S E D A P

A PROGRAM FOR RESEARCH ON

SOCIAL AND ECONOMIC DIMENSIONS OF AN AGING POPULATION

A review of instrumental variables estimation in the applied health sciences

Paul Grootendorst

SEDAP Research Paper No. 215

For further information about SEDAP and other papers in this series, see our web site:
http://socserv.mcmaster.ca/sedap

Requests for further information may be addressed to:
Secretary, SEDAP Research Program
Kenneth Taylor Hall, Room 426
McMaster University
Hamilton, Ontario, Canada
L8S 4M4
FAX: 905 521 8232
e-mail: sedap@mcmaster.ca

A review of instrumental variables estimation in the applied health sciences

Paul Grootendorst

SEDAP Research Paper No. 215

June 2007

The Program for Research on Social and Economic Dimensions of an Aging Population (SEDAP) is an interdisciplinary research program centred at McMaster University with co-investigators at seventeen other universities in Canada and abroad. The SEDAP Research Paper series provides a vehicle for distributing the results of studies undertaken by those associated with the program. Authors take full responsibility for all expressions of opinion. SEDAP has been supported by the Social Sciences and Humanities Research Council since 1999, under the terms of its Major Collaborative Research Initiatives Program. Additional financial or other support is provided by the Canadian Institute for Health Information, the Canadian Institute of Actuaries, Citizenship and Immigration Canada, Indian and Northern Affairs Canada, ICES: Institute for Clinical Evaluative Sciences, IZA: Forschungsinstitut zur Zukunft der Arbeit GmbH (Institute for the Study of Labour), SFI: The Danish National Institute of Social Research, Social Development Canada, Statistics Canada, and participating universities in Canada (McMaster, Calgary, Carleton, Memorial, Montréal, New Brunswick, Queen's, Regina, Toronto, UBC, Victoria, Waterloo, Western, and York) and abroad (Copenhagen, New South Wales, University College London).

A review of instrumental variables estimation in the applied health sciences *

Paul Grootendorst
Faculty of Pharmacy, University of Toronto
Department of Economics, McMaster University

Contact Information:
paul.grootendorst@utoronto.ca
Faculty of Pharmacy, 144 College St, Toronto ON M5S 3M2
416 946 3994

Abstract:
Health scientists often use observational data to estimate treatment effects when controlled experiments are not feasible. A limitation of observational research is non-random selection of subjects into different treatments, potentially leading to selection bias. The 2 commonly used solutions to this problem – covariate adjustment and fully parametric models – are limited by strong and untestable assumptions. Instrumental variables estimation can be a viable alternative. In this paper, I review examples of the application of IV in the health and social sciences, I show how the IV estimator works, I discuss the factors that affect its performance, I review how the interpretation of the IV estimator changes when treatment effects vary by individual, and consider the application of IV to nonlinear models.

Keywords: instrumental variables, treatment effects, health outcomes

JEL Classifications: C31, I12

Résumé :

Quand les expériences contrôlées ne sont pas possibles, les chercheurs du domaine de la santé ont souvent recours à des données d'observation afin d'évaluer les effets de traitement. Une des restrictions de la recherche dépendant de données d'observation est l'affectation non-aléatoire des sujets à divers groupes de programme, ce qui a le potentiel de causer un biais de sélection. Les deux solutions habituelles à ce problème — l'ajustement covarié et les modèles complètement paramétriques — reposent sur des hypothèses fortes et invérifiables. Un estimateur à variables instrumentales (VI) peut représenter une alternative viable. Dans cette étude, je passe en revue des exemples de l'application des VI dans les sciences de la santé et les sciences sociales, je présente le mécanisme de l'estimateur par VI, je discute les facteurs qui affectent sa performance, je résume comment l'interprétation de l'estimateur VI diffère quand l'effet de traitement varie selon les sujets, et je considère l'application de la méthode des VI aux modèles non linéaires.

* Grootendorst acknowledges support from the Premier's Research Excellence Award and the research program into the Social and Economic Dimensions of an Aging Population centred at McMaster University that is primarily funded by the Social Sciences and Humanities Research Council of Canada (SSHRC) and which has received additional support from Statistics Canada.

Background

Much of the empirical work in the applied health sciences attempts to address questions of the sort: what is the effect of x on y? The variable y is typically a dimension of health, such as disease incidence, health related quality of life or mortality, while x could be some health-related behavior, such as cigarette smoking; an individual characteristic, such as income, education or age; the use of health care, such as a new pharmaceutical drug; or exposure to an environmental toxin, such as second-hand smoke. Much of this work relies on observational data, owing to the practical and ethical limitations on the use of controlled experiments in this area. A widely recognized problem in observational research is that, because individuals can sometimes 'choose' different values of x (for instance, the decision to smoke or not), it is unclear to what extent differences in y reflect differences in the level of x and to what extent differences in y reflect differences in the unobserved characteristics of those who choose different levels of x. The recent controversy over the deleterious effects of hormone replacement therapy (HRT) among post-menopausal women illustrates this attribution problem. The observational data clearly indicate that HRT and cardiovascular disease (CVD) are negatively correlated – HRT users have better heart health than non-users. These correlations lead some women, typically those with more education and more income, to initiate HRT as prevention against heart disease. Subsequent experimental evidence, however, indicated that this correlation is entirely due to decisions of healthier women to initiate HRT. Indeed, the causal impact of HRT is to *increase* the risk of heart attack and stroke. In other words, HRT users were healthier than non-users *despite* the deleterious effects of HRT.

Several solutions to this problem have been proposed, but none are entirely satisfactory. The most commonly used of these is to identify, measure and adjust for the behavioural and other factors w that are correlated with both x and y – the so-called 'covariate adjustment' approach. If ones goal is to assess the causal effect of HRT on CVD, for instance, one might adjust for various dimensions of socio-economic status and other potential 'confounding' factors correlated with HRT use that independently affect CVD; one could adjust for these confounding factors using either regression or matching. Regression amounts to imposing a restriction on how w and x affect the mean of y; the most commonly imposed restriction is that the mean of y is linear in a vector of unknown parameters α, β, γ: $E[y|x,w] = \alpha + x\beta + w'\gamma$. Matching compares the values of y among subjects with different levels of x but who share common values of all of the variables in w. A defect of both these techniques is that the analyst might fail to adjust for pertinent confounding variables, because they are either unknown or not readily quantifiable. Regression comes with an additional liability – it requires that one correctly specify a model of the conditional mean of y and it is unclear how specification error affects ones estimate of the impact of x on y.

An alternative to covariate adjustment is to model the correlations between unobserved confounders and outcomes; these are commonly referred to as 'fully parametric' models. The leading example is Heckman's parametric sample selection model. The assumptions embedded in these models are quite restrictive: one needs to specify functional forms for both the conditional mean of y and the joint distribution of the unobserved factors affecting x and y. It is well known that results can be highly sensitive to these assumptions, yet these assumptions cannot be directly verified.

'Instrumental variables' (IV) estimation might be a useful alternative to the covariate adjustment and fully parametric approaches. IV requires one or more instruments z – variables strongly correlated with x but uncorrelated with the unobserved determinants of y. The coin toss in the context of a randomized controlled trial (RCT) on fully compliant subjects is the ideal instrument – the outcome of the coin toss completely determines treatment assignment (x = treatment, control) yet does not directly

affect the outcome y. In an observational study assessing the effects of HRT on CVD, one might use the consumer price of HRT as an instrument, assuming that the consumer price of HRT affects its use but is uncorrelated with unobserved dimensions of CVD. (The latter assumption would rule out, for instance, healthier women having particularly generous drug insurance while less healthy women remaining uninsured.)

In this paper, I review examples of the application of IV in the health and social sciences, I show how the IV estimator works, I discuss the factors that affect its performance, I review how the interpretation of the IV estimator changes when treatment effects vary by individual, and consider the application of IV to nonlinear models.

