

In the paper

- Henmi, M. and Eguchi, S, (2004), A paradox concerning nuisance parameters and projected estimating functions, *Biometrika*, 91(4), 929–941,

the authors explain a paradox that arises in estimation based on estimating functions. In this note I will adopt the same notation as in the paper, but simplify where possible. Here are the components:

1. **Random variable:** X could be multi-dimensional.

2. **Probability model:** $p(x; \theta, \kappa)$

- $\theta = (\beta^\top, \alpha^\top)^\top$ d -dimensional parameter;
 - β is the *parameter of interest*,
 - α is the *nuisance parameter*.
- κ finite or infinite dimensional nuisance parameter, not estimated during the procedure.
- **Score function:** $s(x, \theta, \kappa)$ with

$$s(x, \theta, \kappa) = \frac{\partial}{\partial \theta} \{\log p(x; \theta, \kappa)\} = \begin{bmatrix} s_\beta(x, \theta, \kappa) \\ s_\alpha(x, \theta, \kappa) \end{bmatrix}$$

with s_β and s_α being the components of the score obtained by differentiating with respect to β and α respectively.

- true values of the parameters θ_0 with components β_0 and α_0 , and κ_0 .

3. **Estimating function:** $u(x, \theta)$ ($d \times 1$), with

$$\mathbb{E}[u(X, \theta)] = \mathbf{0} \quad \mathbb{E}[\|u(X, \theta)\|^2] = \mathbb{E}[\{u(X, \theta)\}^\top u(X, \theta)] < \infty$$

Note: It is not assumed here that $u(x, \theta)$ is differentiable with respect to θ ; this is therefore a weaker assumption compared to the one made in the treatment of m -estimation in lectures. However, another regularity assumption (equation (6) in the paper) is made.

4. **Projected estimating function:** The projected estimating function $u_*(x, \theta_0)$ is the solution to the minimization problem

$$u_*(x, \theta_0) = \arg \min_{v=\mathbf{A}u} \mathbb{E}[\|s(X, \theta_0, \kappa_0) - v(X, \theta_0)\|^2]$$

where $\mathbf{A}$ is an arbitrary deterministic $d \times d$ matrix. That is, $u_*(x, \theta_0)$ is the projection of score function $s(x, \theta_0, \kappa_0)$ onto the linear subspace

$$\Lambda = \{\mathbf{A}u(x, \theta_0) : \mathbf{A} \text{ a } d \times d \text{ matrix}\}$$

taken by forming arbitrary matrix multiples of $u(x, \theta)$. By the projection theorem for Hilbert spaces and its required orthogonality condition, we therefore have that

$$\begin{aligned} u_*(x, \theta_0) &= \mathbb{E}[s(X, \theta_0, \kappa_0)u(X, \theta_0)^\top] \{\mathbb{E}[u(X, \theta_0)u(X, \theta_0)^\top]\}^{-1} u(x, \theta_0) \\ &= W(\theta_0, \kappa_0)^\top \{V(\theta_0, \kappa_0)\}^{-1} u(x, \theta_0) \end{aligned} \tag{1}$$

say, with

$$W(\theta_0, \kappa_0) = \mathbb{E}[u(X, \theta_0)s(X, \theta_0, \kappa_0)^\top] \quad V(\theta_0, \kappa_0) = \mathbb{E}[u(X, \theta_0)u(X, \theta_0)^\top].$$

5. **Estimator:** $\hat{\theta}$ is the solution to both estimating equations

$$\sum_{i=1}^n u(x_i, \theta) = \mathbf{0} \quad \text{and} \quad \sum_{i=1}^n u_*(x_i, \theta) = \mathbf{0}.$$

6. **Asymptotic behaviour:** under regularity conditions, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \text{Normal}_d(\mathbf{0}, \text{avar}(\hat{\theta}))$$

where

$$\text{avar}(\hat{\theta}) = \{W(\theta_0, \kappa_0)\}^{-1} V(\theta_0, \kappa_0) \{W(\theta_0, \kappa_0)\}^{-\top} = \{\text{var}[u_*(x, \theta_0)]\}^{-1} \quad (2)$$

