A Class of Semiparametric Tests of Treatment Effect Robust to Confounder Classical Measurement Error

Caleb H. Miles, Joel Schwartz, and Eric J. Tchetgen Tchetgen *

Abstract

When assessing the presence of an exposure causal effect on a given outcome, it is well known that classical measurement error of the exposure can reduce the power of a test of the null hypothesis in question, although its type I error rate will generally remain at the nominal level. In contrast, classical measurement error of a confounder can inflate the type I error rate of a test of treatment effect. In this paper, we develop a large class of semiparametric test statistics of an exposure causal effect, which are completely robust to classical measurement error of a subset of confounders. A unique and appealing feature of our proposed methods is that they require no external information such as validation data or replicates of error-prone confounders. We present a doubly-robust form of this test that requires only one of two models to be correctly specified for the resulting test statistic to have correct type I error rate. We demonstrate validity and power within our class of test statistics through simulation studies. We apply the methods to a multi-U.S.-city, time-series data set to test for an effect of temperature on mortality while adjusting for atmospheric particulate matter with diameter of 2.5 micrometres or less (PM2.5), which is known to be measured with error.

Keywords: Causal inference, Climate change, Double robustness, Environmental health, Measurement error, Semiparametric inference

*Caleb H. Miles is Postdoctoral Fellow, Department of Biostatistics, University of California, Berkeley 94720-7358. Joel Schwartz is Professor, Departments of Environmental Health and Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115. Eric J. Tchetgen Tchetgen is Professor, Departments of Biostatistics and Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115.
1. INTRODUCTION

In observational studies across a number of disciplines, it is common to observe variables measured with error. As noted in Cote and Buckley (1987), "Campbell (1988) has gone so far as to say that measurement error (both random error and method effect) and its confounding influences on research findings cannot be avoided." In the field of causal inference, data on covariates are needed to adjust for confounding in order to make inferences with causal interpretations. While a commonly-cited result states that the ordinary least squares (OLS) coefficient estimate of a single variable subject to classical measurement error in a multiple linear regression will merely be attenuated to the null, and hence produce a valid (albeit conservative) statistic of the null hypothesis of no association, the effects of confounders measured with error can be more harmful. Unaccounted for, mismeasured confounders will produce biased effect estimates and invalid hypothesis tests of a treatment effect in even the simplest of settings. Consider a confounder $X^{*}$ that is measured with classical, nondifferential measurement error $\varepsilon^{*}$ such that $X = X^{*} + \varepsilon^{*}$, where $X$ is the value that is actually observed. A multiple linear regression of outcome $Y$ on exposure $A$ and observed confounders $C$ and $X$ will produce a treatment effect estimate that is biased towards the crude (unadjusted) estimate. Consequently, hypothesis tests concerning the effect of $A$ based on this regression may be invalid in the sense that the corresponding type I error rate will often exceed the nominal level.

In this paper, we present a large class of test statistics of the null hypothesis of no conditional average causal effect that maintain validity when a set of continuous confounders are measured with classical, nondifferential measurement error. We consider three different semiparametric models, all of which specify that the conditional mean of the exposure is linear in the error-prone confounders (on the additive, multiplicative, or logit scale). Beyond this specification, our class of test statistics contains three corresponding subclasses consisting of test statistics with nominal type I error rate within (a) a model that specifies the conditional exposure mean as some possibly-nonlinear function of error-free confounders, (b) a model that specifies the conditional outcome mean as some possibly-nonlinear function of the error-free confounders, and (c) a model that specifies that at least one of (a) or (b) holds. Statistics in (c) are said to be doubly robust.

There has been a great deal of interest in methodology for confounders measured with er-
ror. Ogburn and VanderWeele (2012) and Ogburn and Vanderweele (2013) consider settings in which misclassification of a discrete confounder results in imperfect confounding adjustment, in the sense that the average causal effect will be biased in the direction of the crude (unadjusted) estimate. We will focus on settings in which the error-prone confounders are continuous. When instrumental variables (IVs) for such confounders are available, traditional IV estimators can be used to adjust for measurement error in a linear outcome regression model (Amemiya, 1985; Amemiya et al., 1990; Amemiya, 1990; Buzas and Stefanski, 1996; Carroll and Stefanski, 1994; Carroll et al., 2006; Fuller, 2009; Stefanski and Buzas, 1995, among others). Kuroki and Pearl (2014) give an identification result for a total effect in a linear structural equation model with Gaussian errors when at least two proxies of an error-prone confounder are available. Raykov (2012) proposes a propensity-score estimator for the average causal effect under a latent variable model for confounders, in which at least two error-prone congeneric indicators are measured for each latent confounding variable. McCaffrey et al. (2013) propose an inverse probability of treatment weighted (IPTW) estimator with weights that are functions of the error-prone confounders that is consistent for the average causal effect when the distribution of the measurement error is homoscedastic and known or consistently estimated. In addition to IPTW estimators, Lockwood and McCaffrey (2015b) also consider matching estimators, establishing necessary and sufficient conditions for recovering unconfounded matching estimators based on functions of error-prone confounders. However, they show that these are less likely to hold in practice than conditions for IPTW, and do not provide general guidance in estimating appropriate functions to match on. Lockwood and McCaffrey (2015a) propose a simulation-extrapolation estimator that assumes normality and known or consistently-estimated variance of the measurement error. Cochran and Rubin (1973) derive an analytical expression characterizing the bias incurred by continuous confounders measured with classical error in a linear regression model. Under simplifying assumptions, this bias can be corrected provided the reliability ratio is known. Battistin and Chesher (2014) generalize this work to nonparametric models, allowing for identification of the average causal effect of treatment and the effect of treatment on the treated in a sensitivity analysis on a range of possible values for the variance of measurement error.

The latter method fits into a more general body of measurement error research that does not rely on external data. While most traditional measurement-error methods depend on auxiliary
data such as instrumental variables or data from reliability or validation studies, attention has more recently shifted to developing methods not dependent on such data, which can be expensive to collect or simply unavailable. One such class of methods uses “higher-order” moment restrictions to produce identifying estimating equations for parameters of a regression model with covariates measured with error (Bonhomme and Robin 2009; Cragg 1997; Dagenais and Dagenais 1997; Erickson and Whited 2000, 2002; Kapteyn and Wansbeek 1983; Lewbel 1997, 2012; Pal, 1980; Schennach and Hu 2013; Stuart and Kendall 1979, among others). Another existing method, known as deconvolution, uses external knowledge of the measurement-error distribution to recover the density of the error-free variable (Fan 1991; Fan and Truong 1993, among others). Though this idea is attractive in principle, it is very rare that the distribution of measurement error will be known, and convergence rates tend to be too slow for practical use. When identification conditions are not met, it is possible to compute bounds for the parameter of interest (Frisch 1934; Klepper and Leamer 1984; Schennach 2014). Carroll et al. (2006) and Schennach (2012) survey measurement error literature in which they provide a thorough treatment of methods not requiring external data.

This paper contributes both to the literature on confounder measurement error as well as on measurement error methods not requiring external information. Our proposed class of test statistics is of interest in a variety of practical settings in that it requires neither knowledge of the distribution or variance of the measurement error (as in Battistin and Chesher 2014 or in deconvolution), nor any form of external information. We will assume that the measurement error is mean independent of the error-free confounders and the outcome, i.e. \( E(\varepsilon^* | C, Y) = E(\varepsilon^*) \) over the joint support of \( C \) and \( Y \). Otherwise, no other moment restriction is required not already embedded in the assumptions needed to draw causal inferences. In particular, our statistics directly leverage the no-unobserved-confounding assumption needed for identification of the average causal effect, even in the absence of confounder measurement error.

The governing idea of the proposed approach is that under the null hypothesis of no effect of exposure, the assumption of no unobserved confounding renders the outcome an instrumental variable for the association between the true error-prone covariate \( X^* \) and \( A \) adjusting for \( C \). Thus, as documented in the literature on IV methods for measurement error, \( Y \) can be used to obtain a consistent estimator of the association between \( (C, X^*) \) and \( A \) (Amemiya 1985).
Amemiya et al. [1990], Amemiya [1990], Buzas and Stefanski [1996], Carroll and Stefanski [1994], Carroll et al. [2006], Fuller [2009], Stefanski and Buzas [1995], among others). In this paper, we show that estimation of the conditional association between error-prone covariates and exposure can be accomplished jointly with a test of no treatment effect under a unifying framework of a generalized method of moments test based on overidentifying moment restrictions, known in the econometrics literature as a Sargan test (Sargan, 1958), Hansen test, or J-test (Hansen, 1982).

We demonstrate validity of our test statistics in the presence of measurement error in an extensive simulation study, and compare them with standard outcome-regression and g-estimation tests that do not allow for measurement error. In simulation settings, our tests retain validity while the competing tests break down in the presence of measurement error. We also apply our methods to an environmental health data set to test for a causal effect of same-day temperature on mortality in the United States. We conduct a multi-city analysis with daily information on mortality as well as environmental factors including temperature and concentration of particulate matter with diameter of 2.5 micrometres or less (PM2.5). PM2.5 is known to be a confounder and to be measured with error due to the high level of variability of pollution across monitoring stations (Armstrong, 1990, 2004; Bateson et al., 2007; Kioumourtzoglou et al., 2014; Zeger et al., 2000). Temperature is hypothesized to have a causal association with mortality, allowing us to examine our method to test this hypothesis while being robust to confounder measurement error.

2. A CLASS OF PROPENSITY-SCORE-BASED TEST STATISTICS ROBUST TO MEASUREMENT ERROR

To formalize discussion, we define for each $a$ the counterfactual $Y_a$ to be a subject’s outcome had the subject been assigned, possibly contrary to fact, to exposure level $a$. We link these counterfactuals to the observed variables via the consistency assumption (Robins, 1986), which states that if $A = a$, then $Y_a = Y$ with probability one for each level $a$. Suppose we observe a set of covariates $C$ that are measured without error as well as an additional set of covariates $X$ that are measured with classical error, i.e., additive measurement error. The latter are related to their corresponding, unobserved, true value $X^*$ by $X = X^* + \varepsilon^*$, where $\varepsilon^*$ is the measurement error, assumed to be mean independent of $C$ and $Y$. Further, suppose that given $X^*$ and $C$, there is no
unmeasured confounding of the effect of $A$ on $Y$, which can be formalized as follows:

**Assumption 1.** $Y_a \perp A \mid C, X^*$ for each level $a$ (No unmeasured confounding).

Assume that $A$ is continuous; results are generalized to binary and count exposure in Section [7]. We now present a class of test statistics for the null hypothesis $H_0 : E(Y_a \mid C, X^*) = E(Y_0 \mid C, X^*)$ for all $a$. Intuitively, under the stronger sharp null, $Y_a = Y$ w.p. 1 for all $a$, the assumption of no unmeasured confounding implies that $A \perp Y \mid C, X^*$. Furthermore since $X^*$ is a confounder, we have that $Y \not\perp X^* \mid C$. These two statements formally define $Y$ as an instrumental variable for the conditional association between $X^*$ and $A$ given $C$, and therefore can be used to account for measurement error in estimating a model of $A$ given $C$ and $X$ (Carroll et al., 2006). Although $H_0$ is technically weaker than the sharp null, as we will show the essential idea that $Y$ can nonetheless be used under the null to correct for measurement error in the exposure model remains true, despite $Y$ no longer formally being an IV.

