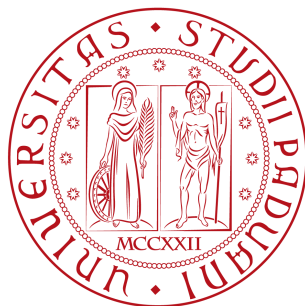


Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Magistrale in  
Scienze Statistiche



## TEST DIREZIONALI PER MODELLI DI REGRESSIONE

Relatore Prof. Nicola Sartori  
Dipartimento di Scienze Statistiche

Laureando: Alice Ruffato  
Matricola N. 1100706

Anno Accademico 2016/2017



# Indice

<b>Introduzione</b>	<b>5</b>
<b>1 Test basati sulla verosimiglianza</b>	<b>7</b>
1.1 Introduzione . . . . .	7
1.2 Test rapporto di verosimiglianza . . . . .	7
1.3 Test direzionale . . . . .	9
1.4 Test $w^*$ di Skovgaard . . . . .	12
<b>2 L'applicazione al modello lineare</b>	<b>15</b>
2.1 Introduzione . . . . .	15
2.2 Il modello di regressione lineare . . . . .	15
2.2.1 Inferenza su un vettore di parametri . . . . .	16
2.3 Test F per modelli di regressione . . . . .	18
2.4 Test direzionale per modelli di regressione . . . . .	20
2.5 Test $w^*$ di Skovgaard per modelli di regressione . . . . .	26
<b>3 Confronto e risultati</b>	<b>29</b>
3.1 Introduzione . . . . .	29
3.2 Soluzione esatta del test direzionale . . . . .	29
3.2.1 Generalizzazione del risultato . . . . .	32
3.3 Soluzione approssimata del test di Skovgaard . . . . .	34
<b>Conclusioni</b>	<b>39</b>
<b>A Proprietà della stima di massima verosimiglianza <math>\hat{\theta}(t)</math></b>	<b>43</b>
<b>B Dimostrazione analitica dell'uguaglianza dei <math>p</math>-value</b>	<b>47</b>



# Introduzione

Il confronto tra modelli è, in ambito statistico, una delle fasi più importanti e delicate dell'analisi di un fenomeno. Esistono vari metodi per operare tale confronto: indici di bontà di adattamento, criteri d'informazione, test statistici, per citarne alcuni. In questa relazione, si analizza il caso di modelli annidati, che differiscono per un insieme di vincoli posti su un vettore di parametri d'interesse. Lo strumento più comune e usato nel confronto di modelli annidati è il test rapporto di verosimiglianza: esso è di facile costruzione, sotto l'ipotesi nulla ha una distribuzione asintotica nota e fornisce risultati soddisfacenti in molti casi. Tuttavia, l'approssimazione che si ottiene usando la distribuzione limite spesso non è accurata, in particolare quando l'informazione campionaria è esigua rispetto alla complessità dei modelli coinvolti. Lo scopo di questo lavoro è di proporre un metodo alternativo al test rapporto di verosimiglianza che fornisca approssimazioni più accurate. In particolare, vengono presentate due alternative: il test di Skovgaard, che è una modifica del test log-rapporto di verosimiglianza, in cui viene introdotto un fattore di correzione che porta ad un miglioramento dell'accuratezza della approssimazione asintotica della distribuzione nulla del test; il test direzionale, che si basa sul condizionamento alla direzione osservata tra il modello nullo e il modello stimato. Contestualmente, viene mostrata l'applicazione dei tre test al modello di regressione lineare, soffermandosi in modo particolare sul test direzionale, oggetto di studio principale della tesi. Nello specifico, si considerano ipotesi lineari sul vettore dei parametri di regressione e si analizza la relazione tra il test direzionale e il test rapporto di verosimiglianza esatto, cioè il test  $F$ .

Il lavoro svolto è così suddiviso: nel primo capitolo vengono presentati i

tre test statistici, ponendo maggiore attenzione al meno noto test direzionale, alla sua costruzione e alle sue caratteristiche.

Il secondo capitolo riguarda l'applicazione al modello di regressione lineare normale: tale modello è ampiamente utilizzato in vari ambiti per la sua facile interpretabilità e specificazione. In particolare è nostro interesse verificare un sistema di ipotesi lineari che specifica delle relazioni tra i coefficienti di regressione. A tale scopo si utilizzano i tre test statistici precedentemente descritti. Quindi, dopo aver definito il modello e il sistema d'ipotesi considerati, si procede con la costruzione del test  $F$ , del test direzionale e del test di Skovgaard.

Infine, nel terzo capitolo si confrontano i risultati ottenuti dal test direzionale e dal test di Skovgaard con la soluzione esatta. Si mostrerà che il livello di significatività osservato del test direzionale coincide con quello esatto, ottenuto utilizzando il test  $F$ . Il test di Skovgaard, invece, fornisce un'approssimazione accurata ad eccezione di casi limite.

# Capitolo 1

## Test basati sulla verosimiglianza

### 1.1 Introduzione

In questa tesi si considera il confronto tra modelli di regressione lineare annidati attraverso verifica d'ipotesi e si forniscono gli strumenti statistici adeguati a tale scopo. Due modelli si dicono annidati quando hanno un'analoga specificazione, ma in uno dei due vengono imposti dei vincoli supplementari su determinati parametri. Più in particolare, definito  $M_1$  il modello completo che contiene  $p$  parametri e  $M_2$  il modello ridotto che contiene  $p - d$  parametri, si ha che  $M_2 \subset M_1$ , ovvero  $M_2$  è la restrizione di  $M_1$  ottenuta imponendo  $d$  vincoli (Azzalini, 2004, Cap.6). In questo capitolo viene presentato inizialmente il test rapporto di verosimiglianza, lo strumento più usato nel confronto di modelli annidati. In seguito si propongono due test alternativi: il test direzionale e il test di Skovgaard; il primo, meno noto, verrà descritto maggiormente nel dettaglio.

### 1.2 Test rapporto di verosimiglianza

Si consideri il vettore delle osservazioni  $y = (y_1, \dots, y_n)$  come una realizzazione di una variabile casuale  $Y$  con distribuzione di probabilità  $p_0(y)$ . Il modello generatore dei dati non è noto, ma si supponga di aver specificato un opportuno modello parametrico  $f(y; \theta)$ , con  $\theta = (\theta_1, \dots, \theta_p)$ . La funzione

di verosimiglianza è definita come:

$$L(\theta; y) = c(y)f(y; \theta)$$

dove  $c(y)$  è una costante che non dipende da  $\theta$ . Solitamente si usa la funzione di log-verosimiglianza, ovvero il logaritmo della funzione di verosimiglianza:

$$l(\theta) = l(\theta; y) = \log L(\theta; y).$$

Massimizzando la quantità precedente, si ottiene la stima di massima verosimiglianza  $\hat{\theta} = \hat{\theta}(y)$ .

Si consideri l'ipotesi nulla  $H_\psi : \psi(\theta) = \psi$ , dove  $\psi$  ha dimensione  $d \leq p$ . Spesso, eventualmente dopo un'opportuna riparametrizzazione, si può scrivere  $\theta = (\psi, \lambda)$ , dove  $\psi$  è il parametro di interesse e  $\lambda$  è un parametro ausiliario di disturbo di dimensione  $p-d$ . Uno strumento adatto alla verifica d'ipotesi, e comunemente usato in ambito statistico, è il *test rapporto di verosimiglianza*:

$$\delta = \delta(\psi) = \frac{\sup_{\theta \in \Theta_0} L(\theta; y)}{\sup_{\theta \in \Theta} L(\theta; y)} = \frac{L(\hat{\theta}_\psi; y)}{L(\hat{\theta}; y)}$$

dove  $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$  è la stima di massima verosimiglianza vincolata di  $\theta$  sotto  $H_\psi$ . Si è soliti utilizzare però un'altra formulazione del test, chiamata anche *test log-rapporto di verosimiglianza*, che è una trasformazione monotona di  $\delta$ :

$$w(\psi) = -2 \log \delta(\psi) = 2\{l(\hat{\theta}) - l(\hat{\theta}_\psi)\}. \quad (1.1)$$

Valori grandi di  $w(\psi)$  sono critici per l'ipotesi nulla. In alcuni casi,  $w(\psi)$  è una funzione monotona crescente di una statistica  $t(y)$  la cui distribuzione sotto  $H_\psi$  è nota. Questo succede ad esempio nel modello di regressione lineare in cui (1.1) è funzione monotona crescente del test  $F$ , di cui è nota la distribuzione esatta (si veda il Paragrafo 2.3). Tuttavia, usualmente si ricorre ad una distribuzione nulla approssimata che, sotto opportune condizioni di regolarità, è  $\chi_d^2$  (Pace e Salvan, 2001, Cap.3). Le approssimazioni che si ottengono usando la distribuzione limite possono non essere accurate in determinati casi: quando, ad esempio, la numerosità campionaria è bassa in relazione alla dimensione  $p$  del parametro o alla dimensione  $d$  dell'ipotesi in esame. Per tale motivo in letteratura sono stati proposti altri test statistici che cercano di risolvere questo problema: due di essi verranno analizzati in seguito.



## 1.3 Test direzionale

I test direzionali furono proposti inizialmente in Fraser e Massam (1985) come strumento per semplificare il calcolo di densità marginali ad alta dimensione sostituendole con densità condizionate uni-dimensionali, nel contesto di inferenza marginale su parametri di regressione in modelli lineari non normali. Skovgaard (1988) li riprese proponendoli attraverso un'approssimazione *saddlepoint* della sua distribuzione, sostenendo il fatto che sono equivalenti al primo ordine al test rapporto di verosimiglianza, ma che forniscono un'approssimazione più accurata del  $p$ -value, specialmente per  $p$ -value piccoli. Successivamente, Davison et al. (2014) hanno proposto l'uso di test direzionali per fare inferenza su un vettore di parametri canonici d'interesse in modelli appartenenti a famiglie esponenziali naturali. Tale utilizzo viene ripreso in questa tesi e applicato al caso di parametri di regressione in un modello lineare normale multivariato. Recentemente, Fraser et al. (2016) hanno esteso l'applicazione dei test direzionali anche a modelli generali e/o a parametri d'interesse non lineari.

L'approccio direzionale si basa sulla determinazione di un  $p$ -value ottenuto valutando quanto sia probabile osservare, sotto l'ipotesi nulla, dei dati più "estremi" rispetto all'ipotesi di quelli osservati, condizionatamente alla direzione che collega i dati osservati ai corrispondenti attesi sotto l'ipotesi nulla. Dal punto di vista tecnico si tratta di ottenere una densità condizionata a tale direzione da utilizzare nel calcolo del  $p$ -value, che si riduce ad una semplice integrazione numerica scalare.

La costruzione del test direzionale per un vettore di parametri si sviluppa in tre fasi: la prima fase consiste nella riduzione della dimensione da quella campionaria,  $n$ , a quella dei parametri,  $p$ ; nel caso il modello appartenga ad una famiglia esponenziale, ciò è possibile considerando la densità marginale della statistica sufficiente. La seconda fase consiste in un'ulteriore riduzione della dimensione, da  $p$  a  $d$ , il numero di parametri d'interesse. Se il parametro d'interesse,  $\psi$ , è lineare nel parametro canonico, allora la riduzione si ottiene condizionandosi alla componente della statistica sufficiente relativa al parametro di disturbo; altrimenti, fissato  $\psi$  dal sistema d'ipotesi, si utilizza una statistica ancillare, la cui distribuzione è approssimativamente libera dal

parametro di disturbo e quindi ci si condiziona ad essa. La terza fase, infine, consiste nel costruire la distribuzione condizionata unidimensionale nella direzione indicata dai dati osservati e ottenere il  $p$ -value direzionale tramite integrazione numerica. Di seguito si presenta nel dettaglio la costruzione del test direzionale, nel caso in cui il parametro d'interesse non sia una forma lineare di quello canonico.

Si assuma che il modello sia una famiglia esponenziale con parametro canonico  $\varphi = \varphi(\theta)$  e con statistica sufficiente  $u = u(y)$ , ovvero

$$f(y; \theta) = h(y) \exp[\varphi(\theta)^T u(y) - K\{\varphi(\theta)\}], \quad (1.2)$$

con associata funzione di log-verosimiglianza  $l(\theta; y) = \varphi(\theta)^T u(y) - K\{\varphi(\theta)\}$ , dove  $K\{\varphi(\theta)\}$  è la *funzione generatrice dei cumulanti*. Per convenienza, si centra la statistica sufficiente  $u$  nel suo punto osservato  $u_{\text{oss}} = u(y_{\text{oss}})$ ; riscrivendo la log-verosimiglianza, si ottiene quindi

$$\begin{aligned} l(\theta; s) &= \varphi(\theta)^T s + l_{\text{oss}}(\theta) = \varphi(\theta)^T (u - u_{\text{oss}}) + l(\theta; u_{\text{oss}}) \\ &= \varphi(\theta)^T s + \varphi(\theta)^T u_{\text{oss}} - K\{\varphi(\theta)\}, \end{aligned} \quad (1.3)$$

dove  $s = u - u_{\text{oss}}$  e  $l_{\text{oss}}(\theta) = l(\theta; s = 0) = l(\theta; u = u_{\text{oss}})$ . L'approssimazione *saddlepoint* (Pace e Salvan, 1996, Paragrafo 10.7) della (1.2) espressa in  $s$  è

$$f(s; \varphi) = \frac{\exp(k'/n)}{(2\pi)^{p/n}} \exp\{l(\varphi; s) - l(\widehat{\varphi}; s)\} |j_{\varphi\varphi}\{\widehat{\varphi}(s)\}|^{-1/2}, \quad (1.4)$$

dove  $\widehat{\varphi}$  è la stima di massima verosimiglianza di  $\varphi$  e l'informazione osservata di Fisher viene calcolata nella parametrizzazione  $\varphi$

$$|j_{\varphi\varphi}\{\widehat{\varphi}(s)\}| = \left| -\frac{\partial^2 l(\varphi; s)}{\partial \varphi \partial \varphi^T} \right|_{\widehat{\varphi}(s)} = \left| -\frac{\partial^2 l_{\text{oss}}(\theta)}{\partial \varphi \partial \varphi^T} \right|_{\widehat{\theta}(s)} = |j_{\theta\theta}\{\widehat{\theta}(s); s\}| |\varphi_{\theta}\{\widehat{\theta}(s)\}|^{-2},$$

dove  $j_{\theta\theta}(\theta; s)$  è la matrice di informazione osservata ottenuta dalla (1.3).

