Augmenting Pre-Analysis Plans with Machine Learning
ONLINE APPENDIX

Jens Ludwig   Sendhil Mullainathan   Jann Spiess

April 2019

Assume $\hat{f}$ with $\hat{f}(z) \in \mathbb{R}$ is fitted using K-fold cross-validation with fixed K, $\hat{f}(z) \to f(z)$ in large samples (in a way we will make precise later), and there are additional covariates $z^0 \in \mathbb{R}^\ell, z^* \in \mathbb{R}^m$. Let

$$v = \begin{pmatrix} z^0 \\ z^* \\ f(z) \end{pmatrix}, \qquad\qquad \hat{v} = \begin{pmatrix} z^0 \\ z^* \\ \hat{f}(z) \end{pmatrix}.$$

Write $\delta = (\alpha', \beta', \gamma)'$ for the population regression coefficient of y on v and $\hat{\delta} = (\hat{\alpha}', \hat{\beta}', \hat{\gamma})'$ for the sample regression coefficient of $Y = (y_1, \ldots, y_n)' \in \mathbb{R}^n$ on $\hat{V} = (\hat{v}_1, \ldots, \hat{v}_n)' \in \mathbb{R}^{n \times (\ell + m + 1)}$. Specifically,

$$\delta = \mathrm{E}[vv']^{-1}\mathrm{E}[vy], \qquad\qquad \hat{\delta} = (\hat{V}'\hat{V})^{-1}\hat{V}'Y.$$

Similarly, write $V = (v_1, \ldots, v_n)' \in \mathbb{R}^{n \times (\ell + m + 1)}, Z^0 = (z_1^0, \ldots, z_n^0)' \in \mathbb{R}^{n \times \ell}, Z^* = (z_1^*, \ldots, z_n^*)' \in \mathbb{R}^{n \times m}, F = (f(z_1), \ldots, f(z_n))' \in \mathbb{R}^n, \hat{F} = (\hat{f}_1(z_1), \ldots, \hat{f}_n(z_n))' \in \mathbb{R}^n, E = (\varepsilon_1, \ldots, \varepsilon_n)'.$

Assumption 1 (Sampling). *$(z_i^0, z_i^*, z_i, \varepsilon_i)$ are sampled iid from some fixed population distribution* P, *where* $\mathrm{E}\varepsilon = 0$ *and, for fixed f, $\mathrm{E}[v\varepsilon] = \boldsymbol{0}$. For $\delta = \delta_n$ (with $\delta_n \to \delta_\infty$), $y_i = v'\delta_n + \varepsilon$, yielding the sampling distribution* P_n *over* $Y, \hat{V}, V$.

Assumption 2 (Fitting). *The sample is randomly partitioned into K (which is fixed) approximately equally sized folds $\bigcup_{k=1}^K I_k = \{1, \ldots, n\}$ (where $\left||I_k| - \frac{n}{K}\right| < 1$). For $i \in I_k$, $\hat{f}_i = \hat{f}^k$, and $\hat{f}$ is a function only of data $\{(y_j, z_j^0, z_j^*, z_j)\}_{j \notin I_k}$ from the other folds.*

Assumption 3 (Convergence). *As $n \to \infty$, for all k, $\mathrm{E}[\mathrm{E}_n[(\hat{f}^k(z) - f(z))^2]] \to 0$ (where the expectations are over the sampling distribution P_n yielding $\hat{f}^k$, and over an additional, independent draw of z from* P*).*

Assumption 4 (Existence of moments). $\mathrm{E}[\epsilon^4], \mathrm{E}[\|v\|^4] < \infty$, *and there exists an integrable* $g(z)$ *with* $(\hat{f}(z) - f(z))^2 \le g(z)$ *for* P*-almost all* z *and* P_n*-a.s., uniformly in* n, *and* $\mathrm{E}[g(z)\epsilon^2], \mathrm{E}[g(z)\|z^0\|^2], \mathrm{E}[g(z)\|z^*\|^2] < \infty$.

