Inverse-Probability-of-Treatment Weighted Estimation of Causal Parameters in the Presence of Error-Contaminated and Time-Dependent Confounders

Di Shu
University of Waterloo

Supervisor: Dr. Grace Yi

August 12, 2016

Introduction

There is an increasing but still scarce literature concerning causal inference with error-prone data.

Marginal structural models (MSM) are widely used to delineate causal effects. In the presence of time-dependent confounding, Robins et al. (2000) proposed the inverse-probability-of-treatment weighted (IPTW) estimation of causal parameters of MSM.

However, their approach has a critical assumption:
measurements are precise!

Objectives:
We consider the case where the time-dependent confounders are mismeasured and propose several methods to accommodate the measurement error. To our knowledge, this problem has not been considered before.

Notation and Framework

- Observed data

  - At visit $k (k = 0, \dots, K)$ for subject $i (i = 1, \dots, n)$:

    * $Z_i(k)$: precisely-observed time-dependent confounders;
    * $X_i(k)$: mismeasured time-dependent confounders;
    * $A_i(k)$: observed binary treatment;
    * $\bar{A}_i(k) = \{A_i(u) : 0 \leq u \leq k,\ u \text{ is integer}\}$: observed treatment history up to visit k;
    * $\bar{A}_i = \bar{A}_i(K)$: observed full treatment history;

  - After visit K:

    Y_i: observed outcome variable;

Notation and Framework

- Potential outcome/counterfactuals:

  - At visit $k (k = 0, \ldots, K)$ for subject $i (i = 1, \ldots, n)$:

    * $a_i(k)$: potential binary treatment;
    * $\bar{a}_i(k) = \{a_i(u) : 0 \leq u \leq k,\ u \text{ is integer}\}$: potential treatment history up to visit k;
    * $\bar{a}_i = \bar{a}_i(K)$: potential full treatment history;

  - After visit K:

    $Y_{\bar{a}_i}$: potential outcome that would have been observed had the subject experienced treatment history $\bar{a}_i$;

Assumptions

- **Assumption 1**: the treatment assignment of subject j has no effect on the potential outcomes of subject i for all $i \neq j$.
- **Assumption 2**: the potential outcome under the observed treatment is equal to the observed outcome, i.e., $Y_{\bar{A}} = Y$.
- **Assumption 3**: $P(\bar{A}|\bar{Z}, \bar{X}, Y_{\bar{a}}) = P(\bar{A}|\bar{Z}, \bar{X})$.
- **Assumption 4**: for $k = 0, 1, \ldots, K$,

$$0 < P\{A(k) = 1|\bar{A}(k-1), \bar{Z}(k), \bar{X}(k)\} < 1.$$

- **Assumption 5**:
$P(\bar{A}|\bar{Z}, \bar{X}) = \prod_{k=0}^{K} P\{A(k)|, \bar{A}(k-1), Z(k), X(k)\}$.

Model Setup

- Marginal structural model:

$$E(Y_{\bar{a}}) = h(\bar{a}; \beta), \tag{1}$$

where β is the causal parameters of interest.
Only one of $\{\bar{a}\}$ is actually observed for an individual.

$$\downarrow$$

We can't fit (1) directly using available data.

- Crude model:

$$E(Y|\bar{A}) = h(\bar{A}; \alpha), \tag{2}$$

where α is the associational parameters.
We can fit (2) directly using available data, as $\bar{A}$ are available.
But $\alpha \neq \beta$ in general (Robins et al. 2000).

Estimation without Measurement Error

To produce a consistent estimator for β, Robins et al. (2000) proposed the IPTW estimation method.

- **Step 1 (Weight Estimation)**:
  For each subject i, determine the weight

$$w_i = \prod_{k=0}^{K} \frac{1}{P\{A_i(k)|, \bar{A}_i(k-1), Z_i(k), X_i(k)\}}, \qquad (3)$$

where

$$\begin{aligned} &\text{logit}[P\{A_i(k) = 1|\bar{A}_i(k-1), Z_i(k), X_i(k)\}] \\ &= \gamma_{0k} + \boldsymbol{\gamma}_{Ak}^T \bar{A}_i(k-1) + \boldsymbol{\gamma}_{Zk}^T Z_i(k) + \boldsymbol{\gamma}_{Xk}^T X_i(k). \end{aligned} \qquad (4)$$

Fitting (4) gives an estimator of $(\gamma_{0k}, \boldsymbol{\gamma}_{Ak}^T, \boldsymbol{\gamma}_{Zk}^T, \boldsymbol{\gamma}_{Xk}^T)^T$ and therefore an estimator $\hat{w}_i$ of w_i using (3).

Estimation without Measurement Error

- **Step 2 (Fitting the Weighted Outcome Model)**:
  Fit model (2) with subject i assigned weights $\hat{w}_i$.

