



Identification of Causal Effects Using Instrumental Variables: Comment

Paul R. Rosenbaum

Journal of the American Statistical Association, Vol. 91, No. 434. (Jun., 1996), pp. 465-468.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28199606%2991%3A434%3C465%3AIOCEUI%3E2.0.CO%3B2-H>

Journal of the American Statistical Association is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

Case of Manpower Training," by J. Heckman and V. J. Hotz, *Journal of the American Statistical Association*, 84, 877-878.

——— (1996), "Selection Bias Adjustment in Treatment-Effect Models as a Method of Aggregation," in *1995 Proceedings of the American Statistical Association*.

Rosenbaum, P., and Rubin, D. (1983), "The Central Role of the Propensity

Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

Roy, A. D. (1951), "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135-146.

Staiger, D., and Stock, J. (1993), "Instrumental Variable Regression With Weak Instruments," mimeo, Harvard University.

Comment

Paul R. ROSENBAUM

1. INTRODUCTION

Angrist, Imbens, and Rubin (AIR) deserve congratulations for a wonderful paper. Linking econometrics with experimental design, they have illuminated both fields. I particularly admire the care they take in defining estimands with few modeling assumptions, in stating assumptions in tangible terms, and in examining the appropriateness and consequences of those assumptions.

In this comment, I would like to slightly restate their argument in terms of an artificial example, then generalize the argument to a larger class of estimators (the Hodges-Lehmann estimators), briefly indicate how one can conduct a sensitivity analysis in a nonrandomized study, and conclude with an observation about the case in which some subjects have unalterable treatment assignments.

2. AN ARTIFICIAL EXAMPLE: ENCOURAGING EXERCISE FOR LUNG DISEASE

The following artificial example illustrates and restates several of the points made by AIR, but its main purpose is to aid in Section 3 in discussing a generalization of the instrumental variables (IV) estimate. Table 1 describes a randomized experiment with 10 subjects suffering from chronic obstructive lung disease (COLD), of whom 5 were randomly selected and encouraged to exercise. So the randomization determines who was encouraged to exercise; that is, Z_i . In fact, not all subjects complied, as indicated by $D_i(Z_i)$. Subjects 1, 2, and 3 exercised as they were encouraged to do, but subjects 4 and 5 ignored the encouragement and did not exercise. Subject 6 was not encouraged to exercise ($Z_6 = 0$) but did so anyway ($D_6(0) = 1$). The outcome is forced expiratory volume (FEV), a measure of lung function, larger values indicating better health, recorded on a convenient integer scale. The quantity $Y_i(0)$ is the outcome that would have been observed from subject i in the absence of exercise. As in AIR's exclusion restriction and in Holland's (1988) discussion of encouragement designs, it is exercise that may have an effect, but encouragement has an effect only if it influences exercise. In Table 1 exercise raises the outcome, FEV, by 3 units for all subjects; that is, the ob-

served response from subject i is $\tilde{Y}_i = Y_i(0) + 3D_i(Z_i)$ (e.g., $\tilde{Y}_1 = Y_1(0) + 3D_1(1) = 5 + 3 = 8$).

Several features of Table 1 are of note. First, the FEV response that would be observed from subject i in the absence of exercise, $Y_i(0)$, is unaffected by encouragement Z_i or exercise $D_i(Z_i)$, and in the theory of randomized experiments, $Y_i(0)$ is a fixed feature of subject i not varying with the random assignment of encouragement Z_i . By good fortune in this artificial example, the distribution of $Y_i(0)$ is perfectly balanced in encouraged ($Z_i = 1$) and control ($Z_i = 0$) groups. Randomization produces such balance in expectation in randomized experiments of all sizes, and in large experiments approximate balance is likely, but the exact balance in Table 1 is an unnecessary but tidy convenience useful in exposition. In short, the randomization worked—without treatment, the two randomized groups ($Z_i = 1$) and ($Z_i = 0$) would have had similar outcomes. The observed responses \tilde{Y}_i are not balanced of course, because encouragement Z_i increases exercise $D_i(Z_i)$, which increases \tilde{Y}_i . Notice also that healthy subjects are more likely to exercise. More precisely, subjects who would have had high FEV absent exercise—subjects with high $Y_i(0)$ —are more likely to have $D_i(Z_i) = 1$. Encouragement appears to increase the amount of exercise, but subjects with low $Y_i(0)$ do not exercise even if encouraged.