Si consideri il sistema di ipotesi  $H_{\psi}$  nel quale si specifica un valore per il parametro d'interesse  $\psi(\theta)$ . Sia  $\lambda$  il parametro di disturbo, tale che  $\theta = (\psi, \lambda)$ , e si definisca il piano  $d$ -dimensionale

$$L_{\psi}^{\text{oss}} = \{s : \widehat{\lambda}_{\psi} = \widehat{\lambda}_{\psi}^{\text{oss}}\},$$

costituito dai punti dello spazio campionario per i quali il valore della stima di massima verosimiglianza vincolata del parametro di disturbo è uguale

a quella osservata. Si può dimostrare che esiste approssimativamente una statistica ancillare per il parametro di disturbo nel modello con  $\psi$  fissato (Fraser e Reid, 1995). La distribuzione condizionata di  $s$ , data tale statistica ancillare approssimata, si può approssimare in modo analogo alla (1.4):

$$f(s; \hat{\varphi}_\psi) = \frac{\exp(k''/n)}{(2\pi)^{(p-d)/n}} \exp\{l(\varphi; s) - l(\hat{\varphi}_\psi; s)\} |j_{(\lambda\lambda)}\{\hat{\varphi}_\psi\}|^{-1/2}, \quad (1.5)$$

con  $s \in L_\psi^{\text{oss}}$  e la matrice di informazione osservata calcolata nella parametrizzazione  $\varphi$

$$|j_{(\lambda\lambda)}\{\hat{\varphi}_\psi\}| = |j_{\lambda\lambda}\{\hat{\varphi}_\psi; s\}| |\partial\varphi(\theta)/\partial\lambda|_{\hat{\theta}_\psi}^{-2}.$$

Il rapporto tra le due densità condizionate approssimate (1.4) e (1.5) fornisce la densità approssimata del modello ridotto in  $\mathbb{R}^d$ ,

$$h(s; \psi) = \frac{\exp(k/n)}{(2\pi)^{d/n}} \exp\{l(\hat{\varphi}_\psi; s) - l(\hat{\varphi}; s)\} |j_{\varphi\varphi}(\hat{\varphi})|^{-\frac{1}{2}} |j_{(\lambda\lambda)}(\hat{\varphi}_\psi)|^{\frac{1}{2}}, \quad s \in L_\psi^{\text{oss}}. \quad (1.6)$$

Infine, si costruisce il modello uni-dimensionale considerando la magnitudo di  $s$ , condizionatamente alla sua direzione. A tale scopo si esamina la curva  $L_\psi^*$ , in  $L_\psi^{\text{oss}}$ , ottenuta congiungendo il valore osservato di  $s$ ,  $s_{\text{oss}} = 0$ , con  $s_\psi$ , il valore atteso di  $s$  in (1.6) sotto  $H_\psi$ , che è la radice della seguente equazione:

$$s + \frac{\partial l_{\text{oss}}\{\theta(\varphi)\}}{\partial\varphi} \Big|_{\hat{\varphi}_\psi^{\text{oss}}} = 0,$$

dove  $l_{\text{oss}}\{\theta(\varphi)\}$  è la funzione di log-verosimiglianza osservata del modello originale. La curva  $L_\psi^*$  può essere parametrizzata, con  $t \in \mathbb{R}$ , nel modo seguente

$$s(t) = s_\psi + t(s_{\text{oss}} - s_\psi) = (1 - t)s_\psi.$$

Al crescere di  $t$  lungo tale linea la stima di massima verosimiglianza vincolata  $\hat{\varphi}_\psi$  rimane sempre la stessa, per costruzione. La stima di massima verosimiglianza globale, invece, cambia al variare di  $t$  e verrà indicata con  $\hat{\varphi}(t)$ . La densità condizionata approssimata, vincolata a  $L_\psi^*$ , si ottiene dalla (1.6), con le opportune sostituzioni

$$h\{s(t); \psi\} = \frac{\exp(k/n)}{(2\pi)^{d/n}} \exp [l\{\hat{\varphi}_\psi; s(t)\} - l\{\hat{\varphi}(t); s(t)\}] |j_{\varphi\varphi}\{\hat{\varphi}(t)\}|^{-\frac{1}{2}} |j_{(\lambda\lambda)}(\hat{\varphi}_\psi)|^{\frac{1}{2}}. \quad (1.7)$$

Infine, dalla (1.7) è possibile calcolare il  $p$ -value direzionale, che fornisce una misura di quanto  $s(t)$  sia più estremo del valore  $s_{\text{oss}}$ , condizionatamente ad essere in  $L_{\psi}^*$

$$p(\psi) = \frac{\int_1^{t_u} t^{d-1} h\{s(t); \psi\} dt}{\int_0^{t_u} t^{d-1} h\{s(t); \psi\} dt}. \quad (1.8)$$

Nella (1.8) il fattore  $t^{d-1}$  deriva dallo Jacobiano della trasformazione di  $s$  in  $(\|s\|, s/\|s\|)$ , e  $t_u$  è il più grande valore di  $t$  per il quale lo stimatore di massima verosimiglianza corrispondente a  $s(t)$  esiste.

Come si può notare, l'approccio direzionale per fare inferenza su un vettore di parametri è sostanzialmente diverso dagli altri metodi basati sul rapporto di verosimiglianza. Tuttavia, esso fornisce risultati molto accurati anche nel caso di dimensioni campionarie piccole o di un numero elevato di parametri d'interesse (si vedano gli esempi in Davison et al., 2014, e Fraser et al., 2016).

## 1.4 Test $w^*$ di Skovgaard

Skovgaard (2001) ha proposto una modifica del test log-rapporto di verosimiglianza che ha la stessa distribuzione approssimata di  $w(\psi)$ , e cioè  $\chi_d^2$ , ma che ha maggiore accuratezza specialmente nelle code della distribuzione. Tale modificazione ha la seguente forma:

$$w^*(\psi) = w(\psi) \left\{ 1 - \frac{\log \gamma(\psi)}{w(\psi)} \right\}^2 \quad (1.9)$$

dove il fattore di modificazione  $\gamma(\psi)$  in famiglie esponenziali è dato da

$$\gamma(\psi) = \frac{\{(u - \tilde{v})^T \tilde{\Sigma}^{-1} (u - \tilde{v})\}^{\frac{d}{2}}}{w(\psi)^{\frac{d}{2}-1} (\hat{\varphi} - \hat{\varphi}_{\psi})^T (u - \tilde{v})} \left( \frac{|\tilde{\Sigma}| |\tilde{j}_{\lambda\lambda}|}{|\hat{\Sigma}| |\tilde{i}_{\lambda\lambda}|} \right)^{\frac{1}{2}}. \quad (1.10)$$

Nella formula (1.10)  $\varphi$  è il vettore dei parametri canonici,  $u$  il vettore delle statistiche sufficienti canoniche corrispondenti,  $v$  e  $\Sigma$  sono rispettivamente il valore atteso e la matrice di varianza-covarianza di  $u$ ,  $j_{\lambda\lambda}$  e  $i_{\lambda\lambda}$  sono i blocchi delle matrici di informazione osservata e attesa relativi a  $\lambda$ . Inoltre, una tilde e un cappello indicano che la quantità è calcolata in  $\hat{\theta}_{\psi}$  o in  $\hat{\theta}$ , rispettivamente. Esiste in realtà una versione più generale valida anche al di fuori di

---

famiglie esponenziali (Skovgaard, 2001). Pur avendo in teoria lo stesso ordine di errore di  $w(\psi)$  nella distribuzione asintotica, e cioè  $O(n^{-1})$ , per  $w^*(\psi)$  l'errore è relativo in una regione delle grandi deviazioni e quindi ci si aspetta una maggiore accuratezza, specialmente nelle code della distribuzione. Analogamente a  $w(\psi)$ , anche il test  $w^*(\psi)$  di Skovgaard fornisce una misura della deviazione dall'ipotesi nulla: più questa distanza aumenta, più si ha evidenza contro  $H_\psi$ .



# Capitolo 2

## L'applicazione al modello lineare

### 2.1 Introduzione

In questo capitolo è presentata l'applicazione dei test descritti nel capitolo precedente al modello di regressione lineare normale. In particolare, si considerano ipotesi lineari sui coefficienti di regressione. Si è deciso di concentrare le analisi sul modello di regressione lineare perchè è un caso in cui esiste un test esatto. È quindi particolarmente interessante studiare la relazione del test direzionale, e del test di Skovgaard, con la soluzione esatta.

### 2.2 Il modello di regressione lineare

Si definisce modello di regressione lineare una qualsiasi relazione esprimibile nella forma

$$Y = X\beta + \varepsilon,$$

in cui  $X$  è una matrice di dimensione  $n \times p$ , con  $n \geq p$ , a rango pieno nella quale ciascuna colonna rappresenta una variabile esplicativa,  $\beta$  è un vettore di parametri ignoti di dimensione  $p \times 1$  ed  $\varepsilon$  è un vettore stocastico di dimensione  $n \times 1$ . La quantità  $X\beta$  è la componente sistematica del modello di regressione e dipende dai valori assunti dalle variabili esplicative, detti anche *regressori*;  $\varepsilon$  invece è la componente aleatoria, detta *errore*, la cui distribuzione determina quella di  $Y$ . Alcuni importanti risultati si possono ottenere senza assumere

una distribuzione specifica del termine d'errore, ma a partire dalle *ipotesi del secondo ordine*. Si richiede cioè che:

- il valore atteso degli errori sia nullo,  $E(\varepsilon) = \mathbf{0}$ ;
- gli errori siano incorrelati e con la medesima varianza costante,  $Var(\varepsilon) = \sigma^2 I_n$ .

Qui e nel seguito  $\mathbf{0}$  indica un vettore o una matrice di zeri, la cui dimensione si desumerà dal contesto. Le due ipotesi implicano che  $E(Y) = X\beta$  e  $Var(Y) = \sigma^2 I_n$  (*omoschedasticità*). Tuttavia, per condurre analisi inferenziali bisogna ipotizzare una distribuzione per l'errore  $\varepsilon$ . Sotto l'ipotesi di normalità si ha che  $\varepsilon$  è distribuito come una normale multivariata

$$\varepsilon \sim N_n(0, \sigma^2 I_n), \quad \text{che implica} \quad Y \sim N_n(X\beta, \sigma^2 I_n).$$

Il modello che ne deriva è detto *modello di regressione lineare normale*.

## 2.2.1 Inferenza su un vettore di parametri

La stima dei coefficienti  $\beta$  è di fondamentale importanza in un modello di regressione lineare, poiché permette di quantificare ed interpretare la relazione tra la variabile risposta e le variabili esplicative. Inoltre, se un coefficiente di regressione  $\beta_j$  è nullo, la corrispondente variabile esplicativa è esclusa dal modello; differenti valori di tale regressore, a parità degli altri, non creano infatti variazioni nella legge di probabilità della variabile risposta (Pace e Salvan, 2001, Cap.9). È quindi necessario individuare quali sono le variabili rilevanti nello spiegare la variabile risposta. Uno strumento utile a tale scopo è verificare la nullità di un singolo o di un gruppo di parametri con un opportuni test statistici; più in generale, si è soliti verificare la validità di specifici vincoli lineari che sono di particolare interesse per l'analisi.

Si consideri quindi il sistema d'ipotesi

$$\begin{cases} H_0 : H\beta = h \\ H_1 : H\beta \neq h \end{cases} \quad (2.1)$$

dove  $H$  è una matrice  $d \times p$  ( $d \leq p$ ) di rango  $d$  formata da costanti specificate e  $h$  è un vettore  $d \times 1$  di costanti note. Il sistema (2.1) è detto costituire un



sistema di *ipotesi lineari* perchè specifica delle relazioni lineari tra le componenti di  $\beta$ . Si noti che ipotesi di nullità di singole, o gruppi di, componenti di  $\beta$  sono casi particolari della (2.1).

Si definisca con  $\theta = (\beta^T, \sigma^2)^T$  il vettore dei parametri del modello di regressione specificato, di dimensione  $(p+1) \times 1$ . La log-verosimiglianza è

$$\begin{aligned} l(\theta) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y - X\beta)^T (y - X\beta) \\ &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} y^T y - \frac{1}{2\sigma^2} \beta^T X^T X \beta + \frac{1}{\sigma^2} \beta^T X^T y. \end{aligned} \quad (2.2)$$

Il modello è una famiglia esponenziale (1.2) con vettore dei parametri canonici  $\varphi(\theta) = \{\beta^T/\sigma^2, -1/(2\sigma^2)\}^T$  e vettore delle statistiche sufficienti ad esso associate  $u = (y^T X, y^T y)^T$ , entrambi di dimensione  $(p+1) \times 1$ . Si noti che l'ipotesi nulla in (2.1) è una funzione lineare dei parametri canonici. Infatti  $H\beta - h = \mathbf{0}$  implica  $H\beta/\sigma^2 - h/\sigma^2 = \mathbf{0}$  e quindi  $H(\beta/\sigma^2) + 2h\{-1/(2\sigma^2)\} = \mathbf{0}$ , ovvero  $\tilde{H}\varphi = \mathbf{0}$  con  $\tilde{H} = (H, 2h)$ .

In questa tesi, e quindi per tutte le analisi successive, si considera un caso particolare della (2.1), ossia

$$\begin{cases} H_0 : \beta_1 = \mathbf{0} \\ H_1 : \beta_1 \neq \mathbf{0} \end{cases} \quad (2.3)$$

in cui si verifica che un gruppo di componenti del vettore  $\beta$ , senza perdita di generalità le prime  $d$ , sia nullo. In questo caso la matrice  $H$ , di dimensione  $d \times p$ , e i vettori  $\beta$  e  $h$ , di dimensione rispettivamente  $(p+1) \times 1$  e  $d \times 1$ , sono definiti come

$$H\beta = [I_d \quad \mathbf{0}] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \beta_1 = h = \mathbf{0},$$

in cui il vettore  $\beta$  è suddiviso nei due blocchi di componenti  $\beta = [\beta_1^T \quad \beta_2^T]^T$ , di dimensione  $d \times 1$  e  $(p-d) \times 1$  rispettivamente. Si noti che ci si può ricondurre al caso generale in (2.1) tramite una semplice riparametrizzazione. Nello specifico, si supponga di voler verificare l'ipotesi nulla

$$H_0 : H\beta = \mathbf{0} \quad (2.4)$$

con  $H$  definita come in (2.1). Sia  $J$  una matrice di costanti note di dimensione  $(p-d) \times p$  e rango  $p-d$ , tale che

$$\bar{H} = \begin{bmatrix} H \\ J \end{bmatrix}$$

sia una matrice di dimensione  $p \times p$  di rango pieno. Anche se non necessario,  $J$  può essere definita, ad esempio, ortogonale rispetto a  $H$ , ovvero tale che  $JH^T = \mathbf{0}$ . Quindi

$$\begin{bmatrix} \psi \\ \lambda \end{bmatrix} = \bar{H}\beta = \begin{bmatrix} H\beta \\ J\beta \end{bmatrix}$$

definisce una riparametrizzazione  $(\psi^T, \lambda^T, \sigma^2)$  di  $(\beta^T, \sigma^2)$  e l'ipotesi nulla (2.4) diventa  $H_0 : \psi = \mathbf{0}$ . La funzione di log-verosimiglianza relativa a tale parametrizzazione è

$$\begin{aligned} l(\psi, \lambda, \sigma^2) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y - X\bar{H}^{-1}\bar{H}\beta)^T (y - X\bar{H}^{-1}\bar{H}\beta) \\ &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y - X^*\bar{H}\beta)^T (y - X^*\bar{H}\beta) \\ &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \left( y - X^* \begin{bmatrix} \psi \\ \lambda \end{bmatrix} \right)^T \left( y - X^* \begin{bmatrix} \psi \\ \lambda \end{bmatrix} \right), \end{aligned}$$

dove  $X^* = X\bar{H}^{-1}$  è la matrice disegno riscalata. Anche il caso con  $h \neq \mathbf{0}$  nella (2.4) è facilmente esprimibile in questa forma e perciò è sempre possibile ricondursi al caso in cui si verifica l'ipotesi che un sottoinsieme di componenti del vettore  $\beta$  sia nulla.

Nei paragrafi successivi il parametro di interesse verrà indicato con  $\psi = \beta_1$  e quello di disturbo con  $\lambda = (\beta_2^T, \sigma^2)^T$ , da cui  $\theta = (\psi^T, \lambda^T)^T$ .