Proposition 1 (Component consistency). *Under Assumptions 1, 2, 3, and 4,*

$$\hat{H} = \frac{1}{n}\hat{V}'\hat{V} \xrightarrow{\mathrm{P}_n} H = \mathrm{E}[vv'], \qquad\qquad \frac{1}{n}\hat{V}'Y \xrightarrow{\mathrm{P}_n} H\delta_\infty$$

as $n \to \infty$.

Proof. Note first that

$$\frac{1}{n}\hat{V}'\hat{V} = \frac{1}{n}(V'V + V'(\hat{V} - V) + (\hat{V} - V)'V + (\hat{V} - V)'(\hat{V} - V)),$$
$$\frac{1}{n}\hat{V}'Y = \frac{1}{n}(V'Y + (\hat{V} - V)'Y).$$

Since $\frac{1}{n}V'Y = \frac{1}{n}V'V\delta_n + \frac{1}{n}V'E$, by the LLN (using the existence of second moments) we know that

$$\frac{1}{n}V'V \xrightarrow{\mathrm{P}_n} \mathrm{E}[vv'] = H, \qquad\qquad \frac{1}{n}V'Y \xrightarrow{\mathrm{P}_n} \mathrm{E}[vv']\delta_\infty = H\delta_\infty.$$

It remains to show that the residual parts vanish. To this end, note that $(\hat{V} - V)'Y = (\hat{V} - V)'V\delta_n + (\hat{V} - V)'E$, and that by sub-multiplicativity of the Frobenius norm $\|\cdot\|$

$$\left\| \frac{1}{n}(\hat{V} - V)'V \right\|^2 \le \underbrace{\left\| (\hat{V} - V)/\sqrt{n} \right\|^2}_{=\|(\hat{F}-F)/\sqrt{n}\|^2} \|V/\sqrt{n}\|^2 = \frac{\|\hat{F} - F\|^2}{n} \frac{\|V\|^2}{n}$$

and similarly $\left\| \frac{1}{n}(\hat{V} - V)'E \right\|^2 \le \frac{\|\hat{F}-F\|^2}{n} \frac{\|E\|^2}{n}$. By the LLN,

$$\frac{\|E\|^2}{n} = \frac{1}{n}\sum_i^n \varepsilon_i^2 \xrightarrow{\mathrm{P}_n} \mathrm{E}[\varepsilon^2] < \infty \qquad\qquad \frac{\|V\|^2}{n} = \frac{1}{n}\sum_i^n \|v_i\|^2 \xrightarrow{\mathrm{P}_n} \mathrm{E}[\|v\|^2] < \infty.$$

The claims of the proposition then follow from

$$\mathrm{E}_n\left[\frac{\|\hat{F} - F\|^2}{n} \right] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}_n(\hat{f}_i(z_i) - f(z_i))^2$$

$$= \frac{1}{n}\sum_{k=1}^K \sum_{i \in I_k} \underbrace{\mathrm{E}_n(\hat{f}^k(z_i) - f(z_i))^2}_{=\mathrm{E}[\mathrm{E}_n[(\hat{f}^k(z)-f(z))^2]]} = \sum_{k=1}^K \frac{|I_k|}{n}\mathrm{E}[\mathrm{E}_n[(\hat{f}^k(z) - f(z))^2]]$$

which vanishes by Assumption 3, so $\frac{\|\hat{F}-F\|^2}{n} \xrightarrow{\mathrm{P}_n} 0$ and all residual parts vanish asymptotically. □