  The resulting IPTW estimator $\hat{\beta}$ is consistent for β, the causal parameters of interest of (1).

  The robust variance estimator (Huber 1967) can be employed (Robins et al. 2000).

Measurement Error Model

Confounders $X_i(k)$ are error-prone.

Let X^*_{ik} be an observed measurement of $X_i(k)$. Assume that conditional on $X_i(k)$,

$$X^*_{ik} = X_i(k) + \epsilon_{ik} \tag{5}$$

for $i = 1, \ldots, n$ and $k = 0, \ldots, K$, where

- the ϵ_{ik} and the $X_i(k)$ are independent, and the ϵ_{ik} are independent across different i and k.
- ϵ_{ik} follow $N(\mathbf{0}, \mathbf{\Sigma}_{\epsilon k})$, with covariance matrix $\mathbf{\Sigma}_{\epsilon k}$.

Adjusting for Measurement Error Effects

Recall the standard IPTW procedure:

- Step 1:

$$w_i = \prod_{k=0}^{K} \frac{1}{P\{A_i(k)|, \bar{A}_i(k-1), Z_i(k), X_i(k)\}},$$

- Step 2: fit $E(Y|\bar{A}) = h(\bar{A}; \alpha)$ with $\hat{w}_i$ in step 1.

Note: $\hat{w}_i$, the estimator for w_i plays a key role. But $X_i(k)$ is unobserved.

Naively ignoring the difference between X_{ik}^* and $X_i(k)$ leads to biased estimation of β.

Idea: estimation of $P\{A_i(k) = 1|\bar{A}_i(k-1), Z_i(k), X_i(k)\}$ adjusted for measurement error.

Regression Calibration (RC)

The conditional probabilities in the denominators of w_i are modeled by

$$\text{logit}[P\{A_i(k) = 1 | \bar{A}_i(k-1), Z_i(k), X_i(k)\}]$$
$$= \gamma_{0k} + \boldsymbol{\gamma}_{Ak}^T \bar{A}_i(k-1) + \boldsymbol{\gamma}_{Zk}^T Z_i(k) + \boldsymbol{\gamma}_{Xk}^T X_i(k).$$

The basic idea of RC (Prentice 1982) is to replace $X_i(k)$ with its conditional expectation $E\{X_i(k)|X_{ik}^*\}$ in the standard analysis.

By Carroll et al. (2006), we estimate $E\{X_i(k)|X_{ik}^*\}$ by

$$\hat{X}_i(k) = \hat{\mu}_k + \hat{\boldsymbol{\Sigma}}_{Xk} \cdot (\hat{\boldsymbol{\Sigma}}_{Xk} + \boldsymbol{\Sigma}_{\epsilon k})^{-1} \cdot (X_{ik}^* - \hat{\mu}_k),$$

where

$$\hat{\mu}_k = \frac{\sum_{i=1}^n X_{ik}^*}{n},$$

$$\hat{\boldsymbol{\Sigma}}_{Xk} = \frac{\sum_{i=1}^n (X_{ik}^* - \hat{\mu}_k)(X_{ik}^* - \hat{\mu}_k)^T - (n-1)\boldsymbol{\Sigma}_{\epsilon k}}{n-1}.$$

Regression Calibration (RC)

Replacing $X_i(k)$ with $\hat{X}_i(k)$ and respectively fitting models (4) gives estimate $(\hat{\gamma}_{0k}, \hat{\gamma}_{Ak}^T, \hat{\gamma}_{Zk}^T, \hat{\gamma}_{Xk}^T)^T$.

The conditional probabilities in the denominators of w_i are estimated by

$$\hat{P}\{A_i(k) = 1 | \bar{A}_i(k-1), Z_i(k), X_i(k)\}$$
$$= \frac{1}{1 + \exp\{-\hat{\gamma}_{0k} - \hat{\gamma}_{Ak}^T \bar{A}_i(k-1) - \hat{\gamma}_{Zk}^T Z_i(k) - \hat{\gamma}_{Xk}^T \hat{X}_i(k)\}},$$

leading to the adjusted IPTW weights which will serve as the input in Step 2 of the standard IPTW estimation.

Regression Calibration (RC)

Remark:

$\hat{X}_i(k)$ is a linear function in X^*_{ik}.

$\downarrow$

naive analysis and RC produce same estimate for w_i.

$\downarrow$

naive analysis and RC produce same estimate for β.

$\downarrow$

naive analysis=RC.

SIMEX-based Methods

- Direct SIMEX (SIMEX_1)
  The basic idea of SIMEX_1 is to obtain causal parameters estimates $\hat{\beta}$ by using SIMEX method.
- Indirect SIMEX (SIMEX_2)
  The basic idea of SIMEX_2 is to obtain logistic parameters estimates $(\hat{\gamma}_{0k}, \hat{\gamma}_{Ak}^T, \hat{\gamma}_{Zk}^T, \hat{\gamma}_{Xk}^T)^T$ by using SIMEX method.

