NBER TECHNICAL WORKING PAPER SERIES

INSTRUMENTAL VARIABLES ESTIMATION OF AVERAGE TREATMENT EFFECTS IN ECONOMETRICS AND EPIDEMIOLOGY

Joshua D. Angrist

Technical Working Paper No. 115

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 1991

The author thanks the Harvard Institute for Economic Research and the National Science Foundation for financial support, and Tom Newman and Norman Hearst for comments and data. Thanks also go to Guido Imbens, Whitney Newey, and participants in the Harvard-MIT econometrics workshop for helpful suggestions and comments, to Bill Greene for help with software, and to Gary Chamberlain and Jamie Robins for a number of particularly helpful discussions. Bill Rodgers provided expert research assistance. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #115
November 1991

INSTRUMENTAL VARIABLES ESTIMATION OF AVERAGE TREATMENT EFFECTS IN ECONOMETRICS AND EPIDEMIOLOGY

ABSTRACT

The average effect of intervention or treatment is a parameter of interest in both epidemiology and econometrics. A key difference between applications in the two fields is that epidemiologic research is more likely to involve qualitative outcomes and nonlinear models. An example is the recent use of the Vietnam era draft lottery to construct estimates of the effect of Vietnam era military service on civilian mortality. In this paper, I present necessary and sufficient conditions for linear instrumental variables techniques to consistently estimate average treatment effects in qualitative or other nonlinear models. Most latent index models commonly applied to qualitative outcomes in econometrics fail to satisfy these conditions, and monte carlo evidence on the bias of instrumental estimates of the average treatment effect in a bivariate probit model is presented. The evidence suggests that linear instrumental variables estimators perform nearly as well as the correctly specified maximum likelihood estimator, especially in large samples. Linear instrumental variables and the normal maximum likelihood estimator are also remarkably robust to non-normality.

Joshua D. Angrist
Department of Economics
Hebrew University of Jerusalem
Jerusalem, Israel 91905
and NBER

The difference in civilian mortality experience between World War Two veterans and nonveterans is a classic example of the confounding impact of selection on estimates of treatment effects. The fact that World War Two veterans have lower mortality than nonveterans the same age is usually attributed to the military screening process (Seltzer and Jablon 1974) and not to any beneficial effects of wartime service. Recently, Alan Krueger and I (Angrist and Krueger 1989) have shown that the higher earnings enjoyed by veterans from World War Two cohorts are also an artifact of selection bias. Using the fact that some World War Two veterans were drafted according to a sequence determined by day of birth, we construct Instrumental Variables (IV) estimates of the effects of World War Two veteran status on earnings that are free of selection bias. These estimates show that World War Two veterans probably earn less than they would have if they had not served in the military.

Whenever some covariates related to both an outcome of interest and the probability of treatment are unobserved or unaccounted for, the likelihood of selection bias renders inferences based on simple comparisons invalid. In such cases, IV estimation provides a powerful and flexible method of correcting for omitted variables bias. IV estimates are constructed by comparing the outcomes of groups with different values of an (instrumental) variable that is related to the outcome of interest solely by virtue of correlation with the probability of treatment. An important recent example of this approach in epidemiology is the work by Hearst, Newman, and Hulley (1986) on the effects of Vietnam Era military service on civilian mortality. Although Vietnam era veteran status is a consequence of both self-selection and military screening, Hearst, Newman and Hulley use the draft lottery to construct IV estimates of the effects of Vietnam

era service that are free of selection bias. Similarly, I have used the Vietnam era draft lottery (Angrist 1990) to estimate the effects of Vietnam era service on civilian earnings.

In econometrics, most applications of IV estimators involve linear models with continuous outcome variables. Applications in epidemiology, such as the smoking study by Permutt and Hebel 1989), can also involve linear models. But applications such as the Vietnam mortality example and Hearst, Buehler, Newman, and Rutherford's (1990) recent study of intravenous drug use among veterans involve limited dependent variables that are usually fitted using nonlinear models. This paper discusses the use of linear IV techniques to estimate average treatment effects in such nonlinear models.

Linear IV techniques are attractive for several reasons. First, the source of identifying information is transparent to the consumer of applied research: the instruments generate a natural experiment that assigns treatment in a manner independent of unobserved covariates. Second, the linear IV estimator does not require observations on individuals; sample covariances are sufficient statistics for estimates of regression parameters. In some applications, (e.g., Angrist and Krueger 1990a, and below), these sample moments are actually taken from different data sets. Third, consistency of IV estimates does not require consistent estimation of the reduced form for endogenous regressors (Kelejian 1971). This is particularly important in an evaluation context, where the reduced form for an endogenous dummy variable is also likely to be nonlinear.

The paper is organized as follows. To further motivate the use of IV in nonlinear qualitative response models, Section 1 presents an illustration using the Hearst, Newman, and Hulley data on Vietnam era

military service and mortality. Section 2 defines the average treatment effect for a class of nonlinear models. Section 3 presents necessary and sufficient conditions for linear IV moment conditions to identify average treatment effects. Section 4 discusses the asymptotic bias of IV estimates of average treatment effects when the identification conditions fail to hold.

In an influential paper in econometrics, Heckman (1978) used a latent index/simultaneous equations model with normally distributed errors to develop estimation strategies for treatment effects in nonlinear limited dependent variable models. A commonly encountered model of this type is bivariate probit (e.g., Ashford and Snowden 1970, Amemiya 1978) which, like most latent variable models, does not satisfy the identification conditions required for linear IV techniques to consistently estimate average treatment effects. Section 5 of the paper contains a detailed study of average treatment effects in bivariate probit. Included in this section are the results of a Monte Carlo sampling experiment that compares the finite sample performance of Maximum Likelihood (ML) and IV estimates of average treatment effects in correctly specified and misspecified models. The paper concludes in Section 6.

1. The Effect of Vietnam Veteran Status on Mortality

In addition to combat-related injuries, military service during the Vietnam war may also affect civilian mortality because of wartime access to narcotics, exposure to toxins like Agent Orange, and causes related to Post-Traumatic Stress Syndrome such as suicide. Hearst, Newman and Hulley (1986) showed that men with sequence numbers that put them at high risk of being drafted in the Vietnam Era draft lotteries had elevated mortality risk after their discharge from the military. They attribute this elevated

risk to a higher probability of military service because between 1970 and 1973, the risk of being drafted was randomly assigned in a series of lotteries based on dates of birth. Each date of birth in the cohorts at risk of being drafted was assigned a Random Sequence Number (RSN) from 1-365. The Selective Service called men for induction by RSN up to a ceiling determined by the Department of Defense. Men born in 1950 were called up to RSN 195, men born in RSN 1951 were called up to RSN 125, and men born in 1952 were called up to RSN 95.

