UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2022

Fragility of the Marginal Treatment Effect

Paul J Devereux, University College Dublin

WP22/04

January 2022

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Fragility of the Marginal Treatment Effect*

Paul J. Devereux

University College Dublin, NHH, IZA, and CEPR

January 25, 2022

Abstract

Many interesting and important economic questions relate to the effects of binary treatments such as starting a college degree or participating in a job training program. The causal effects of these treatments are likely to be heterogeneous and recent research has emphasized the estimation of heterogeneous treatment effects, with a particular focus on Marginal Treatment Effects (MTEs). In this note, I describe why common methods of estimating MTEs of binary treatments can be very sensitive to omitted higher powers of covariates and demonstrate this using simple Monte Carlo simulations. I conclude by discussing approaches that may be useful for researchers to address this problem in practice.

*I thank Ben Elsner for helpful comments.This work was partially supported by the Research Council of Norway through its Centres of Excellence Scheme, FAIR project No 262675.

1 Introduction

Many interesting and important economic questions relate to the effects of binary treatments such as starting a college degree or participating in a job training program. These effects are likely to be heterogeneous; for example, the benefit of going to college may differ greatly based on the cognitive and non-cognitive skills of the individual. Additionally, individuals may decide whether or not to take a particular treatment based on the expected costs and benefits so persons taking the treatment are likely to be those with relatively high returns – referred to as selection into treatment based on gains. While much empirical analysis assumes homogenous treatment effects, recent research has emphasized the estimation of heterogeneous treatment effects, with a particular focus on Marginal Treatment Effects (MTEs).[1]

The Two Stage Least Squares (2SLS) estimator can provide consistent estimates if the treatment effect is constant across the population. MTE estimation relaxes the constant treatment effect assumption and allows treatment effects to differ across the population by quantifying the gain from treatment for individuals shifted into (or out of) treatment by a marginal change in the propensity score (the predicted probability of treatment). By estimating the full distribution of MTEs, researchers can estimate many parameters of interest such as the average treatment effect (ATE), the effect of treatment on the treated (ATT), and various policy related treatment effects. Because of this, these techniques are becoming very widely used in empirical practice.[2]

Estimation of MTEs requires quite strong assumptions and is therefore likely to be sensitive to misspecification. Some previous work (Andresen, 2018) has shown sensitivity of MTE estimates to the specification of the first stage regression that models the probability of receving the treatment. In this note, I focus on a different issue – robustness to

[1] The concept of a marginal treatment effect was first introduced by Björklund and Moffitt (1987) and has been further developed by Heckman and Vytlacil (1999, 2007), Carneiro et al. (2011), and Brinch et al. (2017) among others. Cornelissen et al. (2016) provide a useful introduction for empiricists.

[2] Topics studied include the return to education (Nybom, 2017; Carneiro et al., 2017; Kamhofer et al., 2019), the economics of crime (Agan et al., 2021; Arnold et al., 2018; Bhuller et al., 2020), the effects of childcare (Cornelissen et al., 2018; Felfe and Lalive, 2018; Andresen, 2019), the effects of family size on child education (Brinch et al., 2017) as well as many others.

omitted higher order powers of covariates.[3] Because instrumental variables are often randomly or quasi-randomly assigned, covariates tend to play little role in 2SLS and omitted higher powers of covariates are unlikely to lead to important inconsistencies. However, this is not the case with MTEs.

In this note, I describe why common methods of estimating MTEs can be very sensitive to omitted higher powers of covariates and demonstrate this using simple Monte Carlo simulations. I conclude by discussing what approaches may be useful to address this problem in practice.

2 Definition of a Marginal Treatment Effect

Consider an outcome for individual i, Y_i, and a binary treatment variable, D_i, that equals 0 if the person does not get the treatment and 1 if the person gets the treatment. Y_{ji} refers to the counterfactual outcome when $D_i = j$. We start with the following model:

$$Y_{ji} = \mu_j(X_i) + U_{ji} \qquad for \quad j = 0, 1 \tag{1}$$

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i} \tag{2}$$

$$D_i = I(\mu_D(Z_i) > V_i), \tag{3}$$

where $Z_i = (X_i, Z_i^*)$, and Z_i^* represents one or more excluded instrumental variables. (U_{0i}, U_{1i}, V_i) are error terms, $\mu_j(X_i)$ is a (possibly treatment-varying) function of exogenous covariates, and $\mu_D(Z_i)$ is a function of Z_i. The standard IV assumption for instrument validity is the conditional independence assumption that

$$(U_{0i}, U_{1i}, V_i) \perp Z_i^* | X_i. \tag{4}$$

[3] Indeed any omitted non-linear functions of covariates can be problematic. For simplicity, the focus here is on higher order powers such as omitted quadratic or cubic terms. However, the problem can arise even if quadratic or cubic functions of covariates are included if even higher order terms are incorrectly omitted from the model. Likewise, there can be problems due to omitted interaction terms between covariates.

If V_i has a continuous distribution, we can rewrite the selection equation as

$$D_i = I(P(Z_i) > U_{Di}), \tag{5}$$

where U_{Di} represents quantiles of V_i and $P(Z_i)$ is the propensity score. The Marginal Treatment Effect at each possible value of X_i and U_{Di} is defined as

$$MTE(x,u) = E(Y_{1i} - Y_{0i}|X_i = x, U_{Di} = u),$$

where u is a particular quantile of V.

3 Estimating Marginal Treatment Effects

While it is possible to estimate marginal treatment effects nonparametrically with no further assumptions, this requires full support of the propensity score in both treated and untreated samples for all values of X. Given that this is very unlikely in practice, researchers generally make two further assumptions:

Assumption 1: The model is linear in parameters so

$$\mu_j(X_i) = X_i\beta_j \tag{6}$$

$$\mu_D(Z_i) = Z_i\gamma, \tag{7}$$

where γ and β_j are parameters to be estimated.

Assumption 2:

$$E(Y_{0i}|X_i = x, U_{Di} = u) = x\beta_0 + E(U_{0i}|U_{Di}) \tag{8}$$

$$E(Y_{1i}|X_i = x, U_{Di} = u) = x\beta_1 + E(U_{1i}|U_{Di}). \tag{9}$$

This "linear separability" assumption implies that the MTE is identified over the unconditional support of $P(Z_i)$, jointly generated by the excluded instrument and the covariates, as opposed to the support of $P(Z_i)$ conditional on X_i (Brinch et al., 2017).[4] The linear separability assumption is typically benign in 2SLS as it is only a problem if controls for nonlinear functions of X_i are required to satisfy conditional exogeneity of the instrument Often Z_i^* is randomly or quasi-randomly assigned so this is not an issue. However, as we will see, it is not a benign assumption for MTE estimation.

Given these assumptions, the Marginal Treatment Effect is defined as

$$\begin{aligned} MTE(x,u) &= E(Y_{1i} - Y_{0i}|X_i = x, U_{Di} = u) && (10) \\ &= x(\beta_1 - \beta_0) + E(U_{1i} - U_{0i}|U_{Di} = u) && (11) \\ &= x(\beta_1 - \beta_0) + k(u), && (12) \end{aligned}$$

where

$$k(u) = E(U_{1i} - U_{0i}|U_{Di} = u). \quad (13)$$

In words, the MTE is the treatment effect for an individual who is at the margin of being treated (is indifferent between being treated or not) when $P(Z_i) = u$ and $X_i = x$. These assumptions imply that the intercept of the MTE function depends on X_i but the slope of the function is independent of X_i.[5]

3.1 Local Instrumental Variables (LIV)

The most common method to estimate the MTE is the method of Local Instrumental Variables (LIV; see Heckman and Vytlacil, 1999, 2007; Heckman et al., 2006; Carneiro et al., 2011), which estimates the MTE as the derivative of the outcome with respect to the propensity score, where the outcome has been modeled as a flexible function of the

[4] An alternative but stronger assumption is full independence so $(U_{0i}, U_{1i}, V_i) \perp Z_i^*, X_i$. Both the full independence and the linear separability assumption imply that the marginal treatment effect is additively separable into an observed and an unobserved component (see Cornelissen et al., 2016).

[5] Some early studies assume joint normality of the error terms but, due to the restrictive and somewhat arbitrary nature of the normality assumption, this assumption is now rarely made and I will not consider it in this note.

propensity score. We can write

$$E(Y_i|X_i = x, P(Z_i) = p) = x\beta_0 + x(\beta_1 - \beta_0)p + K(p), \tag{14}$$

where $K(p)$ is a non-linear function of the propensity score. The MTE is then given by

$$MTE(X_i = x, U_{Di} = p) = \frac{\partial E(Y_i|X_i = x, P(Z_i) = p)}{\partial p} = x(\beta_1 - \beta_0) + \frac{\partial K(p)}{\partial p} = x(\beta_1 - \beta_0) + k(p), \tag{15}$$

where $k(p) = \frac{\partial K(p)}{\partial p}$.

A typical approach is to model $K(p)$ as a polynomial in the propensity score, which itself is first estimated from a probit or logit regression of the treatment variable on Z_i.[6] The regression equation is

$$Y_i = X_i\beta_0 + X_i(\beta_1 - \beta_0)\widehat{p} + \sum_{k=2}^{K} \widehat{p}^k \pi_k + U_i \tag{16}$$

where $\widehat{p}$ is the estimated propensity score from the first stage.[7] If the polynomial in $\widehat{p}$ includes only a single quadratic term, the estimated MTE function is linear ($\widehat{k}(p) = 2\widehat{\pi}_2\widehat{p}$) while further higher order terms allow for non-linearities.[8]

3.2 The Separate Approach

As developed by Brinch et al. (2017), this approach uses equations

$$E(Y_{0i}|X_i = x, D_i = 0) = x\beta_0 + E(U_{0i}|U_{Di} \geq P(Z_i)) = x\beta_0 + K_0(p) \tag{17}$$

$$E(Y_{1i}|X_i = x, D_i = 1) = x\beta_1 + E(U_{1i}|U_{Di} < P(Z_i)) = x\beta_1 + K_1(p), \tag{18}$$

[6] The LIV approach with a global polynomial has been used by many studies including Agan et al. (2021), Alessie et al. (2020), Basu et al. (2007), Bhuller et al. (2020), Cornelissen et al. (2018), and Gupta et al. (2021).

[7] The estimated propensity score differs across sample members as they have different values of Z_i. However, to avoid clutter, I omit the i subscript from $\widehat{p}$.