Note: V and W are the analogues of the $\mathcal{I}$ and $\mathcal{J}$ matrices from lecture notes. The definition of W is different from that of $\mathcal{J}$, which would here be

$$\mathcal{J}(\theta_0) = -\mathbb{E}[\dot{u}(X, \theta_0)],$$

as this assumes that $u(x, \theta)$ is differentiable (wrt θ). However, assuming differentiability, we have

$$\begin{aligned} \mathbb{E}[\dot{u}(X, \theta_0)] &= \int \frac{\partial u(x, \theta)}{\partial \theta^\top} \Big|_{\theta=\theta_0} p(x; \theta_0, \kappa_0) dx \\ &= \int \frac{\partial}{\partial \theta^\top} \{u(x, \theta) p(x; \theta, \kappa_0)\} \Big|_{\theta=\theta_0} dx - \int u(x, \theta_0) \frac{\partial p(x; \theta, \kappa_0)}{\partial \theta^\top} \Big|_{\theta=\theta_0} dx \\ &= \frac{\partial}{\partial \theta^\top} \left\{ \int u(x, \theta) p(x; \theta, \kappa_0) dx \right\} \Big|_{\theta=\theta_0} - \int u(x, \theta_0) \frac{\partial \log p(x; \theta, \kappa_0)}{\partial \theta^\top} \Big|_{\theta=\theta_0} p(x; \theta_0, \kappa_0) dx \\ &= \mathbf{0} - \int u(x, \theta_0) \{s(x, \theta_0, \kappa_0)\}^\top p(x; \theta_0, \kappa_0) dx \\ &= -\mathbb{E}[u(X, \theta_0) \{s(X, \theta_0, \kappa_0)\}^\top] \end{aligned}$$

where

- line 2 follows by the chain rule
- line 3 follows exchanging integration and differentiation, and because

$$\frac{d}{d\theta} \{\log g(\theta)\} = \frac{\dot{g}(\theta)}{g(\theta)}$$

- line 4 follows by definition of $s(x, \theta_0, \kappa_0)$, and as

$$\int u(x, \theta) p(x; \theta, \kappa_0) dx = \mathbb{E}[u(X, \theta_0)] = \mathbf{0}.$$

Therefore, if $u(x, \theta)$ is differentiable

$$W(\theta_0, \kappa_0) = \mathbb{E} \left[u(X, \theta_0) \{s(X, \theta_0, \kappa_0)\}^\top \right] = -\mathbb{E} [\dot{u}(X, \theta_0)] = \mathcal{J}(\theta_0).$$

7. **Partitioning the estimating functions:** Write

$$u(x, \theta_0) = \begin{bmatrix} u_\beta(x, \theta_0) \\ u_\alpha(x, \theta_0) \end{bmatrix} \quad u_*(x, \theta_0) = \begin{bmatrix} u_{*\beta}(x, \theta_0) \\ u_{*\alpha}(x, \theta_0) \end{bmatrix}$$

where $u_{*\beta}(x, \theta_0)$ and $u_{*\alpha}(x, \theta_0)$ are the orthogonal projections of $s_\beta(x, \theta_0, \kappa_0)$ and $s_\alpha(x, \theta_0, \kappa_0)$ onto the space spanned by $u(x, \theta_0)$, respectively.

8. **Projecting the estimating function components:** Suppose we want to compute the projection of $u_{*\beta}(x, \theta_0)$ onto the space spanned by $u_{*\alpha}(x, \theta_0)$, that is, the space

$$\mathcal{U}_{*\alpha} = \{\mathbf{B}u_{*\alpha}(x, \theta_0) : \mathbf{B} \text{ an arbitrary deterministic matrix}\}.$$

We have from results in lectures that this projection is

$$\Pi(u_{*\beta}(x, \theta_0)|\mathcal{U}_{*\alpha}) = \mathbb{E}[u_{*\beta}(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top] \{\mathbb{E}[u_{*\alpha}(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top]\}^{-1} u_{*\alpha}(x, \theta_0)$$

and therefore the ‘residual’ is

$$\begin{aligned} u_{*\beta}^{\text{res}}(x, \theta_0) &= u_{*\beta}(X, \theta_0) - \mathbb{E}[u_{*\beta}(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top] \{\mathbb{E}[u_{*\alpha}(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top]\}^{-1} u_{*\alpha}(x, \theta_0) \\ &= u_{*\beta}(X, \theta_0) - \mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha}^{-1} u_{*\alpha}(x, \theta_0) \end{aligned}$$