Proposition 2 (Component Normality). *Under Assumptions 1, 2, 3, and 4, for any fixed matrix $B \in \mathbb{R}^{M \times (\ell + m + 1)}$, writing b for the $\ell + m + 1$-th column of B, with*

   a. *$b = \boldsymbol{0}$ and $\mathrm{E}[Bv\hat{f}^k(z)] = \boldsymbol{0}$ for any k, P_n-almost all $\hat{f}^k$ **or***

   b. *$\limsup_{n \to \infty} \sqrt{n}|\gamma_n| < \infty$ and $\mathrm{E}[\hat{f}^k(z)\varepsilon] = \boldsymbol{0}$ for any k, P_n-almost all $\hat{f}^k$*

we have that

$$\frac{1}{\sqrt{n}} B\hat{V}'(Y - \hat{V}\delta_n) \overset{d}{\longrightarrow} \mathcal{N}(\boldsymbol{0}, BJB') \qquad\qquad (J = \mathrm{E}[\varepsilon^2 vv'])$$

as $n \to \infty$.

Proof. Writing $\Delta = \hat{F} - F \in \mathbb{R}^n$, we first decompose

$$\frac{B\hat{V}'(Y - \hat{V}\delta_n)}{\sqrt{n}} = B\frac{V'(Y - V\delta_n) + V'(V - \hat{V})\delta_n + (\hat{V} - V)'(Y - V\delta_n) + (\hat{V} - V)'(V - \hat{V})\delta_n}{\sqrt{n}}$$

$$= \frac{BV'E}{\sqrt{n}} + \frac{BV'\Delta\gamma_n}{\sqrt{n}} + \frac{B(\hat{V} - V)'E}{\sqrt{n}} + \frac{B(\hat{V} - V)'\Delta\gamma_n}{\sqrt{n}}$$

$$= \frac{BV'E}{\sqrt{n}} + \frac{BV'\Delta\gamma_n}{\sqrt{n}} + \frac{b\Delta'E}{\sqrt{n}} + \frac{b\Delta'\Delta\gamma_n}{\sqrt{n}}.$$

By the CLT,

$$\frac{BV'E}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Bv_i\varepsilon_i \overset{d}{\longrightarrow} \mathcal{N}(\boldsymbol{0}, \mathrm{Var}(Bv\varepsilon)) \qquad (\mathrm{Var}(Bv\varepsilon) = B\mathrm{Var}(v\varepsilon)B' = BJB').$$

It remains to show that the remaining parts vanish asymptotically. As a preliminary step, note that from $\mathrm{E}[x\hat{f}^k(z)] = \boldsymbol{0}$ for some square-integrable random variable x and P_n-almost all $\hat{f}^k$, it follows that $\mathrm{E}[xf(z)] = \boldsymbol{0}$. Indeed, P_n-almost surely

$$\mathrm{E}[xf(z)] = \underbrace{\mathrm{E}[x\hat{f}^k(z)]}_{=0} + \mathrm{E}[x(f(z) - \hat{f}^k(z))]$$

and $\|\mathrm{E}[x(f(z) - \hat{f}^k(z))]\|^2 \leq \mathrm{E}[\|x(f(z) - \hat{f}^k(z))\|]^2 \leq \mathrm{E}[\|x\|^2]\mathrm{E}[(f(z) - \hat{f}^k(z))^2] \to 0$ as $n \to 0$, so we must have $\mathrm{E}[xf(z)] = 0$. Turning now to the two cases of the proposition:

   a. The last two parts are zero. For the second part,

$$\frac{1}{\sqrt{n}} BV'\Delta\gamma_n = \frac{\gamma_n}{\sqrt{n}} \sum_{i=1}^{n} Bv_i(\hat{f}_i(z_i) - f(z_i)) = \frac{\gamma_n}{\sqrt{n}} \sum_{k=1}^{K} \sum_{i \in I_k} Bv_i(\hat{f}^k(z_i) - f(z_i))$$

