University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2002 *Paper* 104

Estimating Causal Parameters in Marginal Structural Models with Unmeasured Confounders Using Instrumental Variables

Tanya A. Henneman* Mark Johannes van der Laan†
Alan E. Hubbard‡

*University of California, Berkeley, henneman@stat.Berkeley.EDU
†University of California, Berkeley, laan@berkeley.edu
‡University of California, Berkeley, hubbard@stat.berkeley.edu

http://biostats.bepress.com/ucbbiostat/paper104

Estimating Causal Parameters in Marginal Structural Models with Unmeasured Confounders Using Instrumental Variables

Tanya A. Henneman, Mark Johannes van der Laan, and Alan E. Hubbard

Abstract

For statisticians analyzing medical data, a significant problem in determining the causal effect of a treatment on a particular outcome of interest, is how to control for unmeasured confounders. Techniques using instrumental variables (IV) have been developed to estimate causal parameters in the presence of unmeasured confounders. In this paper we apply IV methods to both linear and non-linear marginal structural models. We study a specific class of generalized estimating equations that is appropriate to these data, and compare the performance of the resulting estimator to the standard IV method, a two-stage least squares procedure. Our results are applied to simulation studies and a data analysis example comparing treatment procedures for ruptured cerebral aneurysms.

1 Introduction

An important use of statistical methods in health and medical studies is to infer the effects of medical treatments or potentially toxic exposures. Increasingly, observational data arising from hospitals and other medical institutions are being used in statistical analyses to determine the effect of a treatment on a particular outcome of interest. One problem with observational studies is the lack of treatment randomization, thus a big concern in estimating treatment effect using these data is how to control for unmeasured confounders. Even in studies where treatment assignment is randomized, confounding can occur if there is non-compliance. Conventional analysis of this data may produce biased results [5, 8].

Causal inference methodology contains a framework designed to estimate parameters with a causal interpretation. Rubin [22] introduced the assumption that each subject carries with them a range of potential outcomes, one for each type of treatment or exposure being studied. At most one such outcome is observed, the rest are counterfactual outcomes. Robins [21] built upon this idea of modeling counterfactual outcomes to create a class of marginal structural models (MSM). MSMs make use of data with measured confounders, however, situations can arise in which differences between treatment and control groups are due in part to unmeasured confounding variables. For measured confounders the inverse probability of treatment weighted estimator can be used to construct consistent and asymptotically linear estimates of treatment effect [21]. A different approach is needed when unmeasured confounders are present.

In models with unmeasured confounders, techniques from Econometrics using instrumental variables have been developed to construct consistent parameter estimates. Instrumental variables (IV) are defined as variables related to the outcome of interest only through the treatment or exposure variable. Recently, researchers have applied these methods to identifying causal parameters arising from counterfactual models [20, 4, 13, 15]. For linear regression models, the standard IV method is a two-stage least squares regression (TSLS). However, for some non-linear models this approach can be biased. For non-linear models, it has been suggested that an estimating equation (EE) approach should be used [1, 19, 7].

This paper presents a generalized approach to estimating causal parameters in the presence of unmeasured confounders for any marginal structural models, linear or non-linear. We study a class of generalized estimating equations that utilize IVs and compare the performance of the resulting estimator to the TSLS estimator in several simulation studies. Section 2 provides the framework for the marginal structural model used throughout the paper. Section 3 presents two different IV methods, the TSLS and EE approach. Section 4 compares the performance of the IV estimator in simulations using linear and logistic MSMs. Section 5 presents a data analysis example which compares the risk of in-hospital death for two different treatment procedures for ruptured cerebral aneurysms. Section 6 offers a few concluding remarks.

IV methods are an increasingly popular tool in statistics and epidemiology to control for unmeasured confounding when estimating treatment effect [18, 24, 12, 11]. Our aim is to demonstrate that both the TSLS and EE are effective IV estimators for linear MSMs. For binary outcome models both estimators are inconsistent, except under the null of no treatment effect. For logistic models with bounded and continuous outcomes the EE is a consistent estimator, whereas the TSLS is not.

2 Model

Let A represent the treatment variable with $\mathcal{A}$ representing the set of all possible treatment regimes. For the purpose of this paper we consider the case where $\mathcal{A} = \{0, 1\}$. We present our methods for a dichotomous treatment variable, recognizing that our results can be extended for ordinal or continuous treatment variables. Let Y be the outcome variable of interest and V a set of measured covariates. Using Rubin's Causal Model [14] we assume for each subject the existence of treatment specific outcomes Y_a for all $a \in \mathcal{A}$. Let R be an instrument, a variable related to the treatment or exposure variable, but not independently related to the outcome variable. From the counterfactual perspective, this means that the instrument R is independent of all counterfactual outcomes. Let (A_i, Y_i, R_i, V_i) denote the data observed for for $i = 1, \ldots, n$ subjects. The counterfactual response is modeled by

$$Y_a = m(a, V|\beta) + \epsilon, \tag{1}$$

where $\epsilon = Y_0 - E(Y_0|V)$ and $m(A, V|\beta)$ represents a parametric statistical model. The following instrumental variable assumption is made

$$\begin{aligned} E(\epsilon|R, V) &= E(\epsilon|V) \\ &= 0. \end{aligned}$$

For this model β is our causal parameter of interest and we are interested in estimating β in the situation where $E(\epsilon|A, V) \neq 0$. There is a more complete discussion regarding the assumptions needed for causal interpretation of IV estimates for linear models elsewhere [15].

There may arise situations where ϵ is dependent on A and (1) becomes $Y_a = m(a, V|\beta) + \epsilon_a$, but for the purpose of this paper we only consider the case where $\epsilon_a = \epsilon$.

