

# BIOS762 - Notes

Mingwei Fei

December 4, 2022

[Quasi-likelihood, Z-estimate, GEE] **Taylor series and sandwich theorem**

- (a) Quasi-likelihood: We don't know the distribution of  $y_i$ , only know the  $E(y_i), Var(y_i)$ , so we can get the quasi function of  $S_n(\mu_i, \sigma^2)$ , and then we need to know the link between  $\mu_i$  and  $\beta$ . Furthermore, we will use the sandwich estimate.

It usually is used in overdispersion, also two level hierarchical model and random effect models could be used in overdispersion as well.

Quasi-likelihood provides an important method for making statistical inference without making parametric assumption. Suppose independence, so we have covariance matrix only with diagonal element. We only know mean, but variance is unknown  $\sigma^2 = \phi^{-1}$ .  $Cov(Y) = \sigma^2 diag\{v_i(\mu)\}$ , in which  $v_i(\mu)$  is known, but  $\sigma^2$  unknown.

When estimating  $\beta$  using the quasi score function, we don't need to estimate  $\sigma^2$  as it dropped out of the equation. But we can estimate through the variance of Y. Quasi-likelihood can be applied to independent and dependent observations.

- (b) Z-estimate: give the estimating equation of  $\beta$  instead, here we introduce the random effect.  $H(\beta)$  We don't know any information about  $Var(Y)$ . To solve  $\beta$ , we always need score function, so we will need to know the score function from either way. Z-estimator is more generalized than quasi-likelihood estimator.

$$D_\mu(\theta)^T(Y - \mu) = \sum_{i=1}^n \frac{\partial \mu_i(x_i, \beta)}{\partial \beta} (y_i - \mu_i) \quad (1)$$

$$\mu_i = x_i^T \beta \quad (2)$$

- (c) Generalized Estimating Equations: the quasi-likelihood and Z-estimators and GEE could all be considered as they assume the score functions. But GEE could be specialized to handle longitudinal data. GEE is semi-parametric, as it does not need to estimate all the parameters, which

is widely used.

[Godambe efficiency] The quasi-likelihood estimator is Godambe efficient for  $\beta$ .

First, need to understand how the sandwich equation come from. The central limit theorem says that, the sample mean converge to normal distribution with true mean and variance of the true mean.

$$n^{1/2}h_n(\beta^*) \xrightarrow{w} N(0, V(\beta^*)), \quad h_n(\beta^*) = \frac{1}{n}g(X, \beta^*),$$

$$V(\beta^*) = Var_{\beta^*}\{g(x_i, \beta^*)\}$$

The above equation says that the mean of equation function  $h_n(\beta^*)$  with true  $\beta$  converge to normal distribution with mean 0 and variance of the equation function.

Then  $Var(g(X, \beta^*))$  could be calculated from the Taylor expansion.

$$0 = g(X, \hat{\beta}_n) = g(X, \beta^*) + \frac{\partial g(X, \beta^*)}{\partial \beta}(\beta - \beta^*), \quad g(X, \beta) = 0 \text{ for any } \beta$$

$$\hat{\beta}_n = \beta^* - \frac{\partial g(\beta^*)}{\partial \beta}^{-1} g(\beta^*)$$

by LLN,

$$h_n(\beta^*) = E_{\beta^*}[\partial g(X, \beta^*)]$$

$$\hat{\beta}_n = \beta^* - \frac{\partial g(\beta^*)}{\partial \beta}^{-1} g(\beta^*)$$

Then  $V(\beta^*)$  could be written by delta method,

$$V(\beta^*) = \left( \frac{\partial g(X, \beta^*)}{\partial \beta} \right)^{-1} Var(\beta^*) \left( \frac{\partial g(X, \beta^*)}{\partial \beta} \right)^{-1}$$

We need to pay attention that,  $UV^{-1}U^T$  in the above variance function could not be simplified to  $U$  as  $U \neq V$ .  $U = V$  only exists in MLE as the score function of  $\hat{\beta} = 0$ . In the theory of maximum likelihood we have  $V = U$  by the second Bartlett identity. Here  $U$  need not even be a symmetric matrix and even if it is, there is nothing to make  $V = U$  hold and, in general,  $U \neq V$ .

But for  $\left( \frac{\partial g(X, \hat{\beta})}{\partial \beta} \right)$ , we can use the sandwich equation. We do assume that

$U$  and  $V$  have the same dimensions, so  $U$  is square and  $h_n$  is a vector-to-vector function between vector spaces of the same dimension, and there are the same number of estimating equations as parameters in (2). This does not assure that solutions of the estimating equations exist, nor does it assure that solutions

are unique if they exist, but uniqueness is impossible without as many estimating equations as parameters. We also assume that  $U$  and  $V$  are both nonsingular.

By Slutsky theorem,

$$n^{1/2}(\hat{\beta}_n - \beta^*) \xrightarrow{w} \left(\frac{1}{n}\right)^{-1} \left(\frac{\partial g(X, \beta^*)}{\partial \beta}\right)^{-1} n^{-1/2} g(\beta^*)$$

The Slutsky's theorem says that, the product and sum of variables converge to the product and sum of the converged value distribution. If only talk about converge to value, then it is law of large number (weak or strong).

Here we are using the normal distribution of  $n^{-1/2}g(\beta) = \sqrt{n}h_n(\beta)$  and Slutsky, the variance could be written as  $U^{-1}VU^{-1T}$ .

The matrix  $U^{-1}VU^{-1T}$  is called *Godambe information matrix* because its specialization to the theory of maximum likelihood is  $U$  or  $V$  (because  $U = V$  in maximum likelihood), which is the Fisher information matrix, and Godambe initiated the theory of unbiased estimating equations.

We need add assumptions that

$$U_n \xrightarrow{w} U, \quad V_n \xrightarrow{w} V$$

Then by Slutsky's theorem,

$$[U_n^{-1}V_nU_n^{-1T}][n^{1/2}(\hat{\theta}_n - \theta^*)] \xrightarrow{w} Normal(0, I)$$

which we write sloppily as

$$\hat{\theta}_n \approx Normal(\theta^*, U_n^{-1}V_nU_n^{-1T})$$

Another name for the estimator  $U_n^{-1}V_nU_n^{-1T}$  of the asymptotic variance is sandwich estimator (think of  $U_n$  as slices of bread and  $V_n$  as ham).