Applications of IV estimation

Central to the use of IV estimation is the identification of good instruments. Good instruments induce marked variation in exposure to treatments that is incidental to the health outcome under investigation. This variation might be caused by regional differences in health care reimbursement, or regional or time based variation in physician practice styles, or variation in health care spending due to the timing of the business or electoral cycle. Here are several interesting applications of IV in the health and social sciences:

Cholera transmission

Although IV theory has been developed primarily by econometricians, the method originated in epidemiology. IV was used to investigate the route of cholera transmission during the London cholera epidemic of 1853-54. A scientist from that era, John Snow, hypothesized that cholera was waterborne. To test this, he could have tested whether those who drank purer water had lower risk of contracting cholera. In other words, he could have assessed the correlation between water purity (x) and cholera incidence (y). Yet, as Deaton (1997) notes this would not have been convincing: "The people who drank impure water were also more likely to be poor, and to live in an environment contaminated in many ways, not least by the 'poison miasmas' that were then thought to be the cause of cholera." Snow instead identified an instrument that was strongly correlated with water purity yet uncorrelated with other determinants of cholera incidence. This instrument was the identity of the company supplying households with drinking water. At the time, Londoners received drinking water directly from the Thames River. One company, the Lambeth water company, drew water at a point in the Thames above the main sewage discharge; another, the Southwark and Vauxhall company, took water below the discharge. Hence the instrument z was strongly correlated with water purity x. The instrument was also uncorrelated with the other determinants of cholera incidence (y). According to Snow (1855, pages 74-75), the households served by the two companies were quite similar, indeed:

"the mixing of the supply is of the most intimate kind. The pipes of each Company go down all the streets, and into nearly all the courts and alleys. . . . The experiment, too, is on the grandest scale. No fewer than three hundred thousand people of both sexes, of every age and occupation, and of every rank and station, from gentlefolks down to the very poor, were divided into two groups without their choice, and in most cases, without their knowledge; one group supplied with water containing the sewage of London, and amongst it, whatever might have come from the cholera patients, the other group having water quite free from such impurity."

Effectiveness of cardiac catheterization

Newhouse and McClellan (1998) assessed the impact of cardiac catheterization (i.e. bypass surgery or angioplasty) on post myocardial infarction (MI) survival rates. Comparing the health outcomes of those who do and do not get cardiac catheterization is problematic if individuals are selected for the procedure on the basis of disease severity. Instead, Newhouse and colleagues compared the health outcomes of heart attack patients who lived near a hospital with catheterization facilities to outcomes of patients living farther away. After an MI, the patient is typically rushed by ambulance to the nearest hospital, irrespective of whether or not the hospital has catheterization facilities. Hence the catheterization capability status of the hospital closest to the patient's home is a good instrument – it is a strong predictor of whether or not the patient gets the procedure and should be uncorrelated with health outcome (unless those who are more likely to suffer from severe heart attacks move to a home in close proximity to a hospital with catheterization facilities, something that the authors thought was unlikely.)

Effectiveness of health care service use

Fisher *et al* (2003a, b) assessed the impact of the overall level of health care resource use in a region on health outcomes. To do so they could have assessed the correlation between regional health care spending and the average health outcomes of patients in the region. The problem with this approach is that health care spending could be higher in regions in which patient health outcomes are lower (i.e. resources are allocated to regions where needs are greatest). To overcome this problem, they used as an instrument the average end-of-life medical spending in the region. This is a good instrument: it was found to be highly correlated with overall health care use but had no direct effect on patient health outcomes.

Education and earnings

Economists have exploited a variety of sources of quasi-experimental variation in treatment variables. Angrist and Krueger (1991), for example, assessed the role of years of secondary school education on subsequent earnings. Because both years of schooling and earnings are probably correlated with ability or motivation, both of which are latent, the authors compared the earnings of students born in different quarters of the year. Those born earlier in the year have slightly less schooling than those born later in the year due to school start age policies and compulsory schooling laws. Children born in the same calendar year generally are required start school at the same time, i.e., in September of the year they turn 6. Hence school entry age depends on quarter of birth: children born in the fourth quarter enter school at age 5¾ while children born in the first quarter enter school at age 6 ¾. As well, compulsory schooling laws typically require students to stay in school until their 16th birthday, so that the length of schooling varies by up to 12 months with date of birth. Date of birth should therefore be a good instrument, provided that it is uncorrelated with ability. In related research, Oreopoulos (2006) finds that extension of compulsory schooling laws in Canada had marked effects on educational attainment.

The IV estimator

The following example illustrates how the IV estimator adjusts for the influence of confounders. Suppose that we wish to compare the effect on some continuous measure of patient health H of two types of health care, a new type of care and an existing standard type. Let the variable D represent the

type of care used, with $D = 1$ if the patient uses new care and $D = 0$ if the patients uses standard care. Suppose that, unbeknownst to the investigator, the value of D is assigned on the basis of patient frailty F – more frail patients are more likely to receive the new care and also have worse health.
Suppose that treatment D affects health H as follows:

$$H = \beta_0 + \beta_1 D + \varepsilon \quad \text{(H1)}$$

where β_0 and β_1 are unknown parameters and the 'error' ε represents the combined influence of all determinants of H that are not explicitly modeled; one such determinant is F. Interest centers on generating consistent estimates of the treatment effect parameter β_1, the difference in effectiveness of new care and standard care.

Conventional estimators of β_1 will be biased downwards. Consider, for instance, the difference in means (DIM) estimator, which is the sample average H of new care users less the sample average H of standard care users. The expected value of the DIM estimator, b_1^{dm}, is:

$$E\left[b_1^{dm}\right] = E[H|D = 1] - E[H|D = 0] \quad (2)$$

where $E[H|D = 1]$ denotes the expectation of H for those assigned new care. The term $E[H|D = 1]$ can be evaluated by computing the expected value of the health outcomes model (H1) conditional on D.

$$E[H|D] = \beta_0 + \beta_1 D + E[\varepsilon|D] \quad (3)$$

The expected value of H of new care users is:

$$E[H|D = 1] = \beta_0 + \beta_1 + E[\varepsilon|D = 1] \quad (4)$$

and the expected value of H of standard care users is:

$$E[H|D = 0] = \beta_0 + E[\varepsilon|D = 0] \quad (5)$$

Subbing (4) and (5) into (2) yields:

$$E\left[b_1^{dm}\right] = \beta_1 + E[\varepsilon|D = 1] - E[\varepsilon|D = 0] \quad (6)$$

It is clear that the DIM estimator will be unbiased (i.e. $E\left[b_1^{dm}\right] = \beta_1$) if and only if the expected errors are the same in both treatment groups:

$$E[\varepsilon|D = 1] = E[\varepsilon|D = 0] \quad (7)$$

But if frailer patients tend to get new care and be in worse health, then $E[\varepsilon|D = 1] < E[\varepsilon|D = 0]$ and $E\left[b_1^{dm}\right] < \beta_1$. Hence condition (7) can be violated when there are confounding variables. Condition (7) can also be violated if there is 'reverse causality', i.e., if health outcomes H directly affect treatment choice D, or if there is measurement error in D.

An IV estimator for β_1 might work if D, the treatment provided to the patient, depends in part on a variable that is independent of ε. Suppose, for instance, that there are two types of physicians, labeled

C (for conservative) and L (for liberal). Suppose that C physicians tend to use standard care whereas L physicians tend to use new care. The physician's practice style, described by $DocType = \{C, L\}$, is a valid instrument if 2 conditions hold. First, there needs to be pronounced differences in physician practice styles. This means that:

$$E(D|DocType) = Prob(D = 1|DocType) \times 1 + Prob(D = 0|DocType) \times 0$$
$$= Prob(D = 1|DocType)$$

varies with different values of $DocType$. I have assumed that it does; in particular, I have assumed that:

$$Prob(D = 1|DocType = L) > Prob(D = 1|DocType = C) \quad (8)$$

Second, for $DocType$ to be a valid instrument, it needs to be uncorrelated with the error ε, the unexplained determinants of the health outcome:

$$E(\varepsilon|DocType = L) = E(\varepsilon|DocType = C) \quad (9)$$

This condition implies that there are no differences in the quality of care provided by L and C-type physicians that would result in differences in patient health outcomes, nor do sicker patients gravitate selectively towards L or C physicians. A stronger condition required to estimate consistently both β_0 and β_1 is that:

$$E(\varepsilon|DocType) = 0 \quad (10)$$

Conditions (8) and (9) mean that $DocType$ is a valid instrument if it affects health outcomes only through its impact on the likelihood that new care is provided.

To understand how $DocType$ can be used to consistently estimate the treatment effect, take the expectation of H conditional on $DocType$:

$$E(H|DocType) = \beta_0 + \beta_1 E(D|DocType) + E(\varepsilon|DocType) \quad (11)$$

Sub (8) and (10) into (11):

$$E(H|DocType) = \beta_0 + \beta_1 Prob(D = 1|DocType) + 0 \quad (12)$$

Evaluating (12) under the two values of $DocType$ yields:

$$E(H|DocType = L) = \beta_0 + \beta_1 Prob(D = 1|DocType = L)$$
$$E(H|DocType = C) = \beta_0 + \beta_1 Prob(D = 1|DocType = C)$$

These two equations can be solved for β_1:

$$\beta_1 = \frac{\mathrm{E}(H|DocType = L) - \mathrm{E}(H|DocType = C)}{\mathrm{Prob}\ (D=1 \mid DocType = L) - \mathrm{Prob}\ (D=1 \mid DocType = C)} \quad (13)$$

The IV estimator is operationalized by replacing the unknown quantities by sample estimates. Hence, for example, $E(H|DocType = L)$ is replaced by the sample average health of those patients treated by

L-type physicians. $Prob(D = 1|DocType = L)$ is replaced by the sample proportion of patients treated by L-type physicians who are given new care.