[8] Another common approach is semiparametric and uses a local rather than a global polynomial. I discuss this later in the Monte Carlo simulations.

and estimates separate regressions for the conditional expectations of Y_{0i} and Y_{1i} on the sample for which $D_i = 0$ and the sample for which $D_i = 1$, respectively. The control functions ($K_0(p)$ and $K_1(p)$) are specified based on the assumptions made. Then the MTE equals

$$\begin{aligned} MTE(X_i &= x, U_{Di} = p) = E(Y_{1i}|X_i = x, U_{Di} = p) - E(Y_{0i}|X_i = x, U_{Di} = p) \quad (19) \\ &= x(\beta_1 - \beta_0) + k_1(p) - k_0(p), \quad (20) \end{aligned}$$

where $k_j(p) = E(U_{ji}|U_{Di} = p)$. We consider the case where the $k_j(p)$ functions are specified as a polynomial in p.[9] The regression equations are of the form

$$Y_i = X_i\beta_0 + \sum_{k=1}^{K} \widehat{p}^k \pi_{k0} + U_{0i} \text{ if } D_i = 0 \quad (21)$$

$$Y_i = X_i\beta_1 + \sum_{k=1}^{K} \widehat{p}^k \pi_{k1} + U_{1i} \text{ if } D_i = 1, \quad (22)$$

where, as before, $\widehat{p}$ is the estimated propensity score from the first stage.

4 Consequences of Misspecification

4.1 LIV

The MTE estimated using LIV has been shown to be somewhat sensitive to the choice of first stage model (probit, logit, or linear probability model) but perhaps a bigger issue is the fragility of the estimator to other sources of misspecification.[10] Consider once again the LIV estimating equation

$$Y_i = X_i\beta_0 + X_i(\beta_1 - \beta_0)\widehat{p} + \sum_{k=2}^{K} \widehat{p}^k \pi_k + U_i. \quad (23)$$

[9] The separate approach with a global polynomial has been used by Gong et al. (2020), Kowalski (2020), and Sarr et al. (2021) as well as others. As with LIV, a semiparametric approach with a local polynomial can be taken to estimate this model. I consider this approach later in the Monte Carlo simulations.

[10] Andresen (2018) does Monte Carlo simulations to investigate sensitivity to misspecification of the first stage model and to some other misspecifications.

To fix ideas, assume X_i includes a single continuous variable, W_i, in addition to the constant and the order of the polynomial in the propensity score is quadratic. Given this, the estimated MTE is linear and its slope equals $2\widehat{\pi}_2$, twice the coefficient on the squared term. Now, assume that this regression is misspecified in that Y_i depends on W_i^2 as well as W_i but W_i^2 is excluded from the model.[11]

Let's start with the simplest case where $\widehat{p}$ is the predicted value from a linear probability model first stage so

$$\widehat{p} = Z_i\widehat{\gamma} = \widehat{\gamma}_1 + Z_i^*\widehat{\gamma}_z + W_i\widehat{\gamma_w}, \tag{24}$$

where $\widehat{\gamma}_1$ is the estimate of the constant term. Clearly, $\widehat{p}^2$ is correlated with W_i^2. Therefore, given the exclusion of W_i^2 from equation (23), $\widehat{\pi}_2$ will be an inconsistent estimator of π_2 and the estimator of the slope of the MTE will be inconsistent. If, instead, we use a nonlinear first stage such as logit or probit, the same issue arises as $\widehat{p}$ itself is then a nonlinear function of W_i and so both $\widehat{p}$ and $\widehat{p}^2$ are likely to be correlated with W_i^2. Note that this is a fundamental problem unless W_i is an indicator variable.

Some further points are relevant. First, even if W_i and W_i^2 are included in the model, $\widehat{\pi}_2$ will be an inconsistent estimator if Y_i depends on even higher order powers of W_i. Second, the extent of the problem will tend to increase with the number of covariates. Third, this problem continues to exist even if the specified polynomial in $\widehat{p}$ is third order or higher (and may even be exacerbated as included higher powers of $\widehat{p}$ are correlated with excluded higher powers of W_i).[12] So, the LIV estimator of the slope of the MTE will necessarily be very sensitive to the specification of the X_i matrix. The estimator of $\beta_1 - \beta_0$ will also be inconsistent if W_i^2 is excluded from the model as the inconsistency of the estimator for the parameter on $\widehat{p}^2$ will lead to inconsistent estimates of all the other parameters in the model.

[11] Such misspecification could also lead 2SLS to be inconsistent in the constant treatment effects case if the presence of W_i^2 is required to satisfy conditional independence of Z_i^*. However, this may be less likely as (1) instruments are often randomly or quasi-randomly generated and (2) researchers use their knowledge of the situation to appropriately specify X_i to satisfy conditional independence.

[12] Models that assume joint normality and, hence, have a nonlinear parametric function of $\widehat{p}$ in the regression, will also face the same problem as the nonlinear function will be correlated with W_i^2.

4.2 The Separate Approach

Consider once again the estimating equations for the separate approach:

$$Y_i = X_i\beta_0 + \sum_{k=1}^{K} \widehat{p}^k \pi_{k0} + U_{0i} \text{ if } D_i = 0 \quad (25)$$

$$Y_i = X_i\beta_1 + \sum_{k=1}^{K} \widehat{p}^k \pi_{k1} + U_{1i} \text{ if } D_i = 1. \quad (26)$$

As before, consider the case where there is a single included explanatory variable, W_i, Y_{ji} depends on W_i^2 as well as W_i, and the $k_j(p)$ functions are specified as a polynomial in $\widehat{p}$. Start again with the simplest case of a linear MTE function (the regression equations are linear in $\widehat{p}$) and where $\widehat{p}$ is estimated as the predicted value from a linear probability model first stage so $\widehat{p} = Z_i\widehat{\gamma} = \widehat{\gamma}_1 + Z_i^*\widehat{\gamma}_z + W_i\widehat{\gamma_w}$. The omission of W_i^2 leads to omitted variable bias in equations (25) and (26) as W_i and $\widehat{p}$ will generally be correlated with W_i^2. Even in the special case where W_i is uncorrelated with W_i^2 in the full sample, it will be correlated with W_i^2 in each subsample defined by treatment status. The logic is that, since W_i shifts the probability of treatment, the distribution of $\widehat{p}$ and W_i will differ between the subsample with $D_i = 0$ and the subsample with $D_i = 1$, so W_i and $\widehat{p}$ will generally be correlated with W_i^2 in each subsample. These correlations will differ across subsamples, leading to omitted variable biases that differ in size and possibly direction for the estimators of π_{10} and π_{11} and, hence, in the MTE.

With a higher order polynomial in $\widehat{p}$, higher powers of $\widehat{p}$ are correlated with W_i^2 (given $\widehat{p} = \widehat{\gamma}_1 + Z_i^*\widehat{\gamma}_z + W_i\widehat{\gamma_w}$ so $\widehat{p}^2$ is a function of W_i^2) leading to inconsistent estimators of the π_{k0} and π_{k1} parameters and, hence, likely inconsistency of the MTE. With a nonlinear first stage from a Logit or Probit model, there is an additional potential source of inconsistency as $\widehat{p}$ is a nonlinear function of W_i and Z_i^* and, hence, is likely correlated with W_i^2.[13]

[13] These issues also arise when the $K_j(p)$ functions are derived from joint normality and may be exacerbated as these functions are nonlinear in W_i and Z_i^*.

5 Monte Carlo

I now do Monte Carlo experiments to illustrate the effects of misspecification in a very simple setup. The base model assumes

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i} \quad (27)$$

$$Y_{0i} = U_{0i} + \beta_0 X_i + \theta_1 X_i^2 + \theta_2 X_i^3 \quad (28)$$

$$Y_{1i} = U_{1i} + \beta_1 X_i + \theta_1 X_i^2 + \theta_2 X_i^3 \quad (29)$$

$$D_i^* = \pi_0 + \pi_1 Z_i^* + \pi_2 X_i + v_i \quad (30)$$

$$D_i = 1 \text{ if } D_i^* \geq 0 \quad (31)$$

$$D_i = 0 \text{ if } D_i^* < 0 \quad (32)$$

where D_i is a binary endogenous variable, Z_i^* is a continuous instrument that is distributed i.i.d. $N(0,1)$, and X_i is a continuous exogenous variable that is distributed i.i.d. $N(0,1)$. Z_i^* is drawn independently of X_i. We allow for misspecification by excluding X_i^2 or X_i^3 from the estimated model. We vary the values of θ_1 and θ_2, showing estimates without misspecification ($\theta_1 = \theta_2 = 0$) and estimates with misspecification ($\theta_1 \neq 0$ or $\theta_2 \neq 0$). In all cases, we set $\pi_0 = 0$, $\pi_1 = 0.25$, $\pi_2 = 1$ and $\beta_0 = \beta_1 = 1$.

5.1 Linear Model

First, we create a model with a linear MTE function, defining U_{0i} and U_{1i} so as to allow for selection into treatment based on both levels and gains:

$$U_{0i} = 0.5(U_{Di} - 0.5) + \epsilon_i \quad (33)$$

$$U_{1i} = 1.5(U_{Di} - 0.5) + \epsilon_i, \quad (34)$$

where $\epsilon_i \sim N(0, 0.5)$ and U_{Di} is distributed standard uniform. Finally $v_i = F^{-1}(U_{Di})$. By construction, this setup implies that the MTE function is linear.

5.2 Quadratic Model

The second model has a quadratic MTE function:

$$U_{0i} = 1.5(U_{Di} - 0.5) - U_{Di}^2 + \epsilon_i \quad (35)$$

$$U_{1i} = -0.5(U_{Di} - 0.5) + U_{Di}^2 + \epsilon_i, \quad (36)$$

where $\epsilon_i \sim N(0, 0.5)$, U_{Di} is distributed standard uniform, and $v_i = F^{-1}(U_{Di})$. By construction, this setup implies that the MTE function is quadratic.

