

UNIVERZA V LJUBLJANI  
EKONOMSKA FAKULTETA

MAGISTRSKO DELO

BOOTSTRAP HAUSMANOVEGA SPECIFIKACIJSKEGA TESTA ZA  
REGRESIJSKI MODEL S STRUKTURO MOTNJE

Ljubljana, september 2003

PETER REBEC

## Kazalo

Uvod.....	2
1. Model za panelne podatke.....	3
1.1.1 POLS ocena modela za panelne podatke .....	3
1.1.2 GLS ocena parametrov modela za panelne podatke .....	7
1.2 Linearni model za panelne podatke z neopazovanimi učinki .....	12
1.2.1 Model s slučajnimi učinki .....	13
1.2.2 Model s fiksnimi učinki.....	17
2. Hausmanov specifikacijski test za regresijski model s strukturo motnje.....	22
3. Bootstrap .....	24
3.1 Edgeworthov razvoj .....	26
3.2 Splošne statistike .....	26
3.3 Bootstrap in Edgewortov razvoj.....	28
3.4 Vrste intervalov zaupanja.....	30
3.5 Edgeworthov in Cornish-Fisherjev razvoj .....	33
3.6 Lastnosti bootstrap kvantilov in kritičnih točk.....	35
3.7 Dopustni model za Edgeworthov razvoj .....	36
4. Eksperiment.....	39
4.1 Uporaba neparametrične bootstrap metode za linearni regresijski model s strukturo motnje.....	40
4.2 Statistično prizkušanje hipotez z bootstrap metodo .....	41
4.2.1 Bootstrap Hausmanovega specifikacijskega testa.....	42
4.2.2 Specifikacija ničelne in alternativne hipoteze .....	43
4.3 Konstrukcija eksperimenta .....	44
4.3.1 Monte Carlo ponovitve.....	45
4.3.2 Bootstrap ponovitve .....	47
4.3.3 Porazdelitev korelacijskega koeficienta alternativne hipoteze.....	47
5. Rezultati .....	50
6. Zaključek.....	58
Dodatek D1 .....	59
Dodatek D2 .....	61
Literatura .....	69

## Uvod

Pri analizi ekonomskih pojavov pogosto nastopajo podatki, ki imajo časovno in presečno komponento. Glavni razlog za njihovo uporabo je problem manjkajočih spremenljivk, ki lahko povzroči nedosledno oceno parametrov. Dodatna prednost je tudi večja velikost vzorca, ki zato posledično vsebuje več informacije o ocenjevanih parametrih, manjša kolinearnost med pojasnjevalnimi spremenljivkami in možnost proučevanja heterogenih pojavov. Najpogostejša metoda za modeliranje panelnih podatkov je model s strukturo motnje. Razlikujemo dva načina ocenjevanja regresijskih modelov s strukturo motnje (Davidson, MacKinnon, 1993, str. 322). Model s fiksnimi učinki vrednotimo pogojno glede na individualno komponento motnje, zato ni potrebna predpostavka neodvisnosti individualne komponente motnje od neodvisnih spremenljivk. Model s slučajnimi učinki pa ocenjujemo nepogojno. Prednost uporabe modela s slučajnimi učinki, v primeru njegove ustreznosti, je večja učinkovitost pri ocenjevanju parametrov.

Kriterij za odločanje med modelom s fiksnimi in slučajnimi učinki je Hausmanov specifikacijski test (Hausman, 1978). V nalogi smo analizirali odvisnost napake I. vrste in moči asimptotičnega in bootstrap Hausmanovega testa od parametrov modela panelnih podatkov. Preučevani parametri so bili dimenzija panela, varianca spremenljivke neopazovanega učinka in šuma, heteroskedastičnost in časovna avtokorelacija šuma, korelacija časovnega povprečja neodvisne spremenljivke z neopazovanim učinkom in število neodvisnih spremenljivk.

V prvem delu bomo navedli predpostavke modela s slučajnimi in fiksnimi koeficienti, ki mu v drugem delu sledi opis Hausmanovega testa za regresijski model s strukturo motnje. V tretjem delu bomo obravnavali bootstrap in potrebna orodja povezana z Edgeworthovim razvojem, ki jih bomo potrebovali za analiziranje asimptotičnih in bootstrap intervalov pokritja. Četrty in peti del predstavljata jedro magistrske naloge. V četrtem delu je opisan eksperiment, s katerim smo analizirali vpliv parametrov modela panelnih podatkov na asimptotični in bootstrap Hausmanov test. V petem delu komentiramo rezultate, v šestem pa sledi zaključek.

## 1. Model za panelne podatke

V prvem poglavju bomo navedli nekaj splošnih metodoloških elementov pomembnih za modeliranje panelnih podatkov. Pri tem sledimo zapisu, ki ga uporablja Wooldridge (2002).

Vsako spremenljivko presečne enote opazujemo v  $T$  časovnih obdobjih. Naj bo  $\mathbf{x}_t$   $1 \times K$  vektor za  $t = 1, 2, \dots, T$  in naj bo  $\boldsymbol{\beta}$   $K \times 1$  vektor. Populacijski model zapišemo kot:

$$y_t = \mathbf{x}_t \boldsymbol{\beta} + u_t, \quad t = 1, 2, \dots, T \quad (1.1)$$

kjer sta  $y_t$  in  $u_t$  skalarni spremenljivki. Kadar želimo pokazati, da se enačba nanaša na določeno presečno enoto  $i$  za določeno časovno razdobje  $t$ , zapišemo  $y_{it} = \mathbf{x}_{it} \boldsymbol{\beta} + u_{it}$ . V nadaljevanju predpostavljamo, da število presečnih enot  $N \rightarrow \infty$  in je  $T$  fiksni. Indeksa  $i$  ne bomo pisali, kadar to ne bo potrebno. V naslednjem razdelku bomo obravnavali ocenjevanje parametrov modela za panelne podatke z združenimi navadnimi najmanjšimi kvadrati (pooled OLS – POLS).

### 1.1.1 POLS ocena modela za panelne podatke

Slučajni vzorec neodvisnih, identično porazdeljenih presečnih opazovanj, označimo kot  $\{(\mathbf{X}_i, \mathbf{y}_i) : i = 1, 2, \dots, N\}$ , kjer je  $\mathbf{X}_i$   $T \times K$  matrika in  $\mathbf{y}_i$   $T \times 1$  vektor. Vektor  $\mathbf{y}_i$  vsebuje vrednosti odvisne spremenljivke za  $T$  časovnih opazovanj, matrika  $\mathbf{X}_i$  pa vrednosti neodvisnih spremenljivk za ista časovna opazovanja. Od tu naprej predpostavimo, da vse variance in pričakovane vrednosti  $u_{it}$  obstajajo in so končne.

Za dosledno POLS oceno  $\boldsymbol{\beta}$ , ko  $N \rightarrow \infty$ , zadoščata naslednji predpostavki:

$$\text{Predpostavka POLS.1: } E(\mathbf{x}_t^T u_t) = 0, \quad t = 1, 2, \dots, T.$$

Predpostavka POLS.2:  $\text{rang}\left(\sum_{t=1}^T E(\mathbf{x}_t^T \mathbf{x}_t)\right) = K$ .

Predpostavka POLS.1 je pogoj ortogonalnosti, ki je, v primeru vključenosti konstante v  $\mathbf{x}$ , ekvivalentna ničnemu povprečju motnje in nekoreliranosti motnje s pojasnjevalnimi spremenljivkami v tekočem obdobju. Odnos med  $\mathbf{x}_s$  in  $u_t$  za  $s \neq t$  ni omejen, zato predpostavka ne zahteva stroge eksogenosti pojasnjevalnih spremenljivk. Po predpostavkah POLS.1 in POLS.2 je  $\boldsymbol{\beta} = \left(E(\mathbf{X}_i^T \mathbf{X}_i)\right)^{-1} E(\mathbf{X}_i^T \mathbf{y}_i)$ , kar pomeni, da je vektor parametrov  $\boldsymbol{\beta}$  enolično določljiv. Cenilko za  $\boldsymbol{\beta}$  imenujemo POLS cenilka, ker ustreza OLS na opazovanjih združenih po  $i$  in  $t$ :

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T \mathbf{x}_{it} \right)^{-1} \left( \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T y_{it} \right), \quad (1.2)$$

pri že omenjeni predpostavki slučajnega vzorčenja.

IZREK 1.1 (Doslednost POLS): Po predpostavkah POLS.1 in POLS.2 velja  $\text{plim} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ .

Dokaz:

$$\hat{\boldsymbol{\beta}} = \left( N^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T \mathbf{x}_{it} \right)^{-1} \left( N^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T y_{it} \right) = \boldsymbol{\beta} + \left( N^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T \mathbf{x}_{it} \right)^{-1} \left( N^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T u_{it} \right)$$

zapišemo v polni matrični obliki kot  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ , kjer je  $\mathbf{X}$   $NT \times K$  matrika pojasnjevalnih spremenljivk in  $\mathbf{Y}$   $NT \times 1$  vektor odvisne spremenljivke. Po predpostavki POLS.2 je  $\mathbf{X}^T \mathbf{X}$  nesingularna z verjetnostjo, ki se približuje 1, in

$$\text{plim} \left( N^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T \mathbf{x}_{it} \right)^{-1} = \mathbf{A}^{-1},$$

kjer je  $\mathbf{A} \equiv E(\mathbf{x}^T \mathbf{x})$  (glej pomožno trditev D1.1). Po predpostavki POLS.1 velja

$$\text{plim} \left( N^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T u_{it} \right) = E(\mathbf{x}^T u) = 0.$$

Upoštevajoč Slutskyjevo lemo (glej D1.2) sledi  $\text{plim}\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{A}^{-1} \cdot 0 = \boldsymbol{\beta}$ .

Zapišimo asimptotično porazdelitev POLS cenilke:

$$\text{Prepišimo } \sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left( N^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T \mathbf{x}_{it} \right)^{-1} \left( N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T u_{it} \right).$$

Pokazali smo, da

$$\left( N^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T \mathbf{x}_{it} \right)^{-1} - \mathbf{A}^{-1} = o_p(1)$$

(glej definicijo D1.5). Po začetni predpostavki je zaporedje  $\{(\mathbf{x}_{it}^T u_{it}) : i, t = 1, 2, \dots\}$  i.i.d. s povprečjem 0 in ima vsak element končno varianco. Potem sledi iz centralnega limitnega izreka (D1.7), da  $N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T u_{it} \xrightarrow{d} N(0, \mathbf{B})$ , kjer je  $\mathbf{B}$   $K \times K$  matrika  $\mathbf{B} = E(\mathbf{X}_i^T \mathbf{u}_i \mathbf{u}_i^T \mathbf{X}_i)$ .

Ker velja

$$N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T u_{it} = O_p(1)$$

lahko zapišemo

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{A}^{-1} \left( N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T u_{it} \right) + o_p(1), \text{ ker } o_p(1) \cdot O_p(1) = o_p(1).$$

**IZREK 1.2 (Asimptotična normalnost POLS cenilke):** Po predpostavkah POLS.1 in POLS.2 velja  $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1})$ .

Za uporabo običajnih OLS statistik za POLS je potrebna dodatna predpostavka presečne homoskedastičnosti in nekoreliranosti:

Predpostavka POLS.3:

(a)  $E(u_t^2 \mathbf{x}_t^T \mathbf{x}_t) = \sigma^2 E(\mathbf{x}_t^T \mathbf{x}_t)$ ,  $t = 1, 2, \dots, T$ , kjer je  $\sigma^2 = E(u_t^2)$  za  $\forall t$ ;

(b)  $E(u_t u_s \mathbf{x}_t^T \mathbf{x}_s) = 0$ ,  $t \neq s$ ,  $t, s = 1, 2, \dots, T$ .

Zadostni pogoj za prvi del predpostavke je  $E(u_t^2 | \mathbf{x}_t) = \sigma^2$  za  $\forall t$ . To pomeni, da pogojna varianca  $u_t$  ni odvisna od  $\mathbf{x}_t$  in da je absolutna varianca  $u_t$  enaka za vsa časovna obdobja. Drugi del predpostavke omeji pogojne kovariance motnje v času na 0. Ker  $\mathbf{x}_t$  vsebuje tudi konstanto, predpostavka POLS.3b zahteva najmanj  $E(u_t u_s) = 0$ ,  $t \neq s$ . Zadostni pogoj za drugi del predpostavke je  $E(u_t u_s | \mathbf{x}_t, \mathbf{x}_s) = 0$ ,  $t \neq s$ ,  $t, s = 1, 2, \dots, T$ . Predpostavka POLS.3 implicira več kot samo določeno obliko brezpogojne kovariančne matrike motnje  $\mathbf{u} = (u_1, \dots, u_T)^T$ , namreč,  $E(u_t u_t^T) = \sigma^2 \mathbf{I}_T$ . Tako dosežemo konstantne absolutne variance in ničelne brezpogojne kovariance in omejimo pogojne variance in kovariance. Po predpostavki POLS.3 velja  $\mathbf{B} = \sigma^2 \mathbf{A}$ .

IZREK 1.3 (Lastnosti POLS na velikih vzorcih): Po predpostavkah POLS.1 in POLS.2, je cenilka POLS dosledna in asimptotično normalna. Če dodatno predpostavimo še POLS.3, potem velja:

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \sigma^2 \mathbf{A}^{-1}),$$

z asimptotično varianco  $A \text{var}(\hat{\boldsymbol{\beta}}) = \sigma^2 \frac{(E(\mathbf{X}_i^T \mathbf{X}_i))^{-1}}{N}$ ,

Cenilka za  $A \text{var}(\hat{\boldsymbol{\beta}})$  je  $\hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1} = \hat{\sigma}^2 \left( \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}^T \mathbf{x}_{it} \right)^{-1}$ ,

kjer je  $\hat{\sigma}^2$  običajna OLS ocena variance iz združene regresije  $y_{it}$  na  $\mathbf{x}_{it}$ ,  $t = 1, 2, \dots, T$ ,  $i = 1, \dots, N$ .

## 1.1.2 GLS ocena parametrov modela za panelne podatke

Model za panelne podatke

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{v}_i, \quad (1.3)$$

zapisan za posamezno presečno enoto s  $t = 1, 2, \dots, T$  časovnimi opazovanji, kjer je motnja presečne enote v času korelirana, ocenimo z metodo posplošenih najmanjših kvadratov (generalized LS - GLS).

Predpostavka GLS.1:  $E(\mathbf{X}_i \otimes \mathbf{v}_i) = 0$ .

Pri zapisu predpostavke GLS.1 smo uporabili simbol  $\otimes$  za Kroneckerjev produkt. Matrika neodvisnih spremenljivk  $\mathbf{X}_i$  vsebuje stolpec enk, zato velja  $E(\mathbf{v}_i) = 0$ . Predpostavka GLS.1 pomeni nekoreliranost vsakega elementa  $\mathbf{v}_i$  z vsakim elementom  $\mathbf{X}_i$ . Zadostni, vendar ne potrebni, pogoj za predpostavko GLS.1 je  $E(\mathbf{v}_i | \mathbf{X}_i) = 0$ .

Definirajmo simetrično, pozitivno definitno matriko  $T \times T$ :  $\boldsymbol{\Omega} \equiv E(\mathbf{v}_i \mathbf{v}_i^T)$ . Matrika  $\boldsymbol{\Omega}$  je kovariančna matrika  $\mathbf{v}_i$ . Po predpostavki je  $\boldsymbol{\Omega}$  enaka za vse  $i$ . Za doslednost je potrebna še predpostavka o rangi  $K \times K$  matrike  $E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i)$ .

Predpostavka GLS.2: Matrika  $E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i)$  je nesingularna.

GLS oceno dobimo tako, da model (1.3) z leve pomnožimo z  $\boldsymbol{\Omega}^{-1/2}$ :

$$\boldsymbol{\Omega}^{-1/2} \mathbf{y}_i = (\boldsymbol{\Omega}^{-1/2} \mathbf{X}_i) \boldsymbol{\beta} + \boldsymbol{\Omega}^{-1/2} \mathbf{v}_i, \text{ oz. } \mathbf{y}_i^* = \mathbf{X}_i^* \boldsymbol{\beta} + \mathbf{v}_i^* \quad (1.4)$$

in velja  $E(\mathbf{v}_i^* \mathbf{v}_i^{*T}) = \mathbf{I}_T$ . Zapišimo GLS cenilko za  $\boldsymbol{\beta}$ :



$$\boldsymbol{\beta}^* = \left( \sum_{i=1}^N \mathbf{X}_i^{*T} \mathbf{X}_i^* \right)^{-1} \left( \sum_{i=1}^N \mathbf{X}_i^{*T} \mathbf{y}_i^* \right) = \left( \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{y}_i \right). \quad (1.5)$$

Po predpostavki GLS.2 obstaja inverzna matrika v (1.5) z verjetnostjo, ki se približuje 1, ko  $N \rightarrow \infty$ . Vektor  $\boldsymbol{\beta}^*$  zapišemo v matrični obliki kot

$$\boldsymbol{\beta}^* = [\mathbf{X}^T (\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1}) \mathbf{X}]^{-1} [\mathbf{X}^T (\mathbf{I}_N \otimes \boldsymbol{\Omega}^{-1}) \mathbf{Y}], \text{ kjer je } \mathbf{X} \text{ } NT \times K \text{ in } \mathbf{Y} \text{ } NT \times 1.$$

IZREK 1.4 : Po predpostavkah GLS.1 in GLS.2 velja  $\text{plim} \boldsymbol{\beta}^* = \boldsymbol{\beta}$ .