The traditional advice in randomized clinical trials is that the groups formed by randomization should be compared; here, the encouraged ($Z_i = 1$) and control ($Z_i = 0$) groups. The difference in means is $(8 + 7 + 6 + 2 + 1)/5 - (8 + 4 + 3 + 2 + 1)/5 = 6/5 = 1.2$. This is a sensible estimate of the effect of encouragement. Exercise raises FEV by 3, but encouragement raises it only by 1.2 on average, because many subjects do not exercise when encouraged and some exercise without encouragement. A mistaken estimate of the effect of exercise compares those who exercised ($D_i(Z_i) = 1$) to those who did not ($D_i(Z_i) = 0$)—namely, $(8 + 7 + 6 + 8)/4 - (2 + 1 + 4 + 3 + 2 + 1)/6 = 7.250 - 2.167 = 5.083$. This estimate grossly overstates the effect of exercise because healthier subjects were more likely to exercise. The instrumental variables estimate starts with

Table 1. An Experiment Encouraging Exercise for Lung Disease

Subject	Exercise encouraged Z_i	Exercise performed $D_i(Z_i)$	FEV response without exercise $Y_i(0)$	Observed FEV response \tilde{Y}_i
1	1	1	5	8
2	1	1	4	7
3	1	1	3	6
4	1	0	2	2
5	1	0	1	1
6	0	1	5	8
7	0	0	4	4
8	0	0	3	3
9	0	0	2	2
10	0	0	1	1

the 1.2 determined previously by comparing the encouraged and control groups, but it attributes the entire 1.2 to the increase in exercise in the encouraged group; that is, it divides 1.2 by the mean difference in exercise, $(1 + 1 + 1 + 0 + 0)/5 - (1 + 0 + 0 + 0 + 0)/5 = 2/5$, so the estimate is $1.2/(2/5) = 3$.

The point made later is that this argument has nothing to do with means and quickly extends to sturdier estimates, such as the Hodges–Lehmann (HL) (Hodges and Lehmann 1963). Also, with an instrumental variable, exact permutation inferences about an additive effect are obtained using the random assignment of encouragement. Moreover, sensitivity analysis is straightforward in nonrandomized or observational studies. (See Hollander and Wolfe 1973 or Lehmann 1975 for discussion of the standard forms of the HL estimate, and see Maritz 1995, secs. 1 and 8.1 for discussion of the broad scope of HL estimates with various technical results.)

3. THE HODGES–LEHMANN ESTIMATE USING AN INSTRUMENTAL VARIABLE IN A RANDOMIZED EXPERIMENT

Following AIR, assume the stable unit treatment value assumption (SUTVA), the exclusion restriction, and the nonzero average effect of Z on D . For subject i , write \tilde{D}_i for the observed level of exercise, $\tilde{D}_i = Z_i \cdot D_i(1) + (1 - Z_i)D_i(0)$, and write \tilde{Y}_i for the observed outcome, $\tilde{Y}_i = \tilde{D}_i \cdot Y_i(1) + (1 - \tilde{D}_i)Y_i(0)$. The model of an additive effect asserts that $Y_i(1) - Y_i(0) = \tau$ for all i , so $\tilde{Y}_i = Y_i(0) + \tau\tilde{D}_i$, as in Table 1 where $\tau = 3$. In Section 5, it will be seen that the additive model need not hold for all i , that it suffices that additivity holds for subjects who change treatments in response to encouragement, but it is easier to discuss this separately. Write $\mathbf{Z}, \tilde{\mathbf{D}}, \tilde{\mathbf{Y}}$, and \mathbf{Y}_0 for the N -dimensional vectors of Z_i 's, \tilde{D}_i 's, \tilde{Y}_i 's, and $Y_i(0)$'s. Write M for the number of encouraged subjects, $M = \mathbf{Z}^T \mathbf{Z}$. In Table 1, $M = 5$. Write B for the set containing the possible treatment assignments, so B contains $\binom{N}{M}$ vectors of dimension N with M coordinates equal to 1 and $N - M$ coordinates equal to zero. In a randomized experiment, \mathbf{Z} is picked from B at random; that is, $\text{prob}(\mathbf{Z} = \mathbf{z}) = \binom{N}{M}^{-1}$ for each $\mathbf{z} \in B$.