say. Then

$$\begin{aligned} \text{var}[u_{*\beta}^{\text{res}}(X, \theta_0)] &= \mathbb{E} \left[u_{*\beta}^{\text{res}}(X, \theta_0) u_{*\beta}^{\text{res}}(X, \theta_0)^\top \right] \\ &= \mathbb{E} \left[(u_{*\beta}(X, \theta_0) - \mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha}^{-1} u_{*\alpha}(X, \theta_0)) (u_{*\beta}(X, \theta_0) - \mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha}^{-1} u_{*\alpha}(X, \theta_0))^\top \right] \\ &= \mathbb{E} \left[u_{*\beta}(X, \theta_0) u_{*\beta}(X, \theta_0)^\top \right] - \mathbb{E} [u_{*\beta}(X, \theta_0) u_{*\alpha}(X, \theta_0)^\top \mathbf{V}_{\alpha\alpha}^{-1} \mathbf{V}_{\alpha\beta}] \\ &\quad - \mathbb{E} [\mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha} u_{*\alpha}(X, \theta_0) u_{*\beta}(X, \theta_0)^\top] + \mathbb{E} [\mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha} u_{*\alpha}(X, \theta_0) u_{*\alpha}(X, \theta_0)^\top \mathbf{V}_{\alpha\alpha}^{-1} \mathbf{V}_{\alpha\beta}] \\ &= \mathbf{V}_{\beta\beta} - \mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha}^{-1} \mathbf{V}_{\alpha\beta} - \mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha}^{-1} \mathbf{V}_{\alpha\beta} + \mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha}^{-1} \mathbf{V}_{\alpha\alpha} \mathbf{V}_{\alpha\alpha}^{-1} \mathbf{V}_{\alpha\beta} \\ &= \mathbf{V}_{\beta\beta} - \mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha}^{-1} \mathbf{V}_{\alpha\beta} \end{aligned}$$

9. **The asymptotic variance of $\hat{\beta}$:** We have

$$\begin{aligned} \text{avar}(\hat{\theta}) &= \{\text{var}[u_*(X, \theta_0)]\}^{-1} = \left\{ \begin{bmatrix} \mathbb{E}[u_{*\beta}(X, \theta_0)u_{*\beta}(X, \theta_0)^\top] & \mathbb{E}[u_{*\beta}(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top] \\ \mathbb{E}[u_{*\alpha}(X, \theta_0)u_{*\beta}(X, \theta_0)^\top] & \mathbb{E}[u_{*\alpha}(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top] \end{bmatrix} \right\}^{-1} \\ &= \left\{ \begin{bmatrix} \mathbf{V}_{\beta\beta} & \mathbf{V}_{\beta\alpha} \\ \mathbf{V}_{\alpha\beta} & \mathbf{V}_{\alpha\alpha} \end{bmatrix} \right\}^{-1} \end{aligned}$$

say, using the notation from lectures. By extracting the relevant block in the matrix, we obtain the result that

$$\text{avar}(\hat{\beta}) = \{\mathbf{V}_{\beta\beta} - \mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha}^{-1} \mathbf{V}_{\alpha\beta}\}^{-1} \geq \mathbf{V}_{\beta\beta}^{-1} \equiv \{\text{var}[u_{*\beta}(X, \theta_0)]\}^{-1} \quad (3)$$

that is, where the “ $\geq$ ” is interpreted as indicating that

$$\{\mathbf{V}_{\beta\beta} - \mathbf{V}_{\beta\alpha} \mathbf{V}_{\alpha\alpha}^{-1} \mathbf{V}_{\alpha\beta}\}^{-1} - \mathbf{V}_{\beta\beta}^{-1}$$

is non-negative definite. Therefore

$$\text{avar}(\hat{\beta}) = \{\text{var}[u_{*\beta}^{\text{res}}(X, \theta_0)]\}^{-1} \geq \{\text{var}[u_{*\beta}(X, \theta_0)]\}^{-1}$$

with equality if and only if

$$\mathbf{V}_{\beta\alpha} = \mathbb{E}[u_{*\beta}(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top] = \mathbf{0}$$

which is equivalent to the two identities

$$\mathbb{E}[u_{*\beta}(X, \theta_0)s_\alpha(X, \theta_0, \kappa_0)^\top] = \mathbf{0} \quad \mathbb{E}[s_\beta(X, \theta_0, \kappa_0)u_{*\alpha}(X, \theta_0)^\top] = \mathbf{0}.$$

10. **Estimating with α known:** Suppose that $\tilde{\beta}$ is the solution to

$$\sum_{i=1}^n u(x_i, \beta, \alpha_0) = \mathbf{0}.$$

Therefore we have two possible estimators for β : $\hat{\beta}$ and $\tilde{\beta}$.