As was mentioned, the coin toss used to assign subjects into the two treatment groups in the context of a RCT is a special case of IV estimation. Suppose treatments are assigned according to process (T1):

$$D = \begin{cases} 1 \; if \; CoinToss = Heads \\ 0 \; if \; CoinToss = Tails \end{cases} \quad \text{(T1)}$$

In an RCT, the outcome of the coin toss – not $DocType$ or F – assigns individuals to treatments. According to (T1), if $CoinToss = Heads$ then the subject gets the new care ($D = 1$), and if $CoinToss = Tails$ then the subject gets the old care ($D = 0$). Then the estimator of the treatment effect is:

$$\beta_1 = \frac{\mathrm{E}(H|CoinToss = Heads) - \mathrm{E}(H|CoinToss = Tails)}{\mathrm{Prob}\,(D=1 \mid CoinToss = Heads) - \mathrm{Prob}\,(D=1 \mid CoinToss = Tails)} \quad (14)$$

If *all* subjects who get $CoinToss = Heads$ use new care, then $Prob(D = 1 \mid CoinToss = Heads) = 1$. Similarly, if all subjects who get $CoinToss = Tails$ use standard care, then $Prob(D = 1 \mid CoinToss = Tails) = 0$. In this case, the IV estimator simplifies to:

$$\beta_1 = \mathrm{E}(H|CoinToss = Heads) - \mathrm{E}(H|CoinToss = Tails) \quad (15)$$

Replacing the expected values with the sample means gives the standard DIM estimator of the treatment effect. Of course, it is entirely possible that some subjects assigned to use new care will use standard care and likewise, some subjects assigned to standard care will use new. Such non-compliance can be handled by replacing $Prob(D = 1 \mid CoinToss = Heads)$ and $Prob(D = 1 \mid CoinToss = Tails)$ with the proportions in each group who use the new care.

The generalized IV estimator

The IV estimator can be generalized to allow for multiple instruments and explicit modeling of the impact of additional determinants of H. Exploiting multiple sources of independent variation in D can improve the precision of the IV estimator. Moreover, modeling the impact of additional determinants of H will reduce the error (i.e. the amount of unmodelled variation in H) and hence make it easier to satisfy the requirement that the instrument(s) be independent of the error. But by doing so, one incurs the risk of specifying an inaccurate model of the determinants of H. (This is the same risk that one incurs when using regression models.)

Suppose that the error term ε in (H1) can be decomposed as

$$\varepsilon = \boldsymbol{W}'\boldsymbol{\gamma} + \upsilon$$

where $\boldsymbol{W}$ is a set of observed determinants of H, $\boldsymbol{\gamma}$ is a conformable vector of unknown parameters and υ represents the influence of the remaining latent, unmodelled determinants of H. $\boldsymbol{W}$ is assumed to be uncorrelated with υ. Then H is determined by the linear equation:

$$H = \beta_0 + \beta_1 D + \boldsymbol{W}'\boldsymbol{\gamma} + \upsilon \quad \text{(H2)}$$

If $E[\upsilon|D=1] = E[\upsilon|D=0]$, then given a set of n observations on H, D and $\boldsymbol{W}$, the parameters of (H2) can be estimated using ordinary least squares (OLS). If this condition is not satisfied, then OLS is inconsistent. But if one had access to a set of instruments $\boldsymbol{Z}$ which satisfy the conditions

$$E[\upsilon|\boldsymbol{W}, \boldsymbol{Z}] = 0$$

$$\operatorname*{plim}_{n\to\infty} n^{-1}\boldsymbol{Z}'D \neq 0$$

where n is the sample size and plim denotes the probability limit operator, then one can use the generalized IV estimator. This estimator of the parameters β_0, β_1 and $\boldsymbol{\gamma}$ solves the sample moment condition:

$$\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{Z}^*}(\boldsymbol{H} - \beta_0 + \beta_1\boldsymbol{D} + \boldsymbol{W}\boldsymbol{\gamma}) = \boldsymbol{0} \quad \text{(16)}$$

where $\boldsymbol{X} = [\boldsymbol{1}\ \boldsymbol{D}\ \boldsymbol{W}]$ is a matrix consisting of n observations on a constant 1, the treatment indicator D and $\boldsymbol{W}$; (the ith observation of $\boldsymbol{X}$ is denoted $\boldsymbol{X}_i$); $\boldsymbol{H} - \beta_0 + \beta_1\boldsymbol{D} + \boldsymbol{W}\boldsymbol{\gamma} = \boldsymbol{\upsilon}$ is a vector consisting of n error terms; $\boldsymbol{H}$ is the vector of n observations on the health outcome, and $\boldsymbol{P}_{\boldsymbol{Z}^*}$ is the so-called projection matrix:

$$\boldsymbol{P}_{\boldsymbol{Z}^*} = \boldsymbol{Z}^*(\boldsymbol{Z}^{*'}\boldsymbol{Z}^*)^{-\boldsymbol{1}}\boldsymbol{Z}^{*'}$$

where $\boldsymbol{Z}^* = [\boldsymbol{1}\ \boldsymbol{Z}\ \boldsymbol{W}]$ is a matrix consisting of n observations on the constant, the instruments $\boldsymbol{Z}$ and the exogenous or predetermined variables $\boldsymbol{W}$. $\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{Z}^*}$ consists of the predicted values from regressions of each of the columns in $\boldsymbol{X}$ on $\boldsymbol{Z}^*$. (Note that the predicted values from the regressions of $\boldsymbol{1}$ and $\boldsymbol{W}$ on $\boldsymbol{Z}^*$ are the observed values $\boldsymbol{1}$ and $\boldsymbol{W}$.) Hence (16) generalizes condition (10), encountered earlier, that the instruments be orthogonal to the error. Solving the sample moment condition for the unknown parameters yields the generalized IV estimator:

$$\begin{bmatrix} b_0^{iv} \\ b_1^{iv} \\ \boldsymbol{g^{iv}} \end{bmatrix} = (\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{Z}^*}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{Z}^*}\boldsymbol{H} \quad \text{(17)}$$

This estimator can be implemented in two steps. First, one estimates $\widehat{D}$, the predicted values from the regression of D on $\boldsymbol{Z}^*$. $\widehat{D}$ essentially combines the different instruments into a single summary instrument; this summary instrument is orthogonal to the error. Then one estimates the regression of H on $\widehat{D}$ and $\boldsymbol{W}$. The IV treatment effect estimate is the coefficient on $\widehat{D}$.

Are the instruments valid?

It is possible to check whether the two requirements of one's instruments are satisfied.

Assumption 1: The instruments are correlated with the treatment

If they are, then the instruments should have good predictive power in the first stage regression above, i.e. the regression of $\boldsymbol{D}$ on $(\boldsymbol{W}, \boldsymbol{Z})$. This can be tested using an *F* test of the restriction that the instruments are jointly insignificant. Small values of this *F* statistic indicate a violation of Assumption 1.

Assumption 2: The instruments are uncorrelated with the error

This can be tested if there is more than one instrument. One estimates the error term using the residuals from the IV-estimated model:

$$\boldsymbol{v^{iv}} = \boldsymbol{H} - \{b_0^{iv} + b_1^{iv}\boldsymbol{D} + \boldsymbol{W}'\boldsymbol{g^{iv}}\}$$

where b_0^{iv}, b_1^{iv} and $\boldsymbol{g^{iv}}$ denote the IV estimates of the parameters in equation (H2). One then estimates a regression of $\boldsymbol{v^{iv}}$ on $(\boldsymbol{W}, \boldsymbol{Z})$. Large values of $n \times R^2$ from this regression indicates that the instruments in $\boldsymbol{Z}$ explain some of the variation in $\boldsymbol{v^{iv}}$, which is a violation of Assumption 2. If the assumption is satisfied, this test statistic is distributed χ^2 with the number of degrees of freedom equal to the number of instruments minus one. See Davidson and MacKinnon (2003) for details.

Is IV necessary?