3 Methods

IV methods have been utilized to deal with a variety of problems that can arise in experimental studies, such as missing covariates, and unmeasured confounding. One very practical application of these methods is for analyzing data from studies in which there is non-compliance, where one can take as an instrument the treatment originally assigned to the subject. In this section we will discuss two competing IV estimators, the TSLS estimator and an EE estimator.

3.1 Two-Stage Least Squares Approach

For linear models IV methods are well understood. The primary method used is a two-stage least squares regression. To understand this approach, consider the parametric regression model

$$Y = \beta_0 + \beta_1 A + \beta_2 V + \epsilon, \tag{2}$$

where $E(\epsilon|R, V) = 0$, but $E(\epsilon|A, V) \neq 0$. This can represent a situation in which the treatment is related to an unobserved variable that is also associated with the outcome. For instance, doctors may be less likely to proceed with a high-risk procedure on a patient who is more likely to die or experience complications. In this case ordinary least squares (OLS) can produce biased estimates of β.

To illustrate the TSLS procedure consider (2). The first stage involves an OLS regression of A on R and V to obtain $\hat{E}(A|R, V)$. The second stage performs an OLS of Y on $\hat{E}(A|R, V)$ and

V, which yields an estimate of β. One way to see why this approach works for linear models is by taking conditional expectations to (2) which yields

$$E(Y|R,V) = \beta_0 + \beta_1 E(A|R,V) + \beta_2 V.$$

Thus, performing a TSLS regression produces an estimate of the original parameter of interest. One benefit of TSLS is that it is easy to implement using existing statistical software. However, for models that are non-linear in both the parameter and variable, for example a logistic model, regression of Y on $E(A|R,V)$ and V can produce biased estimates of the coefficients [2, 7]. This will be discussed further in section 3.3.

3.2 Estimating Equation Approach

For non-linear models it has been suggested that an estimating equation approach be used [1, 2, 19, 7]. For $\mathcal{A} = \{0,1\}$ consider the MSM given in (1). It can be shown that the estimating function $\phi(R,V)\epsilon(\beta)$ is an unbiased estimating function of β for any function of the instrument, $\phi(R,V)$. This follows from the assumption $E(\epsilon|R,V) = 0$. Thus, the estimator $\hat{\beta}$ solving

$$\frac{1}{n}\sum_{i=1}^{n} \phi(R_i, V_i)\epsilon_i(\beta) = 0 \tag{3}$$

is a consistent estimator of β under standard regularity conditions. For the more general class of models where ϵ is not merely a function of Y_0, this method is inconsistent since the IV assumption will be violated.

Let β be k-dimensional. Then

$$E\left[\phi(R,V)\frac{d}{d\beta}\epsilon(\beta)\right] \tag{4}$$

is a $k \times k$ matrix where the (i,j) element is given by $E\left(\phi_i(R,V)\frac{d}{d\beta_j}\epsilon(\beta)\right)$, for $i,j = 1,\ldots,k$. The influence curve for $\hat{\beta}$ is given as

$$IC(Y|\beta,\phi) = E\left[\phi(R,V)\frac{d}{d\beta}\epsilon(\beta)\right]^{-1} \cdot \phi(R,V)\epsilon(\beta), \tag{5}$$

provided that (4) is invertible. Note, the invertibility condition requires R to be related to A. Under standard regularity conditions and given $Var(IC_j(Y|\beta,\phi)) < \infty$ for $j = 1,\ldots,k$, $\hat{\beta}$ is asymptotically linear with

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta) &= \frac{1}{n}\sum_{i=1}^{n} IC_i(Y|\beta,\phi) + o_p(1) \\ &\Rightarrow N(\overrightarrow{0}, \Sigma), \end{aligned}$$

where $\Sigma = E(IC \cdot IC^t)$. More details regarding the requirements for $\sqrt{n}-$consistency and asymptotic linearity of $\hat{\beta}$ are given in Appendix A.

3.2.1 Implementation

An important consideration when using the EE approach is the selection and estimation of $\phi(R, V)$. In a standard regression setting where in (1) $E(\epsilon|A, V) = 0$, to obtain consistent estimates of β using the least squares method one would solve

$$\frac{1}{n}\sum_{i=1}^{n} h(A_i, V_i)\epsilon_i(\beta) = 0$$

where

$$h(A, V) = \frac{d}{d\beta}m(A, V|\beta) \text{ or } h(A, V) = \frac{\frac{d}{d\beta}m(A, V|\beta)}{Var(\epsilon|A, V)}.$$

This motivates choices for ϕ [2, 19] such as,

$$\phi(R, V) = E\left(\frac{d}{d\beta}m(A, V|\beta)\middle|R, V\right), \tag{6}$$

or another possibility,

$$\phi(R, V) = E\left(\frac{\frac{d}{d\beta}m(A, V|\beta)}{Var(\epsilon|A, V)}\middle|R, V\right). \tag{7}$$

For example, suppose we were interested in estimating parameters from a linear marginal structural model, $Y_a = \beta_0 + \beta_1 a + \beta_2 V + \epsilon$. OLS solves

$$\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\A_i\\V_i\end{pmatrix}\epsilon_i(\beta) = 0.$$

Using the EE approach one could choose ϕ using (6) to obtain,

$$\phi(R, V) = \begin{pmatrix}1\\E(A|R, V)\\V\end{pmatrix}.$$

The influence curve for $\hat{\beta}$ would be

$$IC_i(Y|\beta, \phi) = \left[\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}-1 & -A_i & -V_i\\ -E(A_i|R_i, V_i) & -A_i \cdot E(A_i|R_i, V_i) & -V_i \cdot E(A_i|R_i, V_i)\\ -V_i & -V_i \cdot A_i & -V_i^2\end{pmatrix}\right]^{-1}\begin{pmatrix}1\\A_i\\V_i\end{pmatrix}\epsilon_i(\beta)$$