In their paper, Hearst, Newman and Hulley focus on comparisons of mortality risk by draft-eligibility status. For example, they compare the number of deaths of men born in 1950 with RSN below 195 to the number of deaths of men born in 1950 with RSN above 195. This procedure can be used to provide a valid estimate of the effects of military service on mortality if draft-eligibility is correlated with civilian mortality solely by virtue of its correlation with veteran status. Although not explicitly stated, the assumptions and estimation techniques used by Hearst, Newman, and Hulley can be interpreted as an application of IV estimation to the linear model

$$y_i = \alpha_c + \beta s_i + x_i\gamma \tag{1.1}$$

where y_i is a binary indicator of death in the study interval, s_i indicates veteran status, β is the treatment effect of interest, α_c is a cohort-specific intercept, and x_i is an unobserved confounding variable.[1]

The key identifying assumption justifying IV estimation in this case is that $E[x_i|$ draft-eligibility status$] = 0$. Given this assumption, a consistent estimate of β can be obtained by applying the simplest IV

[1]Similarly, an estimator proposed by Robins (1989) is an application of instrumental variables techniques to a linear model for survival analysis.

estimator -- Wald'a (1940) method of fitting straight lines -- to data for a single cohort and race.[2] To see this, note that because draft-eligibility is uncorrelated with x_1, a consistent Wald estimate can be computed by dividing the data by draft-eligibility status:

$$\hat{\beta} = (\bar{y}^e - \bar{y}^n)/(\hat{p}^e - \hat{p}^n), \tag{1.2}$$

where $\bar{y}$ is the probability of death, $\hat{p}$ is the probability of veteran status, and superscript e and superscript n denote the draft-eligible and draft-ineligible samples. The instrument here is an indicator of draft-eligibility status. Using data in the Appendix Table for white men born in 1950, we have

$$\hat{\beta} = (.0204 - .0195)/(.3527 - .1933) = .00564,$$

so that being a veteran raises mortality risk for this group by half a percentage point -- a 25% increase in risk.[3]

An efficient linear combination of alternative Wald estimates of the same parameter can be computed by Generalized Least Squares (GLS) estimation of the equation

---

[2]The observation that Wald's estimator is also an instrumental variables estimator is usually attributed to Durbin (1954). The results in Hearst, Newman and Hulley (1986) are actually for relative risk, in contrast to estimates of the "risk-difference" generated by Wald's method. It can be shown, however, that the Hearst, Newman, and Hulley relative risk estimates are a simple transformation of Wald risk-difference estimates.

[3]A standard error for the Wald estimate is calculated easily under the null hypothesis that $\beta = 0$. All moments in this example are independent, so the limiting distribution of $(\bar{y}^e - \bar{y}^n)/(\hat{p}^e - \hat{p}^n)$, is $1/(p^e - p^n)$ times the limiting distribution of the numerator. The denominators in $\bar{y}^e$ and $\bar{y}^n$ are $N^e = 127{,}500$ and $N^n = 111{,}200$. The sampling variance of the Wald estimate is therefore

$$[1/(\hat{p}^e-\hat{p}^n)]^2\{\ [\bar{y}^e(1-\bar{y}^e)]/N^e + [\bar{y}^n(1-\bar{y}^n)]/N^n\ \}$$

The estimated standard error is 0.00368 .

$$\bar{y}_{crj} = \alpha_c + \beta_0 R_{crj} + \hat{p}_{crj}\beta_1 + \bar{\epsilon}_{crj}, \tag{1.3}$$

where the subscripts c, r and j index birth cohort, race and draft-eligibility status, and the data are grouped into averages for cells defined by these three variables. Since there are 3 birth cohorts, 2 races, and 2 eligibility groups, equation (1.3) is fit to 12 observations. Angrist (1991) shows that GLS estimates of grouped equations such as (1.3) are the same as Two-Stage Least Squares (TSLS) estimates that efficiently combine all possible IV estimates. In this case, the TSLS instrument set contains dummy variables for the full set of cohort, race and draft-eligibility interactions. Because the estimating equation includes main effects for cohort and race, the effect of veteran status is identified by the exclusion of draft-eligibility status for each race and cohort from the estimating equation.

Figure 1 presents a graphical version of equation (1.3) for suicide, the major cause of death associated with veteran status in these data. The figure graphs residuals from a regression of suicide probabilities on race and cohort dummies, against the corresponding residuals for veteran status probabilities. Thus, the slope of the line in the figure is an estimate of β_1 -- in this case, equal to 0.258 percentage points, with a standard error of 0.06 .

Although Figure 1 clearly shows a strong linear relationship, for a variety of reasons most textbook discussions of limited dependent variable models (e.g., Maddala 1983) argue that the linear model used in this illustration is inappropriate for binary outcomes. Problems include the fact that fitted values in the linear model are not bounded between zero and one. Also, least squares estimation of a linear model does not reflect the fact that a probability distribution is being parameterized, as would,

say, ML estimation of a logit model. Note, however, that linear IV estimation breaks down in logistic or other nonlinear regression models because the IV moment condition fails to hold: the conditional expectation of nonlinear transformations of $\alpha_c + \beta s_i + x_i\gamma$ is not equal to the conditional expectation of the outcome variable, even when conditional variables are mean-independent of unobservables. Such problems notwithstanding, part of the purpose of this paper is to offer formal arguments that help rationalize linear IV estimation in nonlinear models.

2. Nonlinear Models with Omitted Covariates

The general model of interest relates n observations on an outcome variable, y_i, to a treatment indicator, s_i, in the following manner:

$$E[y_i|\ s_i,\ U_i,\ Z_i] = F[s_i,\ U_i;\ \beta] \tag{2.1}$$

$$E[s_i|\ U_i,\ Z_i] = G[Z_i,\ U_i;\ \gamma], \tag{2.2}$$

where U_i is an $r\times 1$ vector of covariates unobserved by the econometrician, Z_i is a $q\times 1$ vector of potential instrumental variables, F and G are functions, and β and γ are parameter vectors. Observed covariates are held constant by estimating in subpopulations (i.e., racial groups). A nonlinear example of (1.1) is given in Rosenbaum and Rubin's (1983) study of the effect of an unobserved binary covariate on estimated average treatment effects in a model for binary outcomes.

The n observations are assumed to be independent and identically distributed. The development that follows can be applied to heterogeneously distributed samples with few modifications. The vector Z_i includes a constant and satisfies rank and independence conditions:

Assumption 1: (i) $E[Z_i'(s_i \ 1)] = \Phi_{zs}$, where the rank of $\Phi_{zs} \geq 2$,

(ii) U_i and Z_i are independent.

Assumption 1(i) is the standard requirement that potential instruments be correlated with regressors. The Z_i are potential instruments because equation (2.1) implies

$$E(y_i - F[s_i, U_i; \beta] \mid Z_i) = 0 \tag{2.3}$$

by the law of iterated expectations, and because Z_i is independent of unobservables by A2(ii).

In many econometric applications, attention is focused on identification strategies for theoretical parameters in the outcome equation (β's) such as marginal rates of substitution. But in evaluation studies, the substantive question motivating applied research concerns the effect of treatment on an outcome, as opposed to the magnitude of structural parameters that arise in economic theory. I therefore focus on the identification of average treatment effects. For model (2.1) the average treatment effect is defined as follows:

$$\pi_1 = E(F[1, U_i; \beta] - F[0, U_i; \beta]). \tag{2.4}$$

The average treatment effect in this model is primarily of interest because of its relationship to the Holland-Rubin (Holland 1986, Rubin 1974) definition of an average causal effect. The Holland-Rubin definition of causality is based on the notion that for each individual we can conceive of outcomes that would occur with and without treatment. The average causal effect is the expectation of an outcome variable when all

individuals in a given population receive treatment minus the expectation of an outcome variable when no individuals receive treatment. The fundamental problem of causal inference is that in practice, we never observe outcomes for any single individual both with and without treatment (Holland 1986). Nevertheless, we can sometimes estimate average causal effects. The parameter π_1 is an average causal effect as long as the average of y_i given s_i and U_i is an unbiased estimate of the average of y_i if all members of the population with a given value of U_i had the same value of s_i. Rosenbaum and Rubin (1983) call this property strong ignorability of s_i given U_i. For example, s_i is strongly ignorable given U_i if it is randomly assigned conditional on U_i.