5.3 Implementation

We estimate the models with 100,000 observations.[14] These large sample sizes imply that we estimate the MTE function with reasonable precision. In both models, the parameter values are chosen to ensure that the instrument is sufficiently strong to provide variation in the treatment across the distribution of the propensity score. The common support across the distribution of the propensity score is shown in Figure 1.[15] We perform 100 Monte-Carlo replications and plot out the average value of the MTE for each percentile of U_{Di}.[16] For each estimator, we also report the average treatment effect (ATE), the effect of treatment on the treated (ATT), and the effect of treatment on the non-treated (ATUT). These treatment effects are calculated from the MTE using appropriate weights as follows (N denotes the number of observations, 100,000 in all cases):

$$\widehat{ATE} = \sum_{i=1}^{N} X_i(\widehat{\beta}_1 - \widehat{\beta}_0) + \sum_{u=0.01}^{0.99} \frac{\widehat{k(u)}}{99} \quad (37)$$

$$\widehat{ATT} = \sum_{i=1}^{N} \frac{\widehat{p}}{\overline{p}} X_i(\widehat{\beta}_1 - \widehat{\beta}_0) + \sum_{u=0.01}^{0.99} (\frac{prop(\widehat{p} > u)}{99\overline{p}})\widehat{k(u)} \quad (38)$$

[14] The models are estimated using the Stata MTEFE command created by Andresen (2018).

[15] Figure 1 shows the distribution of propensity scores for the linear model without misspecification. The analogous distributions are very similar for the quadratic model and for all specifications.

[16] While 100 simulations is not many, it is probably sufficient due to the very large number of observations and, hence, precise estimates. As we will see, even with only 100 replications, the correctly specified models provide average estimates that exactly equal the true parameters.

$$\widehat{ATUT} = \sum_{i=1}^{N} \frac{(1-\widehat{p})}{1-\overline{p}} X_i(\widehat{\beta}_1 - \widehat{\beta}_0) + \sum_{u=0.01}^{0.99} (\frac{prop(\widehat{p} \leq u)}{99(1-\overline{p})})\widehat{k(u)} \tag{39}$$

5.4 Results of Monte Carlo Simulations

5.4.1 Linear Model

Figure 2 shows the estimated marginal treatment effects for the linear MTE model, estimated using the linear polynomial versions of both LIV and the separate method. Because there is limited overlap in the support of the propensity score for the treated and untreated in the tails of the U_{Di} distribution, we trim the pictures and show the MTEs from $U_{Di} = 0.05$ to $U_{Di} = 0.95$. In the absence of misspecification ($\theta_1 = \theta_2 = 0$), both estimators perform very well and we do not show this case in the figure as the estimated MTE line sits on top of the true values of the MTEs. So, we show the true MTEs and those estimated from misspecified models where we consider $\theta_1 = \{0.05, -0.05\}$. The true MTE line has a slope of 1. The slope of the estimated lines differs greatly and is even negative for both estimators when $\theta_1 = 0.05$. Clearly, both estimators are very sensitive to the omitted X_i^2 term.

Table 1 reports the summary treatment effects for each specification. The true value of the ATE is 0, the true ATT is -0.16, and the true ATUT is 0.16. Without misspecification, both approaches are unbiased. However, while both estimators get the ATE right with $\theta_1 \neq 0$, they can be very wrong for the ATT and the ATUT. I also report the average standard errors across the simulations to show that this is not an issue of precision but purely one of bias. Both estimators allow $\widehat{\beta}_1 - \widehat{\beta}_0$ to differ from 0 but, by design, the Monte Carlos have no heterogeneity in this dimension so the estimate of $\widehat{\beta}_1 - \widehat{\beta}_0$ should be 0. However, as can be see in Table 1, when $\theta_1 \neq 0$, this parameter can also be very biased.

Table 1 also shows the R^2 from (1) a regression of Y_i on D_i and X_i and (2) a regression of Y_i on D_i, X_i and X_i^2. Adding X_i^2 to the model has no discernable effect on the R^2, suggesting that it is possible to have sizeable biases even with a relatively small amount of misspecification.[17]

5.4.2 Quadratic Model

Figure 3 shows the estimated marginal treatment effects for the quadratic model. Once again, the correctly specified model performs really well so we do not show it in the pictures. When $\theta_1 \neq 0$, the results differ by approach. The LIV estimates are very biased as can be seen in the first panel of Figure 3. The separate approach gives MTE estimates that are very close to correct, irrespective of the value of θ_1. However, this is a special case. In Figure 4, we show estimates for the quadratic model when we set $\theta_1 = 0$ but $\theta_2 = \{0.02, -0.02\}$, so that the X_i^3 term is incorrectly omitted from the model. In this case, we find large biases using both approaches.

Tables 2 and 3 report the summary treatment effects for each specification of the quadratic model and the biases from misspecification are once again apparent for the ATT, the ATUT, and also for $\widehat{\beta}_1 - \widehat{\beta}_0$. However, the findings differ by specification suggesting that it is unpredictable when the estimators are likely to be unbiased.

5.5 Semi-parametric methods

So far, we have considered estimation methods where $K(p)$ is modelled as a polynomial in the propensity score. This is a parametric approach but is flexible and, quoting French and Taber (2011, p. 578), "By letting the terms in the polynomial get large with the sample size, this can be considered a nonparametric estimator." However, it is also common for researchers to use a semiparametric approach such as the semiparametric LIV approach of Carneiro, Heckman and Vytlacil (2011). This approach retains a probit or logit first stage but estimates the main equation using Robinson's Double Residual approach (Robinson, 1988) and takes a local (rather than global) polynomial approach to estimating the relationship between Y_i and $\widehat{p}$. The LIV version proceeds as follows:[18]

[17] One should not read too much into this finding based on one simple Monte Carlo experiment. Because X_i is drawn from a standard normal distribution, X_i and X_i^2 are uncorrelated which typically implies greater bias from the incorrect omission of X_i^2 than if they were positively correlated.

[18] The separate approach can also be implemented using a similar semiparameteric approach (Brinch et al., 2017).

1. Estimate $\widehat{p}$ using a probit (or logit model).

2. Construct $\widetilde{Y}_i \equiv Y_i - E(Y_i|\widehat{p})$ and $\widetilde{X}_{ik} \equiv X_{ik} - E(X_{ik}|\widehat{p})$ for each component, k of X_i.

3. Do a linear regression of $\widetilde{Y}_i$ on $\widetilde{X}_i$ and $\widehat{p}\widetilde{X}_i$ to get $\widehat{\beta}_1$ and $\widehat{\beta}_0$.

4. Nonparametrically regress $Y_i - X_i\widehat{\beta}_0 - X_i(\widehat{\beta}_1 - \widehat{\beta}_0)\widehat{p}$ on $\widehat{p}$ using a local polynomial approach.

Given that $E(Y_i - X_i\beta_0 - X_i(\beta_1 - \beta_0)\widehat{p}|\widehat{p}) = K(\widehat{p})$, the MTE can be estimated by taking the derivative with respect to $\widehat{p}$:

$$MTE(U_{Di} = \widehat{p}, X_i = x) = X_i(\widehat{\beta}_1 - \widehat{\beta}_0) + \frac{\partial \widehat{K}(\widehat{p})}{\partial \widehat{p}}. \tag{40}$$

Unlike a global polynomial, the MTE can be estimated with this method only over the common support of $\widehat{p}$, which may not span $[0, 1]$ so it may not be possible to estimate the standard treatment parameters (ATE, ATT, ATUT) without making further assumptions and extrapolating. Figure 5 shows results for the linear model and Figures 6 and 7 show results for the quadratic model.[19] Clearly, misspecification can lead to serious biases in both cases. Interestingly, the semiparametric LIV approach is very biased even without misspecification.

6 What Can Researchers Do?

Given the findings that MTE estimators are sensitive to omitted higher powers of X, what are the best approaches to deal with this potential problem? I consider two broad strategies: The first is to use a method that is not vulnerable to this bias; the second is to do robustness checks to evaluate whether the problem is likely present in a particular application.

[19] There are many choices that can be made in estimating the semiparametric model such as the choice of bandwidth and the degree of the local polynomial. I use the default settings in MTEFE which imply a rule of thumb bandwidth and a local quadratic polynomial.

6.1 Methods Without this Problem

6.1.1 Estimate the Model without Covariates

One approach is to omit X from the model so that $\widehat{p}$ is not a function of X. An example of this approach is provided by Brinch et al. (2017) and by Doyle (2007) who both show how the MTE can be estimated without using covariates. However, this approach potentially leads to misspecification of the first stage and inconsistent estimates if identification relies on conditional independence of the instrument. Also, the X variables are often required in the first stage to enable $\widehat{p}$ to span $[0, 1]$ for both treatment and control group so this approach is likely to lack precision in many applications and, because of a limited range of the estimated propensity score, also may heavily rely on extrapolation based on a low order global polynomial.

6.1.2 Saturate the Model in X

Another possible solution to this problem, if feasible, is to saturate the model in X. This can be done if the X variable(s) are discrete by including all possible levels and interactions in the model. If variables are continuous, they could be discretized so that the model can be saturated. With a saturated model, there cannot be any omitted higher order powers so the problem disappears. If X is high-dimensional, this approach may not be feasible or desirable given the large number of resulting variables and hence the loss of degrees-of-freedom. Additionally, continous X variables can be useful to enable $\widehat{p}$ to span $[0, 1]$ for both treatment and control groups. In practice, researchers do not saturate the model in X when estimating marginal treatment effects.

6.1.3 Estimate a Fully Nonparametric Model

If researchers are willing to forego assumption 2 (equations (8) and (9)) and have access to one or more continuous instruments with a large range of variation within cells of $X_i = x$, then an MTE can be estimated for each specific subsample defined by each unique set of values of X, thereby conditioning nonparametrically on X.[20] Unfortunately, while much discussed in the literature, this approach is never feasible with the sets of instruments and covariates available in practice. Recent research suggests promising approaches that aim for interval rather than point estimation of the MTE (Mogstad and Torgovitsky, 2018) but these methodologies are not commonly used by applied researchers.

6.2 Check Robustness To Specifying Covariates More Flexibly

Rather than use a method that is robust to omitted higher order covariate terms, an alternative strategy is to evaluate the robustness of MTE estimates to adding higher order functions of X as additional covariates in equation (23) for LIV and equations (25) and (26) for the Separate Approach. One flexible way of doing this is to add a set of high order polynomials and interactions between them. Assessing stability of the MTE estimates after adding higher order and interaction terms in X provides an indication of whether the estimates are robust to this type of misspecification. While researchers tend to do many robustness checks, assessing robustness to higher order functions of X is very rare in the literature.

