WORKING
PAPER

Inverse Probability Weighting with Error Prone Covariates

DANIEL F. MCCAFFREY
J. R. LOCKWOOD
CLAUDE MESSAN SETODJI

WR-856-1-DEIES

September 2011

RAND EDUCATION

RAND (2011), *pp.* 1–10

September 2011

Inverse Probability Weighting with Error-Prone Covariates

BY DANIEL F. MCCAFFREY, J.R. LOCKWOOD, AND CLAUDE M. SETODJI

RAND Corporation, 4570 Fifth Avenue, Suite 600, Pittsburgh, Pennsylvania, U.S.A.

SUMMARY

Inverse probability weighted estimates are widely used in applications where data are missing due to nonresponse or censoring and in the estimation of causal effects from observational studies. The current estimators rely on ignorability assumptions for response indicators or treatment assignment, and outcomes, conditional on observed covariates which are assumed to be measured without error. However, measurement error is common in variables collected for many applications. For example, in studies of educational interventions, student achievement as measured by standardized tests is almost always used as the key covariate for removing hidden biases but standardized test scores often have substantial measurement errors for many students. We provide several expressions for a weighting function that can yield a consistent estimator for population means using incomplete data and covariates measured with error.

Some key words: Missing observations, causal inference, propensity score, measurement errors.

1. INTRODUCTION

Inverse probability weighting (IPW) estimates are widely used in applications where data are missing due to nonresponse or censoring (Kang & Schafer, 2007; Lunceford & Davidian, 2004; Scharfstein et al., 1999; Robins & Rotnitzky, 1995; Robins et al., 1995) or in observational studies of causal effects where the counterfactuals cannot be observed (Schafer & Kang, 2008; Bang & Robins, 2005; McCaffrey et al., 2004; Robins et al., 2000). This extensive literature has shown the estimators to be consistent and asymptotically normal under very general conditions, and combining IPW with modeling for the mean function yields "doubly robust" estimates which are consistent and asymptotically normal if either, but not necessarily both, the model for the mean or the model for response or treatment is correctly specified (Kang & Schafer, 2007; Bang & Robins, 2005; Lunceford & Davidian, 2004; van der Laan & Robins, 2003; Scharfstein et al., 1999; Robins & Rotnitzky, 1995; Robins et al., 1995). Recent studies have considered estimation of the response or treatment assignment functions (Harder et al., 2011; Lee et al., 2009; McCaffrey et al., 2004; Hirano et al., 2003) and have shown nonparametric and boosting type estimators work well in simulations and applications.

The consistency and asymptotical normality of IPW estimates generally are guaranteed to hold only with data where response or treatment assignment is independent of the outcomes of interest conditional on a set of observed covariates. The extensive IPW literature only considers settings where the covariates are free of measurement error. However, covariates measured with error are common in applications that might use IPW estimation. For instance, psychological scales from surveys are imperfect measures of the underlying constructs and it is likely that outcomes of interest and treatment assignment or response to a survey depend on the individual's underlying psychological state, not on the error-prone measure. Similarly, achievement tests for school students can have very large errors for some students and again it is clear that future achievement depends on a student's true level of achievement and not on the error-prone test scores. Ignoring the measurement error in the covariates can result in bias in IPW estimates (Steiner et al., 2011).

Although there are no papers specifically proposing IPW estimators with error-prone covariates, error-free values can be viewed as missing data and there are IPW estimators with missing regressor data (Robins et al., 1994; Tan, To appear). However, these methods assume that the outcomes are observed for all units with unobserved covariates, which is not true for the estimation of population means with nonresponding units or causal effects when there is measurement error in the covariates. These methods also rely on modeling the conditional mean of the outcomes, which analysts may want to avoid especially in causal inference problems (Rubin, 2001). D'Agostino & Rubin (2000) develop a method for estimating the propensity scores, which could serve as weights for IPW estimation, when covariates are incompletely observed but their method uses pattern-mixture modeling which does not extend to measurement error problems.