We assume an estimating equation for  $\beta$

$$S_n(\mathbf{y}, \beta) = \mathbf{H}^T(\mathbf{y} - \mu(\beta)) = \sum_{i=1}^n \mathbf{h}_i(y_i - \mu_i(\beta))$$

Then by Taylor expansion at true  $\beta_*$

$$\begin{aligned} 0 &= S_n(\mathbf{y}, \tilde{\beta}) = S_n(\mathbf{y}, \beta_*) + \frac{\partial S_n(\mathbf{y}, \beta_*)}{\partial \beta}(\tilde{\beta} - \beta_*)[1 + o_p(n)] \\ \tilde{\beta} - \beta_* &= -\frac{\partial S_n(\mathbf{y}, \beta_*)}{\partial \beta}^{-1} S_n(\mathbf{y}, \beta_*)[1 + o_p(n)] \end{aligned}$$

By WLLN,

$$\frac{1}{n} \frac{\partial S_n(\mathbf{y}, \beta_*)}{\partial \beta} = \frac{1}{n} E_{\theta_*} \left[ \frac{\partial S_n(\mathbf{y}, \beta_*)}{\partial \beta} \right]$$

By Slutsky's theorem,

$$\begin{aligned}\sqrt{n}(\tilde{\beta} - \beta_*) &= - \left( \frac{1}{n} \frac{\partial S_n(\mathbf{y}, \beta_*)}{\partial \beta} \right)^{-1} \frac{1}{\sqrt{n}} S_n(\mathbf{y}, \beta_*) \\ \text{Cov}(\sqrt{n}\tilde{\beta}) &= - \left( \frac{1}{n} \frac{\partial S_n(\mathbf{y}, \beta_*)}{\partial \beta} \right)^{-1} \frac{1}{n} \text{Cov}(S_n(\mathbf{y}, \beta_*)) - \left( \frac{1}{n} \frac{\partial S_n(\mathbf{y}, \beta_*)}{\partial \beta} \right)^{-1} T\end{aligned}$$

We have

$$\begin{aligned}\frac{\partial S_n(\mathbf{y}, \beta_*)}{\partial \beta} &= \frac{\partial \sum_{i=1}^n \mathbf{h}_i(y_i - \mu_i(\beta))}{\partial \beta} = \sum_{i=1}^n \partial \mathbf{h}_i(y_i - \mu_i(\beta)) - \sum_{i=1}^n \mathbf{h}_i \frac{\partial \mu_i}{\partial \beta} = -\mathbf{H}\mathbf{D} \\ \text{Cov}(S_n(\mathbf{y}, \beta_*)) &= E[S_n(\mathbf{y}, \beta_*)^{\otimes 2}] = \frac{1}{n} \left[ \sigma^2 \sum_{i=1}^n \mathbf{h}_i V_i^T \mathbf{h}_i \right] = \sigma^2 \mathbf{H}\mathbf{V}\mathbf{H}^T \\ \text{Cov}(\tilde{\beta}) &= \frac{1}{n} \left( \frac{1}{n} \mathbf{H}\mathbf{D} \right)^{-1} \frac{1}{n} \mathbf{H}^T \mathbf{V} \mathbf{H} \frac{1}{n} \left( \frac{1}{n} \mathbf{H}\mathbf{D} \right)^{-1} T \\ &= \sigma^2 (\mathbf{H}^T \mathbf{D})^{-1} \mathbf{H}^T \mathbf{V} \mathbf{H} (\mathbf{D}^T \mathbf{H})^{-1}\end{aligned}$$

Compare the quasi-likelihood estimator  $\text{Cov}(\tilde{\beta})$  vs.  $\text{Cov}(\hat{\beta})$

$$\begin{aligned}\text{Cov}(\hat{\beta})^{-1} - \text{Cov}(\tilde{\beta})^{-1} &= \sigma^{-2} (\mathbf{D}^T \mathbf{V} \mathbf{D} - (\mathbf{D}^T \mathbf{H})(\mathbf{H}^T \mathbf{V} \mathbf{H})^{-1} (\mathbf{H}^T \mathbf{D})) \\ &= \sigma^{-2} \mathbf{D}^T (\mathbf{V} - \mathbf{H}(\mathbf{H}^T \mathbf{V} \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{D}\end{aligned}$$

$\text{Cov}(\hat{\beta})^{-1} - \text{Cov}(\tilde{\beta})^{-1} \geq 0$  is a non-negative matrix, so  $\text{Cov}(\hat{\beta}) - \text{Cov}(\tilde{\beta})$  is a non-positive matrix. We have  $\text{Var}(\hat{\beta}) < \text{var}(\tilde{\beta})$ , so it is Godambe efficient. [exponential family canonical link] Consider the general exponential family distribution

$$\mathcal{L}n(\theta, \phi) = \sum_{i=1}^n \phi [y_i \theta_i - b(\theta) - c(y_i)] - 0.5 \sum_{i=1}^n s(\phi, y_i) \quad (3)$$

$$\theta_i = x_i^T \beta \quad (4)$$

where

$$\begin{aligned}\mu_i &= \partial_\theta b(\theta) = E(y_i) \\ v_i &= \phi^{-1} \partial_\theta b^2(\theta) = \text{Var}(y_i)\end{aligned}$$

So we have

$$\begin{aligned}\dot{\mathcal{L}}n(\beta) &= \phi \sum_{i=1}^n e_i \dot{\theta}_i, \quad e_i = y_i - \mu_i \\ \ddot{\mathcal{L}}n(\beta) &= -\phi \sum_{i=1}^n v_i \dot{\theta}_i^{\otimes 2} + \phi \sum_{i=1}^n e_i \ddot{\theta}_i\end{aligned}$$

In the exponential family distribution, the parameter of interest (natural or canonical parameter) is  $\theta_i$ ,  $\phi$  is a scale parameter (known or seen as a nuisance) and others are known functions. The n-dimensional vectors of fixed input values for the p explanatory variables are denoted by  $x_1, \dots, x_p$ . We assume that the input vectors influence (3.1) only via a linear function, the linear predictor,

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad (5)$$

upon which  $\theta_i$  depends. As it can be shown that  $\mu, \theta$  dependency is established by connecting the linear predictor  $\eta$  and  $\theta$  **via the mean**. More specifically, the mean  $\mu$  is seen as an invertible and smooth function of the linear predictor, i.e.

$$\eta_i = g(\mu), \quad \mu = g^{-1}(\eta) \quad (6)$$

The function  $g(\cdot)$  is called the link function. If the function connects  $\mu, \eta$  and  $\theta$  such that  $\eta = \theta$ , then this link is called canonical.

(a) Use logistic distribution as example

$$\begin{aligned} \mathcal{L}n(\beta) &= \mathcal{L}np(\beta_2, \beta_3, \dots, \beta_p) \\ &= \sum_{i=1}^N \left[ \sum_{j=2}^I y_{ij} x_i^T \theta_j - n_i \log \left( 1 + \sum_{j=2}^N \exp(x_i^T \beta_j) \right) \right] \\ \partial_{\beta_j} \ln(\beta) &= \sum_{i=1}^N (y_{ij} - n_i \pi_{ij}) x_i \\ &= \sum_{i=1}^N \frac{y_{ij} - E[y_{ij}]}{\text{Var}(y_{ij})} \partial_{\beta_j} \mu_{ij} \\ \partial_{\beta_j}^2 \ln(\beta) &= \sum_{i=1}^N n_i \pi_{ij} (1 - \pi_{ij}) x_i^{\otimes 2} \\ \partial_{\beta_j, \beta_k}^2 \ln(\beta) &= \sum_{i=1}^N n_i \pi_{ij} \pi_{ik} x_i^{\otimes 2}, \quad k \neq j \end{aligned}$$

Note that

$$\begin{aligned}
\pi_{ij} &= \frac{\exp(x_i^T \beta_j)}{1 + \sum_{j=2}^N \exp(x_i^T \beta_j)} \\
E[y_{ij}] &= n_i \pi_{ij} \\
\text{Var}(y_{ij}) &= \phi^{-1} \partial_{\beta} b(\theta) = \partial_{\beta_j}^2 \left[ n_i \log \left( 1 + \sum_{j=2}^N \exp(x_i^T \beta_j) \right) \right]^2 \\
&= \frac{\exp(x_i^T \beta_j) n_i}{\left[ 1 + \sum_{j=2}^N \exp(x_i^T \beta_j) \right]^2} = n_i \pi_{ij} (1 - \pi_{ij}) \\
\mu_i &= \partial_{\theta} = b(\theta) = \partial_{\theta_j} \left[ n_i \log \left( 1 + \sum_{j=2}^N \exp(\theta_j) \right) \right] = \frac{n_i \exp(x_i^T \beta_j)}{1 + \sum_{j=2}^N \exp(x_i^T \beta_j)}
\end{aligned}$$