3

with $\mathrm{E}_n[Bv_i(\hat{f}^k(z_i) - f(z_i))|\hat{f}^k] = \mathrm{E}[Bv(\hat{f}^k(z) - f(z))] = \mathbf{0}$ a.s. and thus

$$\mathrm{Var}_n\left(\sum_{i\in I_k} Bv_i(\hat{f}^k(z_i) - f(z_i))\right) = \mathrm{E}_n\mathrm{Var}_n\left(\sum_{i\in I_k} Bv_i(\hat{f}^k(z_i) - f(z_i))\,\middle|\,\hat{f}^k\right)$$

$$= \mathrm{E}_n\left[|I_k|\mathrm{E}[Bvv'B(\hat{f}^k(z) - f(z))^2]\right].$$

Hence, by Cauchy–Schwarz,

$$\left\|\mathrm{Var}_n\left(\frac{1}{\sqrt{n}}BV'\Delta\right)\right\| \le \frac{K}{n}\sum_{k=1}^{K}\left\|\mathrm{Var}_n\left(\sum_{i\in I_k} Bv_i(\hat{f}^k(z_i) - f(z_i))\right)\right\|$$

$$\le K\mathrm{E}_n\left[\sum_{k=1}^{K} K\frac{|I_k|}{n}\mathrm{E}[\|Bvv'B\|(\hat{f}^k(z) - f(z))^2]\right].$$

By dominated convergence, $\mathrm{E}_n\mathrm{E}[\|Bvv'B\|(\hat{f}^k(z) - f(z))^2] \to 0$. Since also $\gamma_n \to \gamma_\infty \in (-\infty, +\infty)$, $\frac{1}{\sqrt{n}}BV'\Delta\gamma_n \xrightarrow{\mathrm{P}_n} \mathbf{0}$. The statement of the proposition follows.

b. The third term vanishes by the same argument as in **a.**. For the second and fourth term,

$$\frac{BV'\Delta\gamma_n}{\sqrt{n}} = V\frac{V'\Delta}{n}(\sqrt{n}\gamma_n), \qquad\qquad \frac{b\Delta'\Delta\gamma_n}{\sqrt{n}} = b\frac{\Delta'\Delta}{n}(\sqrt{n}\gamma_n),$$

and $\frac{V'\Delta}{n} \xrightarrow{\mathrm{P}_n} \mathbf{0}$, $\frac{\Delta'\Delta}{n} \xrightarrow{\mathrm{P}_n} 0$ as in the proof of Proposition 1.

In both cases, we note that $\left\|\frac{B\hat{V}'(Y - \hat{V}\delta_n)}{\sqrt{n}} - \frac{BV'(Y - V\delta_n)}{\sqrt{n}}\right\| \xrightarrow{\mathrm{P}_n} 0$ together with convergence in distribution of $B\frac{V'(Y - V\delta_n)}{\sqrt{n}}$ implies the statement of the proposition. $\qquad\square$

For the linear regression to be well-behaved, we assume that regressors are (asymptotically) not collinear. This is a technical restriction that puts assumptions on ML-generated regressors that we would like to avoid. We believe that this assumption can be relaxed in the present context (by replacing inverses by generalized inverses and deriving analogous results for that case), but such an extension goes beyond the scope of this appendix.

Assumption 5 (No collinearity). *$H = \mathrm{E}[vv']$ is invertible.*

Putting all results together, we obtain the asymptotic distribution of ML-augmented linear regression in the cases discussed in the main paper:

Proposition 3 (Asymptotic distribution for ML-augmented linear regression). *Under Assumptions 1, 2, 3, 4, and 5, as $n \to \infty$):*

1. $\hat{\delta} \xrightarrow{\mathrm{P}_n} \delta_\infty$;

2. *If* $\limsup_{n\to\infty} \sqrt{n}|\gamma_n| < \infty$ *and* $E[\varepsilon \hat{f}^k(z)] = 0$ *for any* k, P_n*-almost all* $\hat{f}^k$*, then*

$$\sqrt{n}(\hat{\delta} - \delta_n) \xrightarrow{d} \mathcal{N}(\mathbf{0}, H^{-1}JH^{-1});$$

