Online supplement to “Identifying effects of multiple treatments in the presence of unmeasured confounding”

Wang Miao [*], Wenjie Hu [†], Elizabeth L. Ogburn [‡], and Xiaohua Zhou [§]

This supplement includes

- proof of theorems and propositions, useful lemmas,
- discussion and examples on the completeness condition,
- consistency of the least median of squares estimator,
- discussion on identification of a parametric model for a binary outcome,
- details for examples, and
- additional results for simulations and the application.

1. PROOF OF PROPOSITIONS AND THEOREMS

1.1 Proof of Proposition 1

Note that η can be identified by regression of X on Z, then applying lemma 5.1 and theorem 5.1 of Anderson and Rubin (1956) to the factor model for the residuals,

$$X - \eta Z = \alpha U + \varepsilon,$$

we obtain (i) of Proposition 1. The third result of Proposition 1 can be obtained from the well-known completeness property of exponential families, see Theorem 2.2 of Newey and Powell (2003). Here we prove (ii), which rests on the following lemma described by Kotlarski (1967, lemma 1 and remark 5).

---

[*] Department of Probability and Statistics, Peking University, Beijing, PRC; mwfy@pku.edu.cn

[†] Department of Probability and Statistics, Peking University, Beijing, PRC; huwenjie@pku.edu.cn

[‡] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; eogburn@jhsph.edu

[§] Department of Biostatistics and Beijing International Center for Mathematical Research,
Peking University, Beijing, PRC; azhou@math.pku.edu.cn

Lemma 1 (Kotlarski, 1967)**.** *Let U, ε_1, and ε_2 be three independent q-dimensional real random vectors with mean zero, and let $W_1 = U + \varepsilon_1$ and $W_2 = U + \varepsilon_2$. If the joint characteristic function of (W_1, W_2) does not vanish, then the distributions of U, ε_1, and ε_2 are uniquely determined from the joint distribution of (W_1, W_2).*

We apply Kotlarski's lemma to prove (ii) of Proposition 1.

Proof of Proposition 1 (ii). We denote $W = X - \eta Z = \alpha U + \varepsilon$. Note that from (i) of proposition 1, any admissible value for α can be written as $\tilde{\alpha} = \alpha R$ with R an arbitrary $q \times q$ orthogonal matrix, we only need to prove that given $\tilde{\alpha} = \alpha R$, the joint distribution $\tilde{f}(w, u) = f(W = w, U = u; \tilde{\alpha})$ is uniquely determined and $\tilde{f}(w, u) = f(W = w, R^{\mathrm{T}} U = u; \alpha)$.

Because after deleting any row of α there remain two full-rank submatrices of α, there must exist two disjoint square submatrices of α with full rank q. Note that $\tilde{\alpha} = \alpha R$, there must exist two disjoint square submatrices of $\tilde{\alpha}$ with full rank q, which we denote by $\tilde{\alpha}_{\mathcal{I}}$ and $\tilde{\alpha}_{\mathcal{J}}$ with $\mathcal{I}$ and $\mathcal{J}$ denoting the corresponding indices, respectively. Note that $W_{\mathcal{I}} = \tilde{\alpha}_{\mathcal{I}} V + \varepsilon_{\mathcal{I}}$ and $W_{\mathcal{J}} = \tilde{\alpha}_{\mathcal{J}} V + \varepsilon_{\mathcal{J}}$ with $V = R^{\mathrm{T}} U$, we have $\tilde{\alpha}_{\mathcal{I}}^{-1} W_{\mathcal{I}} = V + \tilde{\alpha}_{\mathcal{I}}^{-1} \varepsilon_{\mathcal{I}}$ and $\tilde{\alpha}_{\mathcal{J}}^{-1} W_{\mathcal{J}} = V + \tilde{\alpha}_{\mathcal{J}}^{-1} \varepsilon_{\mathcal{J}}$. According to Lemma 1, the distributions of V, $\tilde{\alpha}_{\mathcal{I}}^{-1} \varepsilon_{\mathcal{I}}$, and $\tilde{\alpha}_{\mathcal{J}}^{-1} \varepsilon_{\mathcal{J}}$ are uniquely determined given $\tilde{\alpha}$, and therefore, the distribution of $\varepsilon = W - \tilde{\alpha} V$ is uniquely determined. As a result, given $\tilde{\alpha}$, there is only one admissible joint distribution, which must be $\tilde{f}(w, u) = f(W = w, R^{\mathrm{T}} U = u \mid \alpha)$. □

1.2 Proof of Theorem 1

Proof. Under the equivalence (Assumption 2 (ii)), given any admissible joint distribution $\tilde{f}(x, u \mid z)$, there must exist some invertible function $V(U)$ such that $\tilde{f}(x, u \mid z) = f\{X = x, V(U) = u \mid z\}$. Because $V(U)$ is invertible, the ignorability assumption 1 ($Y(x) \perp\!\!\!\perp X \mid U$) implies that $Y(x) \perp\!\!\!\perp X \mid V(U)$, the exclusion restriction $Z \perp\!\!\!\perp Y \mid (U, X)$ implies that $Z \perp\!\!\!\perp Y \mid \{X, V(U)\}$, and the completeness (Assumption 2 (iii)) implies that $\tilde{f}(u \mid x, z)$ is also complete in z. Letting $\tilde{f}(y \mid u, x) = f\{y \mid V(U) = u, x\}$, then we have that

$$f\{Y(x) = y\} = \int_u f(y \mid u, x) f(u) du = \int_u \tilde{f}(y \mid u, x) \tilde{f}(u) du, \tag{S.1}$$

$$f(y \mid x, z) = \int_u \tilde{f}(y \mid u, x) \tilde{f}(u \mid x, z) du, \tag{S.2}$$

with $\tilde{f}(u)$ and $\tilde{f}(u \mid x, z)$ derived from $\tilde{f}(x, u \mid z)$. Because $\tilde{f}(u \mid x, z)$ is complete in z, the solution to (S.2) is unique; this is because for any candidate solutions $\tilde{f}_1(y \mid u, x)$ and $\tilde{f}_2(y \mid u, x)$ to (S.2), we must have that $\int_u \{\tilde{f}_1(y \mid u, x) - \tilde{f}_2(y \mid u, x)\}\tilde{f}(u \mid x, z) = 0$, which implies that $\tilde{f}_1(y \mid u, x) = \tilde{f}_2(y \mid u, x)$ by the completeness of $\tilde{f}(u \mid x, z)$ in z. Thus, $\tilde{f}(y \mid u, x)$ is uniquely determined from (S.2), and $f\{Y(x)\}$ is identified by plugging in it into (S.1). $\square$

1.3 **Proof of Theorem 2**

Proof. Under the equivalence (Assumption 3 (ii)), for any admissible joint distribution $\tilde{f}(x, u)$ we must have some invertible function $V(U)$ such that $\tilde{f}(u, x) = f\{X = x, V(U) = u\}$. Letting $\mathcal{C} = \{i : f(u \mid x) \text{ varies with } x_i\}$ and $\tilde{\mathcal{C}} = \{i : \tilde{f}(u \mid x) \text{ varies with } x_i\}$, then we must have $\mathcal{C} = \tilde{\mathcal{C}}$ by noting that $\tilde{f}(u \mid x) = f\{V(U) = u \mid x\}$, i.e., $\mathcal{C}$ can be identified from any admissible joint distribution $\tilde{f}(x, u)$.

Because $V(U)$ is invertible, the ignorability assumption 1 ($Y(x) \perp\!\!\!\perp X \mid U$) implies that $Y(x) \perp\!\!\!\perp X \mid V(U)$, and the completeness (Assumption 3 (iii)) implies that $\tilde{f}(u \mid x)$ is also complete in $x_{\mathcal{S}}$ for any $\mathcal{S} \subset \mathcal{C}$ with cardinality q. Letting $\tilde{f}(y \mid u, x) = f\{y \mid V(U) = u, x\}$, then we have that

$$f\{Y(x) = y\} = \int_u f(y \mid u, x) f(u) du = \int_u \tilde{f}(y \mid u, x)\tilde{f}(u) du, \tag{S.3}$$

$$f(y \mid x) = \int_u \tilde{f}(y \mid u, x)\tilde{f}(u \mid x) du, \tag{S.4}$$

with $\tilde{f}(u)$ and $\tilde{f}(u \mid x)$ obtained from $\tilde{f}(x, u)$.