11. **Comparing the variances:** If the projected estimating equation $u_*(x, \theta_0)$ is **identical** to the original estimating equation $u(x, \theta_0)$, then it follows that

$$\text{avar}(\tilde{\beta}) = \{\text{var}[u_{*\beta}(X, \theta_0)]\}^{-1}.$$

so therefore from equation (3)

$$\text{avar}(\hat{\beta}) \geq \text{avar}(\tilde{\beta}).$$

Now consider the projected estimating equation that would arise by considering $u_\beta(x, \theta_0)$ alone: denote this projected estimating equation $u_{\beta*}(x, \theta_0)$, and recall from the definition that

$$u_{\beta*}(x, \theta_0) = \arg \min_{v = \mathbf{A}u_\beta} \mathbb{E}[\|s_\beta(X, \theta_0, \kappa_0) - v(X, \theta_0)\|^2]$$

where $\mathbf{A}$ is an arbitrary deterministic matrix. That is, $u_{\beta*}(x, \theta_0)$ is the projection of score function $s_\beta(x, \theta_0, \kappa_0)$ onto the linear subspace

$$\Lambda_\beta = \{\mathbf{A}u_\beta(x, \theta_0) : \mathbf{A} \text{ a deterministic matrix}\}$$

Recalling the definition of $u_{*\beta}(x, \theta_0)$ as the projection onto the larger space Λ (that is, $\Lambda_\beta \subset \Lambda$), Figure 1 in the paper helps to visualize the relationship between these functions.

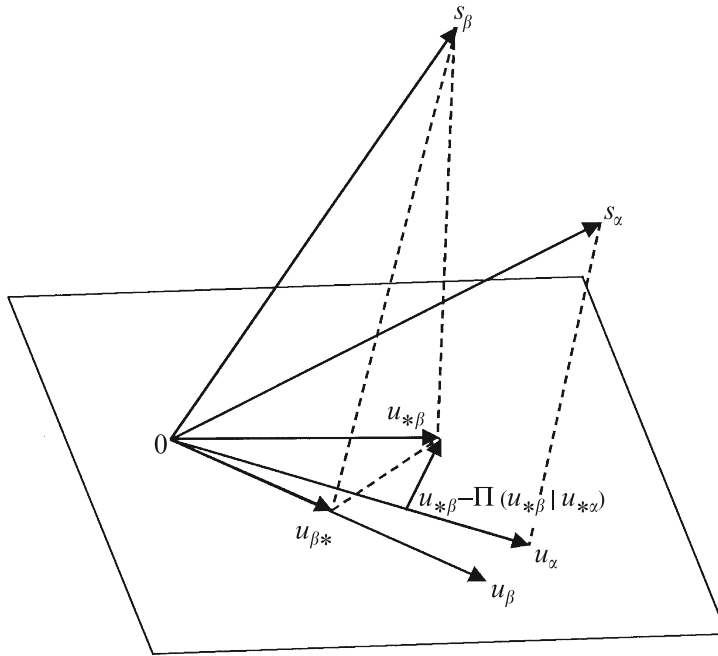


Fig. 1. The relationships between $u_{*\alpha}$, $u_{\beta*}$ and $u_{*\beta}$.

In the Figure, the plane is the linear subspace Λ : evidently $u_{*\beta} \in \Lambda$, and $u_{\beta*} \in \Lambda_\beta \subset \Lambda$.

