

A Probability Index of the Robustness of a Causal Inference

Wei Pan

University of Cincinnati

Kenneth A. Frank

Michigan State University

Causal inference is an important, controversial topic in the social sciences, where it is difficult to conduct experiments or measure and control for all confounding variables. To address this concern, the present study presents a probability index to assess the robustness of a causal inference to the impact of a confounding variable. The information from existing covariates is used to develop a reference distribution for gauging the likelihood of observing a given value of the impact of a confounding variable. Applications are illustrated with an empirical example pertaining to educational attainment. The methodology discussed in this study allows for multiple partial causes in the complex social phenomena that we study, and informs the controversy about causal inference that arises from the use of statistical models in the social sciences.

Keywords: causal inference, confounding variables, linear models, regression coefficients, sensitivity analysis

Causal inference is an important, controversial issue in most fields of the social sciences, such as educational research, behavioral research, psychometrics, econometrics, and sociology as well as epidemiology and biostatistics. In those fields, researchers routinely draw conclusions about causal relationships between dependent variables and independent variables from statistical models using data from observational studies (see Jacobs, Finken, Griffin, & Wright, 1998; Lee, 1999; Okagaki & Frensch, 1998; Portes, 1999, for examples). However, the usual statistical approaches may not lead to valid causal inferences, even if the models are supported by related theories (Abbott, 1998; Cook & Campbell, 1979; Holland, 1986, 1988; McKim & Turner, 1997; Pearl, 2000; Rubin, 1974; Sobel, 1996, 1998;

The authors are grateful to the Editor and the anonymous reviewers for their valuable comments. This research was partially supported by a grant from the University Research Council of the University of Cincinnati to Wei Pan and by grants from the National Institute of Child Health and Human Development (R01 HD40428-02) and from the National Science Foundation (REC-0126167) to Kenneth A. Frank through the Population Research Center, University of Texas at Austin, Chandra Muller (PI).

Winship & Morgan, 1999). The problem mainly comes from the failure to control for the potentially inexhaustible list of all potentially confounding variables.

For instance, Okagaki and Frensch (1998) examined the relationship between parenting style and children's school performance for different ethnic groups, but did not control for the children's age, gender, or socioeconomic status. Jacobs, Finken, Griffin, and Wright (1998) examined the relationships between parent attitudes, intrinsic values of science, peer support, available activities, and preference for future science career for science-talented, rural, adolescent females. However, they also failed to control for some demographics, such as age and socioeconomic status. In a third example, Lee (1999) examined the differences in children's views of the world after they personally experienced a natural disaster for various ethnic, socioeconomic status, and gender groups, but failed to control for pre-world-views. In still another example, Portes (1999) examined the influence of various factors on immigrant students' school achievement and controlled for many demographic and sociopsychological covariates. Nevertheless, we still can ask: "Did he control for all possible sociopsychological factors?" Therefore, the conclusions in each case might not support causal relationships, although the relationships are statistically significant.

To deal with this problem, the literature offers the following solutions:

1. Abandon the use of causal language and emphasize the effects of causes rather than the causes of effects (Holland, 1986, 1988; Sobel, 1996, 1998);
2. Spend more effort on descriptive work (Abbott, 1998; Sobel, 1998);
3. Use alternative models, e.g. randomized and well-controlled nonrandomized studies (Rubin, 1974);
4. Apply causal discovery algorithms which operate on statistical datasets to produce directed causal graphs (Scheines, Spirites, Glymour, Meek, & Richardson, 1998; Spirites, Glymour, & Scheines, 1993);
5. Use instrumental variables to control for selection into treatment and control (Angrist, Imbens, & Rubin, 1996; Bowden & Turkington, 1984);
6. Assess the robustness of a statistical inference to the impact of confounding variables (Frank, 2000).

If we are still interested in exploring causal relationships in the real world, Approaches 1 and 2 are not helpful. Approach 3, random assignment, is often impractical in the social sciences given logical, ethical, and political concerns (Rubin, 1974; Winship & Morgan, 1999). In addition, it is not always possible to measure all confounding variables to be controlled for in statistical analyses. In Approach 4, causal graphs are generated by calculations of conditional statistical dependence or independence among pairs of variables. However, in most cases, the assumptions under which the algorithms operate are not powerful enough to uniquely identify the real causal structure underlining correlational data, rather than some set of statistically equivalent but genuinely alternative representations (Woodward, 1997). Thus, the soundness of the methodology of causal graphs is uncertain. Though Approach 5 is theoretically strong, in practice, it is difficult to

find an instrumental variable that is correlated with assignment to treatment level but not the outcome (Heckman, 1997; Winship & Morgan, 1999).

Approach 6 applies an alternative paradigm. Inspired by the idea of expressing sensitivity of inferences rather than precisely establishing causality, Frank (2000) quantified the impact of a confounding variable on the inference for a regression coefficient. Define X as a predictor of interest, Y as an outcome, and U as a confounding variable. Then, Frank characterized the impact of the confounding variable on the inference for the predictor of interest (including the impact on the both estimated regression coefficient and its standard error) as the product of two dependent correlations: $k = r_{xu} \times r_{yu}$, where r_{xu} is the correlation coefficient between X and U and r_{yu} is the correlation coefficient between Y and U .¹ Afterward, Frank derived a threshold at which the impact of a confounding variable would alter the statistical inference about the predictor if the confounding variable were controlled in the regression model. That is, the initial inference is preserved if the impact of any potential confounding variable does not exceed the corresponding impact threshold for a confounding variable (ITCV) of a given, statistically significant predictor.

Because the ITCV is expressed as the product of two correlations, its metric is directly interpretable in terms of known trends in the social sciences (Cohen & Cohen, 1983). For example, when the ITCV exceeds .25, the observed correlations must be greater than .5, (assuming that the correlations are equal; cf. Frank, 2000, Equation 18), considered a large effect in the social sciences to alter the original inference. ITCV's less than .01 suggests that the component correlations each be less than .1, small effects, to alter the initial inference.

But though the ITCV has an interpretable metric, there is still doubt as to the likelihood of observing a given impact value that exceeds the ITCV in a particular application. For this scenario Frank (2000) drew on a reference distribution, drawing on the observed impacts of existing covariates (defined by the product of two correlations between the covariate and the predictor of interest and the covariate and the outcome), as a basis for evaluating the likelihood of observing an impact value as extreme as the ITCV.

Since the true distribution of the product of two correlations is unknown, Frank (2000) suggested an approximation based on Fisher z transformations of the correlations, and then applied Aroian and colleagues' approximation to the product of two normal variables (Aroian, 1947; Aroian, Taneja, & Cornwell, 1978; Cornwell, Aroian, & Taneja, 1978; Meeker, Cornwell, & Aroian, 1981; Meeker & Escobar, 1994), to define the reference distribution of the impacts of existing covariates. Unfortunately, this doubly asymptotic result is tenuous (see Pan, 2002, p. 60). Thus, the purpose of this present study is to employ a more accurate approximation to the distribution of the product of two dependent correlation coefficients based on Pan (2002) in extending Frank's ITCV to a probability index and then presenting the expression for the index using the more accurate theoretical reference distribution.² We then illustrate the applications of the probability index with an empirical example regarding the effect of father's occupation on educational

attainment (Featherman & Hauser, 1976). In the conclusion, we acknowledge challenges to the logic of the use of the reference distribution.