3.1 Identification Conditions

The identification conditions for IV estimation of π_1 involve the additive separability of F and G, defined as follows:

Assumption 2: (i) $F[s_i, U_i; \beta] = f_1(s_i; \beta) + f_2(U_i; \beta)$,

(ii) $G[Z_i, U_i; \gamma] = g_1(Z_i; \gamma) + g_2(U_i; \gamma)$,

where f_1, f_2, g_1, and g_2 are functions.[4] Assumption 2 characterizes the class of models for which average treatment effects can be estimated using linear IV techniques. This result is formalized in the following proposition:

Proposition 1: Let π_1 be the average treatment effect defined in (2.4).

---

[4] Because s_i is binary, we could replace $f_1(s_i; \beta)$ with $f_{10} + f_{11}s_i$, where f_{10} and f_{11} are constants.

Then for some constant π_0 and for all y_i, U_i and Z_i satisfying (2.1), (2.2), and Assumption 1(ii):

$$E[y_i | Z_i] = \pi_0 + \pi_1 E[s_i | Z_i]$$

if and only if either or both Assumption 2(i) and Assumption 2(ii) holds.

Proof. Sufficiency of 2(i) is immediate because of the independence of U_i and Z_i. To establish sufficiency of 2(ii), write (1.3) as

$$E[y_i | Z_i] = E\{ F[s_i, U_i; \beta] | Z_i \} = \kappa_0 + E[\pi_i s_i | Z_i] \tag{3.1}$$

where

$$\kappa_0 = E\{ F[0, U_i; \beta] | Z_i \} \qquad \pi_i = F[1, U_i; \beta] - F[0, U_i; \beta].$$

Note that

$$\begin{aligned} E[\pi_i s_i | Z_i] &= E\{ \pi_i E[s_i | Z_i, U_i] \mid Z_i \} \\ &= E\{ \pi_i [g_1(Z_i; \gamma) + g_2(U_i; \gamma)] \mid Z_i \} \\ &= g_1(Z_i; \gamma) E[\pi_i | Z_i] + E[\pi_i g_2(U_i; \gamma) | Z_i]. \end{aligned}$$

But independence of U_i and Z_i implies that $E[\pi_i g_2(U_i; \gamma) | Z_i]$ is a constant, say, κ_1. Also, $E[\pi_i | Z_i] = \pi_1$, the average treatment effect, and

$$g_1(Z_i; \gamma) = E[s_i | Z_i] - E[g_2(U_i; \gamma) | Z_i] = E[s_i | Z_i] - \kappa_2.$$

Therefore,

$$E[\pi_i s_i | Z_i] = (E[s_i | Z_i] - \kappa_2)\pi_1 + \kappa_1,$$

so that sufficiency holds for $\pi_0 = \kappa_0 + \kappa_1 - \kappa_2 \pi_1$.

Necessity is established by showing that the proposition cannot hold for some Z_i satisfying (2.1), (2.2) and Condition 1 when neither 2(i) or 2(ii) hold. Suppose that 2(i) does not hold, that $Z_i = [1 \; z_i]'$, where z_i is a single dummy variable, and that 2(ii) does not hold because

$$G[Z_i, U_i; \gamma] = g_1(Z_i; \gamma) + g_2(U_i; \gamma) + z_i g_3(U_i; \gamma),$$

where g_3 is a function of U_i. Then,

(3.2) $$E[y_i|\ Z_i] = \pi_0 + \pi_1 E[s_i|\ Z_i] + z_i E[\pi_i g_3(U_i;\ \gamma)],$$

and $E[\pi_i g_3(U_i;\ \gamma)]$ is a constant that is not generally equal to zero. This completes the proof.

In practice, the plausibility of either Assumption 2(i) or Assumption 2(ii) has to be considered on a case by case basis. For example, latent index models other than the uniform-linear probability model are unlikely to be additively separable. On the other hand, over a limited range that is generally around the median, many cumulative distribution functions are approximately linear. If the variation in the outcome equation or in the expectation of the endogenous regressor is close to this range, linear IV estimation may give a good approximation to the true average treatment effect in models with dummy endogenous variables.

The usefulness of Proposition 1 for econometricians should also be evaluated in light of the fact that economic theory usually provides little guidance as to appropriate parametric distributional assumptions. In some circumstances, restrictions on functional form might be easier to rationalize and test than a distributional assumption. For example, standard instrument-error orthogonality test statistics (e.g., Newey 1985b) may have the power to detect failure of the functional form assumptions required to identify average treatment effects. Moreover, if the underlying distributions or functional forms are misspecified, it becomes an empirical question whether linear IV estimators do a worse job than ML estimators. This is among the questions investigated in Section 5.

It should also be noted that many of the existing non-parametric procedures developed by econometricians for the estimation of latent variable models cannot be used to estimate average treatment effects. For

example, Stoker's (1986) average derivative estimator cannot be used to estimate the effect of discrete explanatory variables. Newey's (1985a) application of Manski's (1975) maximum score estimator to binary response models with endogenous regressors consistently estimates the coefficients on discrete endogenous regressors up to scale. Similarly, Newey's (1986) non-parametric estimator for limited dependent variable models with endogenous regressors consistently estimates ratios of index coefficients. But neither of the Newey procedures recovers enough information to estimate the average effect of treatment on outcomes. Heckman's (1990) results on the nonparametric identifiability of treatment effects may also be of limited practical use because these results require continuously distributed regressors.

4.1 The Large Sample Bias of IV Estimates

Equation (3.1) can be used to make some general statements about the asymptotic bias of IV estimates of the average treatment effect in models where neither F or G are additively separable. Here it is useful to note that, as a consequence of (3.1), the model can be written using a random coefficients notation:[5]

$$y_i = \kappa_i + \pi_i s_i + \epsilon_i, \tag{4.1}$$

where $\kappa_i = F[0, U_i; \beta]$ and $E[\epsilon_i | Z_i] = 0$. Therefore,

---

[5]Heckman and Robb (1985) also discuss instrumental variables estimation of a random treatment effect. In their discussion of random coefficients models for treatment effects, Heckman and Robb suggest that consistent estimates be obtained through by applying a combination of behavioral (latent index) and distributional assumptions that can be used to compute the theoretical conditional expectation of treatment given instruments and covariates.

$$y_i = \kappa_0 + \pi_1 s_i + [\epsilon_i + (\kappa_i - \kappa_0) + s_i(\pi_i - \pi_1)]. \tag{4.2}$$

Since ϵ_i and $(\kappa_i - \kappa_0)$ are uncorrelated with Z_i, TSLS estimates of π_1 converge to

$$\pi_1 + \operatorname*{plim}_{n \to \infty} [\ (\Sigma\ \hat{s}_i^2)^{-1}\ \Sigma\ \hat{s}_i s_i(\pi_i - \pi_1)\ ] \tag{4.3}$$

where $\hat{s}_i = Z_i(Z'Z)^{-1}Z's - \bar{s}$, and $\bar{s}$ is the sample mean of s_i.

The TSLS estimate of π_1 is consistent under A2(i) because π_i is identically equal to π_1. The TSLS estimate of π_1 is consistent under A2(ii) because $\operatorname*{plim}_{n \to \infty} \Sigma\ \hat{s}_i s_i(\pi_i - \pi_1)\ ]$ is zero in this case, even though π_i is not equal to π_1 for all i. To see this, let δ be the vector of population regression coefficients for a regression of s_i on Z_i, and let $\tilde{z}_i$ equal Z_i minus its mean. We have

$$\operatorname*{plim}_{n \to \infty} \Sigma\ [s_i(\pi_i - \pi_1)\hat{s}_i]/n = E(E[s_i(\pi_i - \pi_1)|\ Z_i]\tilde{z}_i)\delta,$$

by the weak law of large numbers. But $E[s_i(\pi_i - \pi_1)|\ Z_i]$ is constant under A2(ii) and $E(\tilde{z}_i) = 0$.

