

Estimating Causal Effects With Matching Methods in the Presence and Absence of Bias Cancellation

THOMAS A. DIPRETE
Duke University

HENRIETTE ENGELHARDT
Austrian Academy of Sciences, Vienna

This article explores the implications of bias cancellation on the estimate of average treatment effects using ordinary least squares (OLS) and Rubin-style matching methods. Bias cancellation (offsetting biases at high and low propensities for treatment in estimates of treatment effects that are uncorrected for nonrandom selection) has been observed when job training is the treatment variable and earnings is the outcome variable. Contrary to published assertions in the literature, bias cancellation is not explainable in terms of the standard selection model, which assumes a symmetric distribution for the errors in the structural and assignment equations. A substantive rationale for bias cancellation is offered, which conceptualizes bias cancellation as the result of a mixture process based on two distinct individual-level decision-making models. While the general properties are unknown, the existence of bias cancellation appears to reduce the average bias in both OLS and matching methods relative to the symmetric distribution case.

Keywords: causal effects; treatment effect; matching methods; bias cancellation

INTRODUCTION

In recent years, economics and other social sciences have paid increased attention to the estimation of causal effects. In the experimental setup (where cases are randomly assigned to treatment and control groups), the difference in the average outcome for otherwise

AUTHORS' NOTE: A version of this article was presented at the August 2000 annual meeting of the American Sociological Association. This research was supported in part by the Max Planck Institute for Human Development, the Max Planck Institute for Demographic Research, Duke University, and the German Institute for Economic Research (DIW, Berlin). We would like to thank Norman Braun, Patricia A. McManus, and the anonymous reviewers for helpful comments on an earlier version.

SOCIOLOGICAL METHODS & RESEARCH, Vol. 32, No. 4, May 2004 501-528
DOI: 10.1177/0049124103260187
© 2004 Sage Publications

statistically identical treatment and control groups forms the basis for estimating the causal effect of the treatment in question. Econometricians and statisticians have long recognized the problems in estimating causal effects with observational data. However, a developing literature stimulated largely by the research of James Heckman and Donald Rubin has produced both new approaches to the problem and a deeper appreciation of the potential and the limitations of standard regression analysis, matching methods, instrumental variable techniques, econometric models of selection bias, or the method of “difference-in-differences” when applied to observational data (e.g., Heckman 1979, 1997; Rosenbaum and Rubin 1984; Rubin and Thomas 1996; Angrist, Imbens, and Rubin 1996; Heckman, Ichimura, and Todd 1997, 1998; Heckman, Ichimura, Smith, et al. 1998).

This article explores the implications for causal estimation that arise under two specific conditions. The first condition is when, in the words of Heckman and Robb (1985), the selection is based on unobservable variables such that methods of estimation that ignore this selection process (which might be termed *naive estimators*) give biased estimates of the causal effect in question. Just as causal effects will generally vary not only at the individual level but also at different points in the distribution of the observed covariates (here called the “X” distribution), the bias arising from any given estimation procedure can also vary across the X distribution. The second condition of interest in this article concerns situations in which these “pointwise” or “point-by-point” biases (by which we mean biases in naive estimates at different points in the X distribution) are offsetting, as observed by Heckman, Ichimura, Smith, et al. (1998) (i.e., they are positive for individuals located in some parts of the X distribution and negative for individuals located at other parts of the X distribution). Social scientists are typically concerned more with the *average* effect of treatment, either for the population or for the treated subpopulation, and with reducing the bias in estimates of average effects than they are with the point-by-point treatment estimates or the biases in these point-by-point treatment estimates. However, in situations with offsetting point-by-point biases, the bias in estimates of the average treatment effect will be reduced to the extent that the offsetting point-by-point biases cancel each other.

This article investigates this possibility in greater depth and explores its implications for estimating treatment effects. Building on the empirical results of Heckman, Ichimura, Smith, et al. (1998), which report evidence of “bias cancellation” in evaluation data for the effects of job training on earnings, we analyze the statistical basis for this phenomenon and show that this empirically observed pattern of bias is inconsistent with the assumptions that underlie the most common approach to the correction of sample selection bias in sociology. We then offer a theoretical rationale as to why the empirical pattern found by Heckman, Ichimura, Smith, et al. in the job-training data may not be an isolated instance. Instead, this empirical pattern may signal that selection bias in observational data often derives from a *mixture* of selection processes, in which individuals select themselves into different states of a “treatment variable” based on multiple logics that can produce differently signed and thus partially offsetting biases in the analysis of causal effects using standard methodologies. Finally, we investigate the empirical properties of matching estimators based on Rosenbaum and Rubin’s (1984) propensity score method in the situation in which bias cancellation does and does not occur. We conduct these empirical investigations using partially simulated data from the German Socioeconomic Panel, in which further training is the treatment variable and earnings is the outcome variable. In these simulations, we intentionally exclude all selection mechanisms based on observable variables to determine how and in what respects the bias cancellation property might allow matching methods to reduce selection bias due to unobservable variables.

BIAS IN THE ESTIMATION OF CAUSAL EFFECTS

Employing a counterfactual approach, Heckman et al. (Heckman, Ichimura, Smith, et al. 1998; Heckman, LaLonde, and Smith 1999) proposed the following structure for understanding causal effects and associated biases. Let (1) be the equation if one is treated and (2) be the equation if one is not treated:

$$Y_1 = g_1(X) + U_1, \quad (1)$$

$$Y_0 = g_0(X) + U_0, \quad (2)$$

where Y is the outcome of interest, X consists of observable variables, and U is the error term. In the traditional regression model, it is assumed that $E(U_1|X) = E(U_0|X) = 0$ and that $g_1(X) = X'\beta_1$ and $g_2(X) = X'\beta_2$. Let $D = 1$ be the indicator of whether one is treated and 0 otherwise. To further illustrate, if one assumes that the effect of treatment is additive, one could rewrite and combine equations (1) and (2) as

$$Y = \gamma D + X'\beta + U_0 + (U_1 - U_0)D.$$

Assume the estimand of interest is the effect of treatment for the treatment group. Then the expected treatment effect δ for those who are treated, conditional on observed covariates, equals

$$E(\delta|X, D = 1) = g_1(X) - g_0(X) + E(U_1 - U_0|X, D = 1), \quad (3)$$

or

$$E(\delta|X, D = 1) = \gamma + E(U_1 - U_0|X, D = 1) \quad (4)$$

in the simple additive case.