3. *If* $E[z^* \hat{f}^k(z)] = \mathbf{0}$ *for any* k, P_n*-almost all* $\hat{f}^k$*, and also* $E[z^0(z^*)'] = \mathbb{O}$*, then*

$$\sqrt{n}(\hat{\beta} - \beta_n) \xrightarrow{d} \mathcal{N}(\mathbf{0}, AH^{-1}JH^{-1}A'),$$

where $A \in \mathbb{R}^{m \times (\ell+m+1)}$ *selects the components* $\beta = A\delta$*.*

Proof. The first statement follows from Proposition 1 and Assumption 5. The second statement is immediate from Proposition 1, Proposition 2 (with B the identity), and Assumption 5 by noting that

$$\sqrt{n}(\hat{\delta} - \delta_n) = \left(\frac{1}{n}\hat{V}'\hat{V}\right)^{-1} \frac{1}{\sqrt{n}}\hat{V}'(Y - \hat{V}\delta).$$

For the third statement, note that $E[z^*\hat{f}^k(z)] = \mathbf{0}$, $E[z^*(z^0)'] = \mathbb{O}$ imply that $B = AH^{-1}$ fulfils the assumptions of Part **a.** in Proposition 2. The conclusion follows by decomposing

$$\sqrt{n}(\hat{\beta} - \beta_n) = A\sqrt{n}(\hat{\delta} - \delta_n) = A\left(\frac{1}{n}\hat{V}'\hat{V}\right)^{-1} \frac{1}{\sqrt{n}}\hat{V}'(Y - \hat{V}\delta)$$

$$= \frac{1}{\sqrt{n}} \underbrace{AH^{-1}}_{=B} \hat{V}'(Y - \hat{V}\delta) + \frac{1}{\sqrt{n}}\left(A\left(\frac{1}{n}\hat{V}'\hat{V}\right)^{-1} - AH^{-1}\right)\hat{V}'(Y - \hat{V}\delta)$$

where Proposition 2 applies to $\frac{1}{\sqrt{n}}B\hat{V}'(Y - \hat{V}\delta)$. It remains to show that the remainder vanishes. To that end, writing $\hat{V} = (\hat{V}_1, \hat{V}_2)$ with $\hat{V}_1 = Z^*, \hat{V}_2 = (Z^0, \hat{F})$ (i.e., changing the order of columns for ease of exposition) we have that (using general results for the inverses of block matrices)

$$A\left(\hat{V}'\hat{V}\right)^{-1} = \left(\left(\hat{V}_1'\hat{V}_1\right)^{-1} \quad \mathbb{O}\right) + \underbrace{(\hat{V}_1'\hat{V}_1)^{-1}\hat{V}_1'\hat{V}_2(\hat{V}_2'\hat{V}_2 - \hat{V}_2'\hat{V}_1(\hat{V}_1'\hat{V}_1)^{-1}\hat{V}_1'\hat{V}_2)^{-1}\left(\hat{V}_2'\hat{V}_1(\hat{V}_1'\hat{V}_1)^{-1} \quad -\mathbb{I}\right)}_{=\hat{r}}.$$

Also writing H_1, H_2 for the respective sub-matrices of H along the diagonal, by Proposition 1

$$\frac{1}{n}\hat{V}_1'\hat{V}_1 \xrightarrow{P_n} H_1, \qquad \frac{1}{n}\hat{V}_1'\hat{V}_2 \xrightarrow{P_n} \mathbb{O} = \mathbb{O}, \qquad \frac{1}{n}\hat{V}_2'\hat{V}_2 \xrightarrow{P_n} H_2.$$

Also, for $\Delta = \hat{F} - F$,

$$\frac{1}{\sqrt{n}}\hat{V}_1'\hat{V}_2 = \left(\frac{(Z^*)'Z^0}{\sqrt{n}} \quad \frac{(Z^*)'F}{\sqrt{n}} + \frac{(Z^*)'\Delta}{\sqrt{n}}\right)$$