We prove that $\tilde{f}(y \mid u, x)$ is uniquely determined from (S.4) given $f(y \mid x)$ and $\tilde{f}(u \mid x)$ by way of contradiction. Suppose two candidate outcome models $\tilde{f}_1(y \mid u, x)$ and $\tilde{f}_2(y \mid u, x)$ satisfy (S.4), then $\int_u \{\tilde{f}_1(y \mid u, x) - \tilde{f}_2(y \mid u, x)\}\tilde{f}(u \mid x) = 0$. Under the null treatments assumption, each of $\tilde{f}_1(y \mid u, x)$ and $\tilde{f}_2(y \mid u, x)$ can depend on only $(|\mathcal{C}| - q)/2$ confounded treatments, and thus the contrast $\{\tilde{f}_1(y \mid u, x) - \tilde{f}_2(y \mid u, x)\}$ can depend on at most $|\mathcal{C}| - q$ confounded treatments. We let $X_{\mathcal{S}}$ denote the rest q confounded treatments that the contrast $\{\tilde{f}_1(y \mid u, x) - \tilde{f}_2(y \mid u, x)\}$ does not depend on, then the completeness (Assumption 3(iii)) implies that $\tilde{f}(u \mid x_{\mathcal{S}}, x_{\bar{\mathcal{S}}})$ is complete in $x_{\mathcal{S}}$, and thus $\{\tilde{f}_1(y \mid u, x) - \tilde{f}_2(y \mid u, x)\} = 0$ almost surely, i.e., $\tilde{f}_1(y \mid u, x) = \tilde{f}_2(y \mid u, x)$ almost surely. Therefore, the solution to (S.4) must be unique. Finally, plugging in $\tilde{f}(y \mid u, x)$ and $\tilde{f}(u)$ into (S.3) identifies the potential outcome distribution. $\square$

1.4 **Proof of Proposition 2**

Proof. We first note that the confounded treatments can be identified under the equivalence assumption, by the argument in the proof of Theorem 2. Note that the candidate solutions depending on more than $(|\mathcal{C}| - q)/2$ confounded treatments contradict the null treatments assumption that at most $(|\mathcal{C}| - q)/2$ confounded ones can affect the outcome, we only focus on solutions that depends on at most $(|\mathcal{C}| - q)/2$ confounded treatments.

We let $A_{\mathcal{C}}$ denotes the number of active ones of the confounded treatments. Consider a solution $\tilde{f}(y \mid u, x_{\mathcal{B}})$ that solves

$$f(y \mid x) = \int_u \tilde{f}(y \mid u, x_{\mathcal{B}})\tilde{f}(u \mid x)du. \tag{S.5}$$

where $|\mathcal{B} \cap \mathcal{C}| \leq (|\mathcal{C}| - q)/2$, i.e., $x_{\mathcal{B}}$ includes at most $(|\mathcal{C}| - q)/2$ confounded treatments. Equation (S.5) is a Fredholm integral equation of the first kind with the kernel $\tilde{f}(u \mid x)$ complete in $x_{\bar{\mathcal{B}}}$, where $x_{\bar{\mathcal{B}}}$ denotes the remaining treatments of x except for $x_{\mathcal{B}}$. For $x_{\mathcal{B}}$ that includes all active treatments, i.e., $x_{\mathcal{A}} \subset x_{\mathcal{B}}$ and $|\mathcal{B} \cap \mathcal{C}| = A_{\mathcal{C}}$, the solution to (S.5) exists and must be unique and equal to $\tilde{f}(y \mid u, x_{\mathcal{A}})$. For $x_{\mathcal{B}}$ that includes $t < A_{\mathcal{C}}$ active ones of the confounded treatments, i.e., $|\mathcal{B} \cap \mathcal{C}| = t < A_{\mathcal{C}}$, the solution to (S.5) does not exist. We prove this by way of contradiction.

Suppose (S.5) has a solution $\tilde{f}(y \mid u, x_{\mathcal{B}})$, then it must also satisfy the following equation

$$f(y \mid x) = \int_u \tilde{f}(y \mid u, x_{\mathcal{B}} \cup x_{\mathcal{A}})\tilde{f}(u \mid x)du; \tag{S.6}$$

where the unknown function $\tilde{f}(y \mid u, x_{\mathcal{B}} \cup x_{\mathcal{A}})$ of u is allowed to depend on all active treatments. Note that $\tilde{f}(y \mid u, x_{\mathcal{B}} \cup x_{\mathcal{A}})$ can depend on at most $(|\mathcal{C}| - q)/2 - t$ null ones of the confounded treatments, Equation (S.6) is a Fredholm integral equation of the first kind, with the kernel $\tilde{f}(u \mid x)$ complete in $x_{\bar{\mathcal{B}}} \cap x_{\bar{\mathcal{A}}} \cap x_{\mathcal{C}}$, i.e., the remaining $q + t$ null ones of the confounded treatments. Therefore, (S.6) can be satisfied by only one function, which in fact is $\tilde{f}(y \mid u, x_{\mathcal{A}})$. This contradicts that $\tilde{f}(y \mid u, x_{\mathcal{B}})$ depend on only $t < A_{\mathcal{C}}$ active ones of the confounded treatments.

As a result, all solutions that depend on at most $(|\mathcal{C}| - q)/2$ confounded treatments must be equal to $\tilde{f}(y \mid u, x_{\mathcal{A}})$, i.e., the solution to (6).

□

1.5 **Proof of Theorem 3**

We first describe a lemma that is useful for proof of Theorem 3.

Lemma 2. *For a $p \times p$ positive-definite matrix Σ_ε and a $p \times q$ matrix α of full column rank with $p > q$, letting $\gamma = (\Sigma_\varepsilon + \alpha\alpha^{\mathrm{T}})^{-1}\alpha$, then $\gamma = \Sigma_\varepsilon^{-1}\alpha\{I_q - \alpha^{\mathrm{T}}(\Sigma_\varepsilon + \alpha\alpha^{\mathrm{T}})^{-1}\alpha\}$.*

Proof. Letting $A = \Sigma_\varepsilon^{-1/2}\alpha$ and $B = \Sigma_\varepsilon^{1/2}\gamma$, then A has full rank, and it is straightforward to verify that

$$B = (I_p + AA^{\mathrm{T}})^{-1}A = A\{I_q - A^{\mathrm{T}}(I_p + AA^{\mathrm{T}})^{-1}A\}.$$

Because $\{I_q - A^{\mathrm{T}}(I_p + AA^{\mathrm{T}})^{-1}A\}\{A^{\mathrm{T}}(I_p + AA^{\mathrm{T}})A\} = A^{\mathrm{T}}A$ and $A^{\mathrm{T}}A$ has full rank of q, then $\{I_q - A^{\mathrm{T}}(I + AA^{\mathrm{T}})^{-1}A\}$ must have full rank of q as a $q \times q$ matrix, i.e., $\{I_q - \alpha^{\mathrm{T}}(\Sigma_\varepsilon + \alpha\alpha^{\mathrm{T}})^{-1}\alpha\}$ has full rank. Thus, we have that

$$\gamma = \{\Sigma_\varepsilon + \alpha\alpha^{\mathrm{T}}\}^{-1}\alpha = \Sigma_\varepsilon^{-1}\alpha\{I_q - \alpha^{\mathrm{T}}(\Sigma_\varepsilon + \alpha\alpha^{\mathrm{T}})^{-1}\alpha\}.$$

In the special case that Σ_ε is diagonal and $q = 1$, we have $\gamma_i = \Sigma_{\varepsilon,i}^{-1}\alpha_i\{1 - \alpha^{\mathrm{T}}(\Sigma_\varepsilon + \alpha\alpha^{\mathrm{T}})^{-1}\alpha\}$, where $\Sigma_{\varepsilon,i}$ is the ith diagonal element of Σ_ε. Therefore, if $\alpha_i \neq 0$, we must have $\gamma_i \neq 0$. ☐

We then prove Theorem 3.

Proof of Theorem 3. Under model (16) and condition (i) of Theorem 3, Σ_ε is identified and any admissible value $\tilde{\alpha}$ is a rotation of the truth (Proposition 1), i.e., $\tilde{\alpha} = \alpha R$ for some $q \times q$ orthogonal matrix R. We let $\gamma = \Sigma_X^{-1}\alpha$ denote the coefficient by linear regression of U on X. Given an admissible value $\tilde{\alpha} = \alpha R$, we let $\tilde{\gamma} = \Sigma_X^{-1}\tilde{\alpha} = \gamma R$ and $\tilde{\delta} = R^{\mathrm{T}}\delta$. We let $\mathcal{C} = \{i : \alpha_i \text{ is not a zero vector}\}$ denote the confounded treatments, where α_i is the ith row of α.

Note that Σ_ε is diagonal under model (16), then according to Lemma 2, we have $\tilde{\gamma}_i = \Sigma_{\varepsilon,i}^{-1}\alpha_i\{I_q - \alpha^{\mathrm{T}}(\Sigma_\varepsilon + \alpha\alpha^{\mathrm{T}})^{-1}\alpha\}R$, where $\tilde{\gamma}_i$ is the ith row of $\tilde{\gamma}$ and $\Sigma_{\varepsilon,i}$ is the ith diagonal element of Σ_ε. Therefore, $\tilde{\gamma}_i$ is a zero vector if and only if α_i is a zero vector, i.e., $\mathcal{C}$ is identified by the set $\{i : \tilde{\gamma}_i \text{ is not a zero vector}\}$.