A Probability Index of the Robustness of a Causal Inference

Suppose we are interested in making a causal inference from the following linear regression model:

$$Y = \beta_0 + \beta_x X + \varepsilon, \quad (1)$$

with the corresponding null hypothesis regarding the coefficient β_x :

$$H_0: \beta_x = 0 \text{ versus } H_1: \beta_x \neq 0.$$

We know that a t -value under the null hypothesis H_0 is

$$t_0 = \frac{r_{xy}}{\sqrt{\frac{1 - r_{xy}^2}{n - 2}}}, \quad (2)$$

where r_{xy} is the sample correlation between X and Y and n is the sample size.

Concerns about a confounding variable suggest that there be a second model that includes a potential confounding variable U :

$$Y = \beta_0 + \beta_x X + \beta_u U + \varepsilon. \quad (3)$$

Then, from Frank (2000), the t value under the null hypothesis H_0 for β_x with respect to Equation 3 is

$$t_u = \frac{r_{xy} - r_{xu}r_{yu}}{\sqrt{\frac{1 - r_{xy}^2 - r_{xu}^2 - r_{yu}^2 + 2r_{xy}r_{xu}r_{yu}}{n - 3}}}, \quad (4)$$

where r_{xu} and r_{yu} are the sample correlations between X and U and between Y and U , respectively.

In order to know how robust the statistical inference about the predictor is to the impact of the confounding variable, we want to know the likelihood that we will retain the primary statistical inference that rejects H_0 , when U is in Equation 3. In other words, for a particular study with an observed t value t_0 ,³ which is larger than the t -critical value t_α at level α ,⁴ we are particularly interested in the following probability W :

$$W = P(t_u > t_\alpha | t_0), \text{ for } t_0 > t_\alpha. \quad (5)$$

Probability Index of the Robustness of Causal Inference

W is the likelihood that we will preserve the original inference that rejects the null hypothesis H_0 , when the potential confounding variable is in Equation 3, given t_0 . Thus, W is the probability of retaining a causal inference (PRCI) given inclusion of a confounding variable. Note that the logic of Equation 5 assumes an inference has already been made—that there is a value t_0 that is larger than t_α . That is, as with sensitivity analysis, we are entering the inference process post-facto. But, the PRCI focuses on the impact on the inference for the regression coefficient rather than on the change in the size of the regression coefficient in sensitivity analysis (e.g., Lin, Psaty, & Kronmal, 1998; Rosenbaum, 1986; see Frank, 2000, pp. 161–163, for a review on the relationship to sensitivity analysis and size rule).

The relationship between the PRCI and sensitivity analysis can be also illustrated by plotting the t value, t_u , versus the impact of the confounding variable, $k = r_{xu} \times r_{yu}$, as shown in Figure 1. The curve was plotted using Frank’s formula (Frank, 2000, Equation 5, p. 154) for $r_{xy} = .30$ and $n = 84$, the medium size of correlation and the corresponding sample size (cf. Cohen, 1988). As shown in Figure 1, when the impact k is larger than the ITCV, the inference for the regression coefficient is altered by the impact of the confounding variable; when the impact k is smaller than the ITCV, the inference for the regression coefficient is unaltered by the impact of the confounding variable. Here, the PRCI is the probability of k smaller than ITCV.

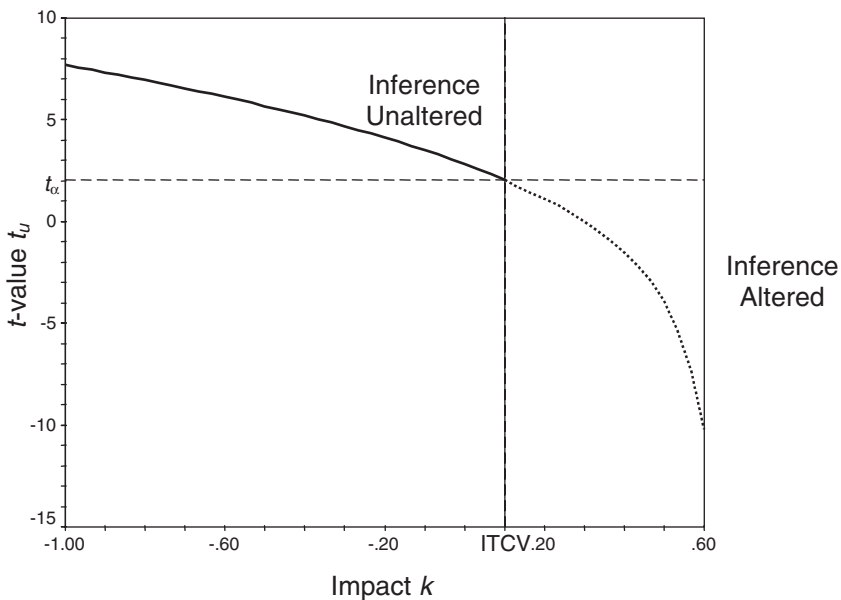


FIGURE 1. *The relationship between the t value, t_u , and the impact of the confounding variable, $k = r_{xu} \times r_{yu}$, for $r_{xy} = .30$ and $n = 84$.*

Note. Since $k = r_{xu} \times r_{yu}$ and since r_{xu} , r_{yu} and r_{xy} are related, when $r_{xy} = .30$, k must be smaller than .65 (cf. Frank, 2000, Equation 5, p. 154).

To develop a computational expression for the PRCI, we employ Frank's (2000) definition of the impact of a confounding variable on the inference for a regression coefficient: $k = r_{xu} \times r_{yu}$. Maximizing the expression for t_u (cf. Equation 4) for a regression coefficient given the constraint $k = r_{xu} \times r_{yu}$ and setting $t_u = t_\alpha$, the t -critical value, given α and the degrees of freedom, Frank defined the impact threshold for a confounding variable (ITCV) on the inference for a regression coefficient as (cf. Frank, 2000, p.155):

$$k = \text{ITCV} = -\frac{t_\alpha^2 + t_\alpha\sqrt{d}}{n-3} + \frac{d + t_\alpha\sqrt{d}}{n-3} r_{xy}, \quad (6)$$

where $d = t_\alpha^2 + n - 3$.

The ITCV indicates a maximum value for the impact of the confounding variable U necessary to make the coefficient $\hat{\beta}_x$ that is statistically significant in Equation 1 become just still statistically significant in Equation 3 (see Frank, 2000, for more information). Thus, when $t_u > t_\alpha$, the coefficient $\hat{\beta}_x$ is significant in Equation 3, which implies that the impact of the confounding variable does not exceed the ITCV. Therefore, the observed value of the PRCI can be obtained as follows in terms of the ITCV:

$$\text{PRCI} = P(K < \text{ITCV}) = \int_{-\infty}^{\text{ITCV}} f(k)dk, \text{ for } t_0 > t_\alpha, \quad (7)$$

where K is the impact of an unmeasured confounding variable with a distribution $f(k)$. Note that the conditional state of t_0 in Equation 5 is implied by the ITCV because the ITCV is a function of r_{xy} (see Equation 6), which is in turn the (inverse) function of t_0 (see Equation 2).

Of course, we cannot always measure confounding variables and, therefore, we cannot always obtain the distribution $f(k)$. That is, there is apparently no empirical basis for obtaining the probability distribution for the impact of an *unobserved* confounding variable U . But Frank (2000) suggested that information from impacts of observed covariates could be used to generate a reference distribution for the impact of the unobserved confounding variable (see Frank, 2000, pp. 172–176, for more information on how to generate the reference distribution). The reference distribution is different from a sampling distribution and we use the term *reference distribution* instead of *sampling distribution* to indicate that the observed estimates are based on the impacts of covariates other than the potentially confounding variable. That is, the impact of the unobserved confounding variable can be characterized by the reference distribution, although it is not actually drawn from the reference distribution. The goal is not to obtain a confidence interval for an observed value of k and then test a null hypothesis, but to evaluate the likelihood of observing k if it can be represented by the impacts of existing covariates.

