SEMIPARAMETRIC THEORY FOR THE LINEAR MODEL

For the $p \times 1$ system of estimating equations

$$\sum_{i=1}^{n} \mathbf{U}_i(\theta) = \mathbf{0}$$

with $\mathbb{E}[\mathbf{U}(\theta_0)] = \mathbf{0}$, we have that

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} Normal_p(\mathbf{0}, \mathbf{V})$$

where

$$\mathbf{V} = \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-\top}$$

with

$$\mathcal{I} = \mathbb{E}[\mathbf{U}(\theta_0)\mathbf{U}(\theta_0)^\top] \qquad \mathcal{J} = \mathbb{E}[\dot{\mathbf{U}}(\theta_0)]$$

both $(p \times p)$ matrices, and

$$\dot{\mathbf{U}}(\theta_0) = -\frac{\partial \mathbf{U}(\theta)}{\partial \theta^\top}\bigg|_{\theta=\theta_0}$$

We proceed by computing estimates, $\widehat{I}_n$ and $\widehat{J}_n$, of the two matrices based on the observed data,

$$\widehat{I}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{U}_i(\widehat{\theta}_n)\{\mathbf{U}_i(\widehat{\theta}_n)\}^\top \qquad \widehat{J}_n = -\frac{1}{n}\sum_{i=1}^{n} \dot{\mathbf{U}}_i(\widehat{\theta}_n)$$

and then computing

$$\widehat{\mathbf{V}} = \widehat{J}_n^{-1}\widehat{I}_n\widehat{J}_n^{-\top}$$

to estimate the asymptotic variance.

In the linear model, suppose that we have parameter $\theta = (\psi^\top, \beta^\top)^\top$, with parameter of interest ψ $(q \times 1)$ and nuisance parameter β $(r \times 1)$. The estimating function (for a single data point) is

$$\mathbf{U}(\theta) = \mathbf{x}^\top(Y - \mathbf{x}\theta) = \begin{bmatrix} \mathbf{x}_2^\top \\ \mathbf{x}_1^\top \end{bmatrix} (Y - \mathbf{x}_2\psi - \mathbf{x}_1\beta)$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are $1 \times r$ and $1 \times q$ row vectors, respectively, with $\mathbf{x} = (\mathbf{x}_2, \mathbf{x}_1)$ is $1 \times p$. We have

$$\dot{\mathbf{U}}(\theta) = \begin{bmatrix} \mathbf{x}_2^\top\mathbf{x}_2 & \mathbf{x}_2^\top\mathbf{x}_1 \\ \mathbf{x}_1^\top\mathbf{x}_2 & \mathbf{x}_1^\top\mathbf{x}_1 \end{bmatrix} = \mathbf{x}^\top\mathbf{x} \equiv \dot{\mathbf{U}}$$

is a $p \times p$ symmetric block matrix that does not depend on θ. We have that

$$\mathcal{I}(\theta_0) = \mathbb{E}[\mathbf{x}^\top(Y - \mathbf{x}\theta)^2\mathbf{x}] = \mathbb{E}[(Y - \mathbf{x}\theta)^2\mathbf{x}^\top\mathbf{x}]$$

as $(Y - \mathbf{x}\theta)$ is a scalar quantity, and

$$\mathcal{J} = \mathbb{E}[\mathbf{x}^\top\mathbf{x}].$$

The estimators would be

$$\widehat{I}_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i\widehat{\theta}_n)^2\mathbf{x}_i^\top\mathbf{x}_i \qquad \widehat{J}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^\top\mathbf{x}_i$$

with $\widehat{\theta}_n$ solving the OLS system.

1. Under the most common (correct specification) homoscedastic (ie constant variance) model, for iid samples from the joint distribution,

$$\mathbb{E}_{Y|X}[Y - \mathbf{x}\theta_0|\mathbf{x}] = 0 \qquad \mathbb{E}_{Y|X}[(Y - \mathbf{x}\theta_0)^2|\mathbf{x}] = \sigma^2$$

so that by iterated expectation

$$\mathcal{I}(\theta_0) = \sigma^2\mathbb{E}[\mathbf{x}^\top\mathbf{x}]$$

and hence

$$\mathbf{V} = \sigma^2\{\mathbb{E}[\mathbf{x}^\top \mathbf{x}]\}^{-1}$$

For the influence function for θ, we have from the theory of m-estimation that

$$\varphi(X,Y) = \{\mathcal{J}(\theta_0)\}^{-1}\mathbf{U}(\theta_0) = \{\mathbb{E}[\mathbf{x}^\top \mathbf{x}]\}^{-1}\mathbf{x}^\top(Y - \mathbf{x}\theta_0)$$