  Then the conditional probabilities are estimated by

$$\hat{P}\{A_i(k) = 1 | \bar{A}_i(k-1), Z_i(k), X_i(k)\}$$
$$= \frac{1}{1 + \exp\{-\hat{\gamma}_{0k} - \hat{\gamma}_{Ak}^T \bar{A}_i(k-1) - \hat{\gamma}_{Zk}^T Z_i(k) - \hat{\gamma}_{Xk}^T \hat{X}_i(k)\}}.$$

Consistent Estimator (CS.Δ)

Previous methods are only approximately consistent.

By adapting Stefanski and Carroll (1987), we directly estimate the conditional probabilities by

$$\hat{P}\{A_i(k) = 1|\bar{A}_i(k-1), Z_i(k), \hat{\Delta}_i(k)\}$$
$$= \frac{1}{1 + \exp\{-\hat{\gamma}_{0k} - \hat{\gamma}_{Ak}^T \bar{A}_i(k-1) - \hat{\gamma}_{Zk}^T Z_i(k) - \hat{\gamma}_{Xk}^T \hat{\Delta}_i(k)\}}, \quad (6)$$

where

- $\hat{\Delta}_i(k) = X_{ik}^* + \{A_i(k) - 1/2\}\boldsymbol{\Sigma}_{\epsilon k}\hat{\gamma}_{Xk}$ is a consistent estimate for $\Delta_i(k) = X_{ik}^* + \{A_i(k) - 1/2\}\boldsymbol{\Sigma}_{\epsilon k}\gamma_{Xk}$.
- $(\hat{\gamma}_{0k}, \hat{\gamma}_{Ak}^T, \hat{\gamma}_{Zk}^T, \hat{\gamma}_{Xk}^T)^T$ is consistent estimator for $(\gamma_{0k}, \gamma_{Ak}^T, \gamma_{Zk}^T, \gamma_{Xk}^T)^T$.

Consistent Estimator (CS.Δ)

Many methods have been developed to obtain consistent estimator $(\hat{\gamma}_{0k}, \hat{\gamma}_{Ak}^T, \hat{\gamma}_{Zk}^T, \hat{\gamma}_{Xk}^T)^T$.

For example, Stefanski and Carroll (1987) proposed the conditional score method to obtain $(\hat{\gamma}_{0k}, \hat{\gamma}_{Ak}^T, \hat{\gamma}_{Zk}^T, \hat{\gamma}_{Xk}^T)^T$ by solving the following estimating equations for $(\gamma_{0k}, \gamma_{Ak}^T, \gamma_{Zk}^T, \gamma_{Xk}^T)^T$:

$$\sum_{i=1}^{n} \left(\left[A_i(k) - \frac{1}{1+\exp\{-\gamma_{0k} - \gamma_{Ak}^T \bar{A}_i(k-1) - \gamma_{Zk}^T Z_i(k) - \gamma_{Xk}^T \Delta_i(k)\}} \right] \cdot \left\{ \begin{array}{c} 1 \\ \bar{A}_i(k-1) \\ Z_i(k) \\ \Delta_i(k) \end{array} \right\} \right) = \mathbf{0}, \tag{7}$$

where $\Delta_i(k) = X_{ik}^* + \{A_i(k) - 1/2\}\mathbf{\Sigma}_{\epsilon k}\gamma_{Xk}$.

Consistent Estimator (CS.Δ)

Theorem: Let $\hat{w}_i$ be the IPTW weights estimated by using (6). The IPTW estimation with weights $\hat{w}_i$ yields consistent estimator for the causal parameters β.

Note: the causal mean $E(Y_{\bar{a}})$ under treatment $\bar{a}$ is consistently estimated by

$$\frac{\sum_{i=1}^{n} \hat{w}_i Y_i I(\bar{A}_i = \bar{a})}{n} \quad \text{or} \quad \frac{\sum_{i=1}^{n} \hat{w}_i Y_i I(\bar{A}_i = \bar{a})}{\sum_{i=1}^{n} \hat{w}_i I(\bar{A}_i = \bar{a})},$$

where $I(\cdot)$ is the indicator function.

Remark: based on the consistent estimators for causal mean $E(Y_{\bar{a}})$ in the theorem, the causal odds ratio (OR), causal risk ratio (RR) and causal risk difference (RD) can be estimated simultaneously to assess the relative effectiveness of two treatment plans.

