

Geometric Aspects of Ordinary Least Squares

The Linear Model

We have

$$\mathbf{Y} = \mathbf{X}\theta + \varepsilon$$

where ε is an $(n \times 1)$ vector of random variables with

$$\mathbb{E}[\varepsilon|\mathbf{X}] = \mathbf{0}_n \quad \text{Var}[\varepsilon|\mathbf{X}] \equiv \mathbb{E}[\varepsilon\varepsilon^\top|\mathbf{X}] = \sigma^2\mathbf{I}_n.$$

We will treat $\mathbf{X}$ as random, and assume that this model is correctly specified.

The Linear Model (cont.)

Suppose we have two sets of predictors that yield matrices

$$\mathbf{X}_\psi : n \times q$$

$$\mathbf{X}_\beta : n \times r$$

Let $\mathbf{X} = [\mathbf{X}_\psi \ \mathbf{X}_\beta]$ be the combined $n \times p$ matrix, and suppose

$$\mathbf{Y} = \mathbf{X}\theta + \epsilon \equiv \mathbf{X}_\psi\psi + \mathbf{X}_\beta\beta + \epsilon$$

with θ , ψ and β are p , q and r -dimensional parameter vector and sub-vectors respectively.

The OLS estimating equations are

$$\begin{pmatrix} \mathbf{X}_{\psi}^{\top} \\ \mathbf{X}_{\beta}^{\top} \end{pmatrix} (\mathbf{Y} - \mathbf{X}_{\psi}\psi - \mathbf{X}_{\beta}\beta) = \mathbf{0}_p$$

These are also the score functions if we use the parametric model where

$$\epsilon \sim \text{Normal}_n(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$$

up to a factor of $1/\sigma^2$, which is not relevant for estimating θ .

Estimating equations (cont.)

That is, if Y and $\mathbf{x}$ ($1 \times p$) comprise a single datum Z ,

$$S_{\theta}(Z, \theta) = \begin{bmatrix} S_{\psi}(Z, \theta) \\ S_{\beta}(Z, \theta) \end{bmatrix} = \mathbf{x}^{\top} \epsilon$$

where

$$S_{\psi}(Z, \theta) = \mathbf{x}_{\psi}^{\top} (Y - \mathbf{x}_{\psi} \psi - \mathbf{x}_{\beta} \beta) = \mathbf{x}_{\psi}^{\top} \epsilon$$

$$S_{\beta}(Z, \theta) = \mathbf{x}_{\beta}^{\top} (Y - \mathbf{x}_{\psi} \psi - \mathbf{x}_{\beta} \beta) = \mathbf{x}_{\beta}^{\top} \epsilon$$

say.

The geometry of score functions

- The space spanned by S_θ is formed by taking linear combinations of the elements of $\mathbf{x}$;
- The space spanned by S_ψ is formed by taking linear combinations of the elements of $\mathbf{x}_\psi$;
 - S_ψ is a $q \times 1$ vector, with zero mean and finite variance, so is an element in the Hilbert space $\mathcal{H}_q$.
- The space spanned by S_β is formed by taking linear combinations of the elements of $\mathbf{x}_\beta$
 - This is the nuisance tangent space, Λ .

The geometry of score functions (cont.)

We consider projecting S_ψ onto the nuisance tangent space.

The projection matrix is computed from the general formula

$$\mathbf{B}_0 = \mathbb{E}[h v^\top] \{ \mathbb{E}[v v^\top] \}^{-1}$$

with h and v set equal to $S_\psi = \mathbf{x}_\psi^\top \epsilon$ and $S_\beta = \mathbf{x}_\beta^\top \epsilon$ respectively, that is

$$\mathbf{B}_0 = \mathbb{E}[S_\psi S_\beta^\top] \{ \mathbb{E}[S_\beta^\top S_\beta] \}^{-1} = \mathbb{E}[\epsilon^2 \mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\epsilon^2 \mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1}$$

The geometry of score functions (cont.)