Dokaz: Po šibkem zakonu o velikih številih (D1.6) velja  $N^{-1} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i \xrightarrow{p} E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i)$ . Iz

predpostavke GLS.2 in Slutskyjeve leme (D1.2) sledi  $\left( N^{-1} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right)^{-1} \xrightarrow{p} \mathbf{A}^{-1}$ , kjer je

$$\mathbf{A} \equiv E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i). \quad (1.6)$$

Potrebno je še pokazati, da  $\text{plim} N^{-1} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i = 0$ . Zaradi zakona o velikih številih je

dovolj, če pokažemo  $E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i) = 0$ . Za matrike ustreznih dimenzij velja  $\text{vec}(\mathbf{DEF}) = (\mathbf{F}^T \otimes \mathbf{D}) \cdot \text{vec}(\mathbf{E})$  (operator  $\text{vec}$  matriko pretvori v vektor stolpec tako, da stolpce matrike razvrsti enega pod drugega). Po predpostavki GLS.1 sledi:

$$\text{vec} E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i) = E(\mathbf{v}_i^T \otimes \mathbf{X}_i^T) \text{vec}(\boldsymbol{\Omega}^{-1}) = E(\mathbf{v}_i \otimes \mathbf{X}_i)^T \text{vec}(\boldsymbol{\Omega}^{-1}) = 0,$$

kjer smo upoštevali dopustnost zamenjave vrstnega reda pričakovanj in  $\text{vec}$  operatorja.

IZREK 1.5 (Asimptotična normalnost GLS cenilke): Po predpostavkah GLS.1 in GLS.2 velja

$$\sqrt{N}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) \overset{a}{\sim} N(0, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}).$$

Dokaz:  $\sqrt{N}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) = \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right)^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i \right)$  in po centralnem limitnem

izreku velja  $N^{-1/2} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i \xrightarrow{d} N(0, \mathbf{B})$ , kjer je

$$\mathbf{B} \equiv E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i \mathbf{v}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i). \quad (1.7)$$

Ker je  $N^{-1/2} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i = O_p(1)$  in  $\left( N^{-1} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right)^{-1} - \mathbf{A}^{-1} = o_p(1)$  zapišemo

$\sqrt{N}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) = \mathbf{A}^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i \right) + o_p(1)$ . Preostanek dokaza sledi iz leme o asimptotični ekvivalenci.

Za GLS oceno  $\boldsymbol{\beta}^*$  je potrebno poznati  $\boldsymbol{\Omega}$  do multipikativnega člena.  $\boldsymbol{\Omega} = \sigma^2 \mathbf{C}$ , kjer je  $\mathbf{C}$  znana  $T \times T$  pozitivno definitna matrika in  $\sigma^2$  neznana konstanta. Ker  $\mathbf{C}$  v splošnem ni znana, pri FGLS oceni nadomestimo neznano matriko  $\boldsymbol{\Omega}$  z dosledno oceno. Zaradi nelinearnosti  $\boldsymbol{\Omega}$  v izrazu za FGLS cenilko, je izpeljava lastnosti FGLS za končne vzorce težavna. Pokazati se da, da je FGLS cenilka asimptotično nepristranska. Asimptotične lastnosti FGLS cenilke, ko  $N \rightarrow \infty$ , so enake kot za GLS cenilko po predpostavkah GLS.1 in GLS.2.

Predpostavimo, da imamo dosledno cenilko  $\hat{\boldsymbol{\Omega}}$  za  $\boldsymbol{\Omega}$ :

$$\text{plim}_{N \rightarrow \infty} \hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}. \quad (1.8)$$

FGLS ocena za  $\boldsymbol{\beta}$  je

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^N \mathbf{X}_i^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^N \mathbf{X}_i^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y}_i \right), \quad (1.9)$$

ki jo v matrični obliki zapišemo kot  $\hat{\boldsymbol{\beta}} = [\mathbf{X}^T (\mathbf{I}_N \otimes \hat{\boldsymbol{\Omega}}^{-1}) \mathbf{X}]^{-1} [\mathbf{X}^T (\mathbf{I}_N \otimes \hat{\boldsymbol{\Omega}}^{-1}) \mathbf{Y}]$ . Ker je GLS cenilka dosledna in, ker  $\hat{\boldsymbol{\Omega}}$  konvergira k  $\boldsymbol{\Omega}$ , je tudi FGLS cenilka dosledna. Velja še več, cenilki FGLS in GLS imata v limiti enako porazdelitev.

Zapišimo asimptotično porazdelitev FGLS cenilke:

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{X}_i^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{v}_i \right). \quad (1.10)$$

Velja

$$N^{-1/2} \sum_{i=1}^N \mathbf{X}_i^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{v}_i - N^{-1/2} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i = \left[ N^{-1/2} \sum_{i=1}^N (\mathbf{v}_i \otimes \mathbf{X}_i)^T \right] \cdot \text{vec}(\hat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1}).$$

Po predpostavki GLS.1 in centralnem limitnem izreku je  $N^{-1/2} \sum_{i=1}^N (\mathbf{v}_i \otimes \mathbf{X}_i) = O_p(1)$ . Ker je

$$O_p(1) \cdot o_p(1) = o_p(1), \text{ sledi}$$

$$N^{-1/2} \sum_{i=1}^N \mathbf{X}_i^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{v}_i = N^{-1/2} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i + o_p(1).$$

Na podoben način pokažemo, da  $N^{-1} \sum_{i=1}^N \mathbf{X}_i^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i - N^{-1} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i = o_p(1)$ . Sledi

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right)^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i \right) + o_p(1). \quad (1.11)$$

Prvi člen na desni v enačbi (1.11) je kar  $\sqrt{N}(\boldsymbol{\beta}^* - \boldsymbol{\beta})$ , kjer je  $\boldsymbol{\beta}^*$  GLS cenilka. Enačbo (1.11) zapišemo kot

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = o_p(1) \quad (1.12)$$

kar kaže, da sta  $\hat{\boldsymbol{\beta}}$  in  $\boldsymbol{\beta}^*$   $\sqrt{N}$ -ekvivalentni. To pomeni, da za asimptotično sklepanje o  $\boldsymbol{\beta}$  na podlagi  $\hat{\boldsymbol{\beta}}$  lahko namesto  $\boldsymbol{\Omega}$  uporabljamo  $\hat{\boldsymbol{\Omega}}$ . Pri tem mora asimptotični približek dovolj dobro ustrezati dejanski porazdelitvi  $\boldsymbol{\beta}^*$ .

IZREK 1.6 (Asimptotična normalnost FGLS): Po predpostavki GLS.1 in GLS.2 velja

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}),$$

kjer je  $\mathbf{A}$  definiran kot v (1.6) in  $\mathbf{B}$  kot v (1.7).

Za asimptotično učinkovitost, ki jo omogoča FGLS ocena, je potrebna dodatna predpostavka.

Predpostavka GLS.3 (Sistemska homoskedastičnost):  $E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i \mathbf{v}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i) = E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i)$ ,  
kjer je  $\boldsymbol{\Omega} \equiv E(\mathbf{v}_i \mathbf{v}_i^T)$ .

Predpostavka GLS.3 je zapisana v najšibkejši obliki. Predpostavko lahko zapišemo tudi kot  $\mathbf{B} = \mathbf{A}$ , kar poenostavi asimptotično varianco iz Izreka 1.5. Zadostni (in ne potrebni) pogoj za predpostavko GLS.3 poenostavi njeno interpretacijo:

$$E(\mathbf{v}_i \mathbf{v}_i^T | \mathbf{X}_i) = E(\mathbf{v}_i \mathbf{v}_i^T). \quad (1.13)$$

Če velja  $E(\mathbf{v}_i | \mathbf{X}_i) = \mathbf{0}$ , potem je predpostavka (1.13) enakovredna  $Var(\mathbf{v}_i | \mathbf{X}_i) = Var(\mathbf{v}_i) = \boldsymbol{\Omega}$  kar pomeni, da morajo biti variance in kovariance elementov  $\mathbf{v}_i$  konstantne pogojno na  $\mathbf{X}_i$ . Ob upoštevanju (1.3) poenostavimo predpostavko GLS.3:

$$\begin{aligned} E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i \mathbf{v}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i) &= E[E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{v}_i \mathbf{v}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i | \mathbf{X}_i)] = \\ &= E[\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} E(\mathbf{v}_i \mathbf{v}_i^T | \mathbf{X}_i) \boldsymbol{\Omega}^{-1} \mathbf{X}_i] = \\ &= E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \mathbf{X}_i) = \\ &= E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i). \end{aligned}$$

Izraz (1.13) olajša prikaz predpostavke o homoskedastičnosti, vendar je v določenih primerih premočna, zato bomo v naslednjem izreku uporabili predpostavko GLS.3:

IZREK 1.6: Po predpostavkah GLS.1 do GLS.3 je asimptotična varianca FGLS cenilke

$$A \text{ var}(\hat{\boldsymbol{\beta}}) = \frac{\mathbf{A}^{-1}}{N} \equiv \frac{E(\mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i)^{-1}}{N}.$$

Oceno za  $A \text{ var}(\hat{\boldsymbol{\beta}})$  dobimo na podlagi dosledne ocene  $\mathbf{A}$  :

$$A \text{ var}(\hat{\boldsymbol{\beta}}) = \frac{\hat{\mathbf{A}}^{-1}}{N} = \left( \sum_{i=1}^N \mathbf{X}_i^T \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right)^{-1}.$$

Pokazati se da, da je, ob upoštevanju predpostavk GLS.1, POLS.2, GLS.2 in (1.13), FGLS cenilka bolj učinkovita kot katerakoli druga cenilka, ki temelji na predpostavki ortogonalnosti  $E(\mathbf{X}_i \otimes \mathbf{v}_i) = 0$ .

## 1.2 Linearni model za panelne podatke z neopazovanimi učinki

Predpostavka nekoreliranosti motnje s pojasnjevalnimi spremenljivkami v istem obdobju je za določene vrste panelnih podatkov prestroga. V tem poglavju obravnavamo linearne modele, ki eksplicitno vsebujejo neopazovani presečni učinek, ki se v času ne spreminja. Zapišimo osnovno obliko modela z neopazovanimi učinki za naključno presečno opazovanje  $i$  :

$$y_{it} = \mathbf{x}_{it} \boldsymbol{\beta} + c_i + u_{it}, \quad t = 1, 2, \dots, T \quad (1.14)$$

kjer je  $\mathbf{x}_{it}$  vektor  $1 \times K$  opazovanih pojasnjevalnih spremenljivk,  $c_i$  pa je neopazovani (lahko slučajni) učinek, ki se v času ne spreminja. Glede na vrsto panelnih podatkov obravnavamo  $c_i$  kot slučajno spremenljivko, ali pa kot parameter, ki je predmet ocenjevanja. Uporabljeni model posledično imenujemo model s slučajnimi učinki, oz. model s fiksnimi učinki.

Za ocenjevanje parametrov modela (1.14) lahko, ob upoštevanju določenih predpostavk, uporabimo metodo POLS. Zapišimo model kot

$$y_{it} = \mathbf{x}_{it} \boldsymbol{\beta} + v_{it}, \quad t = 1, 2, \dots, T \quad (1.15)$$

kjer je  $v_{it} \equiv c_i + u_{it}$ ,  $t = 1, 2, \dots, T$  strukturirana motnja. Za vsak  $t$  je  $v_{it}$  vsota neopazovanega učinka in preostanka motnje. Ocena POLS je dosledna, če  $E(\mathbf{x}_{it}^T v_{it}) = 0$ ,  $t = 1, 2, \dots, T$ . Nekoreliranost med  $\mathbf{x}_{it}$  in  $v_{it}$  je izpolnjena, če:

$$E(\mathbf{x}_{it}^T u_{it}) = 0 \text{ in}$$

$$E(\mathbf{x}_{it}^T c_i) = 0, \quad t = 1, 2, \dots, T. \quad (1.16)$$

Tudi v primeru izpolnitve gornje predpostavke bo strukturirana motnja serijsko korelirana zaradi prisotnosti  $c_i$  v vsakem časovnem obdobju. Za statistično sklepanje je zato potrebna robustna ocena kovariančne matrike. Ker je  $v_{it}$  odvisen od  $c_i$  za vse  $t$ , korelacija med  $v_{it}$  in  $v_{is}$  ne pojenja z naraščanjem  $|t - s|$ , zato je za zanesljivost sklepov na podlagi asimptotičnih porazdelitev POLS ocen pomemben velik  $N$ , glede na fiksen  $T$ .

### 1.2.1 Model s slučajnimi učinki

Slučajni učinek  $c_i$  obravnavamo kot del motnje. Dodaten pogoj glede na predhodnji model s strukturo motnje, ob ortogonalnosti  $c_i$  in  $\mathbf{x}_{it}$ , je stroga eksogenost, ki je potrebna za doslednost FGLS ocene parametrov.

Predpostavka RE.1:

$$(a) \quad E(u_{it} | \mathbf{x}_i, c_i) = 0, \quad t = 1, 2, \dots, T$$

$$(b) \quad E(c_i | \mathbf{x}_i) = E(c_i) = 0$$

kjer je  $\mathbf{x}_i \equiv (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT})$ .

Prvi del predpostavke RE.1 zagotavlja nekoreliranost pojasnjevalnih spremenljivk z  $u_{it}$  za vsako časovno obdobje:  $E(\mathbf{x}_{is}^T u_{it}) = 0$ ,  $s, t = 1, 2, \dots, T$ . Drugi del predpostavke RE.1 je ortogonalnost med  $c_i$  in vsakim  $\mathbf{x}_{it}$ , ki je zagotovljena, če predpostavimo fiksne  $\mathbf{x}_{it}$  in  $E(c_i) = 0$ , ali ob predpostavki neodvisnosti  $c_i$  od  $\mathbf{x}_{it}$ .

Model s slučajnimi učinki uporablja serijsko korelacijo motnje  $v_{it} = c_i + u_{it}$  v ocenjevanju z GLS metodo. Doslednost GLS ocen je zagotovljena s strogo eksogenostjo pojasnjevalnih spremenljivk in sestavljene motnje. Po predpostavki RE.1 lahko zapišemo

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it} \quad (1.17)$$

$$E(v_{it} | \mathbf{x}_i) = 0, \quad t = 1, 2, \dots, T. \quad (1.18)$$

Enačba (1.18) je predpostavka stroge eksogenosti SGLS.1 (po veljavnih predpostavkah za GLS) za model (1.17), ki ga zato lahko ocenimo z metodo GLS. Zapišimo model (1.15) za vsa časovna obdobja  $T$ :

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i. \quad (1.19)$$

$\mathbf{v}_i$  lahko zapišemo kot  $\mathbf{v}_i = c_i\mathbf{j}_T + \mathbf{u}_i$ , kjer je  $\mathbf{j}_T$   $T \times 1$  vektor enk. Definirajmo (brezpogojno) kovariančno matriko motnje  $\mathbf{v}_i$  kot

$$\boldsymbol{\Omega} \equiv E(\mathbf{v}_i\mathbf{v}_i^T) \quad (1.20)$$

pozitivno definitno matriko  $T \times T$ .  $\boldsymbol{\Omega}$  je enaka za vse presečne enote  $i$  zaradi neodvisnosti in enake porazdeljenosti presečnih enot. Za doslednost GLS je potrebna še predpostavka nesingularnosti:

Predpostavka RE.2:  $\text{rang}E(\mathbf{X}_i^T\boldsymbol{\Omega}^{-1}\mathbf{X}_i) = K$ .

Predpostavki RE.1 in RE.2 zadoščata za doslednost GLS in izvedljivih GLS (FGLS) (po veljavnih predpostavkah za GLS). Z dodatnimi pogoji določimo strukturo kovariančne matrike  $\boldsymbol{\Omega}$ . Predpostavimo konstantno brezpogojno varianco  $u_{it}$  v času  $t$ :

$$E(u_{it}^2) = \sigma_u^2, \quad t = 1, 2, \dots, T. \quad (1.21)$$

Druga predpostavka je serijska nekoreliranost  $u_{it}$ :

$$E(u_{it}u_{is})=0 \quad \forall t \neq s. \quad (1.22)$$

(1.21) in (1.22) določata variance in kovariance  $\mathbf{v}_i$  in s tem obliko kovariančne matrike  $\mathbf{\Omega}$ .