Also, therefore, we have that

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} Normal_p(0, \sigma^2\{\mathbb{E}[\mathbf{x}^\top \mathbf{x}]\}^{-1}).$$

2. Focussing on the parameter of interest alone, we seek now a suitable influence or estimating function for estimating ψ. As this is a q-dimensional parameter, we seek a q-dimensional function.

   The theory from lectures suggests that we can derive a suitable estimating equation by projecting the estimating function for ψ onto the nuisance tangent space and taking the residual. In this case, the nuisance tangent space is merely the linear space Λ spanned by $\mathbf{x}_1^\top$, that is

$$\Lambda = \{\mathbf{B}\mathbf{x}_1^\top \ : \ \mathbf{B} \text{ a } q \times r \text{ matrix}\}$$

   Write $\varepsilon = Y - \mathbf{x}_2\psi_0 - \mathbf{x}_1\theta_0$, and

$$\mathbf{U}(\theta_0) = \begin{bmatrix}\mathbf{U}_\psi(\theta_0)\\ \mathbf{U}_\beta(\theta_0)\end{bmatrix} = \begin{bmatrix}\mathbf{x}_2^\top\varepsilon\\ \mathbf{x}_1^\top\varepsilon\end{bmatrix}$$

   The projection of $\mathbf{U}_\psi(\theta_0)$ onto Λ is computed by finding $\mathbf{B}_0$ (a $q \times r$ matrix) such that

$$\mathbb{E}[(\mathbf{x}_2^\top\varepsilon - \mathbf{B}_0\mathbf{x}_1^\top\varepsilon)^\top \mathbf{B}\mathbf{x}_1^\top\varepsilon] = \mathbf{0}$$

   for all $\mathbf{B}$. As per the results on slide p389, we can write explicitly

$$\mathbf{B}_0 = \mathbb{E}[\varepsilon^2\mathbf{x}_2^\top\mathbf{x}_1]\{\mathbb{E}[\varepsilon^2\mathbf{x}_1^\top\mathbf{x}_1]\}^{-1}$$

   and hence the projection as

$$\mathbf{B}_0\mathbf{x}_1^\top\varepsilon = \mathbb{E}[\varepsilon^2\mathbf{x}_2^\top\mathbf{x}_1]\{\mathbb{E}[\varepsilon^2\mathbf{x}_1^\top\mathbf{x}_1]\}^{-1}\mathbf{x}_1^\top\varepsilon$$

   Therefore the residual is

$$\mathbf{x}_2^\top\varepsilon - \mathbb{E}[\varepsilon^2\mathbf{x}_2^\top\mathbf{x}_1]\{\mathbb{E}[\varepsilon^2\mathbf{x}_1^\top\mathbf{x}_1]\}^{-1}\mathbf{x}_1^\top\varepsilon.$$

3. Under the homoscedastic model, by iterated expectation, the residual becomes

$$\mathbf{x}_2^\top\varepsilon - \mathbb{E}[\mathbf{x}_2^\top\mathbf{x}_1]\{\mathbb{E}[\mathbf{x}_1^\top\mathbf{x}_1]\}^{-1}\mathbf{x}_1^\top\varepsilon.$$

   We can estimate the two expectations using sample averages

$$\widehat{\mathbb{E}}[\mathbf{x}_2^\top\mathbf{x}_1] = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_{i2}^\top\mathbf{x}_{i1} = \frac{1}{n}\mathbf{X}_2^\top\mathbf{X}_1$$

   and

$$\widehat{\mathbb{E}}[\mathbf{x}_1^\top\mathbf{x}_1] = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_{i1}^\top\mathbf{x}_{i1} = \frac{1}{n}\mathbf{X}_1^\top\mathbf{X}_1$$

   where

$$\mathbf{X}_1 = \begin{bmatrix}\mathbf{x}_{11}\\ \vdots\\ \mathbf{x}_{n1}\end{bmatrix} \qquad \mathbf{X}_2 = \begin{bmatrix}\mathbf{x}_{12}\\ \vdots\\ \mathbf{x}_{n2}\end{bmatrix}$$

