Imperial College
London

IMPERIAL COLLEGE LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2012

This paper is also taken for the relevant examination for the Associateship.

M2S2

Statistical Modelling

Date: Friday, 11 May 2012 Time: 2 pm – 4 pm

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used.

Statistical tables will not be available.

1. (i) Suppose $\theta \in \mathbb{R}$ is the unknown parameter in a statistical model and suppose that T is an estimator for θ.

   (a) Define the bias and the mean squared error of T.

   (b) How are bias, variance and mean squared error related? Prove this relationship.

   (ii) State the definition of the non-central F-distribution.

   (iii) Can there be a two-variate random vector $\boldsymbol{Y}$ with $\mathrm{cov}(\boldsymbol{Y}) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$?
   Justify your answer.

   (iv) State in one or two sentences the main difference between the Bayesian approach to statistics and the classical approach to statistics.

   (v) In a Bayesian framework, suppose you observe $Y_1, \ldots, Y_n$. Conditionally on τ they are independent and follow a $N(\mu, \frac{1}{\tau})$ distribution. Suppose that $\mu \in \mathbb{R}$ is a known constant and that the a-priori distribution of τ is Gamma(2,2). Identify the posterior distribution of τ given $Y_1, \ldots, Y_n$.
   [Recall that for $\alpha > 0$ and $\beta > 0$ the probability density function (pdf) of a Gamma(α, β)-distributed random variable is $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$.]

2. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables following a Pareto distribution with parameter $\alpha > 0$, i.e. for $i = 1, \ldots, n$,

$$\mathrm{P}(X_i \leq x) = 1 - x^{-\alpha}, \quad x > 1,$$

   and $\mathrm{P}(X_i \leq x) = 0$ for $x \leq 1$.

   (i) Write down the likelihood function.

   (ii) Find the maximum likelihood estimator (MLE) $\widehat{\alpha}_n$ of α.

   (iii) Show that the Fisher information of $X_1, \ldots, X_n$ for α is n/α^2.

   (iv) State a positive lower bound on the variance of any unbiased estimator for α.

   (v) State large sample properties of the MLE $\widehat{\alpha}_n$ as $n \to \infty$.
   You do not need to verify regularity conditions.

   (vi) Construct a two-sided asymptotic 95% confidence interval for α.

3. (i) Consider a linear model of the form $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\mathrm{E}(\boldsymbol{\epsilon}) = 0$ for some $\boldsymbol{\beta} \in \mathbb{R}^p$, where $X \in \mathbb{R}^{n \times p}$ is a deterministic matrix. Furthermore, $n > p$.

(a) State if $\boldsymbol{\epsilon}$ is observable.

(b) State the second order assumptions (SOA).

(c) State the normal theory assumptions (NTA).

(d) Give the definition of a least squares estimator of $\boldsymbol{\beta}$.

(e) Show that any $\widehat{\boldsymbol{\beta}}$ satisfying

$$X^T X\widehat{\boldsymbol{\beta}} = X^T\boldsymbol{Y}$$

is a least squares estimator of $\boldsymbol{\beta}$.

(f) Under what conditions is the least squares estimator unique?

(g) Define the vector of residuals and the leverage of the ith observation. Under second order assumptions, derive the variance of the ith residual in terms of the leverage.

(ii) Suppose that the vector of observation $\boldsymbol{Y}$ satisfies

$$\mathrm{E}(\boldsymbol{Y}) = X\boldsymbol{\beta} + \boldsymbol{Z},$$

for some $\boldsymbol{\beta} \in \mathbb{R}^p$, where $X \in \mathbb{R}^{n \times p}$ is a deterministic matrix of full rank and $\boldsymbol{Z} \in \mathbb{R}^n$ is a deterministic vector. Furthermore, $n > p$.

Suppose that we are interested in estimating $\boldsymbol{c}^T\boldsymbol{\beta}$ for some known $\boldsymbol{c} \in \mathbb{R}^p$.