## 2.3 Test F per modelli di regressione

Si consideri il modello di regressione lineare normale definito nel Paragrafo 2.2, con

$$X = [X_1 \ X_2] \quad e \quad \beta = [\beta_1^T \ \beta_2^T]^T,$$

dove  $X_1$  e  $X_2$  sono matrici di dimensione rispettivamente  $n \times d$  e  $n \times (p-d)$ ,  $\beta_1$  e  $\beta_2$  sono vettori di dimensione rispettivamente  $d \times 1$  e  $(p-d) \times 1$ , tali per

cui vale  $X\beta = X_1\beta_1 + X_2\beta_2$ . Si supponga inoltre di voler verificare il sistema d'ipotesi lineari (2.3). In questo caso specifico, su cui si basano le analisi svolte nell'intera tesi, il test rapporto di verosimiglianza è:

$$\delta = \frac{L(\widehat{\theta}_\psi)}{L(\widehat{\theta})} = \left\{ \frac{\|(I_n - P_2)y\|^2}{\|(I_n - P)y\|^2} \right\}^{-n/2}$$

con  $P = X(X^T X)^{-1}X^T$  e  $P_2 = X_2(X_2^T X_2)^{-1}X_2^T$  matrici di proiezione rispettivamente del modello completo e ridotto sotto l'ipotesi nulla. Solitamente si utilizza la seguente trasformata monotona decrescente del test rapporto di verosimiglianza

$$F = \delta^* \frac{n-p}{d} = \frac{\|(P - P_2)y\|^2/d}{\|(I_n - P)y\|^2/(n-p)}$$

con  $\delta^* = \delta^{-2/n} - 1$ . Il test  $F$  ha distribuzione esatta  $F$  di Fisher con  $d$  e  $n-p$  gradi di libertà, ossia  $F_{d,n-p}$ , sotto l'ipotesi nulla. Infatti, sia sotto l'ipotesi nulla  $H_0$  sia sotto quella alternativa  $H_1$  si ha che

$$\|(I_n - P)y\|^2 \sim \sigma^2 \chi_{n-p}^2,$$

mentre solo sotto l'ipotesi nulla si ha che

$$\|(P - P_2)y\|^2 \sim \sigma^2 \chi_d^2;$$

inoltre  $(I_n - P)$  e  $(P - P_2)$  sono matrici simmetriche, idempotenti, di rango rispettivamente  $n-p$  e  $d$  e si ha che  $(I_n - P)(P - P_2) = P - P_2 - P + PP_2 = \mathbf{0}$ . Quindi risulta che  $\|(I_n - P)y\|^2$  e  $\|(P - P_2)y\|^2$  sono variabili casuali indipendenti.

Si vuole dunque ottenere l'espressione del test  $F$  che consente di verificare i vincoli lineari imposti dal sistema d'ipotesi in questione, nello specifico la nullità di un insieme di componenti del vettore  $\beta$ .

Si consideri la ben nota relazione

$$\|(I_n - P_2)y\|^2 = \|(I_n - P)y\|^2 + \|(P - P_2)y\|^2.$$

Le componenti nello specifico sono

$$\begin{aligned} \|(I_n - P)y\|^2 &= y^T (I_n - P)y = n\widehat{\sigma}^2, \\ \|(I_n - P_2)y\|^2 &= y^T (I_n - P_2)y = n\widehat{\sigma}_\psi^2, \end{aligned}$$

da cui  $\|(P - P_2)y\|^2 = n\widehat{\sigma}_\psi^2 - n\widehat{\sigma}^2$ , dove  $\widehat{\sigma}^2$  e  $\widehat{\sigma}_\psi^2$  sono le stime di massima verosimiglianza globale e vincolata di  $\sigma^2$ . Il test  $F$  risulta quindi essere

$$F = \frac{(n\widehat{\sigma}_\psi^2 - n\widehat{\sigma}^2)/d}{n\widehat{\sigma}^2/(n-p)} = \frac{(\widehat{\sigma}_\psi^2 - \widehat{\sigma}^2)/d}{\widehat{\sigma}^2/(n-p)}. \quad (2.5)$$

La statistica  $F$  esprime la diminuzione relativa nella varianza stimata dell'errore che si ha passando dal modello ridotto a quello completo, corretta per i gradi di libertà. Una grande diminuzione indica che il modello ridotto non è appropriato per rappresentare la relazione tra la variabile risposta e le variabili esplicative (Pace e Salvan, 2001, Cap.9).

## 2.4 Test direzionale per modelli di regressione

Riprendendo la procedura descritta nel Paragrafo 1.3, si vuole ottenere l'espressione del test direzionale per il sistema d'ipotesi (2.3).

La funzione di log-verosimiglianza relativa al modello di regressione lineare normale è la (2.2), con vettore di parametri canonici  $\varphi(\theta) = \{\beta^T/\sigma^2, -1/(2\sigma^2)\}^T$  e vettore delle statistiche sufficienti ad esso associate  $u = (y^T X, y^T y)^T$ . Si centri quest'ultimo nei punti osservati  $u_{\text{oss}} = u(y_{\text{oss}})$ , definendo

$$s = u - u_{\text{oss}} = \begin{pmatrix} X^T y - X^T y_{\text{oss}} \\ y^T y - y_{\text{oss}}^T y_{\text{oss}} \end{pmatrix}$$

tale per cui se  $y = y_{\text{oss}}$  si ha  $s = s_{\text{oss}} = u_{\text{oss}} - u_{\text{oss}} = \mathbf{0}$ . Riscrivendo la (2.2), si ottiene

$$\begin{aligned} l(\theta; s) &= \varphi(\theta)^T s + l_{\text{oss}}(\theta) = \varphi(\theta)^T (u - u_{\text{oss}}) + l(\theta; u_{\text{oss}}) \\ &= \varphi(\theta)^T s + \varphi(\theta)^T u_{\text{oss}} - K\{\varphi(\theta)\}, \end{aligned} \quad (2.6)$$

dove  $K\{\varphi(\theta)\} = \frac{n}{2} \log \sigma^2 + \frac{1}{2\sigma^2} \beta^T X^T X \beta$  è la *funzione generatrice dei cumulanti*. Massimizzando la (2.6), si ottiene la stima di massima verosimiglianza per  $\theta$ ,

$$\widehat{\theta} = (\widehat{\beta}^T, \widehat{\sigma}^2)^T,$$

in cui  $\widehat{\beta} = (X^T X)^{-1} X^T y$  e  $\widehat{\sigma}^2 = (y - X\widehat{\beta})^T (y - X\widehat{\beta})/n = y^T (I_n - P)y/n$ , con  $P$  matrice di proiezione nel sottospazio vettoriale di  $\mathbb{R}^n$  generato dalle

colonne di  $X$ . Ne deriva che la stima di massima verosimiglianza per  $\varphi$  è  $\widehat{\varphi} = \varphi(\widehat{\theta}) = \{\widehat{\beta}^T/\widehat{\sigma}^2, -1/(2\widehat{\sigma}^2)\}^T$ .

Sotto l'ipotesi nulla in (2.3), si ottiene la seguente stima di massima verosimiglianza vincolata per  $\theta$ :

$$\widehat{\theta}_\psi = (\widehat{\beta}_\psi^T, \widehat{\sigma}_\psi^2)^T = (\mathbf{0}, \widehat{\beta}_{2\psi}^T, \widehat{\sigma}_\psi^2)^T,$$

con  $\widehat{\beta}_{2\psi} = (X_2^T X_2)^{-1} X_2^T y$  e  $\widehat{\sigma}_\psi^2 = (y - X_2 \widehat{\beta}_{2\psi})^T (y - X_2 \widehat{\beta}_{2\psi}) / n = y^T (I_n - P_2) y / n$ , dove  $P_2$  è la matrice di proiezione nel sottospazio vettoriale di  $\mathbb{R}^n$  generato dalle colonne di  $X_2$ . Analogamente a prima, la stima di massima verosimiglianza vincolata per  $\varphi$  è  $\widehat{\varphi}_\psi = \varphi(\widehat{\theta}_\psi) = (\mathbf{0}, \widehat{\varphi}_{2\psi}^T, \widehat{\varphi}_{\sigma\psi})^T = \{\mathbf{0}, \widehat{\beta}_{2\psi}^T / \widehat{\sigma}_\psi^2, -1/(2\widehat{\sigma}_\psi^2)\}^T$ . Per il calcolo del  $p$ -value direzionale si necessita della densità in (1.7); si ottiene quindi innanzitutto la quantità  $s_\psi$ , partendo dalla log-verosimiglianza

$$\begin{aligned} l_{\text{oss}}\{\theta(\varphi)\} &= \varphi^T s_{\text{oss}} + \varphi_{\beta_1}^T X_1^T y_{\text{oss}} + \varphi_{\beta_2}^T X_2^T y_{\text{oss}} + \varphi_{\sigma^2} y_{\text{oss}}^T y_{\text{oss}} + \frac{n}{2} \ln(-2\varphi_{\sigma^2}) + \\ &\quad + \frac{1}{4\varphi_{\sigma^2}} (\varphi_{\beta_1}^T X_1^T X_1 \varphi_{\beta_1} + 2\varphi_{\beta_1}^T X_1^T X_2 \varphi_{\beta_2} + \varphi_{\beta_2}^T X_2^T X_2 \varphi_{\beta_2}), \end{aligned}$$

dove  $\varphi = (\varphi_{\beta_1}^T, \varphi_{\beta_2}^T, \varphi_{\sigma^2})^T = \{\beta_1^T/\sigma^2, \beta_2^T/\sigma^2, -1/(2\sigma^2)\}^T$  e  $s_{\text{oss}} = \mathbf{0}$  per costruzione. Derivando nelle tre componenti di  $\varphi$  e calcolando l'opposto della derivata in  $\widehat{\varphi}_\psi$ , si ottiene  $s_\psi = (s_{1\psi}^T, s_{2\psi}^T, s_{3\psi}^T)^T$ , in cui

$$\begin{aligned} s_{1\psi} &= -X_1^T (I_n - P_2) y_{\text{oss}}, \\ s_{2\psi} &= -X_2^T y_{\text{oss}} + X_2^T y_{\text{oss}} = \mathbf{0}, \\ s_{3\psi} &= -y_{\text{oss}}^T y_{\text{oss}} + y_{\text{oss}}^T (I_n - P_2) y_{\text{oss}} + y_{\text{oss}}^T P_2 y_{\text{oss}} = 0. \end{aligned}$$

Quindi si ricava

$$s(t) = (1-t)s_\psi = (1-t)\{-y_{\text{oss}}^T (I_n - P_2) X_1, \mathbf{0}\}^T. \quad (2.7)$$

La funzione di log-verosimiglianza espressa in funzione di  $s(t)$  nella parametrizzazione  $\varphi$  diventa

$$l\{\varphi; s(t)\} = \varphi_\beta^T s_\beta(t) + \varphi_\beta^T X^T y_{\text{oss}} + \varphi_{\sigma^2} y_{\text{oss}}^T y_{\text{oss}} + \frac{n}{2} \ln(-2\varphi_{\sigma^2}) + \frac{1}{4\varphi_{\sigma^2}} \varphi_\beta^T X^T X \varphi_\beta, \quad (2.8)$$

con  $s_\beta(t) = \{s_1^T(t), s_2^T(t)\}^T$  e  $\varphi_\beta = (\varphi_{\beta_1}^T, \varphi_{\beta_2}^T)^T$ , ovvero i vettori  $s(t)$  e  $\varphi$  senza l'ultima componente relativa al parametro  $\sigma^2$ . Massimizzando la (2.8)

si ottiene la stima di massima verosimiglianza  $\hat{\varphi}(t) = \{\hat{\varphi}_\beta^T(t), \hat{\varphi}_{\sigma^2}(t)\}^T$ :

$$\hat{\varphi}_\beta(t) = -\frac{n (X^T X)^{-1} \{s_\beta(t) + X^T y_{\text{oss}}\}}{s_\beta^T(t)(X^T X)^{-1} s_\beta(t) + 2s_\beta^T(t)(X^T X)^{-1} X^T y_{\text{oss}} - y_{\text{oss}}^T (I_n - P) y_{\text{oss}}},$$

$$\hat{\varphi}_{\sigma^2}(t) = \frac{n}{2} \frac{1}{s_\beta^T(t)(X^T X)^{-1} s_\beta(t) + 2s_\beta^T(t)(X^T X)^{-1} X^T y_{\text{oss}} - y_{\text{oss}}^T (I_n - P) y_{\text{oss}}}.$$

Da  $\hat{\varphi}(t)$  si ricava  $\hat{\theta}(t) = \{\hat{\beta}^T(t), \hat{\sigma}^2(t)\}^T$ , con

$$\hat{\beta}(t) = \hat{\beta} + (X^T X)^{-1} s_\beta(t),$$

$$\hat{\sigma}^2(t) = \hat{\sigma}^2 - \frac{1}{n} \{s_\beta^T(t)(X^T X)^{-1} s_\beta(t) + 2s_\beta^T(t)\hat{\beta}\}.$$

È necessario porre un vincolo per  $\hat{\sigma}^2(t)$  in modo che sia sempre positiva

$$\begin{aligned} n \hat{\sigma}^2(t) &= n \hat{\sigma}^2 - s_\beta^T(t)(X^T X)^{-1} s_\beta(t) - 2s_\beta^T(t)\hat{\beta} > 0 \\ &= n \hat{\sigma}^2 - (1-t)^2 s_{\beta\psi}^T (X^T X)^{-1} s_{\beta\psi} - 2(1-t) s_{\beta\psi}^T \hat{\beta} > 0, \end{aligned}$$

con  $s_{\beta\psi} = (s_{1\psi}^T, s_{2\psi}^T)^T$ . Analizzando la parabola che ne deriva, risulta che  $\hat{\sigma}^2(t) > 0$  per  $t < t_u$ , dove

$$t_u = 1 - \{s_{\beta\psi}^T (X^T X)^{-1} s_{\beta\psi}\}^{-1} [-s_{\beta\psi}^T \hat{\beta} - \{\hat{\beta}^T s_{\beta\psi} s_{\beta\psi}^T \hat{\beta} + n \hat{\sigma}^2 s_{\beta\psi}^T (X^T X)^{-1} s_{\beta\psi}\}^{\frac{1}{2}}].$$

Si rimanda in Appendice A la dimostrazione che  $\hat{\theta}(0) = \hat{\theta}_\psi$ , mentre  $\hat{\theta}(1) = \hat{\theta}$ .