Another Two Approaches Based on (6)

For the consistent method:

$$\hat{P}\{A_i(k) = 1 | \bar{A}_i(k-1), Z_i(k), \hat{\Delta}_i(k)\}$$
$$= \frac{1}{1 + \exp\{-\hat{\gamma}_{0k} - \hat{\gamma}_{Ak}^T \bar{A}_i(k-1) - \hat{\gamma}_{Zk}^T Z_i(k) - \hat{\gamma}_{Xk}^T \hat{\Delta}_i(k)\}}.$$

where $(\hat{\gamma}_{0k}, \hat{\gamma}_{Ak}^T, \hat{\gamma}_{Zk}^T, \hat{\gamma}_{Xk}^T)^T$ is consistent estimator.

$(\hat{\gamma}_{0k}, \hat{\gamma}_{Ak}^T, \hat{\gamma}_{Zk}^T, \hat{\gamma}_{Xk}^T)^T$ can also be obtained by approximately consistent method:

- using regression calibration method
- using SIEMX method

We call the two resulting approaches RC.Δ and SIMEX$_2$.Δ.

Selected Simulation Results

	$\boldsymbol{\Sigma}_{\epsilon k} = 0.5^2$		$\boldsymbol{\Sigma}_{\epsilon k} = 1.5^2$	
Method	Bias%	CP%	Bias%	CP%
Naive	2.110	91.4	7.358	15.2
RC	2.110	91.4	7.358	15.2
RC.Δ	0.578	95.5	2.089	89.9
SIMEX_2	1.923	92.2	8.620	5.00
$\text{SIMEX}_2.\Delta$	0.357	95.9	5.513	49.8
CS.Δ	0.032	95.8	0.270	90.6
SIMEX_1	0.394	94.9	5.325	47.4

Bias%: relative bias=bias/true value$\times$100%

CP%: coverage percentage based on robust variance estimate

Selected Simulation Results

CS.Δ is consistent, as expected. But the coverage percentage is 90.6%, not close to 95%. $\rightarrow$ robust variance underestimates the variance.

The robust variance estimate is valid in the absence of measurement error, but not in the presence of measurement error.

By using the jackknife variance estimate (Efron 1982), the resulting CP for CS.Δ is 94.8%. Resampling-based variance is more reliable.

RC.Δ performs the second best.

RC is equivalent to the naive analysis, as expected.

Assumptions Revisit

Assumption 1-4 are standard assumptions for causal inference.

Assumption 5:
$P(\bar{A}|\bar{Z},\bar{X}) = \prod_{k=0}^{K} P\{A(k)|,\bar{A}(k-1), Z(k), X(k)\}$.

It means that the current treatment depends on the previous treatments and current confounders.

As a restriction on the generally true statement $P(\bar{A}|\bar{Z},\bar{X}) = \prod_{k=0}^{K} P\{A(k)|,\bar{A}(k-1), \bar{Z}(k), \bar{X}(k)\}$, this assumption is often violated.

However, this Markov-type assumption is often reasonable, when the previous confounders have no effects on the current treatment assignment, given previous treatments and current confounders.

Summary of Proposed Methods

There is an increasing but still scarce literature concerning causal inference with measurement error.

To adjust for measurement error effects on IPTW estimation, we propose six methods: RC, SIMEX_1, SIMEX_2, CS.Δ, RC.Δ and $\text{SIMEX}_2.\Delta$.

- They are straightforward and easy to implement.
- Ignoring measurement error effects produce biased results.
- SIMEX-based methods are more computation intensive.
- RC and naive analysis is equivalent.
- CS.Δ consistently estimate the causal parameters.
- RC.Δ performs the second best in simulation studies.
- Resampling-based variance estimates are preferred to the invalid robust variance estimates.

Extension

- Previously, we assume $\mathbf{\Sigma}_{\epsilon k}$ is known. This is applicable when conducting sensitivity analysis.
- When $\mathbf{\Sigma}_{\epsilon k}$ is unknown, we can estimate it using additional data sources: validation sample or repeated data. Alternatively, we can use the empirical SIMEX developed by Devanarayan and Stefanski (2002) when repeated measurements are available.
- In the situations where the heteroscedastic measurement error is unknown but repeated measurements are available, the empirical SIMEX is still applicable.

Acknowledgements

Di Shu was partially supported by the Canadian Institutes of Health Research (CIHR) Drug Safety and Effectiveness Network grant TD3-137716 through the scholarship from the CAnadian Network for Advanced Interdisciplinary Methods for comparative effectiveness research (CAN-AIM) team.

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

Some References

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, CRC press.

Cook, J. R. and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, **89,** 1314-1328.

Devanarayan, V. and Stefanski, L. A. (2002). Empirical simulation extrapolation for measurement error models with replicate measurements. *Statistics & Probability Letters* **59,** 219-225.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia: Society for industrial and applied mathematics, Vol. 38.

Some References

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, **1,** 221-233.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, **69,** 331-342.

Robins, J. M., Hernán, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, **11,** 550-560.

Stefanski, L. A. and Carroll. R. J. (1987). Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika*, **74,** 703-716.

Thank you for listening!