Corrigendum to "Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations" [Journal of Econometrics, Volume 189, 123, 2015]*

Max H. Farrell†

February 1, 2018

Abstract

This note provides a correction to Farrell (2015a) (present also in the supplement). Some results on treatment effect inference with a generic first-stage estimator, of Section 5 therein, requires an additional condition to be valid. A full proof of Farrell (2015a, Theorem 3.1) is given. Further, the additional condition is shown to hold for the group lasso estimators as proposed in Section 6 of the paper.

*I am grateful to Whitney Newey for alerting me to this error.

†University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, United States. Tel: +1 773-834-0161; Email: max.farrell@chicagobooth.edu; Web: http://faculty.chicagobooth.edu/max.farrell/.

1 Error in Original Derivation

Farrell (2015a, p. 15) [and Farrell (2015b, p. 3)] states that (see notation defined therein)

$$R_{21} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\mu}_t(x_i) - \mu_t(x_i)) \left(\frac{p_t(x_i) - d_i^t}{p_t(x_i)} \right)$$

is mean-zero conditional on $\{x_i\}_{i=1}^n$, which is false, as, in particular, $\hat{\mu}_t(x_i)$ depends on d_i^t, the treatment status indicator. This invalidates the proof given of Theorem 3.1 "without additional randomness". This is corrected below. Note that the case of "with additional randomness" (Farrell, 2015a, p. 16) is not affected by this error and remains valid as originally stated. See Section 5.1 for the definition and discussion of "additional randomness".

2 Restatement and Discussion of Assumption 3

To remedy the error, we need only add a condition to Assumption 3 (Farrell, 2015a, p. 8). The theorems of Section 5 do not otherwise require restatement. See the original paper for all notation. Assumption 3 should be replaced with the following version, which adds condition (c) but is otherwise identical.

Assumption 3 (First Stage Restrictions). *The estimators $\hat{p}_t(x)$ and $\hat{\mu}_t(x)$ obey the following for a sequence $\{P_n\}$, uniformly in $t \in \overline{\mathbb{N}}_{\mathcal{J}}$.*

(a) $\mathbb{E}_n[(\hat{p}_t(x_i) - p_t(x_i))^2] = o_{P_n}(1)$ *and* $\mathbb{E}_n\left[(\hat{\mu}_t(x_i) - \mu_t(x_i))^2\right] = o_{P_n}(1)$,

(b) $\mathbb{E}_n[(\hat{\mu}_t(x_i) - \mu_t(x_i))^2]^{1/2}\mathbb{E}_n[(\hat{p}_t(x_i) - p_t(x_i))^2]^{1/2} = o_{P_n}(n^{-1/2})$.

(c) $\mathbb{E}_n[(\hat{\mu}_t(x_i) - \mu_t(x_i))(1 - d_i^t/p_t(x_i))] = o_{P_n}(n^{-1/2})$.

The results of Section 5, Theorems 3 and 4 in particular, are now valid as stated under this condition. The "example" result stated in Theorem 1, in the overview section of the original paper, explicitly states conditions on first stage estimation rather than relying on Assumption 3, and thus requires the addition of condition (c) as well.

We note that this new condition is only needed when there is no "additional randomness" in the model selection step (see Section 5.1 of the paper for discussion). The cases of Theorems 3 and 4 "with additional randomness" do not require condition (c), as an alternative proof is used based on Assumptions 4 and 5 in the paper, and thus these cases are correct as originally stated.

3 Proof of Theorem 3.1

For completeness, we now give a proof of Theorem 3.1 "without additional randomness" under this new version of Assumption 3. This proof should replace the one given in Farrell (2015a, p. 15) and Farrell (2015b, p. 3). Assumptions 1 and 2 are given in Sections 3 and 4, respectively, of the original paper.