(b)

[Score test] Consider testing linear hypothesis

$$H_0 : R\xi = b_0$$

Let  $R = (I_r, 0)$ ,  $R\xi = \xi(1)$  and  $\xi^T = (\xi(1)^T, \xi(2)^T)^T$ , in which  $\xi(1), \xi(2)$  are, respectively,  $r \times 1$  and  $(q-r) \times 1$  subvectors of  $\xi$ . Under  $H_0$ , the constrained ML estimate of  $\xi$  is denoted by  $\tilde{\xi} = (b_0^T, \tilde{\xi}(2)^T)^T$ . We define  $\hat{\xi}(2)(\xi(1))$  as a function of  $\xi(1)$ , which maximizes  $\ln(\xi(1), \xi(2))$  for each  $\xi(1)$ . Thus  $\tilde{\xi}(2) = \hat{\xi}(2)(b_0)$ . Thus, the hypothesis test could be written as

$$H_0 : \xi(1) = b_0 \quad H_1 : \xi(1) \neq b_0$$

(a) Z-estimate Score test Let

$$h(\xi, U) = (h_1(\xi, U)^T, h_2(\xi, U)^T)^T$$

Under  $H_0$ ,  $\tilde{\xi} = (b_0, \tilde{\xi}_2^T)^T$ , solve  $S_{n,2}(\xi(2)) = 0$

$$S_n(\hat{\xi}) = \sum_0^n h(\hat{\xi}, U) = 0$$

Write in matrix

$$\dot{S}_n(\xi) = \partial_{\xi} S_n(\xi) = \begin{pmatrix} S_{n,11} & S_{n,12} \\ S_{n,21} & S_{n,22} \end{pmatrix}$$

And

$$\dot{S}_n(\xi)^{-1} = \begin{pmatrix} S_n^{11} & S_n^{12} \\ S_n^{21} & S_n^{22} \end{pmatrix}$$

The Score test statistics

$$\begin{aligned} SC_n &= S_n(\tilde{\xi})^T \text{Var}(S_n(\tilde{\xi}))^{-1} S_n(\tilde{\xi}) = S_n(\tilde{\xi})^T I_n(\tilde{\xi})^{-1} S_n(\tilde{\xi}), \quad \text{if generalized linear model} \\ SC_n &= S_{n,1}(\tilde{\xi})^T \Sigma^{11}(\tilde{\xi}) S_{n,1}(\tilde{\xi}), \end{aligned}$$

As the  $\hat{\xi}(2)$  are the MLE of  $\xi(2)$  given  $\xi(1) = b_0$ ,  $S(\hat{\xi}(2)) = 0$ . Only the  $S(\hat{\xi}(1)) \neq 0$ .

$$\Sigma^{11}(\tilde{\xi}) = \{(I_r, -\tilde{S}_{n,12}\tilde{S}_{n,22}^{-1})\left(\sum_{i=1}^n [h(\tilde{\xi}) - \bar{h}(\tilde{\xi})]^{\otimes 2}\right)(I_r, -\tilde{S}_{n,12}\tilde{S}_{n,22}^{-1})^T\}^{-1}$$

$(\sum_{i=1}^n [h(\tilde{\xi}) - \bar{h}(\tilde{\xi})]^{\otimes 2})$  is actually  $I_n(\xi)$ ,  $\text{Var}([h(\xi, U)])$ . While the  $E[h(\xi, U)] = 0$  in the above cases, but not here. So we need to get the sample mean of  $\bar{h}(\tilde{\xi}) = (1/n) \sum_{i=1}^n [h(\tilde{\xi})]$ .  
Using CLT, we have

$$\sqrt{n}^{-1} S_n(\xi(*)) \rightarrow N(0, \text{Var}[h(\tilde{\xi}, U)])_*$$

Need to pay attention that, use  $I_n(\xi)$  the observed fisher information.  
Show how the  $(I_r, -\tilde{S}_{n,12}\tilde{S}_{n,22}^{-1})$  come in play, using Taylor series

$$\begin{aligned} \tilde{\xi}(2) - \xi(2)_* &= -S_{n,22}^{-1}(S_*) S_{n,2}(\xi_*) + O_p(n^{-1}) \\ S_{n,1}(\tilde{\xi}) &= S_{n,1}(\xi_*) + S_{n,12}(\xi_*)(\tilde{\xi}(2) - \xi(2)_*) + O_p(n) \end{aligned}$$

The above equations include  $(\tilde{\xi}(2) - \xi(2)_*)$ . We could

$$S_{n,1}(\tilde{\xi}) = [I_r, -S_{n,12}S_{n,22}^{-1}] S_n(\xi_*) + O_p(1)$$

(b) Rao's Score test in GLM For a fixed  $\xi(1)$ , we define

$$(\xi(1), \tilde{\xi}(2)(\xi(1))) = \text{argmax}_{\xi(2)} \ln(\xi(1), \xi(2))$$

The quantity  $\ln(\xi(1), \tilde{\xi}(2)(\xi(1)))$  is called the profile likelihood.  
According to the partition  $(\xi(1)^T, \xi(2)^T)$  of  $\xi^T$ , we define

$$\begin{aligned} \partial_{\xi} \ln(\xi)^T &= (\dot{L}_1^T, \dot{L}_2^T), \\ \ddot{L}(\xi) &= \partial_{\xi}^2 \ln = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix}, \quad \ddot{L}(\xi)^{-1} = (\partial_{\xi}^2 \ln)^{-1} = \begin{pmatrix} L^{11} & L^{12} \\ L^{21} & L^{22} \end{pmatrix} \end{aligned}$$

Then Score test statistics

$$\begin{aligned} SC_n &= \partial_{\xi} \ln(\xi)^T \text{Var}(\partial_{\xi} \ln(\xi))^{-1} \partial_{\xi} \ln(\xi) \\ &= (\dot{L}_1^T, \dot{L}_2^T) E[\ddot{L}(\xi)]^{-1} \begin{pmatrix} \dot{L}_1 \\ \dot{L}_2 \end{pmatrix} \end{aligned}$$

We need to calculate  $\partial_\xi \ln(\tilde{\xi})$  and  $\partial_{\tilde{\xi}}^2 \ln(\tilde{\xi})$  (or  $E[\partial_\xi \ln(\tilde{\xi})^{\otimes 2}]$ ). It can be shown that

$$\partial_\xi \ln(\tilde{\xi}) = (\dot{L}_1(\tilde{\xi})^T, 0^T)^T$$

Thus, using equation we can show that

$$\begin{aligned} SC_n &= -(\dot{L}_1(\tilde{\xi})^T, 0^T) \begin{pmatrix} L^{11} & L^{12} \\ L^{21} & L^{22} \end{pmatrix} \begin{pmatrix} \dot{L}_1(\tilde{\xi}) \\ 0 \end{pmatrix} \\ &= -\dot{L}_1(\tilde{\xi})^T L^{11}(\tilde{\xi}) \dot{L}_1(\tilde{\xi}) \end{aligned}$$

- (c) Connection between the ML estimates of  $\xi$  under  $H_0$  and the estimates under the unrestricted space.