Now we have four right-angle triangles to consider

- (1) $0 \rightarrow u_{*\beta} \rightarrow s_\beta$ (as $(s_\beta - u_{*\beta}) \perp \Lambda$)
- (2) $0 \rightarrow u_{\beta*} \rightarrow s_\beta$ (as $(s_\beta - u_{\beta*}) \perp \Lambda_\beta$)
- (3) $u_{\beta*} \rightarrow u_{*\beta} \rightarrow s_\beta$
- (4) $0 \rightarrow u_{\beta*} \rightarrow u_{*\beta}$ (as $(u_{*\beta} - u_{\beta*}) \perp \Lambda_\beta$)

where the right-angle is at the middle corner of the three. Therefore it is evident (and can be proved using Pythagoras's theorem) that

$$u_{\beta*}(x, \theta_0) = \Pi(u_{*\beta}(x, \theta_0) | \Lambda_\beta).$$

Therefore, in full generality, we have that

$$\text{avar}(\tilde{\beta}) = \{\text{var}[u_{\beta*}(X, \theta_0)]\}^{-1} = \{\text{var}[\Pi(u_{*\beta}(X, \theta_0) | \Lambda_\beta)]\}^{-1}$$

In summary, we have

$$\begin{aligned} \text{avar}(\hat{\beta}) &= \{\text{var}[u_{*\beta}(X, \theta_0) - \Pi(u_{*\beta}(X, \theta_0) | \mathcal{U}_{*\alpha})]\}^{-1} \\ \text{avar}(\tilde{\beta}) &= \{\text{var}[u_{\beta*}(X, \theta_0)]\}^{-1} = \{\text{var}[\Pi(u_{*\beta}(X, \theta_0) | \Lambda_\beta)]\}^{-1} \end{aligned}$$

and (as confirmed by inspection of Figure 1), there is no general ordering between these two variances, because the component $u_{*\alpha}(x, \theta_0)$ influences the computation of the former but not the latter. However, if $u(x, \theta_0)$ and $u_*(x, \theta_0)$ are identical, so that $u_\beta(x, \theta_0)$ and $u_{*\beta}(x, \theta_0)$ are identical, then the right-angle triangle (4)

$$0 \rightarrow u_{\beta*}(x, \theta_0) \rightarrow u_{*\beta}(x, \theta_0)$$

demonstrates that

$$\begin{aligned} \|u_{*\beta}(x, \theta_0)\|^2 &= \|u_{\beta*}(x, \theta_0)\|^2 + \|u_{*\beta}(x, \theta_0) - u_{\beta*}(x, \theta_0)\|^2 \\ &\leq \|u_{\beta*}(x, \theta_0)\|^2 \end{aligned}$$

so that

$$\text{var}[u_{*\beta}(X, \theta_0)] \leq \text{var}[u_{\beta*}(X, \theta_0)]$$

that is

$$\text{avar}(\hat{\beta}) \geq \text{avar}(\tilde{\beta})$$

as indicated above.

12. **Theorem 1:** This theorem gives a sufficient condition for when we obtain the reverse result

$$\text{avar}(\tilde{\beta}) \leq \text{avar}(\hat{\beta}).$$

This reverse result holds if the components of the projected estimating equation are orthogonal

$$\mathbf{V}_{\beta\alpha} = \mathbb{E}[u_{*\beta}(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top] = \mathbf{0}. \quad (4)$$

Equality of the two variances holds if and only if

$$\mathbb{E}[u_\beta(X, \theta_0)s_\alpha(X, \theta_0, \kappa_0)^\top] = \mathbf{0}.$$

We have that under the assumption (4),

$$\text{avar}(\tilde{\beta}) - \text{avar}(\hat{\beta}) = \mathbf{D} \text{avar}(\hat{\alpha}) \mathbf{D}^\top$$

where

$$\mathbf{D} = \{\mathbb{E}[u_\beta(X, \theta_0)s_\beta(X, \theta_0, \kappa_0)^\top]\}^{-1} \mathbb{E}[u_\beta(X, \theta_0)s_\alpha(X, \theta_0, \kappa_0)^\top]$$

where this matrix is non-negative definite.