Let $t(\mathbf{Z}, \tilde{\mathbf{Y}})$ be a statistic used to compare the encouraged and control groups. For instance, $t(\mathbf{Z}, \tilde{\mathbf{Y}})$ might be the difference in sample means, say $t_M(\mathbf{Z}, \tilde{\mathbf{Y}}) = \mathbf{Z}^T \tilde{\mathbf{Y}} / M - (1 - \mathbf{Z})^T \tilde{\mathbf{Y}} / (N - M)$, or Wilcoxon's rank sum statis-

tic, $t_W(\mathbf{Z}, \tilde{\mathbf{Y}}) = \mathbf{Z}^T \text{rank}(\tilde{\mathbf{Y}})$, where $\text{rank}(\tilde{\mathbf{Y}})$ is the N -dimensional vector of ranks of the $\tilde{\mathbf{Y}}$ with average ranks used for ties, or the difference between the trimean in the encouraged group and the trimean in the control group, say $t_T(\mathbf{Z}, \tilde{\mathbf{Y}})$. (Recall that the trimean is the sum of the upper and lower quartiles plus twice the median divided by four.)

Let \bar{t} be the expectation of $t(\mathbf{Z}, \mathbf{Y}_0)$ over the randomization distribution of \mathbf{Z} ; that is, the average of $t(\mathbf{z}, \mathbf{Y}_0)$ over the $\binom{N}{M}$ choices $\mathbf{z} \in B$. For the difference in means, $t_M(\mathbf{Z}, \mathbf{Y}_0)$ has expectation $\bar{t}_M = 0$. For the rank sum, $t_W(\mathbf{Z}, \mathbf{Y}_0)$ has expectation $\bar{t}_W = M(N + 1)/2$. For the difference of trimeans, $t_T(\mathbf{Z}, \tilde{\mathbf{Y}})$ has expectation $\bar{t}_T = 0$ if $M = N/2$ and $\bar{t}_T \rightarrow 0$ as $M, N - M \rightarrow \infty$ whether or not $M = N/2$.

Now, $\mathbf{Y}_0 = \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\tau$. The HL estimate using an IV is defined to be the value $\hat{\tau}$ such that $t(\mathbf{Z}, \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\hat{\tau})$ is as close as possible to \bar{t} . For some estimators such as the difference in means, $\hat{\tau}$ may be determined by solving $t(\mathbf{Z}, \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\hat{\tau}) = \bar{t}$. For estimates based on rank statistics that move in discrete steps, there may be no $\hat{\tau}$ such that $t(\mathbf{Z}, \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\hat{\tau}) = \bar{t}$ exactly; then, following HL, $\hat{\tau}$ is defined as the average of the smallest value that is too large and the largest value that is too small, namely

$$\hat{\tau} = \frac{\sup\{\tau: t(\mathbf{Z}, \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\tau) > \bar{t}\} + \inf\{\tau: t(\mathbf{Z}, \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\tau) < \bar{t}\}}{2}$$

Here $\hat{\tau} = \infty$ if there is no finite τ such that $t(\mathbf{Z}, \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\tau) < \bar{t}$ and $\hat{\tau} = -\infty$ if there is no finite $\hat{\tau}$ such that $t(\mathbf{Z}, \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\hat{\tau}) > \bar{t}$.