Using a Taylor's series expansion, we get expansion at  $\tilde{\xi}$

$$0 = \frac{\partial \ln \hat{\xi}}{\partial \xi} = \partial_\xi \ln(\tilde{\xi}) + \ddot{L}(\tilde{\xi})[\hat{\xi} - \tilde{\xi}][1 + O_p(1/\sqrt{n})]$$

Thus, because  $\dot{L}_2(\tilde{\xi}) = 0$ , we have

$$\begin{pmatrix} \dot{L}_1(\tilde{\xi}) \\ 0 \end{pmatrix} = - \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} \hat{\xi} - b_0 \\ \hat{\xi}(2) - \tilde{\xi}(2) \end{pmatrix}$$

Thus

$$\begin{aligned} \hat{\xi} - \tilde{\xi} &= -[\ddot{L}(\tilde{\xi})]^{-1} \dot{L}(\tilde{\xi}) + O_p(1/n) \\ L_{21}(\hat{\xi} - b_0) + L_{22}(\hat{\xi}(2) - \tilde{\xi}(2)) &= 0 \end{aligned}$$

So we can get the connection between  $\hat{\xi}_1$  and  $b_0$

$$\begin{aligned} \hat{\xi}(1) - b_0 &= -L^{11} \dot{L}_1(\tilde{\xi}) + O_p(1/n) \\ \hat{\xi}(2) - \tilde{\xi}(2)(b_0) &= -L_{22}^{-1} L_{21}(\hat{\xi}(1) - b_0) \end{aligned}$$

[MVN]

[Generalized least square projection]

$$V^{-1}(I - A) = (I - A)'V^{-1}(I - A)$$

This could be consider similar to the orthogonal projection  $A^2 = A$ , just we need to add Variance matrix there.

[Making inference of interested parameters]



(a) Transform distribution

Let  $\psi = \mu_1/\mu_2, \lambda = \mu_2$ , transform the distribution above with Jacobian matrix

$$\begin{bmatrix} \frac{\partial \mu_1}{\partial \psi} & \frac{\partial \mu_1}{\partial \lambda} \\ \frac{\partial \mu_2}{\partial \psi} & \frac{\partial \mu_2}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} \lambda & \psi \\ 0 & 1 \end{bmatrix}$$

The transformed distribution of  $\psi, \lambda$

$$P(\psi, \lambda) = |\lambda| \frac{(\psi\lambda)^{y_1} e^{-(\psi\lambda)}}{y_1!} \frac{\lambda^{y_2} e^{-\lambda}}{y_2!} = \frac{\lambda^{y_1+y_2+1} \psi^{y_1} e^{-\lambda(\psi+1)}}{y_1! y_2!}, \quad \lambda > 0$$

(b) reparameterize distribution, create interested parameters, nuisance parameters, link function

[Partitioned test, matrix, model]

**Expectation, Variance Calculation for conditional/joint/marginal distribution** Consider a pair of random variables  $(X, Y)$  and let  $\mu_x = E(X), \sigma_x^2 = Var(X), \mu_y = E(Y), \sigma_y^2 = Var(Y), \sigma_{xy} = Cov(X, Y)$ .

(a) Define  $(\alpha_0, \beta_0)$  as the minimizers of  $E[Y - (\alpha + \beta X)]^2$ , where  $\alpha$  and  $\beta$  are real numbers and the expectation is with respect to the joint distribution of  $(X, Y)$ . Derive explicit expression for  $(\alpha_0, \beta_0)$ .

$$\begin{aligned} Cov(X, Y) &= E(XY) - E(X)E(Y), & E(XY) &= \sigma_{xy} + \mu_x \mu_y \\ Var(Y) &= E(Y^2) - E(Y)^2, & E(Y^2) &= \sigma_y^2 + \mu_y^2 \\ Var(X) &= E(X^2) - E(X)^2, & E(X^2) &= \sigma_x^2 + \mu_x^2 \\ E[Y - (\alpha + \beta X)]^2 &= E[Y^2 + (\alpha + \beta X)^2 - 2(\alpha + \beta X)Y] \\ &= E(Y^2) + (\alpha^2 + \beta^2 E(X^2) + 2\alpha\beta E(X)) - 2(\alpha E(Y) + \beta E(XY)) \\ &= \sigma_y^2 + \mu_y^2 + \alpha^2 + \beta^2(\sigma_x^2 + \mu_x^2) + 2\alpha\beta\mu_x - 2\alpha\mu_y - 2\beta(\sigma_{xy} + \mu_x\mu_y) \\ \frac{\partial g(\alpha, \beta)}{\partial \alpha} &= 2\alpha + 2\beta\mu_x - 2\mu_y = 0 \\ \frac{\partial g(\alpha, \beta)}{\partial \beta} &= 2\beta(\sigma_x^2 + \mu_x^2) + 2\alpha\mu_x - 2(\sigma_{xy} + \mu_x\mu_y) = 0 \\ \alpha &= \mu_y - \beta\mu_x \\ \beta &= \frac{\sigma_{xy}}{\sigma_x^2} \end{aligned}$$

(b) Let  $U$  be a random variable which is independent of  $X$  and  $Y$  for which  $E(U) = 0$  and  $E(U^2) = \sigma_u^2$ . Suppose one is to observe, instead of the pair  $(X, Y)$ , the pair  $(W, Y)$ , where  $W = X + U$ . We may think of  $X$  as being observed with some measurement error  $U$ . Define  $\alpha_0^w$  and  $\beta_0^w$  as the minimizers of  $E([Y - (\alpha + \beta W)]^2)$ , where the expectation is with respect to

the joint distribution of  $(W, Y)$ . Let  $\lambda = \sigma_x^2/(\sigma_x^2 + \sigma_u^2)$ . Derive expression for  $\alpha_0^w$  and  $\beta_0^w$  which only involve  $\alpha_0, \beta_0, \lambda, \mu_x, \beta_0$ .

$$\begin{aligned}
E(WY) &= E((X + U)Y) = E(XY) + E(UY) = (\sigma_{xy} + \mu_x\mu_y), & \text{independence of } U, Y \\
Cov(W, Y) &= E(WY) - E(W)E(Y) = \sigma_{xy} + \mu_x\mu_y - \mu_x\mu_y = \sigma_{xy} \\
Var(Y) &= E(Y^2) - E(Y)^2, & E(Y^2) = \sigma_y^2 + \mu_y^2 \\
Var(W) &= Var(X + U) = Var(X) + Var(U) = \sigma_x^2 + \sigma_u^2, & E(X^2) = \sigma_x^2 + \mu_x^2 \\
E(W) &= E(X) = \mu_x \\
E(W^2) &= Var(W) + E(W)^2 = \sigma_x^2 + 2\mu_x^2 \\
E[Y - (\alpha + \beta W)]^2 &= E[Y^2 + (\alpha + \beta W)^2 - 2(\alpha + \beta W)Y] \\
&= E(Y^2) + (\alpha^2 + \beta^2 E(W^2) + 2\alpha\beta E(W)) - 2(\alpha E(Y) + \beta E(WY)) \\
&= \sigma_y^2 + \mu_y^2 + \alpha^2 + \beta^2(\sigma_x^2 + 2\mu_x^2) + 2\alpha\beta\mu_x - 2\alpha\mu_y - 2\beta(\sigma_{xy} + \mu_x\mu_y) \\
\frac{\partial g(\alpha, \beta)}{\partial \alpha} &= 2\alpha + 2\beta\mu_x - 2\mu_y = 0 \\
\frac{\partial g(\alpha, \beta)}{\partial \beta} &= 2\beta(\sigma_x^2 + 2\mu_x^2) + 2\alpha\mu_x - 2(\sigma_{xy} + \mu_x\mu_y) = 0 \\
\alpha &= \mu_y - \beta\mu_x \\
\beta^w &= \frac{Cov(W, Y)}{Var(W)} = \frac{\sigma_{xy}}{\sigma_x^2 + \sigma_u^2} = \lambda\beta_0
\end{aligned}$$