Various models can be used to estimate $E(A|R, V)$, such as a linear model. A generalized additive model can be used to avoid making any functional form assumptions regarding $E(A|R, V)$. Alternative non-parametric methods and their benefits have been discussed elsewhere [19]. Once $\phi(R, V)$ is selected and estimated from the data a Newton-Raphson algorithm can be used to solve the estimating equation in (3). A starting value, say $\hat{\beta}^0$, can be obtained from performing the TSLS method. To calculate the next estimate of β we compute,

$$\hat{\beta}^1 = \hat{\beta}^0 - \left[\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}(R_i, V_i)\frac{d}{d\beta}\epsilon_i(\hat{\beta}^0)\right]^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}(R_i, V_i)\epsilon_i(\hat{\beta}^0).$$

We continue with an iterative process for $m = 1, 2, 3, \ldots$

$$\hat{\beta}^{m+1} = \hat{\beta}^m - \left[\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}(R_i, V_i)\frac{d}{d\beta}\epsilon_i(\hat{\beta}^m)\right]^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}(R_i, V_i)\epsilon_i(\hat{\beta}^m)$$

until the convergence criteria $||\beta^{m+1} - \beta^m|| < \varepsilon$ and $||\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}(R_i, V_i)\epsilon_i(\beta^{m+1})|| < \varepsilon$ are met for some small ε, where $||\cdot||$ denotes the Euclidean norm. At each iterative step we check to see if $||\frac{1}{n}\sum_{i=1}^{n}\phi(R_i, V_i)\epsilon(\hat{\beta}^{m+1})|| > ||\frac{1}{n}\sum_{i=1}^{n}\phi(R_i, V_i)\epsilon(\hat{\beta}^m)||$. If it is, then we choose a new updated $\hat{\beta}$ by performing a line search in the following manner. For $\varepsilon = (0.1, 0.2, \ldots, 0.9)$, compute

$$\beta^{\varepsilon} = \varepsilon\hat{\beta}^m + (1-\varepsilon)\hat{\beta}^{m+1}.$$

Choose the new β^{m+1} to be the β^{ε} that solves $\min_{\beta^{\varepsilon}} ||\frac{1}{n}\sum_{i=1}^{n}\phi(R_i, V_i)\epsilon_i(\beta^{\varepsilon})||$.

Once $\hat{\beta}$ is determined the influence curve can be used to construct standard error estimates and hence, confidence intervals. The influence curve given in (5) for each subject is estimated as

$$\hat{IC}_i(Y|\hat{\beta}, \hat{\phi}) = \left[\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}(R_i, V_i)\frac{d}{d\beta}\epsilon_i(\hat{\beta})\right]^{-1} \cdot \hat{\phi}(R_i, V_i)\epsilon_i(\hat{\beta}).$$

$\hat{\Sigma}$ is computed by taking the empirical covariance of $\hat{IC}$. A 95% confidence interval for β_1 is $\hat{\beta}_1 \pm 1.95\hat{\sigma}/\sqrt{n}$ where $\hat{\sigma}$ is the appropriate diagonal element of Σ.

3.3 IV Methods for Logistic Models

Logistic models are frequently used to evaluate the effectiveness of a particular treatment. They can be used to model binary or continuous and bounded outcomes. Examples of this type of data can be found in the literature. A logistic model was used to model the average leaf weight per plant in a study comparing the growth patterns of two genotypes of soybean [6]. In a study [9] determining the effect of chloroform and carbon tetrachloride on cell toxicity, a logistic model was used to model the percentage of lactic dehydrogenase enzyme leakage, a surrogate marker for cell toxicity.

Consider the following logistic model,

$$E(Y_a|V) = (1 + \exp(-(\beta_0 + \beta_1 a + \beta_2 V)))^{-1}. \tag{8}$$

Since

$$E[(1 + \exp(-(\beta_0 + \beta_1 A)))^{-1}|R, V] \neq (1 + \exp(-(\beta_0 + \beta_1 E[A|R, V] + \beta_2 V))^{-1},$$

performing a logistic regression of Y on $E(A|R, V)$ and V may yield a biased estimate of β. Thus an adapted version of the TSLS estimator can not guarantee a consistent estimate of treatment effect.

The EE approach only requires $E(\epsilon|R, V) = 0$ to hold to produce consistent estimates. In the case where Y is a bounded and continuous, we can select $\phi(R, V)$ using (7) and (8). This yields

$$\begin{aligned}\phi(R, V) &= E\left(\frac{\frac{d}{d\beta}[1 + \exp(-(\beta_0 + \beta_1 A + \beta_2 V)]^{-1}}{p(1-p)}\middle| R, V\right) \\ &= \begin{pmatrix} 1 \\ E(A|R, V) \\ V \end{pmatrix},\end{aligned}$$

where $p = (1 + exp(-(\beta_0 + \beta_1 A + \beta_2 V)))^{-1}$.

Now, for binary outcome models the EE estimator is also inconsistent. This is because the error term retains a dependence on A, and it would not be possible to find a variable related to treatment and not to the outcome, i.e. the assumption $E(\epsilon|R, V) = 0$ is violated. However, under the assumption of no treatment effect, when $\beta_1 = 0$, both the TSLS and EE estimators are consistent. This means that these procedures can be used for testing purposes.

4 Simulations

In this section we examine the applicability of the IV estimators for different counterfactual MSMs. The performance of the TSLS and EE are evaluated on data sets generated by Monte-Carlo simulations. The IV estimators are also contrasted to the standard estimating procedure used for the respective data types. Three different simulation studies are presented in this section. The first example uses a linear counterfactual model to generate the data. The last two examples use a logistic counterfactual model, one with binary outcomes and the other with continuous outcomes. To see how the estimators performed in finite samples we ran, several simulations varying the sample size and various model parameters. To evaluate the performance we computed for each estimator the bias $= \frac{1}{j}\sum_{i=1}^{j}(\hat{\beta}_i - \beta)$, mean squared error (mse) $= \frac{1}{j}\sum_{i=1}^{j}(\hat{\beta}_i - \beta)^2$, where j is the number of replications and β is the causal parameter of interest. Also computed was the relative mean squared error (rmse) as a ratio of the mse of the estimator to the mse of the EE, and 95% coverage probabilities. All computations were carried out using Splus (Version 3.4).