Note that this quantity is different from what is sometimes called the “treatment effect” in an experimental context. The treatment effect in an experimental context with random assignment of the subjects to the treatment and control group would be

$$E(\delta|X) = g_1(X) - g_0(X). \quad (5)$$

The difference between these two expressions is that the expectation of the conditional (on D) treatment effect contains a term involving the disturbance variables U_0 and U_1 , which is shown in equation (3). In general, the left side of equation (3) cannot be directly estimated because one cannot observe the outcome for the treatment group in the counterfactual case when it does not receive treatment. Therefore, one must estimate the treatment effect by contrasting the outcomes for those in the treatment group with the outcomes for those in the nontreated group. Such a contrast produces the so-called “naive” estimator (so called because it assumes that the treated and

the untreated groups have equal conditional expectations for the error terms):

$$E(\tilde{\delta}|X) = g_1(X) - g_0(X) + E(U_1|X, D = 1) - E(U_0|X, D = 0). \quad (6)$$

The difference between the true treatment effect for the actually treated in (3) and the naive estimator (6) is

$$E(\tilde{\delta}|X) - E(\delta|X, D = 1) = E(U_0|X, D = 1) - E(U_0|X, D = 0) = B(X). \quad (7)$$

$B(X)$ is an expression for the conditional bias of the naive estimator when estimating the treatment effect for the treated, in which the conditioning is done on some function of the observed covariates (this is also referred to as the “point-by-point” bias). When this quantity is equal to zero, we can say that the assignment mechanism is “ignorable” with respect to estimating the treatment effect for the treated. Furthermore, as Rosenbaum and Rubin (1984) show, the bias under ignorable assignment is also zero if one conditions on the propensity score $P(X)$, that is,

$$B(P(X)) = E(U_0|P(X), D = 1) - E(U_0|P(X), D = 0) = 0, \quad (8)$$

where $P(X) = \Pr(D = 1|X)$ is the probability that the case is assigned to (or selects) treatment as a function of the covariates X . The case of nonignorable assignment is equivalent to what Heckman and Robb (1985) refer to as “selection on observables.”

In the case of “selection on unobservables,” the bias shown in equation (7) is no longer zero at every point, and it is this fact that makes estimates based on equation (6) (including ordinary least squares [OLS] and propensity score methods) to be less than ideal strategies for determining the average treatment effect for the treated. As Heckman et al. (Heckman, Ichimura, Smith, et al. 1998; Heckman et al. 1999) have shown, however, one can still reduce the bias if one has access to variables that predict assignment but that have no structural effect on the outcome. Assume that an index function

$$I^* = H(Z) - v \quad (9)$$

determines participation in the treatment group, where $D = 1$ when $I^* > 0$ and $D = 0$ otherwise, Z is a set of observable variables (including perhaps variables in X) that predict assignment to treatment, and (U_0, U_1) is potentially correlated with v .¹ Under these assumptions,

$$\begin{aligned} E(U_0|Z, X, D = 1) &= E(U_0|v < H(Z)), \\ E(U_0|Z, X, D = 0) &= E(U_0|v \geq H(Z)). \end{aligned} \quad (10)$$

As Heckman, Ichimura, Smith, et al. (1998) note, this bias can be expressed as

$$\begin{aligned} B(P(Z)) &= E(U_0|P(Z), D = 1) \\ &\quad - E(U_0|P(Z), D = 0), \end{aligned} \quad (11)$$

where $P(Z)$ is the probability that $D = 1$, given Z . As long as U_0 and v are correlated, the bias in (11) is not pointwise equal to zero in general, and matching methods will generally provide biased estimates for the treatment effect for the treated.

THEORETICAL AND EMPIRICAL FOUNDATIONS FOR BIAS CANCELLATION

While pointwise bias in equation (11) is not equal to zero, the analyst's primary goal is usually to recover accurate estimates of the *average* treatment effect (or, alternatively, the average treatment effect for the treatment group), not the conditional (i.e., "point-by-point") treatment effect or the treatment effect for the treated. The extent of bias in the estimate of the *average* effect will depend both on the way the average is computed and on the nature of the association between bias and observed covariates. In particular, the average bias can be substantially reduced in situations in which the *sign* of the bias varies with observable covariates. In this situation, bias cancellation will occur when an average effect is computed.

Heckman et al. (Heckman, Ichimura, Smith, et al. 1998:1027-29; Heckman et al. 1999:1958-59) have argued that bias reduction is especially large in the case of symmetric distributions, such as the standard multivariate normal distribution. They argued that the average bias

equals zero over intervals around $P = 1/2$ under the conditions that the latent variables v and U_0 are symmetrically distributed around zero. Under these circumstances, they argued that the absolute magnitude of $B(P(Z))$ is symmetrically distributed around $P = 1/2$, while the sign reverses around $P = 1/2$. Consequently, if P itself is symmetrically distributed around $1/2$, then the average bias would equal zero in these symmetric intervals. They further argued that even if P were not symmetrically distributed, the average bias would still be zero within symmetric intervals under an appropriate matching process.

This claim turns out to be incorrect. As we show in Appendix A, the sign of $B(P(Z))$ always remains the same in a symmetric distribution. Contrary to Heckman et al.'s assertion, therefore, the bias does not cancel in symmetric intervals around $P = 1/2$ in the model, which is most commonly assumed by analysts who attempt to correct for selection bias—namely, the bivariate normal case (Heckman, Ichimura, Smith, et al. 1998; Heckman et al. 1999).

Nonetheless, Heckman, Ichimura, Smith, et al.'s (1998) study is very revealing about the empirical variation in bias as a function of observed covariates. The Department of Labor had previously collected experimental data to evaluate four training centers participating in the Job Training Partnership Act (JTPA). In this experiment, a random group of individuals who volunteered to undergo job training were refused participation, and data on these individuals (including their subsequent earnings) were used to estimate the average effect of training on wage change. In addition, data were also collected on a nonexperimental comparison group of individuals who were not trained. Heckman, Ichimura, Smith, et al. used the data for the nonexperimental comparison group and the data for trainees to estimate the average training effect using various selection-bias correcting methods designed for nonexperimental data. Their goal was to compare the performance of these estimators against the presumably unbiased estimate of the average treatment effect from the experimental data. These data also allowed the authors to estimate the relationship between the bias in the estimated training effect using "naive" methods (those that do not attempt to control for selection effects) and observed covariates. Heckman, Ichimura, Smith, et al.'s Table 7 (an extract of which is presented as Table 1 in this article) shows that, as an empirical matter, the bias in the estimated treatment

TABLE 1: Estimated Selection Bias (in Dollars/Month) in Heckman, Ichimura, Smith, et al.'s (1998) Comparison of Experimental and Nonexperimental Evaluation Data for Four Training Centers Participating in the Job Training Partnership Act.