An alternative, which may be infeasible due to the "curse of dimensionality" when there are several X variables, is to estimate these regressions (equation (23) for LIV and equations (25) and (26) for the Separate Approach) using an estimator that is non-parametric in X. One way to avoid the dimensionality problem is by being non-parametric in a single index of X in equation (23) for LIV and equations (25) and (26) for the Separate Approach. A natural choice for that single index is one that uses a similar function of X as is used in the first stage to estimate $\widehat{p}$ as this is most likely to reduce contamination from correlation of $\widehat{p}$ with nonlinear functions of X. Denote the single index as $\widehat{f}(X_i)$. If a probit first stage is used to estimate $\widehat{p}$, $\widehat{f}(X_i)$ could be the predicted value from a probit regression of the treatment status on X. Note that $\widehat{f}(X_i)$ differs from $\widehat{p}$ in that it comes from a regression of D on X rather than a regression of D on X and Z. Then,

[20] As described by Cornelissen et al. (2016, page 55): "If a continuous instrument with a large range of variation within cells of $X_i = x$ is available, then the analysis can proceed in subsamples defined by the values of of $X_i = x$, thus conditioning perfectly and nonparametrically on X, and identifying a separate MTE curve for each value of $X_i = x$."

in equation (23) for LIV and equations (25) and (26) for the Separate Approach, the effect of $\widehat{f}(X_i)$ could be estimated nonparametrically in a partially linear model using, for example, the double residuals estimator introduced by Robinson (1988), which is consistent and efficient. If this procedure led to different MTE estimates compared to the basic model, it would suggest biases from omitted nonlinear functions of X.

A simple approach to assess robustness to higher order powers of X is to add a polynomial in $\widehat{f}(X_i)$ into the equations (equation (23) for LIV and equations (25) and (26) for the Separate Approach) as additional controls. This may make particular sense for the LIV model as the bias comes from correlations between the included polynomial in $\widehat{p}$ and excluded higher powers of X. Once again, if the addition of polynomials in $\widehat{f}(X_i)$ in equation (23) for LIV and equations (25) and (26) for the Separate Approach have little effect on the MTE estimates, this suggests that there is little bias from omitted higher order powers of covariates.

7 Conclusions

This note argues that standard methods used to estimate marginal treatment effects are inconsistent in the presence of omitted higher powers of X and demonstrates this using simple Monte Carlo simulations. The greater information provided by MTEs requires strong assumptions and estimates may be misleading if these do not hold. Assessing the stability of estimates to additional controls for nonlinear functions of X may be useful to address this problem in practice.

References

[1] Agan, A. Y., J. L. Doleac, and A. Harvey (2021, March). Misdemeanor Prosecution. NBER Working Paper 28600.

[2] Alessie RJM, Angelini V, Mierau JO, Viluma L. Moral hazard and selection for voluntary deductibles. *Health Economics.* 2020;29:1251–1269. https://doi.org/10.1002/hec.4134.

[3] Andresen, M. E. (2018). Exploring marginal treatment effects: Flexible estimation using Stata. *Stata Journal*, 18(1):118 158.

[4] Andresen, M. E. (2019). Child care for all? Treatment effects on test scores under essential heterogeneity. Working Paper, Statistics Norway.

[5] Bhuller, M., G. Dahl, K. Loken, and M. Mogstad (2020). Incarceration, Recidivism and Employment. *Journal of Political Economy*, vol. 128, no. 4.

[6] Bjorklund, Anders and Robert Moffitt. The estimation of wage gains and welfare gains in self selection models. *The Review of Economics and Statistics*, 42–49, 1987.

[7] Basu, A., Heckman, J., Navarro-Lozano, S., & Urzua, S. (2007). Use of instrumental variables in the presence of heterogeneity and self-selection: An application to treatment of breast cancer patients. *Health Economics*, 16, 1133–1157.

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

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

[10] Carneiro, P., Lokshin, M., Ridao-Cano, C. and Umapathi, N. Average and marginal returns to upper secondary schooling in Indonesia, *Journal of Applied Econometrics*, Vol. 32, (2017) pp. 16–36.

[11] Cornelissen.T, C. Dustmann, A.Raute and U. Schonberg (2016) From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions, *Labour Economics* 41: 47-60.

[12] Cornelissen, Thomas Christian Dustmann, Anna Raute, and Uta Schonberg. Who benefits from universal child care? estimating marginal returns to early child care attendance. *Journal of Political Economy*, 126(6): 2356–2409, 2018.

[13] Doyle, Joseph, Child Protection and Child Outcomes: Measuring the Effects of Foster Care, *American Economic Review*, 97 (2007), 1583–1610.

[14] Felfe, Christina, and Rafael Lalive. 2018. Does Early Child Care Affect Children's Development? *Journal of Public Economics* 159 (March): 33–53.

[15] French, Eric and Christopher Taber (2011). Identification of Models of the Labor Market. In Ashenfelter, Orley and David Card (eds). *Handbook of Labor Economics*, Vol 4A. Amsterdam, Elsevier: 537-617.

[16] Gong, J., Lu, Y., and Xie, H. (2020). The average and distributional effects of teenage adversity on long-term health. *Journal of Health Economics*, 71:102288.

[17] Gupta, Atul, Sabrina T. Howell, Constantine Yannelis, and Abhinav Gupta. Does private equity investment in healthcare benefit patients? evidence from nursing homes. No. w28474. National Bureau of Economic Research, 2021.

[18] Heckman, J.J., Vytlacil, E., 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proc. Natl. Acad. Sci.* 96 (8), 4730–4734.

[19] Heckman, James J. 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.

[20] Kamhofer DA, Schmitz H, and Westphal M (2019) Heterogeneity in marginal non-monetary returns to higher education. *Journal of the European Economic Association* 17(1): 205–244.

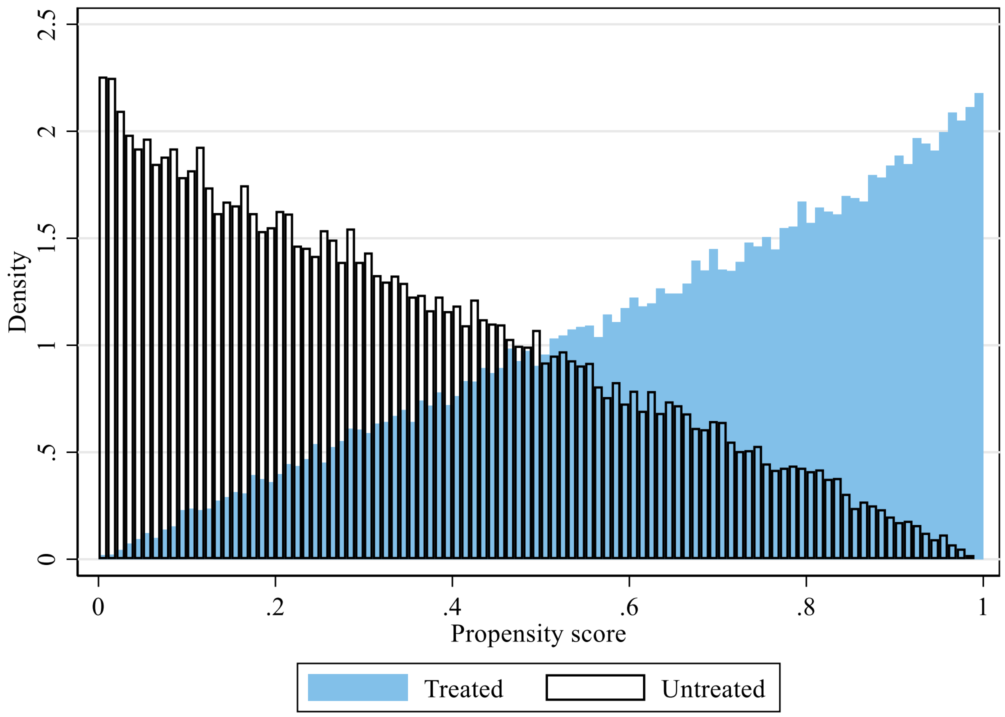
[21] Kowalski, A. E. (2020): Reconciling Seemingly Contradictory Results from the Oregon Health Insurance Experiment and the Massachusetts Health Reform, Working Paper 24647, National Bureau of Economic Research.

[22] Mogstad, Magne and Alexander Torgovitsky. 2018. Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics* 10: 577-613.

[23] Robinson, Peter M. 1988. Root-N-Consistent Semiparametric Regression. *Econometrica*, 56(4): 931-54.

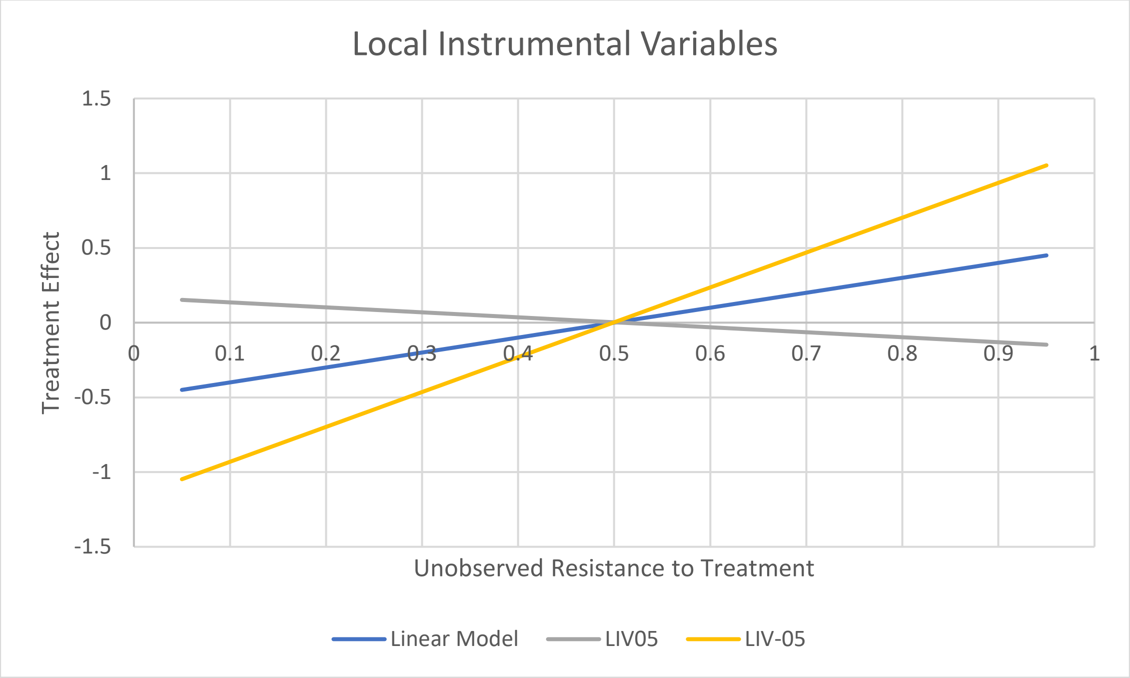
[24] Sarr, M., Ayele, M.B., Kimani, M. E., and Ruhinduka, R. (2021). Who benefits from climate-friendly agriculture? The marginal returns to a rainfed system of rice intensification in Tanzania. *World Development*, 138, 105160.