In this note, we discuss an IPW estimator that is consistent in the presence of measurement error in the covariates. In the next section we give our main result and some straightforward extensions. We then discuss methods to apply the estimator and test those via a simulation study.The note ends with a discussion of future work for applying the estimator in practice.

2. IPW ESTIMATOR WITH ERROR PRONE COVARIATES

Let Y_i, $i = 1, \ldots, n$, be the outcome of primary interest obtained from a sample of units from a population, where interest is in the population mean of Y, μ. IPW commonly is applied to two scenarios where the outcomes are observed for only a portion of the sample. The first scenario is missing data due to survey nonresponse, loss-to-follow-up, or censoring in which sampled units failed to provide requested data. The second scenario involves the estimation of the causal effect of a treatment or treatments in which only one of the possible potential outcomes for each study unit is observed, the outcome corresponding to the unit's assigned treatment, and all other potential outcomes are unobserved. Let R_i be a "response" indicator, i.e., $R_i = 1$ if Y_i is observed and $R_i = 0$ is Y_i is unobserved or missing.

For observational studies, each unit in the population has two potential outcomes one that obtains when assignment to treatment Y_{i1} and one that obtains when assigned to the control condition Y_{i0} Rosenbaum & Rubin (1983). Each study unit also has an observed treatment indicator, T_i, with $T_i = 1$ if the unit i received the treatment and $T_i = 0$ if the unit received the control condition. When estimating the mean of the potential outcomes for treatment, $R_i = T_i$ and $R_i = 1 - T_i$ when estimating the mean of the potential outcomes for control. We will use generic term response indicator but the results apply to both nonresponse and observational studies. In observational studies, we only observe Y_{i1} when $T_i = 1$ and Y_{i0} when $1 - T_i = 1$, and we let $Y_{i,obs} = Y_{i1}T_i + Y_{i0}(1 - T_i)$. We assume that the potential outcomes are well-defined and unique for each unit (i.e., the stable unit treatment assumption, holds, Rosenbaum & Rubin, 1983). Although this assumption may not hold in some studies, this is a problem with or without error-prone covariates and so we assume it holds throughout to focus on the issue of weighting with error-prone data.

For each unit there is a covariate U_i which is unobserved and possibly related to both Y_i and R_i. We observe the covariate $X_i = U_i + \xi_i$, as well as Z_i, another covariate measured without error. We make the following assumptions:

ASSUMPTION 1. *Measurement errors, ξ_i, have a known distribution and are independent of Y_i, R_i and Z_i, given U_i.*

ASSUMPTION 2. $0 < pr(R_i = 1 \mid U_i, Z_i) < 1$ *for all sampled units.*

ASSUMPTION 3. Y_i *is independent of* R_i *conditional on* U_i *and* Z_i.

By Assumption 1, X is a surrogate for U and measurement error is nondifferential (Carroll et al., 2006). Assumptions 2 and 3 are similar to conditions of strong ignorability (Rosenbaum & Rubin, 1983) which also requires Assumption 2. However in the context of causal effect estimation for single treatment, strong ignorability requires the conditional joint distribution of both potential outcomes, (Y_{i0}, Y_{i1}), to be independent of treatment. We require only that each potential outcome be marginally independent of treatment assignment conditional on U_i and Z_i, the weak unconfoundness of Imbens (2000). More importantly, independence is conditional on the error-free variable U_i not the observed error-prone covariate X_i.

THEOREM 1. *Let* $p(u, z) = pr(R = 1 \mid u, z)$ *and let* $W(x, z)$ *be a function that for any* z *in the support of* Z *satisfies*

$$E(W(X, z) \mid U = u, Z = z) = \frac{1}{p(u, z)} \tag{1}$$

Let g *be any function of* Y *such that* $E[g(Y)] = \mu_g$ *and* $E[Rg(Y)W(X, Z)]$ *are finite. Then* $E[Rg(Y)W(X, Z)] = \mu_g$.