We present our first result, which relies on correct specification of a mean regression model for exposure. We will refer to this as the propensity-score model (Rosenbaum and Rubin, 1983). In this vein, consider the semiparametric model $\mathcal{M}_A$ as the set of laws for $(A, X^*, C, Y)$ with sole restriction the parametric model $E(A \mid C, X^*; \alpha) = [1, X^{*T}]g_A(C; \alpha)$, where $g_A = [g_{A,1}(C; \alpha_1), g_{A,2}(C; \alpha_2)^{T}]^{T}$ is a known function of $C$ indexed by the unknown parameter $\alpha = [\alpha_1^{T}, \alpha_2^{T}]^{T}$, where $g_{A,1}$ is real valued and $g_{A,2}$ has the same dimension as $X^*$. We define $p_1$ and $p_2$ to be the dimensions of $\alpha_1$ and $\alpha_2$, respectively, such that $\alpha$ has dimension $p = p_1 + p_2$. We assume throughout that the conditional $X^* - A$ association is linear given $C$, however $g_A(C; \alpha)$, though parametric, can be a nonlinear function of $C$. The case $g_{A,2}(C; \alpha_2) = \alpha_2$ is a constant implies no $X^* - C$ interaction in the model for $A$. Define $\nabla_\alpha$ to be the gradient operator with respect to $\alpha$ and $\mathbb{P}_n$ to be the empirical mean operator.

**Theorem 1.** Let $\ell(C)$ and $m(C)$ be $p + q$-dimensional functions of $C$ for some positive integer $q$, such that the elements of $\ell(C)Y + m(C)$ are linearly independent. Define $U(\alpha) \equiv \{\ell(C)Y + m(C)\} \{A - [1, X^{T}]g_A(C; \alpha)\}$, $\Omega \equiv E\{U(\alpha)U(\alpha)^{T}\}$ and $\hat{U}_n(\alpha) \equiv \mathbb{P}_n U_i(\alpha)$. If $U(\alpha)$ is continuously differentiable, $\nabla_\alpha E\{U(\alpha)\} = E\{\nabla_\alpha U(\alpha)\}$, and $\Omega^{-1} E\{\nabla_\alpha U(\alpha)\}$ has full rank, then for any $\hat{\Omega}_n \overset{p}{\to} \Omega$, the test statistic $\chi^2_{rps} \equiv \min_{\alpha} n\hat{U}_n(\alpha)^{T} \hat{\Omega}_n^{-1}\hat{U}_n(\alpha) \overset{d}{\to} \chi^2_q$ under $\mathcal{M}_A$ and $H_0$. 

6
Thus, we have a valid test of no causal effect of treatment which depends on $X^*$ only through the mismeasured covariate, $X$. Intuitively, standard normal equations for the propensity-score model incur bias due to components that include the product of the residual $A - E(A \mid C, X)$ with the error-prone covariate $X$. However, this can be amended by replacing one of these components with the product of the residual with $Y$. Under $H_0$, this product will form an unbiased estimating equation. Additional unbiased estimating functions can be added simply by multiplying this latter product with any function of $C$, and hence these can be used to form a valid Sargan test statistic. Thus, a simple form of the test in Theorem 1 with $q = 1$ could use $\nabla_\alpha g_A(C; \alpha)$ augmented by $Y$ and the product of $Y$ with an element of $C$ in place of $\ell(C)Y + m(C)$, for instance. In order to ensure linear independence of the elements of $\ell(C)Y + m(C)$, the interaction function $g_{A,2}(C, \alpha_2)$ in the propensity-score model cannot be saturated in $C$.

The following iterative procedure can be used to compute the variance-estimate component $\hat{\Omega}_n$:

initialize $\tilde{\alpha} := \arg\min_\alpha \hat{U}_n(\alpha)^T \hat{\Omega}_n(\alpha)$;
set $\hat{\Omega}_n := \mathbb{P}_n \{U(\tilde{\alpha})U(\tilde{\alpha})^T\}$;
do
  set $\tilde{\alpha} := \arg\min_\alpha \hat{U}_n(\alpha)^T \hat{\Omega}_n^{-1} \hat{U}_n(\alpha)$;
  set $\hat{\Omega}_n := \mathbb{P}_n \{U(\tilde{\alpha})U(\tilde{\alpha})^T\}$;
while convergence not reached;

The first two steps are in fact sufficient for asymptotic validity, however iterating generally improves finite-sample performance. Alternatively, a continuous updating approach can be used, in which $\hat{\Omega}_n$ is indexed by $\alpha$, and $n\hat{U}_n(\alpha)^T \hat{\Omega}_n(\alpha)^{-1} \hat{U}_n(\alpha)$ is minimized in $\alpha$ through both $\hat{U}_n(\alpha)$ and $\hat{\Omega}_n(\alpha)$.

### 3. A CLASS OF DOUBLY-ROBUST TEST STATISTICS

Validity of the test statistic given in the previous section relies on correct specification of $E(A \mid X^*, C) = [1, X^*]^T g_A(C; \alpha)$, however this model may be misspecified. Therefore it is of interest to explore an alternative, potentially more robust approach. Here we present a large class of doubly-robust test statistics. In order to describe this class, let $E(Y \mid C; \gamma) = g_Y(C; \gamma)$ denote
a parametric model for \( E(Y \mid C) \), and consider the semiparametric model \( \mathcal{M}_Y \) with sole restrictions \( E(A \mid C, X^*) - E(A \mid C, X^* = 0) = X^{*T}g_{A,2}(C; \alpha_2) \) and \( E(Y \mid C) = g_Y(C; \gamma) \).

This is a semiparametric model since the association between \( C \) and \( A \) given \( X^* = 0 \) is unrestricted. Further consider the union model \( \mathcal{M}_U \equiv \mathcal{M}_A \cup \mathcal{M}_Y \). We present a class of test statistics for each of these two models, adopting the notation \( \Delta_A(\alpha) = A - [1, X^T]g_{A}(C; \alpha) \) and \( \Delta_Y(\gamma) = Y - g_Y(C; \gamma) \) for the residuals in each model.

**Theorem 2.** Let

\[
U(\alpha_2, \gamma) \equiv \begin{bmatrix} k(C)\Delta_Y(\gamma)\{A - X^Tg_{A,2}(C; \alpha_2)\} \\ S(\gamma) \end{bmatrix},
\]

where \( S(\gamma) \) is a system of estimating equations for \( \gamma \), and \( k(C) \) is a vector-valued function of \( C \) with linearly-independent elements with dimension \( p_2 + q \) for some positive integer \( q \). Suppose \( U(\alpha_2, \gamma) \) is continuously differentiable, \( \nabla_{\alpha_2, \gamma}E\{U(\alpha_2, \gamma)\} = E\{\nabla_{\alpha_2, \gamma}U(\alpha_2, \gamma)\} \), and \( \Omega^{-1}E\{\nabla_{\alpha_2, \gamma}U(\alpha_2, \gamma)\} \) has full rank, where \( \Omega = E\{U(\alpha_2, \gamma)U(\alpha_2, \gamma)^T\} \). Then under \( \mathcal{M}_Y \) and \( H_0 \),

\[
\chi^2_{tor} \equiv \min_{\alpha_2, \gamma} n\hat{U}_n(\alpha_2, \gamma)^T\hat{\Omega}_n^{-1}\hat{U}_n(\alpha_2, \gamma) \xrightarrow{d} \chi^2_q \text{ for any } \hat{\Omega}_n \xrightarrow{p} \Omega.
\]

We also have the result:

**Theorem 3.** Let

\[
U(\alpha, \gamma) \equiv \begin{bmatrix} k(C)\Delta_Y(\gamma)\Delta_A(\alpha) \\ \{\ell(C)Y + m(C)\} \Delta_A(\alpha) \\ S(\gamma) \end{bmatrix},
\]

where \( S(\gamma) \) is a system of estimating equations for \( \gamma \) that is unbiased when \( g_Y(C; \gamma) \) is correctly specified, \( k(C), \ell(C), \) and \( m(C) \) are each vector-valued functions of \( C \) such that \( k(C) \) and \( \ell(C)Y + m(C) \) each consist of linearly-independent elements, and \( \ell \) and \( m \) have dimension \( p_1 \) and \( k \) has dimension \( p_2 + q \) for some positive integer \( q \). Suppose \( U(\alpha, \gamma) \) is continuously differentiable, \( \nabla_{\alpha, \gamma}E\{U(\alpha, \gamma)\} = E\{\nabla_{\alpha, \gamma}U(\alpha, \gamma)\} \), and \( \Omega^{-1}E\{\nabla_{\alpha, \gamma}U(\alpha, \gamma)\} \) has full rank, where \( \Omega = E\{U(\alpha, \gamma)U(\alpha, \gamma)^T\} \). Then under \( \mathcal{M}_U \) and \( H_0 \),

\[
\chi^2_{dr} \equiv \min_{\alpha, \gamma} n\hat{U}_n(\alpha, \gamma)^T\hat{\Omega}_n^{-1}\hat{U}_n(\alpha, \gamma) \xrightarrow{d} \chi^2_q \text{ for any } \hat{\Omega}_n \xrightarrow{p} \Omega.
\]

As before, an appropriate variance estimator \( \hat{\Omega}_n \) can be computed using either an iterated procedure or a continuous-updating approach. An alternative approach would be to first estimate \( \gamma \) by solving \( \mathbb{P}_nS(\gamma) = 0 \), plug this value into \( U(\alpha, \gamma) \) (rendering the \( \gamma \)-estimating-equation
component zero), and use
\[
\hat{\Omega}_n = \frac{1}{n} \sum_{i=1}^{n} \left[ U_i(\hat{\alpha}, \hat{\gamma}) - \left\{ \sum_{j=1}^{n} \nabla_\gamma U_j(\hat{\alpha}, \gamma) \right\} - \left( \sum_{j=1}^{n} \nabla_\gamma S_j(\hat{\gamma}) \right)^{-1} S_i(\hat{\gamma}) \right]\otimes 2
\]
for the variance estimator in the denominator of the test statistic. All estimates of \( \hat{\Omega}_n \) discussed here require that \( k, \ell, \) and \( m \) have not been estimated. Power for both these and the previous tests can be optimized by using appropriate choices of the functions \( \ell(C) \), \( m(C) \), and \( k(C) \) based on the direction of the alternative hypothesis, which we discuss further in Section 6.

4. A SIMULATION STUDY DEMONSTRATING VALIDITY

We now present results from a simulation study drawing samples from the following data generating mechanism. We generate \((Y_0, C)\) under a joint normal model given by \(Y_0 = N(0, 1)\) and \(C = Y_0 + N(0, 1)\), and \(X^*\) and \(A\) under \(X^* = Y_0 + C + Y_0 C + N(0, 1)\) and \(A = C + X^* + N(0, 4)\). To reflect the null hypothesis, we let \(Y = Y_0\). We generate \(X\) from the classical measurement error model \(X = X^* + N\{0, 9(1/\tau - 1)\}\), where \(\tau\) is the reliability ratio, i.e., the ratio of the variability of the true variable \(X^*\) to the variable measured with error \(X\). One may easily verify that Assumption 1 and \(H_0\) are satisfied.