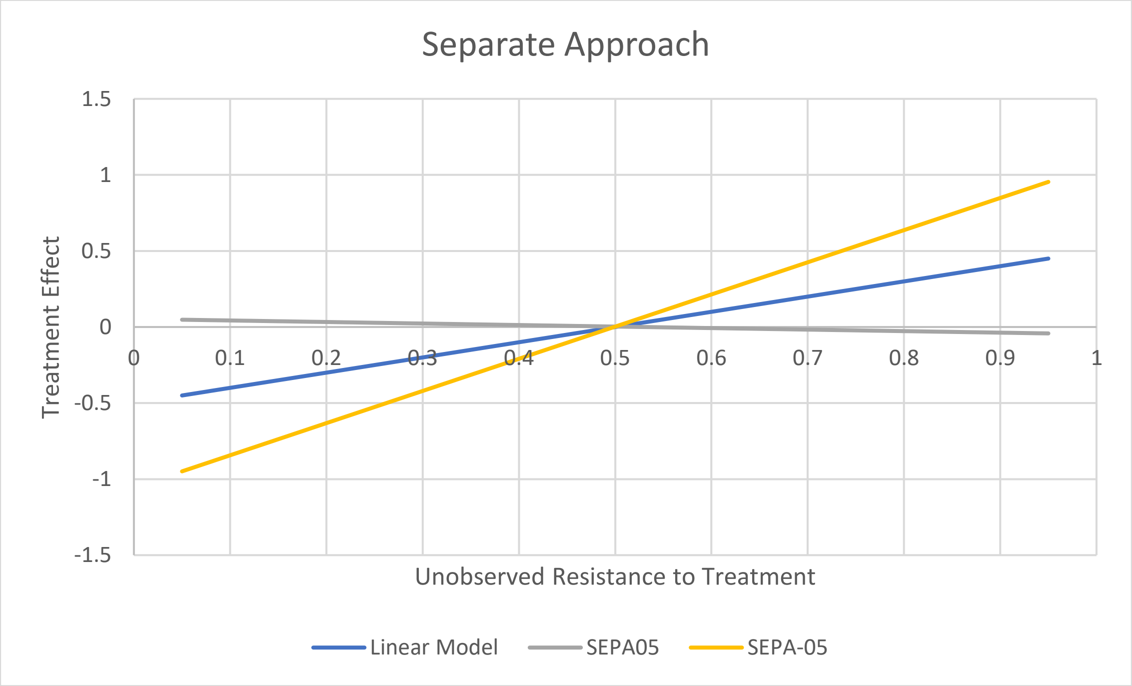
Figure 1: Common Support



Note: This figure shows the degree of common support in the linear model without misspecification. The analogous distributions are very similar for the quadratic model and for all specifications.

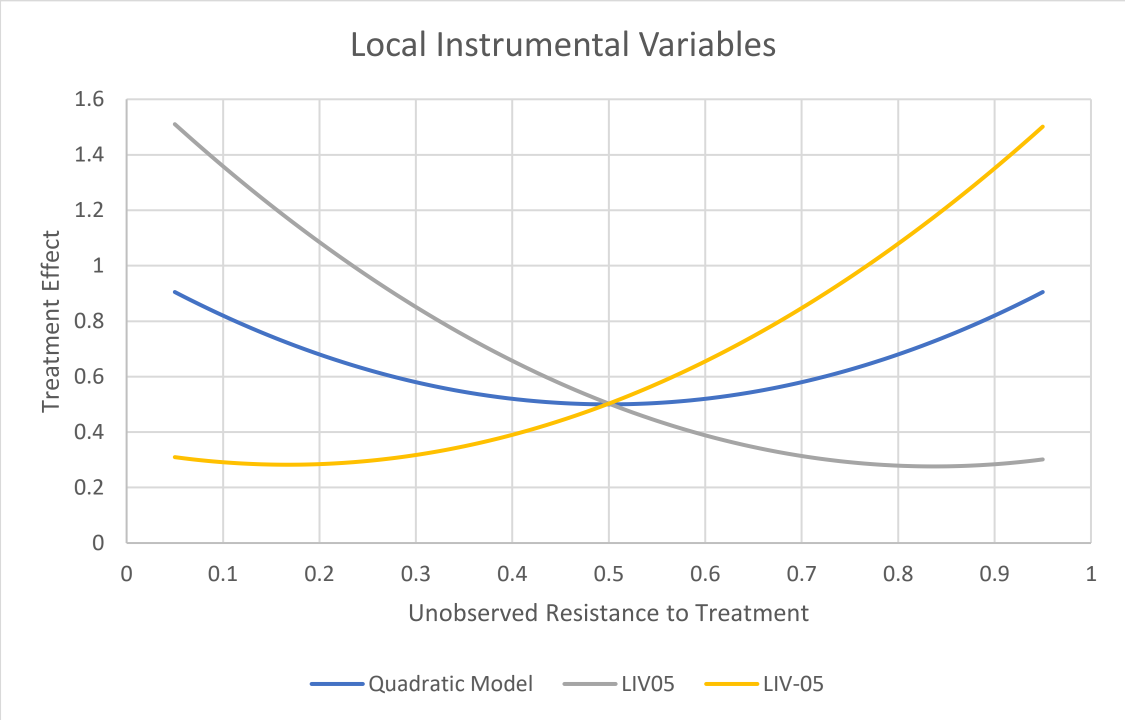
Figure 2: MTE Curve: Linear Model

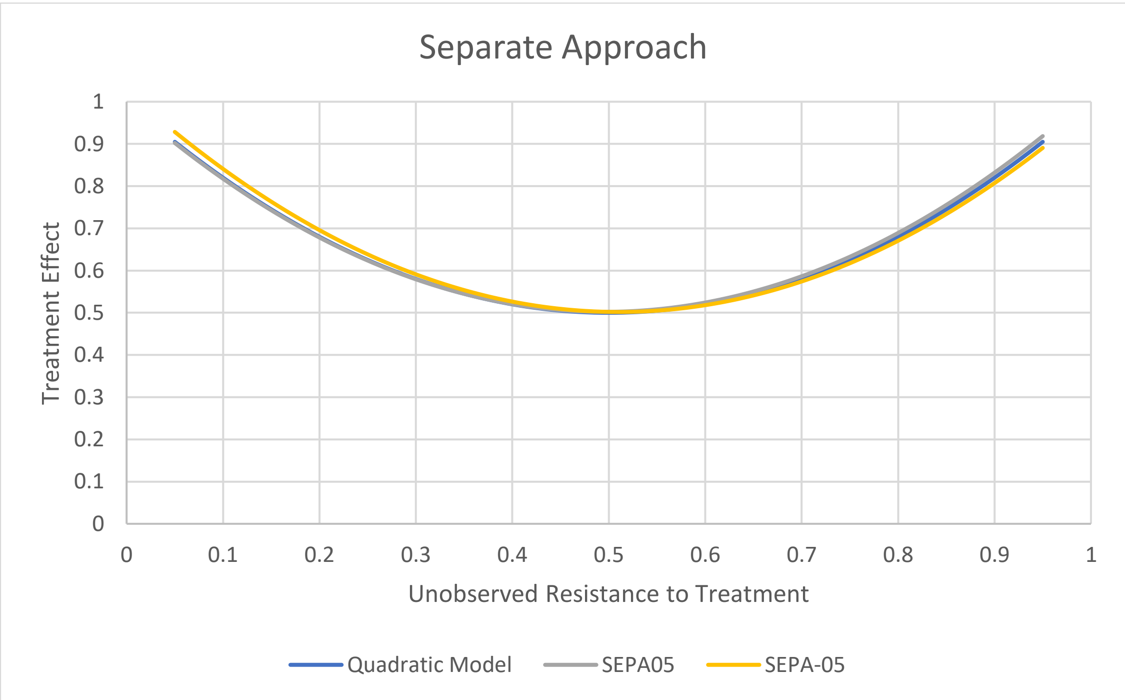




Note: Average estimates from 100 Monte Carlo simulations on a sample size of 100,000. The "Linear Model" line is the true MTE curve. LIV05 denotes Local Instrumental Variables with $\theta_1 = 0.05$, LIV-05 denotes Local Instrumental Variables with $\theta_1 = -0.05$, SEPA05 denotes the Separate Approach with $\theta_1 = 0.05$, SEPA-05 denotes the Separate Approach with $\theta_1 = -0.05$. All estimators impose a linear MTE curve.