If OLS can be used, it should be used because OLS has a much smaller variance than IV and is easier to use. OLS can be used under any of the following circumstances:

1. if $E[\upsilon|\boldsymbol{D} = 1] = E[\upsilon|\boldsymbol{D} = 0]$. If this condition is satisfied then IV and OLS should give similar estimates – they should differ only by chance. A chi-squared based test of the difference in parameter estimates can be used to verify this. An equivalent test is to estimate the residuals, $\hat{e}$, from the regression of D on $(\boldsymbol{W}, \boldsymbol{Z})$. Then one estimates the regression of H on $\boldsymbol{D}$, $\boldsymbol{W}$ and $\hat{e}$. If IV and OLS give similar estimates, the variable $\hat{e}$ will not be statistically different from zero. Johnson and DiNardo (1997, page 339) provide the intuition behind this test. The regression of D on $(\boldsymbol{W}, \boldsymbol{Z})$ splits the variation in D into two parts. One part, the predicted values from this regression (each of which is a linear combination of the variables in $(\boldsymbol{W}, \boldsymbol{Z})$), is uncorrelated with υ, assuming that the instruments are valid. The other part, the residuals, is uncorrelated with υ if the condition is satisfied. If the condition is not satisfied, then these residuals will explain successfully some of the variation in υ, (or, equivalently, some of the variation in H that remains after conditioning on $(\boldsymbol{W}, \boldsymbol{Z})$) and IV estimation may be warranted.
2. if one can sign the bias associated with OLS. Suppose for instance that OLS is known to underestimate a treatment effect. Then if the treatment effect estimate is positive despite this bias, one has good evidence that the treatment effect really is positive.
3. if the sample size is small or the instruments are only weakly correlated with D. The desirable properties of the IV estimator are realized only as the sample size grows infinitely large. (Technically, the IV estimator is biased but consistent.) As I illustrate below, in finite sized samples, the IV estimator can be wildly inaccurate.

Finite sample properties of the IV estimator

Inspecting equation (13), it is clear that the IV estimator of β_1 is in fact a ratio of two unknown quantities: the numerator is the correlation between the instrument and H, the denominator is the correlation between the instrument and D. When the correlation between the instrument and D is weak, then the IV estimator can be highly imprecise. Indeed, when the correlation approaches zero, meaning that the instrument explains none of the variation in D, the estimator approaches infinity. Even when the instrument and D are highly correlated, if the sample size is small then estimates of both correlations can be imprecise, rendering the ratio of these estimated correlations highly imprecise. The degree of imprecision of the IV estimator can be illustrated via Monte Carlo simulation. Suppose that health outcomes H are determined according to the process:

$$H = 100 + 25D - 0.50F - 0.75age + \upsilon, \upsilon \sim U[-10,10] \quad \text{(H3)}$$

where D indicates which form of care is used, F is an index of patient frailty, where a larger value of F means greater frailty, age is patient age in years, and υ is a random variable, representing idiosyncratic factors that affect health, or perhaps measurement error in H. υ can take on any integer value in the interval $[-10,10]$, each with equal probability. The shorthand way of writing this is $\upsilon \sim U[-10,10]$, where "$\sim$" means "distributed as", "U" means the uniform distribution, and $[-10,10]$ defines the range of values υ can assume.[1] Hence H is lower for older, more frail subjects and is 25 units higher for those using new care $(D = 1)$ compared to those using standard care $(D = 0)$.

What remains is to describe how patients are assigned to new and old care. I consider two alternatives:

Randomization with full compliance
Here D is determined by the outcome of a coin toss (T1).

Patients assigned to D on the basis of F and $DocType$.
In this case clinical factors, not a coin toss, determine D.

$$D = \begin{cases} 1 \; if \; Index > 0 \\ 0 \; if \; Index \leq 0 \end{cases}$$
$$Index = -40 + 0.5age + 0.5F - 20dC + \nu, \nu \sim U[-10,10] \quad \text{(T2)}$$

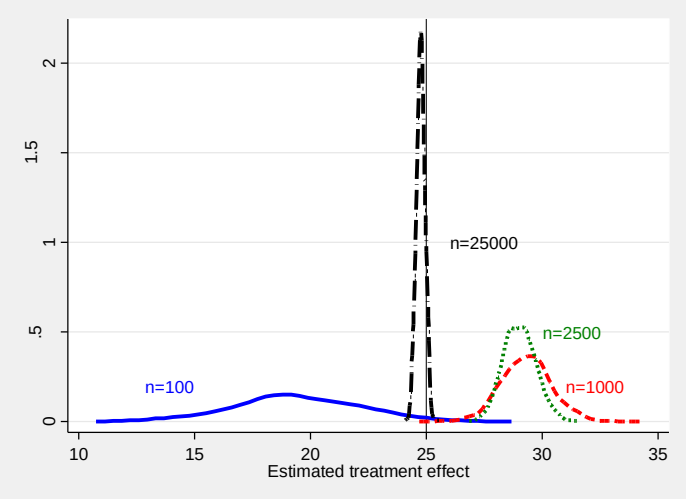
New care is given only if a patient's $Index$ value exceeds some threshold level, arbitrarily set equal to zero. $Index$ is greater, the older is the patient (i.e. the greater is age), the more frail is the patient (i.e. the greater is F), and the greater is the patient-specific idiosyncratic factor (i.e. the greater is the random variable ν, where $\nu \sim U[-10,10]$). High values of ν could reflect other factors that influence treatment assignment, such as a strong patient preference for the new type of care. $Index$ is 20 points lower if the patient is treated by a C-type (conservative) doctor, identified by $dC = 1$, instead of a L-type (liberal) doctor ($dC = 0$). Hence under (T2), sicker patients and those treated by liberal doctors are more likely to receive new care.

[1] Errors drawn from $U[-10,10]$ can take on any one of 21 different integer values: -10, -9, -8, …, -1, 0, 1, …, 8, 9, 10, each occurring with equal probability (=1/21). Subjects with the same values of D, F and age can therefore have health outcomes that differ by as much as 20 units.

I first demonstrate the behavior of the IV estimator as the sample size increase. To do so, I used health outcome process (H3) and treatment assignment process (T2) to generate a sample of size n observations. For each subject in my sample, I arbitrarily assigned values of dC (0 or 1), age (between 25 and 73 years), F (between 1 and 100), and randomly drawn values of υ and ν. I generated $R = 1000$ of such samples, taking independent draws of υ and ν for each sample. I used various values of $n = \{100, 1000, 2500, 25000\}$. For each of the R samples of size n observations, I used the generalized IV estimator that controls for age, and uses dC as an instrument to estimate the treatment effect parameter β_1 (which is equal to 25).

The kernel-smoothed histograms of the R treatment effect estimates are displayed in Figure 1. When $n = 100$ the sampling distribution of the estimator is rather wide (estimates varied from about 10 to 28) and is centered over 18. Increasing n to 1000 decreased the variability of estimates but did not materially improve the bias. The IV estimator is approximately unbiased when the sample size is in the order of $n = 25000$ observations.

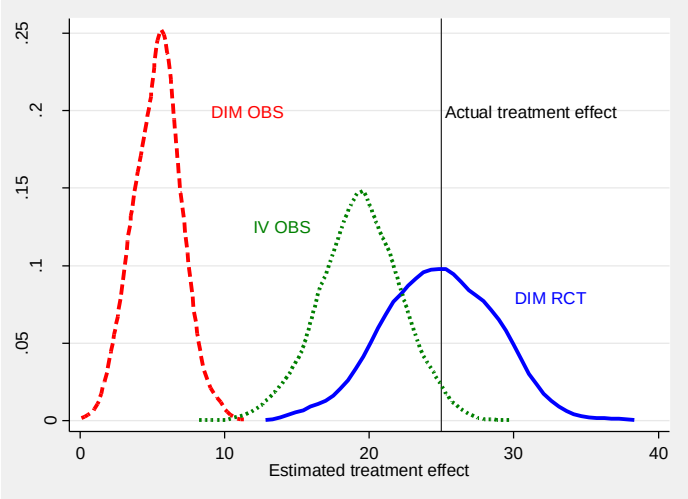
Figure 1 Histogram of 1000 treatment effect estimates generated from IV estimator using various sample sizes.



How do the other estimators compare? Using health outcome process (H3) and one of the treatment assignment processes (T1) and (T2), I generated $R = 1000$ samples of size $n = 100$ observations and, for each sample, used several different estimators to produce estimates of the treatment effect. I considered the difference in means estimator, b_1^{dm}, applied to both treatment assignment process (T2) (i.e., observational data) and treatment assignment (T1) (experimental data). I also reproduced the results for the generalized IV estimator that controls for age, and uses dC as an instrument, shown in Figure 1.