- (c) Suppose  $Y = \alpha_0 + \beta_0 X + \epsilon$ , where  $E(\epsilon) = 0$  and  $Var(\epsilon) = \sigma_\epsilon^2$ . Show that we may write  $Y = \alpha_0^w + \beta_0^w W + \eta$ , where  $E(\eta) = 0$  and  $Var(\eta) = \sigma_\eta^2$ , and find expressions for  $\sigma_\eta^2$  only in terms of  $\sigma_\epsilon^2, \lambda, \sigma_x^2, \beta_0$ .

$$\begin{aligned}
Y &= \alpha_0 + \beta_0 X + \epsilon \\
&= \alpha_0^w + \beta_0^w W + (\alpha_0 - \alpha_w - \beta_0^w W + \beta_0 X + \epsilon) \\
\eta &= \alpha_0 - \alpha_w - \beta_0^w W + \beta_0 X + \epsilon \\
\alpha_0 &= \mu_y - \beta\mu_x \\
\alpha_0^w &= \mu_y - \lambda\beta_0\mu_x, & X = W - U \\
\beta_0 &= \frac{\sigma_{xy}}{\sigma_x^2}, & \beta_0^w = \lambda\beta_0 \\
\eta &= \mu_y - \beta\mu_x - \mu_y + \lambda\beta_0\mu_x + \beta_0(W - U) - \lambda\beta_0 W + \epsilon \\
&= \beta_0\mu_x(\lambda - 1) + \beta_0 W(1 - \lambda) - \beta_0 U + \epsilon \\
E(\eta) &= \beta_0\mu_x(\lambda - 1) + \beta_0(1 - \lambda)\mu_x = 0 \\
\sigma_\eta^2 &= Var(\eta) = Var(\beta_0\mu_x(\lambda - 1) + \beta_0 W(1 - \lambda) - \beta_0 U + \epsilon) \\
&= \beta_0^2(1 - \lambda)^2 Var(W) - \beta_0^2 Var(U) - 2\beta_0^2(1 - \lambda)Cov(W, U) + Var(\epsilon) \\
&= \sigma_\epsilon^2 + \beta_0^2(1 - \lambda)^2(\sigma_x^2 + \sigma_u^2) + \beta_0^2\sigma_u^2 - 2\beta_0^2(1 - \lambda)Cov(X + U, U) \\
Cov(X + U, U) &= Cov(X, U) + Var(U) = \sigma_u^2 \\
\sigma_\eta^2 &= \sigma_\epsilon^2 + \lambda\beta_0^2\sigma_u^2
\end{aligned}$$

- (d) Consider a dataset with  $n$  independent realizations  $(W_1, Y_1), \dots, (W_n, Y_n)$  of  $(W, Y)$  and another dataset with  $n$  independent realizations  $(X_1, \tilde{Y}_1), \dots, (X_n, \tilde{Y}_n)$  of  $(X, Y)$ , where  $W = X + U$ . Let  $\hat{\alpha}^w, \hat{\beta}^w$  be the minimizers of  $\frac{1}{n} \sum_{i=1}^n [Y_i - (\alpha + \beta W_i)]^2$ , so is  $(\hat{\alpha}, \hat{\beta})$ .

- (i) For large  $n$ , give an expression for  $E(\hat{\beta}^w)$  and comment on whether  $\hat{\beta}^w$  is a good estimator of  $\beta_0$  in terms of bias and consistency. Similar as above, we have

$$\begin{aligned} g(\alpha, \beta) &= \frac{1}{n} \sum_{i=1}^n [Y_i - (\alpha + \beta W_i)]^2 \\ \hat{\alpha} &= \bar{Y} - \hat{\beta}^w \bar{X}, \quad \bar{Y} = \frac{1}{n} \sum Y_i, \bar{X} = \frac{1}{n} \sum X_i \\ \hat{\beta}^w &= \frac{\sum_{i=1}^n (W_i - \bar{W})(Y_i - \bar{Y})}{\sum_{i=1}^n (W_i - \bar{W})^2}, \quad \bar{W} = \frac{1}{n} \sum W_i \\ \hat{\beta}^w &= \frac{\hat{\sigma}_{wy}}{\hat{\sigma}_w^2} \xrightarrow{\infty} \frac{\sigma_{wy}}{\sigma_w^2(w)} = \lambda \beta_0 \end{aligned}$$

$\hat{\beta}^w$  is asymptotically bias for  $\beta_0$ , also it is not consistent.

- (ii) Suppose  $\hat{\beta}^w$  is used to test the hypothesis:  $H_0 : \beta_0 = 0$  vs.  $H_1 : \beta_0 \neq 0$ . Using the results in part (c), and for large  $n$ , derive an inequality only in terms of  $\sigma_u^2$ , as to when the t-stat for this hypothesis for the model in part (c) based on measurement error is smaller than the t-stat based on the model without measurement error.
- (iii) Examine the t-stat in part (ii) as a function of  $\sigma_u^2$ , and use the results in (ii) to examine the properties of this t-stat in terms of power and sample size as  $\sigma_u^2 \rightarrow 0$  and  $\sigma_u^2 \rightarrow \infty$  compared to a test stat based on  $\hat{\beta}$ .
- (e) Suppose  $(X, Y)^T \sim N_2(\mu, \Sigma)$ , where  $\mu = (\mu_x, \mu_y)'$  and  $\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$ .
- i) Derive the conditional distribution of  $Y|X$  and show that  $E(Y|X) = \alpha_0 + \beta_0 X$ .  
Construct  $Z = Y + AX$  so that  $Z$  is independent of  $X$ ,  $A = -\frac{\sigma_{xy}}{\sigma_x^2}$ . Then the conditional distribution  $Y|X$  will be only related to  $X$ .

**linear model** Suppose that  $Y$  is a  $4 \times 1$  vector with  $E(Y) = \mu, \mu \in C(E)$ , where  $E$  is the set  $E = \{\mu : \mu' = (\beta_1 + \beta_2 - \beta_3, \beta_2 + \beta_3, -\beta_2 - \beta_3, -\beta_1 - \beta_2 + \beta_3)\}$  where the  $\beta_i$  are real numbers,  $i = 1, 2, 3$ . Further assume that  $Cov(Y) = \sigma^2 I_{4 \times 4}$ , where  $\sigma^2$  is unknown.

- (a) Derive  $\hat{\mu}$ , the ordinary least squares estimate of  $\mu$ , by carrying out the appropriate projection.