Proof.

$$\begin{aligned}
E[Rg(Y)\ W\ (X, Z)] &= \text{pr}(R = 1)E[g(Y)W(X, Z) \mid R = 1] \\
&= \text{pr}(R = 1)\int\int\int\int g(y)w(x, z)f(y, u, z, x \mid R = 1)dxdzdudy \\
&= \text{pr}(R = 1)\int\int\int g(y)\left[\int w(x, z)f(x \mid y, u, z, R = 1)dx\right] f(y, u, z \mid R = 1)dzdudy \\
&= \text{pr}(R = 1)\int\int\int \frac{g(y)}{p(u, z)} f(y, u, z \mid R = 1)dzdudy \qquad (2) \\
&= \text{pr}(R = 1)\int\int\int \frac{g(y)}{p(u, z)} \frac{\text{pr}(R = 1 \mid y, u, z)f(y, u, z)}{\text{pr}(R = 1)} dzdudy \qquad (3) \\
&= \int\int\int \frac{g(y)}{p(u, z)} p(u, z)f(y, u, z)dzdudy \qquad (4) \\
&= \int\int\int g(y)f(y, u, z)dzdudy \\
&= \mu_g
\end{aligned}$$

Equation 2 follows from Assumption 1 and Equation 1. Equation 3 follows from Assumption 2, and Equation 4 follows from Assumption 3.

Theorem 1 guarantees that, when in search of $E(g(Y)) = \mu_g$, it can be recovered using a weighted mean of the observed data, even if Y_i is unobserved for a portion of the sample ($R_i = 0$) and covariates are measured with error, provided the weights $W(X, z)$ derived from the error prone X satisfy Equation 1. The next corollary provides an estimator for μ_g.

COROLLARY 1. *A consistent estimator for μ is*

$$\hat{\mu} = \frac{\sum_{i=1}^{n} R_i Y_i W(X_i, Z_i)}{\sum_{i=1}^{n} R_i W(X_i, Z_i)}. \tag{5}$$

Proof. Divide by N in the numerator and denominator of Equation 5. By Theorem 1 using $g(Y) = Y$ and WLLN, the numerator converges in probability to μ. Similarly using Theorem 1 with $g(Y) = 1$, the denominator converges in probability to 1. By Slutsky's theorem the ratio converges in probability to μ. □

REMARK 1. *Theorem 1 naturally extends to settings with multiple error-prone and error-free covariates.*

In a similar manner, for the estimation of causal effects, Corollary 2 provides a consistent estimator of a treatment effect even in the presence of error-prone covariates.

COROLLARY 2. *Let $\mu_t = E(Y_{it})$, $t = 0, 1$, where expectation is for the entire population, and $\delta = \mu_1 - \mu_0$ equal the average treatment effect. Let $W_1(x, z)$ satisfy the conditions of Theorem 1 with $R = T$ and Y_{i1} and $W_0(x, z)$ satisfy the conditions with $R = 1 - T$ and Y_{i0}. Let $0 < pr(T_i = 1 \mid U_i, Z_i) < 1$ for all sampled units and Y_{it}, $t = 0, 1$, be independent of T_i conditional on U_i and Z_i. Then a consistent estimator of δ is*

$$\hat{\delta} = \frac{\sum_{i=1}^{n} T_i Y_{i,obs} W_1(X_i, Z_i)}{\sum_{i=1}^{n} T_i W_1(X_i, Z_i)} - \frac{\sum_{i=1}^{n} (1 - T_i) Y_{i,obs} W_0(X_i, Z_i)}{\sum_{i=1}^{n} (1 - T_i) W_0(X_i, Z_i)}. \tag{6}$$

Proof. By definition, when $T_i = 1$ $Y_{i,obs} = Y_{i1}$ and when $1 - T_i = 1$ $Y_{i,obs} = Y_{i0}$. Hence, by Corollary 1, the first term on the RHS of Equation 6 converges in probability to μ_1 and the second term converges in probability to μ_0. □

REMARK 2. *Letting $W_{odds}(x, z)$ satisfy $E(W_{odds}(X, z) \mid U = u) = p(u, z)/(1 - p(u, z))$ for any z in the support of Z, then using an approach analogous to the proof of Theorem 1, we can show $E[(1 - T)YW_{odds}(X, Z)] = E(Y_0 \mid T = 1) = \mu_{0|1}$ which can be estimated consistently by*

$$\hat{\mu}_{0|1} = \frac{\sum_{i=1}^{n} (1 - T_i) Y_{i0} W_{odds}(X_i, Z_i)}{\sum_{i=1}^{n} (1 - T_i) W_{odds}(X_i, Z_i)}.$$