The kernel-smoothed histograms of the R treatment effect estimates are displayed in Figure 2. As expected, the sampling distribution of the DIM estimator, b_1^{dm}, estimated using the observational data (labeled as DIM OBS in the figure) is well to the left of 25, indicating that b_1^{dm} is severely downwards biased. Conversely, the DIM estimator estimated using the experimental data, DIM RCT in the figure, is centered over 25, indicating that b_1^{dm} is unbiased in this context. The IV estimator, IV OBS in the figure, is downwards biased, but not to the same degree as DIM OBS.

Figure 2 Histogram of 1000 treatment effect estimates generated from various estimators. Sample size of 100 in each case.



Note: *DIM OBS:* difference in means estimator using observational data. *IV OBS:* instrumental variables estimator using observational data (instrument is *DocType*). *DIM RCT:* difference in means estimator where treatments randomly assigned. Sample size = 100 in each case.

I next examined the properties of the IV estimator using observational data with different values of n and different degrees of correlation between the instrument and treatment assignment. First, I considered the IV estimator using a smaller sample size ($n = 50$). This case is denoted "IV n=50". Then I considered the behavior of the IV estimator where $n = 100$ as before, but where the correlation between dC and D was weakened. Specifically, I modified (T2) so that the coefficient on dC was half as large in absolute value as before:

$$Index = -40 + 0.5age + 0.5F - \mathbf{10}dC + v, v \sim U[-10,10] \quad \text{(T3)}$$

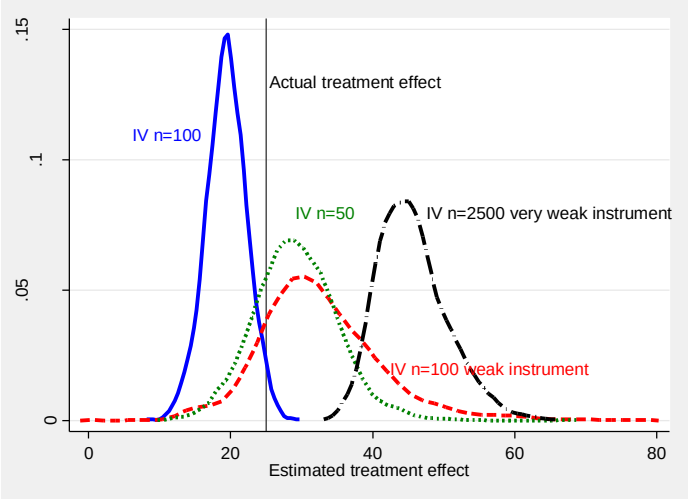
This case was denoted as "IV n=100 weak instrument". Next, I further weakened the influence of dC and D by modifying (T2) so that the coefficient on dC was even smaller:

$$Index = -40 + 0.5age + 0.5F - \mathbf{6}dC + v, v \sim U[-10,10] \quad \text{(T4)}$$

but compensated by increasing the sample size from $n = 100$ to $n = 2500$. This case was denoted as "IV n=2500 very weak instrument".

The resulting sampling distributions are displayed in Figure 3. The IV estimator using $n = 50$ is compared to the IV estimator using $n = 100$ (which appears as "IV OBS" in the previous figure). Note the smaller sample size decreases the precision of the estimator but, ironically, reduces the bias. Weakening the correlation between D and dC while using $n = 100$ resulted in a marked decrease in estimator precision – estimates ranged from less than 0 to over 80. Weakening the correlation even further while using $n = 2500$ produced an IV estimator with roughly the same precision as the "IV n=50" case, but with a marked increase in bias.

Figure 3 Histogram of 1000 treatment effect estimates generated from instrumental variables estimators.



Note: *IV n=100:* instrumental variables estimator with sample size (*n*) = 100. *IV n=50:* instrumental variables estimator with *n* = 50. *IV n=100:* instrumental variables estimator with *n* = 100 and correlation between *DocType* and *D* weakened. *IV n=2500:* instrumental variables estimator with *n* = 2500 and correlation between *DocType* and *D* weakened further. In each case instrument is *DocType.*

The take home message is that small sample sizes or weak correlation between instrument and treatment can adversely affect the performance of the IV estimator. The IV estimator can also behave poorly when an invalid instrument is used, i.e. when there is some correlation between the instrument and the unmodelled determinants of H. Indeed, Bound and colleagues (1995) demonstrate that the adverse effects of weak instruments on IV performance are exacerbated when there is even weak correlation between the instruments and the error. A final determinant of the finite sample performance of the IV estimator is the number of instruments chosen. More instruments are better, but only up to a point. As Kennedy (2003, page 175) notes, as more instruments are added, in small samples, $\widehat{D}$ becomes closer and closer to D "and so begins to introduce the bias that the IV procedure is trying to eliminate. This bias is proportional to the inverse of the *F* test statistic for testing the significance of the instrumental variables in explaining the explanatory variable for which it is to serve as an instrument."

For further discussions of weak instruments and the generalized IV estimator see Staiger and Stock (1997), Hahn and Hausman (2002), and Stock, Wright and Yogo (2002).

The reduced form model

Even if the IV estimator performs poorly, one can learn about the treatment effect by estimating via OLS what is known as the 'reduced form' model. This model relates the health outcome H to the instruments $\boldsymbol{Z}$ and any known exogenous health determinants $\boldsymbol{W}$. To illustrate, suppose that H is determined by the equation:

$$H = f(D, \boldsymbol{W}, \upsilon) \qquad (18)$$

And suppose that treatment assignment D is determined by the equation:

$$D = g(\boldsymbol{Z}, \nu) \tag{19}$$

The reduced form model is derived by substituting (19) into (18):

$$H = f(g(\boldsymbol{Z}, \nu), \boldsymbol{W}, \upsilon) \tag{20}$$

which leads to the estimable regression model:

$$H = \alpha_0 + \boldsymbol{W}'\boldsymbol{\alpha_1} + \boldsymbol{Z}'\boldsymbol{\alpha_2} + \omega$$

where the α's are unknown parameters and ω is the composite error term derived from the two random components: ν and υ. OLS is an unbiased estimator of the α's; the estimates of $\boldsymbol{\alpha_2}$ inform the 'net effect' of each of the instruments on H. The net effect of instrument Z_j on H is comprised of the product of the effect of Z_j on D (which is proportional to the strength of Z_j) and the effect of D on H (i.e. the treatment effect). This can be seen by differentiating (20) w.r.t. Z_j:

$$\frac{\partial H}{\partial Z_j} = \frac{\partial f(D, \boldsymbol{W}, \upsilon)}{\partial D} \frac{\partial g(\boldsymbol{Z}, \nu)}{\partial Z_j}$$

As Angrist and Krueger (2001) note, the sign, magnitude and statistical significance of the estimates of the $\boldsymbol{\alpha_2}$ are informative. If $\boldsymbol{\alpha_{2j}} = 0$, then either $\frac{\partial f(D,\boldsymbol{W},\upsilon)}{\partial D} = 0$ (i.e. the treatment effect is zero), or $\frac{\partial g(\boldsymbol{Z},\nu)}{\partial Z_j} = 0$ (i.e. the instruments are weak). If one can rule out the latter using the results of the first stage regression, then there is evidence that the new treatment does not work better than standard care. Moreover, if one has estimates of $\frac{\partial H}{\partial Z_j}$ and one knows the sign and can bound the magnitude of $\frac{\partial g(\boldsymbol{Z},\nu)}{\partial Z_j}$, then one can learn about the sign and magnitude of $\frac{\partial f(D,\boldsymbol{W},\upsilon)}{\partial D}$.

IV estimation of variable treatment effects

The effect of some treatment x on a health outcome y might vary between individuals. For instance, some medications are ineffective among individuals who lack the ability to metabolize certain compounds. Some individuals apparently do not seem to suffer any adverse consequences from cigarette smoking, while others do. A recent literature has analyzed the properties of the IV estimator when treatment effects vary by individual and individuals use private information to determine which treatment option is best. An illustrative example is Newhouse and McClellan's analysis of the impact of post-MI cardiac revascularization on mortality rates. Recall that they used as an instrument the "differential distance", the additional distance (if any) between the hospital closest to the patient's residence and a hospital with revascularization capacity. While this instrument was highly correlated with the receipt of revascularization, it was the not the only determining factor. Another, unobserved, factor, the patient's suitability for revascularization, also was found to be important. Indeed patients who were ill suited for revascularization would almost certainly not undergo the procedure, no matter how close they lived to a hospital which could perform the procedure. Other patients who were ideal

candidates for the procedure would eventually likely receive it, again, irrespective of their proximity to a hospital. So the IV estimator in this case tells us about the impact of revascularization on the subset of patients occupying the middle ground between being ideal candidates and being ill-suited for the procedure. For such patients, the effectiveness of the procedure is unclear and factors such as geographic proximity to a revascularization facility could be deciding factors in treatment decisions. Note that in this case, the IV treatment effect estimate underestimates the effectiveness of treatment for ideal candidates and overestimates the effectiveness of the treatment among those ill suited for the procedure.