To evaluate expression (4.3) for the general case, note that

$$\operatorname*{plim}_{n \to \infty} \Sigma\ \hat{s}_i^2/n = \delta'E(\tilde{z}_i'\tilde{z}_i)\delta$$

and

$$\operatorname*{plim}_{n \to \infty} \Sigma\ [s_i(\pi_i - \pi_1)\hat{s}_i]/n = E[G(Z_i, U_i;\ \gamma)(\pi_i - \pi_1)\tilde{z}_i]\delta.$$

The asymptotic bias of a linear IV estimate of the average treatment effect is therefore

$$[\delta'E(\tilde{z}_i'\tilde{z}_i)\delta]^{-1}E[G(Z_i, U_i;\ \gamma)(\pi_i - \pi_1)\tilde{z}_i]\delta. \tag{4.4}$$

The bias formula can be simplified further if $G(Z_i, U_i; \gamma)$ is linear and $\tilde{z}_i$ is a scalar:

$$G(Z_i, U_i; \gamma) = \gamma_0 + \gamma_1 \tilde{z}_i + \gamma_2 \tilde{u}_i + \gamma_3 \tilde{u}_i \tilde{z}_i, \tag{4.5}$$

where $\tilde{u}_i$ is U_i minus its mean. Expression (3.4) can now be written,

$$[\gamma_3/\gamma_1] E[\tilde{u}_i(\pi_i - \pi_1)], \tag{4.6}$$

because $\delta = \gamma_1$ in this case. This version of the bias formula is useful because it highlights the role of both the interaction between U_i and Z_i in G, and the covariance of the treatment effect with unobserved characteristics, in determining the asymptotic bias of IV estimates. The interaction term γ_3 is a measure of the non-separability of G, while $E[\tilde{u}_i(\pi_i - \pi_1)]$ is a measure of the non-separability of F. The bias asymptotic is also inversely proportional to a measure of the quality of the instruments, γ_1.

4.2 Optimal Weighting and Choice of Instruments

Properties of the conditional variance of residuals determine the appropriate Generalized Least Squares (GLS) weighting matrix for IV estimators. Write

$$\nu_i = [\epsilon_i + (\kappa_i - \pi_0) + s_i(\pi_i - \pi_1)],$$

for the compound error term in equation (4.2). If the outcome variable is continuous and estimation is based on Assumption 2(i) so that $\nu_i = \epsilon_i$, then it may be reasonable to assume that ν_i is homoscedastic. In this case, conventional TSLS is the most efficient way to use the elements of Z_i

as instruments, and TSLS covariance formulas will give asymptotically correct standard errors for estimates of average treatment effects. But if estimation is justified by A2(ii), the presence of the terms $(\kappa_i - \kappa_0)$ and $(\pi_i - \pi_1)s_i$ in ν_i suggests that ν_i will be heteroscedastic even if ϵ_i is homoscedastic.

Chamberlain's (1987) results on efficient estimation under conditional moment restrictions imply that the optimal instruments are a function of both Z_i and the conditional variance of residuals given Z_i. For estimation of the coefficient on a dummy variable with independent observations, the optimal instruments are (Newey 1989):

$$D[Z_i] = \psi(Z_i)^{-1} * [1 \; E(s_i | Z_i)]', \tag{4.8}$$

where $\psi(Z_i) = E[\nu_i^2 | Z_i]$. Estimates computed using $D[Z_i]$ as instruments asymptotically attain the variance bound for conditional moments estimation. As a practical matter, however, it often infeasible to use the optimal instruments because both $E[s_i | Z_i]$ and $\psi(Z_i)$ are unknown. An important exception to this is when Z_i is discrete with finite support. With discrete Z_i, the optimal IV estimator is a feasible weighted least squares estimator. This estimator can be computed by using a full set of dummy variables to indicate each value of Z_i as instruments in White (1982) TSLS estimation.[6]

---

[6]Suppose that Z_i can take on $j = 1, \ldots, J$ values and let R denote a matrix of J dummy variables that indicate each value of Z. That is, $R_{ij} = 1(Z_i = j)$. Using R as the matrix of instruments, White's TSLS estimator can be interpreted as instrumental variables estimation with instruments equal to $R[\Sigma R_i R_i' \hat{\nu}_i^2/n]^{-1}[R'X/n]$, where $X = [\iota \; s]$. This simplifies to $R * [\bar{x}_j/\hat{\sigma}_j^2]$ where $[\bar{x}_j/\hat{\sigma}_j^2]$ is a matrix with J rows each

5.1 Average Treatment Effects in Bivariate Probit

An important special case of the problem considered in this paper is the estimation of treatment effects in models with qualitative or binary dependent variables. Qualitative response models with endogenous regressors have generated a large theoretical and applied literature. Ashford and Snowden (1970) are usually credited with introducing a bivariate probit model in biometrics and Amemiya (1974) developed a minimum chi-square estimator for this model. Bivariate probit is also among the latent index/simultaneous equations models outlined by Heckman (1978). A variety of econometric aspects of the model are considered by Amemiya (1978) and Newey (1987), and some applications are described in Maddala (1983). The latent index approach to qualitative response analysis is popular in econometrics because the indices correspond to unobserved utilities in the theory of discrete choice.

Bivariate probit with endogenous dummy regressors can be motivated by the following latent index model:

$$(5.1) \qquad y_i = 1 \text{ if } y_i^* = \beta_0 + \beta_1 s_i + [\eta_{1i} - u_i\lambda] > 0$$
$$= 0 \quad \text{otherwise}$$

$$(5.2) \qquad s_i = 1 \text{ if } s_i^* = \gamma_0 + \gamma_1 Z_i + [\eta_{2i} - u_i] > 0$$
$$= 0 \quad \text{otherwise,}$$

where η_{1i}, η_{2i}, and u_i are independent, normally distributed random

---

containing the average of X given Z = j, divided by the variance of X given Z = j. For continuous Z_i in a homoscedastic model, Newey (1989) proposes a number of asymptotic approximations to the optimal instruments.

variables. The treatment indicator, s_i, is endogenous because the compound error terms,

$$\eta_{1i}^{*} = \eta_{1i} - u_i\lambda \quad \text{and} \quad \eta_{2i}^{*} = \eta_{2i} - u_i,$$

are correlated. Note that this formulation preserves the definition of endogeneity implicit in the previous sections: if all relevant covariates were observed, then the effect of treatment on y_i could be estimated using single equation techniques.

The compound error terms in (5.1) and (5.2) are also normally distributed, with covariance matrix:

$$\begin{bmatrix} \sigma_1^2 + \sigma_u^2\lambda^2 & \sigma_u^2\lambda \\ \sigma_u^2\lambda & \sigma_2^2 + \sigma_u^2 \end{bmatrix},$$

where σ_1^2, σ_2^2, and σ_u^2 are the variances of η_{1i}, η_{2i}, and u_i. If $\sigma_1^2 = \sigma_2^2 = \sigma_u^2$, then the correlation between η_{1i}^{*} and η_{2i}^{*} is parameterized by λ as

$$\rho = (1/\sqrt{2}) * [\lambda/\sqrt{(1 + \lambda^2)}]. \tag{5.3}$$

The average effect of treatment on y_i in (5.1) is

$$E\{\ \Phi[(\beta_0 + \beta_1 + u_i\lambda)/\sigma_1] - \Phi[(\beta_0 + u_i\lambda)/\sigma_1]\ \}, \tag{5.4}$$

where $\Phi[\cdot]$ is the standard normal cumulative distribution function. This expression simplifies further because the assumption of bivariate normality leads to a closed form for the expectation. Using the convolution properties of the normal distribution (see, e.g., McFadden and Reid [1975]), expression (5.4) can be written

$$(5.5) \qquad \pi_1^* = \Phi[\ (\beta_0 + \beta_1)/\sqrt{(\sigma_1^2 + \sigma_u^2\lambda^2)}\ ] - \Phi[\ \beta_0/\sqrt{(\sigma_1^2 + \sigma_u^2\lambda^2)}\ ].$$

Note that ML estimation identifies the standardized coefficients, $\beta_0/\sqrt{(\sigma_1^2 + \sigma_u^2\lambda^2)}$ and $\beta_1/\sqrt{(\sigma_1^2 + \sigma_u^2\lambda^2)}$. Therefore, ML estimates of π_1^* can be computed by evaluating (5.5) at the coefficient estimates.