Probability Index of the Robustness of Causal Inference

Suppose we have m observed covariates Z_1, \dots, Z_m . Corresponding to Equations 1 and 3 we have

$$Y = \beta_0 + \beta_x X + \beta_{z_1} Z_1 + \dots + \beta_{z_m} Z_m + \varepsilon, \quad (8)$$

$$Y = \beta_0 + \beta_x X + \beta_u U + \beta_{z_1} Z_1 + \dots + \beta_{z_m} Z_m + \varepsilon. \quad (9)$$

The impact of each covariate Z_i is then $k_{z_i} = r_{xz_i \cdot z_1 \dots z_{i-1} z_{i+1} \dots z_m} \times r_{yz_i \cdot z_1 \dots z_{i-1} z_{i+1} \dots z_m}$ (where the \cdot indicates partialled for). The distribution $f(k)$ can then be evaluated in terms of the distribution of observed impacts k_{z_i} , $i = 1, \dots, m$.

One could simply evaluate $f(k)$ in terms of its order in the observed distribution. Unfortunately, if we have few covariates, then a simple location of $f(k)$ amidst the observed impacts may provide an imprecise evaluation of the likelihood of observing a given impact as extreme as the ITCV. Frank (2000) proposed using the mean values of the observed partial correlations as sufficient statistics for an approximate distribution (cf. Note 2) against which $f(k)$ could be evaluated empirically. Therefore, the probability W in Equation 5 becomes:

$$W = P(t_u > t_\alpha | t_0, \bar{r}_{xz \cdot z}, \bar{r}_{yz \cdot z}), \text{ for } t_0 > t_\alpha, \quad (10)$$

where $\bar{r}_{xz \cdot z} = \frac{1}{m} \sum_{i=1}^m r_{xz_i z_1 \dots z_{i-1} z_{i+1} \dots z_m}$ and $\bar{r}_{yz \cdot z} = \frac{1}{m} \sum_{i=1}^m r_{yz_i z_1 \dots z_{i-1} z_{i+1} \dots z_m}$, the mean values of the partial correlations pertaining to the observed covariates. Then, $\bar{r}_{xz \cdot z}$ and $\bar{r}_{yz \cdot z}$ are used for estimating the population values ρ_{xu} and ρ_{yu} , respectively, based on the assumption that the impacts of existing covariates represent the impact of the confounding variable. Thus, the computational expression for the PRCI in Equation 7 is still operational, once we obtain the theoretical reference distribution $f(k)$ by estimating ρ_{xu} and ρ_{yu} from the two means $\bar{r}_{xz \cdot z}$ and $\bar{r}_{yz \cdot z}$.

As previously noted, Frank's (2000) theoretical reference distribution based on Fisher z transformation was doubly asymptotic, compromising its accuracy. In contrast, Pan (2002) derived a highly accurate approximation to the product of two dependent correlation coefficients by obtaining the first four moments of the product. Then, Pan applied those moments to the Pearson distribution family (Pearson, 1895), obtaining an approximate distribution of the product of two dependent correlation coefficients as a Pearson Type I distribution (see Appendix A).⁵ Ultimately, this yields a reference distribution generated from the measured covariates for the impact of the unmeasured confounding variable. In the next section, we present an application of the PRCI regarding educational attainment, applying Pan's result to generate the reference distribution to inform the causal inference.

An Illustration

As shown in Featherman and Hauser's table (1976, Table 3, p. 469), they estimated the effect of family background, e.g. father's occupation, on educational attainment as .051 with a standard error of .002. From this, Featherman and Hauser concluded that father's occupation had an effect on educational attainment. Recently, Sobel

(1998) argued that both family background and educational attainment are affected by father's education which was not controlled for in the analysis. That is, father's education is a potential confounding variable for the causal relationship between father's occupation and educational attainment (see Figure 2 for an illustration).

Sobel's critique is represented in Figure 2. Begin with the standard representation of the relationship between X and Y , referred to as $\hat{\beta}_x$, associated with the arrow at the top of the figure. Then introduce the concern for the confounding variable in terms of the relationships associated with the confounding variable, r_{xu} and r_{yu} . The impact is then expressed in terms of the arrows emerging from r_{xu} and r_{yu} and converging to represent the product $r_{xu} \times r_{yu}$ which then impacts on $\hat{\beta}_x$. Frank (2000) obtained an ITCV of .228 for father's occupation, indicating that the component correlations (partialled for other covariates) would have to be greater than .477 to alter the inference.

The values of .477 are on the order of large correlations by social science terms (Cohen & Cohen, 1983), but are they likely to be observed in this type of situation? To address this question we use the distribution of the impacts of observed covariates reported in Featherman and Hauser's Table 3 supplemented by covariates listed in correlation matrices reported by Duncan, Featherman and Duncan (1972) for relationships observed in various samples and data sets. Covariates included number of siblings, importance of getting ahead, brother's education, and father in farming. Data sets included the Duncan OCG study and the Family Growth in Metropolitan America study. An empirical distribution of the estimated impacts of the 14 covariates is shown in Figure 3.

To approximate the theoretical reference distribution, we first need to estimate the coefficients of skewness and kurtosis of the reference distribution. We could estimate the coefficients for the reference distribution directly from the sample moments of the product $k_{z_i} = r_{xz_1 z_2 \dots z_{i-1} z_{i+1} \dots z_m} \times r_{yz_1 z_2 \dots z_{i-1} z_{i+1} \dots z_m}$. But, we only

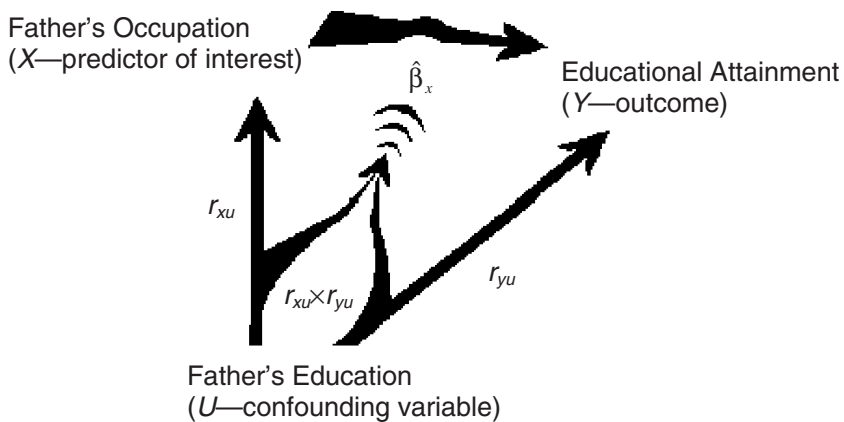


FIGURE 2. *Father's education as a potential confounding variable for the causal relationship between father's occupation and educational attainment.*