To obtain more precise results about the behavior of the IV treatment effect estimator when treatment effects vary, let us modify the health outcome model (H1) slightly:

$$H_i = \beta_0 + \beta_{1i} D_i + \varepsilon_i$$

where i indexes subjects. Hence β_{1i} reflects the treatment effect specific to subject i. Imbens and Angrist (1994) demonstrate that the IV estimator converges to a weighted average of treatment effects where the weights are largest for subjects who vary their treatment choice D_i by the greatest degree in response to changes in the instruments. In particular, following the nomenclature of Auld (2006), if the instrument Z takes G different values, then under fairly general conditions the IV estimate of 'the' causal effect of D on H using Z as an instrument converges to:

$$b_1^{iv} \rightarrow \sum_{g=1}^{G} \lambda_g \beta_{1g}$$

where β_{1g} is the average effect of D on H in subpopulation g and the λ_g's are weights that depend on how much *D* varies with *Z* in subpopulation *g*. Recalling our previous example, one could imagine there being just two values of G: those who live near $(g = 1)$ and those who live far away $(g = 2)$ from a catheterization hospital. In both groups, those whose catheterization treatment decision does not depend on distance will contribute nothing to the treatment effect estimate. Hence the IV estimate reflects the average of the treatment effects in the remaining patients – those whose catheterization treatment decision depends on distance.

The lesson is that when TEs vary, IV estimates will tend to reflect the TEs of those whose treatment decision varies the most with variation in the instrument. One implication is that the interpretation of the IV estimator can depend on the instrument used. Two analysts, each using valid instruments, can legitimately produce different treatment effect estimates. To illustrate, suppose that catheterization treatments were assigned using a coin toss, not differential distance. Hence there would be again be two values of G: those randomized to receive catheterization treatment $(g = 1)$ and those randomized to standard care $(g = 2)$. If the randomization worked as intended, then the mix of patient types should be the same in both groups, implying that the weights λ_g and the average treatment effect β_{1g} should be the same in both groups as well. Moreover because each patient's treatment alllocation is equally dependent on the outcome of the coin toss, each patient will have an identical weight and the IV estimator will converge to a simple average of the β_{1i}. Note that the interpretation of the IV estimator in this context is different than the interpretation when distance was used as an instrument.

IV estimation of treatment effects in non-linear models

The models consider so far have been linear in parameters. Some models, however, are nonlinear. Of these nonlinear models, some can be rendered linear via a suitable transformation. For instance, if health outcomes are determined by the process:

$$H = e^{\beta_0} e^{\beta_1 D} W^{\beta_2} e^{\upsilon} \tag{H4}$$

Then, as Davidson and MacKinnon (2003, page 22) note, the model can be rendered linear in the parameters by taking the logarithm of both sides:

$$\ln H = \beta_0 + \beta_1 D + \beta_2 \ln w + \upsilon \tag{H5}$$

IV estimation of intrinsically nonlinear models can be handled using nonlinear IV estimation. Non-linear IV estimation is suitable when one can formulate one's model in the form of a non-linear regression:

$$H_i = \boldsymbol{x}_i(\boldsymbol{\beta}) + \varepsilon_i \tag{21}$$

where $\boldsymbol{x}_i(\boldsymbol{\beta})$ is a nonlinear regression function that depends on $\boldsymbol{\beta}$, a vector of K unknown parameters, the treatment indicator D_i and any other covariates included in the model. As before, ε_i represents the influence of all other determinants of H_i that are not explicitly modeled, some of which may be associated with D_i. This framework can accommodate a variety of models. Suppose, for example, that H_i is a count variable; perhaps H_i is the number of chronic health problems afflicting subject i. The Poisson model of H_i can be written as:

$$H_i = \exp(\beta_0 + \beta_1 D_i + \boldsymbol{W_i}'\boldsymbol{\gamma}) + \varepsilon_i \tag{22}$$

Conversely H_i might be a binary response; it could be the case that $H_i = 1$ if subject i has high blood pressure and $H_i = 0$ otherwise. The logit model of the probability that $H_i = 1$ can be written as:[2]

$$H_i = \frac{\exp(\beta_0 + \beta_1 D_i + \boldsymbol{W_i}'\boldsymbol{\gamma})}{1 + \exp(\beta_0 + \beta_1 D_i + \boldsymbol{W_i}'\boldsymbol{\gamma})} + \varepsilon_i \tag{23}$$

Nonlinear IV models are somewhat controversial. Terza (2006) is skeptical about the realism of models of the form (21) because they treat observed and latent determinants of health outcomes asymmetrically. While observed determinants are modeled using the non-linear function $\boldsymbol{x}_i(\boldsymbol{\beta})$, latent determinants are relegated to the additive error term. Hence the marginal effects of observed and unobserved covariates on H can be quite different, with no apparent justification. The one exception is the exponential model (22); Terza demonstrates that this model treats observed and unobserved health determinants symmetrically.

[2] As Davidson and MacKinnon (2003, page 456, 476) note, one could improve estimator precision by dividing observations on H_i and $\boldsymbol{x}_i(\boldsymbol{\beta})$ by the square root of the observation's error variance. The error variance in the Poisson case is equal to its conditional mean, while the error variance in the logit case is simply the variance of a Bernoulli distributed random variable: $p_i(1 - p_i)$ where $p_i = \frac{\exp(\beta_0 + \beta_1 D_i + \boldsymbol{W_i}'\boldsymbol{\gamma})}{1 + \exp(\beta_0 + \beta_1 D_i + \boldsymbol{W_i}'\boldsymbol{\gamma})}$.

Consistent estimation of the parameters of non-linear regression models requires that the following 'moment conditions' be satisfied:

$$\boldsymbol{X}(\boldsymbol{\beta})'\big(\boldsymbol{H} - \boldsymbol{x}(\boldsymbol{\beta})\big) = \mathbf{0} \tag{24}$$

where $\boldsymbol{X}(\boldsymbol{\beta}) \equiv \frac{\partial \boldsymbol{x}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ is the matrix of n observations on the K partial derivatives of the regression function with respect to each of the K parameters in $\boldsymbol{\beta}$. These moment conditions are the generalization to the non-linear regression context of the condition in the linear context that the errors be independent of the covariates. For instance, if the ith observation on $\boldsymbol{x}_i(\boldsymbol{\beta})$ is

$$\boldsymbol{x}_i(\boldsymbol{\beta}) = \beta_0 + \beta_1 D_i$$

Then

$$\boldsymbol{X}_i(\boldsymbol{\beta})' \equiv \left(\frac{\partial \boldsymbol{x}_i(\boldsymbol{\beta})}{\partial \beta_0} \quad \frac{\partial \boldsymbol{x}_i(\boldsymbol{\beta})}{\partial \beta_1}\right) = (1 \quad D_i)$$

Hence (24) would require that D and ε be orthogonal (i.e. the errors do not vary with values of D). When the errors are not orthogonal to $\boldsymbol{X}_i(\boldsymbol{\beta})$ then one can use nonlinear IV provided that one has a set of instruments $\boldsymbol{Z}$ that satisfy the condition that:

$$\boldsymbol{X}(\boldsymbol{\beta})'\boldsymbol{P}_{\boldsymbol{Z}^*}\big(\boldsymbol{H} - \boldsymbol{x}(\boldsymbol{\beta})\big) = \mathbf{0} \tag{25}$$

where $\boldsymbol{P}_{\boldsymbol{Z}^*}$ was previously defined and $\boldsymbol{X}(\boldsymbol{\beta})'\boldsymbol{P}_{\boldsymbol{Z}^*}$ are the predicted values from regressions of the columns of $\boldsymbol{X}(\boldsymbol{\beta})$ on $\boldsymbol{Z}^*$. Hence condition (22) states that the summary instruments be independent of the error terms. The non-linear IV estimator is the estimate of $\boldsymbol{\beta}$ that solves (25); this estimator is equivalent to the estimate of $\boldsymbol{\beta}$ that minimizes the criterion function:

$$Q(\boldsymbol{\beta}, \boldsymbol{H}) = \big(\boldsymbol{H} - \boldsymbol{x}(\boldsymbol{\beta})\big)'\boldsymbol{P}_{\boldsymbol{Z}^*}\big(\boldsymbol{H} - \boldsymbol{x}(\boldsymbol{\beta})\big) \tag{26}$$

Recall that IV estimation of the parameters of the linear model can be performed in two steps, wherein D is replaced by $\widehat{D}$, the predicted values from the regression of D on $\boldsymbol{Z}^*$, and this modified model estimated by OLS as per usual. When the model is non-linear, estimates must be derived from minimization of the criterion function (26); replacing D with $\widehat{D}$ and estimating the modified model by non-linear least squares will not yield consistent estimates. Suppose, for instance, that one was analyzing data from a RCT comparing the effectiveness of a new drug vs placebo on the probability of heart failure. Suppose further that compliance with the RCT was not perfect. One might be tempted to estimate (23) via non-linear least squares after replacing the observed D with $\widehat{D}$, the predicted values from a regression of D on $CoinToss$. This temptation should be resisted. One would need to minimize (26) or perhaps use the linear probability model instead.