Under the assumption that $\mathbb{E}[\varepsilon\varepsilon^\top|\mathbf{X}] = \sigma^2\mathbf{I}_n$, we can simplify this by iterated expectation to

$$\mathbf{B}_0 = \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1}$$

and hence

$$\begin{aligned}\Pi(\mathbf{S}_\psi|\Lambda) &\equiv \mathbf{B}_0 \mathbf{v} = \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbf{S}_\beta \\ &= \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbf{x}_\beta^\top \boldsymbol{\epsilon}.\end{aligned}$$

The geometry of score functions (cont.)

The ‘residual’ resulting from the projection is therefore

$$\begin{aligned} S_\psi - \Pi(S_\psi|\Lambda) &= \mathbf{x}_\psi^\top \epsilon - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbf{x}_\beta^\top \epsilon \\ &= (\mathbf{x}_\psi^\top - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbf{x}_\beta^\top) \epsilon \end{aligned}$$

The vector

$$S_\psi - \Pi(S_\psi|\Lambda)$$

lies in the space orthogonal to Λ , that is, for example

$$\mathbb{E}[(S_\psi - \Pi(S_\psi|\Lambda)) S_\beta^\top] = \mathbf{0}_{q \times r}.$$

The geometry of score functions (cont.)

To see this,

$$\begin{aligned}\mathbb{E}[(S_\psi - \Pi(S_\psi|\Lambda))S_\beta^\top] &= \mathbb{E}[(\mathbf{x}_\psi^\top - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{\mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta]\}^{-1} \mathbf{x}_\beta^\top) \epsilon \epsilon^\top \mathbf{x}_\beta] \\ &= \sigma^2 \left(\mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{\mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta]\}^{-1} \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \right) \\ &= \sigma^2 \left(\mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \right) \\ &= \mathbf{0}_{q \times r}.\end{aligned}$$

Thus

$$S_\psi = \Pi(S_\psi|\Lambda) + (S_\psi - \Pi(S_\psi|\Lambda))$$

with the two terms in the sum orthogonal.

The geometry of score functions (cont.)

It is also claimed that

$$S_\psi - \Pi(S_\psi|\Lambda) \equiv \Pi(S_\psi|\Lambda^\perp)$$

that is, the ‘residual’ is also the projection of S_ψ onto $\Lambda^\perp$, the space orthogonal to Λ . By the previous result

$$(S_\psi - \Pi(S_\psi|\Lambda)) \perp \Lambda$$

so therefore

$$\begin{aligned}\Lambda^\perp &= \{\mathbf{B}(S_\psi - \Pi(S_\psi|\Lambda)), \mathbf{B} \text{ a } q \times q \text{ matrix}\} \\ &\equiv \{\mathbf{B}v_\psi, \mathbf{B} \text{ a } q \times q \text{ matrix}\}\end{aligned}$$

say.

The geometry of score functions (cont.)

That is, $\mathbf{v}_\psi = (\mathbf{x}_\psi^\top - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{\mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta]\}^{-1} \mathbf{x}_\beta^\top) \epsilon$ and

$$\Lambda^\perp = \{\mathbf{B}(\mathbf{x}_\psi^\top - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{\mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta]\}^{-1} \mathbf{x}_\beta^\top) \epsilon, \mathbf{B} \text{ a } q \times q \text{ matrix}\}.$$

We now try to compute $\Pi(S_\psi | \Lambda^\perp)$. Using the projection definition, the projection matrix is

$$\mathbf{B}_0 = \mathbb{E}[S_\psi \mathbf{v}_\psi^\top] \{\mathbb{E}[\mathbf{v}_\psi \mathbf{v}_\psi^\top]\}^{-1}.$$

The geometry of score functions (cont.)

Now

$$\mathbf{S}_\psi \mathbf{v}_\psi^\top = \mathbf{x}_\psi^\top \epsilon \epsilon^\top (\mathbf{x}_\psi - \mathbf{x}_\beta \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\psi])$$

so taking expectations we have

$$\mathbb{E}[\mathbf{S}_\psi \mathbf{v}_\psi^\top] = \sigma^2 \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\psi - \mathbf{x}_\psi^\top \mathbf{x}_\beta \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\psi]]$$

The geometry of score functions (cont.)