The proof goes like this:

- For $\tilde{\beta}$: from the analogous result to (2),

$$\begin{aligned} \text{avar}(\tilde{\beta}) &= \{\text{var}[u_{\beta*}(X, \theta_0)]\}^{-1} \\ &= \{\mathbb{E}[u_{\beta}(X, \theta_0)s_{\beta}(X, \theta_0, \kappa_0)^{\top}]\}^{-1}\text{var}[u_{\beta}(X, \theta_0)]\{\mathbb{E}[s_{\beta}(X, \theta_0, \kappa_0)u_{\beta}(X, \theta_0)^{\top}]\}^{-1} \end{aligned}$$

For $\hat{\beta}$: under the assumed orthogonality of $u_{*\beta}(x, \theta_0)$ and $u_{*\alpha}(x, \theta_0)$

$$\text{avar}(\hat{\beta}) = \{\text{var}[u_{*\beta}(X, \theta_0)]\}^{-1}$$

- By assumption, $u_{*\beta}(x, \theta_0)$ and $u_{*\alpha}(x, \theta_0)$ are orthogonal, and form a basis for the linear subspace Λ formed by taking matrix multiples of $u(x, \theta_0)$. Therefore, as $u_{\beta}(x, \theta_0) \in \Lambda$ also, we can express $u_{\beta}(x, \theta_0)$ in terms of the basis vectors as

$$\begin{aligned} u_{\beta}(x, \theta_0) &= \{\mathbb{E}[u_{\beta}(X, \theta_0)u_{*\beta}(X, \theta_0)^{\top}]\}\{\text{var}[u_{*\beta}(X, \theta_0)]\}^{-1}u_{*\beta}(x, \theta_0) \\ &\quad + \{\mathbb{E}[u_{\beta}(X, \theta_0)u_{*\alpha}(X, \theta_0)^{\top}]\}\{\text{var}[u_{*\alpha}(X, \theta_0)]\}^{-1}u_{*\alpha}(x, \theta_0) \end{aligned} \quad (5)$$

by standard linear algebra arguments. Rearranging this we have

$$\begin{aligned} u_{\beta}(x, \theta_0) - \{\mathbb{E}[u_{\beta}(X, \theta_0)u_{*\alpha}(X, \theta_0)^{\top}]\}\{\text{var}[u_{*\alpha}(X, \theta_0)]\}^{-1}u_{*\alpha}(x, \theta_0) \\ = \{\mathbb{E}[u_{\beta}(X, \theta_0)u_{*\beta}(X, \theta_0)^{\top}]\}\{\text{var}[u_{*\beta}(X, \theta_0)]\}^{-1}u_{*\beta}(x, \theta_0). \end{aligned}$$

- Denoting the left-hand side by $r(x, \theta_0)$, that is,

$$r(x, \theta_0) = u_{\beta}(x, \theta_0) - \{\mathbb{E}[u_{\beta}(X, \theta_0)u_{*\alpha}(X, \theta_0)^{\top}]\}\{\text{var}[u_{*\alpha}(X, \theta_0)]\}^{-1}u_{*\alpha}(x, \theta_0)$$

That is, $r(X, \theta_0)$ is the residual obtained after projecting $u_{\beta}(x, \theta_0)$ onto the linear subspace spanned by $u_{*\alpha}(x, \theta_0)$.

- Taking variances on both sides, we have

$$\text{var}[r(X, \theta_0)] = \{\mathbb{E}[u_{\beta}(X, \theta_0)u_{*\beta}(X, \theta_0)^{\top}]\}\{\text{var}[u_{*\beta}(X, \theta_0)]\}^{-1}\{\mathbb{E}[u_{*\beta}(X, \theta_0)u_{\beta}(X, \theta_0)^{\top}]\}$$

It also follows that

$$\begin{aligned} \mathbb{E}[u_{\beta}(X, \theta_0)u_{*\beta}(X, \theta_0)^{\top}] &= \mathbb{E}[u_{\beta}(X, \theta_0)(u_{*\beta}(X, \theta_0) - s_{\beta}(X, \theta_0, \kappa_0))^{\top}] + \mathbb{E}[u_{\beta}(X, \theta_0)s_{\beta}(X, \theta_0, \kappa_0)^{\top}] \\ &= \mathbb{E}[u_{\beta}(X, \theta_0)s_{\beta}(X, \theta_0, \kappa_0)^{\top}] \end{aligned}$$

as the first term is zero (see Figure 1), and so on rearrangement we have that

$$\{\text{var}[u_{*\beta}(X, \theta_0)]\}^{-1} = \{\mathbb{E}[u_{\beta}(X, \theta_0)s_{\beta}(X, \theta_0, \kappa_0)^{\top}]\}^{-1}\text{var}[r(X, \theta_0)]\{\mathbb{E}[s_{\beta}(X, \theta_0, \kappa_0)u_{\beta}(X, \theta_0)^{\top}]\}^{-1}$$