Letting ξ denote the ordinary least squares coefficient by regressing Y on X, then we have

$$\xi = \beta + \tilde{\gamma}\tilde{\delta}. \tag{S.7}$$

By way of contradiction, we prove that $(\beta, \tilde{\delta})$ is uniquely determined from this equation given $(\tilde{\beta}, \tilde{\gamma})$ and under the null treatments assumption. Suppose that two sets of values $(\beta^{(1)}, \tilde{\delta}^{(1)})$ and $(\beta^{(2)}, \tilde{\delta}^{(2)})$ satisfy (S.7), and both $\beta_{\mathcal{C}}^{(1)}$ and $\beta_{\mathcal{C}}^{(2)}$ satisfies the null treatments assumption that at most $(|\mathcal{C}| - q)/2$ entries are nonzero.

For unconfounded treatments, the corresponding rows of $\tilde{\alpha}$ and $\tilde{\gamma}$ must be zero as we show in the above, thus $\beta_{\bar{\mathcal{C}}}^{(1)} = \beta_{\bar{\mathcal{C}}}^{(2)} = \tilde{\beta}_{\bar{\mathcal{C}}}$. We remain to prove that $\beta_{\mathcal{C}}^{(1)} = \beta_{\mathcal{C}}^{(2)}$ and $\tilde{\delta}^{(1)} = \tilde{\delta}^{(2)}$. From (S.7), we have that

$$\beta_{\mathcal{C}}^{(1)} - \beta_{\mathcal{C}}^{(2)} = \tilde{\gamma}_{\mathcal{C}}(\tilde{\delta}^{(2)} - \tilde{\delta}^{(1)}).$$

On the left hand side of this equation, $\beta_{\mathcal{C}}^{(1)} - \beta_{\mathcal{C}}^{(2)}$ has at least q zero entries under assumption (i) of Theorem 3. We use $\mathcal{Z}$ to denote the indices of zero entries of $\beta_{\mathcal{C}}^{(1)} - \beta_{\mathcal{C}}^{(2)}$ and $\tilde{\gamma}_{\mathcal{Z}}$ the corresponding submatrix of $\tilde{\gamma}_{\mathcal{C}}$, then we have that $0 = \tilde{\gamma}_{\mathcal{Z}}(\tilde{\delta}^{(2)} - \tilde{\delta}^{(1)})$. Note that $\tilde{\gamma}_{\mathcal{Z}} = \gamma_{\mathcal{Z}} R$ must have full rank of q under assumption (iii) of Theorem 3, then we must have $\tilde{\delta}^{(2)} = \tilde{\delta}^{(1)}$, and thus $\beta_{\mathcal{C}}^{(1)} = \beta_{\mathcal{C}}^{(2)}$. In summary, $\beta^{(1)} = \beta^{(2)}$, i.e., β is uniquely determined. $\square$

2. DISCUSSION AND EXAMPLES ON THE COMPLETENESS CONDITION

Completeness is a fundamental concept in statistics (see Lehman and Scheffe (1950); Basu (1955)), which is taught in most foundational courses of statistical inference. It has been used to establish the theory for hypothesis testing and unbiased estimation in mathematical statistics (Lehman and Scheffe, 1950), and recently been used to establish identification in causal inference, missing data, and measurement error problems. Nonetheless, it may still be abstract to practitioners. Therefore, it is worth explaining in more detail. We add further explanation and extra examples to facilitate the interpretation and use of the completeness condition in practice. In particular, we illustrate completeness from the following perspectives.

- The role of completeness in identification. Since its prevalent use in statistics, completeness has been widely used to establish identification for a variety of nonparametric and semiparametric models, for instance, the IV regression model (Newey and Powell, 2003; Darolles et al., 2011), IV quantile model (Chernozhukov and Hansen, 2005; Chen et al., 2014), measurement error model (Hu and Schennach, 2008), missing data model (Miao and Tchetgen Tchetgen, 2016; D'Haultfœuille, 2010), and proximal inference (Miao et al., 2018). It has been very well studied by statisticians and economists, primitive conditions are readily available in the literature including very general exponential families of distributions and regression models, and the literature is still growing; see for example, Newey and Powell (2003); D'Haultfœuille (2010, 2011); Darolles et al. (2011); Chen et al. (2014); Hu and Shiu (2018). Our use of completeness falls in this line of work, where the main identifying assumption that captures the underlying causal structure is an IV, auxiliary variables, or null treatments assumption and the completeness is viewed as a regularity condition.

- Intuition and implication of completeness. Completeness is equivalent to the injectivity of the conditional expectation operator (D'Haultfœuille, 2011). Completeness characterizes the informativeness of the auxiliary variable about the confounder and its ability to recover the confounding bias. It is analogous to the relevance condition in the instrumental variable identification. It can be interpreted as a nonparametric rank condition and is easiest understood in the categorical and the linear cases where the outcome model to be identified is parametric. In the categorical case where both U and Z have k levels, completeness means that the matrix $[f(u_i \mid x, z_j)]$ consisting of the conditional probabilities is invertible for any x. This is stronger than dependence of Z and U given X. Roughly speaking, dependence reveals that variability in U is accompanied by variability in Z, and completeness reinforces that any infinitesimal variability in U is accompanied by variability in Z. For instance, if Z is a proxy of U, completeness of $f(u \mid x, z)$ can be interpreted as no coarsening in the measurement Z of the confounder U. As a consequence, completeness fails if the number of levels or dimension of Z is smaller than that of U. For the binary case, completeness holds if U and Z are correlated within each level of X. In the linear model $E(U \mid x, Z) = \gamma_0(x) + \gamma_1(x)Z$, completeness reduces to a rank condition that $\gamma_1(x)$ has full row rank for all x. The rank condition can only hold if the dimension of Z is no smaller than that of U. This argument provides a rationale for measuring a rich set of potential auxiliary variables for the purpose of confounding adjustment. However, if the outcome model is unrestricted, completeness serves as a generic rank condition accommodating both categorical and continuous variables, linear and nonlinear models, although, it can no longer be expressed so concisely as a full rank condition.

- How to assess or test completeness. Completeness can be checked in specific models, for instance, one can check whether the covariance matrix is of full rank in the joint normal model. Unfortunately, Canay et al. (2013) show that for unrestricted models the completeness condition is in fact untestable, even if all relevant variables (X, Z, U in our problem) are observed. Therefore, without restrictions on the distribution, it is impossible to provide empirical evidence in favor of the completeness condition, akin to the ignorability assumption.

- When does completeness hold or fail, and is it a stringent condition? A number of papers (Andrews, 2017; D'Haultfœuille, 2011; Newey and Powell, 2003; Darolles et al., 2011; Chen et al., 2014; Hu and Shiu, 2018) have established genericity results for parametric, semiparametric, and nonparametric distributions satisfying completeness. Andrews (2017) has shown that if Z and U are continuously distributed and the dimension of Z is larger than that of U, then under a mild regularity condition the completeness condition holds generically in the sense that the set of distributions or conditional expectation operators for which completeness fails has a property analogous to having zero Lebesgue measure (Chen et al., 2014; Andrews, 2017). By appealing to such results, completeness holds in a large class of distributions and thus one may argue that it is commonly satisfied.

In short, completeness is one of the most general conditions made in problems of identification. It requires that Z must have sufficient dimensions or levels and variability relative to U. Commonly-used parametric and semiparametric models, such as exponential families (Newey and Powell, 2003, Theorem 2.2) and location-scale families (Mattner, 1992; Hu and Shiu, 2018), and nonparametric additive models (D'Haultfœuille, 2011) satisfy the completeness condition. For nonparametric models, it is not testable but holds in a large class of models.

In the following, we provide extra examples illustrating completeness, see also Lehman and Scheffe (1950) for a variety of parametric examples where completeness holds and also counterexamples. We also refer to Newey and Powell (2003) for completeness of exponential families, Hu and Shiu (2018) for location-scale families, and D'Haultfœuille (2011); Darolles et al. (2011) for additive separable regression models.