Po predpostavki RE.1a velja  $E(c_i u_{it})=0$ ,  $t=1,2,\dots,T$ , zato

$$E(v_{it}^2) = E(c_i^2) + 2E(c_i u_{it}) + E(u_{it}^2) = \sigma_c^2 + \sigma_u^2,$$

kjer je  $\sigma_c^2 = E(c_i^2)$ . Sledi, za  $\forall t \neq s$ ,

$$E(v_{it}v_{is}) = E[(c_i + u_{it})(c_i + u_{is})] = E(c_i^2) = \sigma_c^2.$$

Iz predpostavk RE.1, (1.21) in (1.22) sledi struktura kovariančne matrike  $\mathbf{\Omega}$ :

$$\mathbf{\Omega} = E(\mathbf{v}_i \mathbf{v}_i^T) = \begin{bmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \sigma_c^2 \\ \sigma_c^2 & \cdots & \cdots & \sigma_c^2 + \sigma_u^2 \end{bmatrix}, \quad (1.23)$$

ki jo v lahko zapišemo tudi kot:

$$\mathbf{\Omega} = \sigma_u^2 \mathbf{I}_T + \sigma_c^2 \mathbf{j}_T \mathbf{j}_T^T. \quad (1.24)$$

Kovariančna matrika  $\mathbf{\Omega}$  ima v (1.24) strukturo slučajnih učinkov. V splošnem je struktura kovariančne matrike GLS odvisna od  $T(T+1)/2$  varianc in kovarianc, kovariančna matrika s strukturo slučajnih učinkov pa je odvisna samo od parametrov  $\sigma_c^2$  in  $\sigma_u^2$ , ne glede na dimenzijo  $T$ . Korelacija med  $v_{it}$  in  $v_{is}$  ni odvisna od  $|t-s|$ :  $Corr(v_{is}, v_{it}) = \sigma_c^2 / (\sigma_c^2 + \sigma_u^2) \geq 0$ ,  $s \neq t$ . Korelacija predstavlja tudi razmerje med varianco  $c_i$  in varianco sestavljene motnje, ki je mera za relativen vpliv neopazovanega vpliva  $c_i$ .



Za asimptotično učinkovitost ocene FGLS je potrebna konstantna kovariančna matrika  $\mathbf{v}_i$  pogojno na  $\mathbf{x}_i$  :

$$E(\mathbf{v}_i \mathbf{v}_i^T | \mathbf{x}_i) = E(\mathbf{v}_i \mathbf{v}_i^T). \quad (1.25)$$

Predpostavke (1.21), (1.22) in (1.25) združimo v tretjo predpostavko modela slučajnih učinkov:

Predpostavka RE.3:

$$(a) E(\mathbf{u}_i \mathbf{u}_i^T | \mathbf{x}_i, c_i) = \sigma_u^2 \mathbf{I}_T,$$

$$(b) E(c_i^2 | \mathbf{x}_i) = \sigma_c^2.$$

Na podlagi predpostavke RE.3 velja  $E(u_{it}^2 | \mathbf{x}_i, c_i) = \sigma_u^2$ ,  $t = 1, 2, \dots, T$ , iz česar sledi (1.21), in  $E(u_{it} u_{is} | \mathbf{x}_i, c_i) = 0$ ,  $t \neq s$ ,  $t, s = 1, 2, \dots, T$ , iz česar sledi (1.22). Poleg tega zagotavlja predpostavka RE.3a še konstantne pogojne variance in ničelne pogojne kovariance. Po predpostavkah RE.1b in RE.3b velja  $Var(c_i | \mathbf{x}_i) = Var(c_i)$ , kar je predpostavka homoskedastičnosti neopazovanega učinka  $c_i$ . Na podlagi predpostavke RE.3 velja (1.25) in  $\mathbf{\Omega}$  ima strukturo (1.23).

Če razpolagamo z robustno oceno kovariančne matrike  $\hat{\mathbf{\Omega}}$ , lahko z FGLS ocenimo model s slučajnimi učinki:

$$\hat{\boldsymbol{\beta}}_{RE} = \left( \sum_{i=1}^N \mathbf{X}_i^T \hat{\mathbf{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^N \mathbf{X}_i^T \hat{\mathbf{\Omega}}^{-1} \mathbf{y}_i \right). \quad (1.26)$$

Po predpostavki RE.3 je cenilka modela s slučajnimi učinki učinkovita v razredu cenilk doslednih po predpostavki  $E(\mathbf{v}_i | \mathbf{x}_i) = 0$ .

### 1.2.2 Model s fiksnimi učinki

Obravnavamo linearni model z neopazovanimi učinki za  $T$  časovnih obdobij:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad t = 1, 2, \dots, T. \quad (1.27)$$

Model s slučajnimi učinki za učinkovito oceno  $\boldsymbol{\beta}$  vključi  $c_i$  v motnjo ob predpostavki, da je  $c_i$  ortogonalen na  $\mathbf{x}_{it}$ . Pri tem upošteva serijsko korelacijo v sestavljeni motnji  $v_{it} = c_i + u_{it}$  z GLS oceno. V mnogih primerih pa je smisel uporabe panelnih podatkov ravno v tem, da dopustimo koreliranost med  $c_i$  in  $\mathbf{x}_{it}$ .

Enačbe za  $T$  časovnih obdobij iz modela (1.27) lahko zapišemo kot

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{j}_T + \mathbf{u}_i, \quad (1.28)$$

kjer je  $\mathbf{j}_T$   $T \times 1$  vektor enk. Enačba (1.28) predstavlja slučajno izbrano presečno enoto. Prva predpostavka modela s fiksnimi učinki je stroga eksogenost pojasnjevalnih spremenljivk pogojno na  $c_i$ :

$$\text{Predpostavka FE.1: } E(u_{it} | \mathbf{x}_i, c_i) = 0, \quad t = 1, 2, \dots, T.$$

Predpostavka je identična prvemu delu predpostavke RE.1. Ohranjamo torej strogo eksogenost  $\{\mathbf{x}_{it} : t = 1, 2, \dots, T\}$  pogojno na neopazovani učinek. Bistvena razlika v primerjavi z modelom s slučajnimi učinki je, da ne predpostavimo RE.1b.  $E(c_i | \mathbf{x}_i)$  je lahko poljubna funkcija  $\mathbf{x}_i$ .

Analiza s fiksnimi učinki je glede doslednosti bolj robustna kot analiza s slučajnimi učinki, vendar brez dodatnih predpostavk v  $\mathbf{x}_{it}$  ne moremo vključiti spremenljivk, ki se v času ne spreminjajo. Ker je  $c_i$  lahko poljubno koreliran z vsakim elementom  $\mathbf{x}_{it}$  ne moremo ločiti med učinki časovno nespremenljivih opazovanih spremenljivk in časovno nespremenljivih neopazovanih  $c_i$ .

Namen transformacija enačb je eliminacije neopazovanih učinkov  $c_i$ , ki omogoča oceno  $\beta$  po predpostavki FE.1. V nadaljevanju bomo obravnavali transformacijo fiksnih učinkov (within transformacija). Najprej naredimo povpreček enačbe (1.27) po času  $t=1,2,\dots,T$  tako, da dobimo presečno enačbo

$$\bar{y}_i = \bar{\mathbf{x}}_i \beta + c_i + \bar{u}_i, \quad (1.29)$$

kjer je  $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$ ,  $\bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^T \mathbf{x}_{it}$  in  $\bar{u}_i = T^{-1} \sum_{t=1}^T u_{it}$ . Če odštejemo enačbo (1.29) od enačbe (1.27) za vsak  $t$ , dobimo transformirano enačbo fiksnih učinkov,

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \beta + u_{it} - \bar{u}_i,$$

ki jo zapišemo kot

$$\dot{y}_{it} = \dot{\mathbf{x}}_{it} \beta + \dot{u}_{it}, \quad t = 1, 2, \dots, T \quad (1.30)$$

kjer je  $\dot{y}_{it} \equiv y_{it} - \bar{y}_i$ ,  $\dot{\mathbf{x}}_{it} \equiv \mathbf{x}_{it} - \bar{\mathbf{x}}_i$  in  $\dot{u}_{it} \equiv u_{it} - \bar{u}_i$ . Z odštevanjem časovnega povprečja smo eliminirali individualni specifični učinek  $c_i$ .

Sedaj lahko enačbo (1.30) ocenimo s POLS metodo. Pri tem je enačba (1.30) ocenjevana enačba, interpretacija  $\beta$  pa izhaja iz osnovne enačbe pogojnih pričakovanj  $E(y_{it} | \mathbf{x}_i, c_i) = E(y_{it} | \mathbf{x}_{it}, c_i) = \mathbf{x}_{it} \beta + c_i$ .

Najprej preverimo doslednost POLS ocene enačbe (1.30) na podlagi predpostavke POLS.1:

$$E(\dot{\mathbf{x}}_{it}^T \dot{u}_{it}) = 0, \quad t = 1, 2, \dots, T. \quad (1.31)$$

Za vsak  $t$  lahko levo stran enačbe (1.31) zapišemo kot  $E((\mathbf{x}_{it} - \bar{\mathbf{x}}_i)^T (u_{it} - \bar{u}_i))$ . Po predpostavki FE.1 je  $u_{it}$  nekoreliran z  $\mathbf{x}_{is}$  za vsak  $s, t = 1, 2, \dots, T$ . Sledi, da sta  $u_{it}$  in  $\bar{u}_i$  nekorelirana z  $\mathbf{x}_{it}$  in

$\bar{\mathbf{x}}_i$  za  $t = 1, 2, \dots, T$ . Po predpostavki FE.1 torej velja (1.31) zato pričakujemo, da bo POLS ocena enačbe (1.30) dosledna. Po predpostavki FE.1 velja več kot samo pogoj (1.31), namreč:  $E(\ddot{u}_{it} | \mathbf{x}_i) = E(u_{it} | \mathbf{x}_i) - E(\bar{u}_i | \mathbf{x}_i) = 0$ , kar implicira, da je  $E(\ddot{u}_{it} | \bar{\mathbf{x}}_{i1}, \dots, \bar{\mathbf{x}}_{iT}) = 0$ , ker je vsak  $\ddot{\mathbf{x}}_{it}$  funkcija  $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})$ . S tem smo pokazali, da  $\ddot{\mathbf{x}}_{it}$  zadošča obliki pogojnih pričakovanj predpostavke stroge eksogenosti modela (1.30). Pokažemo lahko, da je cenilka fiksnih učinkov za  $\boldsymbol{\beta}$  po predpostavki FE.1 nepristranska.

Cenilka fiksnih učinkov  $\hat{\boldsymbol{\beta}}_{FE}$  je POLS cenilka iz regresije  $\ddot{y}_{it}$  na  $\ddot{\mathbf{x}}_{it}$ ,  $t = 1, 2, \dots, T$ ;  $i = 1, 2, \dots, N$ . Zapišimo enačbo (1.30) za vsa časovna obdobja

$$\ddot{\mathbf{y}}_i = \ddot{\mathbf{X}}_i \boldsymbol{\beta} + \ddot{\mathbf{u}}_i, \quad (1.32)$$

kjer je  $\ddot{\mathbf{y}}_i$   $T \times 1$ ,  $\ddot{\mathbf{X}}_i$  je  $T \times K$  in  $\ddot{\mathbf{u}}_i$  je  $T \times 1$ . Izraz (1.32) dobimo, če (1.28) pomnožimo z matriko, ki od vsake presečne enote odšteje časovno povprečje. Definirajmo  $\mathbf{Q}_T = \mathbf{I}_T - \mathbf{j}_T (\mathbf{j}_T' \mathbf{j}_T)^{-1} \mathbf{j}_T'$ , ki je simetrična, idempotentna matrika z rangom  $T-1$ :  $\mathbf{Q}_T \mathbf{j}_T = 0$ ,  $\mathbf{Q}_T \mathbf{y}_i = \ddot{\mathbf{y}}_i$ ,  $\mathbf{Q}_T \mathbf{X}_i = \ddot{\mathbf{X}}_i$  in  $\mathbf{Q}_T \mathbf{u}_i = \ddot{\mathbf{u}}_i$ , zato pri množenju (1.28) z leve z  $\mathbf{Q}_T$  dobimo (1.32).

Za asimptotično obnašanje cenilke fiksnih učinkov potrebujemo dodatno predpostavko o rangju matrike pojasnjevalnih spremenljivk, ki imajo odšteta časovna povprečja:

$$\text{Predpostavka FE.2: } \text{rang} \left( \sum_{t=1}^T E(\ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it}) \right) = \text{rang} [E(\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i)] = K.$$

Če  $\mathbf{x}_{it}$  vsebuje element, ki se ne spreminja v času za vsak  $i$ , potem je ustrezní element  $\ddot{\mathbf{x}}_{it}$  enak 0 za vse  $t$  in za vsak  $i$ . V tem primeru  $\ddot{\mathbf{X}}_i$  vsebuje stolpec ničel za vse  $i$  in predpostavka FE.2 ni izpolnjena. Predpostavka FE.2 pri analizi fiksnih učinkov ne dopušča uporabe spremenljivk, ki so časovno nespremenljive (razen v povezavi s časovnimi dummy spremenljivkami).

Cenilko fiksnih učinkov, ki jo imenujemo tudi within cenilka, ker izkorišča časovno variiranje znotraj vsake presečne enote, zapišemo kot

$$\hat{\boldsymbol{\beta}}_{FE} = \left( \sum_{i=1}^N \ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i \right)^{-1} \left( \sum_{i=1}^N \ddot{\mathbf{X}}_i^T \ddot{\mathbf{y}}_i \right) = \left( \sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}_{it}^T \ddot{\mathbf{x}}_{it} \right)^{-1} \left( \sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}_{it}^T \ddot{y}_{it} \right). \quad (1.33)$$

Cenilko, ki izkorišča samo variiranje med posameznimi presečnimi enotami, imenujemo between cenilka. V tem primeru uporabimo OLS metodo na časovno povprečni enačbi (1.29). Between cenilka ni dosledna po predpostavki FE.1, ker  $E(\ddot{\mathbf{x}}_i^T c_i)$  ni vedno enako nič. Doslednost je zagotovljena z RE.1 in standardnim pogojem polnega ranga, vendar je between cenilka manj učinkovita kot cenilka slučajnih učinkov, ker je izgubljena informacija časovnih vrst.

Pokažemo lahko, da je, po predpostavkah FE.1 in FE.2, cenilka  $\hat{\boldsymbol{\beta}}_{FE}$  nepristranska pogojno na  $\mathbf{X}$ .

Naslednja predpostavka zagotavlja, da je cenilka fiksnih učinkov učinkovita glede na predpostavko FE.1:

$$\text{Predpostavka FE.3: } E(\mathbf{u}_i \mathbf{u}_i^T | \mathbf{x}_i, c_i) = \sigma_u^2 \mathbf{I}_T.$$

Predpostavka FE.3 je enaka predpostavki RE.3a. Ker je po predpostavki FE.1  $E(\mathbf{u}_i | \mathbf{x}_i, c_i) = 0$ , lahko FE.3 zapišemo tudi kot  $Var(\mathbf{u}_i | \mathbf{x}_i, c_i) = \sigma_u^2 \mathbf{I}_T$ , če velja FE.1. Predpostavko FE.3 lahko obravnavamo v dveh delih. Velja enakost med pogojno in brezpogojno kovariančno matriko ostankov  $E(\mathbf{u}_i \mathbf{u}_i^T | \mathbf{x}_i, c_i) = E(\mathbf{u}_i \mathbf{u}_i^T)$ , poleg tega ima brezpogojna kovariančna matrika  $E(\mathbf{u}_i \mathbf{u}_i^T)$  obliko  $\sigma_u^2 \mathbf{I}_T$ . To pomeni, da ima motnja  $u_{it}$  konstantno varianco v času  $t$  in je serijsko nekorelirana.