Here $\mu_{0|1}$ is the counterfactual mean of control outcomes for units that receive treatment. The average effect of treatment on the treated, ATT, (Wooldridge, 2002) can be consistently estimated by

$$\frac{\sum_{i=1}^{n} T_i Y_{i1}}{\sum_{i=1}^{n} T_i} - \hat{\mu}_{0|1}.$$

Theorem 1 states the weights need to be unbiased in the sense that the conditional mean equals the inverse probability weight calculated with the error-free U. However, because the conditional density function $f(y|u, z, x, R = 1)$ depends only on u and z, if we could reweight the cases so that the density of u and z for the weighted cases equals $f(u, z)$, the marginal density for the population, then we could obtain a consistent estimate of the expected value of $g(Y)$. This can be formalized with Theorem 2.

THEOREM 2. *Let $\tilde{W}(x, z)$ be a function such that for every u and z, it satisfies*

$$\int \tilde{w}(x, z) f(u, z \mid x, R = 1) f(x \mid R = 1) dx = f(u, z). \tag{7}$$

Let $W(x,z) = \tilde{W}(x,z)/pr(R=1)$*. Let* g *be any function of* Y *such that* $E[g(Y)] = \mu_g$ *and* $E[Rg(Y)W(X,Z)]$ *are finite. Then* $E[Rg(Y)W(X,Z)] = \mu_g$*.*

Proof. $E[Rg(Y)W(X,Z)] = \text{pr}(R=1)E[g(Y)W(X,Z) \mid R=1]$

$$
\begin{aligned}
E[g(Y)\ W\ (X,Z) \mid R=1] &= \int\int\int\int g(y)w(x,z)f(y,x,z,u \mid R=1)dxdzdudy \\
&= \int\int\int\int g(y)w(x,z)f(y \mid x,z,u,R=1)f(u,z \mid x,R=1)f(x \mid R=1)dxdzdudy \\
&= \int\int g(y)f(y \mid u,z)\left[\int \frac{\tilde{w}(x,z)}{\text{pr}(R=1)}f(u,z \mid x,R=1)f(x \mid R=1)dx\right]dzdudy \quad (8) \\
&= \int\int g(y)f(y \mid u,z)\frac{f(u,z)}{\text{pr}(R=1)}dzdudy \quad (9) \\
&= \frac{\mu_g}{\text{pr}(R=1)}
\end{aligned}
$$

Equation 8 follows from Assumption 3 and Equation 9 follows from Equation 7.

Weights solving Equation 7 of Theorem 2 also solve Equation 1 of Theorem 1 and vice versa. Thus, generating a weighting function can be solved by finding weights which are unbiased for the correct weights or weights to reweight the conditional density given $R=1$ to match the marginal density.

REMARK 3. *A weight function* $\tilde{W}(u,z)$ *satisfies Equation 7 if and only if* $W(x,z) = \tilde{W}(x,z)/pr(R=1)$ *satisfies Equation 1. Suppose* $\tilde{W}(u,z)$ *satisfies Equation 7 and* $W(x,z) = \tilde{W}(x,z)/pr(R=1)$*, then*

$$
\begin{aligned}
f(u,z) &= \int \tilde{w}(x,z)f(u,z \mid x,R=1)f(x \mid R=1)dx \\
&= \int \tilde{w}(x,z)f(x \mid u,z,R=1)f(u,z \mid R=1)dx \\
&= \int \tilde{w}(x,z)f(x \mid u,z)dx f(u,z \mid R=1) \\
&= E(\tilde{W}(X,z) \mid u,z)f(u,z \mid R=1) \\
&= E(\tilde{W}(X,z) \mid u,z)\frac{pr(R=1 \mid u,z)f(u,z)}{pr(R=1)} \\
&= E(W(X,z) \mid u,z)p(u,z)f(u,z).
\end{aligned}
$$