Example S.1. *The binary case. Suppose both Z and U are binary, then for any x completeness of $f(u \mid x, z)$ holds as long as $U \not\perp\!\!\!\perp Z \mid X = x$, but otherwise completeness fails if $U \perp\!\!\!\perp Z \mid X = x$.*

Example S.2. *The categorical case. Suppose U has q levels and Z has r levels, then for a given x completeness of $f(u \mid x, z)$ in z holds as long as the matrix*

$$[f(u_i \mid x, z_j)]_{q\times r} = \left\{ \begin{array}{ccc} f(u_1 \mid z_1, x) & \cdots & f(u_1 \mid z_r, x) \\ \vdots & \ddots & \vdots \\ f(u_q \mid z_1, x) & \cdots & f(u_q \mid z_r, x) \end{array} \right\}$$

consisting of the conditional probabilities has full row rank. Therefore, it is necessary that $q \leq r$ and $U \not\perp\!\!\!\perp Z \mid X = x$. Otherwise, completeness fails if either $q > r$ or $U \perp\!\!\!\perp Z \mid X = x$. However for $q > 2$, the full rank condition is stronger than the dependence ($U \not\perp\!\!\!\perp Z \mid X = x$). For instance, if $f(u_1 \mid z, x) \neq f(u_2 \mid z, x)$ and $f(u_3 \mid z, x) = f(u_1 \mid z, x)$ for all z, then $U \not\perp\!\!\!\perp Z \mid X = x$ but the full rank *condition is obviously not met. This is because the variability in U from u_1 to u_3 is not sufficiently captured by Z, i.e., the measure of Z is coarsened if we view it as a proxy of U. Roughly speaking, dependence reveals that variability in U is accompanied by variability in Z, and completeness reinforces that any infinitesimal variability in U is accompanied by variability in Z.*

Example S.3. *Gaussian distributions. Suppose U and Z have dimensions of q and r, respectively, and $f(u, z \mid x)$ is joint normal given x, then completeness of $f(u \mid x, z)$ in z reduces to a rank condition: the coefficient matrix $\gamma_1(x)$ in model $E(U \mid x, Z) = \gamma_0(x) + \gamma_1(x)Z$ has full row rank given x. It is required that the dimension of Z is no smaller than that of U and that the regression coefficients of each confounder on X and Z are not collinear; otherwise, the completeness fails.*

Example S.4. *A scale model. Lehman and Scheffe (1950, example 3.3) presents a counterexample where completeness fails for $f(u \mid x, z) \sim N(0, \sigma^2_{x,z})$. This is because the conditional density is an even function of u and $E\{g(U) \mid x, z\} = 0$ for any square-integrable and odd function g. In this example, the scale or magnitude of variability of U is captured by Z but not the orientation.*

3. CONSISTENCY OF THE LEAST MEDIAN OF SQUARES ESTIMATOR $(\tilde{\delta}^{\text{lms}}, \hat{\beta}^{\text{lms}})$

For the consistency of $(\hat{\delta}^{\text{lms}}, \hat{\beta}^{\text{lms}})$, we need an additional regularity condition that is slightly stronger than assumption (i) of Theorem 3, which is routinely assumed in the least median squares estimation (see Theorem 3 in Chapter 3 of Rousseeuw and Leroy, 2005).

Assumption S.1. *At most $[|\mathcal{C}|/2] - q + 1$ entries of $\beta_{\mathcal{C}}$ are nonzero, where $[x]$ is the largest integer less than or equal to x.*

We show consistency of $(\hat{\delta}^{\text{lms}}, \hat{\beta}^{\text{lms}})$ under this assumption and the assumptions of Theorem 3, given $n^{1/2}$-consistency of $(\hat{\xi}, \hat{\gamma})$, i.e., $n^{1/2}(\hat{\xi} - \xi)$ and $n^{1/2}(\hat{\gamma} - \gamma R)$ are bounded in probability for some unknown orthogonal matrix R. We show that $\hat{\delta}^{\text{lms}} \to R\delta$ and $\hat{\beta}^{\text{lms}} \to \beta$. For notational simplicity, we only consider the special case where R is the identity matrix. For general cases with R unknown, the following proof holds by simply replacing δ with $R\delta$ and γ with γR.

Because $n^{1/2}(\hat{\gamma} - \gamma)$ is bounded in probability, then $||\hat{\gamma}_i||_2^2 \to ||\gamma_i||_2^2 > \log(n)/n$ for $\gamma_i \neq 0$ and $n/\log(n)||\hat{\gamma}_i||_2^2 \to 0$ for $\gamma_i = 0$. Lemma 2 implies $\mathcal{C} = \{i : ||\alpha_i||_2^2 > 0\} = \{i : ||\gamma_i||_2^2 > 0\}$ and therefore,

$f(\hat{\mathcal{C}} \neq \mathcal{C}) \to 0$, i.e., $\hat{\mathcal{C}}$ consistently selects the confounded treatments. Letting

$$\tilde{\delta}^{\text{lms}} = \arg\min_{\delta} \text{median}\ \{(\hat{\xi}_i - \hat{\gamma}_i\delta)^2,\ i \in \mathcal{C}\}, \quad \mathcal{C} = \{i : ||\gamma_i||_2^2 > 0\},$$

we only need to show consistency of $\tilde{\delta}^{\text{lms}}$ because $\tilde{\delta}^{\text{lms}} = \hat{\delta}^{\text{lms}}$ upon $\hat{\mathcal{C}} = \mathcal{C}$.

Note that

$$\begin{aligned} \text{median}\{(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\text{lms}})^2 : i \in \mathcal{C}\} &\leq \text{median}\{(\hat{\xi}_i - \hat{\gamma}_i\delta)^2 : i \in \mathcal{C}\} \\ &\leq \text{median}\{(\hat{\xi}_i - \xi_i - (\hat{\gamma}_i - \gamma_i)\delta + \xi_i - \gamma_i\delta)^2 : i \in \mathcal{C}\} \\ &\leq \text{median}\{(\hat{\xi}_i - \xi_i - (\hat{\gamma}_i - \gamma_i)\delta + \beta_i)^2 : i \in \mathcal{C}\}. \end{aligned}$$

For sufficiently large sample size n, $\hat{\xi}_i - \xi_i$ and $\hat{\gamma}_i - \gamma_i$ are close to zero so that $(\hat{\xi}_i - \xi_i - (\hat{\gamma}_i - \gamma_i)\delta)^2 < (\hat{\xi}_j - \xi_j - (\hat{\gamma}_j - \gamma_j)\delta + \beta_j)^2$ for any i with $\beta_i = 0$ and any j with $\beta_j \neq 0$. Assumption S.1 states that more than half entries of $\beta_{\mathcal{C}}$ are zero, and thus, $\text{median}\{(\hat{\xi}_i - \xi_i - (\hat{\gamma}_i - \gamma_i)\delta + \beta_i)^2 : i \in \mathcal{C}\}$ is attained among the null treatments. Therefore, we have asymptotically

$$\begin{aligned} \text{median}\{(\hat{\xi}_i - \xi_i - (\hat{\gamma}_i - \gamma_i)\delta + \beta_i)^2 : i \in \mathcal{C}\} &\leq \max\{(\hat{\xi}_i - \xi_i - (\hat{\gamma}_i - \gamma_i)\delta)^2 : i \in \mathcal{C} \text{ and } \beta_i = 0\} \\ &\leq \max\{(\hat{\xi}_i - \xi_i - (\hat{\gamma}_i - \gamma_i)\delta)^2 : i \in \mathcal{C}\}. \end{aligned}$$

Hence,

$$\text{median}\{(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\text{lms}})^2 : i \in \mathcal{C}\} \quad \leq \quad \max\{(\hat{\xi}_i - \xi_i - (\hat{\gamma}_i - \gamma_i)\delta)^2 : i \in \mathcal{C}\}.$$

Letting $\Delta = \tilde{\delta}_{\text{lms}} - \delta$, we can show the following result,

Result 1: $\text{median}\{(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\text{lms}})^2 : i \in \mathcal{C}\} \geq \frac{1}{2}\{(\hat{\xi}_i - \xi_i) - (\hat{\gamma}_i - \gamma_i)(\delta + \Delta) - \gamma_i\Delta\}^2$

for at least q elements belonging to the subset $\{i \in \mathcal{C} : \beta_i = 0\}$.

Given Result 1, we have

$$\frac{1}{2}(\hat{\xi}_i - \xi_i - (\hat{\gamma}_i - \gamma_i)(\delta + \Delta) - \gamma_i\Delta)^2 \leq \max\{(\hat{\xi}_i - \xi_i - (\hat{\gamma}_i - \gamma_i)\delta)^2 : i \in \mathcal{C}\}$$

for at least q elements belonging to the subset $\{i \in \mathcal{C} : \beta_i = 0\}$.

Assuming that $(\hat{\xi}, \hat{\gamma})$ are consistent, then the right hand side must converge to zero and thus $\gamma_i\Delta \to 0$ for at least q elements in $\mathcal{C}$. Moreover, any submatrix of $\gamma_{\mathcal{C}}$ consisting of q rows has full rank (Assumption

(iii) of Theorem 3), then Δ must converge to zero, i.e., $\tilde{\delta}_{\rm lms}$ is consistent and as a result $\hat{\delta}_{\rm lms}$ is consistent. Consistency of $\hat{\beta}^{\rm lms} = \hat{\xi} - \hat{\gamma}\hat{\delta}^{\rm lms}$ follows from consistency of $(\hat{\xi}, \hat{\gamma}, \hat{\delta}_{\rm lms})$.