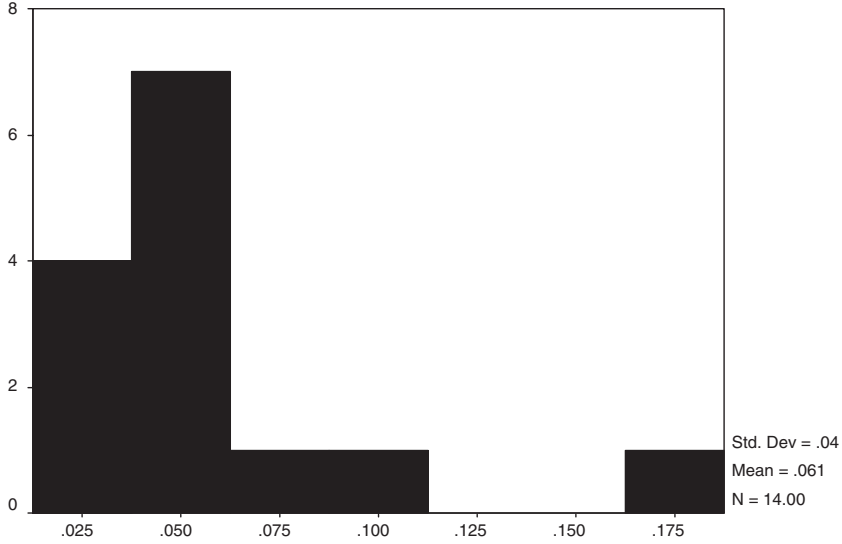


FIGURE 3. An empirical distribution of the impacts of the 14 covariates on father's occupation.

have 14 covariates and the sampling error for the sample moments would be large. Thus, we will estimate the coefficients of skewness and kurtosis for the reference distribution from Pan's (2002) formulae (5), (12), (12'), (12''), and (13).

First, we have the estimated population correlations ρ_{xu} , ρ_{yu} , and ρ_{xy} as follows:

$$\hat{\rho}_{xu} = \bar{r}_{xz \cdot z} = \frac{1}{14} \sum_{i=1}^{14} r_{xz_i \cdot z_1 \dots z_{i-1} z_{i+1} \dots z_{14}} = .235;$$

$$\hat{\rho}_{yu} = \bar{r}_{yz \cdot z} = \frac{1}{14} \sum_{i=1}^{14} r_{yz_i \cdot z_1 \dots z_{i-1} z_{i+1} \dots z_{14}} = .260;$$

$$\hat{\rho}_{xy} = \bar{r}_{xy \cdot z_1 \dots z_{14}} = .325.$$

Substituting those estimated population values into Pan's formulae (5), (12), (12'), (12''), and (13) gives us the estimated values of the coefficients of skewness and kurtosis as $b_1 = .0073$ and $b_2 = 2.995$, respectively. Also, from Equation 14 in Pan, we obtain $\hat{\kappa} = -.1726 < 0$, verifying that the reference distribution can be approximated by a Pearson Type I distribution (cf. Elderton & Johnson, 1969, p. 49). Then, following the formulae in Appendix A (cf. Pan, 2002, p. 26–27), we obtain the estimated distribution function $f(k)$ for $k = r_{xu} \times r_{yu}$ as follows with the graphical presentation in Figure 4:

$$f(k) = 8 \left(1 + \frac{k}{.64} \right)^{114.33} \left(1 - \frac{k}{1.45} \right)^{257.53}, \quad -.64 \leq k \leq 1.45.$$

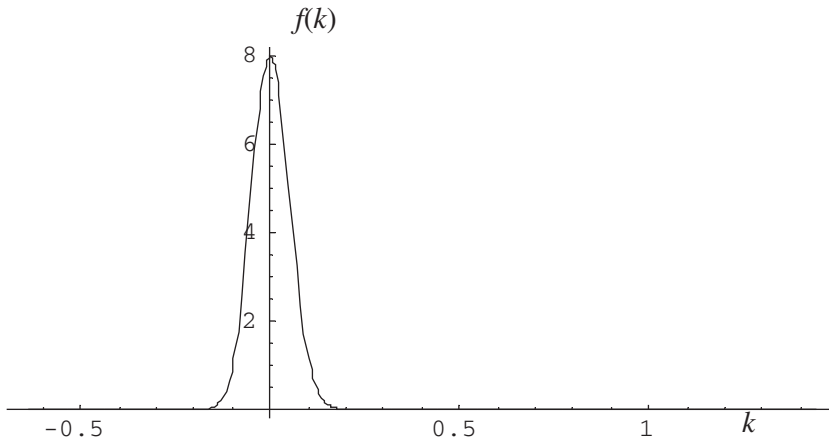


FIGURE 4. The estimated distribution function for $k = r_{xu} \times r_{yu}$.

In contrast to the discrete representation in Figure 3, Figure 4 shows a theoretically continuous function defined by estimates of four moments. Of course, one could use the empirical distribution to make inferences about ITCV. For example, one could compare ITCV to the largest value in the empirical distribution, or to the combined impact of several covariates, or to the overall impact. The advantage of the theoretical distribution is that it combines information across the empirical values to generate a continuous representation. This is especially critical for describing the tails of the distribution for which there is little information in the discrete representation, and yet we suspect the tails extend beyond the discrete cutoffs of .025 and .175, as is shown in Figure 4.

Given the ITCV for father's occupation as .228 (Frank, 2000), we obtain⁶ the
$$PRCI = \int_{-\infty}^{\text{ITCV}} f(k)dk = \int_{-\infty}^{.228} f(k)dk = .999992$$
 via a numerical integration using *Mathematica* (Wolfram Research Inc., 2000; see Appendix B for the *Mathematica* code). To the extent that the impact of father's education is represented by the impacts of the covariates, we could then conclude that the inference regarding father's occupation on educational attainment is very robust to confounding variables similar to those already included in the model. In other words, it is very unlikely that the impact of another confounding variable, such as father's education, will alter our inference about father's occupation, if the impact is similar to the impacts of covariates already in the model.

Use of the PRCI through the ITCV informs the statistical inference and the corresponding debate regarding the inference of an effect of father's occupation on educational attainment. First, the original metric of the ITCV indicates that the correlations associated with any confounding variable would each have to be large by social science terms such that, if included, the confounding variable would alter the original statistical inference regarding father's occupation. But the PRCI helps

us appreciate the likelihood of observing an impact as extreme as the ITCV if the impacts of existing covariates can be considered representative of the impact of unobserved confounding variables. In this case, the likelihood of observing an impact as extreme as the ITCV is very small, and the PRCI is very large. Therefore the inference that father's occupation causes educational attainment is likely to be robust with respect to the impact of a confounding variable. That is, we may still acknowledge that there may be other factors confounded with father's occupation, but it is unlikely that these factors will completely alter the original statistical and causal inferences. Ultimately though we recognize that factors such as father's education may be confounded with father's occupation (Sobel, 1998), we accept Featherman and Hauser's interpretation of father's occupation as a cause of educational attainment.

As it happens, the impact of father's education can be partially assessed in terms of data in Duncan et al.'s correlation table (1972, Table A.1, p. 263). In these data, the correlation between educational attainment and father's education (r_{yu}) was estimated to be .418 and the correlation between father's occupation and father's education (r_{xu}) was estimated to be .494. The unadjusted impact is .206, and similar unadjusted impacts based on the Appendix in Sewell et al. (1980) are .156 for women and .170 for men. Using Sewell et al., adjusting for number of siblings and farm origin, the correlation for men between educational attainment and father's education (r_{yu}) was estimated to be .29 and the correlation between father's occupation and father's education (r_{xu}) was estimated to be .47. The product gives an impact of .136. The impact for women was slightly smaller.

The adjusted, as well as the unadjusted, impacts are less than the ITCV for father's occupation of .228. Not surprisingly, in Sewell et al.'s Tables 8 and 9 (1980, pp. 570–573), father's occupation has a statistically significant direct effect on educational attainment for men or women when controlling for several background characteristics, including father's education as well as parental income, mother's education, mother's employment, farm/rural origin, intact family, number of siblings, mental ability, and so on. Nonetheless, we are not always fortunate enough to have estimates of correlations associated with the confounding variable. The critical point here is that the distribution of the impact of measured covariates can be used to evaluate the PRCI or assess the likelihood that a given impact will exceed the ITCV.