4.1 Example 1: Linear MSM

For this example we are interested in comparing the performance of the EE to the TSLS for linear counterfactual models. The data generating distribution is given by

$$\begin{aligned} Y_0 &= \beta_0 + \epsilon, \epsilon \sim N(0,1) \\ Y_1 &= Y_0 + \beta_1 \\ P(R=1) &= 0.5 \\ logit(P(A=1|R,Y_0)) &= -4 + 8R + Y_0, \end{aligned}$$

with $(\beta_0, \beta_1) = (1, 1)$. The data is generated to ensure that R satisfies the requirements of an instrumental variable. Also, the effect of treatment is confounded by the baseline counterfactual Y_0.

The typical method of estimation for this data is OLS. We refer to the OLS estimator as the naïve estimator. In addition to OLS, the general IV methods presented in sections 3.1 and 3.2 were used with the following specifications. For both the TSLS and EE, $E(A|R)$ was estimated in Splus using the `gam` function with `family=binomial(link=logit)`. The EE used the TSLS estimate of β as an initial estimate. The results in Table 1 are based on 1000 replications for each sample size.

As expected, the naïve estimator produced biased estimates of β_1, while the bias remains small for all sample sizes for the TSLS and EE. There are no remarkable differences in the relative efficiency of both IV estimators. For the larger sample sizes ($n > 500$) the EE has slightly better coverage probabilities.

Table 1: Estimating $\beta_1 = 1$ for linear MSM with varying sample size.

	Naïve			TSLS			EE		
n	RMSE	Bias	Coverage(%)	RMSE	Bias	Coverage(%)	n·MSE	Bias	Coverage(%)
100	1.310	0.144	87.2	1.000	0.001	95.8	4.884	0.001	93.8
500	3.012	0.140	65.7	1.000	-0.004	97.6	4.560	-0.004	96.4
1000	5.138	0.141	39.3	1.000	-0.0002	96.5	4.619	-0.0003	95.0
2000	9.737	0.141	10.3	1.000	-0.0004	96.5	4.455	-0.001	95.2
5000	21.987	0.141	0	1.000	-0.0005	95.8	4.722	-0.001	94.6
7000	29.598	0.141	0	0.999	-0.001	96.6	4.806	-0.001	96.2

4.2 Example 2: Logistic MSM

4.2.1 Binary Outcomes

For this next example, we used a binary outcome model, specifically, a logistic regression model. To evaluate how these procedures can be used for testing purposes, we first considered the situation with no treatment effect. The following data generating distributions were used.

$$\begin{aligned} P(Y_0 = 1) = P(Y_1 = 1) &= (1 + \exp(-\beta_0))^{-1} \\ P(R = 1) &= 0.5 \\ U &\sim N(Y_0, 1) \\ logit(P(A = 1|R, Y_0)) &= -3 + 5R + U, \end{aligned}$$

with $(\beta_0, \beta_1) = (1, 0)$ and where U represents an unmeasured confounder. The typical method of estimation for this model would be a logistic regression estimator which uses an iteratively reweighted least squares approach. We refer to this estimator as the naïve estimator. To analyze the data we also used the adapted TSLS and the EE approach described in section 3.3. with $E(A|R)$ estimated in the same manner as in the previous example. The results are presented in Table 2 and are based on 1000 replications of each sample size.

In Table 2, the naïve method produced biased estimates of β_1 whereas both IV estimators produce estimates with a very little bias. The TSLS estimator was only slightly more efficient than the EE for the smaller sample sizes. Noting the coverage probabilities, we observe that both IV estimators demonstrate their utility in determining if there exists a treatment effect statistically different from zero.

Next we considered the situation where A has a positive treatment effect. To generate data from a logistic counterfactual model with varying treatment effect, the following data generating distribution functions were used.

Table 2: Estimating β_1 for logistic MSM with no treatment effect and varying sample size.

	Naïve			TSLS			EE		
n	RMSE	Bias	Coverage(%)	RMSE	Bias	Coverage(%)	n·MSE	Bias	Coverage(%)
100	0.610	0.189	95.3	0.966	0.006	94.9	37.549	0.003	96.7
500	1.010	0.171	84.5	0.995	-0.006	94.1	35.173	-0.007	94.4
1000	1.457	0.172	76.3	0.998	0.001	95.2	34.323	0.0004	95.1
2000	2.351	0.173	58.4	0.999	-0.002	94.2	34.189	-0.002	94.1
5000	5.176	0.171	22.7	1.000	-0.001	95.2	31.869	-0.001	95.4
7000	6.733	0.172	10.5	1.000	-0.0004	95.9	33.707	-0.0005	95.9

$$\begin{aligned} P(Y_0 = 1) &= (1 + \exp(-\beta_0))^{-1} \\ P(Y_1 = 1) &= (1 + \exp(-(\beta_0 + \beta_1))^{-1} \\ P(R = 1) &= 0.5 \\ U &\sim N(Y_0, 1) \\ logit(P(A = 1|R, Y_0)) &= -3 + 5R + U, \end{aligned}$$

with $\beta_0 = 1$ and $\beta_1 > 0$. The same estimators were used as with the previous situation. The results are presented in Table 3 for varying treatment effect size, and are based on 1000 replications for a fixed sample size of $n = 5000$.