Now we prove Result 1. Note that

$$\mathrm{median}\{(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\rm lms})^2 : i \in \mathcal{C}\} \quad = \quad \mathrm{median}\{(\beta_i + (\hat{\xi}_i - \xi_i) - (\hat{\gamma}_i - \gamma_i)(\delta + \Delta) - \gamma_i\Delta)^2 : i \in \mathcal{C}\}.$$

If $|\mathcal{C}|$ is odd, Assumption S.1 implies that at most $(|\mathcal{C}| - 1)/2 - q + 1$ entries of $\beta_{\mathcal{C}}$ are nonzero. Arranging $(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\rm lms})^2$ in increasing order, then $\mathrm{median}\{(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\rm lms})^2 : i \in \mathcal{C}\}$ is equal to the $(|\mathcal{C}|+1)/2$-th element. Thus, the following inequality holds for at least $(|\mathcal{C}|+1)/2 - \{(|\mathcal{C}|-1)/2 - q + 1\} = q$ elements belonging to the subset $\{i \in \mathcal{C} : \beta_i = 0\}$,

$$\begin{aligned}\mathrm{median}\{(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\rm lms})^2 : i \in \mathcal{C}\} \quad &\geq \quad \{(\hat{\xi}_i - \xi_i) - (\hat{\gamma}_i - \gamma_i)(\delta + \Delta) - \gamma_i\Delta\}^2 \\ &\geq \quad \frac{1}{2}\{(\hat{\xi}_i - \xi_i) - (\hat{\gamma}_i - \gamma_i)(\delta + \Delta) - \gamma_i\Delta\}^2.\end{aligned}$$

If $|\mathcal{C}|$ is even, Assumption S.1 implies that at most $|\mathcal{C}|/2 - q + 1$ entries of $\beta_{\mathcal{C}}$ are nonzero. Arranging $(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\rm lms})^2$ in increasing order, then $\mathrm{median}\{(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\rm lms})^2 : i \in \mathcal{C}\}$ is equal to the average of the $|\mathcal{C}|/2$-th and $(|\mathcal{C}|/2 + 1)$-th elements. Thus, $\mathrm{median}\{(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\rm lms})^2 : i \in \mathcal{C}\}$ is no smaller than half of the $(|\mathcal{C}|/2+1)$-th element. As a result, the following inequality holds for at least $|\mathcal{C}|/2+1-(|\mathcal{C}|/2-q+1) = q$ elements belonging to the subset $\{i \in \mathcal{C} : \beta_i = 0\}$,

$$\mathrm{median}\{(\hat{\xi}_i - \hat{\gamma}_i\tilde{\delta}^{\rm lms})^2 : i \in \mathcal{C}\} \quad \geq \quad \frac{1}{2}\{(\hat{\xi}_i - \xi_i) - (\hat{\gamma}_i - \gamma_i)(\delta + \Delta) - \gamma_i\Delta\}^2.$$

This completes the proof of Result 1.

4. DISCUSSION ON IDENTIFICATION OF A PARAMETRIC BINARY OUTCOME MODEL

Without assist of auxiliary variables and null treatments assumptions, identification is not generally available and depends on specific model assumptions. In a recent note, Kong et al. (2021) consider a binary outcome model with one confounder. Under a factor model for normally distributed treatments and a couple of assumptions such as knowing the sign of confounding bias, they prove identification via a meticulous analysis of the link distribution. However, their identification results do not generalize to the multivariate confounder case as illustrated by the following counterexample.

Example S.5. *Assuming that*

$$X = \alpha U + \varepsilon, \quad f(Y = 1 \mid X, U) = G(\beta_0 + \beta^{\rm T}X + \delta^{\rm T}U), \quad \varepsilon \sim N(0, \Sigma_\varepsilon), \quad U \sim N(0, I_q), \qquad \text{(S.8)}$$

where U is a q-dimensional confounder, Σ_ε is diagonal, and G is a known distribution function relating the outcome mean to a linear model of the treatments and confounder. The unknown parameters (β, δ) capture the treatment effects and the magnitude of confounding, respectively.

Under this setting, one can verify that the observed data distribution $f(x, y)$ is satisfied with $\tilde{\alpha} = \alpha R_1$, $\tilde{\delta} = R_1^{\mathrm{T}} \Sigma^{-1/2} R_2 \Sigma^{1/2} \delta$, and $\tilde{\beta} = \beta + \gamma(I_q - \Sigma^{-1/2} R_2 \Sigma^{1/2})\delta$, where $\Sigma = I_q - \alpha^{\mathrm{T}} \Sigma_X^{-1} \alpha$ and R_1, R_2 are arbitrary $q \times q$ orthogonal matrices.

In the special case where U is univariate, i.e., $q = 1$, there are only two possible values -1 and 1 for orthogonal matrices R_1, R_2; thus, there are at least two possible values for the treatment effect, $\tilde{\beta} = \beta$ and $\tilde{\beta} = \beta + 2\gamma\delta$. If further the signs of δ and at least one entry of α are known, i.e., $R_1 = R_2 = 1$, then $\tilde{\beta} = \beta + 2\gamma\delta$ can be excluded, and in fact, Kong et al. (2021) have shown that $\tilde{\beta} = \beta$ is the only possible value for the treatment effect provided that G is not a normal distribution. However, this argument does not generalize to the multivariate confounder case, because there are infinite number of orthogonal matrices with dimension $q \geq 2$, in which case, it is impossible to specify R_1, R_2.

5. DETAILS FOR EXAMPLES

5.1 Details for Example 1

Note that η can be identified by regression of X on Z. Given η, an arbitrary admissible value $\tilde{\alpha}$, and $\tilde{f}(u \mid x, z) \sim N(\tilde{\gamma}^{\mathrm{T}} x - \tilde{\gamma}^{\mathrm{T}} \eta z, \tilde{\sigma}^2)$ with $\tilde{\gamma} = (\Sigma_{X-\eta Z})^{-1} \tilde{\alpha}$ and $\tilde{\sigma}^2 = 1 - \tilde{\alpha}^{\mathrm{T}} (\Sigma_{X-\eta Z})^{-1} \tilde{\alpha}$, we solve

$$f(y \mid x, z) = \int_u h(y, x, u) \tilde{f}(u \mid x, z) du \tag{S.9}$$

$$\phi \text{ is the probability density function of } N(0,1), \tag{S.10}$$

$$= \int_u h(y, x, u) \cdot \frac{1}{\tilde{\sigma}} \phi \left\{ \frac{u - (\tilde{\gamma}^{\mathrm{T}} x - \tilde{\gamma}^{\mathrm{T}} \eta z)}{\tilde{\sigma}} \right\} du \tag{S.11}$$

for $h(y, x, u)$, which is the outcome model $\tilde{f}(y \mid x, u)$. Following the procedure described by Miao et al. (2018), $h(y, x, u)$ can be represented in Fourier transforms of $\tilde{f}(u \mid x, z)$ and $f(y \mid x, z)$.

By substitution $z' = \{\tilde{\gamma}^{\mathrm{T}} x - \tilde{\gamma}^{\mathrm{T}} \eta z\} / \tilde{\sigma}$, $u' = u / \tilde{\sigma}$, and by letting

$$g(y, x, z') = f \left\{ y \mid x, z = \frac{\tilde{\gamma}^{\mathrm{T}} x - \tilde{\sigma} z'}{\tilde{\gamma}^{\mathrm{T}} \eta} \right\},$$

(S.9) implies that

$$\begin{aligned} g(y,x,z') &= \int_{-\infty}^{+\infty} \frac{1}{\tilde\sigma}\phi(z'-u')\cdot h(y,x,u'\tilde\sigma)du \\ &= \int_{-\infty}^{+\infty} \phi(z'-u')\cdot h(y,x,u'\tilde\sigma)du'. \end{aligned}$$

which is an integral equation of convolution type and can be solved by applying the Fourier transform. Letting h_1 and h_2 denote the Fourier transforms of ϕ and g respectively:

$$\begin{aligned} h_1(t) &= \int_{-\infty}^{+\infty} \exp(-\mathrm{i}tz)\phi(z)dz, \\ h_2(y,x,t) &= \int_{-\infty}^{+\infty} \exp(-\mathrm{i}tz')g(y,x,z')dz' \\ &= -\frac{\tilde\gamma^{\mathrm{T}}\eta}{\tilde\sigma}\int_{-\infty}^{+\infty} \exp\left\{-\mathrm{i}t\frac{\tilde\gamma^{\mathrm{T}}x-\tilde\gamma^{\mathrm{T}}\eta z}{\tilde\sigma}\right\} f(y\mid x,z)dz, \end{aligned}$$

with $\mathrm{i}=(-1)^{1/2}$ the imaginary unity, we have

$$h_2(y,x,t) = h_1(t)\times\int_{-\infty}^{+\infty}\exp(-\mathrm{i}tu')h(y,x,u'\tilde\sigma)du',$$

$$\int_{-\infty}^{+\infty}\exp(-\mathrm{i}tu')h(y,x,u'\tilde\sigma)du' = \frac{h_2(y,x,t)}{h_1(t)};$$