We drew 100,000 samples of size 5000 under four settings with reliability ratios of 50%, 70%, 90%, and 100% (i.e., no measurement error). In each setting, we applied the three testing procedures given in Sections 2 and 3. We compared these tests with two others that ignored the presence of measurement error. The first was an outcome-regression-based test, using the p-value of the regression coefficient for \(A\) when regressing \(Y\) on \(C, X,\) and \(A\) using OLS and using a sandwich variance estimate. Though the outcome model is not correctly specified, the OLS estimate of the coefficient for \(A\) in the absence of measurement error will be unbiased for the slope of \(A\) in \(E(Y \mid A, C, X^*)\) (zero). This is because it is equal to the OLS estimate of the coefficient for the residual obtained from a linear regression of \(A\) on \(C\) and \(X^*\), which is correctly specified. The second comparison test was based on g-estimation (Robins, 1989), using the p-value of the regression coefficient for \(Y\) when regressing \(A\) on \(C, X,\) and \(Y\). All tests used
Table 1: Estimated type 1 error from 100,000 hypothesis tests simulated under the null hypothesis

| Model | Rel. ratio (%) | DR | Robust PS | Robust OR | G-estimation | Standard OR |
|-------|----------------|----|-----------|-----------|--------------|-------------|
| $\mathcal{M}_\cap$ | 50 | 0.0455 | 0.0472 | 0.0485 | 1 | 1 |
| | 70 | 0.0453 | 0.0482 | 0.0497 | 1 | 1.000 |
| | 90 | 0.0533 | 0.0484 | 0.0505 | 0.645 | 0.643 |
| | 100 | 0.0476 | 0.0489 | 0.0517 | 0.0496 | 0.0493 |
| $\mathcal{M}_Y$ | 50 | 0.0452 | 0.714 | 0.0485 | 1 | 1 |
| | 70 | 0.0486 | 0.858 | 0.0497 | 1 | 1.000 |
| | 90 | 0.0527 | 0.939 | 0.0505 | 1 | 0.643 |
| | 100 | 0.0519 | 0.958 | 0.0517 | 0.965 | 0.0493 |
| $\mathcal{M}_A$ | 50 | 0.0463 | 0.0472 | 1 | 1 | 1 |
| | 70 | 0.0495 | 0.0482 | 1 | 1 | 1.000 |
| | 90 | 0.0472 | 0.0484 | 1 | 0.645 | 0.643 |
| | 100 | 0.0497 | 0.0489 | 1 | 0.0496 | 0.0493 |

an $\alpha$ level of 0.05.

All tests with the exception of the standard outcome-regression test were conducted under three different models: the intersection model $\mathcal{M}_\cap$, in which both $g_Y(C; \gamma)$ and $g_A(C; \alpha)$ were correctly specified; $\mathcal{M}_Y$, in which $g_Y(C; \gamma)$ was correctly specified and $g_{A,1}(C; \alpha_1)$ was not; and $\mathcal{M}_A$, in which $g_A(C; \alpha)$ was correctly specified and $g_Y(C; \gamma)$ was not. We used $g_Y(C; \gamma) = \gamma_0 + \gamma_1 C$ and $g_{A,1}(C; \alpha_1) = [1, C] \alpha_1$ for correctly-specified models and $g_Y(C; \gamma) = \gamma_0 + \gamma_1 C^2$ and $g_{A,1}(C; \alpha_1) = [1, C^2] \alpha_1$ for incorrectly-specified models. Though the conditional mean of $Y$ given $C$ and $A$ does not have a simple form, the standard outcome-regression test does not require it to be modeled correctly for validity. Therefore, we show results for the standard outcome-regression test using a misspecified model in all cases for the purposes of comparison.

The index functions for the doubly-robust test used an orthonormalization of $[1, C, C^2, C^3]^T$, with $k(C)$ and $m(C)$ being equal to the first two rows and $\ell(C)$ being equal to the last two. For the robust propensity-score test, we used a Gram-Schmidt orthonormalization of $[1, C, C^2, C^3]^T$ for the function $\ell(C)$, $m(C) = 0$, and $q = 1$. For the robust outcome-regression test, we used $k(C) = [1, C]^T$. The score equations for $\gamma$ in the doubly-robust and robust outcome-regression tests were $S(\gamma) = [1, C^2]^T (Y - \gamma_0 - \gamma_1 C^2)$ under $\mathcal{M}_A$, and $S(\gamma) = [1, C]^T (Y - \gamma_0 - \gamma_1 C)$ otherwise. Results are presented in Table 1.

As expected, the doubly-robust test was approximately valid with correct Monte Carlo type...
error rate under all settings. The robust propensity-score test and robust outcome regression test, on the other hand, were approximately valid under all settings apart from under $\mathcal{M}_Y$ and $\mathcal{M}_A$, respectively. G-estimation and standard outcome-regression tests were not valid in the presence of measurement error, and the standard outcome-regression test was approximately valid in its absence. The $g$-estimation test was approximately valid under no measurement error only when $g_A$ was correctly specified.

5. APPLICATION TO TEST FOR AN EFFECT OF TEMPERATURE ON MORTALITY

As evidence for climate change continues to accumulate, the natural question of whether temperature affects mortality is of increasing importance. While there are many long-term threats posed by rising global temperatures, the immediate effects on mortality also pose a grave public-health concern. When studying this effect, it is vital to control for air pollution as a potential confounder (O’Neill et al., 2003). A common metric of air pollution is PM2.5 concentration, however this is well known to be measured with error (Armstrong, 1990; 2004; Bateson et al., 2007; Kioumourtzoglou et al., 2014; Zeger et al., 2000). In particular, PM2.5 is considered to be contaminated with a mixture of both Berkson error, due to the variability of concentration actually experienced across individuals, and classical error, due to aggregation of measurements across multiple monitoring stations (Kioumourtzoglou et al., 2014; Zeger et al., 2000). The former is benign in the sense that it increases variance but introduces no bias; it is the latter with which we are most concerned. Some studies try to reduce measurement error by using spatial smoothing models (Hoek et al., 2002; Jerrett et al., 2005; Puetz et al., 2009; Sampson et al., 2011; Szpiro et al., 2010; Yanosky et al., 2008), however these rely on geographical data on residency and may induce other forms of error (Gryparis et al., 2009; Sheppard et al., 2012; Szpiro et al., 2011). We implemented our method, and compared it against two methods that ignore the presence of measurement error.

The data set used here consists of time-series mortality data from forty-one U.S. cities measured over the course of 1999 to 2006, though in our analysis, we only considered twenty-four cities with at least eight deaths per day, as cities with lower mortality rates were unlikely to pro-
vide enough power to detect an effect. Data on individual mortality with exact date of death was acquired from the National Center for Health Statistics (NCHS) and from state public health departments (Zanobetti et al., 2009). We excluded accidental deaths (ICD-code 10th revision: V01-Y98, ICD-code 9th revision: 1-799) and deaths of individuals who did not reside in the city in which they died. Temperature data were obtained from the National Oceanic and Atmospheric Administration (NOAA) website, with a city being assigned ambient temperature readings from its nearest monitoring station. PM2.5 data were obtained from the US Environmental Protection Agency’s (EPA) Air Quality System (AQS) database (US EPA 2013). PM2.5 readings were averaged over all monitors in a city whenever multiple readings were available.

On a given day, \( i \), let \( Y_i \) denote the number of deaths, \( A_i \) denote the average temperature in degrees Celsius, \( X_i \) denote the average PM2.5 concentration measurement, and \( C_i \) consist of date, \( t_i \), and dummy variables for day of week. The functions \( g_{A,1}(C; \alpha_1) \) and \( g_Y(C; \gamma) \) in the propensity-score and outcome-regression models used both Fourier bases for time with a period of one year to account for seasonal trends as well as polynomial bases for time to account for secular trends. The dimensions of the Fourier bases were at least four (not including intercept) and the dimensions of the polynomial bases started at zero. Sargan goodness-of-fit tests were used to assess model fit, and more dimensions were added to the bases until the tests no longer rejected at an \( \alpha \) level of 0.10. In particular, we used forms of the test in Theorem 1 with \( \ell(C) = 0 \) (eliminating its power to test for an effect on \( Y \)) and \( m(C) \) equal to \( \nabla_\alpha g_A(C; \alpha) \) augmented by the next two polynomial or Fourier basis functions. Analogous tests were used for the robust outcome-regression model, with the moment functions being equal to the regression residuals multiplied by \( \nabla_\gamma g_Y(C; \gamma) \) augmented by the next two polynomial or Fourier basis functions. Without this step, test rejections could be attributable to model misspecification rather than the presence of a true effect. Both models were linear in these terms as well as the day-of-week dummy variables, and the propensity-score model did not include interaction between \( C \) and \( X \). The outcome-regression model used a log link.

Measurement-error-robust tests from each of the three classes presented in Theorems 1-3 were conducted based on these models. For the robust propensity-score test, we used \([1, t, 0_{p-1}]^T\) as the function \( \ell(C) \), where \( 0_{p-1} \) is a vector of zeroes with length \( p-1 \), and \([0, 0, \nabla_\alpha^T g_{A,1}(C; \alpha_1)]^T\) as the function \( m(C) \). For the robust outcome-regression test, we used \([1, t^n]^T\) as the function
For the doubly-robust test, we used \([1, t]^T\) as the function \(k(C)\), \(\nabla_{\alpha_1} g_{A1}(C; \alpha_1)\) as \(m(C)\), and a vector of zeroes with length \(p_1\) as the function \(\ell(C)\). Thus, in each case \(q = 1\), and we compared resulting test statistics with the corresponding null distribution, \(\chi^2_1\). We used the doubly-robust test for inference, and supplemented our analysis with the other two for an additional check of model fit. As previously mentioned, no model of the relationship between temperature and mortality is needed. This is particularly advantageous in our setting, since this relationship tends to be V or J shaped, and hence not as simple to model.

The two standard statistics considered in the simulation study – based on the g-estimation and standard outcome-regression tests – were implemented. These standard tests do require models for the relationship between temperature and mortality. Consequently, for the standard outcome-regression test, we used a quasi-Poisson model for the outcome-regression model with the deterministic component consisting of the same \(g_Y(C; \gamma)\) described above for the robust test statistics, plus terms for linear-spline basis functions of temperature. We placed a single knot for this spline at 16°C, which is around where vertices of this nonmonotone relationship tend to be. Due to the non-invertibility of this relationship, we conducted two standard g-estimation tests: one based on data from days with mean temperature no higher than 16°C, and the other from days with mean temperature no lower than 16°C. For both, we used OLS estimation of the regression of temperature on the same \(g_A(C; \alpha)\) described above for the robust test statistics plus a linear term for the outcome. Sargan tests were implemented using the gmm package in R. To account for the serially-correlated nature of our data, we used heteroskedasticity and autocorrelation consistent variance estimators from the sandwich package in R for all tests. Results are presented in Table 2.