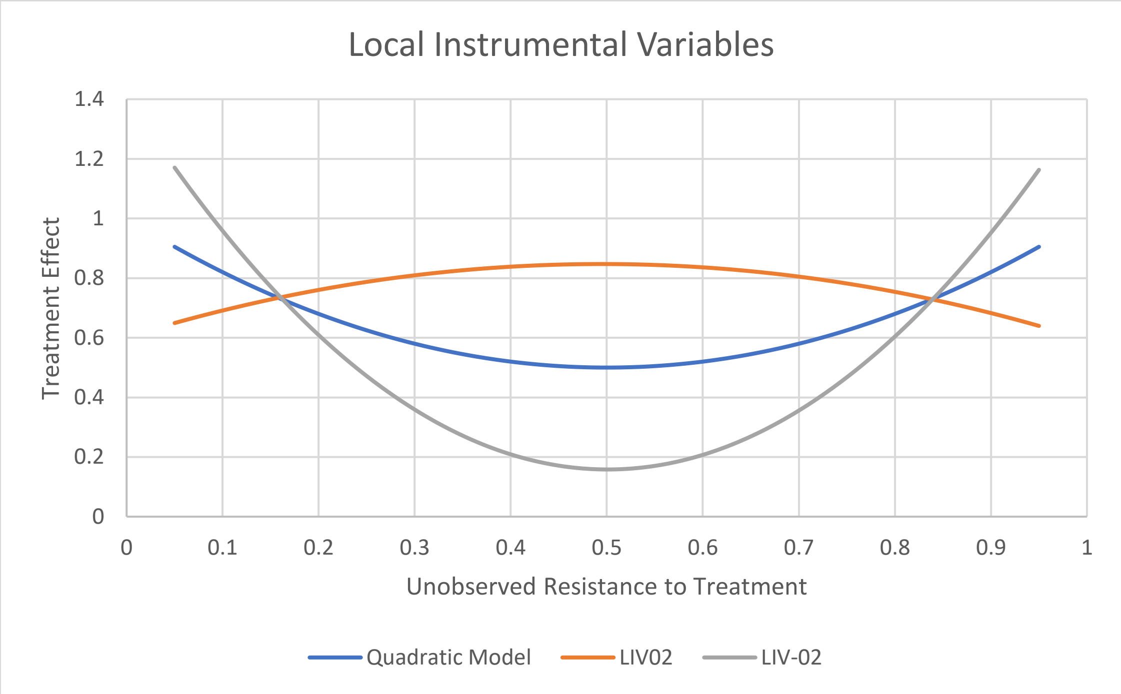
Figure 3: MTE Curve: Quadratic Model (θ_1 varied)

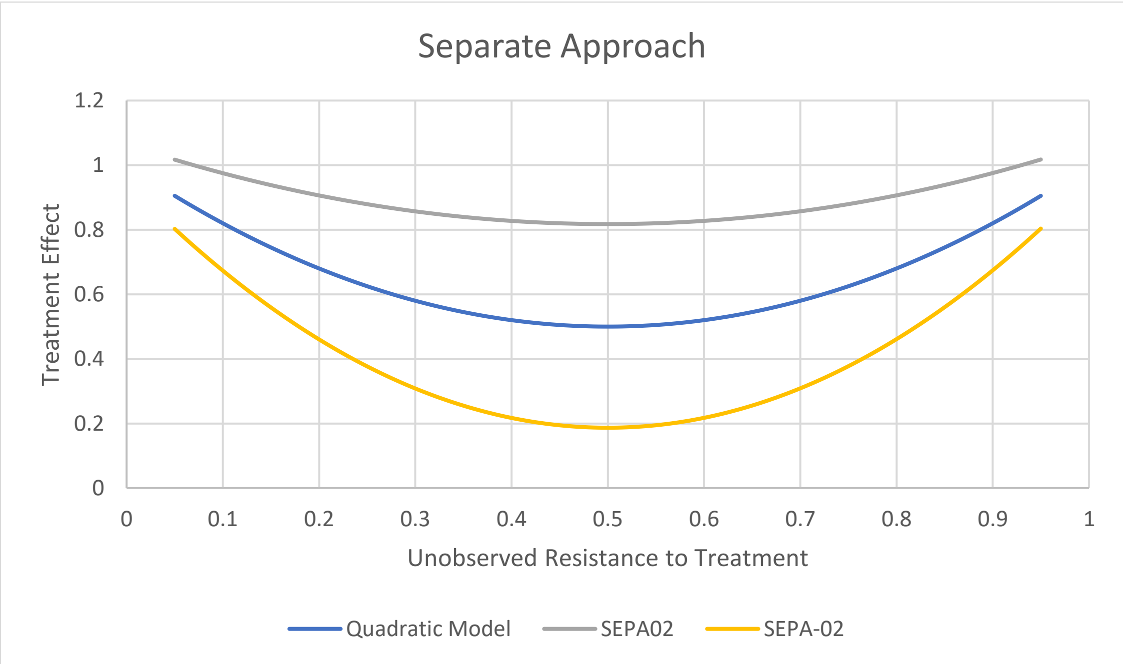




Note: Average estimates from 100 Monte Carlo simulations on a sample size of 100,000. The "Quadratic Model" curve is the true MTE curve. LIV05 denotes Local Instrumental Variables with $\theta_1 = 0.05$, LIV-05 denotes Local Instrumental Variables with $\theta_1 = -0.05$, SEPA05 denotes the Separate Approach with $\theta_1 = 0.05$, SEPA-05 denotes the Separate Approach with $\theta_1 = -0.05$. All estimators impose a quadratic MTE curve.

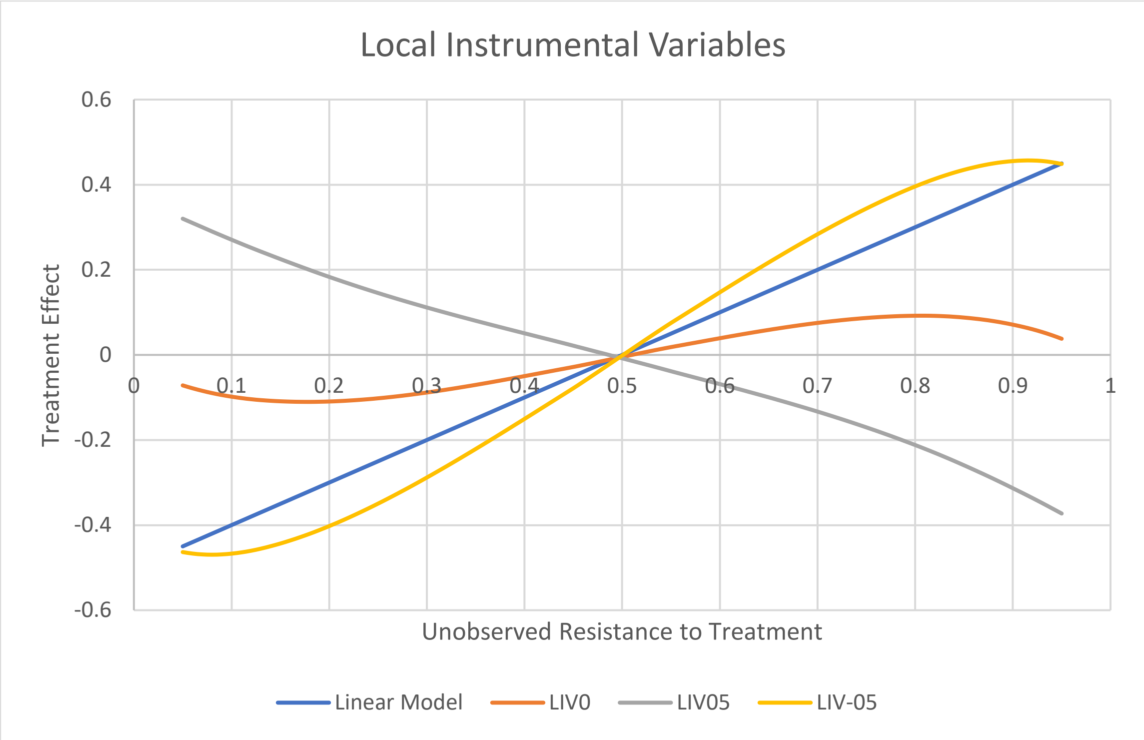
Figure 4: MTE Curve: Quadratic Model (θ_2 varied)

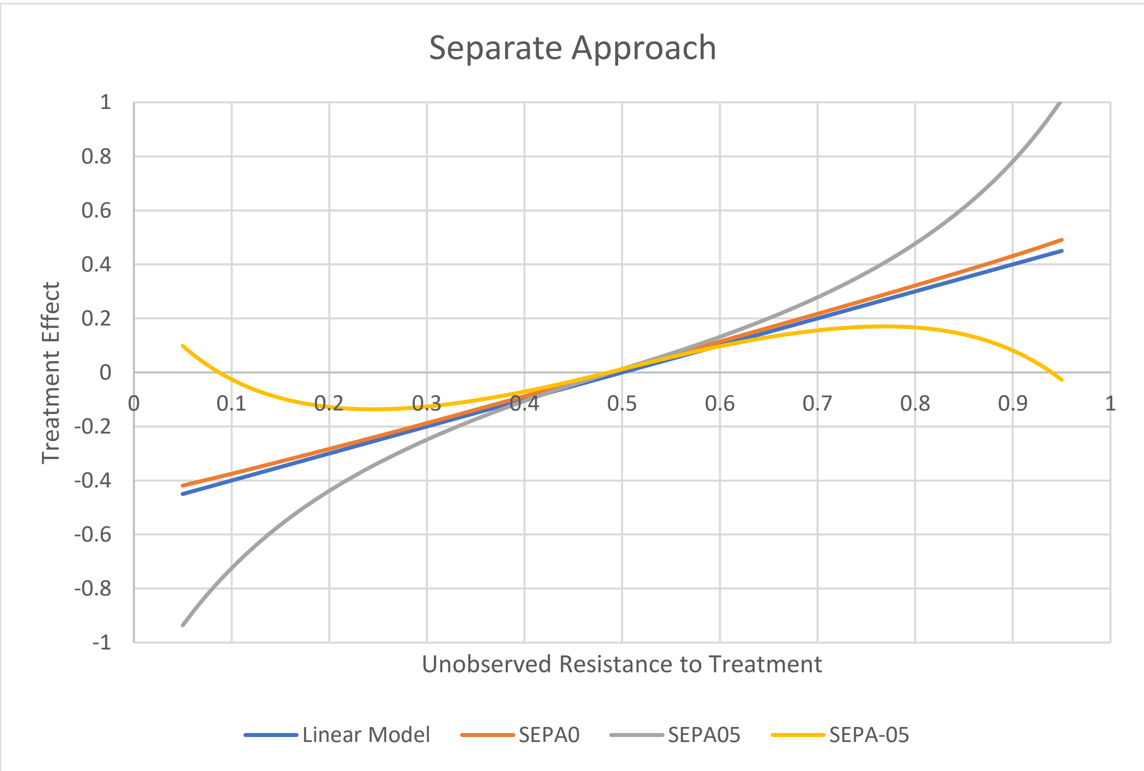




Note: Average estimates from 100 Monte Carlo simulations on a sample size of 100,000. The "Quadratic Model" curve is the true MTE curve. LIV02 denotes Local Instrumental Variables with $\theta_2 = 0.02$, LIV-02 denotes Local Instrumental Variables with $\theta_2 = -0.02$, SEPA02 denotes the Separate Approach with $\theta_2 = 0.02$, SEPA-02 denotes the Separate Approach with $\theta_2 = -0.02$. $\theta_1 = 0$ throughout. All estimators impose a quadratic MTE curve.