Conclusions

Instrumental variables estimation can be a useful alternative to conventional covariate adjustment approaches. Finding good instruments, however, is not easy. Successful application of IV requires either experimental variation in treatment assignment or a source of quasi-experimental variation that

is incidental to the outcome being analysed. Furthermore, the desirable properties of IV are guaranteed to hold only as the sample size approaches infinity. In samples of modest size, IV estimates can be wildly inaccurate if instruments have only a modest effect on treatment or if there is even a weak correlation between instrument and the outcome being modeled. Finally, IV estimation when treatment effects are heterogenous requires careful consideration of the subjects whose treatment status is affected by variation in the instrument. IV reveals nothing about treatment effectiveness among subjects whose treatment status is non-responsive to variation in the instrument.

References

Angrist J, Krueger A. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics* 1991; 106(4):979-1014.

Angrist J, Krueger A. Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives* 2001; 15(4):69-85.

Auld MC. Using observational data to identify the effects of health-related behavior. In Andrew Jones (ed) *Elgar Companion to Health Economics*, Elgar, 2006.

Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. JASA. 1995;90:443– 450.

Davidson R, MacKinnon JG. *Econometric Theory and Methods*. New York: Oxford University Press, 2004.

Deaton A. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Johns Hopkins University Press, 1997

Fisher ES, Wennberg D, Stukel T, Gottlieb D, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care. *Annals of Internal Medicine* 2003a; 138(4):283-287.

Fisher ES, Wennberg D, Stukel T, Gottlieb D, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care. *Annals of Internal Medicine* 2003b; 138(4):288-299.

Hahn J, Hausman J. A new specification test for the validity of instrumental variables. *Econometrica* 2002; 70(1):163–189.

Imbens G, Angrist J. Identification and estimation of local average treatment effects. *Econometrica* 1994; 62(2):467-76.

Johnson J, DiNardo J. *Econometric Methods*, 4th edition. McGraw Hill, 1997.

Kennedy P. *A Guide to Econometrics*, 5th edition. Cambridge MA: MIT Press, 2003.

Newhouse J, McClellan M. Econometrics in outcomes research: The use of instrumental variables. *Annual Review of Public Health* 1998; 19:17-34.

Oreopoulos P. The compelling effects of compulsory schooling evidence from Canada. *Canadian Journal of Economics* 2006; 39(1): 22–52.

Snow J. *On the Mode of Communication of Cholera*. Churchill, London, 1855. [Reprinted (1965) by Hafner, New York.]

Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997; 65:557–586.

Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 2002; 20(4): 518-529.

Terza JV. Estimation of policy effects using parametric nonlinear models: a contextual critique of the generalized method of moments. *Journal of Health Services and Outcomes Research Methodology* 2006; 6:177-198.

Number	Title	Author(s)
(2005)		
No. 124:	Exploring the Use of a Nonparametrically Generated Instrumental Variable in the Estimation of a Linear Parametric Equation	F.T. Denton
No. 125:	Borrowing Constraints, The Cost of Precautionary Saving, and Unemployment Insurance	T.F. Crossley H.W. Low
No. 126:	Entry Costs and Stock Market Participation Over the Life Cycle	S. Alan
No. 127:	Income Inequality and Self-Rated Health Status: Evidence from the European Community Household Panel	V. Hildebrand P. Van Kerm
No. 128:	Where Have All The Home Care Workers Gone?	M. Denton I.U. Zeytinoglu S. Davies D. Hunter
No. 129:	Survey Results of the New Health Care Worker Study: Implications of Changing Employment Patterns	I.U. Zeytinoglu M. Denton S. Davies A. Baumann J. Blythe A. Higgins
No. 130:	Does One Size Fit All? The CPI and Canadian Seniors	M. Brzozowski
No. 131:	Unexploited Connections Between Intra- and Inter-temporal Allocation	T.F. Crossley H.W. Low
No. 132:	Grandparents Raising Grandchildren in Canada: A Profile of Skipped Generation Families	E. Fuller-Thomson
No. 133:	Measurement Errors in Recall Food Expenditure Data	N. Ahmed M. Brzozowski T.F. Crossley
No. 134:	The Effect of Health Changes and Long-term Health on the Work Activity of Older Canadians	D.W.H. Au T. F. Crossley M.. Schellhorn

SEDAP RESEARCH PAPERS: Recent Releases

Number	Title	Author(s)
No. 135:	Population Aging and the Macroeconomy: Explorations in the Use of Immigration as an Instrument of Control	F. T. Denton B. G. Spencer
No. 136:	Users and Suppliers of Physician Services: A Tale of Two Populations	F.T. Denton A. Gafni B.G. Spencer
No. 137:	MEDS-D USERS' MANUAL	F.T. Denton C.H. Feaver B.G.. Spencer
No. 138:	MEDS-E USERS' MANUAL	F.T. Denton C.H. Feaver B.G. Spencer
No. 139:	Socioeconomic Influences on the Health of Older Canadians: Estimates Based on Two Longitudinal Surveys (Revised Version of No. 112)	N.J. Buckley F.T. Denton A.L. Robb B.G. Spencer
No. 140:	Developing New Strategies to Support Future Caregivers of the Aged in Canada: Projections of Need and their Policy Implications	J. Keefe J. Légaré Y. Carrière
No. 141:	Les Premiers Baby-Boomers Québécois font-ils une Meilleure Préparation Financière à la Retraite que leurs Parents? Revenu, Patrimoine, Protection en Matière de Pensions et Facteurs Démographiques	L. Mo J. Légaré
No. 142:	Welfare Restructuring without Partisan Cooperation: The Role of Party Collusion in Blame Avoidance	M. Hering
No. 143:	Ethnicity and Health: An Analysis of Physical Health Differences across Twenty-one Ethnocultural Groups in Canada	S. Prus Z. Lin
No. 144:	The Health Behaviours of Immigrants and Native-Born People in Canada	J.T. McDonald
No. 145:	Ethnicity, Immigration and Cancer Screening: Evidence for Canadian Women	J.T. McDonald S. Kennedy
No. 146:	Population Aging in Canada: Software for Exploring the Implications for the Labour Force and the Productive Capacity of the Economy	F.T. Denton C.H. Feaver B.G. Spencer

Number	Title	Author(s)
(2006)		
No. 147:	The Portfolio Choices of Hispanic Couples	D.A. Cobb-Clark V.A. Hildebrand
No. 148:	Inter-provincial Migration of Income among Canada's Older Population:1996-2001	K.B. Newbold
No. 149:	Joint Taxation and the Labour Supply of Married Women: Evidence from the Canadian Tax Reform of 1988	T.F. Crossley S.H. Jeon
No. 150:	What Ownership Society? Debating Housing and Social Security Reform in the United States	D. Béland
No. 151:	Home Cooking, Food Consumption and Food Production among the Unemployed and Retired Households	M. Brzozowski Y. Lu
No. 152:	The Long-Run Cost of Job Loss as Measured by Consumption Changes	M. Browning T.F. Crossley
No. 153:	Do the Rich Save More in Canada?	S. Alan K. Atalay T.F. Crossley
No. 154:	Income Inequality over the Later-life Course: A Comparative Analysis of Seven OECD Countries	R.L. Brown S.G. Prus
No. 155:	The Social Cost-of-Living: Welfare Foundations and Estimation	T.F. Crossley K. Pendakur
No. 156:	The Top Shares of Older Earners in Canada	M.R. Veall
No. 157:	Le soutien aux personnes âgées en perte d'autonomie: jusqu'où les baby-boomers pourront-ils compter sur leur famille pour répondre à leurs besoins ?	J. Légaré C. Alix Y. Carrière J. Keefe
No. 158:	Les générations X et Y du Québec, vraiment différentes des précédentes ?	J. Légaré P.O. Ménard
No. 159: French	La diversification et la privatisation des sources de revenu de retraite au Canada	L. Mo J. Légaré L. Stone
No. 159: English	The Diversification and the Privatization of the Sources of Retirement Income in Canada	L. Mo J. Légaré L. Stone
No. 160:	Evaluating Pension Portability Reforms: The Tax Reform Act of 1986 as a Natural Experiment	V. Andrietti V.A. Hildebrand