5.2 Finite Sample Behavior

This section compares ML and IV estimates of π_1^* in a small sampling experiment designed to mimic situations encountered in econometric applications. In each experimental design, the true treatment effect is set at 10 percent, which is in the range of the estimated effects of manpower training on employment rates in four social experiments (Ham and LaLonde 1990). The instruments are drawn from a discrete uniform distribution with 8 points of support in increments of 1. Angrist (1990) uses a discrete uniformly distributed instrument to estimated the effects of military service, and Angrist and Krueger (1991) use a discrete uniformly distributed instrument to estimate the monetary returns to education. The number of replications for each experiment is 500, and results are presented for samples of 400 and 800 observations.[7] These sample sizes are in a range commonly encountered in econometric evaluation research (e.g., LaLonde 1986). The resulting bias calculations should provide an upper bound for applications like the mortality example in Section 1, where the treatment effects are so small that the outcome equations are approximately linear.

The base design sets β_0 equal to zero, and β_1 to $\Phi^{-1}(.6) \approx 0.25$,

[7]The computations were made using LIMDEP on microcomputers. Maximum Likelihood estimates were computed using the DFP algorithm from starting values of zero in each replication.

so that the treatment effect consists of a movement from 0.5 to 0.6 . In equation (4.2), γ_0 is set to zero, γ_1 to 0.25, and the instrument ranges from -3.5 to 3.5 . The resulting first-stage equation generates variation in $E[s_i | Z_i]$ from $\Phi(.875) = .81$ to $\Phi(-.875) = .19$. The errors are all normally distributed. The variances σ_1^2, σ_2^2, and σ_u^2 equal 1/2, and λ is equal to 1, so that the compound error terms have unit variance and a correlation coefficient of 1/2. This base design represents a promising scenario for IV estimation: variation in both F and G is close to the median, so that F and G should approximately satisfy Assumptions 2(i) and 2(ii). Moreover, the instrument is highly correlated with the endogenous regressor, tracing out 60 percent of the distribution of η_{2i}^*.

Table 1 presents experimental results for the base design. Columns 1-7 report the mean, standard deviation (SD), root mean squared error (RMSE), mean absolute error (MAE), lower quartile (LQ), median (MD), and upper quartile (UQ) of the estimates from 500 replications. Rows of the table report statistics for maximum likelihood estimates (MLE), just-identified IV estimates using only Z_i and a constant as instruments, TSLS estimates using 8 dummies to indicate each value of Z_i as instruments (Dummy IV), TSLS estimates using 8 dummies in White's (1982) efficient estimator (Efficient IV), and Ordinary least Squares (OLS) estimates of a linear probability model. The MLE's are consistent and efficient for the bivariate probit model. The IV estimator is consistent under Assumptions A2(i) or A2(ii); Dummy IV is consistent and asymptotically efficient under A2(i) if ν_i is homoscedastic; Efficient IV is consistent and asymptotically efficient under A2(i) or A2(ii).[8] The OLS estimates are

[8]Only one set of just-identified IV estimates are presented because instrumental variables estimates in just-identified models are unaffected by the choice of weighting matrix.

identical to estimates of the average treatment effect that would arise from single equation probit estimation using s_i as a regressor.

The means of the IV and ML estimates are within sampling variance of each other in both the 400 and 800 observation samples. The Dummy IV and Efficient IV estimates are biased towards the OLS estimate, which is nearly 4 times larger than the true treatment effect.[9] The bias of the Dummy and Efficient IV estimates is considerably worse in the smaller sample. The contrast between the just-identified IV estimates and the Dummy or Efficient IV estimates illustrates the trade-off between increasing bias and increasing efficiency as the number of instruments increase. In both samples, the MLE's have the lowest RMSE. Other than the OLS estimates, the MLE's are also most efficient, although the various IV estimators have only modestly larger sampling variance. The Efficient IV estimates are always slightly more variable than the asymptotically less efficient Dummy IV estimates. The quartiles do not indicate a significantly larger number of extreme values for the IV estimates than for the MLE's.

The discussion in Section 4.1 suggests that IV estimators should perform more poorly when the treatment effect shifts the distribution of the latent index at a point farther from the median, and when the instruments shift the distribution of the first-stage latent index at a point farther from the median. Table 2 reports results from a design the same as the base design, except that the treatment effect consists of a movement from 0.85 to 0.95 . This constitutes a larger deviation from A2(i) than in the base design, although A2(ii) is still approximately satisfied. All the estimates in Table 2 tend to be somewhat lower than the

---

[9] In a bivariate example, Nelson and Startz (1990) show analytically that the finite-sample central tendency of consistent instrumental variables estimates is biased towards the probability limit of OLS estimates.

true effect, whereas in the base design they are higher. As expected, the IV estimates are farther from the true effect than the MLE's, but the Dummy and Efficient IV estimates are closer. This may be because a positive small sample bias in these estimates offsets negative asymptotic bias.

Table 3 reports results from a design the same as the base design, except that the eight values taken by the instrument shift $E[s_i | Z_i]$ in the lower tail of the latent index distribution, from 0.05 to 0.35 .[10] This constitutes a stronger violation of A2(ii), leaving A2(i) approximately satisfied. All the estimates are now more variable, and the Dummy and Efficient IV estimates are substantially upwards biased in the 400 observation sample. But in the larger sample, in spite of a larger bias, the Dummy and Efficient IV estimates have lower RMSE than the ML estimate. Overall, the evidence from Tables 1, 2 and 3 suggests that IV estimators do not perform appreciably worse than the correctly specified ML estimator, especially in large samples.

Table 4 reports results from a design that combines the upper tail treatment effect of Table 2 with the lower tail first stage of Table 3. The combined violation of assumptions 2(i) and 2(ii) leads to a considerable deterioration in the finite sample performance of all the IV estimators. In fact, OLS now has lower RMSE than any of the IV estimates. However, the MLE performs equally badly, with a mean that tends to be half the size of the true treatment effect. Moreover, roughly half of the MLE replications failed to converge from starting values of zero. MLE's for the convergent subsample (in parentheses) also have a larger MAE than the IV estimates.

Tables 5 and 6 report the results of sampling experiments in which

---

[10] The instruments range from -3.5 to 3.5, $\gamma_0 = -1.02$, and $\gamma_1 = 0.181$

bivariate probit and IV techniques are used to estimate average treatment effects in models where the underlying error distributions are non-normal. Table 5 presents results from the base design where the underlying error distributions are $\chi^2(1)$ instead of $N(0,\sqrt{.5})$. The compound error in this case is distributed as $\chi^2(2)$, because it is the sum of two $\chi^2(1)$ random variables. Table 6 presents results from a modified base design where the treatment effect consists of a movement from 0.4 to 0.5, and the underlying error distributions are uniform on [0,1], so that the compound error terms are triangular on [0,2] with a mode of one.[11] The design used to produce Table 6 comes close to satisfying both A2(i) and A2(ii). The additive separability assumptions are not satisfied exactly in spite of the uniform distribution of η_{1i} and η_{2i} because the indices, $\beta_0 + \beta_1 - u_i\lambda$ and $\gamma_1 + \gamma_1 Z_i - u_i$, are not guaranteed to be between zero and one.