The doubly-robust test rejected the null hypothesis of no effect of temperature on mortality in two cities: New York, NY and Stamford, CT. For New York, the other measurement-error-robust tests also rejected, whereas these tests did not reject for Stamford. In the latter case, this suggests that the other measurement-error-robust tests were underpowered relative to the doubly-robust test. For New York, there was no indication of substantial attenuation due to measurement error, as all standard tests rejected, apart from the g-estimation test on colder days. For Stamford, on the other hand, it does appear that measurement error may have masked an effect from the
Table 2: P-values of hypothesis tests for an effect of temperature on mortality in U.S. cities

| City                  | Robust PS | Robust OR | Doubly Robust | G-est. ≤16 °C | G-est. ≥16 °C | Standard OR |
|-----------------------|-----------|-----------|----------------|--------------|--------------|-------------|
| Albuquerque, NM       | 0.18      | 0.22      | 0.20           | 0.33         | 0.79         | 0.56        |
| Allentown, PA         | 0.016     | 0.15      | 0.92           | 0.88         | 0.58         | 0.25        |
| Annandale, VA         | 0.52      | 0.070     | 0.84           | 0.91         | 0.39         | 0.98        |
| Baltimore, MD         | 0.94      | 0.29      | 0.82           | 0.30         | 0.20         | 0.0018      |
| Boston, MA            | 0.015     | 0.083     | 0.50           | 0.47         | 0.055        | 0.00073     |
| Elizabeth, NJ         | 0.27      | 0.14      | 0.61           | 0.69         | 0.091        | 0.052       |
| Hartford, CT          | 0.19      | 0.20      | 0.13           | 0.36         | 0.22         | 0.033       |
| Lancaster, PA         | 0.17      | 0.067     | 0.16           | 0.21         | 0.16         | 0.54        |
| Melville, NY          | 0.65      | 0.47      | 0.25           | 0.94         | 0.045        | 0.0504      |
| Middlesex, NJ         | 0.48      | 0.32      | 0.18           | 0.91         | 0.44         | 0.38        |
| New Haven, CT         | 0.65      | 0.067     | 0.79           | 0.73         | 0.016        | 0.0047      |
| New York, NY          | 0.013     | 6.1e-4    | 0.0018         | 0.86         | 0.0010       | 6.7e-10     |
| Newark, NJ            | 0.58      | 0.0085    | 0.48           | 0.18         | 0.36         | 0.039       |
| Paterson, NJ          | 0.66      | 0.13      | 0.33           | 0.47         | 0.11         | 0.64        |
| Philadelphia, PA      | 0.0083    | 0.014     | 0.20           | 0.29         | 0.092        | 7.9e-9      |
| Reading, PA           | 0.64      | 0.019     | 0.69           | 0.45         | 0.79         | 0.49        |
| Richmond, VA          | 0.25      | 0.89      | 0.34           | 0.19         | 0.91         | 0.60        |
| Salt Lake City, UT    | 0.56      | 8.8e-4    | 0.77           | 0.014        | 0.10         | 0.0033      |
| Spokane, WA           | 0.55      | 0.95      | 0.75           | 0.082        | 0.011        | 0.011       |
| Stamford, CT          | 0.060     | 0.37      | 0.034          | 0.95         | 0.23         | 0.54        |
| Upper Marlboro, MD    | 0.11      | 0.046     | 0.84           | 0.91         | 0.10         | 0.51        |
| Washington, DC        | 0.042     | 0.013     | 0.80           | 0.20         | 0.97         | 0.0023      |
| Wilmington, DE        | 0.64      | 0.13      | 0.87           | 0.59         | 0.51         | 0.66        |
| York, PA              | 0.27      | 0.17      | 0.28           | 0.35         | 0.92         | 0.68        |
standard tests, as all standard tests failed to reject.

The robust outcome-regression test rejected for Newark, Reading, Salt Lake City, and Upper Marlboro, while neither the doubly-robust test nor the robust propensity-score test did, suggesting that either the outcome-regression model may not be correctly specified for these cities, or that the other measurement-error-robust tests were relatively underpowered. Similarly, the robust propensity-score test rejected for Allentown and Boston, while neither the doubly-robust nor the robust outcome-regression test did, suggesting that either the propensity-score model may not be correctly specified, or that the other measurement-error-robust tests were relatively underpowered. Both the robust propensity-score and robust outcome-regression tests rejected for Philadelphia and Washington, DC, while the doubly-robust test did not. This may reflect the fact that the doubly-robust test had less power to reject than the other tests in these cities. In a number of cities, one or more of the standard tests rejected when the robust tests did not. Because of possible bias induced by measurement error, we cannot discern whether these rejections are indicative of true effects or merely artifacts.

6. POWER AND ESTIMATION UNDER AN ADDITIVE CAUSAL MODEL

We now consider estimation and testing under the alternative hypothesis of an additive causal model,

\[
E(Y_0 \mid A, C, X^*) = E(Y - \psi_0 A \mid A, C, X^*),
\]

(1)

where \( \psi_0 \) is the average causal effect for a unit change in \( A \) such that \( \psi_0 A = E(Y_A - Y_0 \mid C, X^*) \), and \( \psi_0 \) is the causal parameter of interest. The following discussion can be easily adapted to other structural mean models, however.

We conducted a supplementary simulation study varying the value of \( \psi_0 \) to demonstrate the local power of our proposed test statistics, using the same data generating mechanism as in Section 4 in each of the same measurement error settings, but with \( Y = Y_0 + \psi_0 A \) to encode the alternative hypothesis. The robust propensity-score test was conducted on 1000 samples of size
5000 for each value of $\psi_0$. Results are presented in Figure 1. As expected, we observed trends of increasing power with effect size and reliability ratio. Also as expected, power was approximately 0.05 for $\psi_0 = 0$ in all cases. The test achieved an estimated 80% power at $\psi_0 = 0.06$ when $\tau = 0.5$, at $\psi_0 = 0.05$ when $\tau = 0.7$, at $\psi_0 = 0.025$ when $\tau = 0.9$, and at $\psi_0 = 0.02$ when $\tau = 1$. Power appeared to be tending towards unity as $\psi$ increased in all cases. Similar trends were observed in studies with sample sizes of 1000 and 10,000.

Having posited a model for the effect of $A$ on $Y$, our testing approach can be extended for effect estimation. Let $H(\psi) \equiv Y - \psi A$, and define $H \equiv Y - \psi_0 A$ so that $H = H(\psi_0)$. Then $E(H \mid A, C, X^*) = E(Y_0 \mid A, C, X^*)$. Our claim is now that $H$ (instead of $Y$, since $H_0$ is no
longer assumed) behaves like an instrumental variable for the $X^* - A$ association, controlling for $C$. To see this, first note that by randomization of $A$ within $\{C, X^*\}$, we have $E(H \mid A, C, X^*) = E(Y_0 \mid A, C, X^*) = E(Y_0 \mid C, X^*)$, hence $E(H \mid A, C, X^*) = E(H \mid C, X^*)$. Secondly, $\varepsilon^*$ is mean independent of $H$ by assumption. Finally, since $X^*$ is a confounder of the $A-Y$ association, $X^*$ must be correlated with $Y_0$, and hence $H$, by definition.

Replacing $Y$ with $H(\psi)$ in the equations given in Theorems 1-3 when $q = \dim(\psi)$ (one, under model 1) produces a system of estimating equations for $\psi, \alpha$, and (in the doubly-robust and outcome-regression cases) $\gamma$. Unbiasedness of these functions follows analogously to unbiasedness of the moment equations shown in the proofs for Theorems 1-3 in the Appendix. Thus, unknown parameters can be estimated by solving $P_n U(\psi, \alpha) = 0$, $P_n U(\psi, \alpha_2, \gamma) = 0$, or $P_n U(\psi, \alpha_3, \gamma) = 0$. Under certain regularity conditions, the resulting estimator will be consistent and asymptotically normal, and the estimator produced by solving the doubly-robust estimating equations will have these properties provided at least one of $g_A(C; \alpha)$ or $g_Y(C; \gamma)$ is specified correctly. However, even in simple linear models, the “profile estimating equations” in $\psi$ can be highly nonlinear. By profile estimating equations, we mean the equations obtained by solving $\dim(\alpha) + \dim(\gamma)$ of the estimating equations for $\alpha$ and $\gamma$ implicitly in terms of $\psi$, and plugging these into the additional estimating equations. Identifiability issues can be mitigated by using overidentified estimating equations, i.e., $q > \dim(\psi)$, but solving these may remain challenging, and can result in very unstable estimation.

Optimal choices of functions $\ell(C)$ and $m(C)$ for the robust propensity-score estimator are given in the Appendix. These functions also optimize power when used for hypothesis testing as in Section 2. We note, however, that these functions depend on several additional unknown models, and may not necessarily provide efficiency gain if one or more of these additional models is misspecified, even if $g_A$ and $g_Y$ are correct. The additional variability introduced by these parameters must be accounted for in finding a suitable variance estimator $\hat{\Omega}_n$ for both testing and estimation. This can be accomplished by stacking into $U(\psi, \alpha)$ the score or estimating equations used to estimate the nuisance parameters that estimates of the functions $\ell$ and $m$ depend on.
7. EXTENSIONS TO BINARY AND COUNT EXPOSURES

Under stronger conditions, the test statistics described in this paper can be extended to binary- and count-exposure settings. We will now assume there is no interaction between $C$ and $X^*$ in the propensity-score model and that the measurement error is independent of $X^*, C, A,$ and $Y$. Assuming the propensity-score model

$$\logit \Pr(A = 1 \mid C, X^*) = g_A(C; \alpha_1) + \alpha_2^T X^*, \quad (2)$$

for binary $A$ or

$$\log E(A \mid C, X^*) = g_A(C; \alpha_1) + \alpha_2^T X^* \quad (3)$$

for count $A$ is correctly specified, we have the following analogous result.