In Table 3 we observe that the performance of both IV estimators is not as strong as the case in which there is no treatment effect. However the estimators remain fairly unbiased for the smaller treatment effect sizes. Also, for the smaller treatment effect sizes, the MSE and coverage probabilities are similar to what was seen in Table 2. As the assumption of $\epsilon_a = \epsilon$ weakens, the performance of both IV estimators deteriorates with increased bias and variance.

Table 3: Estimating β_1 for logistic MSM with varying treatment effect size for $n = 5000$.

	Naïve			TSLS			EE		
β_1	RMSE	Bias	Coverage(%)	RMSE	Bias	Coverage(%)	n·MSE	Bias	Coverage(%)
0.2	5.389	0.173	23.7	1.003	-0.007	95.4	31.760	-0.008	95.4
0.5	4.762	0.178	25.0	0.996	-0.019	94.5	38.046	-0.020	94.3
0.7	3.849	0.175	30.0	0.990	-0.040	92.5	45.852	-0.040	92.2
1.0	2.633	0.173	35.3	1.016	-0.074	86.8	68.282	-0.070	86.8
2.0	0.570	0.175	61.9	1.466	-0.305	18.0	353.791	-0.238	49.1

4.2.2 Continuous Outcomes

For our final example we considered the case where Y represents a bounded and continuous outcome that is modeled by

$$E(Y_a) = (1 + \exp(-(\beta_0 + \beta_1 a)))^{-1}.$$

To generate this counterfactual data we used the following data generating distribution functions.

$$\begin{aligned} Y_0 &\sim (1+exp(\beta_0))^{-1} + N(\mu=0, \sigma=0.1) \\ Y_1 &= (1+exp(-(\beta_0+\beta_1))^{-1} + Y_0 - E(Y_0) \\ P(R=1) &= 0.5 \\ P(A=1|R,Y_0) &= (1+exp(-(-3+\alpha R+7Y_0))), \end{aligned}$$

with $(\beta_0, \beta_1) = (-1, 2)$. We considered two situations, $\alpha = 2$ and $\alpha = 5$ to compare the performance of the estimators when the correlation between A and R was weakened.

For our naïve estimator we used a non-linear least squares estimator. This estimator uses an iterative procedure to find an estimate that minimizes $||Y-(1+\exp(-(\beta_0+\beta_1 A)))^{-1}||$. The function `nls` in Splus was used to perform this estimation procedure. The TSLS and EE procedures used in 4.2.1 were used in this simulation as well.

In both Tables 4 and 5, the EE clearly outperforms the naïve and TSLS estimators. The EE approach produces estimates that are less biased and more efficient than estimates constructed with the TSLS approach. When the correlation between A and R is weaker, the difference in their performance becomes more remarkable.

Table 4: Estimating $\beta_1 = 2$ for a non-linear MSM with varying sample size.

	Naïve			TSLS			EE		
n	RMSE	Bias	Coverage(%)	RMSE	Bias	Coverage(%)	n·MSE	Bias	Coverage(%)
100	1.581	0.151	71.7	0.964	-0.056	99.9	2.156	-0.009	94.5
500	6.341	0.155	9.7	1.456	-0.051	99.8	2.083	0.00001	95.0
1000	11.717	0.155	0.6	2.100	-0.052	99.5	2.142	-0.001	94.0
2000	21.97	0.155	0	3.175	-0.051	98.1	2.241	-0.001	93.7
5000	56.319	0.155	0	7.037	-0.052	86.2	2.163	-0.001	93.9
7000	75.984	0.155	0	9.292	-0.052	71.1	2.231	-0.001	94.0

Table 5: Estimating $\beta_1 = 2$ for a non-linear MSM with varying sample size and weaker correlation between A and R.

	Naïve			TSLS			EE		
n	RMSE	Bias	Coverage(%)	RMSE	Bias	Coverage(%)	n·MSE	Bias	Coverage(%)
100	1.239	0.277	24.5	0.988	-0.142	100	7.134	-0.020	96.9
500	7.042	0.274	0	2.067	-0.123	100	5.499	0.003	95.5
1000	13.711	0.275	0	3.533	-0.126	100	5.596	-0.001	95.0
2000	27.680	0.273	0	6.645	-0.127	99.2	5.439	-0.002	96.3
5000	64.188	0.274	0	13.710	-0.124	72.5	5.876	0.002	95.6
7000	94.122	0.274	0	20.811	-0.127	35.7	5.588	-0.002	96.0

5 Comparing Treatments for Ruptured Cerebral Aneurysms

We present our findings in the context of an observational study comparing treatment procedures intended to repair ruptured intracranial aneurysms in patients with subarachnoid hemorrhage. The two procedures being compared are surgical clipping, the standard therapy, and endovascular therapy, a newer catheter-based treatment approach. Our analysis is based on data collected on 5383 patients from 70 centers belonging to the University Health System Consortium (UHC). All patients had a primary diagnosis of subarachnoid hemorrhage between the years 1992 and 1997, received either surgical clipping or endovascular therapy, and did not have a secondary diagnosis consistent with a source of hemorrhage other than an aneurysm. More details on the study design can be found elsewhere [17, 16].

Let A be the procedure a patient received; $A = 1$ means the patient received the surgery and $A = 0$ indicates the patient received endovascular therapy. Y, an indicator for in-hospital death, represents our outcome of interest. The probability of death is modeled as

$$logit(P(Y_a = 1)) = \beta_0 + \beta_1 a.$$

The parameter of interest, $\exp(\beta_1)$, is the causal odds ratio which provides a estimate of the ratio of the odds of in-hospital death for who received surgery to those who received the endovascular therapy. A typical analysis would involve logistic regression however, it is believed that treatment decisions could be based on the prognosis of the patient. Meaning, those patients who were at greater risk of dying from the ruptured aneurysm may have been less likely to be offered the surgical procedure. Thus the usual estimation procedure using logistic regression could produce a biased estimate of β_1.