5

where $\frac{(Z^*)'\Delta}{\sqrt{n}} \xrightarrow{\mathrm{P}_n} \mathbf{0}$ as in the proof of Proposition 2, and thus by the CLT under the conditions on moments in Assumption 4 (and orthogonality)

$$\frac{1}{\sqrt{n}}\mathrm{vec}\left(\hat{V}_1'\hat{V}_2\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Sigma_{\hat{V}_1'\hat{V}_2}\right)$$

for an appropriate variance matrix $\Sigma_{\hat{V}_1'\hat{V}_2}$. Consequently, and putting everything together,

$$\frac{1}{\sqrt{n}}\left(A\left(\frac{1}{n}\hat{V}'\hat{V}\right)^{-1} - AH^{-1}\right)\hat{V}'(Y - \hat{V}\delta)$$

$$= \underbrace{\left(\frac{1}{n}\hat{V}_1'\hat{V}_1 - H_1\right)}_{\xrightarrow{\mathrm{P}_n}\mathbf{0}} \overbrace{\frac{1}{\sqrt{n}}A\hat{V}'(Y - \hat{V}\delta)}^{\text{converges by Proposition 2}} + \underbrace{n^{3/2}\hat{r}}_{\text{converges}} \overbrace{\frac{\hat{V}'(Y - \hat{V}\delta)}{n}}^{\xrightarrow{\mathrm{P}_n}\mathbf{0}\text{ by Proposition 1}} \xrightarrow{\mathrm{P}_n} \mathbf{0}$$

where we have applied the above to

$$n^{3/2}\hat{r} = \left(\frac{\hat{V}_1'\hat{V}_1}{n}\right)^{-1} \frac{\hat{V}_1'\hat{V}_2}{\sqrt{n}} \left(\frac{\hat{V}_2'\hat{V}_2}{n} - \frac{\hat{V}_2'\hat{V}_1}{n}\left(\frac{\hat{V}_1'\hat{V}_1}{n}\right)^{-1}\frac{\hat{V}_1'\hat{V}_2}{n}\right)^{-1}\left(\frac{\hat{V}_2'\hat{V}_1}{n}\left(\frac{\hat{V}_1'\hat{V}_1}{n}\right)^{-1} \quad -\mathbb{I}\right).$$

This concludes the proof of the third part, and thus of the proposition. $\qquad\square$

Inference and testing can be performed based on the standard (Eicker–Huber–White) heteroscedasticity-robust estimators of the OLS variance matrix. To that end, let

$$\hat{J} = \frac{1}{n}\sum_{i=1}^{n}\hat{v}_i(y_i - \hat{v}_i'\hat{\delta})^2\hat{v}_i'.$$

The following result establishes valid asymptotic inference based on $\hat{\Sigma} = \hat{H}^{-1}\hat{J}\hat{H}^{-1}$:

Remark 1 (Asymptotic inference)**.** *Under the assumptions of Proposition 3, and also* $\mathrm{E}[g(z)^2] < \infty$ *in Assumption 4,*

$$\hat{H} \xrightarrow{\mathrm{P}_n} H, \qquad\qquad\qquad \hat{J} \xrightarrow{\mathrm{P}_n} J,$$

and thus $\hat{\Sigma} = \hat{H}^{-1}\hat{J}\hat{H}^{-1} \xrightarrow{\mathrm{P}_n} H^{-1}JH^{-1} = \Sigma.$