Similarly

$$\begin{aligned}\mathbb{E}[\mathbf{v}_\psi \mathbf{v}_\psi^\top] &= \sigma^2 \mathbb{E}[(\mathbf{x}_\psi^\top - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{\mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta]\}^{-1} \mathbf{x}_\beta^\top)(\mathbf{x}_\psi - \mathbf{x}_\beta \{\mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta]\}^{-1} \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\psi])]) \\ &= \sigma^2 \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\psi - \mathbf{x}_\psi^\top \mathbf{x}_\beta \{\mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta]\}^{-1} \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\psi]]\end{aligned}$$

after cancelling terms, so therefore

$$\mathbf{B}_0 = \mathbf{I}_q.$$

Hence

$$\Pi(S_\psi | \Lambda^\perp) = \mathbf{B}_0 \mathbf{v}_\psi = (\mathbf{x}_\psi^\top - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{\mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta]\}^{-1} \mathbf{x}_\beta^\top) \epsilon = S_\psi - \Pi(S_\psi | \Lambda)$$

as required.

Deriving the influence function

Recall the Key Theorem: for a q -dimensional parameter of interest $\Psi(\theta)$, let

$$\Gamma(\theta) = \frac{\partial \Psi(\theta)}{\partial \theta^\top} = \begin{bmatrix} \frac{\partial \psi_1}{\partial \theta_1} & \frac{\partial \psi_1}{\partial \theta_2} & \cdots & \frac{\partial \psi_1}{\partial \theta_p} \\ \frac{\partial \psi_2}{\partial \theta_1} & \frac{\partial \psi_2}{\partial \theta_2} & \cdots & \frac{\partial \psi_2}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \psi_q}{\partial \theta_1} & \frac{\partial \psi_q}{\partial \theta_2} & \cdots & \frac{\partial \psi_q}{\partial \theta_p} \end{bmatrix}.$$

Deriving the influence function (cont.)

For example, $q = 2, p = 3$, if

$$\Psi = (\psi_1, \psi_2)^\top = (\theta_1\theta_2, \theta_2 + \theta_3^2)^\top,$$

then

$$\Gamma(\theta) = \begin{bmatrix} \theta_2 & \theta_1 & 0 \\ 0 & 1 & 2\theta_3 \end{bmatrix}$$

Suppose $\hat{\Psi}_n$ is a regular and asymptotically linear (RAL) estimator of Ψ with influence function $\varphi(\cdot)$, such that

$$\mathbb{E}[\{\varphi(Z)\}^\top \varphi(Z)] = \sum_{j=1}^q \mathbb{E}[\varphi_j^2(Z)] < \infty$$

then

$$\mathbb{E}[\varphi(Z) \{S_\theta(Z, \theta_0)\}^\top] = \Gamma(\theta_0).$$

Deriving the influence function (cont.)

If

$$\Psi(\theta) \equiv \psi$$

then

$$\Gamma(\theta_0) = \begin{bmatrix} \mathbf{I}_{q \times q} & \mathbf{0}_{q \times r} \\ \mathbf{0}_{r \times q} & \mathbf{0}_{r \times r} \end{bmatrix}$$

and so

$$(i) \quad \mathbb{E}[\varphi(Z) \{S_\psi(Z, \theta_0)\}^\top] = \mathbf{I}_{q \times q}$$

$$(ii) \quad \mathbb{E}[\varphi(Z) \{S_\beta(Z, \theta_0)\}^\top] = \mathbf{0}_{q \times r}.$$

That is, $\varphi(Z)$ is *orthogonal* to $S_\beta(Z, \theta_0)$.

It is claimed, therefore, that the influence function for estimating ψ takes the form

$$\mathbf{B}_\psi(\mathbf{x}_\psi^\top - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbf{x}_\beta^\top) \epsilon$$

for some matrix $\mathbf{B}_\psi$.