**Theorem 4.** Let $\ell(C)$ and $m(C)$ each be vector-valued functions of $C$ with linearly-independent elements and dimension $p + q$, and let $U(\alpha) \equiv \{\ell(C)Y + m(C)\} \exp(-\alpha_2^T X A) [A - \expit\{g_A(C; \alpha_1)\}]$ for (2) if $A$ is binary or $U(\alpha) \equiv \{\ell(C)Y + m(C)\} [A - \exp\{g_A(C; \alpha_1) + \alpha_2^T X\}]$ for (3) if $A$ is a count. Suppose $U(\alpha)$ is continuously differentiable, $\nabla_\alpha E\{U(\alpha)\} = E\{\nabla_\alpha U(\alpha)\}$, and $\Omega^{-1} E\{\nabla_\alpha U(\alpha)\}$ has full rank, where $\Omega = E\{U(\alpha)U(\alpha)^T\}$. Under $H_0$, the test statistic

$$\chi^2_{\text{rps}} \equiv \min_\alpha \nabla_\alpha \hat{U}_n(\alpha)^T \Omega_n^{-1} \nabla_\alpha \hat{U}_n(\alpha) \xrightarrow{d} \chi^2_q \text{ for any } \Omega_n \xrightarrow{p} \Omega.$$ 

The robust outcome-regression and doubly-robust tests in Section 5 can also be extended to a count-exposure setting under (3). As shown in the Appendix, if $\Delta_A(\alpha)$ is redefined as $A \exp\{-\alpha_2 X\} - \exp\{g_A(C; \alpha_1)\}$, Theorem 3 holds as stated, and Theorem 2 holds if $U(\alpha, \gamma)$ is replaced by

$$\begin{bmatrix} k(C) \Delta_Y(\gamma) A \exp\{-\alpha_2 X\} \\ S(\gamma) \end{bmatrix}.$$ 

Unfortunately, we have no such extensions for (2). Upon specifying a structural conditional-mean model for the causal effect of $A$ on $Y$, the moment functions for each of these tests can be easily adapted to form estimating equations for the average causal effect as was shown for the continuous-exposure case.
8. DISCUSSION

We have developed a large class of statistics for the null hypothesis of no causal effect accounting for confounder classical measurement error. This work contributes to the literature on measurement error not only in causal inference, but also in the absence of external information by leveraging causal assumptions to produce a function of the observed data that behaves as an instrumental variable. The tests presented here do not require a causal model to be specified; they only require specification of a conditional mean model of the exposure, outcome, or both. The doubly-robust test only requires one of these models to be correctly specified. The only assumption required beyond those inherent to the causal inference framework (e.g., no unobserved confounding) is that the conditional mean of exposure is linear in the error-prone confounders, and that the part of the propensity-score model multiplying $X^*$ is not saturated in $C$. The latter condition can in fact be relaxed if $H_0$ is replaced by the sharp null, in which case $Y$ is formally an IV. The functions in the estimating equations involving $Y$ can then be nonlinear in $Y$, and hence the number of linearly-independent elements is no longer restricted by the number of possible covariate patterns in $C$.

Sargan tests behave as goodness-of-fit tests, such that when the appropriate models are correctly specified, the tests presented here are powered to detect whether $H_0$ fits the data. However, the tests are also powered to detect model misspecification, so even in the case where there is no causal effect, the tests may reject in case of model misspecification. Thus, when our tests reject, it is prudent to supplement them with a Sargan goodness-of-fit test for each model used as described in Section 5 in order to ensure the results are not due to poor model fit. Unfortunately, these tests are not useful for detecting nonlinearity in $X^*$.

As this method relies on $Y_0$ behaving as an IV for $X^*$, its performance naturally depends on the strength of this conditional association given $C$. Thus, when this association is weak, $Y_0$ will be a weak instrument, and our methods will have reduced power. However, in this case $X^*$ will be a weak confounder which may not need to be accounted for in any case, and hence there is a bias-variance trade-off to be considered when deciding whether to use this method. An empirical test of the conditional association between $Y$ and $X$ given $C$ can be used as a guideline, however we leave a formal treatment of this issue and development of a hybrid method with an unadjusted
(for \(X^*\)) analysis as a potential avenue for future work.

In the multicity application, we tested for an effect of temperature on mortality while accounting for confounding by an error-prone measurement of PM2.5, and discovered evidence of an effect in New York, NY and Stamford, CT. While results from standard tests agreed with our findings in New York, test results in Stamford disagreed, suggesting these standard tests were biased toward the null in this case. In several other cities, the standard tests showed evidence of an effect, while our measurement-error-robust tests did not. This suggested a possible bias in the standard tests due to measurement error resulting in false positives, and that our method may have protected us against making such an error.

The work presented here is not without limitations. Though the tests presented are robust to measurement error of a subset of confounders, at least one true confounder must be measured correctly. While we have managed to avoid the use of parametric models, the assumption of linearity in the error-contaminated confounders could be unrealistic in certain settings. In our data application, no goodness-of-fit test rejected at an \(\alpha\) level of 0.10 after adding sufficiently many basis functions, however we cannot be certain that the goodness-of-fit tests of the final propensity-score models were powered to detect nonlinearities in the error-contaminated confounders. Finally, we did not test for lagged effects of temperature, which could contribute to the effect of temperature on mortality.

One direction for future work would be to re-analyze these data with nearby cities with similar climates clustered into regions, as in [Schwartz et al.](2015). This is sensible since many of the cities examined are quite close, and differences in test results are likely due to differences in power rather than effect size. This approach would greatly improve power to detect effects of temperature in entire regions. As it stands, the present analysis serves as a practical illustration of the application and interpretation of our method.
APPENDIX

Proof of Theorem 1. Let $\bar{\alpha}$ be the true value of $\alpha$. Under $\mathcal{M}_A$,

$$E\{U(\bar{\alpha})\} = E[\{\ell(C)Y + m(C)\}A - E(A \mid C, X^*) - g_{A,2}(C; \bar{\alpha}_2)^T \varepsilon^*]$$

$$= E[\{\ell(C)Y + m(C)\}A] - E[\{\ell(C)Y + m(C) \mid C, X^*\}E(A \mid C, X^*)] - E[\{\ell(C)Y + m(C)\}g_{A,2}(C; \bar{\alpha}_2)^T]E(\varepsilon^*)$$

$$= E[\{\ell(C)Y + m(C)\}A] - E\{E(\{\ell(C)Y + m(C)\}A \mid A, C, X^* \mid C, X^*)\} = 0,$$

since $\varepsilon^*$ is mean independent of $Y$ and $C$, and $E(Y \mid A, C, X^*) = E(Y \mid C, X^*)$, which is implied by $H_0$ and Assumption 1. The regularity conditions on $U(\alpha)$ are sufficient to ensure that $\bar{\alpha}$ is a local minimum of $n\check{U}_n(\alpha)^T \check{\Omega}_n^{-1}\hat{U}_n(\alpha)$, and hence $\alpha$ is locally identified under $H_0$. Then because $\dim\{U(\alpha)\} = \dim(\alpha) + q$, $E\{U(\alpha)\} = 0$ is an overidentified moment restriction, and the statistic $\chi^2_{rps}$ has a limiting distribution of $\chi^2_{q}$.

$\square$

Proof of Theorem 2. Let $\bar{\gamma}$ and $\bar{\alpha}_2$ be the true values of $\gamma$ and $\alpha_2$, respectively. Under $\mathcal{M}_Y$, $E\{S(\bar{\gamma})\} = 0$ and

$$E\left[k(C) \left\{ Y - g_Y(C; \bar{\gamma}) \right\} \left\{ A - g_{A,2}(C; \bar{\alpha}_2)^T X \right\} \right]$$

$$= E\left[k(C) \left\{ Y - E(Y \mid C) \right\} \left\{ A - E(A \mid C, X^*) + E(A \mid C, X^* = 0) - g_{A,2}(C; \bar{\alpha}_2)^T \varepsilon^* \right\} \right]$$

$$= E\left[k(C) \left\{ Y - E(Y \mid C) \right\} A \right] - E\left[k(C)E\left\{ Y - E(Y \mid C) \mid C, X^* \right\} E(A \mid C, X^*) \right]$$

$$+ E\left[k(C)E\left\{ Y - E(Y \mid C) \mid C \right\} \left\{ E(A \mid C, X^* = 0) - g_{A,2}(C; \bar{\alpha}_2)^T E(\varepsilon^*) \right\} \right]$$

$$= E\left[k(C) \left\{ Y - E(Y \mid C) \right\} A \right] - E\left\{k(C)E(\{Y - E(Y \mid C)\}A \mid A, C, X^* \mid C, X^*)\right\} = 0,$$

since $\varepsilon^*$ is mean independent of $Y$ and $C$, and $E(Y \mid A, C, X^*) = E(Y \mid C, X^*)$, which is implied by $H_0$ and Assumption 1. The regularity conditions on $U(\alpha_2, \gamma)$ are sufficient to ensure that $(\bar{\alpha}_2, \bar{\gamma})$ is a local minimum of $n\check{U}_n(\alpha_2, \gamma)^T \check{\Omega}_n^{-1}\hat{U}_n(\alpha_2, \gamma)$, and hence $(\alpha_2, \gamma)$ is locally identified under $H_0$. Then because $\dim\{U(\alpha_2, \gamma)\} = \dim(\alpha_2) + \dim(\gamma) + q$, $E\{U(\alpha_2, \gamma)\} = 0$ is
an overidentified moment restriction, and the statistic $\chi^2_{tov}$ has a limiting distribution of $\chi^2_q$. \hfill \Box

Proof of Theorem 3. Let $\bar{\gamma}$ be the true value of $\gamma$ under $M_Y$ and $\bar{\alpha}$ be the true value of $\alpha$ under $M_A$. Under $M_A$, there exists some $\bar{\gamma}$ such that $E\{S(\bar{\gamma})\} = 0$. That

$$E \left[ \begin{array}{c} k(C) \Delta_Y(\bar{\gamma}) \Delta_A(\bar{\alpha}) \\ \{\ell(C)Y + m(C)\} \Delta_A(\bar{\alpha}) \end{array} \right] = 0$$

follows from the unbiasedness of $U(\alpha)$ shown in the proof of Theorem 1.

Under $M_Y$, $E\{S(\bar{\gamma})\} = 0$ and for any $\alpha_1$,

$$E \left[ k(C) \{Y - g_Y(C; \bar{\gamma})\} \{A - g_{A,1}(C; \alpha_1) - g_{A,2}(C; \bar{\alpha}_2)^T X\} \right]$$

$$= E \left[ k(C) \{Y - E(Y | C)\} \{A - E(A | C, X^*) + E(A | C, X^* = 0) - g_{A,1}(C; \alpha_1) - g_{A,2}(C; \bar{\alpha}_2)^T \varepsilon^*\} \right]$$

$$= E \left[ k(C) \{Y - E(Y | C)\} A \right] - E \left[ k(C) E \{Y - E(Y | C) | C, X^*\} E(A | C, X^*) \right]$$

$$+ E \left[ k(C) E \{Y - E(Y | C) | C\} \{E(A | C, X^* = 0) - g_{A,1}(C; \alpha_1) - g_{A,2}(C; \bar{\alpha}_2)^T E(\varepsilon^*)\} \right]$$

$$= E \left[ k(C) \{Y - E(Y | C)\} A \right] - E \{k(C) E \{E[Y - E(Y | C)] A | A, C, X^* | C, X^*\}\}$$

$$= 0,$$

since $\varepsilon^*$ is mean independent of $Y$ and $C$, and $E(Y | A, C, X^*) = E(Y | C, X^*)$, which is implied by $H_0$ and Assumption 1. Finally, there exists some $\bar{\alpha}_1$ such that

$$E \left\{ \{\ell(C)Y + m(C)\} \Delta_A(\bar{\alpha}_1, \bar{\alpha}_2) \right\} = 0.$$

Thus, under any law in $M_u$, $E\{U(\alpha, \gamma)\} = 0$ has a solution under $H_0$. The regularity conditions on $U(\alpha, \gamma)$ are sufficient to ensure that $(\bar{\alpha}, \bar{\gamma})$ is a local minimum of $n\hat{U}_n(\alpha, \gamma)^T \hat{\Omega}_n^{-1} \hat{U}_n(\alpha, \gamma)$ under $M_A$ and $(\alpha_0, \bar{\alpha}_1, \bar{\alpha}_2, \bar{\gamma})$ is a local minimum under $M_Y$, and hence $(\alpha, \gamma)$ is locally identified under $M_u$ and $H_0$. Then because $\dim[U(\alpha, \gamma)] = \dim(\alpha) + \dim(\gamma) + q$, $E\{U(\alpha, \gamma)\} = 0$ is an overidentified moment restriction, and the statistic $\chi^2_{dr}$ has a limiting distribution of $\chi^2_q$. \hfill \Box

Proof of Theorem 4. First, reparameterize the propensity-score model as

$$\logit \Pr(A = 1 | C, X^*) = \alpha_0 + g_A(C; \alpha_1) + \alpha_2^T X^*,$$
for model (4) or
\[ \log E(A \mid C, X^*) = \alpha_0 + g_A(C; \alpha_1) + \alpha_2^T X^*, \]
for model (5), where \( g_A(0; \alpha_1) = 0 \), such that \( \alpha_0 \) is a scalar intercept and \( \alpha_1 \) has dimension \( p_1 - 1 \).