Results from the models with chi-square error distributions show the IV and MLE estimators with almost identical means. The MLE still has lower sampling variance, however, and lower mean squared error. Results from the models with uniform error distributions show the IV estimator closer to the true effect of 10 percent than the MLE. In the uniform designs, the median of the IV sampling distribution is also closer to the true effect than the median of the MLE sampling distribution. Again, however, larger sampling variance raises the mean squared error of the IV estimator above that of the MLE.

In both Tables 5 and 6, there is upward bias in the Dummy and Efficient IV estimates, although the bias of the Dummy and IV estimates

---

[11] As in Table 5, for the design reported in Table 6 the compound error terms are the sum of u_i and the η_i's. For Table 5, design parameters are: $\beta_0 = 1.3871$, $\beta_1 = 1.8331$, $\gamma_0 = 0$, $\gamma_1 = 3/8$; for Table 6 design parameters are: $\beta_0 = \sqrt{.8}$, $\beta_1 = 1 - \sqrt{.8}$, $\gamma_0 = 0.5$, $\gamma_1 = 0.1$.

falls with increasing sample size, while the bias of the MLE does not. In smaller samples, the simple IV estimator has lower bias than the MLE but is less efficient. The contrast between the results in Table 4 and the results in Tables 5 and 6 suggests that approximate additive separability is more important than distributional assumptions for the small sample performance of both ML and IV estimates. The MLE is remarkably robust to non-normality in both examples considered here, and remains efficient relative to all the IV estimators.[12] However, in some applications, (e.g., Angrist 1990, Angrist and Krueger 1990a, and the example in this paper) the required micro data are not available and the MLE cannot be computed. In such cases an IV estimator that can be computed from second moments offers an attractive and feasible alternative.

6. Summary and Conclusions

Even with valid exclusion restrictions such as generated by the draft lottery, the ability to answer evaluation questions in any field turns on functional form restrictions or distributional assumptions. This paper outlines functional form restrictions necessary and sufficient for linear IV techniques to provide consistent estimates of average treatment effects. The additive separability restrictions required for linear IV estimates to be consistent are unlikely to hold exactly, even for models with continuous outcome variables. But in many circumstances, the restrictions may hold approximately. This point is illustrated here using a Monte Carlo study of a bivariate probit model. In a number of examples, IV estimates of the

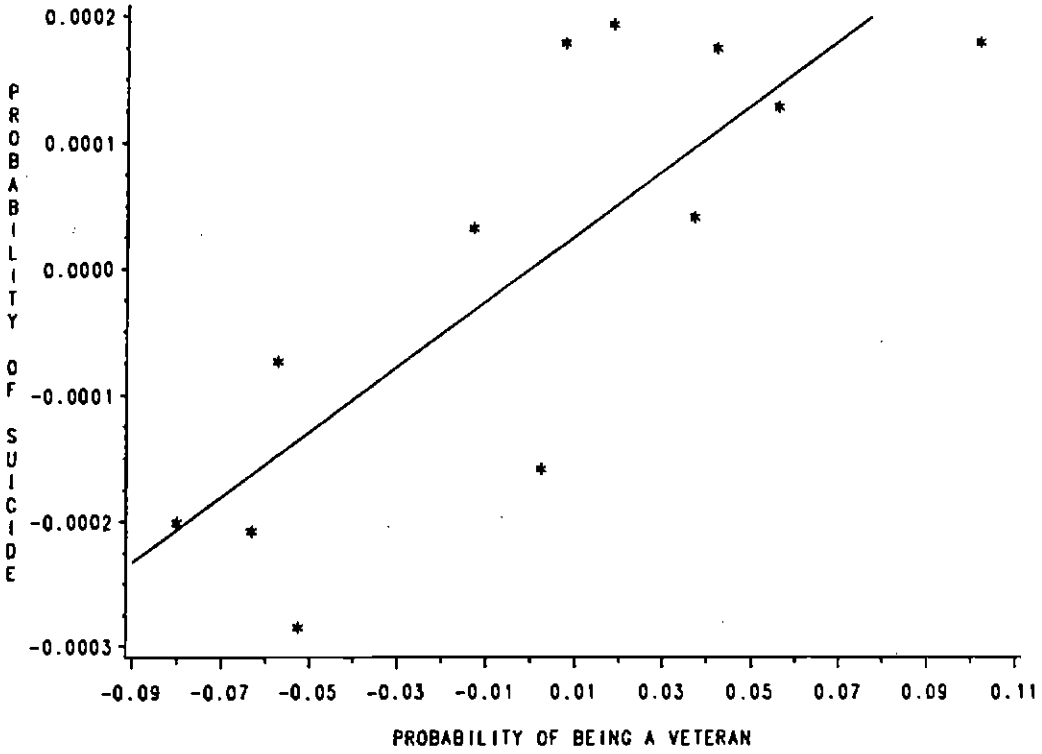
---

[12]The robustness of ML may be related to Ruud's (1986) result showing that for a large class of regressors, maximum likelihood with a misspecified distribution consistently estimates univariate index coefficients up to scale.

average treatment effect do not perform appreciably worse than estimates computed using the correct likelihood function.

These results suggest that linear IV estimation of average treatment effects in nonlinear models can often be justified. Of course, in some applications a variety of estimation strategies are available and all estimation strategies that derive from the same exclusion restrictions should probably be considered. Perhaps the most important reason for presenting IV estimates, however, even when more sophisticated techniques can be used, is that IV techniques are the observational investigator's version of classical experimentation. Instead of an experimenter randomly assigning treatment, the instrumental variables naturally assign treatment in a manner independent of other characteristics related to outcomes. I believe this source of identifying information is easily understood and communicated to non-specialists, and that findings from simple and comprehensible empirical strategies are most likely to affect public health and social policy.

SUICIDE AND VETERAN STATUS BY DRAFT-ELIGIBILITY



PLOT SHOWS RESIDUALS FROM A REGRESSION ON YEAR AND RACE EFFECTS
SOURCE: CA & PA ADMINISTRATIVE MORTALITY DATA, SIPP VETERAN STATUS DATA

.00258
(.00006)

Figure 1

Table 1: Base Design

Sample	Estimator	Mean (1)	SD (2)	RMSE (3)	MAE (4)	LQ (5)	MD (6)	UQ (7)
400	MLE	11.36	10.92	11.00	8.75	3.63	11.12	18.53
	IV	12.06	11.55	11.73	9.33	4.21	11.75	19.65
	Dummy IV	14.13	11.15	11.89	9.40	7.12	13.80	21.70
	Efficient IV	14.54	11.45	12.32	9.77	7.22	14.35	22.30
	OLS	38.04	4.69	28.43	28.04	34.73	38.19	41.35
800	MLE	9.59	7.71	7.72	6.15	4.20	9.75	14.70
	IV	10.38	8.35	8.36	6.68	4.66	11.20	15.80
	Dummy IV	11.66	8.08	8.25	6.66	6.28	12.10	16.70
	Efficient IV	11.84	8.21	8.41	6.80	6.32	12.40	17.10
	OLS	37.58	3.05	27.80	27.60	35.50	37.60	39.80

NOTES: The treatment effect consists of a movement from 0.5 to 0.6 . The instrument (Z_i) is discrete, uniformly distributed with 8 points of support in the range [-3.5, 3.5]. The first stage coefficients, γ_0 and γ_1, were chosen to generate variation in $E(s_i|Z_i)$ from 0.2 to 0.8 .