Nel seguito si ottengono le quantità che compongono la (1.7). In particolare, l'argomento dell'esponenziale è la differenza tra la (2.8) calcolata in  $\hat{\varphi}_\psi$

e la stessa calcolata in  $\widehat{\varphi}(t)$ . Nello specifico si ricava:

$$\begin{aligned}
l\{\widehat{\varphi}_\psi; s(t)\} &= \widehat{\varphi}_{2\psi}^T X_2^T y_{\text{oss}} + \widehat{\varphi}_{\sigma\psi} y_{\text{oss}}^T y_{\text{oss}} + \frac{n}{2} \ln(-2\widehat{\varphi}_{\sigma\psi}) + \frac{1}{4\widehat{\varphi}_{\sigma\psi}} \widehat{\varphi}_{2\psi}^T X_2^T X_2 \widehat{\varphi}_{2\psi} \\
&= \frac{\widehat{\beta}_{2\psi}^T}{\widehat{\sigma}_\psi^2} X_2^T y_{\text{oss}} - \frac{1}{2\widehat{\sigma}_\psi^2} y_{\text{oss}}^T y_{\text{oss}} - \frac{n}{2} \ln \widehat{\sigma}_\psi^2 - \frac{1}{2\widehat{\sigma}_\psi^2} \widehat{\beta}_{2\psi}^T X_2^T X_2 \widehat{\beta}_{2\psi} \\
&= -\frac{n}{2} - \frac{n}{2} \ln \widehat{\sigma}_\psi^2, \\
l\{\widehat{\varphi}(t); s(t)\} &= \widehat{\varphi}_\beta^T(t) \{s_\beta(t) + X^T y_{\text{oss}}\} + \widehat{\varphi}_{\sigma^2}(t) y_{\text{oss}}^T y_{\text{oss}} + \frac{n}{2} \ln\{-2\widehat{\varphi}_{\sigma^2}(t)\} \\
&\quad + \frac{1}{4\widehat{\varphi}_{\sigma^2}(t)} \widehat{\varphi}_\beta^T(t) X^T X \widehat{\varphi}_\beta(t) \\
&= \frac{\widehat{\beta}^T(t)}{\widehat{\sigma}^2(t)} \{s_\beta(t) + X^T y_{\text{oss}}\} - \frac{1}{2\widehat{\sigma}^2(t)} y_{\text{oss}}^T y_{\text{oss}} - \frac{n}{2} \ln \widehat{\sigma}^2(t) + \\
&\quad - \frac{1}{2\widehat{\sigma}^2(t)} \widehat{\beta}^T(t) X^T X \widehat{\beta}(t) \\
&= -\frac{n}{2} - \frac{n}{2} \ln \widehat{\sigma}^2(t),
\end{aligned}$$

da cui  $l\{\widehat{\varphi}_\psi; s(t)\} - l\{\widehat{\varphi}(t); s(t)\} = (n/2) \ln\{\widehat{\sigma}^2(t)/\widehat{\sigma}_\psi^2\}$ .

La matrice di informazione osservata nella parametrizzazione  $\varphi$  calcolata in  $\widehat{\varphi}(t)$ ,  $j_{\varphi\varphi}\{\widehat{\varphi}(t)\}$ , si ottiene derivando opportunamente la (2.8) e risulta pari a

$$j_{\varphi\varphi}\{\widehat{\varphi}(t)\} = \widehat{\sigma}^2(t) \begin{pmatrix} X^T X & 2\{s_\beta(t) + X^T y_{\text{oss}}\} \\ 2\{s_\beta^T(t) + y_{\text{oss}}^T X\} & -2n \widehat{\sigma}^2(t) + 4 y_{\text{oss}}^T y_{\text{oss}} \end{pmatrix}.$$

Per calcolare il determinante della matrice  $j_{\varphi\varphi}\{\widehat{\varphi}(t)\}$  si fa riferimento al seguente noto teorema (Gregorio e Salce, 2010, Cap.4):

**Teorema 1** Sia  $B = \begin{pmatrix} A & b \\ c^T & a \end{pmatrix}$  matrice a blocchi, in cui  $A$  è una matrice non singolare,  $b$  un vettore colonna,  $c^T$  un vettore riga e  $a$  uno scalare. Allora il determinante di  $B$  risulta

$$|B| = |A|(a - c^T A^{-1}b).$$

Si ha quindi

$$\begin{aligned}
|j_{\varphi\varphi}\{\widehat{\varphi}(t)\}| &= \{\widehat{\sigma}^2(t)\}^{p+1} [ |X^T X| \{ -2n\widehat{\sigma}^2(t) + 4y_{\text{oss}}^T y_{\text{oss}} - 4\{s_{\beta}^T(t) + y_{\text{oss}}^T X\} \\
&\quad (X^T X)^{-1} \{s_{\beta}(t) + X^T y_{\text{oss}}\} \} ] \\
&= \{\widehat{\sigma}^2(t)\}^{p+1} |X^T X| \{ -2n\widehat{\sigma}^2(t) + 4y_{\text{oss}}^T (I_n - P)y_{\text{oss}} + \\
&\quad - 4s_{\beta}^T(t)(X^T X)^{-1}s_{\beta}(t) - 8s_{\beta}^T(t)\widehat{\beta} \} \\
&= \{\widehat{\sigma}^2(t)\}^{p+1} |X^T X| \{ -2n\widehat{\sigma}^2(t) + 4n\widehat{\sigma}^2(t) \} \\
&= 2n |X^T X| \{\widehat{\sigma}^2(t)\}^{p+2}.
\end{aligned}$$

Infine si ricava  $j_{(\lambda\lambda)}(\widehat{\varphi}_{\psi})$ , la matrice di informazione osservata ottenuta derivando per il vettore dei parametri di disturbo  $\lambda$  nella parametrizzazione  $\varphi$ , calcolata in  $\widehat{\varphi}_{\psi}$ . Nello specifico:

$$|j_{(\lambda\lambda)}(\widehat{\varphi}_{\psi})| = |j_{\lambda\lambda}\{\widehat{\varphi}_{\psi}; s(t)\}| \left| \frac{\partial \varphi(\theta)}{\partial \lambda} \right|_{\widehat{\theta}_{\psi}}^{-2} \quad (2.9)$$

ma poiché la seconda componente non dipende da  $t$ , non è influente nel calcolo dell'integrale del  $p$ -value direzionale e quindi non viene calcolata. Ci si sofferma dunque sul primo fattore del lato di destra della (2.9): la matrice di informazione osservata viene ricavata nella parametrizzazione  $\theta$  e calcolata poi in  $\widehat{\theta}_{\psi}$ . A partire dalla funzione di log-verosimiglianza (2.8) espressa in funzione di  $\theta$ ,

$$\begin{aligned}
l\{\theta; s(t)\} &= \frac{\beta_1^T}{\sigma^2} \{s_1(t) + X_1^T y_{\text{oss}}\} + \frac{\beta_2^T}{\sigma^2} \{s_2(t) + X_2^T y_{\text{oss}}\} - \frac{1}{2\sigma^2} \{s_3(t) + y_{\text{oss}}^T y_{\text{oss}}\} + \\
&\quad - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\beta_1^T X_1^T X_1 \beta_1 + 2\beta_1^T X_1^T X_2 \beta_2 + \beta_2^T X_2^T X_2 \beta_2),
\end{aligned}$$

si ottengono le derivate prime rispetto a  $\lambda$ :

$$\begin{aligned}
\frac{\partial}{\partial \beta_2} l\{\theta; s(t)\} &= \frac{1}{\sigma^2} \{s_2(t) + X_2^T y_{\text{oss}} - X_2^T X_1 \beta_1 - X_2^T X_2 \beta_2\}, \\
\frac{\partial}{\partial \sigma^2} l\{\theta; s(t)\} &= -\frac{1}{\sigma^4} [\beta_1^T \{s_1(t) + X_1^T y_{\text{oss}}\} + \beta_2^T \{s_2(t) + X_2^T y_{\text{oss}}\}] + \frac{1}{2\sigma^4} \{s_3(t) + \\
&\quad + y_{\text{oss}}^T y_{\text{oss}}\} - \frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (\beta_1^T X_1^T X_1 \beta_1 + 2\beta_1^T X_1^T X_2 \beta_2 + \beta_2^T X_2^T X_2 \beta_2),
\end{aligned}$$



da cui si ricavano le derivate seconde:

$$\begin{aligned}\frac{\partial^2}{\partial\beta_2\partial\beta_2^T}l\{\theta; s(t)\} &= -\frac{1}{\sigma^2}X_2^T X_2, \\ \frac{\partial^2}{\partial\beta_2\partial\sigma^2}l\{\theta; s(t)\} &= -\frac{1}{\sigma^4}\{s_2(t) + X_2^T y_{\text{oss}} - X_2^T X_1\beta_1 - X_2^T X_2\beta_2\}, \\ \frac{\partial^2}{\partial\sigma^4}l\{\theta; s(t)\} &= \frac{2}{\sigma^6}[\beta_1^T\{s_1(t) + X_1^T y_{\text{oss}}\} + \beta_2^T\{s_2(t) + X_2^T y_{\text{oss}}\} - \frac{1}{2}\{s_3(t) + \\ &\quad + y_{\text{oss}}^T y_{\text{oss}}\} - \frac{1}{2}\beta^T X^T X\beta] + \frac{n}{2\sigma^4}.\end{aligned}$$

Calcolando l'opposto delle derivate seconde in  $\hat{\theta}_\psi$ , si ottiene

$$j_{\lambda\lambda}(\hat{\theta}_\psi) = \frac{1}{\hat{\sigma}_\psi^2} \begin{pmatrix} X_2^T X_2 & 0 \\ 0 & \frac{n}{2\hat{\sigma}_\psi^2} \end{pmatrix}.$$

Come si può notare neanche la prima componente della (2.9) dipende da  $t$ , quindi l'intera  $j_{(\lambda\lambda)}(\hat{\varphi}_\psi)$  non influirà nel calcolo del  $p$ -value direzionale. Tale risultato era atteso e deriva dal fatto che l'ipotesi nulla in (2.1), e quindi anche in (2.3), è una funzione lineare dei parametri canonici. In questo caso, per famiglie esponenziali lineari, esiste una distribuzione condizionata ben definita libera dai parametri di disturbo, che può essere usata per fare inferenza sul parametro d'interesse. Nello specifico, la funzione di log-verosimiglianza (1.3) nella parametrizzazione  $\varphi$  può essere scritta come

$$l(\varphi; s) = \psi^T v_1 + \lambda^T v_2 + l_{\text{oss}}(\psi, \lambda),$$

e la distribuzione condizionata di  $v_1$ , dato  $v_2$ , dipende unicamente dal parametro di interesse  $\psi$ :

$$f(v_1|v_2; \psi) = \exp\{\psi^T v_1 - \tilde{K}(\psi)\}\tilde{h}(v_1),$$

con  $\tilde{K}(\cdot)$  e  $\tilde{h}(\cdot)$  funzioni non esplicite. Perciò nella seconda fase di costruzione del test non è necessario condizionarsi ad una statistica ancillare e quindi calcolare il fattore di correzione  $j_{(\lambda\lambda)}(\hat{\varphi}_\psi)$ . Questo è coerente con la teoria generale in Fraser et al. (2016) e Davison et al. (2014) a cui si rimanda per ulteriori approfondimenti. È da sottolineare il fatto che ciò appena detto vale anche nel caso generale del sistema d'ipotesi (2.1), dato che l'ipotesi nulla è

funzione dei parametri canonici. Tale risultato è vantaggioso poiché consente di verificare qualsiasi vincolo lineare nel modello di regressione tramite una formulazione semplificata del test direzionale.

Infine, mettendo assieme tutti gli ingredienti ottenuti in precedenza si ottiene la densità (1.7)

$$\begin{aligned} h\{s(t); \psi\} &\propto \left\{ \frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_\psi^2} \right\}^{\frac{n}{2}} [2n|X^T X| \{\widehat{\sigma}^2(t)\}^{p+2}]^{-1/2} \\ &\propto \left\{ \frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_\psi^2} \right\}^{\frac{n}{2}} \{\widehat{\sigma}^2(t)\}^{-(p/2+1)}, \end{aligned}$$

in cui sono state tralasciate le quantità che non dipendono da  $t$ . Il  $p$ -value direzionale che ne deriva è

$$p(\psi) = \frac{\int_1^{t_u} t^{d-1} \left\{ \frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_\psi^2} \right\}^{\frac{n-p}{2}-1} dt}{\int_0^{t_u} t^{d-1} \left\{ \frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_\psi^2} \right\}^{\frac{n-p}{2}-1} dt}. \quad (2.10)$$

## 2.5 Test $w^*$ di Skovgaard per modelli di regressione

Nel caso di modelli di regressione lineari normali è possibile ottenere una forma esplicita della statistica test  $w^*$  di Skovgaard, in cui il fattore di correzione  $\gamma(\psi)$  è

$$\gamma(\psi) = - \frac{(s_\psi^T \widetilde{\Sigma}^{-1} s_\psi)^{\frac{d}{2}}}{\{n \ln(\widehat{\sigma}_\psi^2 / \widehat{\sigma}^2)\}^{\frac{d}{2}-1} (\widehat{\varphi} - \widehat{\varphi}_\psi)^T s_\psi} \left( \frac{\widehat{\sigma}_\psi^2}{\widehat{\sigma}^2} \right)^{\frac{p}{2}+1}.$$

Tale fattore è composto dalle quantità descritte in (1.10), delle quali si riporta l'espressione qui in seguito.

Il vettore  $v$  è il valore atteso del vettore della statistica sufficiente canonica  $u$

$$\begin{aligned} v = E(u) = K'\{\varphi\} &= \frac{\partial}{\partial \varphi} \left\{ -\frac{n}{2} \ln(-2\varphi_{\sigma^2}) - \frac{1}{4\varphi_{\sigma^2}} \varphi_\beta^T X^T X \varphi_\beta \right\} \\ &= \begin{pmatrix} -\frac{1}{2\varphi_{\sigma^2}} X^T X \varphi_\beta \\ -\frac{n}{2\varphi_{\sigma^2}} + \frac{1}{4\varphi_{\sigma^2}} \varphi_\beta^T X^T X \varphi_\beta \end{pmatrix}, \end{aligned}$$

dove  $K\{\varphi\}$  è la funzione dei cumulanti descritta in (2.6) e le componenti del vettore finale sono la derivata rispetto al vettore  $\varphi_\beta = (\varphi_{\beta_1}^T, \varphi_{\beta_2}^T)^T$  e rispetto a  $\varphi_{\sigma^2}$ . Si ricava quindi  $\tilde{v}$ , ovvero  $v$  calcolato in  $\hat{\varphi}_\psi$ :

$$\tilde{v} = \begin{pmatrix} X^T X \hat{\beta}_\psi \\ n\hat{\sigma}_\psi^2 + \hat{\beta}_\psi^T X^T X \hat{\beta}_\psi \end{pmatrix} = \begin{pmatrix} X_1^T P_2 y_{\text{oss}} \\ X_2^T y_{\text{oss}} \\ n\hat{\sigma}_\psi^2 + y_{\text{oss}}^T P_2 y_{\text{oss}} \end{pmatrix}.$$

Si ha inoltre

$$u - \tilde{v} = \begin{pmatrix} X_1^T y_{\text{oss}} \\ X_2^T y_{\text{oss}} \\ y_{\text{oss}}^T y_{\text{oss}} \end{pmatrix} - \begin{pmatrix} X_1^T P_2 y_{\text{oss}} \\ X_2^T y_{\text{oss}} \\ n\hat{\sigma}_\psi^2 + y_{\text{oss}}^T P_2 y_{\text{oss}} \end{pmatrix} = \begin{pmatrix} X_1^T (I_n - P_2) y_{\text{oss}} \\ \mathbf{0} \\ 0 \end{pmatrix} = -s_\psi.$$

La matrice di covarianza  $\Sigma$  del vettore della statistica sufficiente canonica  $u$  è

$$\Sigma = \text{Var}(u) = K''\{\varphi\} = \begin{pmatrix} -\frac{1}{2\varphi_{\sigma^2}} X^T X & \frac{1}{2\varphi_{\sigma^2}^2} X^T X \varphi_\beta \\ \frac{1}{2\varphi_{\sigma^2}^2} \varphi_\beta^T X^T X & \frac{n}{2\varphi_{\sigma^2}^2} - \frac{1}{2\varphi_{\sigma^2}^3} \varphi_\beta^T X^T X \varphi_\beta \end{pmatrix}.$$