The proof relies on the fact that

$$\frac{1}{a} = \frac{1}{b} + \frac{b-a}{ab} = \frac{1}{b} + \frac{b-a}{b^2} + \frac{(b-a)^2}{ab^2}. \tag{1}$$

To begin, write $\sqrt{n}(\hat{\mu}_t - \mu_t) = \sqrt{n}\mathbb{E}_n[\psi_t(y_i, d_i^t, \mu_t(x_i), p_t(x_i), \mu_t)] + R_1 + R_2$, where

$$R_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} d_i^t (y_i - \mu_t(x_i)) \left(\frac{1}{\hat{p}_t(x_i)} - \frac{1}{p_t(x_i)} \right)$$

and

$$R_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\mu}_t(x_i) - \mu_t(x_i)) \left(1 - \frac{d_i^t}{\hat{p}_t(x_i)} \right).$$

The proof proceeds by showing that both R_1 and R_2 are $o_{P_n}(1)$. Applying the first equality in Eqn. (1), we rewrite R_1 as

$$R_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} d_i^t u_i \left(\frac{p_t(x_i) - \hat{p}_t(x_i)}{\hat{p}_t(x_i) p_t(x_i)} \right).$$

Applying Assumptions 1(b) and 2(c) and the first-stage consistency condition of Assumption 3(a):

$$\mathbb{E}\left[R_1^2 | \{x_i, d_i\}_{i=1}^n\right] = \mathbb{E}_n \left[\frac{d_i^t \sigma_t^2(x_i)}{\hat{p}_t(x_i)^2 p_t(x_i)^2} \left(p_t(x_i) - \hat{p}_t(x_i)\right)^2 \right] \leq C\mathbb{E}_n[(p_t(x_i) - \hat{p}_t(x_i))^2] = o_{P_n}(1).$$

Therefore R_1 is $o_{P_n}(1)$ by Markov's inquality.

Next, again using Eqn. (1), we have $R_2 = R_{21} + R_{22}$, where

$$R_{21} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\mu}_t(x_i) - \mu_t(x_i)) \left(\frac{p_t(x_i) - d_i^t}{p_t(x_i)} \right)$$

and

$$R_{22} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\mu}_t(x_i) - \mu_t(x_i))(\hat{p}_t(x_i) - p_t(x_i)) \left(\frac{d_i^t}{\hat{p}_t(x_i) p_t(x_i)} \right).$$

For the first term, $R_{21} = \sqrt{n}\mathbb{E}_n[(\hat{\mu}_t(x_i) - \mu_t(x_i))(1 - d_i^t/p_t(x_i))] = o_{P_n}(1)$ by Assumption 3(c). Next,

$$|R_{22}| \leq \sqrt{n} \left(\max_{i \leq n} \frac{1}{\hat{p}_t(x_i) p_t(x_i)} \right) \sqrt{\mathbb{E}_n[(\hat{\mu}_t(x_i) - \mu_t(x_i))^2]\mathbb{E}_n[(\hat{p}_t(x_i) - p_t(x_i))^2]} = o_{P_n}(1).$$

by Hölder's inequality, Assumption 1(b) and the rate condition of Assumption 3(b). □

4 Verification of Assumption 3 for Group Lasso Estimators

Under conditions imposed therein, Section 6 of the paper verifies that Assumptions 3(a) and 3(b) hold for the proposed group lasso estimators $\hat{\mu}_t(x_i)$ and $\hat{p}_t(x_i)$. Here we show that 3(c) holds also. No additional assumptions are required.