Write $\hat{\tau}_M, \hat{\tau}_W$, and $\hat{\tau}_T$ for the instrumental HL estimates based on t_M, t_W , and t_T . For the difference in means, simple algebra shows $t_M(\mathbf{Z}, \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\hat{\tau}) = \bar{t} = 0$ if and only if $\hat{\tau}_M$ is the usual instrumental variable estimator discussed by AIR. If encouragement always determines the treatment so $\tilde{\mathbf{D}} = \mathbf{Z}$, then $\hat{\tau}_M$ is the encouraged minus-control difference in sample means, $\hat{\tau}_W$ is the usual HL estimate associated with the rank sum statistic, and $\hat{\tau}_T$ is the difference in trimeans. In short, the estimate $\hat{\tau}$ generalizes both the usual IV estimate and the usual HL estimate.

In the example in Table 1, subtracting 3 from each subject who exercised sets all three statistics, t_M, t_W , and t_T , equal to their null expectations, so $\hat{\tau}_M = \hat{\tau}_W = \hat{\tau}_T = 3$ in this particular case. This is exceptional and reflects the perfect balance of the $Y_i(0)$'s in this constructed example. If \tilde{Y}_1 in Table 1 were replaced by an extremely large positive value,

then $\hat{\tau}_M$ would increase dramatically, $\hat{\tau}_W$ would increase slightly, and $\hat{\tau}_T$ would continue to equal 3.

Exact inference about τ may be based on the randomization distribution of $t(\mathbf{Z}, \mathbf{Y}_0)$ where \mathbf{Y}_0 is fixed. Consider first testing the null hypothesis that $H_0: \tau = \tau^*$. Under the null hypothesis, the fixed responses in the absence of exercise, \mathbf{Y}_0 , are equal to $\tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\tau^*$, which may be computed from the data. The one-sided randomization significance level is the proportion of treatment assignments $\mathbf{z} \in B$ giving a larger value of the test statistic than observed, $|\{\mathbf{z} \in B: t(\mathbf{z}, \mathbf{Y}_0) \geq t(\mathbf{Z}, \mathbf{Y}_0)\}| / \binom{N}{M}$, where $|A|$ denotes the number of elements of the set A . A one-sided confidence interval for τ is obtained by determining all values of τ not rejected by such a test. A two-sided 95% confidence interval is the intersection of two one-sided 97.5% intervals. It is notable that the test of $H_0: \tau = 0$ is identical to the usual randomization test of no effect, as discussed by Fisher (1935) or Kempthorne (1952), but this is not true for $\tau^* \neq 0$.

For instance, in Table 2, to test the false hypothesis $H_0: \tau = 1.5$ using the rank sum test with an instrumental variable, one computes $\tilde{\mathbf{Y}} - 1.5\tilde{\mathbf{D}} = (6.5, 5.5, 4.5, 2, 1, 6.5, 4, 3, 2, 1)$ with ranks $(9.5, 8, 7, 3.5, 1.5, 9.5, 6, 5, 3.5, 1.5)$, so the rank sum is $9.5 + 8 + 7 + 3.5 + 1.5 = 29.5$. Allowing for the ties, the null expectation and variance of the rank sum are 27.5 and 22.5 yielding a standardized deviate of $(29.5 - 27.5) / \sqrt{22.5} = .42$, so the hypothesis $H_0: \tau = 1.5$ is not rejected. Without ties, the familiar exact distribution of the rank sum statistic may be used.

In short, permutation tests, confidence intervals, and HL estimates all use the null distribution of $t(\mathbf{Z}, \mathbf{Y}_0)$, which is the usual randomization distribution (Fisher 1935; Kempthorne 1952, sec. 8.2). If encouragement itself \mathbf{Z} had an additive effect, then one would have $\mathbf{Y}_0 = \tilde{\mathbf{Y}} - \mathbf{Z}\tau$, and the usual procedures for permutation inference would result. What is new with the IV is that exercise $\tilde{\mathbf{D}}$ and not encouragement \mathbf{Z} has the additive effect, so the permutation inference is based on $\mathbf{Y}_0 = \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\tau$, but otherwise permutation methods are unchanged. As it turns out, these considerations extend immediately for sensitivity analysis in observational studies.