Se  $\Sigma$  viene calcolata nella stima di massima verosimiglianza vincolata  $\hat{\varphi}_\psi$  e in quella globale  $\hat{\varphi}$ , si ottiene rispettivamente

$$\begin{aligned} \tilde{\Sigma} &= \hat{\sigma}_\psi^2 \begin{pmatrix} X^T X & 2X^T X \hat{\beta}_\psi \\ 2\hat{\beta}_\psi^T X^T X & 2n\hat{\sigma}_\psi^2 + 4y_{\text{oss}}^T P_2 y_{\text{oss}} \end{pmatrix}, \\ \hat{\Sigma} &= \hat{\sigma}^2 \begin{pmatrix} X^T X & 2X^T y_{\text{oss}} \\ 2y_{\text{oss}}^T X & 2n\hat{\sigma}^2 + 4y_{\text{oss}}^T P y_{\text{oss}} \end{pmatrix}. \end{aligned}$$

Facendo riferimento al Teorema 1, si ricavano i determinanti

$$\begin{aligned} |\tilde{\Sigma}| &= (\hat{\sigma}_\psi^2)^{p+1} |X^T X| \{2n\hat{\sigma}_\psi^2 + 4y_{\text{oss}}^T P_2 y_{\text{oss}} - 4\hat{\beta}_\psi^T X^T X (X^T X)^{-1} X^T X \hat{\beta}_\psi\} \\ &= (\hat{\sigma}_\psi^2)^{p+1} |X^T X| (2n\hat{\sigma}_\psi^2 + 4y_{\text{oss}}^T P_2 y_{\text{oss}} - 4y_{\text{oss}}^T P_2 y_{\text{oss}}) \\ &= 2n(\hat{\sigma}_\psi^2)^{p+2} |X^T X|, \\ |\hat{\Sigma}| &= 2n(\hat{\sigma}^2)^{p+2} |X^T X|. \end{aligned}$$

Infine, dalla funzione di log-verosimiglianza  $l(\theta; s)$ , si ottengono le matrici di informazione osservata  $j_{\lambda\lambda}$  e attesa  $i_{\lambda\lambda}$  che, calcolate nella stima di massima verosimiglianza vincolata  $\hat{\varphi}_\psi$ , risultano essere uguali e pari a

$$i_{\lambda\lambda}(\hat{\varphi}_\psi; s) = j_{\lambda\lambda}(\hat{\varphi}_\psi; s) = \frac{1}{\hat{\sigma}_\psi^2} \begin{pmatrix} X_2^T X_2 & 0 \\ 0 & \frac{n}{2\hat{\sigma}_\psi^2} \end{pmatrix},$$

da cui si ha  $|j_{\lambda\lambda}(\widehat{\varphi}_\psi)| = |i_{\lambda\lambda}(\widehat{\varphi}_\psi)|$ .

Quindi, a partire dal test log-rapporto di verosimiglianza

$$w(\psi) = 2\{l(\widehat{\theta}) - l(\widehat{\theta}_\psi)\} = n \ln(\widehat{\sigma}_\psi^2/\widehat{\sigma}^2),$$

tramite il fattore di correzione  $\gamma(\psi)$  è possibile ottenere la statistica test  $w^*(\psi)$  in (1.9). Essa si distribuisce asintoticamente come un  $\chi_d^2$  sotto l'ipotesi nulla. Skovgaard (2001) mostra che la statistica test  $w^*(\psi)$  fornisce un'eccellente approssimazione nel caso di un parametro d'interesse unidimensionale; nel caso vettoriale Skovgaard sostiene che l'errore di approssimazione sia comunque basso, anche nelle code estreme della distribuzione. Tuttavia in alcuni casi limite, come ad esempio quando si ha un numero elevato di parametri d'interesse sotto l'ipotesi nulla, è stato mostrato che la statistica test  $w^*(\psi)$  non è una valida alternativa al test rapporto di verosimiglianza, poiché tende erroneamente a correggere i  $p$ -value forniti da  $w(\psi)$ ; si veda ad esempio Davison et al. (2014) e Fraser et al. (2016).

# Capitolo 3

## Confronto e risultati

### 3.1 Introduzione

In seguito viene presentato il confronto tra i  $p$ -value relativi ai tre test ottenuti nel capitolo precedente. Dallo studio effettuato emerge un risultato importante: il livello di significatività osservato del test direzionale si dimostra essere uguale a quello esatto, ottenuto dal test  $F$ . La dimostrazione completa viene riportata nell'Appendice B. Il fatto che il test direzionale porti alle stesse conclusioni del test  $F$ , che fornisce una soluzione esatta, conferma l'adeguatezza dell'approccio direzionale e la sua validità come test alternativo a quello basato sul rapporto di verosimiglianza in contesti più generali in cui una soluzione esatta non è disponibile. Inoltre, si mostra che l'uguaglianza tra i  $p$ -value vale anche per il sistema d'ipotesi lineari generale. Per quanto riguarda il test di Skovgaard, esso non fornisce una soluzione esatta; tuttavia l'approssimazione asintotica della distribuzione nulla del test è generalmente molto accurata, ad eccezione del caso in cui la dimensione del parametro d'interesse è molto elevata rispetto all'informazione campionaria.

### 3.2 Soluzione esatta del test direzionale

Dopo aver costruito i test e calcolato i relativi  $p$ -value, si procede con un confronto tra di essi. Il test di riferimento è ovviamente il test  $F$ , poichè

avendo una distribuzione nota sotto l'ipotesi nulla, fornisce una soluzione esatta. Tale soluzione è anche ottimale (Lehmann e Romano, 2005, Cap.7). Di conseguenza, più i risultati degli altri test si avvicinano a quello del test  $F$ , più sono ritenuti accurati. Ci si sofferma ora sul confronto tra il test direzionale e il test  $F$ . Dalle analisi si è giunti al seguente risultato:

**Teorema 2** *Si consideri un modello di regressione lineare normale  $Y \sim N_n(X\beta, \sigma^2 I_n)$ . Si vuole verificare la validità di alcuni vincoli lineari tra le componenti di  $\beta$  (2.1). Allora risulta che il test direzionale fornisce lo stesso  $p$ -value ottenuto con il test  $F$ .*

*Dimostrazione.*

Riprendendo la (2.10), il  $p$ -value direzionale è

$$p_D = \frac{\int_1^{t_u} t^{d-1} \left\{ \frac{\hat{\sigma}^2(t)}{\hat{\sigma}_\psi^2} \right\}^{\frac{n-p}{2}-1} dt}{\int_0^{t_u} t^{d-1} \left\{ \frac{\hat{\sigma}^2(t)}{\hat{\sigma}_\psi^2} \right\}^{\frac{n-p}{2}-1} dt},$$

mentre il  $p$ -value relativo al test  $F$  è

$$p_F = \int_{F_{\text{oss}}}^{+\infty} \frac{1}{B\left(\frac{d}{2}, \frac{n-p}{2}\right)} \frac{1}{t} \left\{ \frac{d^d (n-p)^{n-p} t^d}{(dt + n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dt,$$

dove all'interno dell'integrale c'è la funzione di densità di una  $F$  di Fisher con  $d$  e  $n-p$  gradi di libertà, e  $F_{\text{oss}} = \{(\hat{\sigma}_\psi^2 - \hat{\sigma}^2)/d\}/\{\hat{\sigma}^2/(n-p)\}$  è il valore osservato della statistica test  $F$  (2.5). Si tratta quindi di dimostrare che  $p_D = p_F$ .

Considerando  $p_D$ , si effettui la seguente sostituzione:

$$z = \frac{\{\hat{\sigma}_\psi^2 - \hat{\sigma}^2(t)\}/d}{\hat{\sigma}^2(t)/(n-p)}, \quad (3.1)$$

dove  $z$  è una funzione monotona crescente in  $t$ . Gli estremi dell'integrale diventano:

$$\begin{aligned} t = 0 &\rightarrow z = \frac{(\hat{\sigma}_\psi^2 - \hat{\sigma}_\psi^2)/d}{\hat{\sigma}_\psi^2/(n-p)} = 0, \\ t = 1 &\rightarrow z = \frac{(\hat{\sigma}_\psi^2 - \hat{\sigma}^2)/d}{\hat{\sigma}^2/(n-p)} = F_{\text{oss}}, \\ t = t_u &\rightarrow z = \frac{(\hat{\sigma}_\psi^2 - 0)/d}{0/(n-p)} = +\infty, \end{aligned}$$

dove l'ultimo risultato deriva dal fatto che  $t_u$  è una radice di  $\widehat{\sigma}^2(t)$ .

Dalla (3.1) si ricava inoltre la quantità

$$\frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_\psi^2} = \frac{n-p}{dz+n-p},$$

ed esplicitando  $t$  in funzione di  $z$  si ottiene:

$$t = \sqrt{\frac{dz}{dz+n-p} n\widehat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}}. \quad (3.2)$$

Dalla (3.2), derivando entrambe le parti dell'uguaglianza, si ha

$$dt = \frac{n\widehat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}}{2\sqrt{\frac{dz}{dz+n-p} n\widehat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}}} \frac{n-p}{(dz+n-p)^2} dz.$$

Applicando dunque la sostituzione (3.1) nell'integrale sia al numeratore che al denominatore di  $p_D$ , si ottiene:

$$\begin{aligned} p_D &= \frac{\int_{F_{\text{oss}}}^{+\infty} \frac{(dz)^{d/2-1} (n-p)^{(n-p)/2}}{(dz+n-p)^{(d+n-p)/2}} [n\widehat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}]^{d/2} dz}{\int_0^{+\infty} \frac{(dz)^{d/2-1} (n-p)^{(n-p)/2}}{(dz+n-p)^{(d+n-p)/2}} [n\widehat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}]^{d/2} dz} \\ &= \frac{\int_{F_{\text{oss}}}^{+\infty} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{(dz+n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dz}{\int_0^{+\infty} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{(dz+n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dz}, \end{aligned} \quad (3.3)$$

in cui tutte le quantità che non dipendono da  $z$  sono state portate fuori dall'integrale e semplificate. Si osservi che il denominatore è la funzione *Beta* di parametri  $d/2$  e  $(n-p)/2$  (Wikipedia, 2016). Nello specifico, raccogliendo opportunamente e applicando la sostituzione  $r = dz/(n-p)$ , si ha che

$$\begin{aligned} \int_0^{+\infty} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{(dz+n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dz &= \int_0^{+\infty} \frac{\left(\frac{dz}{n-p}\right)^{d/2-1}}{\left(\frac{dz}{n-p} + 1\right)^{(d+n-p)/2}} \left(\frac{d}{n-p}\right) dz = \\ \int_0^{+\infty} \frac{r^{d/2-1}}{(1+r)^{(d+n-p)/2}} dr &= B\left(\frac{d}{2}, \frac{n-p}{2}\right), \end{aligned}$$

dove  $B$  è la funzione *Beta*. Quindi, portando all'interno dell'integrale al numeratore in (3.3) la funzione *Beta* appena ottenuta, si conclude che

$$p_D = \int_{F_{\text{oss}}}^{+\infty} \frac{1}{B\left(\frac{d}{2}, \frac{n-p}{2}\right)} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{(dz + n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dz = p_F .$$

Si è così dimostrata l'uguaglianza dei  $p$ -value relativi al test  $F$  e al test direzionale. Per ulteriori approfondimenti sulla dimostrazione si rimanda all'Appendice B.

È da sottolineare che tale risultato si è ottenuto nel caso di un modello di regressione lineare normale in cui si è definito un sistema d'ipotesi lineari come in (2.3). Infatti, l'intera dimostrazione si basa sulle quantità ottenute nel capitolo precedente in cui si sono ricavate le statistiche test per verificare la nullità di alcune componenti del vettore  $\beta$ . Tuttavia, si è dimostrato che il risultato è valido anche nel caso generale in cui si considera il sistema d'ipotesi (2.1).

### 3.2.1 Generalizzazione del risultato

Si consideri il sistema d'ipotesi generale (2.1),

$$\begin{cases} H_0 : H\beta = h \\ H_1 : H\beta \neq h \end{cases}$$

in cui si vogliono verificare  $d$  vincoli lineari del vettore  $\beta$ . A partire dalla funzione di log-verosimiglianza (2.6), si ottengono le stime di massima verosimiglianza globale e vincolata di  $\theta$ , che sono rispettivamente

$$\hat{\theta} = (\hat{\beta}^T, \hat{\sigma}^2)^T \quad \text{e} \quad \hat{\theta}_H = (\hat{\beta}_H^T, \hat{\sigma}_H^2)^T ,$$

in cui

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad \text{e} \quad \hat{\beta}_H = \hat{\beta} - (X^T X)^{-1} H^T K (H\hat{\beta} - h)$$

sono le stime dei coefficienti di regressione, con  $K = \{H(X^T X)^{-1} H^T\}^{-1}$ , e

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} (y - X\hat{\beta})^T (y - X\hat{\beta}) , \\ \hat{\sigma}_H^2 &= \frac{1}{n} (y - X\hat{\beta}_H)^T (y - X\hat{\beta}_H) = \hat{\sigma}^2 + \frac{1}{n} (H\hat{\beta} - h)^T K (H\hat{\beta} - h) , \end{aligned}$$



sono le stime della varianza. Per ottenere il  $p$ -value direzionale, si deve calcolare la densità (1.7): essendo l'ipotesi lineare nei parametri canonici, anche in questo caso non è necessario ricavare il fattore  $|j_{(\lambda\lambda)}(\widehat{\varphi}_\psi)|$ . Seguendo la procedura riportata nel Paragrafo 2.4, si ha

$$s(t) = (1-t) \begin{pmatrix} -H^T K(H\widehat{\beta} - h) \\ -2(H\widehat{\beta} - h)^T K h \end{pmatrix};$$

massimizzando quindi la funzione di log-verosimiglianza espressa in funzione di  $s(t)$  si ottiene  $\widehat{\theta}(t) = \{\widehat{\beta}^T(t), \widehat{\sigma}^2(t)\}^T$ :

$$\begin{aligned} \widehat{\beta}(t) &= \widehat{\beta} - (1-t)\{(X^T X)^{-1} H^T K(H\widehat{\beta} - h)\}, \\ \widehat{\sigma}^2(t) &= \widehat{\sigma}^2 + \frac{1}{n}(1-t^2)\{(H\widehat{\beta} - h)^T K(H\widehat{\beta} - h)\}. \end{aligned}$$

Affinché  $\widehat{\sigma}^2(t)$  sia positivo, deve valere il seguente vincolo:

$$t < t_u = \sqrt{1 - n\widehat{\sigma}^2\{(H\widehat{\beta} - h)^T K(H\widehat{\beta} - h)\}}.$$

Infine risulta che le rimanenti quantità che compongono la (1.7) sono analoghe a quelle ottenute nel Paragrafo 2.4, per cui si ha che

$$\begin{aligned} h\{s(t); \psi\} &\propto \left\{ \frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_H^2} \right\}^{\frac{n}{2}} [2n|X^T X|\{\widehat{\sigma}^2(t)\}^{p+2}]^{-1/2} \\ &\propto \left\{ \frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_H^2} \right\}^{\frac{n}{2}} \{\widehat{\sigma}^2(t)\}^{-(p/2+1)}, \end{aligned}$$

in cui sono state tralasciate le quantità che non dipendono da  $t$ . Quindi il  $p$ -value relativo al test direzionale per il sistema d'ipotesi generale è

$$p_{D_H} = \frac{\int_1^{t_u} t^{d-1} \left\{ \frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_H^2} \right\}^{\frac{n-p}{2}-1} dt}{\int_0^{t_u} t^{d-1} \left\{ \frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_H^2} \right\}^{\frac{n-p}{2}-1} dt}.$$

Per quanto riguarda la statistica test  $F$  associata al sistema d'ipotesi (2.1), analogamente a quanto riportato nel Paragrafo 2.3, si ha che

$$F = \frac{(n\widehat{\sigma}_H^2 - n\widehat{\sigma}^2)/d}{n\widehat{\sigma}^2/(n-p)} = \frac{(\widehat{\sigma}_H^2 - \widehat{\sigma}^2)/d}{\widehat{\sigma}^2/(n-p)}, \quad (3.4)$$

con distribuzione nulla  $F_{d,n-p}$ . Il  $p$ -value relativo a tale test è

$$p_{F_H} = \int_{F_{\text{oss}}}^{+\infty} \frac{1}{B\left(\frac{d}{2}, \frac{n-p}{2}\right)} \frac{1}{t} \left\{ \frac{d^d (n-p)^{n-p} t^d}{(dt + n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dt,$$

in cui  $F_{\text{oss}}$  è il valore osservato della (3.4), ovvero

$$F_{\text{oss}} = \frac{(n\hat{\sigma}_H^2 - n\hat{\sigma}^2)/d}{n\hat{\sigma}^2/(n-p)} = \frac{\{(H\hat{\beta} - h)^T K(H\hat{\beta} - h)\}/d}{s^2}.$$

Applicando la sostituzione

$$z = \frac{\{\hat{\sigma}_H^2 - \hat{\sigma}^2(t)\}/d}{\hat{\sigma}^2(t)/(n-p)}$$

agli integrali di  $p_{D_H}$ , la dimostrazione è analoga alla precedente per il sistema d'ipotesi (2.3), per cui si conclude che  $p_{D_H} = p_{F_H}$ .