When \( A \) is binary, for the true value \( \bar{\alpha} \) of \( \alpha \), we have

\[
E \left( \{ \ell(C)Y + m(C) \} \exp \left( -\bar{\alpha}_2^T X^* A \right) \left[ A - \expit \{ \alpha_0^* + g_A(C; \bar{\alpha}_1) \} \right] \right)
= E \left( \{ \ell(C)Y + m(C) \} \exp \left( -\bar{\alpha}_2^T X^* A - \bar{\alpha}_2^T \varepsilon^* A \right) \left[ A - \expit \{ \alpha_0^* + g_A(C; \bar{\alpha}_1) \} \right] \right)
= E \left( \{ \ell(C)Y + m(C) \} \exp \left( -\bar{\alpha}_2^T X^* A \right) E \left\{ \exp \left( -\bar{\alpha}_2^T \varepsilon^* A \right) \mid A, Y, C, X^* \right\} \right. \\
\left. \times \left[ A - \expit \{ \alpha_0^* + g_A(C; \bar{\alpha}_1) \} \right] \right)
= E \left( \{ \ell(C)Y + m(C) \} \exp \left( -\bar{\alpha}_2^T X^* A \right) \exp \left( KA \right) \left[ A - \expit \{ \alpha_0^* + g_A(C; \bar{\alpha}_1) \} \right] \right)
\]

where \( \exp \left( K \right) = E \left\{ \exp \left( -\bar{\alpha}_2^T \varepsilon^* \right) \right\} \) is the moment generating function of \( \varepsilon^* \) evaluated at \( -\bar{\alpha}_2 \).

We then note that the joint density of \( (A, X^*) \) given \( C \) can be expressed as

\[
f(A, X^* \mid C) = \frac{f(X^* \mid A = 0, C) \exp \left( \bar{\alpha}_2^T X^* A \right) f(A \mid X^* = 0, C)}{t(C)}
\]

where \( t(C) \) is a normalizing constant. We then have that

\[
E \left( \{ \ell(C)Y + m(C) \} \exp \left( -\bar{\alpha}_2^T X^* A \right) \exp \left( KA \right) \left[ A - \expit \left( \alpha_0^* + g(C; \bar{\alpha}_1) \right) \right] \right)
= E \int_x \sum_a f(x \mid A = 0, C) \expit \left( \bar{\alpha}_2^T xa \right) f(a \mid X = 0, C) \left( \left[ \ell(C)E \{ Y \mid a, x, C \} + m(C) \right] \exp \left( -\bar{\alpha}_2^T xa \right) \expit \left( Ka \right) \left[ a - \expit \left( \alpha_0^* + g(C; \bar{\alpha}_1) \right) \right] \right) dx
= E \int_x \left[ \ell(C)E \{ Y \mid x, C \} + m(C) \right] f(x \mid A = 0, C) t(C)^{-1} dx \left( \sum_a f(a \mid X = 0, C) \expit \left( Ka \right) \left[ a - \expit \left( \alpha_0^* + g(A(C; \bar{\alpha}_1)) \right) \right] \right) \]
\[
= E \int_x \left( \left[ \ell(C)E \{ Y \mid x, C \} + m(C) \right] f(x \mid A = 0, C) t(C)^{-1} \left( \sum_a \expit \left( \alpha_0^* + g(A(C; \bar{\alpha}_1))a \right) \left[ a - \expit \left( \alpha_0^* + g(A(C; \bar{\alpha}_1)) \right) \right] \right) \right) dx \left( \frac{1 + \expit \left( \alpha_0^* + g(A(C; \bar{\alpha}_1)) \right)}{1 + \expit \left( \alpha_0^* + g(A(C; \bar{\alpha}_1)) \right)} \right) \]
\[
= 0
\]
where \( \alpha_0^* = K + \bar{\alpha_0} \).

When \( A \) is a count, let \( \exp(K) = E \{ \exp(\bar{\alpha}_2^T \varepsilon^*) \} \), i.e., the moment generating function of \( \varepsilon^* \) evaluated at \( \bar{\alpha}_2 \). Under model (5), for the true value \( \bar{\alpha} \) of \( \alpha \), we have

\[
E \left( \{ \ell(C)Y + m(C) \} \left[ A - \exp \{ \alpha_0^* + g_A(C; \bar{\alpha}_1) + \bar{\alpha}_2^T X \} \right] \right) \\
= E \left( \{ \ell(C)Y + m(C) \} A \right) - E \left( \{ \ell(C)Y + m(C) \} \exp \{ \alpha_0^* + g_A(C; \bar{\alpha}_1) + \bar{\alpha}_2^T X \} \right) \\
\times E \left( \exp(\bar{\alpha}_2^T \varepsilon^*) \right) \\
= E \left( \{ \ell(C)Y + m(C) \} A \right) - E \left( \{ \ell(C)Y + m(C) \} \exp \{ \alpha_0^* + g_A(C; \bar{\alpha}_1) + \bar{\alpha}_2^T X \} \right) \\
= E \left( \{ \ell(C)Y + m(C) \} A \right) - E \left( \{ \ell(C)E(Y | C, X^*) + m(C) \} A | C, X^* \right) \\
= E \left( \{ \ell(C)Y + m(C) \} A \right) - E \left( \{ \ell(C)E(Y | C, X^*, A) + m(C) \} A | C, X^* \right) \\
= E \left( \{ \ell(C)Y + m(C) \} A \right) - E \left( \{ \ell(C)Y + m(C) \} A | C, X^*, A \right) | C, X^* \\
= 0,
\]

where \( \alpha_0^* = K + \bar{\alpha}_0 \).

Thus, in either case, \( E \{ U(\alpha_0^*, \bar{\alpha}_1, \bar{\alpha}_2) \} = 0 \) under \( H_0 \). The regularity conditions on \( U(\alpha) \) are sufficient to ensure that \( (\alpha_0^*, \bar{\alpha}_1, \bar{\alpha}_2) \) is a local minimum of \( n\bar{U}_n(\alpha)^T \hat{\Omega}_n^{-1} \bar{U}_n(\alpha) \), and hence \( \alpha \) is locally identified under \( H_0 \). Then because \( \dim \{ U(\alpha) \} = \dim(\alpha) + q \), \( E \{ U(\alpha) \} = 0 \) is an overidentified moment restriction, and the statistic \( \chi_{\alpha}^2 \) has a limiting distribution of \( \chi_q^2 \). \( \square \)

**Theorem 5.** Let \( \hat{\beta}(\ell, m) \) be the estimator solving \( \mathbb{P}_n U(\ell, m; \beta) = 0 \) corresponding to the moment functions in Theorem 1, and define

\[
d(C) \equiv E \{ \Delta(\alpha)^2 H(\psi)^2 | C \} E \{ \Delta(\alpha)^2 | C \} - E \{ \Delta(\alpha)^2 H(\psi) | C \} E \{ \Delta(\alpha)^2 | C \},
\]

\[
\ell^*(C) \equiv d(C)^{-1} \begin{bmatrix}
E \{ \Delta(\alpha)^2 | C \} E \{ \Delta(\alpha) A | C \} \\
- \text{Cov} \{ \Delta(\alpha)^2, H(\psi) | C \} \nabla_{\alpha_1} g_{A,1}(C; \alpha_1) \\
\left[ E \{ \Delta(\alpha)^2 | C \} E(H(\psi) X^T | C) - E \{ \Delta(\alpha)^2 H(\psi) | C \} E(X^T | C) \right] \\
\times \nabla_{\alpha_2} g_{A,2}(C; \alpha_2)
\end{bmatrix},
\]

24
and

\[
m^*(C) \equiv d(C)^{-1} \begin{bmatrix}
- E \{\Delta(\alpha)^2 H(\psi) \mid C\} & E \{\Delta(\alpha) A \mid C\} \\
Cov\{\Delta(\alpha)^2 H(\psi), H(\psi) \mid C\} \nabla_{\alpha_1 g_{A,1}}(C; \alpha_1) \\
- \left[ E \{\Delta(\alpha)^2 H(\psi) \mid C\} E(H(\psi)X^T \mid C) - E \{\Delta(\alpha)^2 H(\psi)^2 \mid C\} E(X^T \mid C) \right] \\
\times \nabla_{\alpha_2 g_{A,2}}(C; \alpha_2)
\end{bmatrix}.
\]

Under \(\mathcal{M}_A\) and the causal model defined by equation (3), \(\hat{\beta}(\ell^*, m^*)\) achieves the minimum asymptotic variance of all estimators in the class of estimators defined by the estimating equations in Theorem 1. The corresponding variance is \(E\left[U(\ell^*, m^*; \beta)U(\ell^*, m^*; \beta)^T\right]\).