MLE estimates are maximum likelihood estimates for bivariate probit. IV estimates use Z_i and a constant as instruments. Dummy IV estimates use a full set of dummies for each value of Z_i as instruments. Efficient IV estimates uses a full set of dummies in the optimally weighted two-stage least squares estimator. The OLS estimate is the Ordinary Least Squares coefficient from a regression of y_i on s_i, and is the same as the single-equation probit estimate of the average treatment effect.

Table 2: Upper Tail Treatment Effect

Sample	Estimator	Mean (1)	SD (2)	RMSE (3)	MAE (4)	LQ (5)	MD (6)	UQ (7)
400	MLE	8.96 (9.12)	8.58 (8.75)	8.64 (8.79)	6.56 (6.66)	4.42 (4.26)	9.37 (9.89)	14.7 (14.9)
	IV	8.88	8.16	8.24	6.47	3.38	9.07	14.1
	Dummy IV	9.89	7.84	7.84	6.17	4.62	10.30	14.8
	Efficient IV	9.81	8.41	8.41	6.70	4.24	10.20	15.4
	OLS	21.3	3.25	11.80	11.30	19.10	21.20	23.6
800	MLE	9.71	5.26	5.27	4.02	6.86	10.20	13.1
	IV	8.94	5.30	5.40	4.23	5.61	9.10	12.6
	Dummy IV	9.49	5.22	5.24	4.12	6.26	9.49	13.0
	Efficient IV	9.70	5.45	5.46	4.29	6.38	9.80	13.4
	OLS	21.3	2.31	11.50	11.30	19.60	21.40	22.8

NOTES: The treatment effect consists of a movement from 0.85 to 0.95. Other design features are as described in Table 1.

In the sample of 400 observations, 23 out of 500 maximum likelihood replications failed to converge. Reported statistics are evaluated for all observations, including the last completed iteration for non-convergent estimates. Results for the convergent subsample are reported in parentheses.

Table 3: Lower Tail First Stage

Sample	Estimator	Mean (1)	SD (2)	RMSE (3)	MAE (4)	LQ (5)	MD (6)	UQ (7)
400	MLE	11.76 (11.76)	21.50 (21.60)	21.6 (21.7)	17.7 (17.7)	-3.50 (-3.50)	10.6 (10.6)	27.2 (27.2)
	IV	9.50	26.60	26.6	20.1	-6.14	10.8	24.5
	Dummy IV	15.19	22.50	23.1	17.8	2.00	15.5	29.4
	Efficient IV	15.42	22.90	23.5	18.1	1.98	15.9	29.2
	OLS	40.82	5.18	31.3	30.8	37.40	40.9	44.6
800	MLE	10.40	17.40	17.4	14.2	-1.25	11.7	22.5
	IV	9.48	18.10	18.1	14.4	-1.97	10.5	21.0
	Dummy IV	12.70	16.80	17.0	13.6	1.60	13.3	23.5
	Efficient IV	12.80	16.90	17.1	13.7	1.60	13.6	23.9
	OLS	40.60	3.64	30.8	30.6	38.00	40.8	43.4

NOTES: The treatment effect consists of a movement from 0.5 to 0.6 . The instrument (Z_i) is discrete, uniformly distributed with 8 points of support in the range [-3.5, 3.5]. The first stage coefficients, γ_0 and γ_1, were chosen to generate variation in $E(s_i|Z_i)$ from 0.05 to 0.35. Other features of the design are as in Table 1.

In the sample of 400 observations, one out of 500 maximum likelihood replications failed to converge. Reported statistics are evaluated for all observations, including the last completed iteration for non-convergent estimates. Results for the convergent subsample are reported in parentheses.

Table 4: Upper Tail Treatment Effect and Lower Tail First Stage

Sample	Estimator	Mean (1)	SD (2)	RMSE (3)	MAE (4)	LQ (5)	MD (6)	UQ (7)
400	MLE	4.87 (5.42)	13.7 (20.9)	14.6 (21.4)	9.39 (14.4)	3.90 (2.69)	4.66 (13.6)	9.15 (19.4)
	IV	1.43	19.2	21.0	16.3	-10.5	2.01	13.7
	Dummy IV	4.72	16.8	17.6	13.6	-5.00	5.52	15.7
	Efficient IV	4.31	17.3	18.2	14.3	-6.88	5.26	16.2
	OLS	17.0	2.30	7.34	6.98	15.4	16.9	18.7
800	MLE	3.97 (3.55)	13.8 (18.9)	15.1 (20.0)	9.60 (13.2)	3.96 (0.33)	4.62 (10.93)	12.1 (16.4)
	IV	3.67	13.5	14.9	11.7	-4.51	3.62	12.0
	Dummy IV	5.38	12.3	13.1	10.4	-2.54	5.15	13.1
	Efficient IV	4.76	12.8	13.8	10.9	-3.01	4.73	13.3
	OLS	17.0	1.53	7.17	6.97	15.9	17.0	18.1

NOTES: The treatment effect consists of a movement from 0.85 to 0.95 . The instrument (Z_i) is discrete, uniformly distributed with 8 points of support in the range [-3.5, 3.5]. The first stage coefficients, γ_0 and γ_1, were chosen to generate variation in $E(s_i|Z_i)$ from 0.05 to 0.35. This design combines features of the designs in Tables 2 and 3.

In the samples with 400 observations, only 216 out of 500 maximum likelihood replications converged. In the sample with 800 observations, only 266 out of 500 maximum likelihood replications converged. Reported statistics are evaluated for all observations, including the last completed iteration for non-convergent estimates. Results for the convergent subsample are reported in parentheses.

Table 5: Chi-Square Error Distributions

Sample	Estimator	Mean (1)	SD (2)	RMSE (3)	MAE (4)	LQ (5)	MD (6)	UQ (7)
400	MLE	12.8	11.5	11.8	9.44	4.62	12.5	20.1
	IV	12.7	12.4	12.7	10.3	4.43	13.1	20.8
	Dummy IV	15.0	11.8	12.8	10.4	6.43	14.9	23.5
	Efficient IV	15.6	12.3	13.5	10.9	6.71	15.5	24.2
	OLS	39.3	4.82	29.7	29.3	36.1	39.5	42.5
800	MLE	12.6	7.99	8.41	6.62	7.42	12.2	17.6
	IV	12.9	8.60	9.08	7.21	7.28	12.9	18.6
	Dummy IV	14.1	8.33	9.28	7.46	8.35	14.3	19.6
	Efficient IV	14.4	8.47	9.54	7.65	8.55	14.5	20.1
	OLS	39.4	3.22	29.6	29.4	37.0	39.40	41.5

NOTES: The treatment effect consists of a movement from 0.5 to 0.6. The instrument (Z_1) is discrete, uniformly distributed with 8 points of support in the range [$\frac{1}{8}$, 8]. All error distributions are $\chi^2(1)$, so that the compound errors, η_{1i} and η_{2i}, are $\chi^2(2)$. The first stage coefficients, γ_0 and γ_1, generate variation in $E[s_1|Z_1]$ from 0.17 to 0.78.