Proof. Convergence of $\hat{H}$ is established in Proposition 1. For $\hat{J}$, as a preliminary step,

$$\hat{J} - \frac{1}{n}\sum_{i=1}^{n}\hat{v}_i(y_i - \hat{v}_i'\delta)^2\hat{v}_i' = \frac{1}{n}\sum_{i=1}^{n}\hat{v}_i\left(2(y_i - \hat{v}_i'\delta)\hat{v}_i'(\hat{\delta} - \delta) + (\hat{v}_i'(\hat{\delta} - \delta))^2\right)\hat{v}_i',$$

so the norm of this difference is bounded by

$$\|\hat{\delta} - \delta\| \frac{1}{n} \sum_{i=1}^{n} 2\|\hat{v}_i\|^3 \|y_i - \hat{v}_i'\delta\| + \|\hat{\delta} - \delta\|^2 \frac{1}{n} \sum_{i=1}^{n} \|\hat{v}_i\|^4.$$

Since $\|\hat{\delta} - \delta\| \xrightarrow{\mathrm{P}_n} 0$ by Proposition 3 and the sum terms have uniformly bounded expectation, the difference converges to zero in probability by Markov's inequality. We can therefore consider directly

$$\frac{1}{n} \sum_{i=1}^{n} \hat{v}_i \underbrace{(y_i - \hat{v}_i'\delta)^2}_{=\hat{\varepsilon}_i} \hat{v}_i' - \frac{1}{n} \underbrace{\sum_{i=1}^{n} v_i \varepsilon_i^2 v_i'}_{\xrightarrow{\mathrm{P}_n} J}$$

$$= \frac{1}{n} \sum_{i=1}^{n} (\hat{v}_i \hat{v}_i' - v_i v_i')(\hat{\varepsilon}_i^2 - \varepsilon_i^2) + v_i v_i'(\hat{\varepsilon}_i^2 - \varepsilon_i^2) + (\hat{v}_i \hat{v}_i' - v_i v_i')\varepsilon^2$$

where, writing $\Delta = \hat{F} - F$ as in earlier proofs,

$$\hat{\varepsilon}_i - \varepsilon_i = \Delta_i \gamma$$
$$\hat{\varepsilon}_i^2 - \varepsilon_i^2 = (\hat{\varepsilon}_i - \varepsilon_i)(\hat{\varepsilon}_i + \varepsilon_i) = (\hat{\varepsilon}_i - \varepsilon_i)^2 + 2(\hat{\varepsilon}_i - \varepsilon_i)\varepsilon_i = \Delta_i^2 \gamma^2 + 2\Delta_i \gamma \varepsilon_i$$
$$\|\hat{v}_i \hat{v}_i' - v_i v_i'\| = \Delta_i^2 + 2|\Delta_i|\|v_i\|.$$

Consequently,

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \hat{v}_i (y_i - \hat{v}_i'\delta)^2 \hat{v}_i' - \frac{1}{n} \sum_{i=1}^{n} v_i \varepsilon_i^2 v_i' \right\|$$
$$\leq \frac{1}{n} \sum_{i=1}^{n} (\Delta_i^2 + 2|\Delta_i|\|v_i\|)(\Delta_i^2 \gamma^2 + 2|\Delta_i|\|\gamma\|\varepsilon_i|) + v_i v_i'(\Delta_i^2 \gamma^2 + 2|\Delta_i|\|\gamma\|\varepsilon_i|) + (\Delta_i^2 + 2|\Delta_i|\|v_i\|)\varepsilon^2.$$

Since $\Delta_i^2 \leq g(z_i)$ and the expectation of the resulting upper bound for the above difference is uniformly bounded, this difference converges to zero in mean (and thus in probability) by dominated convergence. Hence,

$$\hat{J} = \frac{1}{n} \sum_{i=1}^{n} v_i \varepsilon_i^2 v_i' + o_{\mathrm{P}_n} \xrightarrow{\mathrm{P}_n} J,$$

finishing the proof. □

Remark 2 (Asymptotic power and variance). *Write $\delta_n^b = ((\alpha_n^b)', (\beta_n^b)')'$ for the baseline population regression coefficient of y on $v^b = ((z^0)', (z^*)')'$, and $\hat{\delta}^b = ((\hat{\alpha}^b)', (\hat{\beta}^b)')'$ for their sample analog, and maintain the assumptions of Proposition 3.*