4. SENSITIVITY ANALYSIS USING INSTRUMENTAL VARIABLES IN OBSERVATIONAL STUDIES

In the experiment in Table 1, random assignment of encouragement tended to balance the distribution of $Y_i(0)$ in encouraged ($Z_i = 1$) and control ($Z_i = 0$) groups. In an observational study or nonrandomized experiment, subjects might have differing chances of receiving encouragement to exercise; that is, it may be quite wrong to assume that $\text{prob}(\mathbf{Z} = \mathbf{z}) = \binom{N}{M}^{-1}$ for each $\mathbf{z} \in B$. Perhaps the severely ill would be less likely to receive encouragement than the less severely ill.

It is possible to study the sensitivity of permutation inferences to departures from random assignment of treatments (see, for instance, Rosenbaum 1993, 1995). These techniques replace $\text{prob}(\mathbf{Z} = \mathbf{z}) = \binom{N}{M}^{-1}$ with a range of distri-

butions of treatment assignments, thereby obtaining a range of null distributions for $t(\mathbf{Z}, \mathbf{Y}_0)$. Using these techniques with an IV is straightforward; one calculates $\mathbf{Y}_0 = \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\tau$ as in Section 3 and applies the sensitivity analysis to the result. For instance, the sensitivity analysis gives not one null expectation $\bar{t}_W = M(N+1)/2$ for the rank sum statistic, but rather a range of expectations $[\bar{t}_{W,\text{low}}, \bar{t}_{W,\text{high}}]$ depending on a sensitivity parameter Γ . This yields a range of instrumental HL estimates obtained by approximately solving $t(\mathbf{Z}, \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\hat{\tau}) = \bar{t}_{W,\text{low}}$ and $t(\mathbf{Z}, \tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\hat{\tau}) = \bar{t}_{W,\text{high}}$.

5. AVOIDING SPECULATION ABOUT SUBJECTS WHO IGNORE ENCOURAGEMENT

AIR carefully focus attention on subjects who do not ignore encouragement; see, for instance, their proposition 1. They call subjects who ignore encouragement "always-takers" if $D_i(1) = D_i(0) = 1$ or "never-takers" if $D_i(1) = D_i(0) = 0$. They argue, in effect, that one can say little about subjects who ignore encouragement because nothing that the experimenter does will change the treatment they receive. This final section briefly observes that the argument in Section 3 continues to hold if nothing is assumed about subjects who ignore encouragement.

Following AIR, consider a randomized experiment and assume that Z is an instrumental variable in the sense of their definition 3, so $D_i(1) \geq D_i(0)$. In addition, assume that the treatment has an additive effect for compliers only; that is, $Y_i(1) - Y_i(0) = \tau$ whenever $D_i(1) = 1 > 0 = D_i(0)$. No assumption is made about $Y_i(1) - Y_i(0)$, when $D_i(1) = D_i(0)$. Let τ^* be a hypothesized value for τ . Then the adjusted responses are

$$\tilde{Y}_i - \tilde{D}_i\tau^* = \begin{cases} Y_i(1) - \tau^* & \text{if } D_i(1) = D_i(0) = 1, \\ Y_i(0) + (\tau - \tau^*)Z_i & \text{if } D_i(1) = 1 > 0 = D_i(0), \\ Y_i(0) & \text{if } D_i(1) = D_i(0) = 0. \end{cases}$$

Note first that the adjusted responses $\tilde{Y}_i - \tilde{D}_i\tau^*$ will be independent of encouragement Z_i if and only if the hypothesized τ^* equals the true τ . Also, if $\tau^* < \tau$, then the adjusted responses $\tilde{Y}_i - \tilde{D}_i\tau^*$ for encouraged subjects ($Z_i = 1$) will tend to be somewhat higher than those for control ($Z_i = 0$) subjects, and conversely if $\tau^* > \tau$. As a consequence, to render the adjusted responses independent of encouragement, one must have the correct τ^* , and a test statistic such as the rank sum statistic that is consistent when one distribution is stochastically larger than another will, in sufficiently large sample sizes, reject any fixed $\tau^* \neq \tau$, thereby yielding consistent tests, confidence intervals, and point estimates. In short, the procedures in Section 3 describe subjects who comply with no assumptions about those who ignore encouragement.

ADDITIONAL REFERENCES

- Hodges, J., and Lehmann, E. (1963), "Estimates of Location Based on Rank Tests," *Annals of Mathematical Statistics*, 34, 598-611.