SEDAP RESEARCH PAPERS: Recent Releases

Number	Title	Author(s)
No. 161:	Estimating a Collective Household Model with Survey Data on Financial Satisfaction	R. Alessie T.F. Crossley V.A. Hildebrand
No. 162:	Physician Labour Supply in Canada: A Cohort Analysis	T.F. Crossley J. Hurley S.H. Jeon
No. 163:	Tax Incentives and Household Portfolios: A Panel Data Analysis	S. Alan S. Leth-Petersen
No. 164:	The Healthy Immigrant Effect and Immigrant Selection: Evidence from Four Countries	S. Kennedy J.T. McDonald N. Biddle
No. 165:	Well-Being Throughout the Senior Years: An Issues Paper on Key Events and Transitions in Later Life	M. Denton K. Kusch
No. 166:	Satisfied Workers, Retained Workers: Effects of Work and Work Environment on Homecare Workers' Job Satisfaction, Stress, Physical Health, and Retention	I.U. Zeytinoglu M. Denton
No. 167:	Contrasting Inequalities: Comparing Correlates of Health in Canada and the United States	H. Armstrong W. Clement Z. Lin S. Prus
(2007)		
No. 168:	Health human resources planning and the production of health: Development of an extended analytical framework for needs-based health human resources planning	S. Birch G. Kephart G. Tomblin-Murphy L. O'Brien-Pallas R. Alder A. MacKenzie
No. 169:	Gender Inequality in the Wealth of Older Canadians	M. Denton L. Boos
No. 170:	The Evolution of Elderly Poverty in Canada	K. Milligan
No. 171:	Return and Onwards Migration among Older Canadians: Findings from the 2001 Census	K.B. Newbold
No. 172:	Le système de retraite américain: entre fragmentation et logique financière	D. Béland

SEDAP RESEARCH PAPERS: Recent Releases

Number	Title	Author(s)
No. 173:	Entrepreneurship, Liquidity Constraints and Start-up Costs	R. Fonseca P.-C. Michaud T. Sopraseuth
No. 174:	How did the Elimination of the Earnings Test above the Normal Retirement Age affect Retirement Expectations?	P.-C. Michaud A. van Soest
No. 175:	The SES Health Gradient on Both Sides of the Atlantic	J. Banks M. Marmot Z. Oldfield J.P. Smith
No. 176:	Pension Provision and Retirement Saving: Lessons from the United Kingdom	R. Disney C. Emmerson M. Wakefield
No. 177:	Retirement Saving in Australia	G. Barrett Y.-P. Tseng
No. 178:	The Health Services Use Among Older Canadians in Rural and Urban Areas	H. Conde J.T. McDonald
No. 179:	Older Workers and On-the-Job Training in Canada: Evidence from the WES data	I.U. Zeytinoglu G.B. Cooke K. Harry
No. 180:	Private Pensions and Income Security in Old Age: An Uncertain Future – Conference Report	M. Hering M. Kpessa
No. 181:	Age, SES, and Health: A Population Level Analysis of Health Inequalitites over the Life Course	S. Prus
No. 182:	Ethnic Inequality in Canada: Economic and Health Dimensions	E.M. Gee K.M. Kobayashi S.G. Prus
No. 183:	Home and Mortgage Ownership of the Dutch Elderly: Explaining Cohort, Time and Age Effects	A. van der Schors R.J.M. Alessie M. Mastrogiacomo
No. 184:	A Comparative Analysis of the Nativity Wealth Gap	T.K. Bauer D.A. Cobb-Clark V. Hildebrand M. Sinning
No. 185:	Cross-Country Variation in Obesity Patterns among Older Americans and Europeans	P.C. Michaud A. van Soest T. Andreyeva

Number	Title	Author(s)
No. 186:	Which Canadian Seniors Are Below the Low-Income Measure?	M.R. Veall
No. 187:	Policy Areas Impinging on Elderly Transportation Mobility: An Explanation with Ontario, Canada as Example	R. Mercado A. Páez K. B. Newbold
No. 188:	The Integration of Occupational Pension Regulations: Lessons for Canada	M. Hering M. Kpessa
No. 189:	Psychosocial resources and social health inequalities in France: Exploratory findings from a general population survey	F. Jusot M. Grignon P. Dourgnon
No. 190:	Health-Care Utilization in Canada: 25 Years of Evidence	L.J. Curtis W.J. MacMinn
No. 191:	Health Status of On and Off-reserve Aboriginal Peoples: Analysis of the Aboriginal Peoples Survey	L.J. Curtis
No. 192:	On the Sensitivity of Aggregate Productivity Growth Rates to Noisy Measurement	F.T. Denton
No. 193:	Initial Destination Choices of Skilled-worker Immigrants from South Asia to Canada: Assessment of the Relative Importance of Explanatory Factors	L. Xu K.L. Liaw
No. 194:	Problematic Post-Landing Interprovincial Migration of the Immigrants in Canada: From 1980-83 through 1992-95	L. Xu K.L. Liaw
No. 195:	Inter-CMA Migration of the Immigrants in Canada: 1991-1996 and 1996-2001	L. Xu
No. 196:	Characterization and Explanation of the 1996-2001 Inter-CMA Migration of the Second Generation in Canada	L. Xu
No. 197:	Transitions out of and back to employment among older men and women in the UK	D. Haardt
No. 198:	Older couples' labour market reactions to family disruptions	D. Haardt
No. 199:	The Adequacy of Retirement Savings: Subjective Survey Reports by Retired Canadians	S. Alan K. Atalay T.F. Crossley
No. 200:	Underfunding of Defined Benefit Pension Plans and Benefit Guarantee Insurance - An Overview of Theory and Empirics	M. Jametti
No. 201:	Effects of 'authorized generics' on Canadian drug prices	P. Grootendorst

SEDAP RESEARCH PAPERS: Recent Releases

Number	Title	Author(s)
No. 202:	When Bad Things Happen to Good People: The Economic Consequences of Retiring to Caregive	P.L. McDonald T. Sussman P. Donahue
No. 203:	Relatively Inaccessible Abundance: Reflections on U.S. Health Care	I.L. Bourgeault
No. 204:	Professional Work in Health Care Organizations: The Structural Influences of Patients in French, Canadian and American Hospitals	I.L. Bourgeault I. Sainsaulieu P. Khokher K. Hirschkorn
No. 205:	Who Minds the Gate? Comparing the role of non physician providers in the primary care division of labour in Canada & the U.S.	I.L. Bourgeault
No. 206:	Immigration, Ethnicity and Cancer in U.S. Women	J.T. McDonald J. Neily
No. 207:	Ordinary Least Squares Bias and Bias Corrections for *iid* Samples	L. Magee
No. 208:	The Roles of Ethnicity and Language Acculturation in Determining the Interprovincial Migration Propensities in Canada: from the Late 1970s to the Late 1990s	X. Ma K.L. Liaw
No. 209:	Aging, Gender and Neighbourhood Determinants of Distance Traveled: A Multilevel Analysis in the Hamilton CMA	R. Mercado A. Páez
No. 210:	La préparation financière à la retraite des premiers boomers : une comparaison Québec-Ontario	L. Mo J. Légaré
No. 211:	Explaining the Health Gap between Canadian- and Foreign-Born Older Adults: Findings from the 2000/2001 Canadian Community Health Survey	K.M. Kobayashi S. Prus
No. 212:	“Midlife Crises”: Understanding the Changing Nature of Relationships in Middle Age Canadian Families	K.M. Kobayashi
No. 213:	A Note on Income Distribution and Growth	W. Scarth
No. 214:	Is Foreign-Owned Capital a Bad Thing to Tax?	W. Scarth
No. 215:	A review of instrumental variables estimation in the applied health sciences	P. Grootendorst