Assume that you estimate $\boldsymbol{\beta}$ from the (incorrect) linear model

$$\mathrm{E}(\boldsymbol{Y}) \in \mathrm{span}(X)$$

with second order assumptions using the least squares estimator $\widehat{\boldsymbol{\beta}}$.

Compute the bias and the mean squared error of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}$ as an estimator for $\boldsymbol{c}^T\boldsymbol{\beta}$.

4. (i) State the Fisher-Cochran Theorem.

(ii) Describe the usual hypotheses considered in the F-test.

(iii) State the test statistic of the F-test in the form that uses residual sums of squares. State its distribution under the null hypothesis.

(iv) Describe the decision rule for an F-test to the level 5%.

(v) Rewrite the test statistic of the F-test using projection matrices.
Starting from the Fisher-Cochran Theorem, prove the distribution of the test statistic of the F-test under the null hypothesis.
You do not need to show that certain matrices are projection matrices and work out their rank; simply state which matrices are projection matrices and give their respective rank without proof.

(vi) What is the distribution of the test statistic of the F-test if the null hypothesis is not necessarily true? Prove this.

Imperial College
London

IMPERIAL COLLEGE LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2012

This paper is also taken for the relevant examination for the Associateship.

M2S2

Statistical Modelling (Solutions)

Setter's signature	Checker's signature	Editor's signature
.....	.....	.....

1. (i) (a) $\text{bias}_\theta(T) = \mathrm{E}_\theta(T) - \theta$.
$\text{MSE}_\theta(T) = \mathrm{E}_\theta((T-\theta)^2)$.

seen ⇓

2

(b) $\text{MSE}_\theta(T) = \mathrm{Var}_\theta(T) + (\text{bias}_\theta(T))^2$

2

Let $X = T - \theta$. Then $\mathrm{Var}_\theta(X) = \mathrm{E}_\theta(X^2) - (\mathrm{E}_\theta X)^2 = \text{MSE}_\theta(T) - (\text{bias}_\theta(T))^2$. Rearraning shows the relationship.

3

(ii) If $W_1 \sim \chi^2_{n_1}(\delta)$, $W_2 \sim \chi^2_{n_2}$ independently then

$$F = \frac{W_1/n_1}{W_2/n_2}$$

is said to have a *non-central F* distribution with (n_1, n_2) d.f. and non-centrality parameter δ.

3

(iii)

sim. seen ⇓

No, as the matrix is not positive semidefinite. To see the latter,

$$(1, -1)\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2 < 0.$$

3

(iv) In the classical approach, the parameter is an unknown constant, whereas in the Bayesian approach, it is the realisation of a random variable.

2

(v)

$\pi(\tau) \propto \tau e^{-2\tau}$ and $f(Y_i|\tau) \propto \sqrt{\tau} e^{-\tau \frac{(Y_i-\mu)^2}{2}}$. Hence,

$$\pi(\tau)\prod_{i=1}^n f(Y_i|\tau) \propto \tau e^{-2\tau}\prod_{i=1}^n \sqrt{\tau} e^{-\tau\frac{(Y_i-\mu)^2}{2}} = \tau^{(2+n/2)-1} e^{-(2+\sum_{i=1}^n (Y_i-\mu)^2/2)\tau}$$

Hence, $\tau|Y_1, \ldots, Y_n \sim \text{Gamma}(2 + \frac{n}{2}, 2 + \frac{1}{2}\sum_{i=1}^n (Y_i - \mu)^2)$

5

2. (i) The pdf of the individual random variables is $f_\alpha(x) = \frac{\partial}{\partial \alpha} \mathrm{P}(X_i \leq x) = \alpha x^{-\alpha-1}$ and thus the likelihood is sim. seen ⇓

$$L(\alpha) = \prod_{i=1}^{n} f_\alpha(x_i) = \alpha^n \prod_{i=1}^{n} x_i^{-\alpha-1}.$$

3

(ii) The log-likelihood $l(\alpha)$ is

$$l(\alpha) = n \log(\alpha) - (\alpha + 1) \sum_{i=1}^{n} \log(x_i).$$

Taking derivatives gives

$$\frac{\partial}{\partial \alpha} l(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^{n} \log(x_i).$$