Alternatively, suppose $W(x,z)$ satisfies Equation 1 and let $\tilde{W}(u,z) = W(x,z)pr(R=1)$, then

$$\begin{aligned}\int \tilde{w}(x,z)f(u,z \mid x, R=1)f(x \mid R=1)dx &= pr(R=1)\int w(x,z)f(x \mid u,z,R=1)f(u,z \mid R=1)dx \\ &= pr(R=1)\int w(x,z)f(x \mid u,z)dx f(u,z \mid R=1) \\ &= pr(R=1)E(W(X,z) \mid u,z)f(u,z \mid R=1) \\ &= \frac{pr(R=1)}{pr(R=1 \mid u,z)}f(u,z \mid R=1) \\ &= \frac{pr(R=1)}{pr(R=1 \mid u,z)}\frac{pr(R=1 \mid u,z)f(u,z)}{pr(R=1)} \\ &= f(u,z).\end{aligned}$$

3. ESTIMATION

3·1. Estimators

Corollaries 1 and 2 provide consistent estimators in the presence of error-prone covariates, but they treat the propensity scores and weighting functions, $p(u,z)$ and $W(x,z)$, as known. In practice they will need to be estimated from the observed data. One approach is to first estimate $p(u,z)$ using available techniques for consistent estimation of models with error-prone covariates. For example, if a logistic regression model is used for $p(u,z)$, then the conditional score approach (Carroll et al., 2006) could be used without any assumptions on the distribution of U. If the distribution of U is known or can be well-approximated, then likelihood or Bayesian methods could be used with various parametric models for $p(u,z)$, or gradient boosting corrected for measurement error could be used for nonparametric modeling (Sexton & Laake, 2008).

The next step is to calculate $W(x,z)$ using the estimated $\hat{p}(u,z)$ for the unknown propensity score function. Fourier transforms provide a way to estimate $W(x,z)$ through its inverse properties. Assuming $l(u,z) = 1/p(u,z)$ is an integrable function, then $\phi_l(t,z) = \int_{-\infty}^{\infty} e^{-itu} l(u,z)du$ is its Fourier transform for a given value of z. The Fourier transform is related to the characteristic function for a random variable ξ with density function f_ξ; the characteristic function of ξ is $E(e^{it\xi}) = \phi_{f_\xi}(-t)$. Fourier inversion (Diggle & Hall, 1993) allows for the recovery of $l(u,z)$ with $l(u,z) = (2\pi)^{-1}\int_{-\infty}^{\infty} e^{itu}\phi(t,z)dt$. It is straightforward to show that a weight function

$$W(x,z) = \frac{1}{2\pi}\int e^{itx}\frac{\phi_l(t)}{\phi_{f_\xi}(-t)}dt,$$

satisfies $E(W(X,z) \mid U=u, Z=z) = 1/p(u,z)$, thus potentially providing a way of estimating μ_g via Corollary 1 by plugging in $\hat{p}(u,z)$ for $p(u,z)$ and calculating $\phi_l(t,z)$ by numerical integration.

Alternatively, one can directly estimate an approximation to $W(x,z)$ using simulated data and the estimated $\hat{p}(u,z)$. For example, one could generate a grid of u_j $j=1,\ldots,J$ values and simulate ξ_b $b=1,..,B$ error terms from the density of ξ and set $X_{jb} = u_j + \xi_b$. If we assume that for a given z, $W(x,z) = \sum_{k=1}^{K}\beta_{zk}\eta_k(x)$ for a set of basis functions $\eta_k(x)$, then we can approximate $E(W(x,z) \mid U=u_j, Z=z)$ by $B^{-1}\sum_{b=1}^{B}\sum_{k=1}^{K}\beta_{zk}\eta_k(x_{jb})$=$\sum_{k=1}^{K}\beta_{zk}\bar{\eta}_{kj}$ where $\bar{\eta}_{kj} = B^{-1}\sum_{b=1}^{B}\eta_k(x_{jb})$. We can estimate the β coefficients by fitting a linear regression of $1/\hat{p}(u_j,z)$ on $\bar{\eta}_{kj}$, $k=1,\ldots,K$.