Proof. By Theorem 5.3 in Newey and McFadden (1994), if an optimal estimator \(\hat{\beta}(\tilde{\ell}, \tilde{m})\) exists within the class \(\{\tilde{\beta}(\ell, m) : \ell \in \mathcal{L}, m \in \mathcal{M}\}\), the functions \(\ell\) and \(m\) are guaranteed to satisfy

\[
- E \left[ \frac{\partial}{\partial \beta} U(\ell, m; \beta) \right] = E \left[ U(\ell, m; \beta)U(\tilde{\ell}, \tilde{m}; \beta)^T \right]
\]

for all functions \(\ell\) and \(m\), and the estimator will have variance equal to \(E\left[U(\tilde{\ell}, \tilde{m}; \beta)U(\tilde{\ell}, \tilde{m}; \beta)^T\right]\). Thus it suffices to show that \(\ell^*(C)\) and \(m^*(C)\) satisfy (2). We have

\[
- E \left[ \frac{\partial}{\partial \beta} U(\ell, m; \beta) \right] = E \left[ \ell(C)A\Delta(\alpha), \{\ell(C)H(\psi) + m(C)\} \nabla_{\alpha_1 g_{A,1}}(C; \alpha_1), \{\ell(C)H(\psi) + m(C)\} X^T \nabla_{\alpha_2 g_{A,2}}(C; \alpha_2) \right]
\]

where \(V = [V_1, V_2, V_3] \equiv [A\Delta(\alpha), H(\psi)\nabla_{\alpha_1 g_{A,1}}(C; \alpha_1), H(\psi)X^T \nabla_{\alpha_2 g_{A,2}}(C; \alpha_2)]\) and \(W = [W_1, W_2, W_3] \equiv [0, \nabla_{\alpha_1 g_{A,1}}(C; \alpha_1), X^T \nabla_{\alpha_2 g_{A,2}}(C; \alpha_2)]\). If we partition the components of the functions \((\ell^*, m^*)\) into \((\ell^*_1, m^*_1)\), \((\ell^*_2, m^*_2)\), and \((\ell^*_3, m^*_3)\), where \(\ell^*_1\) and \(m^*_1\) are scalar functions, \(\ell^*_2\) and \(m^*_2\) are \(p_1\) dimensional, and \(\ell^*_3\) and \(m^*_3\) are \(p_2\) dimensional, then

\[
E\left[U(\ell, m; \beta)U(\ell^*, m^*; \beta)^T\right] = E\left[\Delta(\alpha)^2\{\ell(C)H(\psi) + m(C)\}\{\ell^*(C)H(\psi) + m^*(C)\}^T\right]
\]

\[
= E\left[\Delta(\alpha)^2\{\ell(C)H(\psi) + m(C)\}\left[\ell^*_1(C)H(\psi) + m^*_1(C), \{\ell^*_2(C)H(\psi) + m^*_2(C)\}^T, \{\ell^*_3(C)H(\psi) + m^*_3(C)\}^T\right]\right].
\]
Thus, we can solve (1) by partitioning it into four independent equations corresponding to the partition of $\ell^*$ and $m^*$: For all $k \in \{1, 2, 3\}$,

$$
E \left[ (\Delta(\alpha))^2 \{\ell(C)H(\psi) + m(C)\} \{\ell_k^*(C)H(\psi) + m_k^*(C)\} \right] = E \{\ell(C)V_k + m(C)W_k\}.
$$

$$
E \left[ (\Delta(\alpha))^2 \{\ell_k^*(C)H(\psi) + m_k^*(C)\} \{H(\psi) | C\} \ell(C) \right.
\left. -E \left[ (\Delta(\alpha))^2 \{\ell_k^*(C)H(\psi) + m_k^*(C)\} | C\right] m(C) \right] = E \{\ell(C)V_k + m(C)W_k\}
$$

$$
\Leftrightarrow
E \left[ (\Delta(\alpha))^2 \{\ell_k^*(C)H(\psi) + m_k^*(C)\} \{H(\psi) - V_k | C\} \ell(C) \right.
\left. -E \left[ (\Delta(\alpha))^2 \{\ell_k^*(C)H(\psi) + m_k^*(C)\} | C\right] m(C) \right] = 0
$$

$$
\Leftrightarrow
E[\Delta(\alpha)^2H(\psi)^2 | C]\ell_k^*(C) + E[\Delta(\alpha)^2H(\psi) | C]m_k^*(C) - E(V_k | C) = 0
$$

$$
E[\Delta(\alpha)^2H(\psi) | C]\ell_k^*(C) + E[\Delta(\alpha)^2 | C]m_k^*(C) - E(W_k | C) = 0
$$

$$
\Leftrightarrow
\begin{bmatrix}
\ell_k^*(C) \\
m_k^*(C)
\end{bmatrix}
= \begin{bmatrix}
E[\Delta(\alpha)^2H(\psi)^2 | C] & E[\Delta(\alpha)^2H(\psi) | C]

E[\Delta(\alpha)^2H(\psi) | C] & E[\Delta(\alpha)^2 | C]
\end{bmatrix}^{-1}
\begin{bmatrix}
E(V_k | C) \\
E(W_k | C)
\end{bmatrix}
$$
given that $\Pr\{d(C) = 0\} = 0$. The second implication can be seen to hold by recognizing the necessity of the first equation when $\ell(C) = E[\Delta(\alpha)^2 \{\ell_k^*(C)H(\psi) + m_k^*(C)\} H(\psi) - V_k | C]$ and $m(C) = 0$ and the necessity of the second equation when $\ell(C) = 0$ and

$$
m(C) = E[\Delta(\alpha)^2 \{\ell_k^*(C)H(\psi) + m_k^*(C)\} - W_k | C].
$$
Thus, (2) is solved by

\[ \ell^*_1(C) = d(C)^{-1} E \left\{ \Delta(\alpha)^2 \mid C \right\} E \left\{ \Delta(\alpha)A \mid C \right\} \]
\[ m^*_1(C) = -d(C)^{-1} E \left\{ \Delta(\alpha)^2 H(\psi) \mid C \right\} E \left\{ \Delta(\alpha)A \mid C \right\} \]
\[ \ell^*_2(C) = -d(C)^{-1} \text{Cov}\left\{ \Delta(\alpha)^2, H(\psi) \mid C \right\} \nabla_{\alpha_1} g_{A,1}(C; \alpha_1) \]
\[ m^*_2(C) = d(C)^{-1} \text{Cov}\left\{ \Delta(\alpha)^2 H(\psi), H(\psi) \mid C \right\} \nabla_{\alpha_1} g_{A,1}(C; \alpha_1) \]
\[ \ell^*_3(C) = d(C)^{-1} \left[ E \left\{ \Delta(\alpha)^2 \mid C \right\} E \left( H(\psi)X^T \mid C \right) \nabla_{\alpha_2} g_{A,2}(C; \alpha_2) \right. \]
\[ - E \left\{ \Delta(\alpha)^2 H(\psi) \mid C \right\} E \left( X^T \mid C \right) \nabla_{\alpha_2} g_{A,2}(C; \alpha_2) \right] \]
\[ m^*_3(C) = -d(C)^{-1} \left[ E \left\{ \Delta(\alpha)^2 H(\psi) \mid C \right\} E \left( H(\psi)X^T \mid C \right) \nabla_{\alpha_2} g_{A,2}(C; \alpha_2) \right. \]
\[ - E \left\{ \Delta(\alpha)^2 H(\psi)^2 \mid C \right\} E \left( X^T \mid C \right) \nabla_{\alpha_2} g_{A,2}(C; \alpha_2) \right] . \]

Proof of validity of the robust outcome-regression test under model (5). Let \( \bar{\gamma} \) and \( \bar{\alpha}_2 \) be the true values of \( \gamma \) and \( \alpha_2 \) respectively. Under \( \mathcal{M}_Y \), \( E \{ S(\bar{\gamma}) \} = 0 \) and

\[ E \left[ k(C) \left\{ Y - g_Y(C; \gamma) \right\} A \exp(-\bar{\alpha}_2^T X) \right] \]
\[ = E \left[ k(C)E \left\{ Y - E \left( Y \mid C \right) \mid C, X^*, A \right\} A \frac{E \left( A \mid C, X^* = 0 \right)}{E \left( A \mid C, X^* \right)} E \left\{ \exp(-\bar{\alpha}_2^T \varepsilon^*) \right\} \right] \]
\[ = E \left[ k(C)E \left\{ Y - E \left( Y \mid C \right) \mid C, X^* \right\} A \frac{E \left( A \mid C, X^* = 0 \right)}{E \left( A \mid C, X^* \right)} \right] e^K \]
\[ = E \left[ k(C)E \left\{ Y - E \left( Y \mid C \right) \mid C, X^* \right\} E \left( A \mid C, X^* \right) \frac{E \left( A \mid C, X^* = 0 \right)}{E \left( A \mid C, X^* \right)} \right] e^K \]
\[ = E \left( k(C)E \left\{ Y - E \left( Y \mid C \right) \mid C \right\} E \left( A \mid C, X^* = 0 \right) \right] e^K \]
\[ = 0, \]

where \( e^K = E \left\{ \exp(-\bar{\alpha}_2^T \varepsilon^*) \right\} \), since \( \varepsilon^* \perp \{ X^*, C, A, Y \} \), and \( E(Y \mid A, C, X^*) = E(Y \mid C, X^*) \), which is implied by \( H_0 \) and Assumption 1. The regularity conditions on \( U(\alpha_2, \gamma) \) are sufficient to ensure that \( (\bar{\alpha}_2, \bar{\gamma}) \) is a local minimum of \( n\hat{U}_n(\alpha_2, \gamma)^T \hat{\Omega}_n^{-1} \hat{U}_n(\alpha_2, \gamma) \), and hence \( (\alpha_2, \gamma) \) is locally identified under \( H_0 \). Then because \( \dim\{U(\alpha_2, \gamma)\} = \dim(\alpha_2) + \dim(\gamma) + q \), \( E\{U(\alpha_2, \gamma)\} = 0 \) is an overidentified moment restriction, and the statistic \( \chi^2_{ror} \) has a limiting distribution of \( \chi^2_q \). \( \square \)
Proof of validity of the doubly-robust test under model (5). First, reparameterize the propensity-score model as

\[
\log E(A \mid C, X^*) = \alpha_0 + g_A(C; \alpha_1) + \alpha_2^T X^*,
\]

where \(g_A(0; \alpha_1) = 0\), such that \(\alpha_0\) is a scalar intercept and \(\alpha_1\) has dimension \(p_1 - 1\). Let \(\exp(K) = E\{\exp(\tilde{\alpha}_2^T \varepsilon^*)\}\), i.e., the moment generating function of \(\varepsilon^*\) evaluated at \(\tilde{\alpha}_2\), \(\tilde{\gamma}\) be the true value of \(\gamma\) under \(M_Y\), and \(\tilde{\alpha}\) be the true value of \(\alpha\) under \(M_A\). Under \(M_A\), there exists some \(\tilde{\gamma}\) such that \(E\{S(\tilde{\gamma})\} = 0\), and for any such \(\tilde{\gamma}\),

\[
E \left[ \begin{array}{c}
  k(C) \Delta_Y(\tilde{\gamma}) \Delta_A(\tilde{\alpha}) \\
  \{\ell(C)Y + m(C)\} \Delta_A(\tilde{\alpha})
\end{array} \right] = E \left[ \begin{array}{c}
  k(C) \{Y - g_Y(C; \tilde{\gamma})\} \\
  \ell(C)Y + m(C)
\end{array} \right] [A \exp(-\tilde{\alpha}_2^T X) - \exp \{\alpha_0^* + g_A(C; \tilde{\alpha}_1)\}] \\
= E \left[ \begin{array}{c}
  k(C) \{E(Y \mid C, X^*, A) - g_Y(C; \tilde{\gamma})\} \\
  \ell(C)E(Y \mid C, X^*) + m(C)
\end{array} \right] \left[ A \frac{E(A \mid C, X^* = 0)}{E(A \mid C, X^*)} - E(A \mid C, X^* = 0) \right] \\
= e^K E \left[ \begin{array}{c}
  k(C) \{E(Y \mid C, X^*) - g_Y(C; \tilde{\gamma})\} \\
  \ell(C)E(Y \mid C, X^*) + m(C)
\end{array} \right] \left[ A \frac{E(A \mid C, X^* = 0)}{E(A \mid C, X^*)} - E(A \mid C, X^* = 0) \right] \\
= 0.
\]
Under $\mathcal{M}_Y$, $E\{S(\gamma)\} = 0$, and for any $\alpha_1$,