Quindi, utilizzare il test direzionale o il test  $F$  per verificare vincoli lineari di un gruppo di coefficienti di regressione, per un modello di regressione lineare normale, è essenzialmente identico, poiché portano al medesimo risultato. Il fatto che il test direzionale fornisca soluzioni esatte conferma la validità dell'approccio sottostante, basato sul condizionamento alla direzione osservata tra il modello nullo e il modello stimato.

### 3.3 Soluzione approssimata del test di Skovgaard

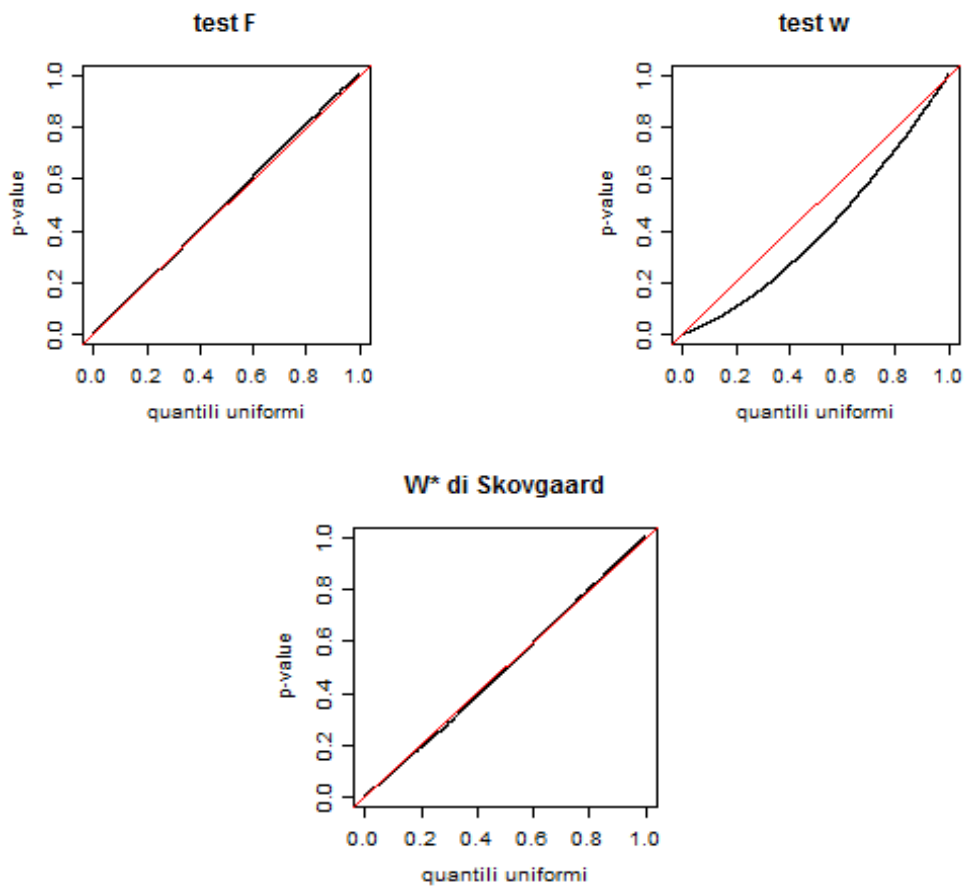
Infine si conclude con l'analisi dei risultati prodotti dal test  $w^*$  di Skovgaard. Al contrario del test direzionale, il test  $w^*$  non fornisce soluzioni esatte. Si è quindi verificata l'accuratezza dell'approssimazione asintotica della distribuzione nulla del test tramite degli studi di simulazione. Nello specifico, dal modello sotto l'ipotesi nulla si sono generati tanti campioni diversi, per ciascuno dei quali si è stimato il modello di regressione lineare sotto l'ipotesi nulla e quello sotto l'ipotesi alternativa, si sono ottenute le statistiche test e si sono calcolati i  $p$ -value relativi ad esse. Al termine della simulazione, si hanno dunque tanti  $p$ -value quanti sono i campioni generati. Tali  $p$ -value simulati sotto l'ipotesi nulla, se fossero esatti, si distribuirebbero uniformemente. Si confronta quindi la loro distribuzione empirica con quella uniforme

tramite dei grafici quantile-quantile: la linea bisettrice di colore rosso rappresenta i quantili di una distribuzione uniforme, mentre la linea nera indica la distribuzione dei  $p$ -value simulati. Se le due linee si sovrappongono, l'approssimazione fornita dal test è accurata; questo si verifica a meno dell'errore Montecarlo sempre quando le soluzioni del test sono esatte. Se la linea nera è sotto la linea rossa, significa che la statistica test tende a produrre  $p$ -value più piccoli, mentre viceversa se la linea nera è al di sopra, i  $p$ -value sono più grandi.

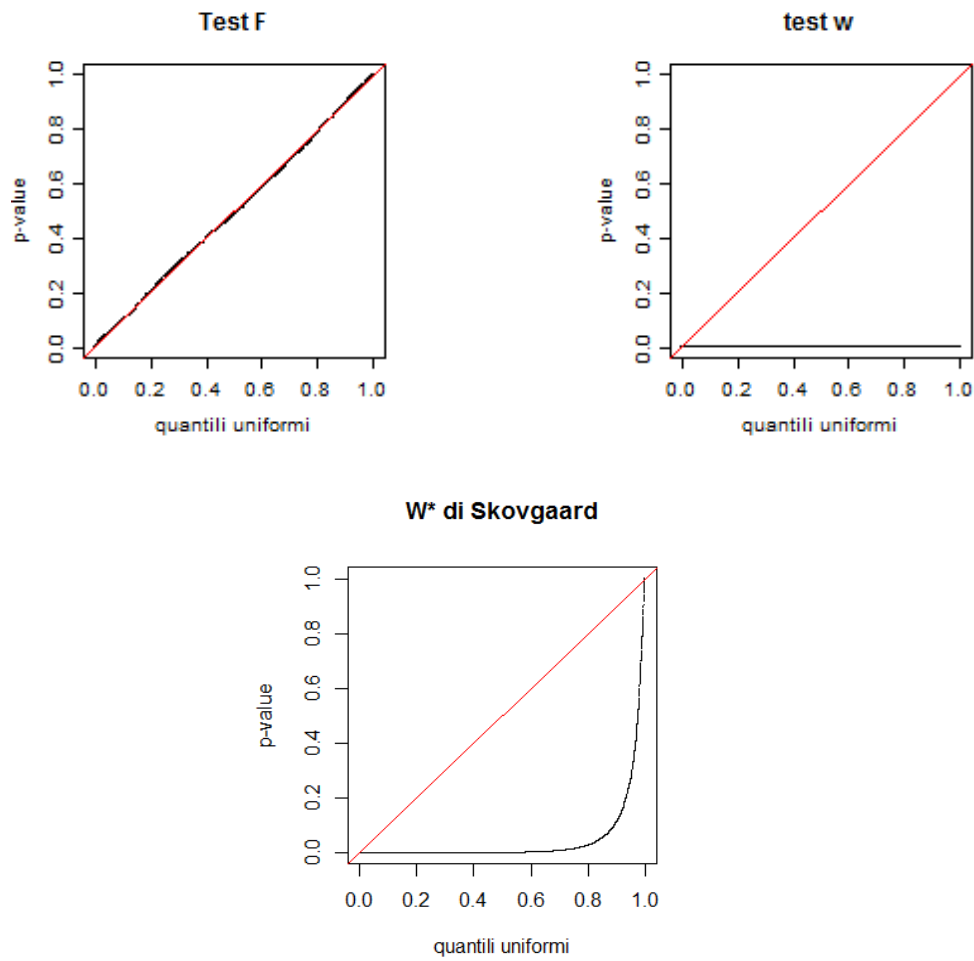
Si analizzano due casi distinti. Nel primo si sono effettuate 10000 simulazioni, per ciascuna delle quali si è generato un campione di dati di numerosità  $n = 100$  da una variabile casuale normale con media  $X\beta$  e varianza unitaria, con  $X$  di dimensione  $100 \times 20$  e  $\beta$  di dimensione  $20 \times 1$ . La matrice  $X$  è formata dall'intercetta e da 19 variabili esplicative generate uniformemente in  $(0,1)$ ; il vettore  $\beta$  è stato fissato pari ad un vettore di costanti comprese tra  $-5$  e  $+5$ , e nel modello ridotto si sono posti 8 coefficienti pari a zero. Si hanno quindi 8 parametri di interesse e 13 di disturbo, includendo anche  $\sigma^2$ . In Figura 3.1 si confrontano i  $p$ -value empirici simulati sotto l'ipotesi nulla con i quantili della distribuzione uniforme. Si analizzano i  $p$ -value del test  $F$  e quelli ottenuti con l'approssimazione  $\chi^2_8$  della distribuzione del test log-rapporto di verosimiglianza  $w$  e del test  $w^*$  di Skovgaard. Come si può notare, i  $p$ -value simulati del test  $F$  giacciono sulla bisettrice e ciò è una conseguenza del fatto che le soluzioni fornite sono esatte. L'approssimazione asintotica della distribuzione nulla del test  $w^*$  di Skovgaard risulta essere estremamente accurata e migliora nettamente quella relativa al test log-rapporto di verosimiglianza. I  $p$ -value prodotti da quest'ultimo, infatti, sono più piccoli di quello che dovrebbero essere. Quindi, in questo caso, la correzione apportata da  $w^*$  al test  $w$  permette di avere una distribuzione nulla asintotica molto più accurata, con risultati paragonabili a quelli esatti.

In Figura 3.2 si riporta l'analisi di un caso limite in cui il numero di parametri d'interesse è particolarmente elevato rispetto alla numerosità campionaria. Nello specifico si sono effettuate 5000 simulazioni in modo analogo alle precedenti: si sono considerati campioni di numerosità pari a  $n = 600$  e  $p = 500$  variabili esplicative. Si vuole verificare in questo caso la nullità di 400 coefficienti di regressione, per cui il parametro di interesse ha dimensione

$d = 400$ . Il test  $F$ , analogamente a prima, fornisce soluzioni esatte quindi i  $p$ -value si distribuiscono uniformemente. I  $p$ -value calcolati con l'approssimazione asintotica per la distribuzione di  $w$  e  $w^*$ , al contrario, sono molto più piccoli di quello che dovrebbero essere. Ne consegue che il test log-rapporto di verosimiglianza e la versione corretta di Skovgaard non sono adeguati per verificare un tale sistema di ipotesi, nel caso in cui l'informazione campionaria sia esigua in relazione al numero di parametri inclusi nel modello.



**Figura 3.1:** Confronto tra i quantili della distribuzione uniforme e i  $p$ -value simulati sotto l'ipotesi nulla relativi ai tre test per verificare la nullità di  $d = 8$  coefficienti di regressione, nel caso di un modello di regressione lineare normale di numerosità  $n = 100$  con  $p = 20$  variabili esplicative.



**Figura 3.2:** Confronto tra i quantili della distribuzione uniforme e i  $p$ -value simulati sotto l'ipotesi nulla relativi ai tre test per verificare la nullità di  $d = 400$  coefficienti di regressione, nel caso di un modello di regressione lineare normale di numerosità  $n = 600$  con  $p = 500$  variabili esplicative.



# Conclusioni

Lo scopo di questa tesi è studiare il comportamento di test basati sulla verosimiglianza nel contesto del modello di regressione normale. L'approssimazione ottenuta dalla distribuzione nulla del test rapporto di verosimiglianza in alcuni casi limite, ad esempio quando la numerosità campionaria è esigua rispetto al numero di parametri del modello, non è accurata. Si sono studiate quindi due alternative: il test direzionale e il test di Skovgaard. L'approccio direzionale si basa sulla determinazione di un  $p$ -value ottenuto valutando quanto sia più probabile osservare sotto l'ipotesi nulla dei dati più estremi rispetto all'ipotesi di quelli osservati, condizionatamente alla direzione osservata tra il modello nullo e il modello stimato. Il test di Skovgaard, invece, è una modifica del test log-rapporto di verosimiglianza in cui si introduce un fattore di correzione che porta ad un miglioramento dell'accuratezza.

Per analizzare dal punto di vista pratico l'adeguatezza di questi metodi, essi sono stati applicati al modello di regressione lineare normale per verificare un sistema di ipotesi lineari sul vettore dei coefficienti di regressione. Per tale modello esiste un test esatto, il test  $F$ , che è una trasformazione monotona del test rapporto di verosimiglianza, la cui distribuzione nulla è una  $F$  di Fisher. Si sono quindi analizzate le relazioni del test direzionale e del test di Skovgaard con la soluzione esatta fornita dal test  $F$ .

La prima fase dell'analisi ha riguardato la costruzione delle statistiche test. Il fatto che il modello appartenga ad una famiglia esponenziale lineare ha portato ad una semplificazione della specificazione di entrambi i test in esame. Inoltre, per il test direzionale, essendo l'ipotesi nulla funzione dei parametri canonici, è stato possibile condizionarsi ad una distribuzione ben definita libera dai parametri di disturbo per fare inferenza sul parametro

d'interesse. Una volta ottenute le statistiche test, si sono ricavati i relativi  $p$ -value e si sono confrontati con quelli del test esatto.

È risultato che il livello di significatività osservato del test direzionale coincide con quello esatto, ottenuto utilizzando il test  $F$ . Si è dimostrato in particolare che tale risultato vale sia nel caso in cui si verifica la nullità di un gruppo di coefficienti di regressione, sia nel caso più generale di un qualsiasi vincolo lineare tra di essi. Quindi, in questa specifica situazione si possono utilizzare indifferentemente il test direzionale e il test  $F$ , poiché portano al medesimo risultato. Le ragioni di questo risultato possono essere dovute all'esattezza dell'approssimazione *saddlepoint* nel modello normale. Inoltre, resta da investigare se, in questo esempio, la distribuzione della magnitudo dello scostamento dall'ipotesi nulla possa essere indipendente dalla direzione di allontanamento come succede nel caso con varianza nota (Skovgaard, 1988). Il fatto che il test direzionale fornisca soluzioni esatte conferma la validità dell'approccio teorico sottostante. Ne consegue, inoltre, che può essere ritenuto un'adeguata alternativa al test rapporto di verosimiglianza in contesti più generali.