Discussion

Causal inference is a controversial topic in the social sciences, where we are often unable to conduct a randomized experiment or statistically control for all possible confounding variables. In the literature, there are some attempts to deal with the controversy about causal inference, but most approaches have practical or theoretical limitations. Frank (2000) shifted the focus to quantifying the impact of a confounding variable, expanding the paradigm of sensitivity analysis by deriving the single valued threshold, the ITCV, at which the impact would alter an inference (assuming the impact is maximized). Though the metric of ITCV is interpretable in

terms of the product of two correlation coefficients, little is known in any particular example regarding the likelihood of observing an impact as extreme as the ITCV. Frank suggested using the impacts of observed covariates as a basis for a reference distribution, but Frank's doubly asymptotic theoretical reference distribution is suspect. In this study we applied Pan's (2002) more accurate approximation to the product of two dependent correlations to generate the reference distribution and directly exploit Frank's observation by using the ITCV and the reference distribution in the integral to obtain the probability index PRCI.

Our overall approach can be considered an interpretation of sensitivity analysis with a focus on inference. With sensitivity analysis one would represent a distribution of possible estimates for β_x given a broad set of alternative conditions. In particular, one can observe how sensitive the estimate of β_x is to inclusion of a confounding variable. Ultimately sensitivity analyses help researchers develop nuanced interpretations of the effect of x on y , given the distribution of possible estimates of β_x .

Drawing on Frank's representation of the impact of a confounding variable through a single parameter, k , allows us to extend sensitivity analysis in two key ways. First, because of the one-to-one correspondence between k and β_x we can obtain the value of k required to change β_x a given amount or reduce β_x below a given threshold (see Frank, 2000, p. 182). Accounting for the impact of the confounding variable on the standard error of β_x , we then determine the value of k that would make t_u just statistically significant. This then defines the impact threshold for β_x (ITCV). By identifying the threshold for altering inference, the PRCI focuses attention on the conditions necessary to alter the single inference already made, rather than implying a distribution of possible value impacts and corresponding inferences as in sensitivity analysis. Thus the PRCI draws on statistical inference to extract a single value from the distribution of β_x as in a sensitivity analysis.

Second, if sensitivity were defined by multiple parameters it would be highly complex to evaluate the threshold relative to impacts of existing covariates. But given a relatively small number of covariates and applying Pan's approximation we can generate a reference distribution for k . We can then use the reference distribution to interpret the ITCV for β_x as a probability index PRCI.

Where the ITCV indicates the impact necessary to alter a statistical inference, the PRCI is the *probability* of observing such an impact. Probability was introduced to causation through statements such as: "Assuming there is no common cause of A and B, if A happens then B will happen with increased probability p " (Davis, 1988; Einhorn & Hogarth, 1986; Suppes, 1970). But because probable causes do not absolutely link cause and effect, probable causes are open to challenges of alternative explanations associated with confounding variables. This has motivated many theories of causation (Dowe, 1992; Mackie, 1974; Reichenbach, 1956; Salmon, 1984, 1994; Sober, 1988). Alternatively, the PRCI provides a probabilistic response to challenges to "probable causes." In the Featherman and Hauser example, $\text{PRCI} = .999992$ or the likelihood of observing an impact greater than the ITCV was less than .00001. Thus there is only a very small probability

that an alternate cause would account for the observed relationship between father's occupation and educational attainment (assuming the impacts of observed covariates represent the impact of the unobserved covariate). In circumstances such as these, one needs not contort theory nor conduct experiments to defend the interpretation of a coefficient as an effect. This is consistent with how we approach causal inference from a cognitive perspective drawing on the philosophical positions of multiple causes and probable causes (Einhorn & Hogarth, 1986; Mackie, 1974; Meehl, 1978; Mill, [1843] 1973).

But, how small must be the likelihood of an impact exceeding the ITCV (or how large must be the PRCI) to conclude that an inference is robust? It may be tempting to set a cut-off value for the PRCI. Instead we offer the following guideline for interpreting the magnitude of the PRCI. Specifically, if $PRCI > .95$, this indicates that the probability of sustaining the original inference is large and we can say that the statistical inference is very robust with respect to concerns about confounding variables. If $.8 < PRCI \leq .95$, the statistical inference is fairly robust, but we may still need to check some possible confounding variables, and we should interpret the causal inference regarding X with caution. When $PRCI \leq .8$, we would claim that the inference is not robust and researchers must consider the possibility that the inference is not robust with respect to a confounding variable. Note that .95 and .8 for the PRCI are arbitrary, as is .05 for the significance level or .1, .3, and .5 for small, medium, and large effect sizes. Researchers can make their own judgments based on what is studied.

Using of the PRCI does not absolutely rebut a skeptic's claim that the inclusion of a confounding variable would alter the inference of a regression coefficient. Rather, it allows researchers to quantitatively assess the skeptic's claim that the impact of the confounding variable is large enough to alter an inference (see Cordray, 1986, for a similar, albeit non-quantitative, argument). If the skeptic's arguments are not compelling, one can more strongly argue that a statistically significant coefficient is indicative of a causal relationship (although the size of the effect may still be undetermined). In this sense, causal inferences are neither absolutely affirmed nor rejected, but are statements that are asserted and debated (e.g., Abbott, 1998, pp. 164, 168; Cohen & Nagel, 1934; Einhorn & Hogarth, 1986; Gigerenzer, 1993; Sober, 1988; Thompson, 1999).

Of course, causal inference cannot be asserted based on statistical criteria alone (see Dowe's, 1992, critique of Salmon, 1984). A statistical relation combines with a general theory of causal processes (e.g., Salmon, 1984, 1994) as well as a theory of a specific causal mechanism (see McDonald, 1997; Scheines et al., 1998; Spirites et al., 1993) to establish what Suppes (1970) described as a *prima facie* cause. Although Featherman and Hauser (1976) focused on gender differences, they generally argued that family background and resources could provide opportunities to pursue status attainment (including education). This theory combines with the statistically significant coefficients to establish family background as a *prima facie* cause of educational attainment. *Prima facie* causes are then separated into spurious causes and genuine causes depending on whether the effect can be attributed

to a confounding variable. In the Featherman and Hauser example, the PRCI of .999992, far larger than .95, for father's occupation is consistent with the assertion that father's occupation is a genuine cause of educational attainment.

Following Frank (2000) we have defined the PRCI relative to significance tests. Led by Jacob Cohen, many have recently questioned the use of, or called for abandoning, significance tests (Cohen, 1990, 1994; Gigerenzer, 1993; Hunter, 1997; Oakes, 1986; Schmidt, 1996; Sohn, 1998),⁷ although many of the arguments are not new (Bakan, 1966; Carver, 1978; Meehl, 1978; Morrison & Henkel, 1970; Rozenboom, 1960). Those against significance testing argue that using an arbitrary cut-off to evaluate a null hypothesis inaccurately represents the data and falsely dichotomizes decision making. Instead we should use confidence intervals to represent the lack of certainty in our belief about our data, power analysis to assess the probability of a Type II error, and effect sizes to represent the magnitude of effects. Those in favor of significance testing respond that making policy and determining courses of treatment require binary decisions, that an α level can be agreed upon for making such a decision, and that the conservative stance of an unknown relationship being nil accurately represents resistance to the implementation of a new program or treatment (Abelson, 1997; Chow, 1988; Cortina & Dunlap, 1997; Frick, 1999; Harris, 1997; Wainer, 1999).