1. *Then, $\hat{\delta}^b \xrightarrow{\mathrm{P}_n} \delta_\infty^b = \lim_{n\to\infty} \delta_n^b$ where*

$$\delta_n^b = \begin{pmatrix} \alpha_n \\ \beta_n \end{pmatrix} + \mathrm{E}^{-1}[v^b(v^b)']\mathrm{E}[v^b f(z)]\gamma_n$$

and also

$$\sqrt{n}(\hat{\delta}^b - \delta_n^b) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (H^b)^{-1}J^b(H^b)^{-1}), \quad H^b = \mathrm{E}[v^b(v^b)'], \ J^b = \mathrm{E}[\underbrace{(y - (v^b)'\delta_\infty^b)}_{\varepsilon + \left(f(z) - (v^b)'(H^b)^{-1}\mathrm{E}[v^b f(z)]\right)\gamma_\infty}{}^2 v^b(v^b)'].$$

Specifically, this implies for the relative performance of the augmented OLS regression:

2. *Under the assumptions of **2.** in Proposition 3 and Remark 1, for $\beta_n = \frac{\beta}{\sqrt{n}}, \gamma_n = \frac{\gamma}{\sqrt{n}}$ (implying $\beta_n^b = \frac{\beta^b}{\sqrt{n}}$), the Wald test statistics*

$$\hat{x} = \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix}' \hat{\Sigma}_{\beta,\gamma}^{-1} \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix}, \qquad\qquad \hat{x}^b = (\hat{\beta}^b)' \left(\hat{\Sigma}_\beta^b\right)^{-1} \hat{\beta}^b$$

based on the corresponding (robust) variance estimates $\hat{\Sigma}, \hat{\Sigma}^b$ have asymptotic distributions

$$\hat{x} \xrightarrow{d} \chi_{m+1}^2\left(\begin{pmatrix} \beta \\ \gamma \end{pmatrix}' \Sigma_{\beta,\gamma}^{-1} \begin{pmatrix} \beta \\ \gamma \end{pmatrix}\right), \qquad\qquad \hat{x}^b \xrightarrow{d} \chi_m^2\left((\beta^b)' \left(\Sigma_\beta^b\right)^{-1} \beta^b\right).$$

Since we can verify that

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix}' \Sigma_{\beta,\gamma}^{-1} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} \geq (\beta^b)' \left(\Sigma_\beta^b\right)^{-1} \beta^b,$$

the maximal cost in power from including the additional regressor corresponds to one additional degree of freedom in the χ^2 distribution that is used to construct critical values.

3. *Under the assumptions of **3.** in Proposition 3, with $\gamma_n \to \gamma = \gamma_\infty$,*

$$\sqrt{n}(\hat{\beta} - \beta_n) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_\beta), \qquad\qquad \sqrt{n}(\hat{\beta}^b - \beta_n) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_\beta^b)$$

with

$$\Sigma_\beta = \mathrm{E}^{-1}[z^*(z^*)']\mathrm{E}[z^*\varepsilon^2(z^*)']\mathrm{E}^{-1}[z^*(z^*)']$$

$$\Sigma_\beta^b = \mathrm{E}^{-1}[z^*(z^*)']\mathrm{E}[z^*(\varepsilon + \underbrace{(f(z) - (z^0)'\mathrm{E}^{-1}[z^0(z^0)^{-1}]\mathrm{E}[z^0 f(z)])}_{=f^\perp(z)}\gamma)^2(z^*)']\mathrm{E}^{-1}[z^*(z^*)'],$$

so the asymptotic variance decreases, $\Sigma_\beta \prec \Sigma_\beta^b$ (wrt the partial ordering implied by positive definiteness) relative to the baseline OLS regression iff $\mathrm{E}[z^*\varepsilon^2(z^*)'] \prec \mathrm{E}[z^*(\varepsilon + f^\perp(z)\gamma)^2(z^*)']$.