Table 6: Uniform/Triangular Error Distributions

Sample	Estimator	Mean (1)	SD (2)	RMSE (3)	MAE (4)	LQ (5)	MD (6)	UQ (7)
400	MLE	8.97	11.4	11.4	9.32	4.59	7.96	16.4
	IV	9.55	12.4	12.4	10.1	5.60	9.14	18.0
	Dummy IV	12.2	11.5	11.7	9.47	4.39	12.6	20.0
	Efficient IV	12.6	11.9	12.2	9.80	4.26	12.8	20.4
	OLS	38.4	4.37	28.7	28.4	35.2	38.6	41.5
800	MLE	8.91	8.18	8.25	6.64	3.07	8.69	14.3
	IV	9.88	8.99	8.99	7.18	3.85	10.2	15.6
	Dummy IV	11.3	8.78	8.88	7.09	5.56	11.3	17.0
	Efficient IV	11.5	8.91	9.04	7.22	5.54	11.5	17.1
	OLS	38.9	3.17	29.1	28.9	36.8	38.9	40.9

NOTES: The treatment effect consists of a movement from 0.4 to 0.5. The instrument (Z_1) is discrete, uniformly distributed with 8 points of support in the range [$\frac{1}{8}$, 8]. All error distributions are UN[0,1] so that the compound errors, η_{1i} and η_{2i}, are Triangular on [0,2], with a mode of 1. The first stage coefficients, γ_0 and γ_1, generate variation in $E[s_1|Z_1]$ from 0.18 to 0.76.

Appendix

Mortality and Veteran Status by Race, Year of Birth, and Draft-Eligibility

Race	Year	Draft-Eligibility[a]	Number of Deaths (Suicides)[b] (1)	Probability of Death (Suicide)[c] (2)	Probability of Military Service[d] (3)
White	1950	yes	2601 (436)	.0204 (.0034)	.3527
		no	2169 (352)	.0195 (.0032)	.1933
	1951	yes	1494 (279)	.0170 (.0032)	.2831
		no	2823 (480)	.0168 (.0029)	.1468
	1952	yes	1079 (207)	.0154 (.0029)	.2310
		no	2978 (514)	.0149 (.0026)	.1257
Non-White	1950	yes	536 (60)	.0346 (.0039)	.1957
		no	493 (46)	.0365 (.0034)	.1354
	1951	yes	350 (33)	.0376 (.0035)	.2014
		no	612 (63)	.0344 (.0035)	.1514
	1952	yes	235 (26)	.0309 (.0034)	.1449
		no	663 (66)	.0309 (.0031)	.1287

[a] Determined by lottery number (RSN) cutoff: RSN 195 for men born in 1950, RSN 125 for men born in 1951, RSN 95 for men born in 1952.

[b] From California and Pennsylvania administrative records, all deaths 1974-1983. Data sources and methods documented in Hearst, Newman and Hulley (1986). NOTE: Sample sizes differ from Hearst, et al., because non-US-born are included.

[c] Equal to number of deaths divided by Population At Risk (PAR) estimated from the 1970 census, 1% public use sample State files. PAR is the number of men in each birth cohort and race group. Estimates of PAR by draft-eligibility are computed assuming a uniform distribution of lottery numbers.

[d] Relative frequencies estimated from the Survey of Income and Program Participation. Data sources and methods are documented in Angrist (1990).

[e] Suicide is ICD 950-959.9. Total suicides = 2268 whites, 294 non-whites.

References

Angrist, J.D. (1990). "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," American Economic Review, June 1990.

_____ (1991), "Grouped Data Estimation and Testing in Simple Labor Supply Models," Journal of Econometrics 47, 243-266.

_____, and Alan Krueger (1989), "Why Do World War Two Veterans Earn More than Nonveterans?," NBER Working Paper.

_____, and Alan Krueger (1990a), "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," Harvard Institute of Economic Research Discussion paper No. 1475, March.

_____, and _____ (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?," Quarterly Journal of Economics, November.

Amemiya, T. (1974), "Bivariate Probit Analysis: Minimum Chi-Square Methods," Journal of the American Statistical Association 69: 940-944.

_____ (1978), "The Estimation of a Simultaneous Equation Generalized Probit Model," Econometrica 46: 1193-205.

Ashford, J.R., and R.R. Snowden (1970), "Multivariate Probit Analysis," Biometrics 26: 535-646.

_____ (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," Journal of Econometrics 34: 305-334.

Ham, J.C. and R.J. LaLonde (1990), "Using Social Experiments to Estimate the Effect of Training on Transition Rates," in Panel Data and Labor Market Studies, edited by J. Hartog, G. Ridder, and J. Theeuwes, Amsterdam: North-Holland.

Hearst, N., T. Newman and S.B. Hulley (1986), "Delayed Effects of the Military Draft on Mortality: A Randomized Natural Experiment," New England Journal of Medicine 314 (March 6, 1986): 620-624.

Hearst, N., J. W. Buehler, T. Newman, and G.W. Rutherford (1990), "The Draft Lottery and AIDS: Evidence Against Increased Intravenous Drug Use by Vietnam-era Veterans," Mimeo, Department of Epidemiology and Biostatistics, University of California, San Francisco, March.

Heckman, J.J. (1978), "Dummy Endogenous Variables in a Simultaneous Equations System," Econometrica 46(6): 931-959.

_____ (1990), "Varieties of Selection Bias," American Economic Review 80(2) : 313-318.

_____, and Richard Robb (1985), "Alternative Methods for Evaluating the Impact of Interventions," in Heckman and Singer, eds., Longitudinal Analysis of Labor Market Data, New York: Cambridge University Press.

Kelejian, H. (1971), "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables," Journal of the American Statistical Association 66: 373-374.

Lalonde, R.J. (1986), " Evaluating the Econometric Evaluations of Training Programs Using Experimental Data," American Economic Review 76(4): 602-620.

Maddala, G.S. (1983), Limited Dependent and Qualitative Variables in Econometrics, New York: Cambridge University Press.

Manski, C. (1975), "Maximum Score Estimation of the Stochastic Utility Model of Choice," Journal of Econometrics 3: 205-228.

McFadden, D., and F. Reid (1975), "Aggregate Travel Demand Forecasting from Disaggregated Behavioral Models," Transportation Research Board Record No. 534: 24-37.

Nelson, C.R., and R. Startz (1990), "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," Econometrica 58(4): 967-976.

Newey, Whitney K. (1985a), "Semiparametric Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables," Annales De L'INSEE 59/60: 219-235.

_____ (1985b), "Generalized Method of Moments Estimation and Testing," Journal of Econometrics 29: 229-256.

_____ (1986), "Linear Instrumental Variables Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables," Journal of Econometrics 32: 127-141.

_____ (1987), "Efficient Estimation of Limited Dependent Variables with Endogenous Explanatory Variables," Journal of Econometrics 36: 231-250.

_____ (1989), "Efficient Instrumental Variables Estimation of Nonlinear Models," Princeton University Econometric Research Program Memorandum No. 341, March 1989.

Permutt, Thomas and J. Richard Hebel (1989), "Simultaneous-Equation Estimation in a Clinical Trial of the Effect of Smoking on Birth Weight," Biometrics 45 (June): 619-622.

Robins, James M. (1989), "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," Health Service Research Methodology: A Focus on AIDS, edited by L. Sechrest, H. Freeman, and A. Bailey, NCHSR, U.S. Public Health Service.

Rosenbaum, P.R. and D.B. Rubin (1983), "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome," Journal of the Royal Statistical Society Series B 45(2): 212-218.

Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," Journal of Educational Psychology 66(5): 688-701.

Ruud, Paul (1986), "Consistent Estimation of Limited Dependent Variables Models Despite Misspecification of Distribution," Journal of Econometrics 32: 157-187.

Seltzer, C.C. and S. Jablon (1974), "Effects of Selection on Mortality," American Journal of Epidemiology 100: 367-372.

Stoker, T.M. (1986), "Consistent Estimation of Scaled Coefficients," Econometrica 54(6): 1461-1481.

White, H. (1982), "Instrumental Variables Regression with Independent Observations," Econometrica 50: 482-499.