Equating this to 0 gives $\widehat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log(x_i)}$. 3

This indeed a maximum likelihood estimator as

$$\left(\frac{\partial}{\partial \alpha}\right)^2 l(\alpha) = -\frac{n}{\alpha^2} < 0$$

2

(iii) First, we compute the Fisher information $I_1(\alpha)$ for a sample of size 1.

$$\log f_\alpha(x) = \log(\alpha) - (\alpha + 1) \log(x),$$

$$\left(\frac{\partial}{\partial \alpha}\right)^2 \log f_\alpha(x) = \frac{\partial}{\partial \alpha} \left(\frac{1}{\alpha} - \log(x)\right) = -\frac{1}{\alpha^2},$$

and thus

$$I_1(\alpha) = -\,\mathrm{E}\left[\left(\frac{\partial}{\partial \alpha}\right)^2 \log f_\alpha(X_1)\right] = \frac{1}{\alpha^2}$$

Hence, due to independence, the Fisher information $I_n(\alpha)$ for α based on $X_1, \ldots, X_n$ is $\frac{n}{\alpha^2}$. 3

[Alternatively, one can compute $I_1(\alpha)$ via

$$I_1(\alpha) = \mathrm{E}\left[\left(\frac{\partial}{\partial \alpha} \log f_\alpha(X_1)\right)^2\right] = \mathrm{E}\left[\left(\frac{1}{\alpha} - \log(X_1)\right)^2\right].$$

However, working out the expectation is somewhat lengthy.]

(iv) By the Rao-Cramer inequality, any unbiased estimator T for α satisfies

$$\mathrm{Var}_\alpha(T) \geq \frac{1}{I_n(\alpha)} = \frac{\alpha^2}{n}.$$

3

(v) $\sqrt{n}(\widehat{\alpha} - \alpha) \xrightarrow{d} N(0, \underbrace{\frac{1}{I_1(\alpha)}}_{=\alpha^2})$ as $n \to \infty$. 2

(vi) From Part (v), we get

$$\sqrt{n}(\widehat{\alpha}/\alpha - 1) \xrightarrow{d} N(0,1) \quad (n \to \infty)$$

Let c_1 and c_2 be such that $\Phi(c_1) = 0.025$ and $\Phi(c_2) = 0.975$, where Φ is the cdf of a standard normal distribution. Then

$$\mathrm{P}(c_1 \leq \sqrt{n}(\frac{\widehat{\alpha}}{\alpha} - 1) \leq c_2) \to 0.95.$$

Solving the inequalities for α gives

$$\frac{\widehat{\alpha}}{n^{-\frac{1}{2}}c_2 + 1} \leq \alpha \leq \frac{\widehat{\alpha}}{n^{-\frac{1}{2}}c_1 + 1}$$

implying the approximate confidence interval

$$\left[\frac{\widehat{\alpha}}{n^{-\frac{1}{2}}c_2 + 1}, \frac{\widehat{\alpha}}{n^{-\frac{1}{2}}c_1 + 1}\right]$$

4

unseen ⇓

Alternative solution: *First, show that* $\widehat{\alpha} \to \alpha$*: The survival function of* $\log(X_i)$ *is*

$$\mathrm{P}(\log(X_i) > t) = \mathrm{P}(X_i > \exp(t)) = \exp(t)^{-\alpha} = \exp(-\alpha t).$$

Thus, by the law of large numbers $\frac{1}{n}\sum_{i=1}^{n} \log(X_i) \to \mathrm{E}\log(X_1) = \int_0^\infty e^{-\alpha x} dx = 1/\alpha$.