by Fourier inversion, we have

$$h(y,x,u'\tilde\sigma) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp(\mathrm{i}tu')\frac{h_2(y,x,t)}{h_1(t)}dt;$$

by substitution $u=u'\tilde\sigma$, we obtain

$$h(y,x,u) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp\left\{\frac{\mathrm{i}tu}{\tilde\sigma}\right\}\frac{h_2(y,x,t)}{h_1(t)}dt,$$

and the potential outcome distribution is

$$f\{Y(x)=y\} = \int_{-\infty}^{+\infty} h(y,x,u)\phi(u)du.$$

5.2 A confounder proxy example for the auxiliary variables approach

For simplicity, we consider a binary confounder, p binary treatments, and a binary proxy Z of the confounder. We assume that

$$X_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp X_p \mid U, \tag{S.12}$$

$$\text{at least three treatments are correlated with } U; \tag{S.13}$$

$$Z \not\perp\!\!\!\perp U, \quad Z \perp\!\!\!\perp (X,Y)\mid U; \tag{S.14}$$

where the last independence is known as the nondifferentially error assumption (Ogburn and VanderWeele, 2013; Carroll et al., 2006). Under (S.12)–(S.14), completeness of $f(u \mid x, z)$ in z holds as long as Z is correlated with U, because $E\{g(U) \mid x, z\} = 0 \Leftrightarrow \sum_u g(u)f(x, u)f(z \mid u) = 0 \Leftrightarrow g(u)f(x, u) = 0 \Leftrightarrow g(u) = 0$.

According to Kuroki and Pearl (2014), under (S.12)–(S.13), any admissible joint distribution $\tilde{f}(x, u)$ equals the joint distribution of X and some label switching of U. Given $\tilde{f}(x, u)$, we solve $f(z, x) = \sum_u \tilde{f}(z \mid u)\tilde{f}(x, u)$ to obtain $\tilde{f}(z \mid u)$ and $\tilde{f}(x, z, u) = \tilde{f}(x, u)\tilde{f}(z \mid u)$. We then obtain $\tilde{f}(y \mid u, x)$ by solving $f(y \mid x, z) = \sum_u \tilde{f}(y \mid u, x)\tilde{f}(u \mid x, z)$, and finally the potential outcome distribution is identified by $f\{Y(x) = y\} = \sum_u \tilde{f}(y \mid u, x)\tilde{f}(u)$. The identification result can be generalized to the categorical setting, and we refer to Kuroki and Pearl (2014) for details of factor analysis in this case.

5.3 A normal mixture model that satisfies the equivalence assumption

Example S.6. *(Yakowitz and Spragins, 1968, Proposition 2). Suppose* U *has* q *categories with* $f(U = u_i) = \pi_i$ *and* X *is a* p *dimensional vector with* $f(x \mid u_i) \sim N(\mu_i, \Sigma_i)$*. Assuming that the pairs* (μ_i, Σ_i) *are all distinct, then* $f(x)$ *has a unique representation in normal mixtures* $f(x) \sim \sum_{i=1}^{q'} p_i' N(\mu_i', \Sigma_i')$*: we must have* $q' = q$ *and for each* i *there must exist some* j *such that* $\pi_i' = \pi_j$ *and* $(\mu_i', \Sigma_i') = (\mu_j, \Sigma_j)$*. That is, the equivalence holds and* $f(x, u)$ *is identified up to a label switching of the confounder.*

6. ADDITIONAL RESULTS FOR SIMULATIONS AND THE APPLICATION

6.1 Results for simulations

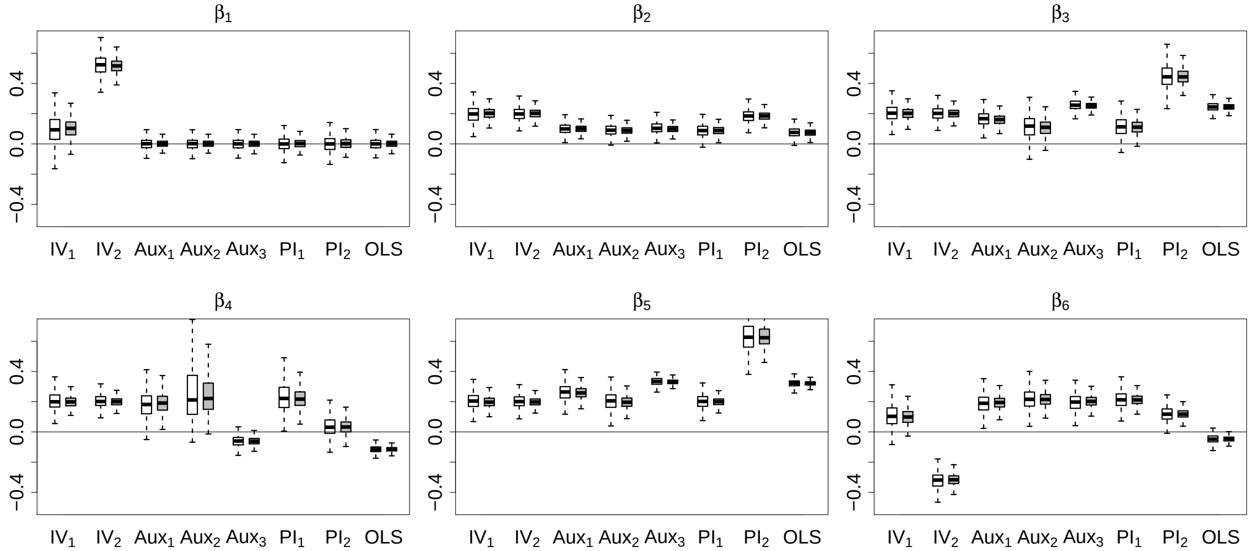

Figure S.1: Bias of estimators when the exclusion restriction fails in the auxiliary variables setting. White boxes are for sample size 1000 and gray ones for 2000.

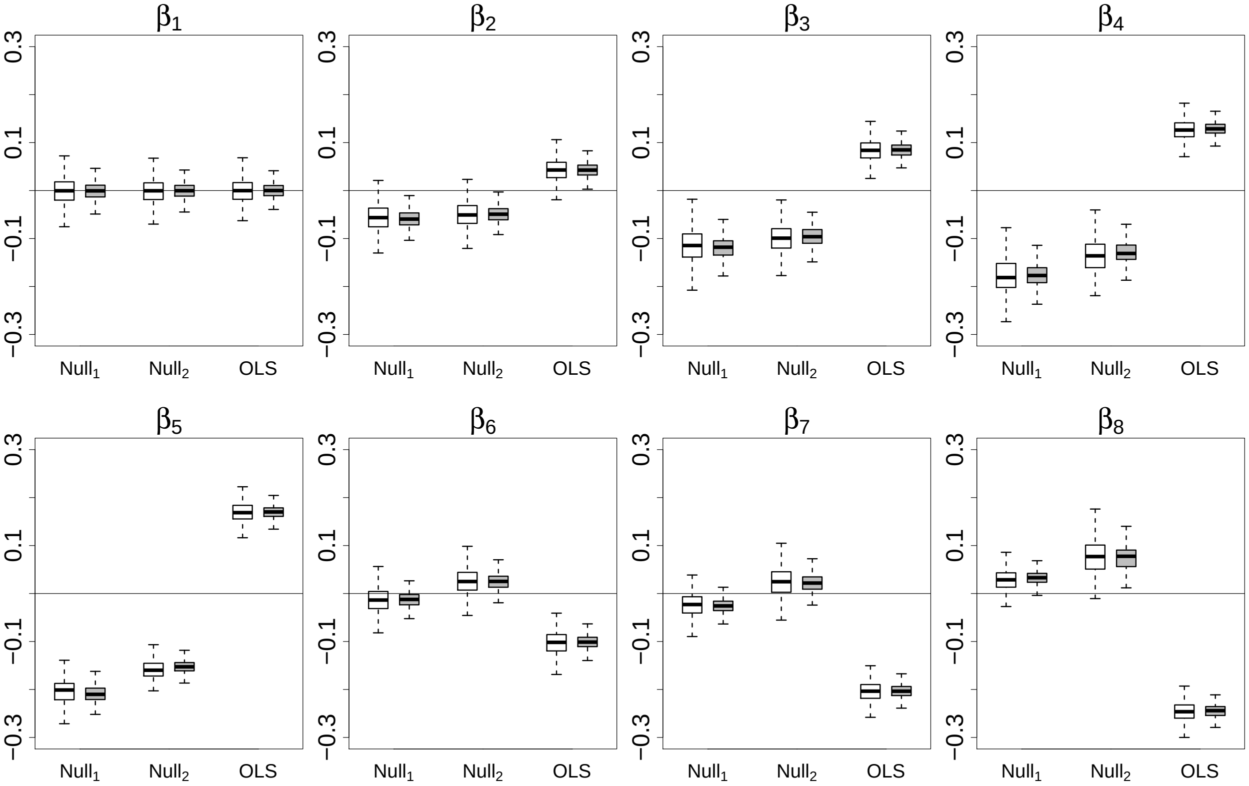

Figure S.2: Bias of estimators in Case 2 of the null treatments setting. White boxes are for sample size 2000 and gray ones for 5000.