- Holland, P. (1988), "Causal Inference, Path Analysis, and Recursive Structural Equation Models," in *Sociological Methodology*, Washington, DC: American Sociological Association.
- Hollander, M., and Wolfe, D. (1973), *Nonparametric Statistical Methods*, New York: John Wiley.
- Kemphorne, O. (1952), *Design and Analysis of Experiments*, New York: John Wiley.
- Lehmann, E. (1975), *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden Day.
- Maritz, J. S. (1995), *Distribution-Free Statistical Methods*, London: Chapman and Hall.
- Rosenbaum, P. R. (1993), "Hodges-Lehmann Point Estimates of Treatment Effect in Observational Studies," *Journal of the American Statistical Association*, 88, 1250-1253.
- (1995), *Observational Studies*, New York: Springer-Verlag.

Rejoinder

Joshua D. ANGRIST, Guido W. IMBENS, and Donald B. RUBIN

We thank Heckman, Greenland and Robins, Moffitt, and Rosenbaum for their stimulating comments on our paper. After making two general remarks, we address specific points in each comment.

Both Heckman and Greenland and Robins stress that LATE is the average causal effect for a subpopulation that cannot be identified in the sense that we cannot label all individual units in the population as compliers or noncompliers. Greenland and Robins suggest that attention should focus on the population average treatment effect, whereas Heckman is more interested in the average effect for those who receive treatment, also the estimand of interest in Peters (1941), Belson, (1951), Cochran (1969), and Rubin (1973a,b, 1977). For policy purposes, one may indeed be interested in averages for the entire population, or for specific subpopulations other than compliers. Within the context of a particular study with a specific instrument, however, the data are not directly informative about average effects for subpopulations other than compliers. A key insight from our work is that compliers are the *only* group with members observed taking the treatment and members observed not taking the treatment. Always-takers are always observed taking the treatment, so the data simply cannot be informative about average treatment effects for this group, and similarly for never-takers. In the same vein, a clinical trial restricted to young men is not informative about treatment effects for adult women. Yet Heckman and Greenland and Robins appear to criticize us precisely because we limit our discussion of causal effects to the only subpopulation about which the data are directly informative.

Following a core analysis focused on the directly estimable effect, one may wish to extend the conclusions to broader groups. Such extensions are routine in the interpretation of clinical trials, which are seldom based on representative samples of the overall target population. Our approach makes it clear, however, that in instrumental variables (IV) contexts, extensions to groups other than compliers can only be extrapolations.

The second issue raised by multiple discussants is the propriety of our example. Clearly, an example with a binary randomized instrument is not representative of economic applications of IV techniques where candidate in-

struments are rarely based on actual randomization. A major reason for using this example was to stress that randomization alone does not make a candidate instrument a valid one because randomization does not make the exclusion restriction more plausible. The fact that economists do not always make a clear distinction between ignorability and exclusion restrictions is evidenced by Moffitt's incorrect comment that randomization makes the draft lottery "by *necessity* an obvious and convincing instrument" (italics ours) for the effect of the military service. In fact, one contribution of our approach is to provide a framework that clearly separates ignorability and exclusion assumptions. Both statisticians and economists should find this separation useful and clarifying.

HECKMAN

Heckman begins by arguing that the RCM is a version of the widely used econometric switching regression model. We view the term Rubin causal model (coined by Holland [1986] for work by Rubin [1974, 1978]) as referring to a model for causal inference where causal effects are defined explicitly by comparing potential outcomes. This comparison can be in the context of a randomized experiment or an observational study. Any element of the set of the potential outcomes *could* have been observed by manipulation of the treatment of interest, even though ex-post only one of them is actually observed. Moreover, the RCM defines the assignment mechanism, which determines which potential outcomes are observed, as the conditional probability of each possible treatment assignment given the potential outcomes and possibly other variables. In contrast, the switching regression model as expounded by Quandt (1958, 1972) is a time series model where the first part of the sample comes from one regression model and the second part from a separate regression model with an unknown switching point.

A second example mentioned by Heckman is Roy (1951), who studied the distribution of observed incomes in a world