There is variation across hospitals in the utilization of endovascular therapy, but it is believed that the quality of care is comparable for all hospitals, and thus should not have a direct effect on an individual's outcome. For this reason we selected the hospital at which the patient was treated for their ruptured aneurysm to serve as our instrument. However, we note that correlation between the hospital and treatment received was small (r=0.18).

To account for potential confounders V = (Age, Race, Source and Type of Admission, Institutional Volume, and Year of Admission) we were interested in fitting the following model,

$$logit(P(Y_a = 1|V)) = \beta_0 + \beta_1 a + \beta_2 \text{Age} + \beta_3 \text{Race} + \beta_4 \text{Source} + \beta_5 \text{Type} + \beta_6 \text{Volume} + \beta_7 \text{Year}.$$

The data was analyzed with the three estimators described in section 4.2.1, a standard logistic regression estimator, an adapted version of the TSLS estimator and the EE estimator with

$$\phi(R, V) = \begin{pmatrix} 1 \\ E(A|R, V) \\ \text{Age} \\ \text{Race} \\ \text{Source} \\ \text{Type} \\ \text{Volume} \\ \text{Year} \end{pmatrix}.$$

The results of this analysis are presented in Table 6. The naïve approach, a standard logistic regression estimator, yields a nearly statistically significant conclusion that those patients receiving endovascular therapy have an increased risk of in-hospital death compared to those receiving the

surgical clipping. However, both IV estimators produce OR estimates in the opposite direction of the naïve estimator. After accounting for potential confounders both the TSLS and EE approach yield the conclusion that there does not exist a statistically significant difference between the two treatment procedures.

Table 6: Comparing surgery and endovascular therapy

Estimator	OR	95% CI
Unadjusted		
Naïve	0.78	(0.60,1.02)
TSLS	1.87	(1.04,3.38)
EE	2.04	(0.95,4.38)
Adjusted		
Naïve	0.74	(0.56,0.97)
TSLS	1.46	(0.80,2.66)
EE	1.50	(0.48,4.71)

6 Discussion

Health outcomes research using non-experimental data from medical providers is being used to determine the effectiveness of various treatments [18, 12]. However, the lack of randomization in assigning treatment prevents researchers from using standard statistical techniques to compare outcomes amongst different treatment groups. The use of instrumental variables will allow researchers to expand the use of already existing data.

For linear MSM both the TSLS and EE are consistent estimators of treatment effect, and there are no remarkable differences in the performance of the two estimators. Logistic models are commonly found in epidemiological studies, biological modeling and medical treatment evaluations. For binary outcome models, although both IV estimators are in general inconsistent, there are still informative results that a researcher can obtain applying these IV methods for testing purposes. With continuous outcomes the EE remains consistent as long as $\epsilon_a = \epsilon$ is a reasonable assumption. In regression problems where one expects dependence between the residual and the treatment variable, estimation can be salvaged by finding a variable unrelated to the residual but related to the treatment.

7 Appendix A: Consistency and Asymptotic Linearity of the Estimating Equation Estimator

Consider the scenario described in Section 4.2.2. Let A and R be dichotomous random variables. $Y_a = m(a|\beta) + \epsilon$, where $m(A|\beta) = (1 + \exp(-\beta_0 - \beta_1 A))^{-1}$ and $\epsilon \sim N(0, \sigma^2)$. Here we only treat a special case, but the proof also holds for bounded ϵ. The counterfactuals are defined more specifically by,

$$\begin{aligned} Y_0 &\sim [1 + \exp(\beta_0)]^{-1} + N(0, \sigma^2) \\ Y_1 &= [1 + \exp(-(\beta_0 + \beta_1)]^{-1} + Y_0 - E(Y_0). \end{aligned}$$

Let $X_i = (Y_i, A_i, R_i)$ for $i = 1, \ldots n$ represent the observed data with $X \sim P_0$ and $\beta^0 = \beta(P_0)$. Let P_n represent the empirical probability distribution. Define $U(X|\beta) \equiv \phi(R)\epsilon(\beta)$ with

$$\phi(R) = \begin{pmatrix} 1 \\ E(A|R) \end{pmatrix}.$$

We showed in Section 3.2 that $U(X|\beta)$ is an unbiased estimating function provided $E(\epsilon|R, V) = 0$. In this section we give the conditions under which the estimator solving

$$\frac{1}{n}\sum_{i=1}^{n} \phi(R_i)\epsilon_i(\beta) = 0$$

exists and is a $\sqrt{n}-$consistent and asymptotically linear estimator of β^0.

Let $\mathcal{F} = \{U(X|\beta) : \beta\}$. Consider the following set of assumptions.

(A1) $E_{P_0}U(X|\beta) = 0$ implies $\beta = \beta_0$, i.e. β is identifiable.

(A2) $\beta \to E_{P_0}U(X|\beta)$ is a continuous and differentiable mapping.

(A3) $\frac{d}{d\beta}E_{P_0}U(X|\beta)$ is invertible.

(A4) $\mathcal{F}$ is a P-Donsker class of functions.

(A5) There exists a $\hat{\beta}_n$ such that $E_{P_n}U(X|\hat{\beta}_n) = 0$.

It is reasonable to assume that the first three assumptions hold. The remaining two assumptions require further discussion. Let $||\cdot||_v^*$ be as defined in van der Laan (1995). If $f : \Re^3 \to \Re$ has the property that $||f||_v^* < \infty$, then f is bounded over all section specific variations.