6.2 Results for the application

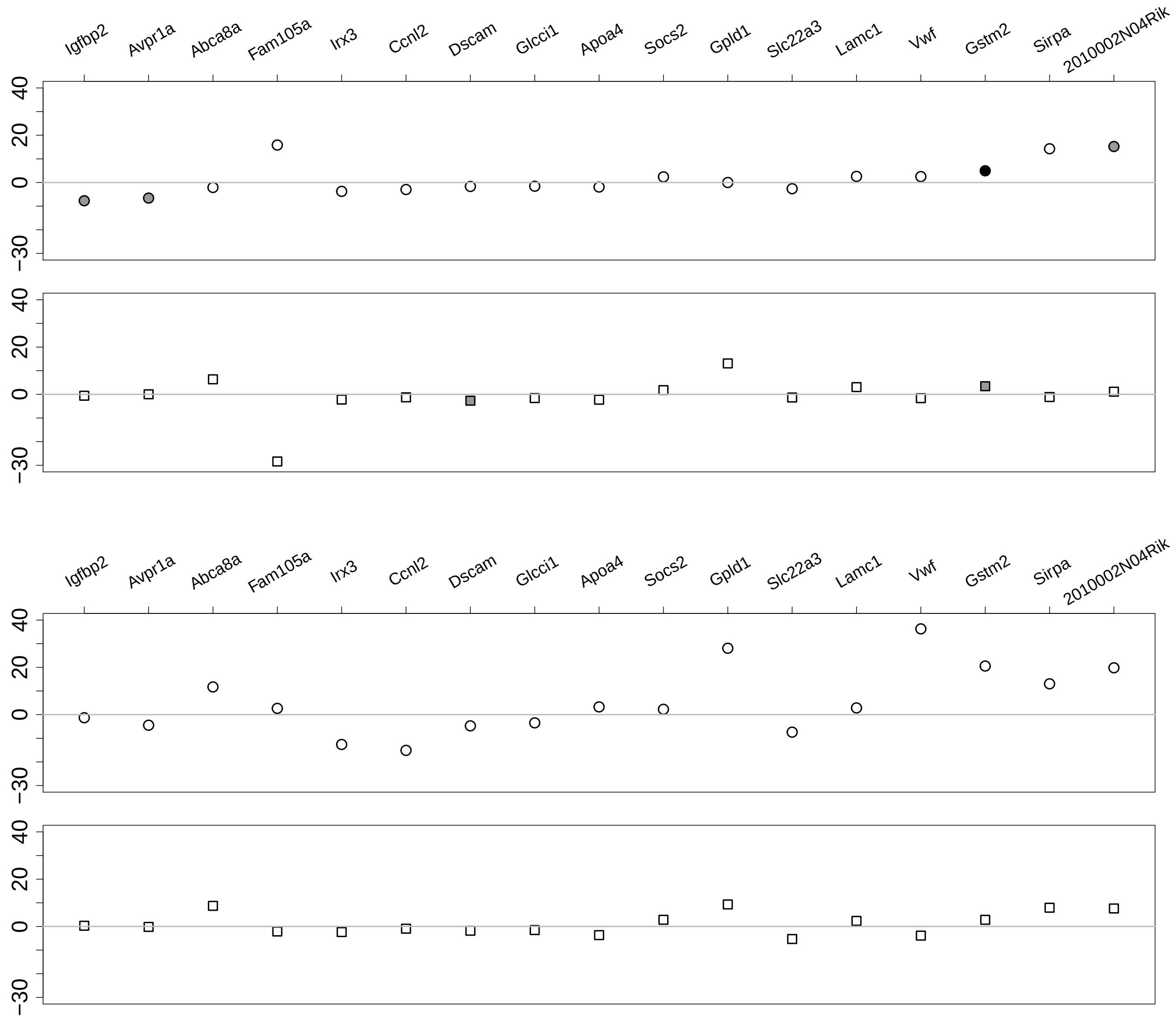

Figure S.3: Effect estimates for 17 genes when two (the first two panels) or three (the last two panels) factors are used in analyses. The first and third panels are for the auxiliary variables estimation, and the second and fourth for the null treatments estimation. Black points are for significant estimates at level of 0.05, gray ones for 0.1, and white ones for estimates not significant at 0.1.

Point and confidence interval estimates

Bootstrap percentiles: 2.5%, 97.5%, 5%, 95%; Significance codes: "**" for significant at level of 0.05, "*" for 0.1, and "0" for not significant at level 0.1.

Estimation with one confounder

```
1. Results for the auxiliary variables approach

            estimates    2.5%  97.5%      5%    95% significance
Igfbp2        -10.626 -15.614 -5.266 -14.764 -5.719     **
Avpr1a         -8.296 -14.568 -1.544 -13.494 -2.670     **
Abca8a         -5.664 -12.368  1.072 -10.887 -0.032      *
Fam105a        -4.325 -13.812  4.983 -12.137  2.933      0
Irx3           -3.793  -8.033  0.593  -7.297 -0.154      *
Ccnl2          -3.421  -7.810  0.917  -6.966  0.246      0
Dscam          -2.442  -6.703  0.276  -5.072 -0.514      *
Glcci1         -1.429  -6.999  3.139  -5.964  2.360      0
Apoa4          -0.287  -4.226  3.821  -3.569  3.299      0
Socs2           1.162  -1.800  3.909  -1.276  3.516      0
Gpld1           1.456  -6.520 10.827  -4.847  8.951      0
Slc22a3         1.724  -4.953  7.274  -3.915  6.177      0
Lamc1           3.111  -1.267  8.967  -0.661  8.204      0
Vwf             3.410  -2.669  9.452  -1.400  8.589      0
Gstm2           5.717   1.181 10.045   1.864  9.304     **
Sirpa           7.283   1.312 13.743   2.263 12.774     **
2010002N04Rik  10.293   3.012 17.203   4.351 16.110     **
```

```
2. Results for the null treatments approach

           estimates    2.5%  97.5%      5%    95% significance
Igfbp2        -1.194  -9.369  2.800  -8.045  1.840      0
Avpr1a        -0.430  -7.574  3.999  -6.821  2.678      0
Abca8a         6.032  -3.043 13.693  -1.842 12.352      0
```

```
Fam105a       -6.726 -15.875  3.118 -14.526  1.234      0
Irx3          -2.465  -7.292  1.475  -6.661  0.988      0
Ccnl2         -1.349  -6.031  2.855  -5.338  2.083      0
Dscam         -2.255  -7.043 -0.145  -5.540 -0.390     **
Glcci1        -1.536  -6.826  3.343  -6.151  2.567      0
Apoa4         -2.595  -6.378  2.136  -5.713  1.285      0
Socs2          2.117  -1.100  4.992  -0.533  4.513      0
Gpld1          8.681  -0.709 16.796   0.703 14.594      *
Slc22a3       -1.803  -7.358  4.809  -6.227  3.932      0
Lamc1          2.664  -1.792  8.506  -1.235  7.696      0
Vwf           -1.072  -5.372  6.179  -4.309  5.445      0
Gstm2          3.688  -0.806  9.612   0.171  8.948      *
Sirpa          4.972  -0.704 11.731   0.285 10.628      *
2010002N04Rik  6.272   0.096 13.682   1.353 12.599     **
```

3. Results for the crude estimation

```
           estimates    2.5%  97.5%      5%    95% significance
Igfbp2        -8.445 -12.797 -4.586 -12.124 -5.292     **
Avpr1a        -6.607 -11.779 -1.634 -11.094 -2.395     **
Abca8a        -3.236  -8.644  2.049  -7.674  1.429      0
Fam105a       -5.017 -14.030  3.880 -12.662  2.108      0
Irx3          -3.494  -8.028  0.878  -7.284  0.263      0
Ccnl2         -2.962  -7.357  0.860  -6.716  0.269      0
Dscam         -2.410  -7.427 -0.220  -5.806 -0.640     **
Glcci1        -1.522  -7.485  3.178  -6.389  2.303      0
Apoa4         -0.973  -4.563  2.935  -3.992  2.045      0
Socs2          1.518  -1.109  4.289  -0.743  3.828      0
```

```
Gpld1          3.919  -2.720 10.907  -1.583  9.980      0
Slc22a3        0.609  -5.184  6.260  -4.205  4.983      0
Lamc1          3.174  -1.191  9.643  -0.598  8.337      0
Vwf            2.032  -3.505  7.896  -2.520  6.781      0
Gstm2          4.767   0.251  9.300   1.194  8.647     **
Sirpa          6.935   0.858 13.027   1.636 12.190     **
2010002N04Rik  9.226   2.454 15.538   3.235 14.574     **
```