Za asimptotično učinkovitost POLS ocene morajo biti  $\{u_{it} : t = 1, 2, \dots, T\}$  homoskedastični v času  $t$  in serijsko nekorelirani. Varianco  $u_{it}$  zapišemo kot

$$E(\ddot{u}_{it}^2) = E[(u_{it} - \bar{u}_i)^2] = E(u_{it}^2) + E(\bar{u}_i^2) - 2E(u_{it} \bar{u}_i) = \sigma_u^2 + \frac{\sigma_u^2}{T} - \frac{2\sigma_u^2}{T} = \sigma_u^2 \left(1 - \frac{1}{T}\right), \quad (1.34)$$

s čimer smo pokazali homoskedastičnost v času  $t$ . Kovarianco med  $\ddot{u}_{it}$  in  $\ddot{u}_{is}$  za  $t \neq s$  zapišemo kot

$$E(\ddot{u}_{it}\ddot{u}_{is}) = E(u_{it}u_{is}) - E(u_{it}\bar{u}_i) - E(u_{is}\bar{u}_i) + E(\bar{u}_i^2) = 0 - \frac{\sigma_u^2}{T} - \frac{\sigma_u^2}{T} + \frac{\sigma_u^2}{T} = -\frac{\sigma_u^2}{T} < 0. \quad (1.35)$$

Iz (1.34) in (1.35) sledi za  $t \neq s$

$$\text{Corr}(\ddot{u}_{it}, \ddot{u}_{is}) = \frac{-1}{T-1}, \quad (1.36)$$

kar kaže, da je motnja, od katere je odšteto njeno časovno povprečje, negativno serijsko korelirana. Za velike  $T$  korelacija teži k 0.

Izkaže se, da je serijska koreliranost  $\ddot{u}_{it}$ , ki nastane zaradi odštetja časovnega povprečja, ob upoštevanju predpostavke FE.3, manjšega pomena. Po preureditvi (1.33) zapišemo  $\hat{\boldsymbol{\beta}}_{FE}$  kot

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) = \left( N^{-1} \sum_{i=1}^N \ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i \right)^{-1} \left( N^{-1/2} \sum_{i=1}^N \ddot{\mathbf{X}}_i^T \mathbf{u}_i \right), \quad (1.37)$$

kjer smo upoštevali  $\ddot{\mathbf{X}}_i^T \ddot{\mathbf{u}}_i = \mathbf{X}_i^T \mathbf{Q}_T \mathbf{u}_i = \ddot{\mathbf{X}}_i^T \mathbf{u}_i$ . Po predpostavki FE.3  $E(\mathbf{u}_i \mathbf{u}_i^T | \ddot{\mathbf{X}}_i) = \sigma_u^2 \mathbf{I}_T$ . Iz analize POLS sledi asimptotična porazdelitev  $\hat{\boldsymbol{\beta}}_{FE}$

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) \stackrel{a}{\sim} N\left( \mathbf{0}, \sigma_u^2 [E(\ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i)]^{-1} \right). \quad (1.38)$$

Asimptotična varianca  $\hat{\boldsymbol{\beta}}_{FE}$ :

$$A \text{ var}(\hat{\boldsymbol{\beta}}_{FE}) = \frac{\sigma_u^2 [E(\ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i)]^{-1}}{N}. \quad (1.39)$$

Če je  $\hat{\sigma}_u^2$  dosledna cenilka za  $\sigma_u^2$ , potem lahko (1.39) ocenimo tako, da  $E(\ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i)$  zamenjamo z vzorčno vrednostjo  $N^{-1} \sum_{i=1}^N \ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i$ :

$$A \text{ var}(\hat{\boldsymbol{\beta}}_{FE}) = \hat{\sigma}_u^2 \left( \sum_{i=1}^N \ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i \right)^{-1} = \hat{\sigma}_u^2 \left( \sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}_{it}^T \ddot{\mathbf{x}}_{it} \right)^{-1}. \quad (1.40)$$

## 2. Hausmanov specifikacijski test za regresijski model s strukturo motnje

Glede na obravnavo neopazovanih učinkov  $c_i$  ločimo dve alternativni specifikaciji modela za panelne podatke. Ocena parametrov modela s fiksnimi učinki je dosledna tudi, če sta  $c_i$  in  $\mathbf{x}_{it}$  korelirana, ocena parametrov modela s slučajnimi učinki pa je v tem primeru nedosledna. Statistično značilno razliko v ocenah parametrov interpretiramo kot kršitev predpostavke RE.1b.

Hausmanov specifikacijski test temelji na predpostavki, da v ničelni hipotezi obstaja dosledna, asimptotično normalna in asimptotično učinkovita cenilka, kjer učinkovitost pomeni doseganje asimptotične Cramér-Raove spodnje meje. V alternativni hipotezi napačne specifikacije pa je cenilka pristranska in nedosledna. Za specifikacijski test je potreben obstoj cenilke, ki je dosledna tudi v primeru napačne specifikacije, vendar ni asimptotično učinkovita v ničelni hipotezi. Razlika  $\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}$ , kjer je  $\hat{\boldsymbol{\beta}}_{RE}$  učinkovita v ničelni hipotezi in  $\hat{\boldsymbol{\beta}}_{FE}$  dosledna v alternativni hipotezi, ima verjetnostno limito enako nič, če ni prisotna napačna specifikacija. V primeru napačne specifikacije se bo verjetnostna limita razlikovala od nič. Za definiranje testne statistike je ključna naslednja lema:

LEMA 2.1: Če velja ničelna domneva, sta cenilki  $\hat{\boldsymbol{\beta}}_{RE}$  in  $\hat{\boldsymbol{\beta}}_{FE}$  dosledni in asimptotično normalno porazdeljeni,  $\hat{\boldsymbol{\beta}}_{RE}$  dosega asimptotično Cramér-Rao spodnjo mejo,  $\sqrt{N}(\hat{\boldsymbol{\beta}}_{RE} - \boldsymbol{\beta}) \overset{a}{\sim} N(0, V_{RE})$  in  $\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) \overset{a}{\sim} N(0, V_{FE})$ . Naj bo  $\hat{\boldsymbol{q}} = \hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}$ . Potem imata

limitni porazdelitvi slučajnih spremenljivk  $\sqrt{N}(\hat{\beta}_{RE} - \beta)$  in  $\sqrt{N}\hat{q}$  kovariančno matriko enako 0,  $C(\hat{\beta}_{RE}, \hat{q}) = 0$ .

Dokaz: Naj  $\hat{\beta}_{RE}$  in  $\hat{q}$  ne bosta ortogonalna. Ker je  $\text{plim}\hat{q} = 0$  definirajmo novo cenilko  $\hat{\beta}_O = \hat{\beta}_{RE} + rA\hat{q}$ , kjer je  $r$  skalar in  $A$  poljubna matrika. Nova cenilka je dosledna in asimptotično normalna z asimptotično varianco

$$V(\hat{\beta}_O) = V(\hat{\beta}_{RE}) + rAC(\hat{\beta}_{RE}, \hat{q}) + rC^T(\hat{\beta}_{RE}, \hat{q})A^T + r^2AV(\hat{q})A^T. \quad (2.1)$$

Razlika med asimptotično varianco nove cenilke in asimptotično učinkovite cenilke  $\hat{\beta}_{RE}$  je

$$F(r) = V(\hat{\beta}_O) - V(\hat{\beta}_{RE}) = rAC + rC^T A^T + r^2AV(\hat{q})A^T. \quad (2.2)$$

Odvajajmo razliko (2.2) po  $r$ :

$$F'(r) = AC + C^T A^T + 2rAV(\hat{q})A^T. \quad (2.3)$$

Naj bo  $A = -C^T$  in ker je  $C$  simetrična dobimo

$$F'(r) = -2C^T C + 2rC^T V(\hat{q})C. \quad (2.4)$$

Pri  $r = 0$  je  $F'(0) = -2C^T C$  nepozitivno definitna. Ker  $F(0) = 0$  je za majhne  $r$   $F(r) < 0$  kar je protislovje, če  $C = C(\hat{\beta}_{RE}, \hat{q})$  ni enako 0.

POMOŽNI IZREK 2.1:  $V(\hat{q}) = V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE})$  je nenegativno definitna.

Dokaz: Ker  $\hat{q} + \hat{\beta}_{RE} = \hat{\beta}_{FE}$  iz Leme 2.1 sledi  $V(\hat{q}) + V(\hat{\beta}_{RE}) = V(\hat{\beta}_{FE})$  in  $\hat{\beta}_{RE}$  dosega asimptotično Cramér-Rao spodnjo mejo.



Hausmanovo statistiko definiramo kot Mahalanobisovo razdaljo  $m = N\hat{q}^T V(\hat{q})^{-1} \hat{q}$ . Ker je  $C(\hat{\beta}_{RE}, \hat{q}) = 0$ , velja  $V(\hat{q}) = V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE})$  in ker je  $\hat{q}$  razlika normalnih spremenljivk, je Hausmanova statistika  $m$  asimptotično porazdeljena po  $\chi_K^2$ , kjer je  $K$  dimenzija vektorja  $\beta$ .

IZREK 2.1: V ničelni hipotezi je testna statistika  $m = N\hat{q}^T \hat{V}(\hat{q})^{-1} \hat{q} \stackrel{a}{\sim} \chi_K^2$ , kjer je  $\hat{V}(\hat{q})$  dosledna ocena (po ničelni hipotezi)  $V(\hat{q})$  glede na  $\hat{\beta}_{FE}$  in  $\hat{\beta}_{RE}$ .

Dokaz:  $\sqrt{N}\hat{q} = \sqrt{N}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \stackrel{a}{\sim} N(0, V(\hat{q}))$ , zato je  $N\hat{q}^T \hat{V}(\hat{q})^{-1} \hat{q}$  asimptotično porazdeljena po  $\chi_K^2$ , ker je  $\hat{V}(\hat{q})$  dosledna ocena  $V(\hat{q})$ .

Predpostavka stroge eksogenosti RE.1a mora veljati v ničelni in alternativni hipotezi. Korelacija med  $x_{is}$  in  $u_{it}$  za vsak  $s$  in  $t$  povzroči nedoslednost ocen parametrov modela s fiksnimi in slučajnimi učinki. Veljavnost predpostavke RE.3 v ničelni hipotezi zagotavlja učinkovitost cenilke slučajnih učinkov glede na cenilko fiksnih učinkov.

### 3. Bootstrap

V nalogi želimo, med drugim, z Monte Carlo metodo analizirati lastnosti intervalov zaupanja bootstrap porazdelitve Hausmanove statistike glede na intervale zaupanja asimptotične porazdelitve Hausmanove statistike, zato bomo uporabili Edgeworthov razvoj. V nadaljevanju bomo navedli potrebne elemente Edgeworthovega in Cornish-Fisherjevega razvoja, ki jih bomo potrebovali za definiranje napak pokritja intervalov zaupanja. Kjer obravnavamo omenjene razvoje porazdelitev sledimo zapisu, ki ga uporablja Hall (1992).

S pomočjo vzorca, ki ga slučajno izberemo iz populacije, tvorimo s ponovljenim slučajnim vzorčenjem iz originalnega vzorca bootstrap vzorce. Vzorec  $\mathfrak{N} = \{X_1, X_2, \dots, X_n\}$  je izbor  $n$  števil (ponavadi skalarjev, lahko tudi vektorjev), ki so slučajno izbrani iz populacije. Slučajno pomeni, da so  $X_i$  neodvisno in enako porazdeljene slučajne spremenljivke, ki imajo populacijsko porazdelitveno funkcijo  $F$ . Pri neparametričnem ponovljenem vzorčenju

predstavlja  $\aleph^* = \{X_1^*, X_2^*, \dots, X_n^*\}$  neurejen izbor  $n$  enot, ki so slučajno izbrani iz  $\aleph$  s ponavljanjem. Verjetnost, da bo  $X_i^*$  enak kateremukoli  $X_j$ , je enaka  $n^{-1}$ :

$$P(X_i^* = X_j | \aleph) = n^{-1}, 1 \leq i, j \leq n. \quad (3.1)$$

$X_i^*$ -ji so neodvisno in enako porazdeljeni pogojno na  $\aleph$ . Vzorec  $\aleph^*$  lahko vsebuje elemente, ki se ponavljajo.

Statistično sklepanje se pogosto nanaša na opis odnosa med vzorcem in populacijo, iz katere je bil vzorec potegnjeno. Če imamo dan funkcional  $f_t$  iz razreda  $\{f_t : t \in T\}$ , želimo določiti vrednost  $t_0$  iz  $t$ , ki reši enačbo

$$E\{f_t(F_0, F_1) | F_0\} = 0, \quad (3.2)$$

kjer  $F_0$  označuje populacijsko porazdelitveno funkcijo in  $F_1$  vzorčno empirično porazdelitveno funkcijo. Pogojevanje na  $F_0$  poudarja, da je pričakovanje glede na populacijsko porazdelitveno funkcijo  $F_0$ . Enačba (3.2) je populacijska enačba, ker za njeno rešitev potrebujemo lastnosti populacije.

Osnovna ideja bootstrapa je dobiti približno rešitev populacijske enačbe z nadomestitvijo  $(F_0, F_1)$  v (3.2) z  $(F_1, F_2)$ :

$$E\{f_t(F_1, F_2) | F_1\} = 0, \quad (3.3)$$

kjer je  $F_2$  empirična porazdelitvena funkcija vzorca izbranega iz  $F_1$ .

Enačba (3.3) je vzorčna enačba, ki jo lahko rešimo, če poznamo vzorčno porazdelitveno funkcijo  $F_1$ . Rešitev vzorčne enačbe naj bi bila čim boljši približek za rešitev populacijske enačbe, ker slednja ni poznana.

### 3.1 Edgeworthov razvoj

Edgeworthov razvoj je pripomoček za analiziranje porazdelitev testnih statistik. Iz vzorca velikosti  $n$  ocenimo parameter  $\hat{\theta}$ . Predpostavimo, da se  $n^{1/2}(\hat{\theta} - \theta_0)$  asimptotično normalno porazdeljuje s povprečjem 0 in varianco  $\sigma^2$ . Potem lahko, ob izpolnitvi predpostavk, porazdelitev  $n^{1/2}(\hat{\theta} - \theta_0)$  razvijemo v potenčno vrsto po  $n^{-1/2}$ :

$$P(n^{1/2}(\hat{\theta} - \theta_0)/\sigma \leq x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + \dots + n^{-j/2}p_j(x)\phi(x) + \dots, \quad (3.4)$$

kjer je  $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$  gostota standardne normalne porazdelitve in  $\Phi(x) = \int_{-\infty}^x \phi(u)du$  standardna normalna porazdelitev. Izraz (3.4) imenujemo Edgeworthov razvoj. Funkcije  $p_j$  so polinomi s koeficienti, ki so odvisni od kumulant od  $\hat{\theta} - \theta_0$ .

Kumulante  $\kappa_n$  so definirane z

$$\ln \chi(t) = \sum_{n=0}^{\infty} \kappa_n \frac{(it)^n}{n!},$$

kjer je

$$\chi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

karakteristična funkcija.

### 3.2 Splošne statistike

Najosnovnejša oblika Edgeworthovega razvoja se nanaša na primer, ko je  $\hat{\theta}$  vzorčna ocena povprečja in  $\theta_0$  populacijsko povprečje. Edgeworthov razvoj lahko uporabimo tudi za splošnejše statistike. Naj ima statistika  $S_n$  v limiti standardno normalno porazdelitev kot  $S_n = n^{1/2}(\hat{\theta} - \theta_0)/\sigma$ , ali  $S_n = n^{1/2}(\hat{\theta} - \theta_0)/\hat{\sigma}$ , kjer je  $\hat{\sigma}^2$  dosledna ocena asimptotične

variance  $\sigma^2$  od  $n^{1/2}\hat{\theta}$ . Zapišimo karakteristično funkcijo  $\chi_n$  za  $S_n$  in naj bo  $\kappa_{j,n}$  j-ta kumulanta. Potem je

$$\chi_n(t) = E(e^{itS_n}) = e^{\kappa_{1,n}it + \frac{1}{2}\kappa_{2,n}(it)^2 + \dots + \frac{1}{j!}\kappa_{j,n}(it)^j + \dots} \quad (3.5)$$

Kumulanta  $\kappa_{j,n}$  je v mnogo za uporabo zanimivih primerih reda  $n^{-\frac{(j-2)}{2}}$  in jo lahko razvijemo v potenčno vrsto po  $n^{-1}$ :

$$\kappa_{j,n} = n^{-\frac{(j-2)}{2}} (k_{j,1} + n^{-1}k_{j,2} + n^{-2}k_{j,3} + \dots), \quad j \geq 1, \quad (3.6)$$

kjer je  $k_{1,1} = 0$  in  $k_{2,1} = 1$ . Zadnji dve zvezi sta posledica  $\kappa_{1,n} = E(S_n) \rightarrow 0$  in  $\kappa_{2,n} = \text{var}(S_n) \rightarrow 1$ . Dokaz za splošen primer je zelo težaven (glej Hall, 1992, str. 52), v enostavnem primeru vsote neodvisnih slučajnih spremenljivk pa lahko (3.6) zapišemo kot

$$\kappa_{j,n} = n^{-(j-2)/2} \kappa_j, \quad j \geq 2, \quad (3.7)$$

kjer so  $\kappa_j$  kumulante osnovne populacije. Z združitvijo (3.5) in (3.6) dobimo

$$\begin{aligned} \chi_n(t) &= e^{-\frac{1}{2}t^2 + n^{-1/2}(k_{1,2}it + \frac{1}{3!}k_{3,1}(it)^3) + n^{-1}(\frac{1}{2!}k_{2,2}(it)^2 + \frac{1}{4!}k_{4,1}(it)^4) + \dots} = \\ &= e^{-\frac{t^2}{2}} \left( 1 + n^{-1/2} \left( k_{1,2}it + \frac{1}{3!}k_{3,1}(it)^3 \right) + \right. \\ &\quad \left. + n^{-1} \left( \frac{1}{2!}k_{2,2}(it)^2 + \frac{1}{4!}k_{4,1}(it)^4 + \frac{1}{2} \left( k_{1,2}it + \frac{1}{3!}k_{3,1}(it)^3 \right)^2 \right) + O(n^{-3/2}) \right). \end{aligned} \quad (3.8)$$

V splošnem (3.5) in (3.6) zapišemo kot

$$\chi_n(t) = e^{-t^2/2} + n^{-1/2}r_1(it)e^{-t^2/2} + n^{-1}r_2(it)e^{-t^2/2} + \dots + n^{-j/2}r_j(it)e^{-t^2/2} + \dots, \quad (3.9)$$

kjer je  $r_j$  polinom stopnje največ  $3j$ , sod za sode  $j$  in lih za lihe  $j$ . Če uporabimo

$$\int_{-\infty}^{\infty} e^{itx} d\{-He_{j-1}(x)\phi(x)\} = (it)^j e^{-t^2/2}, \quad (3.10)$$

kjer je  $He_{j-1}$  Hermitov polinom, lahko obrnemo karakteristično funkcijo (3.9) in dobimo

$$P(S_n \leq x) = \Phi(x) + n^{-1/2} p_1(x)\phi(x) + n^{-1} p_2(x)\phi(x) + \dots + n^{-j/2} p_j(x)\phi(x) + \dots, \quad (3.11)$$

kjer je  $p_j$  polinom stopnje največ  $3j-1$ , lih za sode  $j$  in sod za lihe  $j$ . Polinom  $p_1$  je popravek za asimetričnost,  $p_2$  pa popravek za sploščenost in sekundarni učinek asimetričnosti.