3·2. *Simulation Study*

We conducted a simulation study to test the feasibility of such a two-step approach using the estimator of Corollary 2 for estimating treatment effects. We used maximum likelihood to estimate the parameters of the propensity score function and the function-approximation, simulation approach to obtain solutions to Equation 1.

For each of 100 Monte Carlo iterations, we generated a sample of 1000 independent draws (U, Z_1) from a bivariate normal distribution in which both U and Z_1 are mean zero, variance one and their correlation is 0.3. For each unit in the sample we also drew a Bernoulli random variable Z_2 with mean 0.5, independent of (U, Z_1) and across units, and a Bernoulli treatment indicator T with propensity score $p(u, z_1, z_2) = \mathrm{pr}(T = 1 \mid U = u, Z_1 = z_1, Z_2 = z_2) = G(0.5 + 1.2u + 0.5z_1 - 1.0z_2 + 0.7uz_2)$, where G is the cumulative distribution function (CDF) of Cauchy random variable. We use the Cauchy CDF for the inverse link function rather than the traditional logistic or probit functions to reduce the extreme values among the reciprocals of the $p(u, z_1, z_2)$ and $1 - p(u, z_1, z_2)$ (Ridgeway & McCaffrey, 2007). We also generated error-prone variables $X = U + \xi$, where ξ are iid $N(0, .09)$ so that the reliability of X is roughly 0.92, similar to the reliability of student achievement test scores used for school accountability (Pennsylvania Department of Education, 2010), for instance.

For each iteration, we estimated the coefficients of $p(u, z_1, z_2)$ by maximizing the likelihood via SAS®Proc NLMIXED using the correctly-specified functional form for the propensity score, and the measurement model for X (Rabe-Hesketh et al., 2003), which included that U was normally distributed with unknown variance, estimated from the data, and that X given $U = u$ was $N(u, 0.09)$.

After obtaining an estimate of $\hat{p}(u, z_1, z_2)$, for each observation within an iteration, we approximated $W(x, z_1, z_2)$ by generating a sequence of 800 equally-spaced pseudo-u values between -5 and 5, and for each u we generated 500 random $N(0, 0.09)$ measurement errors and error-prone pseudo-X values. We approximated W by cubic B-spline basis functions with 31 knots and its expectation by the average of the basis functions evaluated at the pseudo-X values, and solved for the unknown parameters of the function approximation by linear regression as described above. We then used the resulting function evaluated at X, Z_1 and Z_2 as the weight for the observation in the estimator given by Equation 6. We refer to this as the "weighting function estimator." The settings for approximating were chosen via exploratory analysis which indicated that the approximation error with these settings was small and did not improve appreciably with larger samples of Us or Xs in the computationally feasible range, or with additional knots in the spline.

For comparison, we also consider three other estimators. The first, which we call the "ideal estimator," is the standard IPW estimator with the unknown propensity scores estimated using the correct functional form and the error-free U. Obviously, this estimator would not be available in applications and is included as benchmark for assessing the cost of measurement error, as this would be the standard estimator had U been observed. The second alternative estimator we consider is standard IPW estimator with the unknown propensity scores estimated using the correct functional form and the error-prone X. We call this the "naïve estimator" and include it because this estimator is commonly used in practice even though it will in general be biased and the estimated coefficients for the propensity score are inconsistent for the true values. The final estimator we consider, which we call the "BLUP estimator," is another IPW estimator with the coefficients of the propensity scores estimated by maximizing the likelihood for treatment and the measurement model for X (i.e., the same coefficient estimates used in deriving the weighting functions) and the propensity scores estimated by evaluating the estimated function at the observed values of Z_1 and Z_2 and the best linear unbiased predictor (BLUP) or the conditional mean of U estimated from the model. The BLUP estimator acknowledges the measurement error in X and attempts to correct for it in the estimation of the propensity scores and using a prediction of the error-free variable.