Deriving the influence function (cont.)

To deduce the influence function from first principles, we refer back to the OLS solution:

$$\hat{\theta}_n = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$$

hence

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta_0) &= \sqrt{n} \left((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} - \theta_0 \right) \\ &= \sqrt{n} \left((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\theta_0) \right) \\ &= \sqrt{n} (\mathbf{X}^\top \mathbf{X})^{-1} \sum_{i=1}^n \mathbf{x}_i^\top (Y_i - \mathbf{x}_i \theta_0) \\ &= \mathbb{E}[\mathbf{x}^\top \mathbf{x}]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i^\top (Y_i - \mathbf{x}_i \theta_0) + o_p(1) \end{aligned}$$

This suggests that the influence function for estimating θ is

$$\varphi(Z) = \mathbb{E}[\mathbf{x}^\top \mathbf{x}]^{-1} \mathbf{x}^\top (Y - \mathbf{x}\theta_0)$$

which is equal to

$$\varphi(Z) = \mathbb{E}[\mathbf{x}^\top \mathbf{x}]^{-1} \mathbf{x}^\top \epsilon.$$

Note that in this case

$$\mathbb{E}[\varphi(Z)S_{\theta}^{\top}] = \mathbb{E}[\mathbf{x}^{\top}\mathbf{x}]^{-1}\mathbb{E}[\mathbf{x}^{\top}\epsilon\epsilon\mathbf{x}] = \sigma^2\mathbf{I}_p$$

and so the Key Theorem for RAL estimators holds apart from the scaling term σ^2 .

This suggests that actually the scores should be scaled by a factor of $1/\sigma^2$; this factor would be present in the case of a Normal distribution assumption for ϵ .

The influence function for ψ

Let

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} = \mathbf{A}^{-1}$$

say, where

$$\mathbf{A} = \mathbb{E}[\mathbf{x}^\top \mathbf{x}] = \begin{bmatrix} \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\psi] & \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \\ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\psi] & \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \end{bmatrix} \equiv \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

say.

The influence function for ψ (cont.)

By block matrix inversion results, we have that

$$\begin{aligned} \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} &= \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \mathbf{I}_q & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I}_r \end{bmatrix} \\ &= \mathbf{D} \begin{bmatrix} \mathbf{I}_q & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I}_r \end{bmatrix} \end{aligned}$$

say.

The influence function for ψ (cont.)

Therefore

$$\begin{aligned}\varphi(Z) &= \mathbf{A}^{-1} \mathbf{x}^\top \epsilon \\ &= \mathbf{D} \begin{bmatrix} \mathbf{I}_q & -\mathbf{A}_{12} \mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{21} \mathbf{A}_{11}^{-1} & \mathbf{I}_r \end{bmatrix} \begin{bmatrix} \mathbf{x}_\psi^\top \epsilon \\ \mathbf{x}_\beta^\top \epsilon \end{bmatrix} \\ &= \mathbf{D} \begin{bmatrix} \mathbf{x}_\psi^\top \epsilon - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{x}_\beta^\top \epsilon \\ \mathbf{x}_\beta^\top \epsilon - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{x}_\psi^\top \epsilon \end{bmatrix}\end{aligned}$$

of which the first q components are

$$(\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})^{-1} \left(\mathbf{x}_\psi^\top \epsilon - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{x}_\beta^\top \epsilon \right)$$

The influence function for ψ (cont.)

As

$$\mathbf{A}_{12} = \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \quad \mathbf{A}_{22} = \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta]$$

this yields

$$\varphi(Z) = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} \left(\mathbf{x}_\psi^\top \epsilon - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbf{x}_\beta^\top \epsilon \right)$$

which matches the earlier claim, with the matrix factor

$$\mathbf{B}_\psi = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}.$$

The asymptotic variance

Given the influence function, we conclude that

$$\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{d} \text{Normal} \left(\mathbf{0}_q, \mathbb{E}[\varphi(Z)\{\varphi(Z)\}^\top] \right)$$

with

$$\varphi(Z) = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} \left(\mathbf{x}_\psi^\top \epsilon - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{x}_\beta^\top \epsilon \right).$$

Thus the asymptotic variance is

$$\mathbf{B}_\psi \mathbb{E}[(\mathbf{x}_\psi^\top - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{x}_\beta^\top)\epsilon\epsilon^\top(\mathbf{x}_\psi - \mathbf{x}_\beta\mathbf{A}_{22}^{-1}\mathbf{A}_{21})] \mathbf{B}_\psi^\top$$

where, in fact, $\mathbf{B}_\psi^\top = \mathbf{B}_\psi$.