### 3.3 Bootstrap in Edgewortov razvoj

Po definiciji je statistika asimptotično pivotalna, če njena limitna porazdelitev ni odvisna od neznanih parametrov. Prednost pivotalnih statistik lahko prikažemo z Edgeworthovim razvojem. Če je pivotalna statistika  $T$  asimptotično normalno porazdeljena, potem lahko ob izpolnjenih predpostavkah njeno porazdelitveno funkcijo razvijemo kot

$$G(x) = P(T \leq x) = \Phi(x) + n^{-1/2} q(x)\phi(x) + O(n^{-1}), \quad (3.12)$$

kjer je  $q$  kvadraten polinom,  $\Phi$  in  $\phi$  pa sta standardna normalna porazdelitev in gostota. Kot primer vzamemo  $T = n^{1/2}(\hat{\theta} - \theta_0)/\hat{\sigma}$ , kjer je  $\hat{\theta}$  ocena neznanega parametra  $\theta_0$ ,  $\hat{\sigma}^2$  pa je ocena asimptotične variance  $\sigma^2$  od  $n^{1/2}\hat{\theta}$ .

Podobno lahko razvijemo porazdelitev bootstrap ocene  $G$ :

$$\hat{G}(x) = P(T^* \leq x | \mathfrak{N}) = \Phi(x) + n^{-1/2} \hat{q}(x)\phi(x) + O_p(n^{-1}), \quad (3.13)$$

kjer je  $T^*$  bootstrap oblika  $T$ , ocenjena s ponovljenim vzorčenjem iz  $\aleph^*$ , polinom  $\hat{q}$  pa dobimo iz  $q$  z nadomestitvijo neznanek z bootstrap ocenami. Ocene koeficientov  $\hat{q}$  se razlikujejo od ustreznih koeficientov  $q$  za  $\hat{q} - q = O_p(n^{-1/2})$ . Z odštevanjem (3.12) in (3.13) dobimo

$$P(T^* \leq x | \aleph) - P(T \leq x) = O_p(n^{-1}). \quad (3.14)$$

Napaka bootstrap približka za  $G$  je reda  $n^{-1}$ , kar je znatna izboljšava glede na približek normalne porazdelitve, ki ima napako reda  $n^{-1/2}$ .

Pri nepivotalni statistiki  $U = n^{-1/2}(\hat{\theta} - \theta_0)$  je napaka bootstrap približka porazdelitve reda  $n^{-1/2}$  namesto  $n^{-1}$ . Zaradi (3.12) velja

$$H(x) = P(U \leq x) = \Phi(x/\sigma) + n^{-1/2}p(x/\sigma)\phi(x/\sigma) + O(n^{-1}),$$

zaradi (3.13) pa

$$(3.15)$$

$$\hat{H}(x) = P(U^* \leq x | \aleph) = \Phi(x/\hat{\sigma}) + n^{-1/2}\hat{p}(x/\hat{\sigma})\phi(x/\hat{\sigma}) + O_p(n^{-1}),$$

kjer je  $p$  polinom,  $\hat{p}$  pa dobimo tako, da v  $p$  nadomestimo neznane parametre z njihovimi bootstrap ocenami. Varianca  $\sigma^2$  je asimptotična varianca od  $U$ ,  $\hat{\sigma}^2$  pa njena bootstrap ocena. Statistika  $U^*$  je bootstrap oblika  $U$ . Ponovno je  $\hat{p} - p = O_p(n^{-1/2})$  in  $\hat{\sigma} - \sigma = O_p(n^{-1/2})$ , zato

$$\hat{H}(x) - H(x) = \Phi(x/\hat{\sigma}) - \Phi(x/\sigma) + O_p(n^{-1}). \quad (3.16)$$

Razlika med  $\hat{\sigma}$  in  $\sigma$  je reda  $n^{-1/2}$  (saj se  $n^{1/2}(\hat{\sigma} - \sigma)$  v limiti porazdeljuje normalno  $N(0, \zeta^2)$  za nek  $\zeta > 0$ ). Razlika  $\Phi(x/\hat{\sigma}) - \Phi(x/\sigma)$  je zato v splošnem reda velikosti  $n^{-1/2}$ . Iz izraza (3.16) sledi, da je napaka bootstrap približka za  $H$  reda  $n^{-1/2}$ . Razlog za slabši

bootstrap približek porazdelitve statistike  $U$  je prisotnost  $\sigma$  v limitni porazdelitveni funkciji  $\Phi(x/\sigma)$ , zaradi česar je  $U$  nepivotalna.

### 3.4 Vrste intervalov zaupanja

Naj bo  $F_0$  populacijska porazdelitvena funkcija in  $\theta(\cdot)$  funkcional definiran na porazdelitvenih funkcijah tako, da je  $\theta_0 = \theta(F_0)$  točna vrednost parametra. Označimo z  $F_1$  porazdelitveno funkcijo vzorca  $\aleph$  iz populacije. Za neparametrični primer je  $F_1$  empirična porazdelitvena funkcija  $\aleph$ . Bootstrap ocena  $\theta_0$  je  $\hat{\theta} = \theta(F_1)$ . Definirajmo  $F_2$  kot porazdelitveno funkcijo ponovljenega vzorca  $\aleph^*$  iz 'populacije' s porazdelitveno funkcijo  $F_1$ .

Teoretični  $\alpha$ -kvantilni interval zaupanja za  $\theta_0$ ,

$$I_1 = (-\infty, \hat{\theta} + t_0),$$

dobimo z rešitvijo populacijske enačbe (3.2) za  $t = t_0$  s funkcijo

$$f_t(F_0, F_1) = P\{\theta(F_0) \leq \theta(F_1) + t\} - \alpha.$$

Vrednost  $t_0$  je določena z  $E\{f_t(F_0, F_1) | F_0\} = 0$ , torej

$$P(\theta \leq \hat{\theta} + t_0) = \alpha.$$

Ker pri bootstrapu empirična porazdelitvena funkcija  $F_1$  nastopa namesto  $F_0$  in porazdelitev ponovnega vzorčenja  $F_2$  namesto empirične porazdelitve  $F_1$ , je bootstrap različica intervala zaupanja je  $\hat{I}_1 = (-\infty, \hat{\theta} + \hat{t}_0)$ , kjer je  $t = \hat{t}_0$  rešitev vzorčne enačbe (3.3), torej

$$P(\theta(F_1) \leq \theta(F_2) + \hat{t}_0 | F_1) = \alpha.$$

Interval zaupanja  $\hat{I}_1$  imenujemo bootstrap kvantilni interval zaupanja.

Za sestavo t-kvantilnega intervala zaupanja za  $\theta_0$  definirajmo  $\sigma^2(F_0)$ , ki je asimptotična varianca od  $n^{1/2}\hat{\theta}$  in  $\hat{\sigma}^2 = \sigma^2(F_1)$ . Teoretični  $\alpha$  t-kvantilni interval zaupanja je  $J_1 = (-\infty, \hat{\theta} + t_0\hat{\sigma})$ , kjer je  $t_0$  določen z

$$P(\theta_0 \leq \hat{\theta} + t_0\hat{\sigma}) = \alpha.$$

To je le konkreten zapis za funkcional  $f_t(F_0, F_1)$  (glej (3.2)), torej

$$f_t(F_0, F_1) = I\{\theta(F_0) \leq \theta(F_1) + t\sigma(F_1)\} - \alpha.$$

Bootstrap interval zaupanja dobimo z reševanjem ustrezne vzorčne enačbe in je enak  $\hat{J}_1 = (-\infty, \hat{\theta} + \hat{t}_0\hat{\sigma})$ , kjer je  $\hat{t}_0$  določen z

$$P(\hat{\theta} \leq \theta(F_2) + \hat{t}_0\sigma(F_2) | F_1) = \alpha.$$

Zaradi poenostavitve zapisa bomo v prihodnje  $\theta(F_2)$  označevali z  $\hat{\theta}^*$  in  $\sigma(F_2)$  z  $\hat{\sigma}^*$ .

Izrazimo  $t_0$  in  $\hat{t}_0$  s kvantili  $u_\alpha$ ,  $v_\alpha$ ,  $\hat{u}_\alpha$  in  $\hat{v}_\alpha$ , ki so definirani z

$$P(n^{1/2}(\theta(F_1) - \theta(F_0)) / \sigma(F_0) \leq u_\alpha | F_0) = \alpha, \quad (3.17)$$

$$P(n^{1/2}(\theta(F_1) - \theta(F_0)) / \sigma(F_1) \leq v_\alpha | F_0) = \alpha, \quad (3.18)$$

$$P(n^{1/2}(\theta(F_2) - \theta(F_1)) / \sigma(F_1) \leq \hat{u}_\alpha | F_1) = \alpha \quad (3.19)$$

in

$$P(n^{1/2}(\theta(F_2) - \theta(F_1)) / \sigma(F_2) \leq \hat{v}_\alpha | F_1) = \alpha. \quad (3.20)$$

Zapišimo  $\sigma = \sigma(F_0)$  in  $\hat{\sigma} = \sigma(F_1)$ , ter definirajmo naslednje intervale zaupanja:



$$I_1 = (-\infty, \hat{\theta} - n^{-1/2} \sigma u_{1-\alpha}), \quad J_1 = (-\infty, \hat{\theta} - n^{-1/2} \hat{\sigma} v_{1-\alpha}), \quad (3.21)$$

$$\hat{I}_1 = (-\infty, \hat{\theta}^* - n^{-1/2} \hat{\sigma} \hat{u}_{1-\alpha}), \quad \hat{J}_1 = (-\infty, \hat{\theta}^* - n^{-1/2} \hat{\sigma}^* \hat{v}_{1-\alpha}). \quad (3.22)$$

Vsi intervali zaupanja so za  $\theta_0$  z verjetnostmi pokritja približno enakimi  $\alpha$ . Interval zaupanja  $\hat{I}_1$  je kvantilni,  $\hat{J}_1$  pa t-kvantilni.

Dvostranski interval zaupanja, ki ima enako tveganje v obeh repih, sestavimo kot presek dveh enostranskih intervalov. Dvostranska analoga za  $I_1$  in  $J_1$  sta

$$I_2 = (\hat{\theta} - n^{-1/2} \sigma u_{(1+\alpha)/2}, \hat{\theta} - n^{-1/2} \sigma u_{(1-\alpha)/2})$$

in

$$J_2 = (\hat{\theta} - n^{-1/2} \hat{\sigma} v_{(1+\alpha)/2}, \hat{\theta} - n^{-1/2} \hat{\sigma} v_{(1-\alpha)/2}),$$

ter njuni bootstrap različici

$$\hat{I}_2 = (\hat{\theta} - n^{-1/2} \hat{\sigma} \hat{u}_{(1+\alpha)/2}, \hat{\theta} - n^{-1/2} \hat{\sigma} \hat{u}_{(1-\alpha)/2})$$

in

$$\hat{J}_2 = (\hat{\theta} - n^{-1/2} \hat{\sigma} \hat{v}_{(1+\alpha)/2}, \hat{\theta} - n^{-1/2} \hat{\sigma} \hat{v}_{(1-\alpha)/2}).$$

Intervala zaupanja  $I_2$  in  $J_2$  imata enako tveganje v obeh repih,

$$P(\theta_0 \leq \hat{\theta} - n^{-1/2} \sigma u_{(1+\alpha)/2}) = P(\theta_0 > \hat{\theta} - n^{-1/2} \sigma u_{(1-\alpha)/2}) = \frac{1}{2}(1-\alpha),$$

intervala zaupanja  $\hat{I}_2$  in  $\hat{J}_2$  pa imata približno enako tveganje v obeh repih. Vsi zgoraj definirani intervali imajo vsaj asimptotično kritje enako  $\alpha$  v smislu, da, če je  $I$  katerikoli od intervalov, potem

$$P(\theta_0 \in I) \rightarrow \alpha,$$

ko  $n \rightarrow \infty$ . Verjetnost  $\alpha$  imenujemo ciljano pokritje (nominal coverage) intervala zaupanja  $I$ . Napaka pokritja za  $I$  je razlika med dejanskim in ciljanim pokritjem,

$$\text{napaka pokritja} = P(\theta_0 \in I) - \alpha.$$

### 3.5 Edgeworthov in Cornish-Fisherjev razvoj

Naj bosta

$$P(n^{1/2}(\hat{\theta} - \theta_0)/\sigma \leq x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + \dots + n^{-j/2}p_j(x)\phi(x) + \dots \quad (3.23)$$

in

$$P(n^{1/2}(\hat{\theta} - \theta_0)/\hat{\sigma} \leq x) = \Phi(x) + n^{-1/2}q_1(x)\phi(x) + \dots + n^{-j/2}q_j(x)\phi(x) + \dots \quad (3.24)$$

Edgeworthova razvoja za nestudentizirano in studentizirano statistiko. Polinoma  $p_j$  in  $q_j$  sta stopnje največ  $3j-1$  in sta lihi ali sodi funkciji glede na sodost ali lihost  $j$ . Tako kot lahko z Edgeworthovim razvojem za vsak kvantil dobimo ustrezno vrednost porazdelitvene funkcije (pokritje), lahko s Cornish-Fisherjevim razvojem dobimo obratno, torej za vsako pokritje (vrednost kumulativne funkcije) ustrezen kvantil. Z razvojem (3.23) in (3.24) namreč dobimo Cornish-Fisherjev razvoj za kvantila  $u_\alpha$  in  $v_\alpha$ , definirana v (3.17) in (3.18):

$$u_\alpha = z_\alpha + n^{-1/2}p_{11}(z_\alpha) + n^{-1}p_{21}(z_\alpha) + \dots + n^{-j/2}p_{j1}(z_\alpha) + \dots, \quad (3.25)$$

$$v_\alpha = z_\alpha + n^{-1/2}q_{11}(z_\alpha) + n^{-1}q_{21}(z_\alpha) + \dots + n^{-j/2}q_{j1}(z_\alpha) + \dots, \quad (3.26)$$

kjer je  $z_\alpha$  rešitev  $\Phi(z_\alpha) = \alpha$ . Polinoma  $p_{j1}$  in  $q_{j1}$  sta stopnje največ  $j+1$  in sta lihi funkciji za sod  $j$  in sodi funkciji za lih  $j$ .

Koeficienti polinomov  $p_j$  in  $q_j$  (tudi  $q_{j1}$  in  $q_{j1}$ ) so odvisni od kumulant porazdelitve  $(\hat{\theta} - \theta_0)/\sigma$  in  $(\hat{\theta} - \theta_0)/\hat{\sigma}$ .