We evaluated the performance of the four estimators by assessing means and standard deviations of the differences in the weighted means of U, Z_1, and Z_2 for the treatment and control groups. Unbiased estimators will have mean zero. We use bias and variance in the group differences in the covariates to assess the estimators because bias and variance in treatment effects for outcomes interest will generally correspond to group differences in the covariates and by focusing on the covariates we do not need to chose an arbitrary function for the outcomes.

Figure 1 presents the results of the simulation study. As expected, the ideal estimator has small bias for all three covariates. Also, as expected the naïve estimator has substantial bias for U of 0.11 (recall U has variance 1). It has very small bias for Z_1 or Z_2. The BLUP estimator is similar, with notable bias for U and negligible bias for Z_1 and Z_2. The weighting function estimator has small bias for all three covariates as would be expected by Corollary 2 if the weighting functions were known, and which appears to hold when they are approximated using the function-approximation, simulation method with estimated propensity scores. The weighting function estimator is about as efficient as the ideal estimator but both are somewhat less efficient than the naïve and BLUP estimators (about 84 and 92 percent respectively). The weighting function estimator (MSE) has slightly smaller mean square error than the ideal estimator for U and Z_1 but somewhat larger for Z_2. It has much smaller MSE than the naïve or BLUP estimators (less than one fourth or one third as large) for U but somewhat larger values for Z_1 and Z_2 (20 or 14 percent larger) because it is somewhat less efficient.

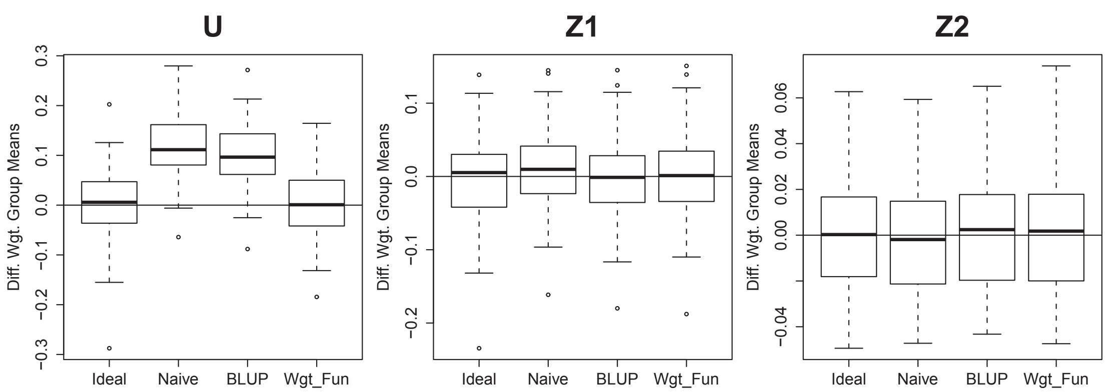

Fig. 1. Box plots of the differences in the weighted means of U, Z_1, and Z_2 for four weighted estimators.

4. DISCUSSION

IPW estimators are increasingly being used in causal model applications and have been shown to work well in simulation studies (Lee et al., 2009). However, these methods can be biased when the covariates are measured with errors, as can other methods that rely on propensity scores such as matching or subclassification (Steiner et al., 2011). Although there are studies of the bias in IPW estimators with error-prone covariates, the literature did not contain any consistent estimators, not even if the distribution of the measurement errors is known. This could be a serious limitation of IPW estimation in applied social science and policy analysis because measurement error is very common in such applications.

We show that such an estimator exists and that the weights are derived by evaluating a weighting function at the observed data, provided the weighting function is unbiased for the true weight conditional on the error-free covariate. We also show that the estimator can be applied to data using a consistent estimator of the propensity score function and simulation-based approximation for the weighting function. The estimator performed nearly as well as the infeasible IPW estimator based on unobservable error-free data and substantially better than a naïve estimator that ignores measurement error and an estimator that attempts to correct for measurement error using the linear predictor of the unobserved error-free covariate.