$E(Y)$  is in the column space of  $C(E)$ , we need to find the o.p.o on  $C(E)$ . Also  $Cov(Y) = \sigma^2 I_{4 \times 4}$ , we can use ordinary least squares estimator as i.i.d.

$$\mu' = (\beta_1 + \beta_2 - \beta_3, \beta_2 + \beta_3, -\beta_2 - \beta_3, -\beta_1 - \beta_2 + \beta_3) = X\beta$$

$$\beta = (\beta_1, \beta_2, \beta_3)^T$$

$$X = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$$

$$C(X) = X_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & -1 \\ -1 & -1 \end{bmatrix}$$

$$M_\mu = X_1(X_1'X_1)^{-1}X_1^T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix}$$

$$\hat{\mu} = M_\mu Y = 1/2 \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} (y_1, y_2, y_3, y_4)^T = 1/2(y_1 - y_4, y_2 - y_3, y_3 - y_2, y_4 - y_1)^T$$

- (b) Find the BLUE of  $\beta_2 - \beta_3$  or show that it is nonestimable.

$$\lambda = (0, 1, -1)^T, \quad \lambda^T \beta = \beta_2 - \beta_3$$

$$\lambda^T = \rho^T X = (\rho_1, \rho_2, \rho_3, \rho_4) \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & -1 & 1 \end{bmatrix} = (0, 1, -1)$$

$$\rho_1 = \rho_4, \quad \rho_2 - \rho_3 = 1, \quad \rho_2 - \rho_3 = -1$$

The contradict of  $\rho_2 - \rho_3$  indicate that  $\beta_2 - \beta_3$  is not estimable.

- (c) Consider testing  $H_0 : \beta_2 + \beta_3 = 0$  versus  $H_1 : \beta_2 + \beta_3 \neq 0$ . Let  $E_0$  denote the set  $E$  assuming that  $H_0$  is true. Explicitly give the sets  $E_0$  and  $E \cap E_0^\perp$ .

Find the  $M_0, M_1$  are the o.p.o onto  $C(X)$  for  $H_0, H_1$ . **Not on  $C(Y)$** . Then the  $C(M_0), C(M_1)$  and the sets  $E_0$  and  $E \cap E_0^\perp$  relationship needs atten-

tion.

$$\begin{aligned}\lambda &= (0, 1, 1)^T, & \lambda^T \beta &= \beta_2 + \beta_3 \\ \lambda^T &= \rho^T X = (\rho_1, \rho_2, \rho_3, \rho_4) \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & -1 & 1 \end{bmatrix} = (0, 1, 1) \\ \rho &= (1, 2, 1, 1)^T\end{aligned}$$

$\beta_2 + \beta_3$  is estimable with one  $\rho = (1, 2, 1, 1)^T$ . Then we can have  $H_0 : \rho^T M Y = 0$  that  $\rho^T M \perp C(E_0)$ .

$$\begin{aligned}M_1 &= (M\rho)[(M\rho)^T(M\rho)]^{-1}(M\rho)^T \\ M\rho &= \rho_N = (0, 1, -1, 0)^T \\ M_1 &= \rho_N[(\rho_N)^T(\rho_N)]^{-1}(\rho_N)^T = 1/2 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ &= 1/2(0, 1, -1, 0)^T\end{aligned}$$

And the complement of  $M_0 = M - M_1$

$$\begin{aligned}M_0 &= M - M_1 = 1/2 \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \\ &= 1/2(1, 0, 0, -1)^T\end{aligned}$$

Also we can just look at  $C(Y)$  column space

$$E_0 = \text{span}\{(\beta_1 + 2\beta_2, 0, 0, -\beta_1 - 2\beta_2)^T\} = \text{span}\{(1, 0, 0, -1)^T\}$$

- (d) Assuming normality for  $Y$ , construct the simplest possible expression for the F statistic for the hypothesis  $H_0 : \mu \in E_0$  versus  $H_1 : \mu \notin E_0$ , where  $E_0$  is specified in part (c), and give the distribution of the F statistic under the null and alternative hypotheses.

$$\begin{aligned}M\rho &= \rho_N = (0, 1, -1, 0)^T \in M, & r(\rho_N) &= 1 \\ M_\rho &= \rho_N[(\rho_N)^T(\rho_N)]^{-1}(\rho_N)^T = 1/2 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ MSE &= \|(I - M)Y\| = 1/2(y_1 + y_4)^2 + 1/2(y_2 + y_4)^2 \\ F &= \frac{Y^T M_\rho Y / r(\rho)}{MSE} = \frac{2(y_2 - y_3)^2}{(y_1 + y_4)^2 + (y_2 + y_4)^2} \sim F(1, 2, \gamma), & r(M - M_\rho) &= 1, r(I - M) = 2\end{aligned}$$

In which, under  $H_0, \gamma = 0$ , and under  $H_1$ .

$$\begin{aligned}\gamma &= \frac{\|(M_1)X\beta\|}{2\|(I-M)Y\|/2} \\ &= \frac{(\beta_2 + \beta_3)^2}{\sigma^2}\end{aligned}$$

- (e) Assuming normality for  $Y$ , construct an exact 95% confidence interval for  $\beta_2 + \beta_3$ .

From part(d), we have

$$\begin{aligned}\lambda' &= (0, 1, 1) \\ \rho &= (1, 1, 0, 1)^T \\ \lambda^T \beta &= \rho^T MY = M_1 Y = 1/2(y_2 - y_3) \\ F &= \frac{\|\lambda^T \beta\|/r(\rho)}{\sigma^2} = \frac{\lambda' \beta [\lambda' [X'X]^{-1} \lambda]^{-1} (\lambda' \beta)^T}{\sigma^2} \sim F(1, 2, \gamma) \\ [\lambda' [X'X]^{-1} \lambda]^{-1} &= 2 \\ \{\beta : \frac{\lambda' \beta [\lambda' [X'X]^{-1} \lambda]^{-1} (\lambda' \beta)^T}{\sigma^2} &\leq F(0.95, 1, 2)\}\end{aligned}$$

Consider the linear model  $Y = X\beta + Z\gamma + \epsilon$ , where  $E(\epsilon) = 0$  and  $Cov(\epsilon) = V$ ,  $V$  is assumed known and positive definite, and  $(\beta, \gamma)$  are unknown. Further, let  $A = X(X'V^{-1}X)^{-1}X'V^{-1}$ ,  $X$  is  $n \times p$ ,  $Z$  is  $n \times q$ , and both  $X$  and  $Z$  may be less than full rank. Let  $C(H)$  denote the usual label for the column space of an arbitrary matrix  $H$ .

- (a) Show that  $(I - A)'V^{-1}(I - A) = (I - A)'V^{-1} = V^{-1}(I - A)$ .  
Let's review the o.p.o onto  $C(X)$ , and a projection  $A = X(X'V^{-1}X)^{-1}X'V^{-1}$ .

$$\begin{aligned}V &= QQ^T \\ P &= Q^{-1}X[(Q^{-1}X)'(Q^{-1}X)]^{-1}(Q^{-1}X)^T = Q^{-1}X[X'V^{-1}X]^{-1}X^TQ^{-1'}\end{aligned}$$

We need to transform  $P$  to  $A$ , so by definition,  $AX = X$ .

$$\begin{aligned}PQ^{-1}X &= Q^{-1}X \\ Q^{-1}X[(Q^{-1}X)'(Q^{-1}X)]^{-1}(Q^{-1}X)^TQ^{-1}X &= Q^{-1}X \\ Q^{-1}X[X'V^{-1}X]^{-1}X^TQ^{-1'}Q^{-1}X &= Q^{-1}X \\ Q^{-1}AX &= Q^{-1}X\end{aligned}$$

Then we have  $AX = X$ , then  $A$  is a projection onto  $C(X)$ . **Need attention that  $A$  is not a o.p.o. with respect to  $x'y$  (not symmetric) it is only a projection. But is o.p.o to  $x'V^{-1}y$ .**