Po predpostavkah modela gladke funkcije (glej razdelek 3.7), kjer sta  $\hat{\theta}$  in  $\hat{\sigma}$  gladki funkciji povprečja vzorca  $\mathfrak{N} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$ , so koeficienti polinomov  $p_j$  in  $q_j$  funkcije momentov vektorja  $\mathbf{X}$ . Pri prehodu od razvoja (3.23) in (3.24), ki temelji na populaciji, k razvoju, ki temelji na vzorcu, nadomesti populacijsko porazdelitveno funkcijo  $F_0$  vzorčna porazdelitvena funkcija  $F_1$ , populacijske momente koeficientov  $p_j$  in  $q_j$ , pa nadomestijo vzorčni momenti. Naj bosta  $\hat{p}_j$  in  $\hat{q}_j$  polinoma, ki jih dobimo iz  $p_j$  in  $q_j$  z nadomestitvijo populacijskih momentov z vzorčnimi momenti in zapišimo  $\hat{\theta}^* = \theta(F_2)$  in  $\hat{\sigma}^* = \sigma(F_2)$ . Potem sta vzorčni obliki razvojev (3.23) in (3.24)

$$P(n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma} \leq x | \mathfrak{N}) = \Phi(x) + n^{-1/2} \hat{p}_1(x)\phi(x) + \dots + n^{-j/2} \hat{p}_j(x)\phi(x) + \dots \quad (3.27)$$

in

$$P(n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \leq x | \mathfrak{N}) = \Phi(x) + n^{-1/2} \hat{q}_1(x)\phi(x) + \dots + n^{-j/2} \hat{q}_j(x)\phi(x) + \dots \quad (3.28)$$

Oba razvoja obravnavamo kot asimptotični vrsti. Ob upoštevanju ustreznih predpostavk o momentih in gladkosti velja

$$P(n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma} \leq x | \mathfrak{N}) = \Phi(x) + n^{-1/2} \hat{p}_1(x)\phi(x) + \dots + n^{-j/2} \hat{p}_j(x)\phi(x) + o_p(n^{-j/2}). \quad (3.29)$$

Ostanek  $o_p(n^{-j/2})$  je enakomeren za  $x$ :

$$n^{j/2} \sup_{-\infty < x < \infty} |P(n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma} \leq x | \mathfrak{N}) - (\Phi(x) + n^{-1/2} \hat{p}_1(x)\phi(x) + \dots + n^{-j/2} \hat{p}_j(x)\phi(x))| \rightarrow 0,$$

ko  $n \rightarrow \infty$ . Za dokaz glej Hall, 1992.

Če naredimo obratni razvoj Edgeworthovega razvoja (3.27) in (3.28) dobimo vzorčno obliko Cornish-Fisherjevega razvoja (3.21) in (3.22):

$$\hat{u}_\alpha = z_\alpha + n^{-1/2} \hat{p}_{11}(z_\alpha) + n^{-1} \hat{p}_{21}(z_\alpha) + \dots + n^{-j/2} \hat{p}_{j1}(z_\alpha) + \dots, \quad (3.30)$$

$$\hat{v}_\alpha = z_\alpha + n^{-1/2} \hat{q}_{11}(z_\alpha) + n^{-1} \hat{q}_{21}(z_\alpha) + \dots + n^{-j/2} \hat{q}_{j1}(z_\alpha) + \dots \quad (3.31)$$

Polinoma  $\hat{p}_{j_1}$  in  $\hat{q}_{j_1}$  se od  $p_{j_1}$  in  $q_{j_1}$  razlikujeta po tem, da je  $F_0$  nadomeščena z  $F_1$ , zato so populacijski momenti nadomeščeni z vzorčnimi. Cornish-Fisherjev razvoj je asimptotična vrsta in velja enakomerno za  $\alpha$ , ki zavzema vrednosti na odprtem intervalu  $(0,1)$ :

$$n^{j/2} \sup_{\varepsilon < \alpha < 1-\varepsilon} |\hat{u}_\alpha - (z_\alpha + n^{-1/2} \hat{p}_{11}(z_\alpha) + \dots + n^{-j/2} \hat{p}_{j1}(z_\alpha))| \rightarrow 0, \quad (3.32)$$

ko  $n \rightarrow \infty$  za vsak  $0 < \varepsilon < \frac{1}{2}$ . Bistvena predpostavka za omenjene relacije je ‘pohlevnost’ vzorčne porazdelitve. Za dokaz glej Hall, 1992.

### 3.6 Lastnosti bootstrap kvantilov in kritičnih točk

V nadaljevanju nas zanima natančnost bootstrap ocene kvantilov porazdelitev  $S$  in  $T$  in njen vpliv na kritične točke bootstrap intervalov zaupanja.

Naj bosta  $u_\alpha$  in  $v_\alpha$   $\alpha$ -kvantila porazdelitev  $S$  in  $T$ , ter  $\hat{u}_\alpha$  in  $\hat{v}_\alpha$  njuni bootstrap oceni. Zapišimo razliko (3.25) in (3.30), ter (3.26) in (3.31):

$$\hat{u}_\alpha - u_\alpha = n^{-1/2}(\hat{p}_{11}(z_\alpha) - p_{11}(z_\alpha)) + n^{-1}(\hat{p}_{21}(z_\alpha) - p_{21}(z_\alpha)) + \dots \quad (3.33)$$

in

$$\hat{v}_\alpha - v_\alpha = n^{-1/2}(\hat{q}_{11}(z_\alpha) - q_{11}(z_\alpha)) + n^{-1}(\hat{q}_{21}(z_\alpha) - q_{21}(z_\alpha)) + \dots \quad (3.34)$$

Polinom  $\hat{p}_{j_1}$  dobimo tako, da v  $p_{j_1}$  nadomestimo, kot rečeno, populacijske momente z vzorčnimi, ki se razlikujejo reda  $O_p(n^{-1/2})$ . To je tudi red\* razlike med  $\hat{p}_{j_1}$  in  $p_{j_1}$ . Sledi

$$\hat{u}_\alpha - u_\alpha = O_p(n^{-1/2} \cdot n^{-1/2} + n^{-1}) = O_p(n^{-1})$$

in podobno  $\hat{v}_\alpha - v_\alpha = O_p(n^{-1})$ .

---

\* Konvergenco cenilke s hitrostjo  $n^{-1/2}$  imenujemo konvergenco prvega reda.

S tem je pokazano, da sta napaki bootstrap ocen kvantilov  $u_a$  in  $v_a$  reda  $n^{-1}$ . Istočasno je napaka približka normalne porazdelitve za  $u_a$  in  $v_a$  je reda  $n^{-1/2}$ :

$$z_a - u_a = z_a - (z_a + n^{-1/2} p_{11}(z_a) + \dots) = -n^{-1/2} p_{11}(z_a) + O(n^{-1}).$$

Ker je  $\hat{v}_{1-a} = v_{1-a} + O_p(n^{-1})$ , se zgornja meja intervala zaupanja  $\hat{J}_1 = (-\infty, \hat{\theta} - n^{-1/2} \hat{\sigma}_{\hat{v}_{1-a}})$  razlikuje od zgornje meje  $J_1 = (-\infty, \hat{\theta} - n^{-1/2} \hat{\sigma}_{v_{1-a}})$  samo za  $O_p(n^{-3/2})$ . Interval zaupanja  $\hat{J}_1$  je točen do drugega reda za  $J_1$  in  $\hat{\theta} - n^{-1/2} \hat{\sigma}_{\hat{v}_{1-a}}$  je točna do drugega reda za  $\hat{\theta} - n^{-1/2} \hat{\sigma}_{v_{1-a}}$ , ker se količini ne razlikujeta v členih, ki so reda velikosti do vključno  $(n^{-1/2})^2 = n^{-1}$ . Interval zaupanja  $\hat{I}_1 = (-\infty, \hat{\theta} - n^{-1/2} \hat{\sigma}_{\hat{u}_{1-a}})$  pa je za  $I_1 = (-\infty, \hat{\theta} - n^{-1/2} \hat{\sigma}_{u_{1-a}})$  točen do prvega reda, ker se zgornji meji ne razlikujeta samo za člene, ki so manjši od reda  $n^{-1/2}$ . Ugotovitve o kritičnih točkah veljajo enako tudi za dvostranske intervale zaupanja.

Lastnosti intervalov zaupanja v kritičnih točkah vplivajo na napako pokritja. Interval zaupanja  $\hat{J}_1$ , ki je točen do drugega reda, ima napako pokritja reda  $n^{-1}$ , interval zaupanja  $\hat{I}_1$ , ki je točen do prvega reda, pa ima napako pokritja reda  $n^{-1/2}$ .

### 3.7 Dopustni model za Edgeworthov razvoj

Naj bodo  $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \dots$  neodvisni in enako porazdeljeni slučajni vektorji dimenzije  $d$  s povprečjem  $\boldsymbol{\mu}$  in  $\bar{\mathbf{X}} = n^{-1} \sum \mathbf{X}_i$ . Naj bo  $A: \mathbb{R}^d \rightarrow \mathbb{R}$  gladka funkcija in  $A(\boldsymbol{\mu}) = 0$ . Z  $A$  mislimo na funkcije oblike

$$A(\mathbf{x}) = (g(\mathbf{x}) - g(\boldsymbol{\mu})) / h(\boldsymbol{\mu}), \quad (3.35)$$

kjer je  $\theta_0 = g(\boldsymbol{\mu})$  parameter ocenjen kot  $\hat{\theta} = g(\bar{\mathbf{X}})$  in  $h(\boldsymbol{\mu})^2$  je asimptotična varianca  $n^{1/2} \hat{\theta}$ ; ali

$$A(\mathbf{x}) = (g(\mathbf{x}) - g(\boldsymbol{\mu})) / h(\mathbf{x}), \quad (3.36)$$

kjer je  $h(\bar{\mathbf{X}})$  ocena  $h(\boldsymbol{\mu})$  (Predpostavljamo torej, da je  $h$  znana funkcija).

Gornji model gladke funkcije (smooth function model) nam omogoča Edgeworthovo obravnavo tudi bolj splošnih problemov, kjer je  $\theta_0$  povprečje, varianca, razmerje povprečij in varianc, razlika povprečij in varianc, korelacijski koeficient in podobno.

Za izrek, ki je ključen za Edgeworthov razvoj statistik, ki so gladke funkcije, moramo definirati nekaj količin. Označimo  $i$ -ti element  $d$  dimenzionalnega vektorja  $\mathbf{v}$  kot  $v^{(i)}$  ali  $(\mathbf{v})^{(i)}$ ,

$$a_{i_1 \dots i_j} = \left( \partial^j / \partial x^{(i_1)} \dots \partial x^{(i_j)} \right) A(\mathbf{x}) \Big|_{\mathbf{x}=\boldsymbol{\mu}} \quad (3.37)$$

in

$$\mu_{i_1 \dots i_j} = E\left( (\mathbf{X} - \boldsymbol{\mu})^{(i_1)} \dots (\mathbf{X} - \boldsymbol{\mu})^{(i_j)} \right), \quad j \geq 1. \quad (3.38)$$

Zapišimo še

$$\sigma^2 = \sum_{i=1}^d \sum_{j=1}^d a_i a_j \mu_{ij}, \quad (3.39)$$

$$A_1 = \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d a_{ij} \mu_{ij}, \quad (3.40)$$

$$A_2 = \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d a_i a_j a_k \mu_{ijk} + 3 \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d a_i a_j a_k \mu_{ik} \mu_{jl}. \quad (3.41)$$

IZREK 3.1: Naj bo funkcija  $A$   $j+2$ -krat zvezno odvedljiva v okolici  $\boldsymbol{\mu} = E(\mathbf{X})$ ,  $A(\boldsymbol{\mu}) = 0$ ,  $E(\|\mathbf{X}\|^{j+2}) < \infty$  in za karakteristično funkcijo  $\chi$  vektorja  $\mathbf{X}$  velja Cramérjev pogoj

$$\limsup_{\|t\| \rightarrow \infty} |\chi(t)| < 1. \quad (3.42)$$

Predpostavimo, da  $\sigma > 0$ . Za  $j \geq 1$  potem velja

$$P(n^{1/2}A(\bar{\mathbf{X}})/\sigma \leq x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + n^{-1}p_2(x)\phi(x) + \dots + n^{-j/2}p_j(x)\phi(x) + o(n^{-j/2}). \quad (3.43)$$

Relacija (3.43) velja enakomerno za  $x$ , kjer je  $p_j$  polinom stopnje največ  $3j-1$ , lih za sode  $j$  in sod za lihe  $j$ , s koeficienti, ki so odvisni od momentov  $\mathbf{X}$  do stopnje  $j+2$ . Za  $j=1$  je  $p_j$

$$p_1(x) = -\left(A_1\sigma^{-1} + \frac{1}{6}A_2\sigma^{-3}(x^2-1)\right). \quad (3.44)$$

Za dokaz izreka glej Hall, 1992.

V nadaljevanju bomo funkcije, ki nastopajo v Izreku 3.1, imenovali družina gladkih funkcij. Sledi obravnava bootstrap različice gornjega izsledka. Oceniti želimo porazdelitveno funkcijo  $(\hat{\theta} - \theta_0)/\sigma$  ali  $(\hat{\theta} - \theta_0)/\hat{\sigma}$ . Obe statistiki lahko zapišemo splošneje kot  $A(\bar{\mathbf{X}})$ , kjer je  $A(\mathbf{x})$  definiran kot v (3.35) in (3.36). Za bootstrap oceno porazdelitev naj bo  $\mathfrak{N}^* = \{\mathbf{X}_1^*, \mathbf{X}_2^*, \dots, \mathbf{X}_n^*\}$  slučajni vzorec s ponavljanjem iz  $\mathfrak{N}$  in  $\bar{\mathbf{X}}^* = n^{-1} \sum \mathbf{X}_i^*$  vzorčno povprečje. Definirajmo

$$\hat{\theta}^* = g(\bar{\mathbf{X}}^*) \text{ in } \hat{\sigma}^* = h(\bar{\mathbf{X}}^*).$$

Statistiki  $(\hat{\theta} - \theta_0)/\sigma$  in  $(\hat{\theta} - \theta_0)/\hat{\sigma}$  lahko izrazimo z  $\hat{A}(\bar{\mathbf{X}}^*)$ , kjer je

$$\hat{A}(\mathbf{x}) = (g(\mathbf{x}) - g(\bar{\mathbf{X}}))/h(\bar{\mathbf{X}}), \quad (3.45)$$

$$\hat{A}(\mathbf{x}) = (g(\mathbf{x}) - g(\bar{\mathbf{X}}))/h(\mathbf{x}). \quad (3.46)$$

Bootstrap porazdelitev  $A(\bar{\mathbf{X}})$  (kjer je  $A(\bar{\mathbf{X}}) = (g(\bar{\mathbf{X}}) - g(\boldsymbol{\mu}))/h(\boldsymbol{\mu})$  in

$A(\bar{\mathbf{X}}) = (g(\bar{\mathbf{X}}) - g(\boldsymbol{\mu}))/h(\bar{\mathbf{X}})$  v pivotalni obliki) je podana s porazdelitvijo  $\hat{A}(\bar{\mathbf{X}}^*)$ , pogojno na  $\mathfrak{N}$ .

Naj bo  $\hat{A}$  definiran kot v (3.45) ali (3.46), ter  $\pi_j = p_j$  v prvem primeru in  $\pi_j = q_j$  v drugem primeru, in naj bo  $\hat{\pi}_j$  polinom, ki ga dobimo tako, da v  $\pi_j$  nadomestimo populacijske

momente z ustreznimi vzorčnimi momenti. Vzorčni moment, ki ustreza populacijskemu momentu

$$E\left(X^{(i_1)} \dots X^{(i_j)}\right) \text{ je } n^{-1} \sum_{k=1}^n X_k^{(i_1)} \dots X_k^{(i_j)} .$$

IZREK 3.2: Naj bo  $\lambda > 0$  dan in naj bo  $l = l(\lambda)$  zadosti veliko pozitivno število. Predpostavimo, da sta  $g$  in  $h$   $\nu + 3$ -krat omejeno odvedljivi v okolici  $\boldsymbol{\mu}$ ,  $E(\|\mathbf{X}\|^l) < \infty$  in velja Cramérjev pogoj (3.42). Potem obstaja konstanta  $C > 0$  tako, da

$$P\left[\sup_{-\infty < x < \infty} |P(n^{1/2} \hat{A}(\bar{\mathbf{X}}^*) \leq x | \mathfrak{N}) - \Phi(x) - \sum_{j=1}^{\nu} n^{-j/2} \hat{\pi}_j(x) \phi(x)| > C n^{-(\nu+1)/2}\right] = O(n^{-\lambda}), \quad (3.47)$$

$$P\left[\max_{1 \leq j \leq \nu} \sup_{-\infty < x < \infty} (1 + |x|)^{-(3j-1)} |\hat{\pi}_j(x)| > C\right] = O(n^{-\lambda}). \quad (3.48)$$

Za dokaz izreka glej Hall, 1992.