	<i>Decile of the Empirical Distribution for the Probability of Choosing to Be Trained</i>								
	1	2	3	4	5	6	7	8	9
	[.0002, .0023)	[.0023, .0087)	[.0087, .0152)	[.0152, .0269)	[.0269, .0410)	[.0410, .0822)	[.0822, .0983)	[.0983, .1337)	[.1337, .2534)
Average bias across six quarters	-282 (116)	-188 (91)	-118 (81)	-63 (79)	3 (98)	168 (130)	169 (117)	81 (147)	488 (281)

SOURCE: Heckman, Ichimura, Smith, et al. (1998:1049, Table 7, row 6). Bootstrap standard errors are shown in parentheses. Heckman, Ichimura, Smith, et al. report that the data were too sparse to estimate the bias in the 10th decile.

effect varied with the estimated probability that an individual would choose to be treated. In particular, the bias changed sign in the training evaluation data, being *positive* at relatively high probabilities of choosing treatment (i.e., at high probabilities of being treated, the naive predictor was biased upward—it gave too large an estimate of the treatment effect) and *negative* at low probabilities.²

The empirical pattern found in the Heckman, Ichimura, Smith, et al. (1998) data suggests that the distribution of the error variables in the JTPA training data was far from the bivariate normal distribution typically assumed in the literature on sample selection models. We speculate that this observed pattern occurred because the choice to be trained may have been generated by a *mixture* of assignment mechanisms, rather than the single assignment mechanism that is commonly assumed to operate in nonexperimental data. If the mixture process had the property that the correlation structure for those with low probabilities of receiving training was different from the correlation structure for those with high probabilities of receiving training, the result would be similar to what Heckman, Ichimura, Smith, et al. observed in the JTPA data.

To motivate the theoretical plausibility of such a mixing distribution, it is useful to consider the substantive meaning of the correlation between U_0 (the disturbance in the not-treated structural equation)

and ν (the error in the assignment equation) that gives rise to the bias (see equation (11)). For simplicity, let us assume that individuals know their U_0 —in other words, they can perfectly predict outcomes in the situation in which they chose not to be treated (the assumption of perfect knowledge is unnecessary—we could instead assume that their estimate of U_0 is highly correlated with the true U_0). We make the further assumption that individuals make their decisions to participate (i.e., their choice of ν) in response to their knowledge of U_0 . We then consider the following two decision-making scenarios as plausibly operating at the same time. In statistical terms, these scenarios differ in the assumed correlation between U_0 and ν . In behavioral terms, these scenarios differ in the nature of the decision-making process.

*SCENARIO A: POSITIVE CORRELATION
BETWEEN U_0 AND ν*

Consider the case in which U_0 and ν are positively correlated. From the definition of the selection function, the probability of selection into the treatment *increases* as ν *decreases*. A positive correlation means that (after controlling for measured variables) the higher the earnings in the absence of treatment, the lower the probability of treatment. A negative correlation means that (after controlling for measured variables) the higher the earnings in the absence of treatment, the higher the probability for selection into treatment. Recalling that the gain from treatment for the treated equals the structural effect of treatment plus the difference between the disturbance in the treatment and the nontreatment equations (equation (3)), we also need to acknowledge the possible impact of the correlation between U_0 and U_1 on our interpretation. If U_0 equaled U_1 , then (conditional on X) everyone would get the same benefit from treatment. In the more plausible scenario, however, U_0 and U_1 are positively correlated but not perfectly so. In this latter scenario, individuals with high (and positive) U_0 tend to benefit less from the treatment than do individuals with low (and negative) U_0 . Thus, a positive correlation between U_0 and ν means that (after controlling for measured variables) individuals who would experience lower than average gains from treatment are less likely to select themselves into treatment than are individuals who would experience higher than average gains from treatment.

As equation (11) shows, a positive correlation between U_0 and v corresponds to a negative bias in the naive estimate of the treatment effect. Those in the treatment group would have an average value of U_0 that is less than zero, while those in the control group would have an average value of U_0 that is greater than zero. The naive estimator of the treatment effect, which essentially differences the observed outcomes for the group selected into treatment from the observed outcomes for the group selected into nontreatment, would not take into account the selection effect and thus would arrive at a downwardly biased estimate of the treatment effect.

*SCENARIO B: NEGATIVE CORRELATION
BETWEEN U_0 AND v*

Suppose that the correlation between U_0 and v is *negative*. In this scenario (after controlling for measured variables), the higher the earnings in the absence of treatment, the higher the probability for selection into the treatment. In this situation, the group of individuals who select themselves into treatment actually would experience a smaller average impact of the treatment than would the group of individuals who do not select themselves into treatment.³ This behavior might seem to be irrational, but it could be interpreted as consistent with what Kahneman and Tversky (1979) referred to as “prospect theory.” Prospect theory argues that the negative effects of losses on a utility function are greater than the positive effects of equivalent-size gains and that most people therefore exhibit what they referred to as “loss aversion.” If the correlation between U_0 and U_1 is positive (and this seems very likely), then those individuals who (conditional on X) would have higher than average earnings if they were not treated (i.e., $E(U_0) > 0$) would also have higher than average earnings if they were treated (i.e., $E(U_1) > 0$), even though their *gains* from treatment would be lower than average (i.e., $E(U_1 - U_0) < 0$). If individuals with higher than average earnings in the absence of treatment developed a strong interest in “preserving their gains,”⁴ then they would be especially likely to choose to be treated. For these individuals, a naive estimator would overstate the “true” effect of treatment.