For  $\rho'AY = \rho'MY$ , that  $M = X(X'X)^{-1}X^T$ , we need to have  $C(X) =$

$C(VX), C(X) = C(V^{-1}X)$ , then  $A = X(X'V^{-1}X)^{-1}X'V^{-1}$ .  
Show that  $A, I-A$  are projections

$$\begin{aligned} A^2 &= A \\ X(X'V^{-1}X)^{-1}X'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} &= X(X'V^{-1}X)^{-1}X'V^{-1} \\ (I-A)^2 &= I-A \\ (I-A)(I-A) &= I-A-A+A^2 = I-A \end{aligned}$$

Then transform the equation

$$\begin{aligned} (I-A)'V^{-1}(I-A)(I-A) &= (I-A)'V^{-1}(I-A), \\ (I-A)'V^{-1}(I-A) &= (I-A)'V^{-1} \end{aligned}$$

To show  $(I-A)'V^{-1}(I-A) = V^{-1}(I-A)$

$$\begin{aligned} (I-A)'(I-A)' &= [(I-A)(I-A)]^T = (I-A)^T \\ (I-A)'(I-A)'V^{-1}(I-A) &= (I-A)'V^{-1}(I-A) \\ (I-A)'V^{-1}(I-A) &= V^{-1}(I-A) \end{aligned}$$

- (b) Show that  $A$  is the projection operator onto  $C(X)$  along  $C(V^{-1}X)^\perp$ . We need to show  $AX = X, x \in C(X)$ , and  $Aw = 0, w \in C(V^{-1}X)^\perp, C(V^{-1}X) = C(X)$ . We have shown that  $AX = X$  in above (a), then let  $w \in C(V^{-1}X)^\perp$

$$\begin{aligned} (V^{-1}X)'w &= 0, & X'V^{-1}w &= 0 \\ Aw &= w, & X'V^{-1}Aw &= 0 \\ X'V^{-1}A &= 0, & A &\perp C(V^{-1}X) \end{aligned}$$

- (c) Let  $B$  denote the projection operator onto  $C(X, Z)$  along  $C(V^{-1}(X, Z))^\perp$ . Assume that all matrix inverses exist. Show that

$$B = A + (I-A)Z[Z'(I-A)'V^{-1}(I-A)Z]^{-1}Z'(I-A)'V^{-1}$$

To show  $B$  is a projection,  $B^2 = B$ , but if need to show projection onto  $C(X, Z)$ , then show  $BX = X, x \in C(X, Z)$ . If need to show projection along, then show  $Bw = 0, w \in C(V^{-1}(X, Z))^\perp$ .

$$\begin{aligned} B^2 &= (A + (I-A)Z[Z'(I-A)'V^{-1}(I-A)Z]^{-1}Z'(I-A)'V^{-1}) (A + (I-A)Z[Z'(I-A)'V^{-1}(I-A)Z]^{-1}Z'(I-A)'V^{-1}) \\ &= (A + (I-A)Z[Z'(I-A)'V^{-1}(I-A)Z]^{-1}Z'(I-A)'V^{-1}) = B \\ A(I-A) &= 0 \end{aligned}$$

Show  $B(X, Z) = (BX, BZ) = (X, Z)$ .

$$\begin{aligned} BX &= (A + (I-A)Z[Z'(I-A)'V^{-1}(I-A)Z]^{-1}Z'(I-A)'V^{-1}) X \\ &= AX + (I-A)Z[Z'(I-A)'V^{-1}(I-A)Z]^{-1}Z'(I-A)'V^{-1}(I-A)X = X, \quad \text{part (a)} \\ BZ &= (A + (I-A)Z[Z'(I-A)'V^{-1}(I-A)Z]^{-1}Z'(I-A)'V^{-1}) Z \\ &= AZ + (I-A)Z[Z'(I-A)'V^{-1}(I-A)Z]^{-1}Z'(I-A)'V^{-1}(I-A)Z = AZ + (I-A)Z = Z, \quad \text{part (b)} \end{aligned}$$

Next show projection along  $C(V^{-1}(X, Z))^{\perp}$ .

$$\begin{aligned}
(V^{-1}(X, Z))'w &= 0, & X'V^{-1}w &= 0, & Z'V^{-1}w &= 0 \\
Bw &= (A + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1})w = 0 \\
&= Aw + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}w = 0 \\
(X, Z)'V^{-1}B &= 0, & B &\perp C(V^{-1}(X, Z))
\end{aligned}$$

- (d) Show that  $(\hat{\gamma}, \hat{\beta})$  are generalized BLUE's for the linear model, where  $(\hat{\gamma}, \hat{\beta})$  satisfy

$$\begin{aligned}
\hat{\gamma} &= [Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}(I - A)Y \\
X\hat{\beta} &= A(Y - Z\hat{\gamma})
\end{aligned}$$

To show projection operator onto  $C(Z)$  is  $[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}$ , also it is along  $C(V^{-1}Z)^{\perp}$ .

By Gauss-Markov theorem, the least square estimates of  $\gamma, \beta$  are BLUE's.

$$\begin{aligned}
Y &= X\beta + Z\gamma \\
AY &= AX\beta + AZ\gamma = X\beta + AZ\gamma \\
(I - A)Y &= (I - A)Z\gamma
\end{aligned}$$

We can construct projection operation regarding  $C(Z)$ .

$$\begin{aligned}
Q^{-1}(I - A)Y &= Q^{-1}(I - A)Z\gamma \\
P_z &= Q^{-1}(I - A)Z[(Q^{-1}(I - A)Z)'(Q^{-1}(I - A)Z)]^{-1}(Q^{-1}(I - A)Z)' \\
P_z Q^{-1}(I - A)Z &= Q^{-1}(I - A)Z
\end{aligned}$$

$$[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}Z = Z, \quad \text{also use part (a)}$$

$$C = [Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}, \quad CZ = Z$$

We have  $C$  a projection operator onto  $C(Z)$ , also we can prove orthogonality to  $C(V^{-1}Z)^{\perp}$ . Then we have proved.

- (e) Suppose that  $\epsilon \sim N_n(0, V)$  and  $V$  is known. Further, suppose that  $\gamma, \beta$  are both estimable. From first principles, derive the likelihood ratio test for the hypothesis  $H_0 : \gamma = 0$ , where  $\gamma, \beta$  are both unknown, and state the exact distribution of the test statistic under the null and alternative hypotheses.