Al contrario del test direzionale, il test di Skovgaard non fornisce soluzioni esatte. Tramite degli studi di simulazione si è verificata l'accuratezza dell'approssimazione asintotica della distribuzione nulla del test. In particolare, si sono analizzati due casi distinti, che differiscono per numerosità campionaria e dimensionalità dei parametri del modello. È risultato che, se la complessità del modello è contenuta e l'informazione campionaria è sufficientemente elevata, l'approssimazione asintotica della distribuzione nulla del test risulta essere estremamente accurata e migliora nettamente quella relativa al test log-rapporto di verosimiglianza. Quindi, in questo caso, la correzione apportata da  $w^*$  al test  $w$  permette di avere risultati paragonabili a quelli esatti. Tuttavia, se la numerosità campionaria è esigua rispetto alla complessità del modello coinvolto, il test di Skovgaard produce  $p$ -value troppo piccoli, analogamente al test rapporto di verosimiglianza.

In conclusione, quindi, per il caso specifico di applicazione analizzato in questa relazione, risulta che il test direzionale è una valida alternativa al test rapporto di verosimiglianza in modelli generali, poiché fornisce soluzioni esatte in casi notevoli come il modello lineare normale. Per quanto riguarda



il test di Skovgaard, invece, può essere ritenuto tale solo in casi non limite.



# Appendice A

## Proprietà della stima di massima verosimiglianza $\widehat{\theta}(t)$

Nel Paragrafo 2.4 si è ottenuta la stima di massima verosimiglianza

$$\widehat{\theta}(t) = \{\widehat{\beta}^T(t), \widehat{\sigma}^2(t)\}^T,$$

in cui

$$\widehat{\beta}(t) = \widehat{\beta} + (X^T X)^{-1} s_{\beta}(t), \quad (\text{A.1})$$

$$\widehat{\sigma}^2(t) = \widehat{\sigma}^2 - \frac{1}{n} \{s_{\beta}^T(t)(X^T X)^{-1} s_{\beta}(t) + 2s_{\beta}^T(t)\widehat{\beta}\}. \quad (\text{A.2})$$

Si mostra che vale:  $\widehat{\theta}(1) = \widehat{\theta}$  e  $\widehat{\theta}(0) = \widehat{\theta}_{\psi}$ .

Per quanto riguarda la prima uguaglianza, dalla (2.7) si ha che  $s_{\beta}(1) = \mathbf{0}$ . Quindi, sostituendo in (A.1) e (A.2), si ottiene il primo risultato.

Si consideri la seconda uguaglianza. Se  $t = 0$ , si può riscrivere la (A.1) come segue:

$$\widehat{\beta}(0) = \widehat{\beta} + (X^T X)^{-1} s_{\beta}(0) = \widehat{\beta} + (X^T X)^{-1} \{-y_{\text{oss}}^T (I_n - P_2) X_1, \mathbf{0}\}^T. \quad (\text{A.3})$$

Il vettore  $s_{\beta}(0) = (s_{1\psi}^T, s_{2\psi}^T)^T$ , di dimensione  $p \times 1$ , ha le ultime  $p - d$  componenti nulle; di conseguenza solo il prodotto tra le prime  $d$  colonne della matrice  $(X^T X)^{-1}$  e le prime  $d$  componenti del vettore  $s_{\beta}(0)$  è diverso da zero. Si esprime dunque la matrice  $(X^T X)^{-1}$  nella sua forma a blocchi, utilizzando il seguente teorema:

**Teorema 3** Sia  $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$  una matrice a blocchi, in cui le sotto-matrici sulla diagonale ( $A_{11}$  e  $A_{22}$ ) sono quadrate e non singolari. Allora l'inversa di  $A$  risulta

$$A^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix} = \begin{pmatrix} A_{11}^{-1}(I + A_{12}B_{22}A_{21}A_{11}^{-1}) & -A_{11}^{-1}A_{12}B_{22} \\ -B_{22}A_{21}A_{11}^{-1} & B_{22} \end{pmatrix},$$

dove  $I$  è una matrice identità di ordine appropriato e  $B_{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$ .

In questo caso  $A = X^T X = \begin{pmatrix} X_1^T X_1 & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 \end{pmatrix}$ , quindi  $A^{-1} = (X^T X)^{-1}$ , con

$$\begin{aligned} A^{11} &= (X_1^T X_1)^{-1} + (X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T X_1 (X_1^T X_1)^{-1}, \\ A^{12} &= -(X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1}, \\ A^{21} &= -(X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T X_1 (X_1^T X_1)^{-1} = (A^{12})^T, \\ A^{22} &= (X_2^T X_2 - X_2^T P_1 X_2)^{-1}. \end{aligned}$$

Il prodotto nella (A.3) dà come risultato il vettore  $\begin{pmatrix} A^{11} s_{1\psi} \\ A^{21} s_{1\psi} \end{pmatrix}$  di dimensione  $p \times 1$ , in cui:

$$\begin{aligned} A^{11} s_{1\psi} &= -(X_1^T X_1)^{-1} X_1^T (I_n - P_2) y_{\text{oss}} + \\ &\quad - (X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 (I_n - P_2) y_{\text{oss}}, \\ A^{21} s_{1\psi} &= (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 (I_n - P_2) y_{\text{oss}}. \end{aligned}$$

Il vettore  $\widehat{\beta}$  in (A.3) può anch'esso essere riscritto evidenziando le prime  $d$  componenti dalle restanti:  $\widehat{\beta} = (\widehat{\beta}_1^T, \widehat{\beta}_2^T)^T$ . Tramite la scomposizione della matrice  $X = [X_1 \ X_2]$  e della matrice inversa  $(X^T X)^{-1}$  si ottiene

$$\begin{aligned} \widehat{\beta}_1 &= (X_1^T X_1)^{-1} X_1^T y_{\text{oss}} - (X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1) y_{\text{oss}}, \\ \widehat{\beta}_2 &= (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1) y_{\text{oss}}. \end{aligned}$$

Quindi riprendendo la (A.3), si ha  $\widehat{\beta}(0) = \begin{pmatrix} \widehat{\beta}_1 + A^{11} s_{1\psi} \\ \widehat{\beta}_2 + A^{21} s_{1\psi} \end{pmatrix}$ , in cui:

$$\begin{aligned} \widehat{\beta}_1 + A^{11} s_{1\psi} &= (X_1^T X_1)^{-1} X_1^T P_2 y_{\text{oss}} - (X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} + \\ &\quad + (X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}}. \end{aligned}$$

Aggiungendo e sottraendo la quantità

$$\{(X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_2 y_{\text{oss}}\}$$

e raccogliendo opportunamente si ottiene:

$$\begin{aligned} \widehat{\beta}_1 + A^{11} s_{1\psi} &= + (X_1^T X_1)^{-1} X_1^T P_2 y_{\text{oss}} + \\ &\quad - (X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} \{X_2^T (I_n - P_1) X_2\} (X_2^T X_2)^{-1} X_2^T y_{\text{oss}} + \\ &\quad - (X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_2) y_{\text{oss}} \\ &= + (X_1^T X_1)^{-1} X_1^T P_2 y_{\text{oss}} - (X_1^T X_1)^{-1} X_1^T P_2 y_{\text{oss}} = \mathbf{0}. \end{aligned}$$

Per quanto riguarda la seconda componente di  $\widehat{\beta}(0)$ , si ha:

$$\widehat{\beta}_2 + A^{21} s_{1\psi} = (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} - (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}}.$$

Analogamente a prima, aggiungendo e sottraendo la quantità

$$(X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_2 y_{\text{oss}}$$

e raccogliendo opportunamente si ottiene:

$$\begin{aligned} \widehat{\beta}_2 + A^{21} s_{1\psi} &= (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} - (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_2 y_{\text{oss}} + \\ &\quad + (X_2^T X_2 - X_2^T P_1 X_2)^{-1} \{X_2^T (I_n - P_1) X_2\} (X_2^T X_2)^{-1} X_2^T y_{\text{oss}} \\ &= (X_2^T X_2)^{-1} X_2^T y_{\text{oss}} = \widehat{\beta}_{2\psi}. \end{aligned}$$

Quindi  $\widehat{\beta}(0) = \begin{pmatrix} \mathbf{0} \\ \widehat{\beta}_{2\psi} \end{pmatrix}$  e ne consegue che  $\widehat{\sigma}^2(0) = \widehat{\sigma}_{\psi}^2$ , da cui si ottiene  $\widehat{\theta}(0) = \widehat{\theta}_{\psi}$ .



# Appendice B

## Dimostrazione analitica dell'uguaglianza dei $p$ -value

Le analisi svolte hanno evidenziato che, nel caso di un modello di regressione lineare normale in cui si verifica la nullità di alcuni parametri  $\beta$ , i  $p$ -value relativi al test direzionale e al test  $F$  sono esattamente uguali. In seguito si riporta nel dettaglio la dimostrazione di tale uguaglianza.

Il  $p$ -value relativo al test  $F$  è

$$p_F = \int_{F_{\text{oss}}}^{+\infty} \frac{1}{B\left(\frac{d}{2}, \frac{n-p}{2}\right)} \frac{1}{t} \left\{ \frac{d^d (n-p)^{n-p} t^d}{(dt + n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dt$$

con  $F_{\text{oss}} = \{(\hat{\sigma}_\psi^2 - \hat{\sigma}^2)/d\}/\{\hat{\sigma}^2/(n-p)\}$  il valore osservato della statistica test  $F$  (2.5), mentre il  $p$ -value relativo al test direzionale è

$$p_D = \frac{\int_1^{t_u} t^{d-1} \left\{ \frac{\hat{\sigma}^2(t)}{\hat{\sigma}_\psi^2} \right\}^{\frac{n-p}{2}-1} dt}{\int_0^{t_u} t^{d-1} \left\{ \frac{\hat{\sigma}^2(t)}{\hat{\sigma}_\psi^2} \right\}^{\frac{n-p}{2}-1} dt}. \quad (\text{B.1})$$

Si tratta quindi di dimostrare che  $p_D = p_F$ .

Considerando  $p_D$ , si effettua la seguente sostituzione:

$$z = \frac{\{\hat{\sigma}_\psi^2 - \hat{\sigma}^2(t)\}/d}{\hat{\sigma}^2(t)/(n-p)}, \quad (\text{B.2})$$

dove  $z$  è una funzione monotona crescente in  $t$ . Gli estremi dell'integrale diventano:

$$\begin{aligned} t = 0 &\rightarrow z = \frac{(\widehat{\sigma}_\psi^2 - \widehat{\sigma}_\psi^2)/d}{\widehat{\sigma}_\psi^2/(n-p)} = 0, \\ t = 1 &\rightarrow z = \frac{(\widehat{\sigma}_\psi^2 - \widehat{\sigma}^2)/d}{\widehat{\sigma}^2/(n-p)} = F_{\text{oss}}, \\ t = t_u &\rightarrow z = \frac{(\widehat{\sigma}_\psi^2 - 0)/d}{0/(n-p)} = +\infty. \end{aligned}$$

Dalla (B.2) si ottiene la quantità  $\frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_\psi^2}$ :

$$\begin{aligned} z &= \frac{n-p}{d} \frac{\widehat{\sigma}_\psi^2 - \widehat{\sigma}^2(t)}{\widehat{\sigma}^2(t)}, \\ dz \widehat{\sigma}^2(t) &= (n-p) \{ \widehat{\sigma}_\psi^2 - \widehat{\sigma}^2(t) \}, \\ \widehat{\sigma}^2(t)(dz + n-p) &= (n-p) \widehat{\sigma}_\psi^2, \\ \frac{\widehat{\sigma}^2(t)}{\widehat{\sigma}_\psi^2} &= \frac{n-p}{dz + n-p}. \end{aligned}$$

Rimane quindi da ricavare  $t$  in funzione di  $z$ , così da poter sostituire  $t$  e  $dt$  con rispettivamente  $z$  e  $dz$  in (B.1). Isolando le componenti che dipendono da  $t$  si ottiene

$$\begin{aligned} n\widehat{\sigma}(t)^2 &= \frac{n-p}{dz + n-p} n\widehat{\sigma}_\psi^2, \\ n\widehat{\sigma}^2 - (1-t)^2 s_\psi^T (X^T X)^{-1} s_\psi - 2(1-t) s_\psi^T \widehat{\beta} &= \frac{n-p}{dz + n-p} n\widehat{\sigma}_\psi^2, \\ (t^2 - 2t + 1) s_\psi^T (X^T X)^{-1} s_\psi + (2-2t) s_\psi^T \widehat{\beta} &= n\widehat{\sigma}^2 - \frac{n-p}{dz + n-p} n\widehat{\sigma}_\psi^2, \\ t^2 s_\psi^T (X^T X)^{-1} s_\psi - 2t \{ s_\psi^T (X^T X)^{-1} s_\psi + s_\psi^T \widehat{\beta} \} &= n\widehat{\sigma}^2 - \frac{n-p}{dz + n-p} n\widehat{\sigma}_\psi^2 - 2s_\psi^T \widehat{\beta} + \\ &\quad - s_\psi^T (X^T X)^{-1} s_\psi. \quad (\text{B.3}) \end{aligned}$$

Si considera l'espressione  $s_\psi^T (X^T X)^{-1} s_\psi + s_\psi^T \widehat{\beta}$ : si esplicita un addendo alla volta, a partire da  $s_\psi^T (X^T X)^{-1} s_\psi$ . Riprendendo i risultati ottenuti in



Appendice A, il prodotto si può riscrivere come

$$\begin{aligned}
& s_{\psi}^T (X^T X)^{-1} s_{\psi} = s_{1\psi}^T A^{11} s_{1\psi} = \\
& = y_{\text{oss}}^T (I_n - P_2) X_1 \{ (X_1^T X_1)^{-1} + (X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T X_1 \\
& \quad (X_1^T X_1)^{-1} \} X_1^T (I_n - P_2) y_{\text{oss}} = \\
& = y_{\text{oss}}^T (I_n - P_2) P_1 (I_n - P_2) y_{\text{oss}} + \\
& \quad + y_{\text{oss}}^T (I_n - P_2) P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 (I_n - P_2) y_{\text{oss}} = \\
& = y_{\text{oss}}^T (I_n - P_2) P_1 (I_n - P_2) y_{\text{oss}} + y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& \quad - y_{\text{oss}}^T P_2 P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& \quad - y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} + \\
& \quad + y_{\text{oss}}^T P_2 P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} \\
& = y_{\text{oss}}^T (I_n - P_2) P_1 (I_n - P_2) y_{\text{oss}} + y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& \quad - 2y_{\text{oss}}^T P_2 P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \tag{B.4}
\end{aligned}$$

$$+ y_{\text{oss}}^T P_2 P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} . \tag{B.5}$$

Considerando singolarmente la (B.4), si aggiunge e si toglie ad essa la quantità  $\{ y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} \}$

$$\begin{aligned}
& - y_{\text{oss}}^T P_2 P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& + y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& - y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} = \\
& = y_{\text{oss}}^T P_2 (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& \quad - y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} = \\
& = y_{\text{oss}}^T X_2 (X_2^T X_2)^{-1} \{ X_2^T (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} \} X_2^T P_1 y_{\text{oss}} + \\
& \quad - y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} = \\
& = y_{\text{oss}}^T P_2 P_1 y_{\text{oss}} - y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} .
\end{aligned}$$