The key point here is that the PRCI represents a middle ground. Based on the significance test, the framework of the PRCI applies to binary decisions. But like the confidence interval, the PRCI contextualizes a given probability value. Like the effect size, the numerical value of the PRCI indicates an aspect of the strength of the relationship between X and Y ; the stronger the relationship between X and Y , the greater must be the robustness of the causal inference to a confounding variable. In fact, use of the PRCI is very much in the spirit of the recent guidelines for statistical methods in psychology journals, whose authors, including many of the most prominent statisticians in the social sciences, declined to call for a ban on statistical tests (Wilkinson et al., 1999, pp. 602–603). Instead the report recommends results of statistical tests be reported in their full context. And the PRCI is part of that context. Perhaps it is best to consider the PRCI like other statistical tools that should be used based on the consideration of the researcher, referee, and editor (Grayson, 1998).

Most generally, we presented the PRCI in the context of statistical inference because contemporary theories of causation provide a sound basis for social scientists to use statistical inference. For all intents and purposes, unmeasurable differences among people force social scientists to accept probable causes and statistical relationships just as theoretical uncertainty in physical measurement forces physical scientists to accept the same (e.g., Suppes, 1970). Philosophers of science then turned to probabilistic and statistical relations as essential and irreplaceable aspects causality (Salmon, 1998). Nonetheless for social science to progress we must recognize that a statistical inference is not a final proclamation (e.g., Hunter, 1997; Rozenboom, 1960; Sohn, 1998). This caution is consistent with Fisher's qualified interpretation of the p value (see Gigerenzer, 1993, p. 329). Therefore instead

of abandoning statistical inference we should expand upon it to recognize the limitations of the inference and the robustness of the inference with respect to alternative explanations. It is in this vein that we developed the PRCI and that we use it here.

Some researchers may be uncomfortable with the use of measured covariates to generate a reference distribution for the impact of an unknown confounding variable. But we acknowledge that this use of the reference distribution is only as valid as is the set of covariates on which it is based, which is no different from any other inference from a sample that must be representative of the population. In this light, the impacts of existing covariates represent important information by which to evaluate the PRCI.

Another way to verify the legitimacy of the use of the reference distribution to evaluate the PRCI is to assess the sensitivity of the PRCI to deviations from the assumption that the impact of an unknown confounding variable follows the reference distribution generated from the impacts of measured covariates. In the present article, we assume that the impact of an unknown confounding variable can be represented by the impacts of measured covariates, even if the impact of an unknown confounding variable is not actually drawn from the reference distribution. Thus, it is desirable to examine the sensitivity of the PRCI to various assumptions about the impact of an unknown confounding variable.

In the empirical example pertaining to educational attainment, we obtained the observed correlation $r_{xy} = .325$. Thus, for this sensitivity analysis for the PRCI, we fix the population correlation ρ_{xy} at .325 and hypothesize that the unobservable correlations ρ_{xu} and ρ_{yu} could take small (.10), medium (.30), and large (.50) values, according to Cohen (1988). Then, the corresponding PRCI values, referred as to PRCI*, are listed in Table 1. As can be seen in Table 1, we do not include every possible pair of .10, .30, and .50, because ρ_{xu} and ρ_{yu} are symmetric in the mathematical expressions of the reference distribution. Therefore, we have removed duplicate cases. In addition, without loss of generality we can change the sign of the relevant variable to have all correlations positive. Therefore, there are only six pairs of positive correlations used for this sensitivity analysis.

TABLE 1
Sensitivity Analysis for PRCI at $\rho_{xy} = .325$

Hypothetical ρ_{xu} values ^a	Hypothetical ρ_{yu} values ^a	PRCI*	Percent Change ^b
.10	.10	.999988	-.0004
.10	.30	.999989	-.0003
.10	.50	.999995	.0003
.30	.30	.999990	-.0002
.30	.50	.999994	.0002
.50	.50	.999995	.0003

^aThe hypothetical values are the small, medium, and large correlations, according to Cohen (1988).

^b% of change = $100 \times (\text{PRCI}^* - \text{PRCI}) / \text{PRCI}$, where $\text{PRCI} = .999992$.

From Table 1 we can see that the changes of the PRCI* values from the original PRCI value (.999992) are extremely small, within only $\pm 0.0004\%$, which shows that the PRCI is very robust to various assumptions about the impact of an unknown confounding variable. Thus, the PRCI is very insensitive to the deviations from the assumption that the impact of an unknown confounding variable follows the reference distribution that is generated from the impacts of measured covariates. The integration of sensitivity analysis to the use of the PRCI strengthens the methodology of the PRCI proposed in the current study and gains greater understanding of the validity of one's inferences.

Three other caveats regarding the use of the reference distribution in developing the PRCI are critical. First, for the impacts of the measured covariates to have a tractable distribution that is representative of the impact of an unmeasured confounding variable, we assume the impacts of the covariates are homogeneous. That is, we must assume that the impacts of observed and unobserved covariates come from a single distribution. Heterogeneity can be assessed by generating a P-P plot of the observed impacts against the theoretical distribution. When the empirical distribution of the impacts is heterogeneous, researchers need to evaluate the sources of impacts according to substantive theory to identify the sources of large and small impacts. For example, certain types of factors may have stronger causes, or effects may be stronger for certain subpopulations. When the impacts are heterogeneous, it is reasonable, but arguable, to use the maximum impact to obtain a reference distribution.

Second, one may be also concerned about the influence of small values of population correlations on obtaining the value of the PRCI through Equation 7. The fact that Pan's (2002) approximation to the distribution of the product of two dependent correlation coefficients is comparatively poor for small values of ρ_{xu} and ρ_{yu} becomes more of a concern if the partial correlations of observed covariates with the outcome and with the predictor of interest are smaller than their zero-order correlations, which usually occurs in the social sciences because the covariates are often correlated with one another. On the other hand, in the case of confounding, given the generally negative relationship between the t value for the predictor of interest and the product of the correlations with respect to the covariates, when the impacts of the covariates are small, the inferences about the predictor of interest through the t test are more likely to be robust. Thus, we are more likely to retain the primary inference when impacts are small, although we may have some difficulty in characterizing the distribution for the small, partialled impacts of covariates. In other words, the poor approximation for small correlations would only result in more conservative decisions.

Third and last, note that the current approach assumes that all dependent and independent variables are measured without error. This concern does not apply directly to the confounding variable which is assumed to be perfectly measured to maximize impact. But, it does apply to the distribution of impacts of covariates used to generate the reference distribution. To the extent that the covariates are unreliably measured, their impacts will underestimate their true impacts. When reliabilities are known, a correlation disattenuation is recommended. That is, one

can conduct all analyses on a correlation matrix that has been adjusted for attenuation. It is especially important to use a correlation disattenuation when the impact of covariates comprises two partial correlations, because partial correlations produce underestimated small impacts.