   which are $n \times r$ and $n \times q$ matricex respectively. This inspires the estimating equation

$$\sum_{i=1}^n \left(\mathbf{x}_{i2}^\top\varepsilon_i - \mathbf{X}_2^\top\mathbf{X}_1(\mathbf{X}_1^\top\mathbf{X}_1)^{-1}\mathbf{x}_{i1}^\top\varepsilon_i\right) = \mathbf{0}$$

   or equivalently

$$\left(\mathbf{X}_2^\top - \mathbf{X}_2^\top\mathbf{X}_1(\mathbf{X}_1^\top\mathbf{X}_1)^{-1}\mathbf{X}_1^\top\right)\epsilon = \mathbf{0}$$

   where $\epsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$.

4. Defining

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

we see that the estimating equation is

$$\left(\mathbf{X}_2^\top - \mathbf{X}_2^\top \mathbf{H}_1\right)\epsilon = \mathbf{0}$$

or equivalently

$$\mathbf{X}_2^\top \left(\mathbf{I}_n - \mathbf{H}_1\right)(\mathbf{y} - \mathbf{X}_2\psi - \mathbf{X}_1\beta) = \mathbf{0}.$$

This is precisely the estimating equation arrived at by the two-step approach that solves for β for fixed ψ, and then plugs the solution back into the estimating equation to solve for ψ (lecture slides pp 368–378). As

$$\left(\mathbf{I}_n - \mathbf{H}_1\right)\mathbf{X}_1 = \mathbf{0} \qquad n \times r$$

the equation can also be written

$$\mathbf{X}_2^\top \left(\mathbf{I}_n - \mathbf{H}_1\right)(\mathbf{y} - \mathbf{X}_2\psi) = \mathbf{0}.$$

The solution is

$$\widehat{\psi} = \left(\mathbf{X}_2^\top \left(\mathbf{I}_n - \mathbf{H}_1\right)\mathbf{X}_2\right)^{-1}\mathbf{X}_2^\top \left(\mathbf{I}_n - \mathbf{H}_1\right)\mathbf{y}.$$

5. In this case, we get the **same** solution for ψ whether we solve the joint estimating equation or the projected estimating equation. Under the homoscedastic model, using the joint estimating equation, we have asymptotic variance

$$\mathbf{V} = \sigma^2\{\mathbb{E}[\mathbf{x}^\top \mathbf{x}]\}^{-1}$$

where

$$\mathbb{E}[\mathbf{x}^\top \mathbf{x}] = \mathbb{E}\begin{bmatrix}\mathbf{x}_2^\top \mathbf{x}_2 & \mathbf{x}_2^\top \mathbf{x}_1 \\ \mathbf{x}_1^\top \mathbf{x}_2 & \mathbf{x}_1^\top \mathbf{x}_1\end{bmatrix} = \begin{bmatrix}\mathbf{V}_{22} & \mathbf{V}_{21} \\ \mathbf{V}_{12} & \mathbf{V}_{11}\end{bmatrix}$$

say, so therefore by results for block matrices, the asymptotic variance for ψ is

$$\sigma^2\left(\mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{21}\right)^{-1}.$$

Direct from the projected estimating function,

$$\mathbf{U}_\psi^{\text{proj}}(\theta) = \mathbf{x}_2^\top\varepsilon - \mathbb{E}[\varepsilon^2\mathbf{x}_2^\top \mathbf{x}_1]\{\mathbb{E}[\varepsilon^2\mathbf{x}_1^\top \mathbf{x}_1]\}^{-1}\mathbf{x}_1^\top\varepsilon. = \left(\mathbf{x}_2^\top - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{x}_1^\top\right)\epsilon$$

we have that the asymptotic variance is

$$\sigma^2\left\{\mathbb{E}\left[\left(\mathbf{x}_2^\top - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{x}_1^\top\right)\left(\mathbf{x}_2^\top - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{x}_1^\top\right)^\top\right]\right\}^{-1}.$$