Trditev (3.47), ki velja za bootstrap, ustreza trditvi (3.43). Omejenost stopnje polinoma pri Edgeworthovem razvoju za gladke funkcije bootstrap ocene (3.48) ustreza trditvi o stopnji polinoma  $p_j$  največ  $3j - 1$ . Ker pa je bootstrap statistika ocenjena pogojno na vzorec  $\mathfrak{N}$ , veljajo seveda zaključki Izreka 3.2 v verjetnosti.

#### 4. Eksperiment

V četrtem poglavju bo najprej opisana izvedba bootstrap metode za regresijski model s strukturo motnje in bootstrap preizkušanja statističnih hipotez. Sledi podroben opis eksperimenta.



#### 4.1 Uporaba neparametrične bootstrap metode za linearni regresijski model s strukturo motnje

Linearni regresijski model s strukturo motnje izrazimo kot

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T.$$

Predpostavljamo, da sta komponenti motnje  $u_{it}$  in  $c_i$  slučajna vzorca iz neznanih verjetnostnih porazdelitev in da imata pričakovanji enaki 0. Model verjetnosti za linearni regresijski model s strukturo motnje je določen z dvema skupinama parametrov. Prva je vektor parametrov  $\boldsymbol{\beta}$ , druga pa so parametri verjetnostnih porazdelitev obeh komponent strukturirane motnje. V primeru, ko bi bil vektor  $\boldsymbol{\beta}$  znan, bi lahko izračunali ocenjeni komponenti motnje  $\hat{u}_{it} = (y_{it} - \bar{y}_i) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta}$  in  $\hat{c}_i = \bar{y}_i - \bar{\mathbf{x}}_i\boldsymbol{\beta}$ , pri čemer gre  $\bar{u}_i$  v verjetnosti proti 0. Ker vektor parametrov  $\boldsymbol{\beta}$  ni znan, za izračun obeh komponent ostankov uporabimo njegovo dosledno oceno  $\hat{\boldsymbol{\beta}}$ :

$$\hat{u}_{it} = (y_{it} - \bar{y}_i) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\hat{\boldsymbol{\beta}}, \quad (4.1)$$

$$\hat{c}_i = \bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}. \quad (4.2)$$

Oceni verjetnostnih porazdelitev komponent motnje sta empirični porazdelitvi  $\hat{u}_{it}$  in  $\hat{c}_i$ , zato pri bootstrapu predpostavljamo, da je verjetnost vseh ocenjenih motenj enaka

$$\frac{1}{NT} \text{ za } \forall \hat{u}_{it}, i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T \quad (4.3)$$

in

$$\frac{1}{N} \text{ za } \forall \hat{c}_i, \quad i = 1, 2, \dots, N. \quad (4.4)$$

S pomočjo modela verjetnosti lahko generiramo podatke za bootstrap oceno parametrov regresijskega modela s strukturo motnje. Najprej izberemo slučajni vzorec bootstrap komponent ostankov:

$$\{u_{11}^*, u_{12}^*, \dots, u_{1T}^*, u_{21}^*, u_{22}^*, \dots, u_{2T}^*, \dots, u_{N1}^*, u_{N2}^*, \dots, u_{NT}^*\} \quad (4.5)$$

in

$$\{c_1^*, c_2^*, \dots, c_N^*\}, \quad (4.6)$$

nato pa generiramo bootstrap vrednost odvisne spremenljivke:

$$y_{it}^* = \mathbf{x}_{it} \hat{\boldsymbol{\beta}} + c_i^* + u_{it}^*, \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T. \quad (4.7)$$

Vrednosti vektorjev neodvisnih spremenljivk so enake za vse bootstrap ponovitve, ker neodvisne spremenljivke obravnavamo kot fiksne.

## 4.2 Statistično prizkušanje hipotez z bootstrap metodo

Statistično preizkušanje hipotez je formalno sredstvo za razlikovanje med verjetnostnimi porazdelitvami na podlagi slučajnih spremenljivk (statistik) pripadajočih eni izmed verjetnostnih porazdelitev. Neyman-Pearsonov pristop sprejemanja ali zavračanja ničelne hipoteze vključuje možnost dveh tipov napak. Napaka I. vrste nastopi v primeru zavrnitve ničelne hipoteze, ki je pravilna. Verjetnost za napako I. vrste imenujemo tudi velikost testa in ju bomo v nadaljevanju uporabljali kot sinonima. Napaka II. vrste pa nastopi v primeru sprejetja ničelne hipoteze čeprav je ta napačna. Verjetnost zavrnitve ničelne hipoteze, ki je napačna, imenujemo moč testa.

Običajno preizkušanje parametričnih hipotez poteka tako, da iz vzorca ocenimo  $\hat{\theta}$  neznan parameter  $\theta$ . Za preizkušanje hipotez o oceni parametra uporabimo verjetnostno porazdelitev testne statistike, ki je v ničelni hipotezi določena vsaj asimptotično. Pri bootstrap testu obravnavamo vzorec kot populacijo, za katero poznamo točno vrednost iskanega parametra, ki je kar ocena  $\hat{\theta}$ . Nato s ponovljenim slučajnim vzorčenjem s ponavljanjem iz prvotnega vzorca generiramo bootstrap vzorce, ki imajo enako velikost kot prvotni vzorec, vendar drugačno sestavo. Iz bootstrap vzorca nato ocenimo bootstrap oceno  $\hat{\theta}^b$ , ki, ob ponovljenem slučajnem vzorčenju, pri šibkih pogojih konvergira v verjetnosti k  $\hat{\theta}$ . Na ta način generiramo

verjetnostno porazdelitev testne statistike, ki določa značilnosti prvotnega vzorca in ničelne hipoteze  $\hat{\theta}^b = \hat{\theta}$ . Kvantilna meja določa mejo zaupanja bootstrap testa.

Glavni razlog za uporabo bootstrapa pri preizkušanju hipotez je odstopanje porazdelitve asimptotičnih testov od predpostavljene (limitne) porazdelitve na majhnih vzorcih. Z bootstrpom ocenjene testne statistike pod določenimi pogoji hitreje konvergirajo k pravi vrednostim kot asimptotične. Osnovni pogoj za hitrejšo konvergenco glede na asimptotične teste je pivotalnost testne statistike. Verjetnostna porazdelitev pivotalne statistike ni odvisna od neznanih parametrov. Za večino testov je pogoj pivotalnosti zadoščen asimptotično.

#### 4.2.1 Bootstrap Hausmanovega specifikacijskega testa

Hausmanova statistika se porazdeljuje po  $\chi^2$  porazdelitvi asimptotično. Za moč testa na majhnih vzorcih je še zlasti pomemben drugi člen razlike  $V(\hat{q}) = V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE}) = \sigma_u^2 (X^T Q X)^{-1} - (X^T \Omega^{-1} X)^{-1}$  zaradi inverzne matrike  $\Omega^{-1}$ , ki je ključna determinanta konvergence (v skrajnem primeru ima ocenjena Hausmanova statistika lahko negativno vrednost). Na hitrost približevanja teoretični porazdelitvi vplivajo različni faktorji, npr. dimenzije panela, časovna in presečna avtokorelacija šuma, heteroskedastičnost šuma, avtokoreliranost neodvisnih spremenljivk in podobno.

Izboljšavo Hausmanovega testa za regresijski model s strukturo motnje na majhnih vzorcih lahko dosežemo z bootstrap metodo tako, da meje značilnosti ocenimo z empirično porazdelitvijo Hausmanove statistike. Z ocenami koeficientov modela ocenimo komponente motnje  $\hat{c}_i$  in  $\hat{u}_{it}$ . S ponovnim vzorčenjem iz  $\{\hat{c}_i\}$  in  $\{\hat{u}_{it}\}$ , neodvisno od  $\{X_{it}\}$ , generiramo vrednosti odvisne spremenljivke. Na ta način dobimo s ponovnim vzorčenjem iz  $\{\hat{c}_i\}$  in  $\{\hat{u}_{it}\}$  empirično porazdelitev Hausmanove statistike. Ponovno vzorčenje iz  $\{\hat{c}_i\}$  in  $\{\hat{u}_{it}\}$ , neodvisno od  $\{X_{it}\}$ , zaradi "prekinitve" povezave med  $\hat{c}_i$  in  $X_{it}$ , povzema osnovno idejo Hausmanovega testa nekoreliranosti časovno invariantne komponente ostanka z neodvisnimi spremenljivkami. Bootstrap Hausmanovega testa primerjamo z osnovnim testom glede na napako I. vrste in glede na moč.

Napako I. vrste analiziramo tako, da glede na ciljano napako I. vrste, za teoretično predpostavljeni model s slučajnimi učinki ocenimo delež zavrženih ničelnih hipotez glede na  $\chi^2$  porazdelitev in glede na empirično porazdelitev Hausmanove statistike. Moč testa preizkušamo v dveh korakih. Najprej ocenimo napako I. vrste asimptotičnega testa, ki jo uporabimo kot stopnjo tveganja empirične porazdelitve bootstrap Hausmanove statistike pri predpostavljenem modelu s fiksnimi učinki, s čimer dosežemo primerljivost moči asimptotičnega in bootstrap testa (Davidson, MacKinnon, 2000). Nato ocenimo delež zavrženih hipotez glede na  $\chi^2$  porazdelitev in glede na korigirano empirično porazdelitev Hausmanove statistike.

#### 4.2.2 Specifikacija ničelne in alternativne hipoteze

Kriterij za odločanje med cenilkama fiksnih in slučajnih učinkov je Hausmanov specifikacijski test, ki temelji na razliki  $\hat{\beta}_{RE}$  in  $\hat{\beta}_{FE}$ . Obe cenilki sta dosledni v ničelni hipotezi:

$$H_0: E(c_i | \mathbf{x}_i) = 0, \quad (4.8)$$

(glej predpostavko RE.1b); pri tem je  $\hat{\beta}_{FE}$  dosledna tudi v primeru, če ničelna hipoteza ne velja, asimptotično učinkovita  $\hat{\beta}_{RE}$  pa je dosledna samo, če velja ničelna hipoteza. Predpostavka stroge eksogenosti RE.1a mora veljati v ničelni in alternativni hipotezi. Navadno tudi predpostavljamo veljavnost RE.3 v ničelni hipotezi, kar zagotavlja večjo učinkovitost cenilke slučajnih učinkov glede na cenilko fiksnih učinkov.

Specifikacijo alternativne hipoteze povzemamo po Arellanu (1993, str.90):

$$H_1: E(c_i | \mathbf{x}_i) = E(c_i | \bar{\mathbf{x}}_i) = \bar{\mathbf{x}}_i^T \boldsymbol{\gamma}, \quad (4.9)$$

tako, da v primeru ničelne hipoteze velja  $\gamma = 0$ . Slučajna spremenljivka neopazovanega učinka  $c_i$  je torej korelirana s časovnim povprečjem neodvisne spremenljivke.

### 4.3 Konstrukcija eksperimenta

Analiza je bila narejena za osnovni model in več alternativ, ki so se od osnovnega modela razlikovale v enem parametru. Namen alternativnih oblik modela je bilo preizkusiti vpliv različnih parametrov procesa generiranja podatkov na Hausmanov specifikacijski test za regresijski model s strukturo motnje. Analizirali smo vpliv na napako I. vrste in moč testa. Panelni podatki so bili generirani z modelom

$$y_{it} = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}.$$

Značilnosti osnovne oblike modela:

1. Dimenzija panela:  $N = 25$ ,  $T = 10$
2. Matrika strukture: 1 neodvisna spremenljivka  $x_{it} \sim N(0,1)$
3. Spremenljivka časovno invariantnega neopazovanega učinka  $c_i$  neodvisno in enako porazdeljena  $c_i \sim N(0,0.5)$
4. Čisti šum neodvisno in enako porazdeljen  $u_{it} \sim N(0,0.5)$
5. Korelacija časovnega povprečja neodvisne spremenljivke in spremenljivke neopazovanih učinkov v alternativni hipotezi  $r(c_i, \bar{x}_i) = 0.5$

Glede na osnovni model so bile oblikovane naslednje alternative:

- A01. Enakomerna porazdelitev spremenljivke neopazovanega učinka  $c_i \sim U(0,0.5)$
- A02. Enakomerna porazdelitev čistega šuma  $u_{it} \sim U(0,0.5)$
- A03. Varianca čistega šuma  $\text{var}(u_{it}) = 1$
- A04. Časovna avtokorelacija čistega šuma  $u_{it} : u_{it} = 0.8u_{it-1} + u_{2it}$ ,  $u_{2it} \sim N(0,0.18)$
- A05. Časovna avtokorelacija čistega šuma  $u_{it} : u_{it} = -0.8u_{it-1} + u_{2it}$ ,  $u_{2it} \sim N(0,0.18)$
- A06. Presečna heteroskedastičnost čistega šuma  $u_{it}$
- A07. Matrika strukture: 2 neodvisni spremenljivki, druga časovno avtokorelirana

A08. Podvojitev presečnih enot  $N = 50$ ,  $r = 0.3$

A09. Podvojitev časovnih opazovanj  $T = 20$

A10. Podvojitev obeh dimenzij panela  $N = 50$ ,  $T = 20$

A11.  $r(c_i, \bar{x}_i) = 0.3$ ,  $r(c_i, \bar{y}_i) = 0.7$  in ustrezne negativne korelacije

Poudariti je potrebno, da smo glede na osnovni model v različnih preučevanih variantah vedno spreminjali samo po en parameter, kar omogoča natančno ugotavljanje vpliva manipuliranega parametra.

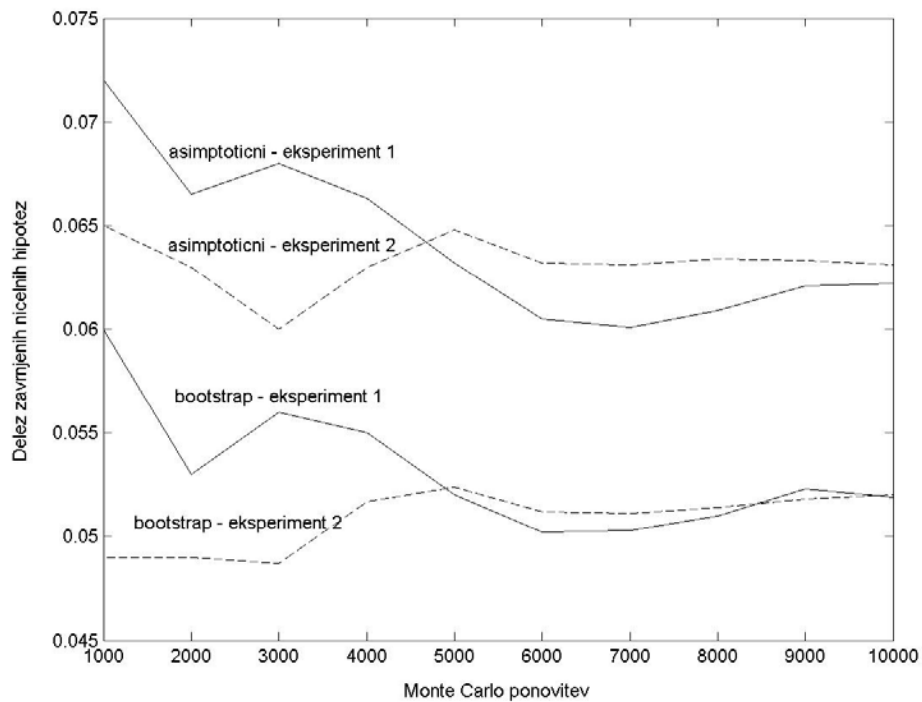
Vsak eksperiment je bil sestavljen iz dveh delov. V prvem delu smo ocenjevali napako I. vrste, zato smo predpostavili model s slučajnimi učinki, kjer časovno povprečje neodvisne spremenljivke ni bilo korelirano s spremenljivko neopazovanih vplivov. V drugem delu smo predpostavili model s fiksnimi učinki in preizkušali moč testa.

#### 4.3.1 Monte Carlo ponovitve

Analiza intervalov zaupanja za asimptotično in bootstrap Hausmanovo statistiko je bila narejena s pomočjo Monte Carlo metode. Za vsako Monte Carlo ponovitev sta bili generirani obe komponenti motnje, s pomočjo katerih je bila, ob fiksnih neodvisnih spremenljivkah, generirana odvisna spremenljivka. Poleg tega je bila ob vsaki ponovitvi simulirana bootstrap porazdelitev Hausmanove statistike.