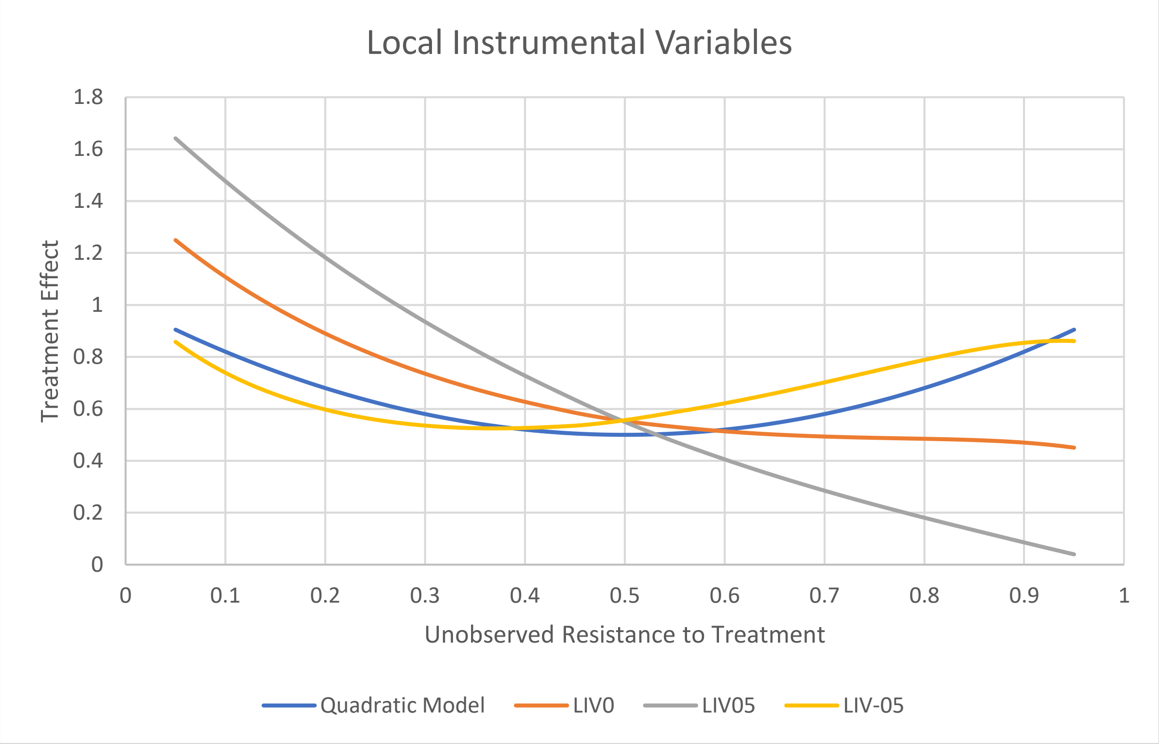
Figure 5: MTE Curve: Linear Model using Semiparametric Method

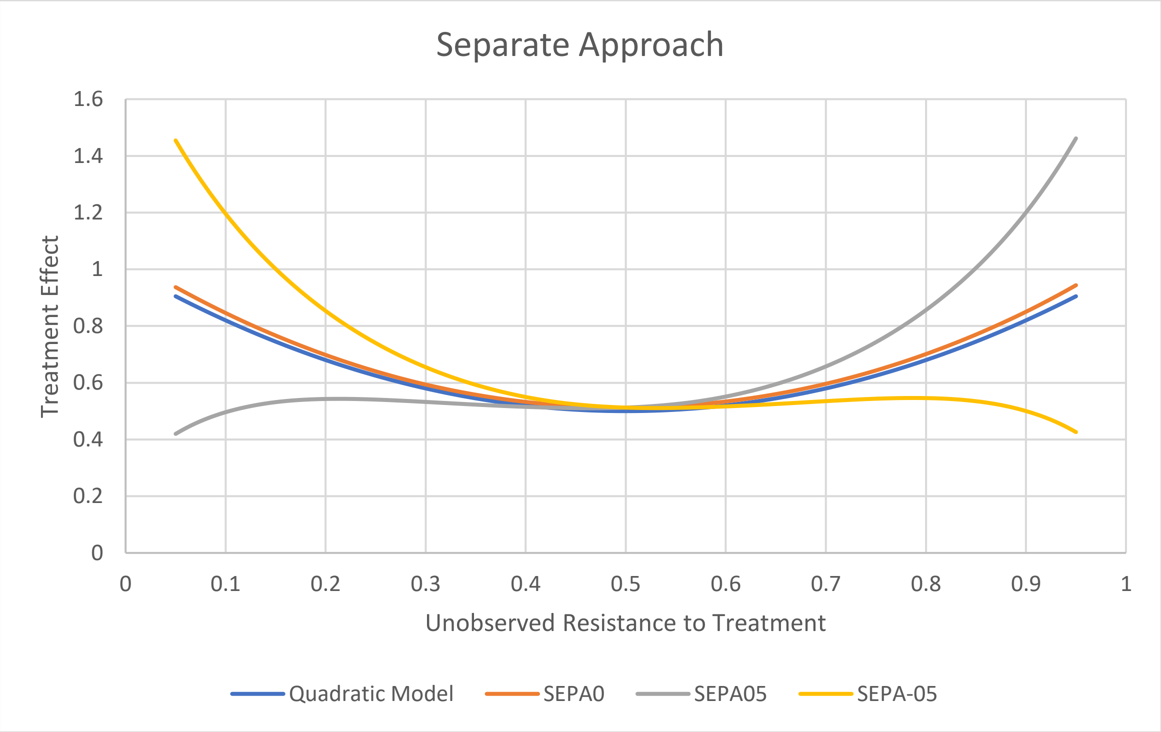




Note: Average estimates from 100 Monte Carlo simulations on a sample size of 100,000. The "Linear Model" line is the true MTE curve. LIV0 denotes Local Instrumental Variables with $\theta_1 = 0$, LIV05 denotes Local Instrumental Variables with $\theta_1 = 0.05$, LIV-05 denotes Local Instrumental Variables with $\theta_1 = -0.05$, SEPA0 denotes the Separate Approach with $\theta_1 = 0$, SEPA05 denotes the Separate Approach with $\theta_1 = 0.05$, SEPA-05 denotes the Separate Approach with $\theta_1 = -0.05$.

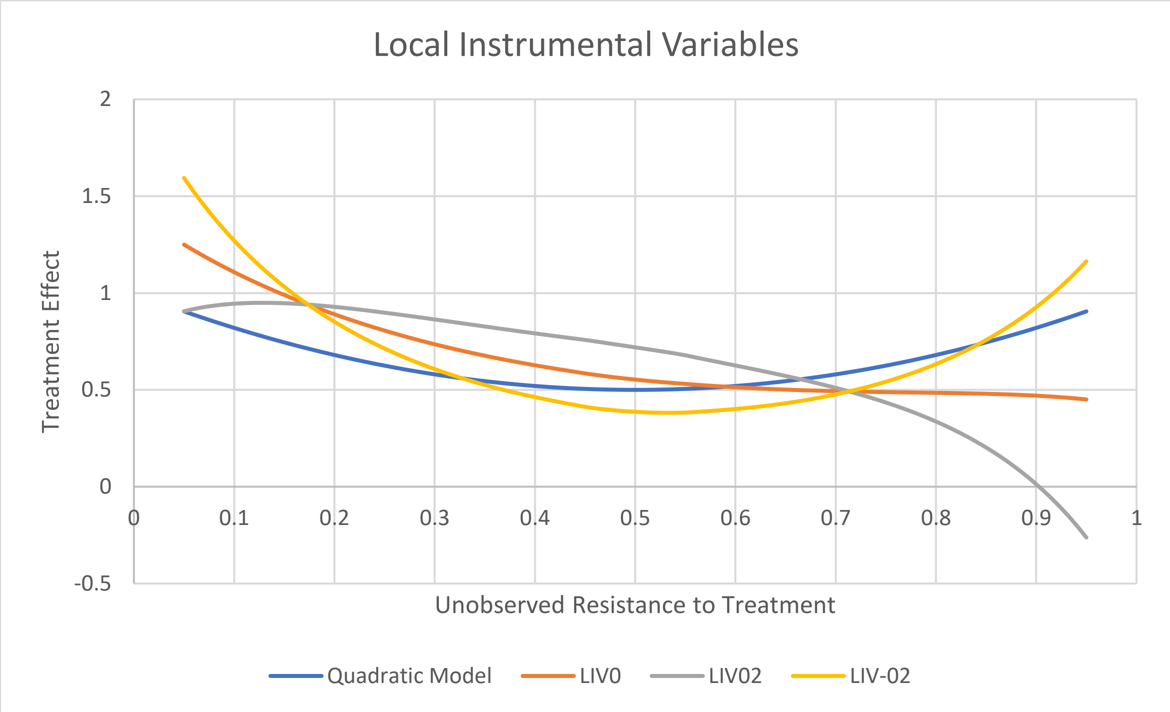
Figure 6: MTE Curve: Quadratic Model using Semiparametric Method (θ_1 varied)