Si procede allo stesso modo con (B.5), togliendo e aggiungendo la quantità

$$\begin{aligned}
& \{y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}}\} \\
& + y_{\text{oss}}^T P_2 P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} + \\
& - y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} + \\
& + y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} = \\
& = - y_{\text{oss}}^T P_2 (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} + \\
& + y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} = \\
& = - y_{\text{oss}}^T X_2 (X_2^T X_2)^{-1} \{X_2^T (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1}\} X_2^T P_1 P_2 y_{\text{oss}} + \\
& + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} = \\
& = - y_{\text{oss}}^T P_2 P_1 P_2 y_{\text{oss}} + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} .
\end{aligned}$$

Quindi, sostituendo i risultati ottenuti, si ha

$$\begin{aligned}
s_\psi^T (X^T X)^{-1} s_\psi & = y_{\text{oss}}^T (I_n - P_2) P_1 (I_n - P_2) y_{\text{oss}} + \\
& + y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& - 2 \{y_{\text{oss}}^T P_2 P_1 y_{\text{oss}} - y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}}\} + \\
& - y_{\text{oss}}^T P_2 P_1 P_2 y_{\text{oss}} + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} \\
& = y_{\text{oss}}^T P_1 y_{\text{oss}} + y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& - 2 y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} . \tag{B.6}
\end{aligned}$$

Si ripete nuovamente lo stesso procedimento in (B.6) con la quantità

$$\begin{aligned}
& \{y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_2 y_{\text{oss}}\} \\
& + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 P_2 y_{\text{oss}} \\
& - y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_2 y_{\text{oss}} \\
& + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_2 y_{\text{oss}} = \\
& = - y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1) P_2 y_{\text{oss}} + \\
& + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_2 y_{\text{oss}} = \\
& = - y_{\text{oss}}^T X_2 \{(X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1) X_2\} (X_2^T X_2)^{-1} X_2^T y_{\text{oss}} + \\
& + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} = \\
& = - y_{\text{oss}}^T P_2 y_{\text{oss}} + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} .
\end{aligned}$$

Concludendo risulta che

$$\begin{aligned}
s_\psi^T (X^T X)^{-1} s_\psi &= y_{\text{oss}}^T P_1 y_{\text{oss}} - y_{\text{oss}}^T P_2 y_{\text{oss}} + \\
&\quad + [y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
&\quad - 2y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
&\quad + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}}] \\
&= y_{\text{oss}}^T (P_1 - P_2) y_{\text{oss}} + \\
&\quad + y_{\text{oss}}^T (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1) y_{\text{oss}} .
\end{aligned}$$

Si procede esplicitando  $s_\psi^T \hat{\beta}$ :

$$\begin{aligned}
s_\psi^T \hat{\beta} &= s_\psi^T (X^T X)^{-1} X^T y_{\text{oss}} \\
&= -y_{\text{oss}}^T (I_n - P_2) X_1 \{ (X_1^T X_1)^{-1} X_1^T y_{\text{oss}} + (X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} \\
&\quad X_2^T P_1 y_{\text{oss}} - (X_1^T X_1)^{-1} X_1^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} \} \\
&= -y_{\text{oss}}^T (I_n - P_2) P_1 y_{\text{oss}} - y_{\text{oss}}^T (I_n - P_2) P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1) y_{\text{oss}} \\
&= -y_{\text{oss}}^T (I_n - P_2) P_1 y_{\text{oss}} + y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} + \\
&\quad - y_{\text{oss}}^T P_2 P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} + \tag{B.7}
\end{aligned}$$

$$\begin{aligned}
&\quad - y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
&\quad + y_{\text{oss}}^T P_2 P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} . \tag{B.8}
\end{aligned}$$

Considerando singolarmente la (B.7), si aggiunge e si toglie ad essa la quantità  $\{y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}}\}$

$$\begin{aligned}
&\quad - y_{\text{oss}}^T P_2 P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} + y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} + \\
&\quad - y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} = \\
&= y_{\text{oss}}^T P_2 (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} + \\
&\quad - y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} = \\
&= y_{\text{oss}}^T X_2 (X_2^T X_2)^{-1} \{ X_2^T (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} \} X_2^T y_{\text{oss}} + \\
&\quad - y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} = \\
&= y_{\text{oss}}^T P_2 y_{\text{oss}} - y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} .
\end{aligned}$$

Analogamente in (B.8) si sottrae e si somma la quantità

$$\begin{aligned}
& \{y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}}\} \\
& y_{\text{oss}}^T P_2 P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} - y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& + y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} = \\
& = -y_{\text{oss}}^T P_2 (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + \\
& + y_{\text{oss}}^T P_2 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} = \\
& = -y_{\text{oss}}^T X_2 (X_2^T X_2)^{-1} \{X_2^T (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1}\} X_2^T P_1 y_{\text{oss}} + \\
& + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} = \\
& = -y_{\text{oss}}^T P_2 P_1 y_{\text{oss}} + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} .
\end{aligned}$$

Quindi, sostituendo i risultati ottenuti, si ha

$$\begin{aligned}
s_{\psi}^T \hat{\beta} &= -y_{\text{oss}}^T (I_n - P_2) P_1 y_{\text{oss}} + y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} + \\
& + y_{\text{oss}}^T P_2 y_{\text{oss}} - y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} + \\
& - y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} \\
& - y_{\text{oss}}^T P_2 P_1 y_{\text{oss}} + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} \\
& = -y_{\text{oss}}^T P_1 y_{\text{oss}} + y_{\text{oss}}^T P_2 y_{\text{oss}} + \\
& + \{y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} - y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T y_{\text{oss}} + \\
& - y_{\text{oss}}^T P_1 X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}} + y_{\text{oss}}^T X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T P_1 y_{\text{oss}}\} \\
& = -y_{\text{oss}}^T (P_1 - P_2) y_{\text{oss}} - y_{\text{oss}}^T (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1) y_{\text{oss}} \\
& = -s_{\psi}^T (X^T X)^{-1} s_{\psi} .
\end{aligned}$$

Di conseguenza  $s_{\psi}^T (X^T X)^{-1} s_{\psi} + s_{\psi}^T \hat{\beta} = 0$ .

Riprendendo la (B.3) e considerando il risultato appena ricavato si ha

$$t^2 s_{\psi}^T (X^T X)^{-1} s_{\psi} = n\hat{\sigma}^2 - \frac{n-p}{dz+n-p} n\hat{\sigma}_{\psi}^2 + s_{\psi}^T (X^T X)^{-1} s_{\psi} . \quad (\text{B.9})$$

Esplicitando  $n\hat{\sigma}^2$  si ottiene

$$\begin{aligned}
n\hat{\sigma}^2 &= y_{\text{oss}}^T (I_n - P) y_{\text{oss}} \\
&= y_{\text{oss}}^T y_{\text{oss}} - y_{\text{oss}}^T \{P_1 + (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1)\} y_{\text{oss}} \\
&= y_{\text{oss}}^T (I_n - P_1) y_{\text{oss}} - y_{\text{oss}}^T (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1) y_{\text{oss}} ,
\end{aligned}$$

in cui la matrice  $P$  è stata esplicitata utilizzando il Teorema 3 e la scomposizione della matrice  $(X^T X)^{-1}$ . Nello specifico

$$P = X^T (X^T X)^{-1} X = X_1 (A^{11} + A^{12}) X_1^T + X_2 (A^{21} + A^{22}) X_2^T,$$

dove  $A^{11}, A^{12}, A^{21}, A^{22}$  sono gli elementi di  $(X^T X)^{-1}$ . La quantità  $n\widehat{\sigma}^2 + s_\psi^T (X^T X)^{-1} s_\psi$  risulta quindi essere

$$\begin{aligned} n\widehat{\sigma}^2 + s_\psi^T (X^T X)^{-1} s_\psi &= y_{\text{oss}}^T (I_n - P_1) y_{\text{oss}} \\ &\quad - y_{\text{oss}}^T (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1) y_{\text{oss}} + \\ &\quad + y_{\text{oss}}^T (P_1 - P_2) y_{\text{oss}} + \\ &\quad + y_{\text{oss}}^T (I_n - P_1) X_2 (X_2^T X_2 - X_2^T P_1 X_2)^{-1} X_2^T (I_n - P_1) y_{\text{oss}} \\ &= y_{\text{oss}}^T (I_n - P_2) y_{\text{oss}} = n\widehat{\sigma}_\psi^2. \end{aligned}$$

Riprendendo la (B.9) si ottiene

$$\begin{aligned} t^2 s_\psi^T (X^T X)^{-1} s_\psi &= n\widehat{\sigma}_\psi^2 - \frac{n-p}{dz+n-p} n\widehat{\sigma}_\psi^2, \\ t^2 s_\psi^T (X^T X)^{-1} s_\psi &= \frac{dz}{dz+n-p} n\widehat{\sigma}_\psi^2, \\ t^2 &= \frac{dz}{dz+n-p} n\widehat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}, \\ t &= \sqrt{\frac{dz}{dz+n-p} n\widehat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}}. \end{aligned} \quad (\text{B.10})$$

Dalla (B.10) si ricava  $dt$ :

$$dt = \frac{n\widehat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}}{2\sqrt{\frac{dz}{dz+n-p} n\widehat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}} (dz+n-p)^2} dz.$$

Avendo ottenuto tutte le componenti, si applica la sostituzione proposta (B.2)

agli integrali in (B.1):

$$\begin{aligned}
p_D &= \frac{\int_1^{t_u} t^{d-1} \left\{ \frac{\hat{\sigma}^2(t)}{\hat{\sigma}_\psi^2} \right\}^{(n-p)/2-1} dt}{\int_0^{t_u} t^{d-1} \left\{ \frac{\hat{\sigma}^2(t)}{\hat{\sigma}_\psi^2} \right\}^{(n-p)/2-1} dt} \\
&= \frac{\int_{F_{\text{oss}}}^{+\infty} \frac{(dz)^{d/2-1} (n-p)^{(n-p)/2}}{(dz+n-p)^{(d+n-p)/2}} [n\hat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}]^{d/2} dz}{\int_0^{+\infty} \frac{(dz)^{d/2-1} (n-p)^{(n-p)/2}}{(dz+n-p)^{(d+n-p)/2}} [n\hat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}]^{d/2} dz} \\
&= \frac{\frac{1}{d} [n\hat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}]^{d/2} \int_{F_{\text{oss}}}^{+\infty} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{(dz+n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dz}{\frac{1}{d} [n\hat{\sigma}_\psi^2 \{s_\psi^T (X^T X)^{-1} s_\psi\}^{-1}]^{d/2} \int_0^{+\infty} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{(dz+n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dz} \\
&= \frac{\int_{F_{\text{oss}}}^{+\infty} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{(dz+n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dz}{\int_0^{+\infty} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{(dz+n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dz}. \tag{B.11}
\end{aligned}$$

Si osserva che il denominatore è la funzione *Beta* di parametri  $d/2$  e  $(n-p)/2$ . Nello specifico si ha

$$\begin{aligned}
&\int_0^{+\infty} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{(dz+n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dz = \\
&\int_0^{+\infty} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{\left(\frac{dz}{n-p} + 1\right)^{d+n-p} (n-p)^{d+n-p}} \right\}^{\frac{1}{2}} dz = \\
&\int_0^{+\infty} \frac{\left(\frac{dz}{n-p}\right)^{d/2-1}}{\left(\frac{dz}{n-p} + 1\right)^{(d+n-p)/2}} \left(\frac{d}{n-p}\right) dz = \\
&\int_0^{+\infty} \frac{r^{d/2-1}}{(1+r)^{(d+n-p)/2}} dr = B\left(\frac{d}{2}, \frac{n-p}{2}\right),
\end{aligned}$$

dove nell'ultimo integrale si è applicata la sostituzione  $r = dz/(n-p)$ .

---

Quindi, portando all'interno dell'integrale al numeratore in (B.11) la funzione *Beta* appena ottenuta, si conclude che

$$p_D = \int_{F_{\text{oss}}}^{+\infty} \frac{1}{\text{B}\left(\frac{d}{2}, \frac{n-p}{2}\right)} \frac{1}{z} \left\{ \frac{d^d (n-p)^{n-p} z^d}{(d z + n - p)^{d+n-p}} \right\}^{\frac{1}{2}} dz = p_F .$$





# Bibliografia

- Azzalini, A. (2004). *Inferenza Statistica*. Milano: Springer.
- Davison, A. C., Fraser, D. A. S., Reid, N. e Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families. *Journal of the American Statistical Association* **109**, 302–314.
- Fraser, D.A.S. e Massam, H. (1985). Conical tests: Observed levels of significance and confidence regions. *Statistische Hefte* **26**, 1–17.
- Fraser, D.A.S. e Reid, N. (1995). Ancillaries and third order significance. *Utilitas Mathematica* **7**, 33–53.
- Fraser, D.A.S., Reid, N. e Sartori, N. (2016). Accurate directional inference for vector parameters. *Biometrika* **103**, 625–639.
- Gregorio, E. e Salce, L. (2010). *Algebra lineare*. Edizioni Libreria Progetto.
- Lehmann, E. L. e Romano, J. P. (2005). *Testing Statistical Hypotheses*. New York: Springer.
- Pace, L. e Salvan, A. (1996). *Teoria della Statistica: metodi, modelli, approssimazioni asintotiche*. Padova: Cedam.
- Pace, L. e Salvan, A. (2001). *Introduzione alla statistica II: Inferenza, verosimiglianza, modelli*. Padova: Cedam.
- Skovgaard, I. M. (1988). Saddlepoint expansions for directional test probabilities. *Journal of the Royal Statistical Society B* **50**, 3–32.
- Skovgaard, I. M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics* **28**, 3–32.
- Wikipedia (2016). *Funzione beta di Eulero*. URL: [\url{//it.wikipedia.org/w/index.php?title=Funzione\\_beta\\_di\\_Eulero&oldid=83382256}](https://it.wikipedia.org/w/index.php?title=Funzione_beta_di_Eulero&oldid=83382256).



# Ringraziamenti

*Alla fine di questo percorso credo sia giusto ringraziare chi ne ha fatto parte e l'ha reso possibile.*

*Un ringraziamento speciale va al professore Nicola Sartori, relatore di questa tesi, che mi ha affiancato con professionalità e disponibilità rendendo questa esperienza piacevole e interessante.*

*Un grazie in particolare ai miei genitori, a cui dedico il raggiungimento di questo traguardo. Senza di loro tutto ciò non sarebbe stato possibile. Grazie per avermi supportata e sopportata in questi anni, per la vostra pazienza e il vostro immancabile aiuto. Un grazie anche ai nonni, agli zii e ai cugini per il sostegno dato.*

*Grazie a Francesca, Francesca, Silvia, Susanna e Tatiana per essermi state vicine in ogni momento, soprattutto nei più difficili, per le risate e le serate passate assieme. Anche se ognuna ha intrapreso la propria strada, l'amicizia che ci lega rimarrà per sempre.*

*Grazie ai miei compagni di corso, per aver reso unica ed indimenticabile questa esperienza universitaria. Porterò con me il ricordo dei giorni passati a studiare assieme, delle feste e delle serate in cui ci siamo divertiti, delle trasferte che abbiamo organizzato. Ringrazio anche i miei colleghi tutor, Edoardo Lucia e Mattia, per il bellissimo percorso affrontato assieme da cui è nata una forte intesa.*

*E infine un ringraziamento speciale va a Mattia, per aver creduto in me e soprattutto in noi sin dall'inizio, per avermi sostenuta in ogni momento, per avermi aiutata e confortata. Abbiamo passato assieme un anno fantastico e sono sicura che i prossimi lo saranno ancor di più.*