Appendix A

Let X , Y , and Z be trivariate normal variables. Based on Pan (2002), the distribution of the product of two dependent correlation coefficients, $r_{xz} \times r_{yz}$, can be approximated by a Pearson Type I distribution, and the density function is (cf. Elderton and Johnson, 1969; Kendall & Stuart, 1977)

$$f(k) = f_0 \left(1 + \frac{k}{a_1}\right)^{m_1} \left(1 - \frac{k}{a_2}\right)^{m_2}, \quad -a_1 \leq k \leq a_2; \quad \frac{m_1}{a_1} = \frac{m_2}{a_2},$$

where $k = r_{xz} \times r_{yz}$, and

$$f_0 = \frac{a_1^{m_1} a_2^{m_2}}{(a_1 + a_2)^{m_1 + m_2 + 1}} \cdot \frac{1}{B(m_1 + 1, m_2 + 1)},$$

$$a_1 + a_2 = \frac{1}{2} \sqrt{\mu_2 [\beta_1 (s + 2)^2 + 16(s + 1)]},$$

$$s = \frac{6(\beta_2 - \beta_1 - 1)}{6 + 3\beta_1 - 2\beta_2},$$

and m_1 and m_2 are given by

$$\frac{s - 2}{2} \pm \frac{s(s + 2)}{2} \sqrt{\frac{\beta_1}{\beta_1 (s + 2)^2 + 16(s + 1)}},$$

with $m_1 < m_2$ if $\mu_3 > 0$ and $m_1 > m_2$ if $\mu_3 < 0$. Note that β_1 and β_2 are the coefficients of skewness and kurtosis.

Appendix B

The following is the *Mathematica* code for computing the PRCI:

$$\text{PRCI} = \int_{-\infty}^{\text{ITCV}} f(k) dk = \int_{-\infty}^{.228} f(k) dk = .999992.$$

```
In[1]:= b1 = 0.0073
        b2 = 2.995
        mu2 = .002482
```

Pan and Frank

Out[1]= 0.0073

Out[2]= 2.995

Out[3]= 0.002482

```
In[4]:= s = 6 * (b2 - b1 - 1) / (6 + 3 * b1 - 2 * b2)
m1 = (s - 2) / 2 - s * (s + 2) * Sqrt[b1 / (b1 * (s + 2) ^ 2 + 16 * (s + 1))] / 2
m2 = (s - 2) / 2 + s * (s + 2) * Sqrt[b1 / (b1 * (s + 2) ^ 2 + 16 * (s + 1))] / 2
asum = Sqrt[mu2 * (b1 * (s + 2) ^ 2 + 16 * (s + 1))] / 2
a1 = asum * m1 / (m1 + m2)
a2 = asum - a1
f0 = (a1 ^ m1) * (a2 ^ m2) / ((a1 + a2) ^ (m1 + m2 + 1))
* Beta[m1 + 1, m2 + 1]
```

Out[4]= 373.862

Out[5]= 114.33

Out[6]= 257.533

Out[7]= 2.08843

Out[8]= 0.642091

Out[9]= 1.44634

Out[10]=7.99786

```
In[11]:= f[k_] := f0 * (1 + k / a1) ^ m1 * (1 - k / a2) ^ m2
PRCI[p_] := NIntegrate[f[k], {k, -a1, p}]
PRCI[.228]
```

Out[13]=0.999992

Converted by *Mathematica* October 13, 2002

Notes

¹In the case of having covariates, the impact is the product of two *partial* correlation coefficients (see Frank, 2000, p. 166).

²According to Frank (2000), the reference distribution for the impact of the unobserved confounding variable can be characterized from the impacts of the observed covariates, where each of the impacts is the product of two dependent correlations (between the covariate and the predictor and between the covariate and the outcome). Since the true distribution of the product of two dependent correlations is unknown, the theoretical reference distribution is also unknown. Instead of estimating the reference distribution directly from the impacts, Frank proposed that the reference distribution could be approximated by transforming each component correlation to an approximate normal (using the Fisher z transformation) and then using the result from Aroian et al. (1978) to approximate the distribution of two dependent normally distributed variables. But, his double approximation is tenuous. On the other hand, Pan (2002) approximated the distribution of the product of two dependent correlations as a Pearson Type I distribution (cf. Appendix A) by applying the first four moments of the product of two dependent correlations to the Pearson distribution family; any deviation from this approximation would be in the fifth moment and higher. Thus, based on Frank's proposal and Pan's approxima-

tion, the theoretical reference distribution can be more accurately approximated by the Pearson Type I distribution using the mean values of the component correlations as the sufficient statistics for the parameters, ρ_{xu} and ρ_{yu} , of the theoretical reference distribution.

³Without loss of generality, $t_0 > 0$ is assumed, because the case of $t_0 < 0$ is symmetrical.

⁴Different t_{α} s might be used for Equations 1 and 3, because the degrees of freedom are different. One is $n - 2$, and $n - 3$ for the other. But, the two t_{α} s are very close, and when n fairly large, they are almost identical. For simplicity, we used the same symbol for the both t_{α} s.

⁵For obtaining the reference distribution, here we tactically use Pan's approximation to the product of two dependent correlation coefficients. A reviewer pointed out that Pan's approximation is for the product of two dependent correlation coefficients when a sample of (X, Y, Z) is drawn from a trivariate *normal* distribution, and that the independent variables X and Z are not necessarily normally distributed in regression analysis. The use of Pan's approximation is reasonable because the focal variables in the reference distribution are correlations, not original variables $X, Y,$ and Z . Also, our approach is as adequate as is the analysis of variables of mixed types using correlation-based procedures, such as regression and structural equation modeling.

⁶There are also two other methods to obtain the PRCI: (a) Programming using Bowman and Shenton's (1979) approach; and (b) Looking up the probability value in Pearson and Hartley's table (1972) with interpolation.

⁷Perhaps it is not surprising that recent concerns about p -values have been voiced more strongly by psychologists than sociologists. Sociologists have *always* had to be more circumspect about interpreting p values because of the limited possibilities for experimentation in sociology.

References

- Abbott, A. (1998). The causal devolution. *Sociological Methods and Research*, 27(2), 148–181.
- Abelson, R. (1997). On the surprising longevity of flogged horses. *Psychological Science*, 8(1), 12–15.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444–455.
- Aroian, L. A. (1947). The probability function of the product of two normally distributed variables. *The Annals of Mathematical Statistics*, 18(2), 265–271.
- Aroian, L. A., Taneja, V. S., & Cornwell, L. W. (1978). Mathematical forms of the distribution of the product of two normal variables. *Communications in Statistics: Part A—Theory and Methods*, 7(2), 165–172.
- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, 66, 423–437.
- Bowden, R. J., & Turkington, D. A. (1984). *Instrumental variables*. Cambridge, UK: Cambridge University Press.