Then, using, by Slutsky's Lemma,

$$\sqrt{n}\left(1 - \frac{\alpha}{\widehat{\alpha}}\right) = \frac{\sqrt{n}(\widehat{\alpha} - \alpha)}{\widehat{\alpha}} = \sqrt{n}(\widehat{\alpha} - \alpha)\frac{1}{n}\sum_{i=1}^{n} \log(X_i) \xrightarrow{d} N(0,1) \quad (n \to \infty)$$

Let c_1 *and* c_2 *be such that* $\Phi(c_1) = 0.025$ *and* $\Phi(c_2) = 0.975$*, where* Φ *is the cdf of a standard normal distribution. Then*

$$\mathrm{P}(c_1 \leq \sqrt{n}(1 - \frac{\alpha}{\widehat{\alpha}}) \leq c_2) \to 0.95.$$

Solving the inequalities for α *gives*

$$(1 - n^{-\frac{1}{2}}c_1)\widehat{\alpha} \geq \alpha \geq (1 - n^{-\frac{1}{2}}c_2)\widehat{\alpha}$$

giving the approximate confidence interval $[(1 - n^{-\frac{1}{2}}c_2)\widehat{\alpha}, (1 - n^{-\frac{1}{2}}c_1)\widehat{\alpha}]$.

3. (i) (a) No, $\boldsymbol{\epsilon}$ is not observable. seen ⇓ 1

(b) Second Order Assumption (SOA): $\mathrm{cov}(\boldsymbol{\epsilon}) = \sigma^2 I_n$ for some $\sigma^2 > 0$. 1

(c) Normal theory assumptions (NTA): $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$ for some $\sigma^2 > 0$. 1

(d) Let $S(\boldsymbol{\beta}) = (\boldsymbol{Y} - X\boldsymbol{\beta})^T(\boldsymbol{Y} - X\boldsymbol{\beta})$. Any p-variate random vector $\widehat{\beta}$ such that

$$S(\widehat{\boldsymbol{\beta}}) = \min_{\boldsymbol{\beta} \in \mathbb{R}^p} S(\boldsymbol{\beta})$$

is called a least squares estimator. 2

(e) Let $\widehat{\boldsymbol{\beta}}$ satisfy $X^T X\widehat{\boldsymbol{\beta}} = X^T\boldsymbol{Y}$. Then for any $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$\begin{aligned} S(\boldsymbol{\beta}) &= (\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}} + X\widehat{\boldsymbol{\beta}} - X\boldsymbol{\beta})^T(\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}} + X\widehat{\boldsymbol{\beta}} - X\boldsymbol{\beta}) \\ &= S(\widehat{\boldsymbol{\beta}}) + \underbrace{(X\widehat{\boldsymbol{\beta}} - X\boldsymbol{\beta})^T(X\widehat{\boldsymbol{\beta}} - X\boldsymbol{\beta})}_{\geq 0} + 2\underbrace{(X\widehat{\boldsymbol{\beta}} - X\boldsymbol{\beta})^T(\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}})}_{=(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T(X^T\boldsymbol{Y} - X^T X\widehat{\boldsymbol{\beta}}) = 0} \\ &\geq S(\widehat{\boldsymbol{\beta}}) \end{aligned}$$

Hence, $\widehat{\boldsymbol{\beta}}$ minimises S and is therefore a least squares estimator. 3

(f) The least squares estimator is unique iff X is of full rank. 2

(g) The vector $\boldsymbol{e}$ of residuals is defined by $\boldsymbol{e} = \boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}$.
Let P be the projection matrix onto the space spanned by the columns of X.
The leverage of the ith observation is P_{ii}.
As $\boldsymbol{e} = \boldsymbol{Y} - P\boldsymbol{Y} = (I - P)\boldsymbol{Y}$, we have

$$\mathrm{cov}(\boldsymbol{e}) = \mathrm{cov}((I - P)\boldsymbol{Y}) = \sigma^2(I - P)(I - P)^T$$

Using that $(I - P)$ is a projection matrix this simplifies to

$$\mathrm{cov}(\boldsymbol{e}) = \sigma^2(I - P)$$

Hence, $\mathrm{Var}\, e_i = \sigma^2(1 - P_{ii})$. 5

(ii)

$$\begin{aligned} \mathrm{E}\, \boldsymbol{c}^T\widehat{\boldsymbol{\beta}} &= \mathrm{E}[\boldsymbol{c}^T(X^T X)^{-1}X^T\boldsymbol{Y}] = \boldsymbol{c}^T(X^T X)^{-1}X^T\, \mathrm{E}\,[\boldsymbol{Y}] \\ &= \boldsymbol{c}^T(X^T X)^{-1}X^T(X\boldsymbol{\beta} + \boldsymbol{Z}) = \boldsymbol{c}^T\boldsymbol{\beta} + \boldsymbol{c}^T(X^T X)^{-1}X^T\boldsymbol{Z} \end{aligned}$$