Theorem 1 *The class of functions* $\mathcal{F} = \{f : \Re^3 \to \Re : ||f||_v^* \leq M\}$ *is a Donsker class.*

The proof of this theorem is given in van der Laan (1995) [23]. Clearly $m(A|\beta)$ has compact support on $[0, 1]$. However, it is possible to have Y such that $|Y - m(A|\beta)| > 1$. Let $M_1 < \infty$ be such that when $|Y| \leq M_1$, $|Y - m(A|\beta)| \leq 1$. Define

$$\mathcal{F}_1 = \{f = (f_1, f_2), f_1 = [Y - m(A|\beta)]I(|Y| \leq M_1) \text{ and } f_2 = E(A|R)[Y - m(A|\beta)]I(|Y| \leq M_1) : \beta\}$$

Define

$$\mathcal{F}_2 = \left\{f = (f_1, f_2), f_1 = \left(\frac{Y - m(A|\beta)}{Y}\right) \text{ and } f_2 = E(A|R)\left(\frac{Y - m(A|\beta)}{Y}\right) : \beta\right\}$$

Each component of f in $\mathcal{F}_1$ and $\mathcal{F}_2$ belongs to a class of functions that have uniform bounded sectional variation. For $j = 1, 2$ each of the following mappings are of bounded variation where $f = (f_1, f_2)$ in $\mathcal{F}_1$ or $\mathcal{F}_2$.

$$\begin{aligned}
Y &\to f^{(1)}(Y, A, R), f^{(1)}(Y, A, R) = \frac{d}{dY}f_j(X|\beta) \\
A &\to f^{(2)}(Y, A, R), f^{(2)}(Y, A, R) = \frac{d}{dA}f_j(X|\beta) \\
R &\to f^{(3)}(Y, A, R), f^{(3)}(Y, A, R) = \frac{d}{dR}f_j(X|\beta) \\
(Y, A) &\to f^{(4)}(Y, A, R), f^{(4)}(Y, A, R) = \frac{d^2}{dYdA}f_j(X|\beta) \\
(Y, R) &\to f^{(5)}(Y, A, R), f^{(5)}(Y, A, R) = \frac{d^2}{dYdR}f_j(X|\beta) \\
(A, R) &\to f^{(6)}(Y, A, R), f^{(6)}(Y, A, R) = \frac{d^2}{dAdR}f_j(X|\beta) \\
(Y, A, R) &\to f^{(7)}(Y, A, R), f^{(7)}(Y, A, R) = \frac{d^3}{dYdAdR}f_j(X|\beta)
\end{aligned}$$

Therefore, $\mathcal{F}_1$ and $\mathcal{F}_2$ are a Donsker class of functions.

The next needed theorem follows from Gill et al. (1995) [10].

Theorem 2 *If* $E(g^{2+\delta}) < \infty$ *and* $\mathcal{F}$ *is a Donsker class of functions, then* $g\mathcal{F}$ *is Donsker.*

By Theorem 2 $Y \cdot I(|Y| > M_1) \cdot \mathcal{F}_2$ is a Donsker class of function. Finally,

$$\begin{aligned}
\mathcal{F} &= \{U(X|\beta) : \beta\} \\
&= \mathcal{F}_1 + Y \cdot I(|Y| > M_1)\mathcal{F}_2.
\end{aligned}$$

Thus (A4) holds.

To prove the final assumption holds we need the following lemma.

Lemma 1 *Define* $H(P,\beta) \equiv E_p[U(X|\beta)]$. *Suppose* $\frac{d}{d\beta}H(P,\beta)$ *is invertible at the true value of* $\beta = \beta^0$ *and that* $H(P_n,\beta)$ *is continuous in* β *in a neighborhood* N_{β^0} *of* β^0. *Assume the regularity condition that* $\mathcal{F} = \{U(X|\beta) : \beta\}$ *is a Donsker class of functions. Then there exists a solution* $\beta = \hat{\beta}_n$ *in* N_{β^0} *of* $H(P_n,\hat{\beta}_n)$ *with probability tending to one.*

Proof: The following proof is given in [3]. First note that $H(P,\beta_0) = 0$. The invertibility condition implies the existence of β^1 and β^2 in N_{β_0} such that for each component $H(P,\beta^1) < 0$ and $H(P,\beta^2) > 0$. Under the regularity condition, $H(P_n,\beta) \xrightarrow{P} H(P,\beta)$ for all β. Therefore, $H(P_n,\beta^1) < 0$ and $H(P_n,\beta^2) > 0$ with probability tending to one. With the continuity condition we have the existence of a solution in N_{β_0} of $H(P_n,\hat{\beta}_n) = 0$ with probability tending to one.

Theorem 3 *If (A1)-(A5) hold then* $\hat{\beta}_n$ *solving* $E_{P_n}U(X|\beta) = 0$ *is a* $\sqrt{n}-$*consistent and asymptotically linear estimator of* β^0.

Proof of consistency: As a consequence of $\mathcal{F}$ being Donsker, and hence Glivenko-Cantelli,

$$E_{P_n}U(X|\beta) \xrightarrow{P} E_{P_0}U(X|\beta).$$

Since $E_{P_n}U(X|\hat{\beta}_n) = 0$,

$$E_{P_0}U(X|\hat{\beta}_n) \xrightarrow{P} 0 \text{ as } n \to \infty.$$

By the Bolzano-Weierstrass property there exists a subsequence of $\hat{\beta}_n$ such that $\beta_{n(k)} \to \beta_\infty$. Along with (A2) this gives us $E_{P_0}U(X|\beta_{n(k)}) \to E_{P_0}U(X|\beta_\infty) = 0$. Using (A1) we get $\beta_\infty = \beta^0$ and $\beta_{n(k)} \xrightarrow{P} \beta^0$. Therefore $\hat{\beta}_n$ is a consistent estimator of β^0.