The two forms of behavior described above in Scenario A and Scenario B produce offsetting biases. The pattern observed in

Table 1 would result if the individuals who behaved according to “conventional rationality” tended to be clustered at the lower values of the linear predictor for treatment, while those who behaved according to the “prospect theory” model were clustered at the higher values of the linear predictor for treatment. If the distribution of the disturbance variables was perfectly “mirror symmetrical” around some value of $H(Z)$ (in the sense that the correlations between U_0 and ν were equal and of the opposite sign at values of $H(Z)$ that were equidistant from some specific reference point), then the complete bias cancellation written about by Heckman, Ichimura, Smith, et al. (1998) would occur, and the average bias would be zero if the data were perfectly balanced around this point. If the offsetting biases did not perfectly balance, then the average bias would not be eliminated, but it would be smaller than in the case in which all point-by-point biases had the same sign. The extent of bias cancellation, or whether the bias even changes sign as a function of $H(Z)$, is an empirical question, just as is the question of whether an error distribution is multivariate normal. It should be noted, however, that the standard framework for eliminating bias due to unobservables (corresponding to the two-step Heckman selectivity correction approach) simply assumes a bivariate normal distribution for the errors in the structural and in the assignment equations on the basis of no empirical data. The data presented by Heckman, Ichimura, Smith, et al., however, are important because they suggest that in at least one important context, bias cancellation does occur. Our theoretical argument above is important because it shows that the Heckman empirical result is inconsistent with the standard assumptions underlying the usual approach to sample selection; also, our argument provides a theoretical rationale as to why the empirical pattern found by Heckman, Ichimura, Smith, et al. in the job-training data may not be an isolated instance.

*THE ROLE OF U_1 IN THE ESTIMATION
OF TREATMENT EFFECTS*

The above discussion concerned the possibility of bias that arises out of possible correlation between U_0 and ν . It is also important to consider the implications for bias that arise from possible correlation

between U_1 and (U_0, ν) . Technically, U_1 plays no role in the structure of the point-by-point bias for the estimate of the treatment effect for those assigned or self-assigned to treatment. This can be seen in equation (3) or in equation (11).⁵ Nonetheless, U_1 can play an important indirect role in the estimation process because, as we suggested in our proposed justification for bias cancellation, the extent and nature of bias cancellation can depend on the relationship between U_0 and U_1 . Consequently, even if $B(X)$ is unaffected by U_1 , the *average* of $B(X)$ across the range of X might be influenced by U_1 and by its correlation with the other disturbance variables.

DATA AND METHODS

We illustrate the potential impact of matching estimates in the presence and absence of bias cancellation using panel data on additional job training. The use of real data for studying the extent of bias in alternative estimators has disadvantages that arise from our inability to test the underlying exclusion restrictions concerning the selection equations, as well as from our inability to directly observe the shape of the joint error distribution. To overcome this limitation, we use simulated data in which the observed treatment variable and the observed dependent variable are replaced by measures simulated from a predetermined model. The advantage of simulated data is that we know the model that generates the data, and we can use this knowledge to evaluate our estimation methods.

The starting point for our simulations was a subset of data drawn from the German Socioeconomic Panel (GSOEP). We used the 1989 and 1993 special modules on any continuous training undertaken by sample members in the previous three years. We combined demographic, employment, income, and wage information from the 1985 to 1995 waves for respondents between the ages of 18 and 60 who were living in West Germany in 1989 or 1993 with the information on continuous training to obtain the data that formed the basis for our simulations. Our working data set consisted of 9,259 observations that were nonmissing on our predictor and dependent variables. In the sample for which we had complete information, 12.5 percent reported further training. We applied a three-step simulation procedure.

First, we simulate the correlation of U_1 , U_0 , and v following a standard symmetric distribution and a distribution that produces bias cancellation. While we believe that the substantive force producing bias cancellation is a mixture process corresponding to the two scenarios described above, we use the more expedient simulation strategy of assuming that the correlation between U_0 and v_1 reversed beyond a specific point in the distribution.⁶ We intentionally simulated data in which the errors were independent of the measured covariates to determine the impact of bias cancellation on the average treatment effect as a function of the particular pattern of the joint error distributions.

Next we replaced the observed treatment variable with a simulated treatment variable, which was constructed by combining the prediction from probit estimations for further training with the simulated stochastic disturbance v (the results of the probit estimation can be found in Table A1 in Appendix B). This simulated treatment variable by construction has the same binary distribution as the original variable. We then produced a second simulated further training variable by adding unity to all the linear predictors from the probit equation to get a variable whose probability distribution was closer to being symmetrical, around 0.5. With this modification, the proportion of the sample with further training rose to 41.6 percent. By combining our simulated training and disturbance variables in different ways, we obtained the six cases listed in Table 2, which (see especially the notes to Table 2) provides a precise description of the disturbance structure in our simulations.

The third step was to simulate the outcome variable. We estimated a wage equation using OLS, in which the two-year difference in the natural logarithm of the wage was the dependent variable, and the right-hand side included a set of standard covariates (see Table A1 in Appendix B), specifically including further training as a predictor. We then replaced the product of observed further training multiplied by its estimated effect with our simulated training variable multiplied by an effect that we specified (we enhanced the effect of training to be three times as large as the estimated effect in the OLS equation). We added the simulated stochastic disturbance to the estimated log(wage) to obtain a simulated outcome variable.

TABLE 2: Correlation Structures Used in Simulations.

	U_0	U_1	ν
Case 1: Symmetric distribution— U_1 is independent of U_0 and ν			
U_0	1		
U_1	-0.0079	1	
ν_1	0.5984	-0.0116	1
Case 2a: Bias cancellation— U_1 is independent of U_0 and ν ; proportion trained in simulation equals proportion trained in observed data			
U_0	1		
U_1	0.0063	1	
ν_1	0.2999	0.0090	1
Case 2b: Bias cancellation— U_1 is independent of U_0 and ν ; proportion trained in simulation is enhanced			
U_0	1		
U_1	0.0012	1	
ν_1	0.2814	-0.0062	1
Case 3: Symmetric distribution— U_1 is correlated with U_0 and ν			
U_0	1		
U_1	0.6949	1	
ν_1	0.5928	0.4905	1
Case 4a: Bias cancellation— U_1 is correlated with U_0 and ν ; proportion trained in simulation equals proportion trained in observed data			
U_0	1		
U_1	0.6950	1	
ν_1	0.3029	0.2440	1
Case 4b: Bias cancellation— U_1 is correlated with U_0 and ν ; proportion trained in simulation is enhanced			
U_0	1		
U_1	0.6949	1	
ν_1	0.2817	0.2275	1