Thus $\mathrm{bias}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}) = \boldsymbol{c}^T(X^T X)^{-1}X^T\boldsymbol{Z}$.
Furthermore,

$$\begin{aligned} MSE(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}) &= \mathrm{Var}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}) + (\mathrm{bias}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}))^2 = \boldsymbol{c}^T\, \mathrm{cov}(\widehat{\boldsymbol{\beta}})\boldsymbol{c} + (\mathrm{bias}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}))^2 \\ &= \boldsymbol{c}^T(X^T X)^{-1}\boldsymbol{c}\sigma^2 + (\boldsymbol{c}^T(X^T X)^{-1}X^T\boldsymbol{Z})^2 \end{aligned}$$

5

4. (i) If $A_1, \ldots, A_k$ are $n \times n$ projection matrices such that $\sum_{i=1}^{k} A_i = I_n$, and if $\boldsymbol{Z} \sim N(\boldsymbol{\mu}, I_n)$ then $\boldsymbol{Z}^T A_1 \boldsymbol{Z}, \ldots, \boldsymbol{Z}^T A_k \boldsymbol{Z}$ are independent and

seen ⇓

$$\boldsymbol{Z}^T A_i \boldsymbol{Z} \sim \chi^2_{r_i}(\delta_i), \quad \text{where } r_i = \operatorname{rank} A_i \text{ and } \delta_i^2 = \boldsymbol{\mu}^T A_i \boldsymbol{\mu}.$$

3

(ii) We want to test whether a sub-model of a linear model $\mathrm{E}\,\boldsymbol{Y} = X\boldsymbol{\beta}$ is true, i.e. we want to test

$$H_0 \,:\, \mathrm{E}\,\boldsymbol{Y} \in \operatorname{span}(X_0) \text{ against } H_1 \,:\, \mathrm{E}\,\boldsymbol{Y} \notin \operatorname{span}(X_0)$$

for some matrix X_0 with $\operatorname{span}(X_0) \subset \operatorname{span}(X)$.

2

(iii) Under $H_0 \,:\, \mathrm{E}\,\boldsymbol{Y} \in \operatorname{span}(X_0)$,

$$F = \frac{\mathrm{RSS}_0 - \mathrm{RSS}}{\mathrm{RSS}} \cdot \frac{n-r}{r-s} \sim F_{r-s,n-r}$$

where $r = \operatorname{rank} X$, $s = \operatorname{rank} X_0$, RSS= residual sum of squares in the full model $\mathrm{E}\,\boldsymbol{Y} = X\boldsymbol{\beta}$ and RSS= residual sum of squares in the sub-model $\mathrm{E}\,\boldsymbol{Y} = X_0\widetilde{\boldsymbol{\beta}}$.

3

(iv) The null is rejected for large values of F. As we want a test to the level 0.05, we reject if

$$F > c,$$

where c is such that $\mathrm{P}(X \geq c) = 0.05$ for $X \sim F_{r-s,n-r}$.

2

(v) Let P be the projection matrix onto $\operatorname{span} X$, and let $Q = I - P$ the projection matrix onto $(\operatorname{span} X)^\perp$.
Likewise, let P_0 be the projection matrix onto $\operatorname{span} X_0$, and let $Q_0 = I - P_0$ the projection matrix onto $(\operatorname{span} X_0)^\perp$. Then,

$$\mathrm{RSS} = \boldsymbol{Y}^T Q \boldsymbol{Y}, \quad \mathrm{RSS}_0 = \boldsymbol{Y}^T Q_0 \boldsymbol{Y}.$$