Our simulation study clearly demonstrates that the results of this paper yield a feasible estimator which can perform well in applications. The conditions of our study are moderately challenging, involving multiple covariates and interactions between the error-prone variable and other covariates in the propensity score model. Future research will need to explore methods for tuning the simulation in the approximation of the weighting function and to develop methods for modeling building for the propensity score function when the functional form is unknown. Common approaches use differences between treatment and control in the distributions of covariates to select variables, terms and the functional form of the propensity score function (Dehejia & Wahba, 1999). However, it is not clear if the distribution of the error prone covariates can proxy for those of the error-free variables.

An additional area of future research is the application of the weighting function estimator in the presence of heteroskedastic measurement error; e.g. if $X_i = U_i + \xi_i$ where $\xi_i \sim N(0, \sigma_i^2)$. In education applications, for example, individual test scores have measurement error variance that is larger for achievement scores in the extremes of the distribution than for those near the middle of the distribution. Theorem 1 and its corollaries extend naturally to this setting by allowing the weighting function to depend on i. However, this case introduces additional practical challenges for both steps of our two-step estimation procedure: heteroskedastic measurement error potentially can degrade the finite-sample performance of consistent estimators of $p(u, z)$ and acceptable approximation of the weighting function may require tailoring to individual observations depending on the size of σ_i^2. Additional research is necessary to determine what combinations of approaches to both problems produce the estimators with the best properties.

ACKNOWLEDGEMENT

This research was supported in part by grants from the Institute of Education Sciences and the National Institute on Drug Abuse.

REFERENCES

BANG, H. & ROBINS, J. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–972.

CARROLL, R., RUPPERT, D., STEFANSKI, L. & CRAINICEANU, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. Baco Raton, FL: Chapman & Hall/CRC.

D'AGOSTINO, JR., R. & RUBIN, D. B. (2000). Estimating and using propensity scores with partially missing data. *Journal of the American Statistical Association* **95**, 749–759.

DEHEJIA, R. H. & WAHBA, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* **94**, 1053–1062.

DIGGLE, P. & HALL, P. (1993). A fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society. Series B (Methodological)* **55**, 523–531.

HARDER, V., STUART, E. & ANTHONY, J. (2011). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods* **15**, 234–249.

HIRANO, K., IMBENS, G. W. & RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71**, 1161–1189.

IMBENS, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika* **87**, 706–710.

KANG, J. & SCHAFER, J. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22**, 523–539.

LEE, B., LESSLER, J. & STUART, E. (2009). Improving propensity score weighting using machine learning. *Statistics in Medicine* **29**, 337–346.

LUNCEFORD, J. K. & DAVIDIAN, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine* **23**, 2937–2960.

MCCAFFREY, D., RIDGEWAY, G. & MORRAL, A. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods* **9**, 403–425.

Pennsylvania Department of Education (2010). Technical Report for the 2010 Pennsylvania System of School Assessment. Prepared by Data Recognition Corporation.

RABE-HESKETH, S., PICKLES, A. & SKRONDAL, A. (2003). Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Statistical Modeling* **3**, 215–232.

RIDGEWAY, G. & MCCAFFREY, D. (2007). Comment: Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22**, 540–543.

ROBINS, J. & ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association* **90**, 122129.

ROBINS, J., ROTNITZKY, A. & ZHAO, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846866.

ROBINS, J., ROTNITZKY, A. & ZHAO, L. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106121.

ROBINS, J. M., HERNAN, M. A. & BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560.

ROSENBAUM, P. & RUBIN, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 4155.

RUBIN, D. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology* **2**, 169–188.

SCHAFER, J. & KANG, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods* **13**, 279–313.

SCHARFSTEIN, D. O., ROTNITZKY, A. & ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (C/R: P1121-1146). *Journal of the American Statistical Association* **94**, 1096–1120.

SEXTON, J. & LAAKE, P. (2008). Logitboost with errors-in-variables. *Comput. Stat. Data Anal.* **52**, 2549–2559.

STEINER, P., COOK, T. & SHADISH, W. (2011). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Journal of Educational and Behavioral Statistics* **36**, 213–236.

TAN, Z. (To appear). Estimating and using propensity scores with partially missing data. *Biometrika* .

VAN DER LAAN, M. & ROBINS, J. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer.

WOOLDRIDGE, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.