To derive the likelihood ratio test, we link with F-test and construct independent chi-square statistics. The nominator is the square difference



between two models, and the denominator is MSE.

$$\begin{aligned} Y &= X\beta + Z\gamma + \epsilon, & Y &= W\delta + \epsilon, \\ W &= (X, Z), \delta = (\beta, \gamma)^T, & V &= QQ^T, X_{n \times p}, Z_{n \times q}, \end{aligned}$$

Assume  $\text{rank}(W) = p+q$ , **Pay attention to define matrix rank**

$$\begin{aligned} Q^{-1}Y &= W = Q^{-1}X\beta + Q^{-1}Z\gamma \sim N(0, I) \\ Q^{-1}Y &= \tilde{Y}, Q^{-1}Z = \tilde{Z}, & Q^{-1}X &= \tilde{X} \\ \tilde{Y} &= \tilde{W}\delta + \tilde{\epsilon}, & \tilde{\epsilon} &\sim N(0, I) \end{aligned}$$

The likelihood ratio test

$$LRT = \frac{\text{Sup}_{H_0} L(\delta|W)}{\text{Sup}_{H_1} L(\delta|W)}$$

The likelihood under  $H_0$

$$\begin{aligned} \delta_0 &= (\beta_0, 0)', & W_0 &= (\tilde{X}) \\ L(\beta_0|W) &= \frac{1}{\sqrt{2\pi}} \exp \left( \frac{-1}{2} (\tilde{Y} - W_0\delta_0)' (\tilde{Y} - W_0\delta_0) \right) \\ &= \frac{1}{\sqrt{2\pi}} \exp \left( \frac{-1}{2} (\tilde{Y} - \tilde{X}\hat{\beta})' (\tilde{Y} - \tilde{X}\hat{\beta}) \right) \\ \tilde{X}\hat{\beta} &= M_0W_0, & \hat{\beta} &\text{ satisfy} \\ M_0 &= \tilde{X}[(\tilde{X})'(\tilde{X})]^{-1}\tilde{X}' \end{aligned}$$

Thus the numerator is

$$L(\beta_0|W) = \frac{1}{\sqrt{2\pi}} \exp \left( \frac{-1}{2} (\tilde{Y}'(I - M_0)\tilde{Y}) \right)$$

The likelihood under  $H_1$

$$\begin{aligned} L(\delta|W) &= \frac{1}{\sqrt{2\pi}} \exp \left( \frac{-1}{2} (\tilde{Y} - W\delta)' (\tilde{Y} - W\delta) \right) \\ \tilde{W}\hat{\delta} &= M\tilde{W} & \hat{\delta} &\text{ satisfy} \\ M &= \tilde{W}[(\tilde{W})'(\tilde{W})]^{-1}\tilde{W}' \end{aligned}$$

Thus the denominator is

$$L(\delta|W) = \frac{1}{\sqrt{2\pi}} \exp \left( \frac{-1}{2} (\tilde{Y}'(I - M)\tilde{Y}) \right)$$

The LRT test

$$-2\Lambda = \frac{(\tilde{Y}'(I - M_0)\tilde{Y})}{(\tilde{Y}'(I - M)\tilde{Y})} = (\tilde{Y}'(M - M_0)\tilde{Y})$$

Thus we reject the  $H_0$ , if  $\left(\tilde{Y}'(M - M_0)\tilde{Y}\right) > c$ .  $M - M_0$  is an orthogonal projection as  $(M - M_0)(M - M_0) = (M - M_0)$  and  $M - M_0$  is symmetric. So  $r(M - M_0) = q$  when  $X$  and  $Z$  are full rank.

$$\begin{aligned}\left(\tilde{Y}'(M - M_0)\tilde{Y}\right) &\sim (H_0)\chi^2(q) \\ \left(\tilde{Y}'(M - M_0)\tilde{Y}\right) &\sim (H_1)\chi^2(q, \theta) \\ \theta &= \frac{\|(M - M_0)\tilde{W}\delta\|}{2}\end{aligned}$$

Thus we reject  $H_0$  at level  $\alpha$  if  $\left(\tilde{Y}'(M - M_0)\tilde{Y}\right) > \chi_\alpha^2(q)$

- (f) Suppose that  $\epsilon \sim N_n(0, \sigma^2 R)$ , where  $R$  is known and positive definite, and  $\beta, \gamma, \sigma^2$  are all unknown. Further, assume that  $\beta, \gamma$  are both estimable. Derive an exact joint 95% confidence region for  $\beta, \gamma, \sigma^2$ . To write the distribution of  $\beta, \gamma, \sigma^2$ , such as F distribution or Chi-square (if  $\sigma^2$  is known) distribution  $M_\beta/\sigma^2, M_\gamma/\sigma^2$ .

$$\begin{aligned}Y &= X\beta + Z\gamma + \epsilon, & Y &= W\delta + \epsilon, \\ W &= (X, Z), \delta = (\beta, \gamma)^T, & R &= QQ^T, X_{n \times p}, Z_{n \times q},\end{aligned}$$

Assume  $\text{rank}(W) = p+q$ , **Pay attention to define matrix rank**

$$\begin{aligned}Q^{-1}Y &= W = Q^{-1}X\beta + Q^{-1}Z\gamma \sim N(0, \sigma^2 I) \\ Q^{-1}Y &= \tilde{Y}, Q^{-1}Z = \tilde{Z}, & Q^{-1}X &= \tilde{X} \\ \tilde{Y} &= \tilde{W}\delta + \tilde{\epsilon}, & \tilde{\epsilon} &\sim N(0, \sigma^2 I)\end{aligned}$$

We have the F test for  $\beta$

$$\begin{aligned}\lambda' &= (1, 0, 0), & \rho'\tilde{W} &= \lambda', \\ F_\beta &= \frac{(\tilde{Y} - \tilde{W}b)'M_{MP}(\tilde{Y} - \tilde{W}b)/r(M_{MP})}{\sigma^2} \\ &= \frac{(\tilde{Y} - \tilde{W}b)'(M\rho)[\rho'M\rho]^{-1}(M\rho)'(\tilde{Y} - \tilde{W}b)/r(M_{MP})}{\sigma^2} \\ &= \frac{(\rho'M\tilde{Y} - \rho'M\tilde{W}b)'[\rho'M\rho]^{-1}(\rho'M\tilde{Y} - \rho'M\tilde{W}b)/r(M_{MP})}{\sigma^2} \\ &= \frac{(\lambda'\hat{\delta} - \lambda'\delta)'[\rho'M\rho]^{-1}(\lambda'\hat{\delta} - \lambda'\delta)/r(M_{MP})}{\sigma^2} \\ &= \frac{(\hat{\beta} - \beta)'[\rho'M\rho]^{-1}(\hat{\beta} - \beta)/r(M_{MP})}{\sigma^2}, & r(M_{MP}) &= p\end{aligned}$$

We have the chi-square distribution for  $\sigma^2$

$$\frac{Y'(I - M)Y}{\sigma^2} \chi^2(n - p - q)$$

The degrees of freedom is  $n - p - q$  since both  $(\beta, \gamma)$  are estimable.  
 Furthermore, we have the chi-square distribution for  $\delta$ , because  $M \perp (I - M)$

$$\begin{aligned} \frac{\|MY - \tilde{W}\delta\|}{\sigma^2} &\sim \chi^2(p + q) \\ P\{\chi_{a/2}^2(p + q) \leq \frac{\|MY - \tilde{W}\delta\|}{\sigma^2} \leq \chi_{1-a/2}^2(p + q), \quad \chi_{b/2}^2(n - p - q) \leq \frac{\|(I - M)Y\|}{\sigma^2} \leq \chi_{1-b/2}^2(n - p - q)\} \\ &= 1 - \alpha \end{aligned}$$

Such that  $(1 - a)(1 - b) = 1 - \alpha$ .  
 The region is given

$$\begin{aligned} \{(\delta, \sigma^2) : \chi_{a/2}^2(p + q) \leq \frac{\|MY - \tilde{W}\delta\|}{\sigma^2} \leq \chi_{1-a/2}^2(p + q), \quad \chi_{b/2}^2(n - p - q) \leq \frac{\|(I - M)Y\|}{\sigma^2} \leq \chi_{1-b/2}^2(n - p - q)\} \\ = 1 - \alpha \end{aligned}$$

\* ... or skip a few, using the `\setcounter{exercise}{x}` command.  
 For more information, refer to <https://github.com/gijs-pennings/latex-homework>.