y,Z_1)}{Cov(Z_1,\hat{D}_{TSLS}(Z_1))}$. Same applies for Z_2. $\hat{D}_{LIV}$ is estimated using a probit model. $\hat{\mu}_V(Z_1)$ is the estimated linear predictor of the probit regression conditioned Z_1. $\hat{\mu}_V(Z_2)$ and $\hat{\mu}_V(Z_1,Z_2)$ are defined analogously.