The asymptotic variance (cont.)

Expanding the brackets, we have

$$\mathbf{x}_\psi^\top \mathbf{x}_\psi - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{x}_\beta^\top \mathbf{x}_\psi - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{x}_\beta^\top \mathbf{x}_\beta \mathbf{A}_{22}^{-1} \mathbf{A}_{21} + \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{x}_\beta^\top \mathbf{x}_\beta \mathbf{A}_{22}^{-1} \mathbf{A}_{21}$$

and, as

$$\mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\psi] = \mathbf{A}_{11} \quad \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\psi] = \mathbf{A}_{21} \quad \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] = \mathbf{A}_{22}$$

we can rewrite the overall expectation as

$$\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21} \equiv \mathbf{B}_\psi^{-1}.$$

Hence

$$\mathbb{E}[\varphi(Z)\{\varphi(Z)\}^\top] = \sigma^2 \mathbf{B}_\psi = \sigma^2 (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})^{-1}.$$

The asymptotic variance (cont.)

Note that this is precisely the *same* asymptotic variance as we would have deduced by carrying out joint estimation of parameter vector $\theta = (\psi, \beta)$, and then taking the block matrix from the asymptotic variance of $\hat{\theta}_n$, which is

$$\sigma^2 \mathbf{V} = \sigma^2 \mathbf{A}^{-1}.$$

The score and the influence function

Note that for estimating θ we have the score function

$$S_{\theta} = \mathbf{x}^{\top} (Y - \mathbf{x}\theta) = \mathbf{x}^{\top} \epsilon$$

and the influence function

$$\varphi(Z) = \{\mathbb{E}[\mathbf{x}^{\top} \mathbf{x}]\}^{-1} \mathbf{x}^{\top} \epsilon$$

so the two functions satisfy

$$\varphi(Z) = \{\mathbb{E}[\mathbf{x}^{\top} \mathbf{x}]\}^{-1} S_{\theta}.$$

The score and the influence function (cont.)

For estimating ψ we have the *projected* score

$$S_\psi - \Pi(S_\psi|\Lambda) = \left(\mathbf{x}_\psi^\top - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbf{x}_\beta^\top \right) \epsilon$$

and the influence function

$$\varphi(Z) = (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})^{-1} \left(\mathbf{x}_\psi^\top - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbf{x}_\beta^\top \right) \epsilon$$

so the two functions satisfy

$$\varphi(Z) = (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})^{-1} (S_\psi - \Pi(S_\psi|\Lambda)).$$

Linear models: a recap

We may write the linear model

$$(\mathbf{Y} - \mathbf{X}_\psi \psi) = \mathbf{X}_\beta \beta + \varepsilon$$

so for any ψ , we solve for β

$$\tilde{\beta} = (\mathbf{X}_\beta^\top \mathbf{X}_\beta)^{-1} \mathbf{X}_\beta^\top (\mathbf{y} - \mathbf{X}_\psi \psi).$$

We then have that

$$\mathbf{X}_\beta \tilde{\beta} = \mathbf{X}_\beta (\mathbf{X}_\beta^\top \mathbf{X}_\beta)^{-1} \mathbf{X}_\beta^\top (\mathbf{y} - \mathbf{X}_\psi \psi) = \mathbf{H}_1 (\mathbf{y} - \mathbf{X}_\psi \psi)$$

say, with

$$\mathbf{H}_1 = \mathbf{X}_\beta (\mathbf{X}_\beta^\top \mathbf{X}_\beta)^{-1} \mathbf{X}_\beta^\top$$

so that $\mathbf{H}_1 \mathbf{H}_1 = \mathbf{H}_1^\top \mathbf{H}_1 = \mathbf{H}_1$.