$E[k(C)\Delta_Y(\gamma)\Delta_A(\alpha_1, \bar{\alpha}_2)]$

$= E\left(k(C)\{Y - E(Y \mid C)\} \left[A \exp(-\bar{\alpha}_2^T X) - \exp \{g_A(C; \alpha_1)\}\right]\right)$

$= E\left(k(C)E\{Y - E(Y \mid C) \mid C, X^*\} \left[A \frac{E(Y \mid C, X^* = 0)}{E(Y \mid C, X^*)} \exp(-\bar{\alpha}_2^T \varepsilon^*) - \exp \{g_A(C; \alpha_1)\}\right]\right)$

$= E\left(k(C)E\{Y - E(Y \mid C) \mid C, X^*\} \left[E(Y \mid C, X^*) \frac{E(Y \mid C, X^* = 0)}{E(Y \mid C, X^*)} e^K - \exp \{g_A(C; \alpha_1)\}\right]\right)$

$= E\left(k(C)E\{Y - E(Y \mid C) \mid C\} \left[E(Y \mid C, X^* = 0) e^K - \exp \{g_A(C; \alpha_1)\}\right]\right)$

$= 0,$

since $\varepsilon^* \perp \perp \{X^*, C, A, Y\}$ and $E(Y \mid A, C, X^*) = E(Y \mid C, X^*)$, which is implied by $H_0$ and Assumption 1. Finally, there exists some $\tilde{\alpha}_1$ such that

$E\left[\{\ell(C)Y + m(C)\} \Delta_A(\alpha_0^*, \tilde{\alpha}_1, \bar{\alpha}_2)\right] = 0.$

Thus, under any law in $\mathcal{M}_U$, $E\{U(\alpha, \gamma)\} = 0$ has a solution under $H_0$. The regularity conditions on $U(\alpha, \gamma)$ are sufficient to ensure that $(\tilde{\alpha}, \tilde{\gamma})$ is a local minimum of $n\tilde{U}_n(\alpha, \gamma)^T \hat{\Omega}_n^{-1} \tilde{U}_n(\alpha, \gamma)$ under $\mathcal{M}_A$ and $(\alpha_0^*, \tilde{\alpha}_1, \bar{\alpha}_2, \tilde{\gamma})$ is a local minimum under $\mathcal{M}_Y$, and hence $(\alpha, \gamma)$ is locally identified under $\mathcal{M}_U$ and $H_0$. Then because $\dim[U(\alpha, \gamma)] = \dim(\alpha) + \dim(\gamma) + q$, $E\{U(\alpha, \gamma)\} = 0$ is an overidentified moment restriction, and the statistic $\chi^2_{dr}$ has a limiting distribution of $\chi^2_q.$
REFERENCES

Amemiya, Y. (1985). Instrumental variable estimator for the nonlinear errors-in-variables model. *Journal of Econometrics*, 28(3):273–289.

Amemiya, Y. (1990). Two-stage instrumental variables estimators for the nonlinear errors-in-variables model. *Journal of Econometrics*, 44(3):311–332.

Amemiya, Y., Brown, P., and Fuller, W. (1990). Instrumental variable estimation of the nonlinear measurement error model. *Statistical Analysis of Measurement Error Models and Applications*, pages 147–156.

Armstrong, B. G. (1990). The effects of measurement errors on relative risk regressions. *American Journal of Epidemiology*, 132(6):1176–1184.

Armstrong, B. G. (2004). Exposure measurement error: consequences and design issues. *Exposure Assessment in Occupational and Environmental Epidemiology*.

Bateson, T. F., Coull, B. A., Hubbell, B., Ito, K., Jerrett, M., Lumley, T., Thomas, D., Vedal, S., and Ross, M. (2007). Panel discussion review: Session three – issues involved in interpretation of epidemiologic analyses – statistical modeling. *Journal of Exposure Science and Environmental Epidemiology*, 17:S90–S96.

Battistin, E. and Chesher, A. (2014). Treatment effect estimation with covariate measurement error. *Journal of Econometrics*, 178(2):707–715.

Bonhomme, S. and Robin, J.-M. (2009). Consistent noisy independent component analysis. *Journal of Econometrics*, 149(1):12–25.

Buzas, J. S. and Stefanski, L. A. (1996). Instrumental variable estimation in generalized linear measurement error models. *Journal of the American Statistical Association*, 91(435):999–1006.

Campbell, D. T. (1988). *Methodology and Epistemology for Social Sciences: Selected Papers*. University of Chicago Press.
Carroll, R. and Stefanski, L. (1994). Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. *Statistics in Medicine*, 13(12):1265–1282.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective*. CRC press.

Cochran, W. G. and Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 417–446.

Cote, J. A. and Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research*, pages 315–318.

Cragg, J. G. (1997). Using higher moments to estimate the simple errors-in-variables model. *RAND Journal of Economics*, pages S71–S91.

Dagenais, M. G. and Dagenais, D. L. (1997). Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics*, 76(1):193–221.

Erickson, T. and Whited, T. M. (2000). Measurement error and the relationship between investment and q. *Journal of Political Economy*, 108(5):1027–1057.

Erickson, T. and Whited, T. M. (2002). Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Theory*, 18(03):776–799.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, pages 1257–1272.

Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics*, pages 1900–1925.

Frisch, R. (1934). *Statistical confluence analysis by means of complete regression systems*, volume 5. Universitetets Økonomiske Instituut.

Fuller, W. A. (2009). *Measurement Error Models*, volume 305. John Wiley & Sons.

Gryparis, A., Paciorek, C. J., Zeka, A., Schwartz, J., and Coull, B. A. (2009). Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*, 10(2):258–274.
Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.

Hoek, G., Brunekreef, B., Goldbohm, S., Fischer, P., and van den Brandt, P. A. (2002). Association between mortality and indicators of traffic-related air pollution in the Netherlands: A cohort study. *The Lancet*, 360(9341):1203–1209.

Jerrett, M., Burnett, R. T., Ma, R., Pope III, C. A., Krewski, D., Newbold, K. B., Thurston, G., Shi, Y., Finkelstein, N., Calle, E. E., et al. (2005). Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology*, 16(6):727–736.

Kapteyn, A. and Wansbeek, T. (1983). Identification in the linear errors in variables model. *Econometrica: Journal of the Econometric Society*, pages 1847–1849.

Kioumourtzoglou, M.-A., Spiegelman, D., Szpiro, A. A., Sheppard, L., Kaufman, J. D., Yanosky, J. D., Williams, R., Laden, F., Hong, B., and Suh, H. (2014). Exposure measurement error in PM2.5 health effects studies: A pooled analysis of eight personal exposure validation studies. *Environmental Health*, 13(1):2.

Klepper, S. and Leamer, E. E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica: Journal of the Econometric Society*, pages 163–183.

Kuroki, M. and Pearl, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2):423–437.

Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica: Journal of the Econometric Society*, pages 1201–1213.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, 30(1):67–80.

Lockwood, J. and McCaffrey, D. (2015a). Simulation-extrapolation for estimating means and causal effects with mismeasured covariates. *Observational Studies*, 1:241–290.
Lockwood, J. and McCaffrey, D. F. (2015b). Matching and weighting with functions of error-prone covariates for causal inference. *Journal of the American Statistical Association*, (just-accepted):1–24.

McCaffrey, D. F., Lockwood, J., and Setodji, C. M. (2013). Inverse probability weighting with error-prone covariates. *Biometrika*, page ast022.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.

Ogburn, E. L. and VanderWeele, T. J. (2012). On the nondifferential misclassification of a binary confounder. *Epidemiology (Cambridge, Mass.)*, 23(3):433.

Ogburn, E. L. and Vanderweele, T. J. (2013). Bias attenuation results for nondifferentially mismeasured ordinal and coarsened confounders. *Biometrika*, 100(1):241–248.

O’Neill, M. S., Zanobetti, A., and Schwartz, J. (2003). Modifiers of the temperature and mortality association in seven US cities. *American Journal of Epidemiology*, 157(12):1074–1082.

Pal, M. (1980). Consistent moment estimators of regression coefficients in the presence of errors in variables. *Journal of Econometrics*, 14(3):349–364.

Puett, R. C., Yanosky, J. D., Hart, J. E., Paciorek, C. J., Schwartz, J. D., Suh MacIntosh, H. H., Speizer, F. E., and Laden, F. (2009). Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses’ Health Study. *Environmental Health Perspectives*.

Raykov, T. (2012). Propensity score analysis with fallible covariates a note on a latent variable modeling approach. *Educational and Psychological Measurement*, 72(5):715–733.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period–application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512.

Robins, J. M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: A focus on AIDS*, 113:159.
Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Sampson, P. D., Szpiro, A. A., Sheppard, L., Lindström, J., and Kaufman, J. D. (2011). Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data. *Atmospheric Environment*, 45(36):6593–6606.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society*, pages 393–415.

Schennach, S. M. (2012). Measurement error in nonlinear models: A review. Technical report, cemmap working paper, Centre for Microdata Methods and Practice.

Schennach, S. M. (2014). Entropic latent variable integration via simulation. *Econometrica*, 82(1):345–385.

Schennach, S. M. and Hu, Y. (2013). Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *Journal of the American Statistical Association*, 108(501):177–186.

Schwartz, J. D., Lee, M., Kinney, P. L., Yang, S., Mills, D., Sarofim, M. C., Jones, R., Streeter, R., Juliana, A. S., Peers, J., et al. (2015). Projections of temperature-attributable premature deaths in 209 US cities using a cluster-based Poisson approach. *Environmental Health*, 14(1):85.

Sheppard, L., Burnett, R. T., Szpiro, A. A., Kim, S.-Y., Jerrett, M., Pope III, C. A., and Brunekreef, B. (2012). Confounding and exposure measurement error in air pollution epidemiology. *Air Quality, Atmosphere & Health*, 5(2):203–216.

Stefanski, L. and Buzas, J. (1995). Instrumental variable estimation in binary regression measurement error models. *Journal of the American Statistical Association*, 90(430):541–550.

Stuart, A. and Kendall, M. G. (1979). *The Advanced Theory of Statistics*. Macmillan, New York, 4th edition.

Szpiro, A. A., Sampson, P. D., Sheppard, L., Lumley, T., Adar, S. D., and Kaufman, J. D. (2010). Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics*, 21(6):606–631.

34
Szpiro, A. A., Sheppard, L., and Lumley, T. (2011). Efficient measurement error correction with spatially misaligned data. *Biostatistics*, page kxq083.

Yanosky, J., Paciorek, C., and Suh, H. (2008). Predicting chronic fine and coarse particulate exposures using spatio-temporal models for the Northeastern and Midwestern U.S. *Epidemiology*, 19(6):S211–S212.

Zanobetti, A., Schwartz, J., et al. (2009). The effect of fine and coarse particulate air pollution on mortality: A national analysis. *Environmental Health Perspectives*, 117(6):898–903.

Zeger, S. L., Thomas, D., Dominici, F., Samet, J. M., Schwartz, J., Dockery, D., and Cohen, A. (2000). Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environmental Health Perspectives*, 108(5):419.