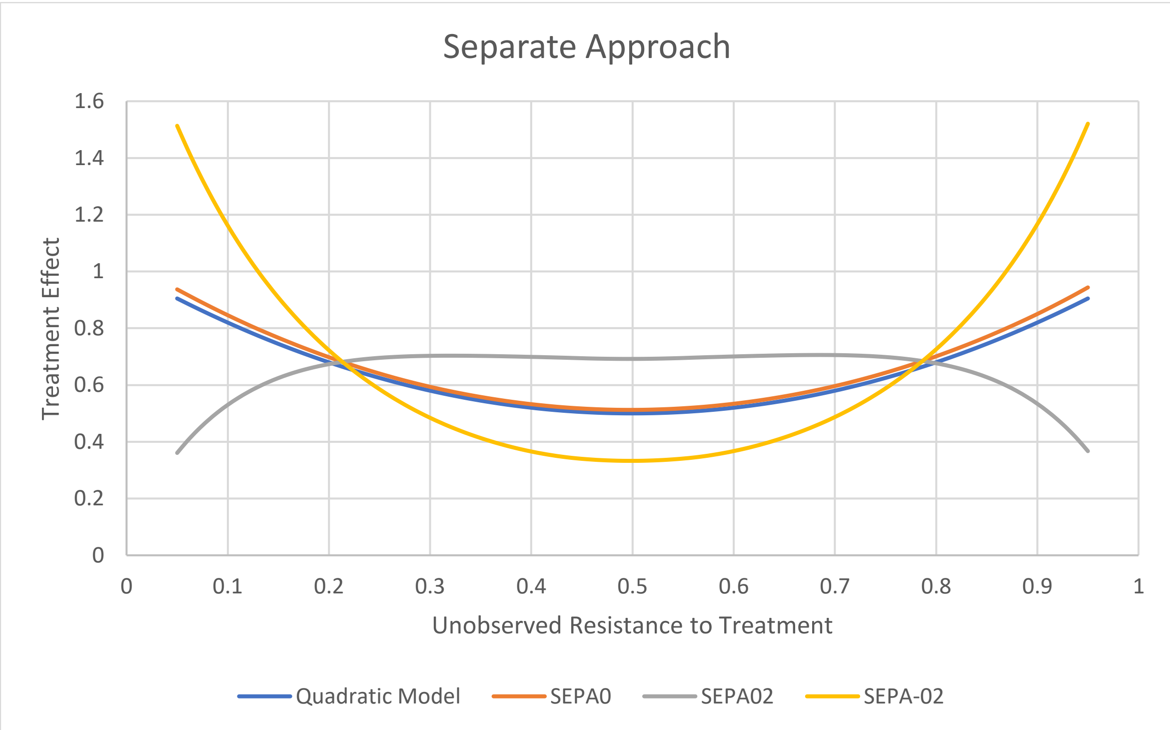




Note: Average estimates from 100 Monte Carlo simulations on a sample size of 100,000. The "Quadratic Model" line is the true MTE curve. LIV0 denotes Local Instrumental Variables with $\theta_1 = 0$, LIV05 denotes Local Instrumental Variables with $\theta_1 = 0.05$, LIV-05 denotes Local Instrumental Variables with $\theta_1 = -0.05$, SEPA0 denotes the Separate Approach with $\theta_1 = 0$, SEPA05 denotes the Separate Approach with $\theta_1 = 0.05$, SEPA-05 denotes the Separate Approach with $\theta_1 = -0.05$.

Figure 7: MTE Curve: Quadratic Model using Semiparametric Method (θ_2 varied)





Note: Average estimates from 100 Monte Carlo simulations on a sample size of 100,000. The "Quadratic Model" line is the true MTE curve. LIV0 denotes Local Instrumental Variables with $\theta_2 = 0$, LIV02 denotes Local Instrumental Variables with $\theta_2 = 0.02$, LIV-02 denotes Local Instrumental Variables with $\theta_2 = -0.02$, SEPA0 denotes the Separate Approach with $\theta_2 = 0$, SEPA02 denotes the Separate Approach with $\theta_2 = 0.02$, SEPA-02 denotes the Separate Approach with $\theta_2 = -0.02$. $\theta_1 = 0$ throughout.

Table 1: Average Estimates from the Linear Model

	True Value	LIV	LIV	LIV	SEPA	SEPA	SEPA
θ_1		0	0.05	-0.05	0	0.05	-0.05
θ_2		0	0	0	0	0	0
MTE Slope	1.00	1.00	-0.33	2.33	1.01	-0.10	2.11
ATE	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ATT	-0.16	-0.16	0.26	-0.58	-0.16	0.17	-0.49
ATUT	0.16	0.16	-0.26	0.58	0.16	-0.16	0.49
$\hat{\beta}_1 - \hat{\beta}_0$	0.00	0.00	-0.37	0.37	0.00	-0.27	0.27
Standard Error ATE		0.02	0.02	0.02	0.02	0.02	0.02
Standard Error ATT		0.04	0.04	0.04	0.03	0.03	0.03
Standard Error ATUT		0.04	0.04	0.04	0.03	0.03	0.03
Standard Error MTE Slope		0.14	0.14	0.14	0.08	0.08	0.08
Standard Error $\hat{\beta}_1 - \hat{\beta}_0$		0.02	0.02	0.02	0.01	0.01	0.01
R^2 without X^2 included		0.77	0.77	0.77	0.77	0.77	0.77
R^2 when X^2 included		0.77	0.77	0.77	0.77	0.77	0.77

Note: Average estimates from 100 Monte Carlo simulations on a sample size of 100,000. LIV denotes Local Instrumental Variables. SEPA denotes the separate approach to MTE estimation.

Table 2: Average Estimates from the Quadratic Model (θ_1 varied)

	True Value	LIV	LIV	LIV	SEPA	SEPA	SEPA
θ_1		0	0.05	-0.05	0	0.05	-0.05
θ_2		0	0	0	0	0	0
ATE	0.67	0.67	0.67	0.67	0.67	0.67	0.67
ATT	0.67	0.67	1.09	0.25	0.67	0.77	0.57
ATUT	0.67	0.66	0.24	1.08	0.67	0.57	0.77
$\hat{\beta}_1 - \hat{\beta}_0$	0.00	0.00	-0.37	0.37	0.00	-0.19	0.19
Standard Error ATE		0.03	0.03	0.03	0.02	0.02	0.02
Standard Error ATT		0.04	0.04	0.04	0.03	0.03	0.03
Standard Error ATUT		0.04	0.04	0.04	0.03	0.03	0.03
Standard Error $\hat{\beta}_1 - \hat{\beta}_0$		0.02	0.02	0.02	0.01	0.01	0.01
R^2 without X^2 included		0.74	0.73	0.73	0.74	0.73	0.73
R^2 when X^2 included		0.74	0.74	0.74	0.74	0.74	0.74

Note: Average estimates from 100 Monte Carlo simulations on a sample size of 100,000. LIV denotes Local Instrumental Variables. SEPA denotes the separate approach to MTE estimation.

Table 3: Average Estimates from the Quadratic Model (θ_2 varied)

	True Value	LIV	LIV	LIV	SEPA	SEPA	SEPA
θ_1		0	0	0	0	0	0
θ_2		0	0.02	-0.02	0	0.02	-0.02
ATE	0.67	0.67	0.77	0.57	0.67	0.90	0.44
ATT	0.67	0.67	0.77	0.57	0.67	0.90	0.44
ATUT	0.67	0.66	0.76	0.56	0.67	0.90	0.44
$\hat{\beta}_1 - \hat{\beta}_0$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Standard Error ATE		0.03	0.03	0.03	0.02	0.02	0.02
Standard Error ATT		0.04	0.04	0.04	0.03	0.03	0.03
Standard Error ATUT		0.04	0.04	0.04	0.03	0.03	0.03
Standard Error $\hat{\beta}_1 - \hat{\beta}_0$		0.02	0.02	0.02	0.01	0.01	0.01
R^2 without X^2 or X^3 included		0.74	0.76	0.71	0.74	0.76	0.71
R^2 when X^2 and X^3 are included		0.74	0.76	0.71	0.74	0.76	0.71

Note: Average estimates from 100 Monte Carlo simulations on a sample size of 100,000. LIV denotes Local Instrumental Variables. SEPA denotes the separate approach to MTE estimation.

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP21/07 Ellen Ryan and Karl Whelan: 'A Model of QE, Reserve Demand and the Money Multiplier' February 2021
WP21/08 Cormac Ó Gráda and Kevin Hjortshøj O!Rourke: 'The Irish Economy During the Century After Partition' April 2021
WP21/09 Ronald B Davies, Dieter F Kogler and Ryan Hynes: 'Patent Boxes and the Success Rate of Applications' April 2021
WP21/10 Benjamin Elsner, Ingo E Isphording and Ulf Zölitz: 'Achievement Rank Affects Performance and Major Choices in College' April 2021
WP21/11 Vincent Hogan and Patrick Massey: 'Soccer Clubs and Diminishing Returns: The Case of Paris Saint-Germain' April 2021
WP21/12 Demid Getik, Marco Islam and Margaret Samahita: 'The Inelastic Demand for Affirmative Action' May 2021
WP21/13 Emmanuel P de Albuquerque: 'The Creation and Diffusion of Knowledge - an Agent Based Modelling Approach' May 2021
WP21/14 Tyler Anbinder, Dylan Connor, Cormac Ó Gráda and Simone Wegge: 'The Problem of False Positives in Automated Census Linking: Evidence from Nineteenth-Century New York's Irish Immigrants' June 2021
WP21/15 Morgan Kelly: 'Devotion or Deprivation: Did Catholicism Retard French Development?' June 2021
WP21/16 Bénédicte Apouey and David Madden: 'Health Poverty' July 2021
WP21/17 David Madden: 'The Dynamics of Multidimensional Poverty in a Cohort of Irish Children' August 2021
WP21/18 Vessela Daskalova and Nicolaas J Vriend: 'Learning Frames' August 2021
WP21/19 Sanghamitra Chattopadhyay Mukherjee: 'A Framework to Measure Regional Disparities in Battery Electric Vehicle Diffusion in Ireland' August 2021
WP21/20 Karl Whelan: 'Central Banks and Inflation: Where Do We Stand and How Did We Get Here?' August 2021
WP21/21 Tyler Anbinder, Cormac Ó Gráda and Simone Wegge: '"The Best Country in the World": The Surprising Social Mobility of New York's Irish Famine Immigrants' August 2021
WP21/22 Jane Dooley and David Madden: 'Ireland's Post Crisis Recovery, 2012-2019: Was It Pro-Poor?' September 2021
WP21/23 Matthew Shannon: 'The Impact of Victimisation on Subjective Well-Being' September 2021
WP21/24 Morgan Kelly: 'Persistence, Randomization, and Spatial Noise' October 2021 *(For revised version of this paper see WP21/25)*
WP21/25 Morgan Kelly: 'Persistence, Randomization, and Spatial Noise' November 2021
WP21/26 Eliane Badaoui and Frank Walsh: 'Productivity, Non-Compliance and the Minimum Wage' November 2021
WP21/27 Annette Broocks and Zuzanna Studnicka: 'Gravity and Trade in Video on Demand Services' December 2021
WP21/28 Linda Mastrandrea: 'Linking Retail Electricity Pricing and the Decarbonisation of the Energy Sector: a Microeconomic Approach' December 2021
WP22/01 Doina Caragea, Theodor Cojoianu, Mihai Dobri, Andreas Hoepner and Oana Peia 'Competition and Innovation in the Financial Sector: Evidence from the Rise of FinTech Start-ups' January 2022
WP22/02 Sandra E Black, Paul J Devereux, Fanny Landaud and Kjell G Salvanes: 'The (Un)Importance of Inheritance' January 2022
WP22/03 Claes Ek and Margaret Samahita: 'Pessimism and Overcommitment: An Online Experiment with Tempting YouTube Content' January 2022

UCD Centre for Economic Research
Email economics@ucd.ie