NOTE: Case 1: U_0 and ν follow a bivariate normal distribution. The specified positive correlation corresponds to Scenario A. U_1 is assumed to be independent from the other errors (the observed correlation is not zero due to sampling fluctuations). This gives a symmetric distribution for the error variables, with no bias cancellation. Case 2a: The correlation structure is the same as in Case 1 for 75 percent of cases with the lowest propensity scores but then reverses, so that U_0 and ν have a negative correlation above the 75 percent point of the propensity score distribution. This corresponds to Scenario A above for the lower $3/3$ of the distribution and to Scenario B for the upper $1/4$ of the distribution. The simulation model is based on a proportion trained that matches the actual data. Case 2b: The correlation structure is the same as in Case 2a. The simulation model is based on an enhanced proportion trained (about 42 percent as opposed to 12.5 percent in the actual sample), and the point of sign reversal is the midpoint of the propensity scores. Case 3: U_0 , U_1 , and ν are symmetrically distributed. There is no bias cancellation in this case. Case 4a: U_0 and ν follow a "mirror-symmetric" bivariate normal distribution around $P(\text{train}) = 1/2$; in addition, ν is correlated with U_1 . Proportion trained equals the observed proportion. Case 4b: The correlation structure is the same as in Case 4a. Proportion trained is enhanced as in Case 2b.

In our examples, we assume that the “true” structural effect of training is fixed in the population. The “true” total effect of training varies across the sample, however, because the total effect includes an error component in addition to the structural effect, and this error component is not fixed in the population (see equation (3)). Because we know the underlying structure of the simulated data, we can compute the average bias for the entire sample or for any subsample (e.g., for the treated). This allows us to compare the pattern of bias in our data with the pattern that Heckman, Ichimura, Smith, et al. (1998) observed in the training data. Table 3 shows the relationship between the bias and the propensity score in our simulated data. As can be seen, the simulated data based on symmetric error distributions look quite different from the empirical pattern observed in the Heckman, Ichimura, Smith, et al. data. In contrast, the patterns in the data where the correlation reverses at a specific value of the propensity score (i.e., a specific value of $P(\text{train})$) are qualitatively similar to the empirical pattern found in the Heckman, Ichimura, Smith, et al. data.

We analyzed these simulated data using two methods. The first was OLS. The second involved matching data between those in the treatment group with respondents who were not treated. Our matching procedure is a variant of the propensity/Mahalanobis metric matching method proposed by Rosenbaum and Rubin (1985) and Rubin (1991) and applied by Lechner (1999) (see Appendix B).

For the training-enhanced data in which bias cancellation was (induced to be) present, we also included illustrative estimates based on a two-step matching procedure that was designed to take maximum advantage of the induced bias cancellation in our simulated data. In the first step, we focused on the group of respondents who had been treated, and we matched cases whose probability of training was greater than 0.5 with cases whose propensity score (specifically, the linear component of the probit model—see Appendix B for further clarification) had the same magnitude but was opposite in sign. The resulting matching yielded a distribution of treated cases that was approximately symmetrically distributed around the probability of 0.5. We then matched each subset of treated cases with its most closely matching counterpart in the control

TABLE 3: Average Bias (Percent Change in Wages) as a Function of the Probability of Receiving Training and the Correlation Structure of the Error Distribution: Simulated Data Derived From the German Socioeconomic Panel.

<i>Error Distribution</i>	<i>Fraction Trained</i>	<i>Probability of Receiving Training</i>							<i>Total</i>
		≤ 0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0			
Independent errors $(U_0)(U_1)(v)$	Observed	-.0009 (7,739)	.0010 (1,520)						-.0005
	Enhanced	.0033 (424)	-.0002 (4,104)	-.0009 (3,784)	.0013 (947)				-.0004
Symmetric $(U_0v)(U_1)$	Observed	-.0424 (7,739)	-.0378 (1,520)						-.0403
	Enhanced	-.0423 (424)	-.0371 (4,104)	-.0369 (3,784)	-.0357 (947)				-.0349
Symmetric $(U_0v)U_1$	Observed	-.0424 (7,739)	-.0371 (1,520)						-.0402
	Enhanced	-.0441 (424)	-.0373 (4,104)	-.0362 (3,784)	-.0380 (947)				-.0349
Bias canceling $(U_0v)(U_1)$	Observed	-.0294 (7,739)	.0402 (1,520)						-.0091
	Enhanced	-.0423 (424)	-.0371 (4,104)	-.0082 (3,784)	.0357 (947)				-.0158
Bias canceling $(U_0v)U_1$	Observed	-.0276 (7,739)	.0371 (1,520)						-.0091
	Enhanced	-.0441 (424)	-.0373 (4,104)	-.0077 (3,784)	.0380 (947)				-.0155

NOTE: Sample size is in parentheses. Sample size is zero for cells without entries. Error terms that are grouped together in column 1 in the same parentheses are assumed to be correlated with each other.

sample and used the resulting matched subsample in the estimation procedure.

RESULTS

Table 4 shows the results from our analyses of the simulated data in the situation in which U_1 is independent of the other error variables (Cases 1, 2a, and 2b in Table 2). Table 5 contains results for the cases in which all three error variables are correlated (Cases 3, 4a, and 4b in Table 2).

Tables 4 and 5 demonstrate an assertion made by Heckman and Robb (1985) that is still underappreciated in the social sciences—namely, that the “true” experimental effect differs from the “true” treatment effect for individuals who are selected (or self-selected) through a nonrandom assignment process. Row 1 contains the “true” average effect of training for all members of the sample. We set the effect to be the same value for all observations in the simulated data—namely, three times as great as the OLS estimate for the training effect in the GSOEP subsample from the simple specification that we report in Table A1. The effect in row 2 in Tables 4 and 5 is the “true” average effect for the self-selected sample; that is, it is the average δ from equation (3) for the treated subsample. This value, which is intrinsically unobservable in real data, equals the average of equation (3) for sample members. It is “true” in the sense that it is the average difference between the outcome that *was* experienced by the treated population and the outcome they *would have* experienced if they had not been treated (one can only have “access” to this information in a simulated world; in the “real” world, such information is intrinsically missing). This outcome is the sum of the “true” experimental effect of the treatment and the difference in the error in the “treated” and the “not-treated” equation. Row 3 reports the “naive” estimate of the causal effect, which is shown in equation (6), where we substitute sample averages for expectations.