Proof of asymptotic normality: Suppose $\hat{\beta}_n$ is a consistent estimator of β^0. We have by definition that

$$E_{P_n}U(X|\hat{\beta}_n) - E_{P_0}U(X|\beta^0) = 0. \tag{9}$$

By adding and subtracting $E_{P_0}U(X|\hat{\beta}_n)$ to (9) we get

$$E_{P_0}\left[U(X|\hat{\beta}_n) - U(X|\beta^0)\right] = -E_{P_n-P_0}U(X|\hat{\beta}_n). \tag{10}$$

Using Taylor's expansion on the l.h.s. of (10)

$$E_{P_0}\left[U(X|\hat{\beta}_n) - U(X|\beta^0)\right] = \frac{d}{d\beta}E_{P_0}U(X|\beta)|_{\beta=\beta^0}(\hat{\beta}_n - \beta^0) + o(||\hat{\beta}_n - \beta^0||). \tag{11}$$

Combining (10) and (11),

$$-E_{P_n-P_0}U(X|\hat{\beta}_n) = \frac{d}{d\beta}E_{P_0}U(X|\beta)|_{\beta=\beta^0}(\hat{\beta}_n - \beta^0) + o(||\hat{\beta}_n - \beta^0||).$$

Then,

$$\begin{aligned}\hat{\beta}_n - \beta^0 &= -\left[\frac{d}{d\beta}E_{P_0}U(X|\beta)|_{\beta=\beta^0}\right]^{-1} E_{P_n-P_0}U(X|\hat{\beta}_n) + o(||\hat{\beta}_n - \beta^0||) \\ &= -E_{P_n-P_0}\left[\frac{d}{d\beta}E_{P_0}U(X|\beta)|_{\beta=\beta^0}\right]^{-1} U(X|\hat{\beta}_n) + o(||\hat{\beta}_n - \beta^0||)\end{aligned}$$

The Donsker class assumption gives us that

$$-E_{P_n-P_0}\left[\frac{d}{d\beta}E_{P_0}U(X|\beta)|_{\beta=\beta^0}\right]^{-1}U(X|\hat{\beta}_n)=O_p(1/\sqrt{n}).$$

Thus $\hat{\beta}_n-\beta^0=O_p(1/\sqrt{n})+o(||\hat{\beta}_n-\beta^0||)$, which implies $||\hat{\beta}_n-\beta^0||=O_p(1/\sqrt{n})$. So,

$$\hat{\beta}_n-\beta^0=-E_{P_n-P_0}\left[\frac{d}{d\beta}E_{P_0}U(X|\beta)|_{\beta=\beta^0}\right]^{-1}U(X|\hat{\beta}_n)+o_p(1/\sqrt{n}).$$

Since $\mathcal{F}$ is Donsker, and $U(X|\hat{\beta}_n)\in\mathcal{F}$ with probability tending to one, it follows that $\hat{\beta}_n$ is asymptotically linear.

References

[1] T. Amemiya. The nonlinear two-stage least-squares estimator. *Journal of Econometrics*, 2:105–110, 1974.

[2] T. Amemiya. The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. *Econometrica*, 45(4):955–968, 1977.

[3] C. Andrews, M. van der Laan, and J. Robins. Locally efficient estimation of regression parameters using current status data. Technical report, University of California, Berkeley, 1999.

[4] J. D. Angrist and G. W. Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442, 1995.

[5] A. Balke and J. Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997.

[6] M. Davidian and D. M. Giltinan. *Nonlinear models for repeated measurement data*. Chapman and Hall, 1995.

[7] E. M. Foster. Instrumental variables for logistic regression: An illustration. *Social Science Research*, 26:487–504, 1997.

[8] C. E. Frangakis and D. B. Rubin. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86(2):365–379, 1999.

[9] C. Gennings, V. Chinchilli, and W. Carter. Response surface analysis with correlated data: A nonlinear model approach. *Journal of the American Statistical Association*, 84:805–809, 1989.

[10] R. D. Gill, M. J. van der Laan, and J. A. Wellner. Inefficient estimators of the bivariate survival function for three models. *Annales de L'I.H.P. Probabilites et Statistiques*, 3(3):545–597, 1995.

[11] S. Greenland. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, 29:722–729, 2000.

[12] K. M. Harris and D. K. Remler. Who is the marginal patient? Understanding instrumental variables estimates of treatment effects. *Health Services Research*, 33(5):1337–1360, 1998.

[13] J. J. Heckman. Instrumental variables: A cautionary tale. Technical Report 185, National Bureau of Economic Research, 1995.

[14] P. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81:945–970, 1986.

[15] G. Imbens, J. Angrist, and D. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91:444–455, 1996.

[16] S. C. Johnston. Combining ecological and individual variables to reduce confounding by indication: Case study-subarachnoid hemorrhage treatment. *Journal of Clinical Epidemiology*, 53:1236–1241, 2000.

[17] S. C. Johnston. Effect of endovascular services and hospital volume on cerebral aneurysm treatment outcomes. *Stroke*, 31:111–117, 2000.

[18] M. Mc Clellan. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *Journal of the American Medical Association*, 272(11):859–866, 1994.

[19] W. K. Newey. Efficient instrumental variables estimation of nonlinear models. *Econometrica*, 58(4):809–837, 1990.

[20] J. Robins. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics*, 23(8):2379–2412, 1994.

[21] J. Robins. *Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference. Statistical Models in Epidemiology: The Environment and Clinical Trials*, pages 95–134. M.E. Halloran and D. Berry (eds.) Springer-Verlag, 1999.

[22] D. B. Rubin. Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

[23] M. van der Laan. *Efficient and inefficient estimation in semiparametric models*. PhD thesis, C.W.I. tract, ISBN 90-393-0339-8, 1995.

[24] N. Zohoori. Does endogeneity matter? A comparison of empirical analyses with and without control for endogeneity. *Annals of Epidemiology*, 7(4):258–266, 1997.