For any values of β and ψ

$$\{(\mathbf{I}_n - \mathbf{H}_1)(\mathbf{y} - \mathbf{X}_\psi\psi)\}^\top \{\mathbf{X}_\beta\beta\} = (\mathbf{y} - \mathbf{X}_\psi\psi)^\top (\mathbf{I}_n - \mathbf{H}_1)^\top \mathbf{X}_\beta\beta = 0$$

so that

$$(\mathbf{I}_n - \mathbf{H}_1)(\mathbf{y} - \mathbf{X}_\psi\psi)$$

is orthogonal to the space spanned by $\mathbf{X}_\beta\beta$.

For estimating ψ , we have the reduced form model

$$(\mathbf{Y} - \mathbf{X}_\psi \psi) = \mathbf{X}_\beta \tilde{\beta} + \varepsilon$$

or equivalently

$$(\mathbf{Y} - \mathbf{X}_\psi \psi) = \mathbf{H}_1(\mathbf{Y} - \mathbf{X}_\psi \psi) + \varepsilon$$

that is

$$(\mathbf{I}_n - \mathbf{H}_1)\mathbf{Y} = (\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_\psi \psi + \varepsilon.$$

Linear models: a recap (cont.)

Note that the estimating equation is

$$\mathbf{X}_\psi^\top (\mathbf{I}_n - \mathbf{H}_1)(\mathbf{y} - \mathbf{X}_\psi \psi) = \mathbf{0}_q$$

implying orthogonality between

$$\mathbf{X}_\psi \quad \text{and} \quad (\mathbf{I}_n - \mathbf{H}_1)(\mathbf{y} - \mathbf{X}_\psi \psi).$$

and for any β

$$(\mathbf{I}_n - \mathbf{H}_1)(\mathbf{y} - \mathbf{X}_\psi \psi) = (\mathbf{I}_n - \mathbf{H}_1)(\mathbf{y} - \mathbf{X}_\beta \beta - \mathbf{X}_\psi \psi)$$

as

$$(\mathbf{I}_n - \mathbf{H}_1)\mathbf{X}_\beta = (\mathbf{I}_n - \mathbf{H}_1)^\top \mathbf{X}_\beta = \mathbf{0}_q.$$

Thus we have for any β

$$\mathbf{X}_\psi^\top (\mathbf{I}_n - \mathbf{H}_1)(\mathbf{y} - \mathbf{X}_\beta \beta - \mathbf{X}_\psi \psi) = \mathbf{X}_\psi^\top (\mathbf{I}_n - \mathbf{H}_1) \mathbf{e} = \mathbf{0}_q$$

say, where $\mathbf{e} = (\mathbf{y} - \mathbf{X}_\beta \beta - \mathbf{X}_\psi \psi)$, so that

$$\begin{aligned} \mathbf{X}_\psi^\top (\mathbf{I}_n - \mathbf{H}_1) \mathbf{e} &= (\mathbf{X}_\psi^\top - \mathbf{X}_\psi^\top \mathbf{H}_1) \mathbf{e} \\ &= (\mathbf{X}_\psi^\top - \mathbf{X}_\psi^\top \mathbf{X}_\beta (\mathbf{X}_\beta^\top \mathbf{X}_\beta)^{-1} \mathbf{X}_\beta^\top) \mathbf{e} \end{aligned}$$

which is the sample-based version of the projected score

$$S_\psi - \Pi(S_\psi | \Lambda) = \left(\mathbf{x}_\psi^\top - \mathbb{E}[\mathbf{x}_\psi^\top \mathbf{x}_\beta] \{ \mathbb{E}[\mathbf{x}_\beta^\top \mathbf{x}_\beta] \}^{-1} \mathbf{x}_\beta^\top \right) \epsilon.$$