Table 4 shows that—under the assumptions of Cases 1, 2a, and 2b—the true average effect for the self-selected sample is greater than the experimental effect. This relationship arises from the positive correlation between U_0 and v (which implies that individuals who

TABLE 4: Results From Simulated Data (Standard Errors in Parentheses): Imposed Correlations between U_0 and ν as in Panels 1, 2a, and 2b of Table 2.

	<i>Fraction Trained Is as Observed</i>		<i>Fraction Trained Is Enhanced</i>	
	<i>Symmetric Distribution (No Bias Cancellation)</i>	<i>Distribution-Produced Bias Cancellation^a</i>	<i>Symmetric Distribution (No Bias Cancellation)</i>	<i>Distribution-Produced Bias Cancellation^b</i>
1. "True" experimental effect	.0599	.0599	.0599	.0599
2. "True" average effect for the self-selected sample	.0965	.0666	.0805	.0686
3. "Naive" estimate	.0563	.0575	.0456	.0528
4. Ordinary least squares (OLS)	.0553 (.0012)	.0589 (.0012)	.0443 (.0008)	.0552 (.0008)
5. Matching	.0526 (.0018)	.0597 (.0018)	.0438 (.0010)	.0561 (.0010)
6. Matching, balance on $P(\text{train})$	NA	NA	NA	.0581 (.0012)
7. Percent bias reduction vs. OLS	Negative	10 percent	Negative	21 percent ^c

NOTE: NA = Not Applicable.

a. Correlation reversed at the 75th percentile.

b. Correlation reversed at $P(\text{train}) = .50$.

c. Computations based on prior balancing on $P(\text{train})$.

TABLE 5: Results From Simulated Data (Standard Errors in Parentheses): Imposed Correlations Between U_0 , U_1 , and v as in Panels 3, 4a, and 4b of Table 2.

	<i>Fraction Trained Is as Observed</i>		<i>Fraction Trained Is Enhanced</i>	
	<i>Symmetric Distribution (No Bias Cancellation)</i>	<i>Distribution-Produced Bias Cancellation^a</i>	<i>Symmetric Distribution (No Bias Cancellation)</i>	<i>Distribution-Produced Bias Cancellation^b</i>
1. "True" experimental effect	.0599	.0599	.0599	.0599
2. "True" average effect for the self-selected sample	.0651	.0621	.0634	.0612
3. "Naive" estimate	.0531	.0545	.0442	.0458
4. Ordinary least squares (OLS)	.0238 (.0012)	.0529 (.0012)	.0263 (.0008)	.0452 (.0008)
5. Matching	.0245 (.0016)	.0556 (.0019)	.0251 (.0011)	.0459 (.0011)
6. Matching, balance on $P(\text{train})$	NA	NA	NA	.0524 (.0013)
7. Percent bias reduction vs. OLS	2 percent	29 percent	Negative	45 percent ^c

NOTE: NA = Not Applicable.

a. Correlation reversed at the 75th percentile.

b. Correlation reversed at $P(\text{train}) = .50$.

c. Computations based on prior balancing on $P(\text{train})$.

anticipate negative shocks if they proceed along their current path are more likely to choose the alternative path of treatment), coupled with the independence between U_0 and U_1 . Furthermore, at least for our examples, the increment in the total effect that comes from the difference between U_1 and U_0 is much greater in the symmetric case than in the bias cancellation case.

In our examples, the OLS estimate (row 4) was (not surprisingly) quite close to the “naive” estimate that arises from substituting the control group’s outcome in the absence of treatment for the outcome that the treatment group would have experienced if it were not treated. The OLS bias was more significantly enhanced for data generated from a symmetric distribution. We note again that the OLS specification was “perfect” in that it exactly matched the specification that was used to generate the data.

Row 5 shows the results we obtained using the matching procedure based on the probit propensity score for selection into further training. When the data were generated by a symmetric distribution, the matching method slightly underperformed the OLS estimates. When the data were generated by a bias-canceling distribution, matching outperformed OLS. Finally, two-step matching (row 6) as described above led to a still greater reduction in bias, relative to the OLS estimator.

Table 5 shows results for Cases 3, 4a, and 4b from Table 2, in which (unlike for Table 4) we imposed a positive correlation between U_0 and U_1 . The first major consequence of this positive correlation is to reduce the size of the “true” average effect of selecting training for the sample that was nonrandomly selected into training (row 2). The second major consequence is to increase the downward bias of the OLS estimate and of the corresponding matching estimates in the symmetric distribution cases (row 4, columns 1 and 3, and row 5). As in Table 4, the matching estimate was inferior to the OLS estimate in the symmetric distribution case (again, recall that the OLS equation has no misspecification bias—it exactly matches the equation used to simulate the dependent variable). Also, as in Table 4, the matching estimators outperformed the OLS estimator in both data cases. However, in the simulated data, in which the proportion trained equaled the observed fraction trained, the matching estimator was inferior to the OLS estimator, which in this case was actually rather close to the “true” value.

DISCUSSION AND CONCLUSION

Taking as a starting point Heckman, Ichimura, Smith, et al.'s (1998) finding that bias cancellation can exist in data in which individuals can select themselves into a treatment, we have made three contributions to the literature on causal estimation. First, we have shown that Heckman, Ichimura, Smith, et al.'s theoretical justification for bias cancellation (the assertion that bias cancellation arises naturally from symmetrically distributed stochastic disturbances in the structural and assignment equations) is incorrect; in fact, bias cancellation never occurs when the errors are symmetrically distributed. Second, we have proposed a behavioral explanation for the observed bias cancellation, one that could also produce a bias-canceling, data-generating process in other situations in which individuals are able to select themselves into a potentially beneficial state or activity (i.e., a "treatment"). Third, we have generated and analyzed simulated data on further training and wage changes, and we have shown how bias can emerge from the joint distribution of the error variables in the structural and assignment equations.

Using correlation structures that appear to correspond reasonably well with the observed bias pattern that Heckman, Ichimura, Smith, et al. (1998) found for evaluation data of the JTPA, we find that bias cancellation in the error distribution reduces the bias of estimates for the average treatment effect, even when OLS is used to produce the estimate. Generally speaking (though not in all instances), a simple matching estimator does an even better job than OLS, even when the OLS estimate derives from an equation that was perfectly specified. This is a potentially important result because the relatively strong performance of the matching estimator is occurring in a situation (selection on unobservables) in which matching estimators were thought to have no comparative advantage. We further found that a two-step balancing procedure (first on the cases treated and then a matching between treated and control cases) produces superior estimates when the axis for matching in the first step roughly matches the point at which bias reversal occurs in the data.