- Bowman, K. O., & Shenton, L. R. (1979). Approximate percentage points for Pearson distributions. *Biometrika*, *66*, 147–151.
- Carver, R. (1978). The case against statistical significance testing. *Harvard Educational Review*, *48*, 378–399.
- Chow, S. L. (1988). Significance test or effect size. *Psychological Bulletin*, *103*(1), 105–110.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304–1312.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997–1003.
- Cohen, J., & Cohen, P. (1983). *Applied multivariate regression/correlation analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.
- Cohen, M. R., & Nagel, E. (1934). *An introduction to logic and the scientific method*. New York: Harcourt Brace.
- Cook, T., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.
- Cordray, D. S. (1986). Quasi-experimental analysis: A mixture of methods and judgment. *New Directions for Program Evaluation*, *31*, 9–27.
- Cornwell, L. W., Aroian, L. A., & Taneja, V. S. (1978). Numerical evaluation of the distribution of the product of two normal variables. *Journal of Statistical Computing and Simulation*, *7*, 123–131.
- Cortina, J. M., & Dunlap, W. P. (1997). On the logic and purpose of significance testing. *Psychological Methods*, *2*, 161–172.
- Davis, W. A. (1988). Probabilistic theories of causation. In J. H. Fetter (Ed.), *Probability and causality* (pp. 133–160). Dordrecht: D. Reidel.
- Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science*, *59*, 195–216.
- Duncan, O. D., Featherman, D. L., & Duncan, B. (1972). *Socioeconomic background and achievement*. New York: Seminar Press.
- Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Review*, *99*(1), 3–19.
- Elderton, W. P., & Johnson, N. L. (1969). *Systems of frequency curves*. London, UK: Cambridge University Press.
- Featherman, D. L., & Hauser, R. M. (1976). Sexual inequalities and socioeconomic achievement in the U.S. 1962–1973. *American Sociological Review*, *41*, 462–483.
- Frank, K. A. (2000). The impact of a confounding variable on a regression coefficient. *Sociological Methods and Research*, *29*(2), 147–194.
- Frick, R. W. (1999). Defending the statistical status quo. *Theory and Psychology*, *9*, 183–189.
- Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for the data analysis in the behavioral sciences: Methodological issue* (pp. 311–339). Hillsdale, NJ: Lawrence Erlbaum.
- Grayson, D. A. (1998). The frequentist façade and the flight from evidential inference. *British Journal of Psychology*, *89*, 325–345.
- Harris, R. J. (1997). Significance tests have their place. *Psychological Science*, *8*(1), 8–11.
- Heckman, J. J. (1997). Instrumental variable: A study of implicit behavioral assumptions used in making program evaluation. *Journal of Human Resources*, *32*, 441–462.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–970.

Probability Index of the Robustness of Causal Inference

- Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models (With discussion). In C. C. Clogg (Ed.), *Sociological methodology* (pp. 449–493). Washington, DC: American Sociological Association.
- Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science*, 8(1), 3–7.
- Jacobs, J. E., Finken, L. L., Griffin, N. L., & Wright, J. D. (1998). The career plans of science-talented rural adolescent girls. *American Educational Research Journal*, 35, 681–704.
- Kendall, M., Sir, & Stuart, A. (1977). *The advanced theory of statistics, Vol. 1, Distribution theory* (4th Ed.). New York: Macmillan.
- Lee, O. (1999). Science knowledge, world views, and information sources in social and cultural contexts: Making sense after a nature disaster. *American Educational Research Journal*, 36, 187–219.
- Lin, D. Y., Psaty, B. M., & Kronmal, R. A. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*, 54, 948–936.
- Mackie, J. (1974). *The cement of the universe*. Oxford, UK: Oxford University Press.
- McDonald, R. P. (1997). Haldane's lungs: A case study in path analysis. *Multivariate Behavioral Research*, 32, 1–38.
- McKim, V. R., & Turner, S. P. (Eds.). (1997). *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences*. Notre Dame, IN: University of Notre Dame Press.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald and the slow progress of soft psychology. *Journal of Counseling and Clinical Psychology*, 46, 806–834.
- Meeker, W. Q., Cornwell, L. W., & Aroian, L. A. (1981). The product of two normally distributed random variables. In W. Kenney & R. Odeh (Eds.), *Selected tables in mathematical statistics* (Vol. VII, pp. 1–256). Providence, RI: American Mathematical Society.
- Meeker, W. Q., & Escobar, L. A. (1994). An algorithm to compute the CDF of the product of two normal random variables. *Communications in Statistics: Part B—Simulation and Computation*, 23(1), 271–280.
- Mill, J. S. [1843] (1973). A system of logic: Ratiocinative and inductive. In J. M. Robson (Ed.), *The collected works of John Stuart Mill* (vols. 7, 8). Toronto: University of Toronto Press.
- Morrison, D. E., & Henkel, R. E. (1970). *The significance test controversy*. Chicago, IL: Aldine.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.
- Okagaki, L., & Frensch, P. A. (1998). Parenting and children's school achievement: A multiethnic perspective. *American Educational Research Journal*, 35, 123–144.
- Pan, W. (2002). The distribution of the product of two dependent correlation coefficients with applications in causal inference. *Dissertation Abstract International*, 62(12), 4137A. (UMI No. AAT 3036725).
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press.
- Pearson, E. S., & Hartley, H. O. (1972). *Biometrika tables for statisticians, Vol. II*. New York: Cambridge University Press.
- Pearson, K. (1895). Contributions to the mathematical theory of evolution, II. Skew variations in homogeneous material. *Philosophical Transactions of the Royal Society of London, Series A*, 186, 343–414.

- Portes, P. R. (1999). Social and psychological factors in the academic achievement of children of immigrants: a cultural history puzzle. *American Educational Research Journal*, 36, 489–507.
- Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of California Press.
- Rosenbaum, P. R. (1986). Dropping out of high school in the United States: An observational study. *Journal of Educational Statistics*, 11, 207–224.
- Rozenboom, W. W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin*, 57, 416–428.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66(5), 688–701.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
- Salmon, W. (1994). Causality without counterfactuals. *Philosophy of Science*, 61, 397–312.
- Salmon, W. (1998). *Causality and explanation*. New York: Oxford University Press.
- SAS Institute Inc. (2001). *The SAS System for Windows (Version 8.02)* [Computer software]. Cary, NC: SAS Institute Inc.
- Scheines, R., Spirites, P., Glymour, C., Meek, C., & Richardson, T. (1998). The TETRAD project: Constraint based aids to causal model specification. *Multivariate Behavioral Research*, 33, 65–117.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1, 115–129.
- Sewell, W. H., Hauser, R. M., & Wolf, W. C. (1980). Sex, schooling, and occupational status. *American Journal of Sociology*, 86, 551–583.
- Sobel, M. E. (1996). An introduction to causal inference. *Sociological Methods and Research*, 24(3), 353–379.
- Sobel, M. E. (1998). Causal inference in statistical models of the process of socioeconomic achievement: A case study. *Sociological Methods and Research*, 27(2), 318–348.
- Sober, E. (1988). The principle of the common cause. In J. H. Fetter (Ed.), *Probability and causality* (p. 211–228). Dordrecht: D. Reidel.
- Sohn, D. (1998). Statistical significance and replicability. *Theory and Psychology*, 8, 291–311.
- Spirites, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search* [Lecture Notes in Statistics 81]. New York: Springer-Verlag.
- Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam, NE: North-Holland.
- Thompson, B. (1999). If statistical significance tests are broken/misused, what practices should supplement or replace them? *Theory and Psychology*, 9, 165–181.
- Wainer, H. (1999). One cheer for null hypothesis significance testing. *Psychological Methods*, 4, 212–213.
- Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanation. *American Psychologist*, 54, 594–604.
- Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observed data. *Annual Review of Sociology*, 25, 659–706.
- Wolfram Research Inc. (2000). *Mathematica (Version 4.1.0.0)* [Computer software]. Champaign, IL: Wolfram Research Inc.
- Woodward, J. (1997). Causal models, probabilities, and invariance. In V. R. McKim & S. P. Turner (Eds.), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 265–315). Notre Dame, IN: University of Notre Dame Press.

Authors

WEI PAN is Assistant Professor of Quantitative Educational Research, Division of Educational Studies, University of Cincinnati, P.O. Box 210002 Cincinnati, OH 45221-0002; wei.pan@uc.edu. His areas of specialization are quantitative research methodology, causal inference, and applications of structural and multilevel modeling in education and behavioral sciences.

KENNETH A. FRANK is Associate Professor of Measurement and Quantitative Methods, Department of Counseling, Education Psychology, and Special Education, Michigan State University, 460 Erickson Hall, East Lansing, MI 48824-1034; kenfrank@msu.edu. His areas of specialization are social networks, causal inference, and hierarchical linear modeling.

Received November 2002

Revision Received January 2003

Accepted February 2003