Now

$$\left(\mathbf{x}_2^\top - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{x}_1^\top\right)\left(\mathbf{x}_2^\top - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{x}_1^\top\right)^\top = \mathbf{x}_2^\top \mathbf{x}_2 - \mathbf{x}_2^\top \mathbf{x}_1\mathbf{V}_{11}^{-1}\mathbf{V}_{12} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{x}_1^\top \mathbf{x}_2 + \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{x}_1^\top \mathbf{x}_1\mathbf{V}_{11}^{-1}\mathbf{V}_{12}$$

and taking expectations yields

$$\mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} + \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{11}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} = \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}$$

and the two asymptotic variances match.

```
#Numerical verification
set.seed(1293)
library(mvnfast)
muX<-c(-1,2,1)
Sig<-0.9^abs(outer(1:3,1:3,'-'))
n<-1000
X<-rmvn(n,muX,Sig)
th<-c(1,1,1,1)
Xm<-cbind(1,X)
Y<- Xm %*% th + rnorm(n)*2

coef(summary(lm(Y~X)))
```

```
+               Estimate Std. Error  t value     Pr(>|t|)
+ (Intercept) 1.1650961  0.4581732 2.542917 1.114348e-02
+ X1          0.9968153  0.1503912 6.628150 5.559852e-11
+ X2          0.8739335  0.1997438 4.375272 1.340581e-05
+ X3          1.0751411  0.1469218 7.317776 5.184399e-13
```

```
X1<-Xm[,1:2]  #Take beta to be the first two elements of theta
X2<-Xm[,3:4]  #Take psi to be the last two elements of theta

#Joint estimating equation
th.hat<-solve(t(Xm) %*% Xm) %*% (t(Xm) %*% Y)
th.hat
```

```
+           [,1]
+ [1,] 1.1650961
+ [2,] 0.9968153
+ [3,] 0.8739335
+ [4,] 1.0751411
```

```
Y.hat0<- Xm %*% th.hat
sigsq.hat<-sum((Y-Y.hat0)^2)/(n-length(th.hat))
sigsq.hat
```

```
+ [1] 4.079885
```

```
#Variance
V<-sigsq.hat * solve((t(Xm) %*% Xm)/n)/n
V
```

```
+             [,1]          [,2]        [,3]          [,4]
+ [1,]  0.20992264  0.0629881266 -0.08040989  0.0172747704
+ [2,]  0.06298813  0.0226175095 -0.02011278 -0.0003756986
+ [3,] -0.08040989 -0.0201127803  0.03989759 -0.0194391372
+ [4,]  0.01727477 -0.0003756986 -0.01943914  0.0215860284
```

```
sqrt(diag(V)) #Matches the standard error column above
```

```
+ [1] 0.4581732 0.1503912 0.1997438 0.1469218
```

```
#Variance calc with I and J matrices
Ihat<-(t(Xm) %*% (Xm * matrix(Y-Y.hat0,nrow=n,ncol=4)^2))/n
Jhat<-(t(Xm) %*% Xm)/n
Vhat<-(solve(Jhat) %*% Ihat %*% solve(t(Jhat)))/n
Vhat
```

```
+             [,1]          [,2]        [,3]          [,4]
+ [1,]  0.19948463  0.0602972084 -0.07637099  0.0168832616
+ [2,]  0.06029721  0.0220906118 -0.01860799 -0.0009084687
+ [3,] -0.07637099 -0.0186079919  0.03918979 -0.0202802382
+ [4,]  0.01688326 -0.0009084687 -0.02028024  0.0229538245
```

```
sqrt(diag(Vhat))
```

```
+ [1] 0.4466370 0.1486291 0.1979641 0.1515052
```

```
########################################################################
#Projected estimating equation
I.n<-diag(rep(1,n))
H1<- X1 %*% solve(t(X1) %*% X1) %*% t(X1)
psi.hat<-solve(t(X2) %*% (I.n-H1) %*% X2) %*% (( t(X2)%*%(I.n-H1)) %*% Y)
psi.hat
```

```
+           [,1]
+ [1,] 0.8739335
+ [2,] 1.0751411
```

```r
#Variance estimation
V11<-(t(X1) %*% X1)/n
V12<-(t(X1) %*% X2)/n
V21<-t(V12)
V22<-(t(X2) %*% X2)/n

V<-sigsq.hat*solve(V22-(V21 %*% solve(V11) %*% V12))/n
sqrt(diag(V))

+ [1] 0.1997438 0.1469218

#Check one of the claimed orthogonalities
M<-(I.n-H1) %*% X1
apply(M,2,mean)

+ [1] -3.015206e-16  1.177847e-15

apply(M,2,var)

+ [1] 1.081672e-30 4.425538e-30
```