Because bias cancellation is a property of the distribution of an unobservable variable, it is difficult to know whether Heckman,

Ichimura, Smith, et al. (1998) discovered a common pattern or an unusual one. It is also difficult to know whether their finding that bias reversal took place near the midpoint of distribution is common or unusual. In our opinion, the answers to these questions require first an understanding of *why* bias cancellation occurs. In this article, we have proposed what we believe is a plausible answer to this question. Further theoretical refinement and further empirical analysis of cases in which experimental and nonexperimental data can be directly compared are probably the best hope for greater understanding of this phenomenon.

One final observation should be made. The standard Heckman two-step correction for sample selection bias typically assumes that the structural and assignment errors follow a bivariate normal distribution. This distribution is very different from a distribution that would produce bias cancellation. If bias cancellation is common, then the theoretical justification for applying the Heckman two-step method becomes more questionable. The practical implications of bias cancellation for the Heckman two-step method, as for other techniques, depend on the specific distributions both of the disturbances of the structural and selection equations and of the frequency distribution of cases across the covariate distribution. Further research is needed to establish the robustness of various estimators of causal effects across the range of plausible distributions to better understand the potential implications of bias cancellation for the estimation of causal effects using nonexperimental data.

APPENDIX A

BIAS CANCELLATION AND SYMMETRIC DISTRIBUTIONS

Assume that the joint distribution of v and U_0 is symmetrical. Assume further that for a group of individuals indexed by i , $H(Z_i) = a$, while for a second group of individuals indexed by i' , $H(Z_{i'}) = -a$. Then for group i , equation (11) can be reexpressed as

$$B(P(Z)|H(Z) = a) = E(U_0|v < a) - E(U_0|v \geq a), \quad (12)$$

while for group i' ,

$$B(P(Z)|H(Z) = -a) = E(U_0|v < -a) - E(U_0|v \geq -a). \quad (13)$$

But in a bivariate symmetric distribution, $E(U_0|v < a) = -E(U_0|v \geq -a)$. Therefore,

$$\begin{aligned} B(P(Z)|H(Z) = a) + B(P(Z)|H(Z) = -a) &= E(U_0|v < a) - E(U_0|v \geq a) \\ &\quad + E(U_0|v < -a) - E(U_0|v \geq -a) \\ &= 2 \cdot E(U_0|v < a) - 2 \cdot E(U_0|v \geq a) \\ &= 2 \cdot B(P(Z)|H(Z) = a) \neq 0. \end{aligned} \quad (14)$$

To see this another way, note (e.g., Heckman, Ichimura, Smith, et al. 1998, note 25), that

$$\begin{aligned} E(U_0|P(X)) &= E(U_0|P(X), D = 1)P(D = 1) \\ &\quad + E(U_0|P(X), D = 0)P(D = 0) \\ &= E(U_0|P(X), D = 1)P(X) \\ &\quad + E(U_0|P(X), D = 0)(1 - P(X)), \end{aligned} \quad (15)$$

where $D = 1$ when an individual is treated, and $D = 0$ when an individual is not treated. $P(X)$, the propensity score, gives the probability that an individual is treated. If we assume that $E(U_0|P(X)) = 0$, then, by rearranging terms, we obtain

$$E(U_0|P(X), D = 1) = -\frac{1 - P(X)}{P(X)}E(U_0|P(X), D = 0). \quad (16)$$

Recalling that $B(X) = E(U_0|X, D = 1) - E(U_0|X, D = 0)$, it follows from (16) that

$$B(P(X)) = \frac{-1}{P(X)}E(U_0|P(X), D = 0). \quad (17)$$

This expression can only change sign if $E(U_0|P(X), D = 0)$ changes sign for some value of $P(X)$, but this never happens in a symmetric distribution.

APPENDIX B
METHOD FOR MATCHING

The steps of the matching procedure used in this article, which are similar to the procedure used by Lechner (1999), are as follows:

1. Split the observations into two pools: a treatment group T (further training) and a comparison group C (no further training). Estimate a probit model for participation in the treatment group.
2. Based on the estimated probit model, compute the propensity score $\hat{b}'X_T$ and the variance $var(\hat{b}'X_T)$ for all treated persons T . Construct for all treated persons the interval $\hat{b}'X_T \pm w\sqrt{var(\hat{b}'X_T)}$, and choose w , such that one obtains a confidence interval of the desired size around $\hat{b}'X_T$.⁷
3. Randomly select a treated person from the treatment group n_T .
4. Find observations in the control group that obey $\hat{b}'X_C \in \hat{b}'X_T \pm w\sqrt{var(\hat{b}'X_T)}$.
 - (a) If there is *no* observation of the control group lying between the given limits of the confidence interval, the selected person will not be considered further, and Step 3 has to be repeated.
 - (b) If there is *only one* observation between the given limits of the confidence interval, go to Step 6.
 - (c) If there is *more than one* observation in the confidence interval, proceed as follows. Compute additional match variables in relation to the start date of observation n_T and a subset of variables already included in the estimation of the propensity score. Denote these variables as a_T and a_C . Evaluate the distance $d(T, C) = (\hat{b}'X_T, a_T)' - (\hat{b}'X_C, a_C)'$ between each nontreated and treated observation. Choose the nontrainee who is the "closest neighbor" of the trainee T in terms of the Mahalanobis distance, defined as $md = d(T, C)' cov^{-1} d(T, C)$, where cov is the estimated sample covariance matrix of $(\hat{b}'X, a)'$ in the group of nontrainees.
5. Remove the treated and nontreated (now matched control) observations from their respective groups.
6. If there are any observations left in the trainee group, begin again with Step 3. In the illustrative cases reported in this article, we matched on the propensity score by itself; no additional covariates were used.

TABLE A1: Probit Equation That Predicted Further Training and Was Used to Simulate the Further Training Variable and the Wage Change Equation in the Actual German Socioeconomic Panel Data.