Estimation with two confounders

```
1. Results for the auxiliary variables approach


           estimates    2.5%  97.5%      5%    95% significance
Igfbp2        -7.707 -18.416  1.113 -16.599 -0.978      *
Avpr1a        -6.586 -15.364  0.383 -13.920 -1.049      *
Abca8a        -2.100 -15.114  8.107 -12.496  5.177      0
Fam105a       15.842 -43.428 59.749 -34.904 49.521      0
Irx3          -3.752  -8.337  0.682  -7.363  0.020      0
Ccnl2         -2.959  -7.808  1.329  -7.060  0.594      0
Dscam         -1.671  -6.691  1.455  -5.357  0.861      0
Glcci1        -1.570  -7.136  3.210  -5.886  2.482      0
Apoa4         -1.913  -7.064  4.566  -5.579  3.704      0
Socs2          2.382  -2.566  6.707  -1.939  5.493      0
Gpld1          0.004  -9.004 11.469  -6.927  9.564      0
Slc22a3       -2.617 -15.306 11.915 -11.466  9.435      0
Lamc1          2.585  -1.511  9.001  -1.086  8.108      0
Vwf            2.503  -3.897  9.630  -2.735  8.523      0
Gstm2          4.962   0.284 10.218   1.271  9.119     **
Sirpa         14.276  -6.498 30.434  -2.951 26.550      0
2010002N04Rik 15.232  -2.550 31.459   0.999 25.530      *
```

```
2. Results for the null treatments approach
```

```
            estimates    2.5%  97.5%      5%    95% significance
Igfbp2         -0.584   -9.401  2.190  -7.986  1.276      0
Avpr1a          0.021   -8.330  2.998  -7.046  2.156      0
Abca8a          6.359   -3.449 13.216  -2.241 11.147      0
Fam105a       -28.398 -318.341  1.951 -73.696  0.404      0
Irx3           -2.211   -7.370  1.553  -6.523  1.022      0
Ccnl2          -1.270   -6.192  2.666  -5.442  2.121      0
Dscam          -2.692   -7.175  0.208  -5.986 -0.262      *
Glcci1         -1.582   -6.935  3.447  -5.972  2.602      0
Apoa4          -2.275   -6.315  2.239  -5.450  1.679      0
Socs2           1.784   -1.301  5.008  -0.813  4.409      0
Gpld1          13.077   -0.993 19.818  -0.040 17.053      0
Slc22a3        -1.340   -9.239  5.508  -7.669  4.482      0
Lamc1           3.058   -1.893  8.703  -1.245  7.451      0
Vwf            -1.619   -5.267  6.335  -4.201  5.598      0
Gstm2           3.426   -0.295  9.968   0.438  9.129      *
Sirpa          -1.141   -6.001 14.311  -3.833 12.380      0
2010002N04Rik   1.116   -2.534 14.654  -1.160 13.157      0
```

Estimation with three confounders

```
1. Results for the auxiliary variables approach
```

```
            estimates    2.5%  97.5%      5%    95% significance
Igfbp2         -1.312 -17.987  6.862 -15.951  4.212      0
Avpr1a         -4.481 -18.904  5.022 -15.353  2.897      0
```

```
Abca8a         11.713 -17.207 34.518 -12.953 28.557      0
Fam105a         2.648 -71.350 81.166 -49.248 62.241      0
Irx3          -12.628 -33.805  4.268 -24.106  1.251      0
Ccnl2         -15.119 -23.126  4.215 -18.271  1.616      0
Dscam          -4.751 -15.444  1.729 -11.027  0.785      0
Glcci1         -3.485  -9.837  4.176  -7.731  3.013      0
Apoa4           3.268  -7.404 10.427  -5.519  7.904      0
Socs2           2.249  -5.602  8.652  -3.800  6.912      0
Gpld1          28.073 -50.673 59.428 -19.273 42.508      0
Slc22a3        -7.413 -23.712 12.482 -18.348  9.988      0
Lamc1           2.841  -2.863 13.409  -1.291 10.471      0
Vwf            36.274 -11.822 56.152  -4.906 43.160      0
Gstm2          20.551  -4.118 30.792  -0.593 25.547      0
Sirpa          13.029  -6.271 33.161  -3.244 26.817      0
2010002N04Rik  19.805  -6.478 69.222  -0.918 53.812      0
```

2. Results for the null treatments approach

```
            estimates    2.5%  97.5%      5%    95% significance
Igfbp2         0.304    -9.903   1.264   -8.630  0.689      0
Avpr1a        -0.188    -8.836   3.601   -7.821  2.650      0
Abca8a         8.709    -4.579  13.954   -3.275 10.415      0
Fam105a       -2.115 -1527.892  42.311 -900.908  6.914      0
Irx3          -2.364    -7.044   4.262   -6.313  2.759      0
Ccnl2         -0.936    -5.863   4.479   -5.029  3.555      0
Dscam         -1.812    -6.532   1.339   -5.410  0.149      0
Glcci1        -1.523    -6.817   3.630   -5.850  2.827      0
Apoa4         -3.665    -7.561   2.568   -6.635  1.677      0
```

```
Socs2           2.812    -1.284   5.211   -0.670  4.579       0
Gpld1           9.310    -4.096  19.869   -2.003 15.753       0
Slc22a3        -5.310    -7.903   5.516   -6.701  4.519       0
Lamc1           2.353    -2.483   9.319   -1.486  7.890       0
Vwf            -3.877   -15.989   5.917  -11.843  4.612       0
Gstm2           2.807    -3.831   9.466   -2.007  8.653       0
Sirpa           7.923    -4.153  13.763   -2.283 12.411       0
2010002N04Rik  7.609   -19.287 178.871   -2.461 14.933       0
```

REFERENCES

Anderson, T. W. and Rubin, H. “Statistical inference in factor analysis.” In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*, 111–150. Berkeley, Calif.: University of California Press (1956).

Andrews, D. W. “Examples of L^2-complete and boundedly-complete distributions.” *Journal of Econometrics*, 199:213–220 (2017).

Basu, D. “On statistics independent of a complete sufficient statistic.” *Sankhya*, 15:377–380 (1955).

Canay, I. A., Santos, A., and Shaikh, A. M. “On the testability of identification in some nonparametric models with endogeneity.” *Econometrica*, 81:2535–2559 (2013).

Carroll, R., Ruppert, D., Stefanski, L., and Crainiceanu, C. *Measurement Error in Nonlinear Models: A Modern Perspective*. Boca Raton: Chapman & Hall/CRC, 2nd edition (2006).

Chen, X., Chernozhukov, V., Lee, S., and Newey, W. K. “Local identification of nonparametric and semiparametric models.” *Econometrica*, 82:785–809 (2014).

Chernozhukov, V. and Hansen, C. “An IV model of quantile treatment effects.” *Econometrica*, 73:245–261 (2005).

Darolles, S., Fan, Y., Florens, J. P., and Renault, E. “Nonparametric instrumental regression.” *Econometrica*, 79:1541–1565 (2011).

D'Haultfœuille, X. "A new instrumental method for dealing with endogenous selection." *Journal of Econometrics*, 154:1–15 (2010).

—. "On the completeness condition in nonparametric instrumental problems." *Econometric Theory*, 27:460–471 (2011).

Hu, Y. and Schennach, S. M. "Instrumental variable treatment of nonclassical measurement error models." *Econometrica*, 76:195–216 (2008).

Hu, Y. and Shiu, J.-L. "Nonparametric identification using instrumental variables: Sufficient conditions for completeness." *Econometric Theory*, 34:659–693 (2018).

Kong, D., Yang, S., and Wang, L. "Multi-cause causal inference with unmeasured confounding and binary outcome." *Biometrika*, in press (2021).

Kotlarski, I. "On characterizing the gamma and the normal distribution." *Pacific Journal of Mathematics*, 20:69–76 (1967).

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

Lehman, E. and Scheffe, H. "Completeness, Similar Regions and Unbiased Tests. Part I." *Sankhya*, 10:219–236 (1950).

Mattner, L. "Completeness of location families, translated moments, and uniqueness of charges." *Probability Theory and Related Fields*, 92:137–149 (1992).

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. "Identifying causal effects with proxy variables of an unmeasured confounder." *Biometrika*, 105:987–993. (2018).

Miao, W. and Tchetgen Tchetgen, E. J. "On varieties of doubly robust estimators under missingness not at random with a shadow variable." *Biometrika*, 103(2):475–482 (2016).

Newey, W. K. and Powell, J. L. "Instrumental variable estimation of nonparametric models." *Econometrica*, 71:1565–1578 (2003).

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

Rousseeuw, P. J. and Leroy, A. M. *Robust Regression and Outlier Detection*, volume 589. John wiley & sons (2005).

Yakowitz, S. J. and Spragins, J. D. “On the identifiability of finite mixtures.” *The Annals of Mathematical Statistics*, 209–214 (1968).