- Writing

$$\mathbf{C} = \{\mathbb{E}[u_{\beta}(X, \theta_0)s_{\beta}(X, \theta_0, \kappa_0)^{\top}]\}^{-1}$$

the difference $\text{avar}(\tilde{\beta}) - \text{avar}(\hat{\beta})$ is

$$\mathbf{C}(\text{var}[u_{\beta}(X, \theta_0)] - \text{var}[r(X, \theta_0)])\mathbf{C}^{\top}$$

Now, as $r(X, \theta_0)$ is the residual obtained after projecting $u_{\beta}(X, \theta_0)$ onto the linear subspace spanned by $u_{*\alpha}(X, \theta_0)$, it follows by Pythagoras's theorem that

$$\text{var}[u_{\beta}(X, \theta_0)] - \text{var}[r(X, \theta_0)] \geq 0$$

and as $\mathbf{C}$ is positive definite, we have that $\text{avar}(\tilde{\beta}) \geq \text{avar}(\hat{\beta})$.

- We have more precisely that

$$\begin{aligned} \text{var}[r(X, \theta_0)] &= \text{var}[u_\beta(X, \theta_0)] \\ &\quad - \{\mathbb{E}[u_\beta(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top]\}\{\text{var}[u_{*\alpha}(X, \theta_0)]\}^{-1}\{\mathbb{E}[u_{*\alpha}(X, \theta_0)u_\beta(X, \theta_0)^\top]\} \end{aligned}$$

so therefore $\text{var}[u_\beta(X, \theta_0)] - \text{var}[r(X, \theta_0)]$ equals

$$\{\mathbb{E}[u_\beta(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top]\}\{\text{var}[u_{*\alpha}(X, \theta_0)]\}^{-1}\{\mathbb{E}[u_{*\alpha}(X, \theta_0)u_\beta(X, \theta_0)^\top]\}$$

and $\text{avar}(\tilde{\beta}) - \text{avar}(\hat{\beta})$ simplifies to

$$\mathbf{D}\{\text{var}[u_{*\alpha}(X, \theta_0)]\}^{-1}\mathbf{D}^\top$$

where

$$\begin{aligned} \mathbf{D} &= \mathbf{C}\mathbb{E}[u_\beta(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top] \\ &= \{\mathbb{E}[u_\beta(X, \theta_0)s_\beta(X, \theta_0, \kappa_0)^\top]\}^{-1}\mathbb{E}[u_\beta(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top] \\ &= \{\mathbb{E}[u_\beta(X, \theta_0)s_\beta(X, \theta_0, \kappa_0)^\top]\}^{-1}\mathbb{E}[u_\beta(X, \theta_0)s_\alpha(X, \theta_0, \kappa_0)^\top] \end{aligned}$$

where the last line follows by (5): using Figure 1,

$$\begin{aligned} \mathbb{E}[u_\beta(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top] &= \mathbb{E}[u_\beta(X, \theta_0)(u_{*\alpha}(X, \theta_0) - s_\alpha(X, \theta_0, \kappa_0))^\top] + \mathbb{E}[u_\beta(X, \theta_0)s_\alpha(X, \theta_0, \kappa_0)^\top] \\ &= \mathbb{E}[u_\beta(X, \theta_0)s_\alpha(X, \theta_0, \kappa_0)^\top]. \end{aligned}$$

13. **Asymptotic independence:** The paradox in Theorem 1 holds if

$$\mathbb{E}[u_{*\beta}(X, \theta_0)u_{*\alpha}(X, \theta_0)^\top] = \mathbf{0}$$

that is, if $\hat{\beta}$ and $\hat{\alpha}$ are asymptotically independent.

14. **Asymptotic expansion:** We have by Taylor expansion the asymptotic equivalence

$$\sqrt{n}(\tilde{\beta} - \beta_0) = \sqrt{n}(\hat{\beta} - \beta_0) + \sqrt{n}\mathbf{D}(\hat{\alpha} - \alpha_0) + o_p(1).$$

and if the paradox holds the two terms on the right hand side are asymptotically independent.