<i>Variable</i>	<i>Probit: Further Training</i> (n = 9, 259)		<i>Ordinary Least Squares:</i> <i>Wage Change</i> (n = 9, 288)	
	<i>Coefficient</i>	<i>Standard Error</i>	<i>Coefficient</i>	<i>Standard Error</i>
Age (years)	-.0143	.0022	-.0021	.0002
Female (0/1)	-.1845	.0486	.0107	.0043
German born (0/1)	.3210	.1077		
Partner employed (0/1)	.0461	.0447		
Number of kids	.0267	.0201		
Single (0/1)	-.0996	.0680		
Widowed (0/1)	-.2128	.2852		
Divorced (0/1)	-.0984	.1008		
Separated (0/1)	.1492	.1523		
Highest degree:				
<i>Hauptschule</i> (0/1)	-.2588	.1972	.0149	.0248
Highest degree:				
<i>Realschule</i> (0/1)	.1937	.1968	.0169	.0249
Highest degree:				
<i>Fachhochschule</i> (0/1)	.2403	.2073	.0542	.0263
Highest degree:				
<i>Abitur</i> (0/1)	-.0383	.2187	.0307	.0277
Highest degree:				
Other (0/1)	a		.0362	.0437
Highest degree:				
University (0/1)	.5230	.2143	.0549	.0277
Part-time employed (0/1)	-.3090	.0685		
Further training (0/1)			.0200	.0062
Federal land:				
Schleswig-Holstein (0/1)	-.3694	.1178		
Federal land:				
Hamburg (0/1)	-.1812	.1435		
Federal land:				
Niedersachsen (0/1)	-.1870	.0944		
Federal land:				
Bremen (0/1)	-.7159	.2687		
Federal land:				
Nordrhein-Westfalen (0/1)	-.3407	.0882		
Federal land:				
Hessen (0/1)	-.2843	.09865		
Federal land:				
Rheinland-Pfalz (0/1)	-.3336	.1075		

(Continued)

TABLE A1: Continued

Variable	Probit: Further Training (n = 9, 259)		Ordinary Least Squares: Wage Change (n = 9, 288)	
	Coefficient	Standard Error	Coefficient	Standard Error
Federal land:				
Baden-Wuerttemberg (0/1)	-.1991	.09224		
Federal land:				
Bayern (0/1)	-.2995	.09175		
Constant	-.3513	.2640	.1613	.0256

NOTE: Reference categories: male, not German born, partner not employed, married, highest degree: no degree, full-time employed, no further training, and federal land: Berlin.

a. Category dropped because it predicts failure perfectly.

NOTES

1. Algebraically, it makes no difference whether the sign of the error term v for the index function is positive or negative.

2. Interestingly, the sign reversal in Heckman et al.'s (1998) data took place around the median probability of training observed in the data. This point at which the bias changed direction in Heckman's data, at about $P = 0.04$, was far below the 0.5 probability level that figured in Heckman, Ichimura, Smith, et al.'s (1998) theoretical justification for the bias cancellation phenomenon.

3. At the individual level, the effect of the treatment is equal to $\delta_i = g_1(X_i) - g_0(X_i) + (U_{1i} - U_{0i})$ or $\gamma + (U_{1i} - U_{0i})$ in the simple additive case (see equations (3) and (4)). A positive correlation between U_0 and U_1 implies that $U_1 - U_0$ is negatively related to U_0 .

4. Their "preserved gain" is $g_1(X) - g_0(X)$, or γ in the additive case. See equations (3) and (4).

5. As Heckman (1997) shows, the bias does depend on U_1 as well as U_0 when the quantity to be estimated is the effect of treatment for a randomly selected sample from the population. This dependency on U_1 disappears, however, when the quantity to be estimated is the effect of treatment on the sample selected into treatment.

6. Note that in Heckman, Ichimura, Smith, et al.'s (1998) data, the probability value at which the sign of the bias reverses is far lower than 0.5; indeed, it appears to be roughly in the center of the data. This is a highly fortuitous event because it implies that substantial cancellation would occur by averaging over the entire sample. In our simulations, we instead specified that the sign reversal would occur at the point where the probability of receiving the treatment is 0.5 for the enhanced data, in which 41.6 percent of sample members were trained. For the simulated data in which the proportion trained equals the observed proportion, we specified that the sign reversal would occur at the 75th percentile of the propensity scores.

7. Rubin (1991), for example, defined $w = 0.25$, while Lechner (1999) defined $w = 1.65$. We, like Rubin, set $w = 0.25$.

REFERENCES

- Angrist, Joshua D., Guido D. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91:444-55.
- Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153-61.
- . 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32:441-62.
- Heckman, James J., Robert J. LaLonde, and Jeffrey A. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." Pp. 1865-2097 in *Handbook of Labor Economics*, vol. 3, edited by O. Ashenfelter and D. Card. Amsterdam: North-Holland.
- Heckman, James J., Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66:1017-98.
- Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64:605-54.
- . 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 65:261-94.
- Heckman, James J. and Richard Robb Jr. 1985. "Alternative Methods for Evaluating the Impact of Interventions." Pp. 156-245 in *Longitudinal Analysis of Labor Market Data*, edited by J. J. Heckman and B. Singer. Cambridge, UK: Cambridge University Press.
- Kahneman, Daniel and Amos Tversky. 1979. "Prospective Theory: An Analysis of Decision Under Risk." *Econometrica* 47:263-91.
- Lechner, Michael. 1999. "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany After Unification." *Journal of Business & Economic Statistics* 17:74-90.
- Rosenbaum, Paul R. and Donald B. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79:516-24.
- . 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician* 39:33-38.
- Rubin, Donald B. 1991. "Practical Implications of Models of Statistical Inference for Causal Effects and the Critical Role of the Assignment Mechanism." *Biometrics* 47:1213-34.
- Rubin, Donald B. and Neal Thomas. 1996. "Matching Using Estimated Propensity Scores: Relating Theory to Practice." *Biometrics* 52:249-61.

Thomas A. DiPrete is a professor of sociology at Duke University. His research interests include social stratification, demography, and quantitative methodology. His current research projects include methods for assessing bias in the estimation of causal effects, the comparative structure of employment insecurity in European and American labor markets, the impact of institutional arrangements that influence the level of incompatibility between work and family on fertility rates in the United States and several European countries, and the investigation of why women have overtaken men in rates of college enrollment and completion in the United States and many other industrialized nations.

Henriette Engelhardt is a research scientist at the Vienna Institute for Demographic Research at the Austrian Academy of Sciences. Her research interests include social demography and quantitative methodology, with a special focus on the estimation of causal effects. Her most recent research project investigates the changing correlation between fertility and female labor force participation over space and time, applying pooled cross-sectional time-series analysis. A paper on the causal relationship between fertility and female employment is forthcoming in Population Studies.