Recall that rather than generic $\hat{\mu}_t(x)$ and $\mu_t(x_i)$, as above, we are now explicitly considering high-dimensional approximately sparse linear models for $\mu_t(x_i)$. In this context, we add and subtract the pseudotrue values to write

$$\sqrt{n}\mathbb{E}_n[(\hat{\mu}_t(x_i) - \mu_t(x_i))(1 - d_i^t/p_t(x_i))] = A_1 + A_2,$$

where

$$A_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_i^* \beta_t^* - \mu_t(x_i))\left(\frac{p_t(x_i) - d_i^t}{p_t(x_i)}\right) \qquad \text{and} \qquad A_2 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_i^{*\prime}\hat{\beta}_t - x_i^* \beta_t^*)\left(\frac{p_t(x_i) - d_i^t}{p_t(x_i)}\right).$$

For the first term, $\mathbb{E}[A_1|\{x_i\}_{i=1}^n] = 0$ holds as β_t^* is nonrandom. From Assumption 1(b) and the definition of the bias term b_s^y we find

$$\mathbb{E}[A_1^2|\{x_i\}_{i=1}^n] = \frac{1}{n}\sum_{i=1}^{n}(x_i^*\beta_t^* - \mu_t(x_i))^2 \mathbb{E}\left[\left(\frac{p_t(x_i) - d_i^t}{p_t(x_i)}\right)^2\right] \leq C\mathbb{E}_n(x_i^*\beta_t^* - \mu_t(x_i))^2 \leq C(b_s^y)^2.$$

Therefore $|A_1| = O_{P_n}(b_s^y) = o_{P_n}(1)$, where the second equality is assumed in the bias condition of Assumption 4 and the first equality follows from Markov's inequality.

For the second term, define $\tilde{\Sigma}_{t,j} = \mathbb{E}\left[(x_{i,j}^*)^2(d_i^t - p_t(x_i))^2/p_t(x_i)^2\right]$ and then proceed as follows:

$$\begin{aligned}
A_1 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{p_t(x_i) - d_i^t}{p_t(x_i)}\right)\sum_{j\in\hat{S}_Y} x_{i,j}^*(\hat{\beta}_{t,j} - \beta_{t,j}^*) \\
&= \sum_{j\in\hat{S}_Y}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{x_{i,j}^*(p_t(x_i) - d_i^t)/p_t(x_i)}{\tilde{\Sigma}_{t,j}^{1/2}}\right\}\tilde{\Sigma}_{t,j}^{1/2}(\hat{\beta}_{t,j} - \beta_{t,j}^*) \\
&\leq \left(\max_{j\in\mathbb{N}_p}\tilde{\Sigma}_{t,j}^{1/2}\right)\left(\max_{j\in\mathbb{N}_p}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{x_{i,j}^*(p_t(x_i) - d_i^t)/p_t(x_i)}{\tilde{\Sigma}_{t,j}^{1/2}}\right)\left\|\hat{\beta}_t - \beta_t^*\right\|_1 \\
&= O(1)O_{P_n}(\log(p))\left\|\hat{\beta}_t - \beta_t^*\right\|_1.
\end{aligned}$$

This quantity is $o_{P_n}(1)$ by Corollary 5 in the original paper, which among other results, gives a rate for the ℓ_1 norm of the estimated coefficients. For the final equality, Assumptions 1(b), 2(b), and 2(c) imply that $\max_{j\in\mathbb{N}_p}\tilde{\Sigma}_{t,j} = O(1)$, while the center factor is bounded by applying the moderate deviation theory for self-normalized sums of de la Peña, Lai, and Shao (2009, Theorem 7.4) and in particular Belloni, Chen, Chernozhukov, and Hansen (2012, Lemma 5). To apply this theory, first note that the summand of the center factor has bounded third moment and second moment bounded away from zero from Assumptions 1(b) and 2. $\Sigma_{t,j}$ normalizes the second moment, and the theory applies under Assumption 4.

References

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen, 2012, Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica, 80, 2369–2429.

de la Peña, V. H., T. L. Lai, and Q.-M. Shao, 2009, Self-Normalized Processes: Limit Theory and Statistical Applications, Probability and Its Applications. Springer.

Farrell, M. H., 2015a, Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations. Journal of Econometrics, 189, 1–23.

——— , 2015b, Supplement to 'Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations'. Journal of Econometrics, 189.