Using this gives

$$F = \frac{\boldsymbol{Y}^T Q_0 \boldsymbol{Y} - \boldsymbol{Y}^T Q \boldsymbol{Y}}{\boldsymbol{Y}^T Q \boldsymbol{Y}} \cdot \frac{n-r}{r-s} = \frac{\boldsymbol{Y}^T (P - P_0) \boldsymbol{Y}/\sigma^2}{\boldsymbol{Y}^T (I-P) \boldsymbol{Y}/\sigma^2} \cdot \frac{n-r}{r-s}$$

2

Let $\boldsymbol{Z} = \boldsymbol{Y}/\sigma$.
Let $A_1 = I - P$, $A_2 = P - P_0$, $A_3 = P_0$.
Clearly, $A_1 + A_2 + A_3 = I$. Furthermore, A_1, A_2, and A_3 are projection matrices.
The Fisher-Cochran theorem now implies

* $\boldsymbol{Z}^T (P - P_0) \boldsymbol{Z}$ and $\boldsymbol{Z}^T (I - P) \boldsymbol{Z}$ are independent,
* $\boldsymbol{Z}^T (P - P_0) \boldsymbol{Z} \sim \chi^2_{\operatorname{rank}(P-P_0)} (\mathrm{E}\, \boldsymbol{Z}^T (P - P_0) \,\mathrm{E}\, \boldsymbol{Z})$,
* $\boldsymbol{Z}^T (I - P) \boldsymbol{Z} \sim \chi^2_{\operatorname{rank}(I-P)} (\mathrm{E}\, \boldsymbol{Z}^T (I - P) \,\mathrm{E}\, \boldsymbol{Z})$.

We show that the non-centrality parameters are 0.
Under H_0, we know $\mathrm{E}\, \boldsymbol{Z} = \frac{1}{\delta} \mathrm{E}\, \boldsymbol{Y} \in \operatorname{span}(X_0) \subset \operatorname{span}(X)$ Thus,

$$(P - P_0) \,\mathrm{E}\, \boldsymbol{Z} = \underbrace{P \,\mathrm{E}\, \boldsymbol{Z}}_{=\mathrm{E}\, \boldsymbol{Z}} - \underbrace{P_0 \,\mathrm{E}\, \boldsymbol{Z}}_{=\mathrm{E}\, \boldsymbol{Z}} = \boldsymbol{0}.$$

Hence, $\mathrm{E}\, \boldsymbol{Z}^T (P - P_0) \,\mathrm{E}\, \boldsymbol{Z} = 0$. Furthermore,

$$E \boldsymbol{Z}^T (I - P) \,\mathrm{E}\, \boldsymbol{Z} = E \boldsymbol{Z}^T (\mathrm{E}\, \boldsymbol{Z} - \underbrace{P \,\mathrm{E}\, \boldsymbol{Z}}_{=\mathrm{E}\, \boldsymbol{Z}}) = 0.$$

Without proof: $\operatorname{rank}(I - P) = n - r$, $\operatorname{rank}(P - P_0) = r - s$.
To summarise, we have shown

$$\boldsymbol{Z}^T (P - P_0) \boldsymbol{Z} \sim \chi^2_{r-s},\ \boldsymbol{Z}^T (I - P) \boldsymbol{Z} \sim \chi^2_{n-r} \text{ independently.}$$

Thus, by definition, $F \sim F_{r-s,n-r}$. 5

(vi) If H_0 is not true then the proof is still valid, except for the non-centrality parameter of $\boldsymbol{Z}^T (P - P_0) \boldsymbol{Z}$. Now,

$$\mathrm{E}\, \boldsymbol{Z}^T (P - P_0) \,\mathrm{E}\, \boldsymbol{Z} = \frac{1}{\sigma^2} \boldsymbol{\beta}^T X^T (P - P_0) X \boldsymbol{\beta}.$$

Thus, without assuming H_0, we get

$$F \sim F_{r-s,n-r}(\delta), \text{ where } \delta^2 = \frac{1}{\sigma^2} (X\boldsymbol{\beta})^T (P - P_0) X \boldsymbol{\beta}.$$

3