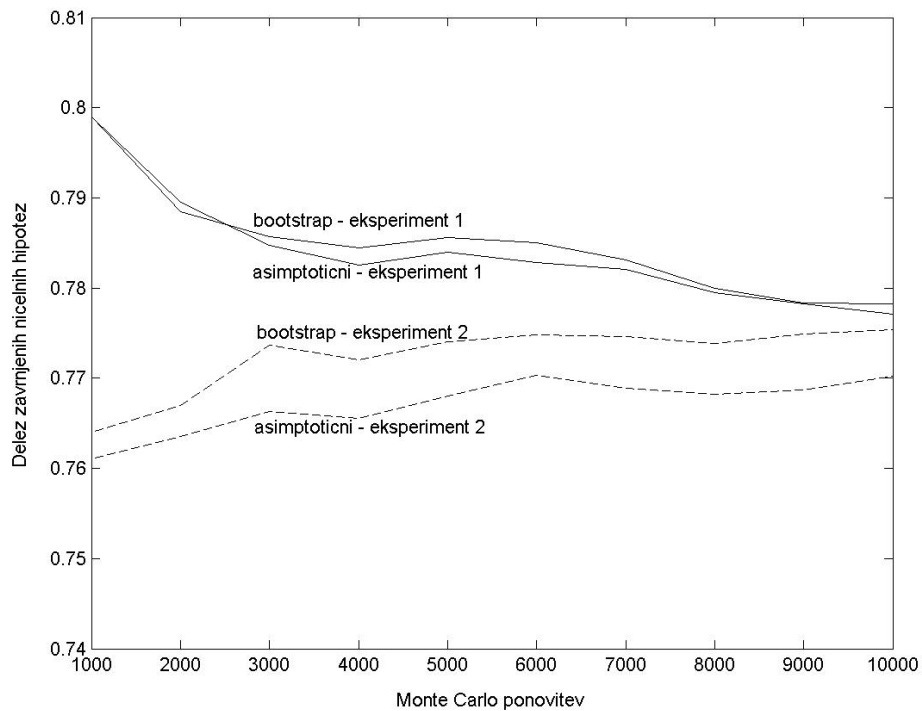
Izbor števila Monte Carlo ponovitev je prikazan na Sliki 1, ki kaže odvisnost deleža zavrnjenih ničelnih hipotez (nekoreliranosti povprečja neodvisne spremenljivke s spremenljivko neopazovanega učinka) od števila Monte Carlo ponovitev za predpostavljeni osnovni model s slučajnimi učinki. Prikazani so rezultati dveh eksperimentov z enakim procesom generiranja podatkov (data generating process). Iz slike ocenjujemo, da je konvergenca obeh eksperimentov dosežena pri okrog 10000 ponovitvah. Podoben sklep lahko naredimo tudi na podlagi Slike 2, ki prikazuje delež zavrnjenih ničelnih hipotez za asimptotični in bootstrap Hausmanov test ob predpostavljenem osnovnem modelu s fiksnimi učinki. Enako število ponovitev je bilo uporabljeno tudi v analizi Anderssona in Karlssona

**Slika 1:** Ilustracija potrebnega števila Monte Carlo ponovitev



Pripomba: Konvergenca Hausmanove statistike za napako I. vrste  
Vir: Lastni izračun

**Slika 2:** Ilustracija potrebnega števila Monte Carlo ponovitev



Pripomba: Konvergenca Hausmanove statistike za moč testa  
Vir: Lastni izračun

(2001), ki sta uporabila bootstrap metodo za regresijski model s komponento motnje za preizkušanje hipotez o regresijskih koeficientih.

#### **4.3.2 Bootstrap ponovitve**

Empirična porazdelitev Hausmanove statistike je bila ocenjena s 400 bootstrap ponovitvami. Ker je funkcija moči testa (glede na njegovo velikost) konkavna (Slika 7), bi povečanje števila bootstrap ponovitev prednost bootstrapa samo povečala (Davidson in MacKinnon, 2000, str. 4). V istem članku sta avtorja s simulacijo pokazala, da pri izbranem številu bootstrap ponovitev (400) znaša izguba največ 1 odstotek moči limitnega testa pri 5 odstotnem tveganju pri ničelni hipotezi. Število bootstrap ponovitev  $B$  naj bi bilo izbrano tako, da je  $\alpha(B+1)$  celo število, vendar empirično nismo ugotovili razlike med 399 in 400 bootstrap ponovitvami. Na Sliki 3 je prikazana empirična porazdelitev Hausmanove statistike za 400 bootstrap ponovitev ob predpostavljenem modelu s slučajnimi učinki. Označen je 5 odstotni kvantil glede na katerega ocenjujemo bootstrap napako I. vrste. Ta je bila pri vseh eksperimentih bližje ciljanemu 5 odstotnemu tveganju, kot pri asimptotičnih ocenah. Manjša napaka I. vrste bootstrap Hausmanovega testa glede na asimptotični test je bila tudi glavna opažena prednost bootstrapa za enostavno strukturno matriko.

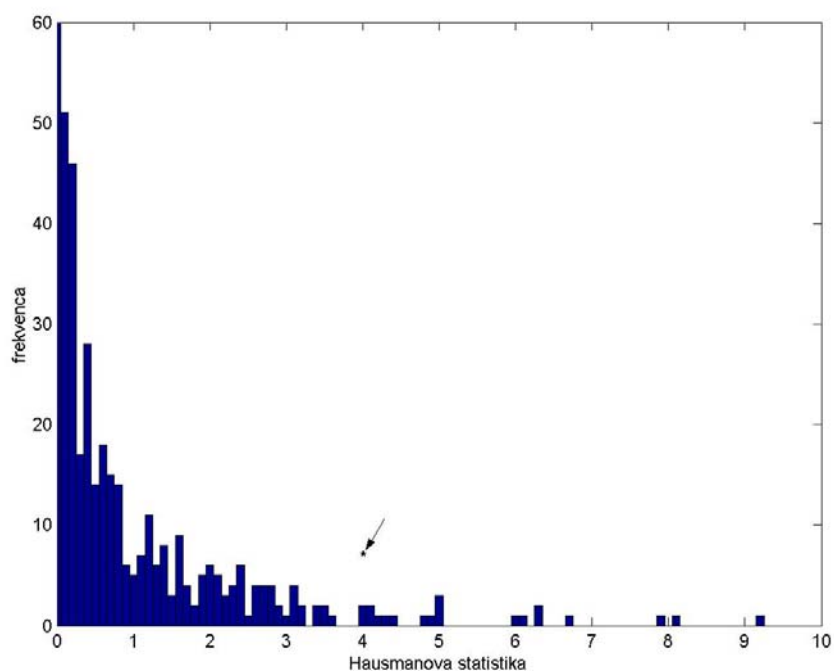
Z asimptotično oceno napake I. vrste smo nato popravili velikost pri ocenjevanju moči bootstrapa Hausmanovega testa v drugem delu eksperimenta, kjer je bil predpostavljen model s fiksnimi učinki. Tako smo naredili oba testa primerljiva. Teoretična pričakovana moč obeh testov je s tem postala enaka (Davidson, MacKinnon, 2000, str. 6). Ker je bila ničelna hipoteza napačna smo primerjali delež zavrnjenih ničelnih hipotez (moč testa) za asimptotični in bootstrap Hausmanov test.

#### **4.3.3 Porazdelitev korelacijskega koeficienta alternativne hipoteze**

Alternativna hipoteza je specificirana s koreliranostjo časovnega povprečja neodvisne spremenljivke in spremenljivke neopazovanega učinka. V ničelni hipotezi je koreliranost



**Slika 3:** Histogram bootstrapa Hausmanove statistike

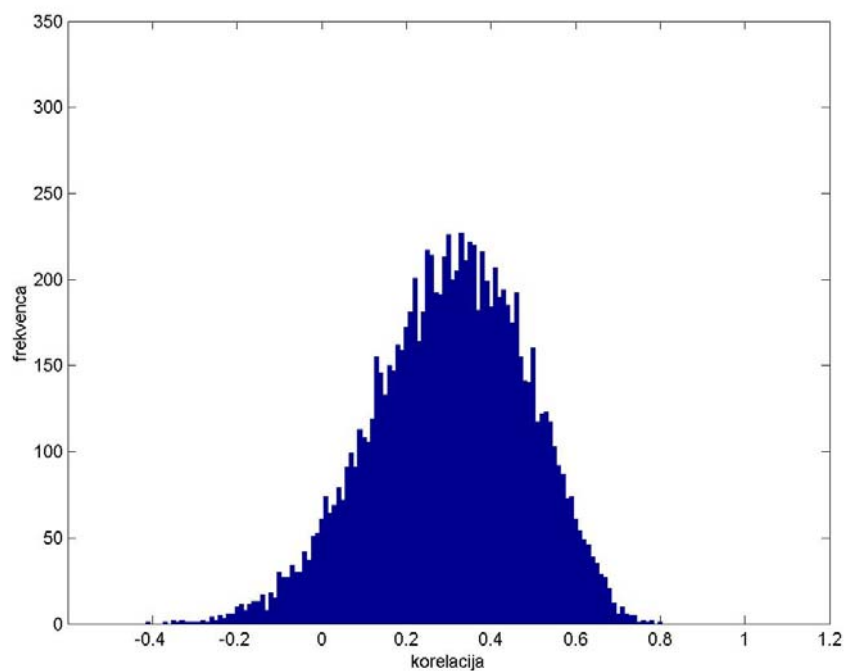


\*kvantilna meja

Pripomba: Histogram za 400 bootstrap Hausmanovih statistik

Vir: Lastni izračun

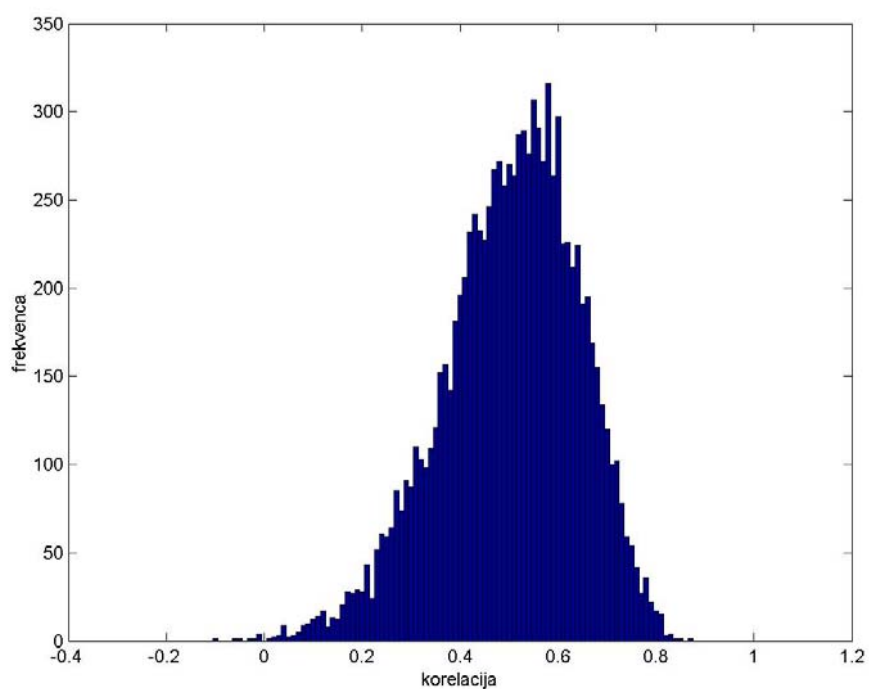
**Slika 4:** Histogram porazdelitve korelacijskega koeficienta za ciljano korelacijo 0.3



Pripomba: Korelacija med časovnim povprečjem neodvisne spremenljivke in slučajno spremenljivko neopazovanega učinka za 10000 Monte Carlo ponovitev

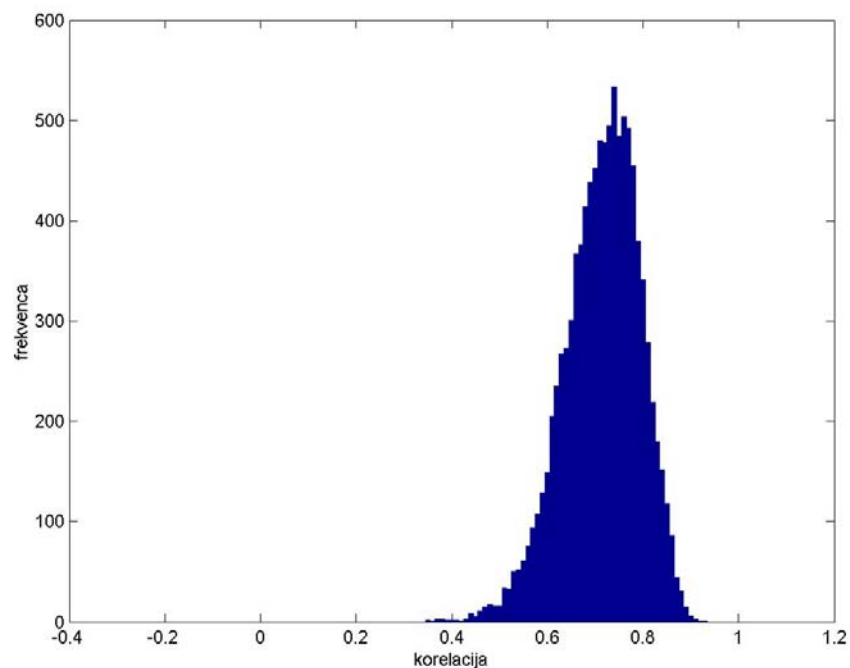
Vir: Lastni izračun

**Slika 5:** Histogram porazdelitve korelacijskega koeficienta za ciljano korelacijo 0.5



Pripomba: Korelacija med časovnim povprečjem neodvisne spremenljivke in slučajno spremenljivko neopazovanega učinka za 10000 Monte Carlo ponovitev  
Vir: Lastni izračun

**Slika 6:** Histogram porazdelitve korelacijskega koeficienta za ciljano korelacijo 0.7



Pripomba: Korelacija med časovnim povprečjem neodvisne spremenljivke in slučajno spremenljivko neopazovanega učinka za 10000 Monte Carlo ponovitev  
Vir: Lastni izračun

enaka nič, v alternativni pa smo preizkušali vplive različnih jakosti koreliranost na moč asimptotičnega in bootstrap Hausmanovega testa. Na Slikah 4, 5 in 6 so prikazani histogrami empiričnih porazdelitev korelacijskih koeficientov za vrednosti korelacije 0.3, 0.5 in 0.7. Iz slik je razvidno, da je variabilnost korelacijskega koeficienta obratno sorazmerna z njegovo velikostjo, kar sicer sledi tudi za njegov teoretični približek za velike vzorce:

$$\text{var}(r) = \frac{(1 - \rho^2)^2}{n},$$

kjer je  $r$  vzorčni korelacijski koeficient,  $\rho$  populacijski,  $n$  pa velikost vzorca. Približek velja za  $n > 500$ .

## 5. Rezultati

V Tabeli 1 je prikazan vpliv preučevanih parametrov na napako I. vrste in na moč testa. V čelu tabele so navedeni osnovni model in spremenjeni parameter alternative. Rezultati učinkov na napako I. vrste so bili naslednji:

- Vse napake I. vrste ustreznega testa so približno enake.
- Asimptotični test v vseh primerih precenjuje ciljano napako I. vrste, bootstrap test pa podcenjuje samo v primeru negativne avtokorelacije  $u_{it}$ .
- Relativna napaka velikosti asimptotičnega testa je v vseh primerih večja od relativne napake velikosti bootstrap testa.
- Napaka I. vrste se bistveno zmanjša s podvojitvijo presečnih enot.
- Napaka I. vrste se poveča z vključitvijo avtokorelirane neodvisne spr. v matriko strukture; pri asimptotičnem bolj kot pri bootstrap testu.

**Tabela 1:** Napaka I. vrste in moč testa

	<i>napaka I. vrste*</i>	
	<i>asimptotični</i>	<i>bootstrap</i>
<i>osnovni model</i>	0.0622	0.0519
$c_i \sim U(0,1/2)$	0.0637	0.0547
$u_{it} \sim U(0,1/2)$	0.0620	0.0520
$\text{var}(u_{it}) = 1$	0.0612	0.0534
<i>časovna avtokorelacija <math>u_{it}</math>: 0.8</i>	0.0616	0.0522
<i>časovna avtokorelacija <math>u_{it}</math>: -0.8</i>	0.0569	0.0481
<i>presečna heteroskedastičnost <math>u_{it}</math></i>	0.0606	0.0518
$N=50$	0.0571	0.0514
$T=20$	0.0616	0.0523
$N=50, T=20$	0.0545	0.0522
<i>2 neodvisni spr.</i>	0.0682	0.0527

	<i>moč testa</i>	
	<i>asimptotični</i>	<i>bootstrap**</i>
<i>osnovni model (<math>r = 0.5</math>)</i>	0.7771	0.7782
$c_i \sim U(0,1/2)$	0.7653	0.7709
$u_{it} \sim U(0,1/2)$	0.7680	0.7719
$\text{var}(u_{it}) = 1$	0.7218	0.7205
<i>časovna avtokorelacija <math>u_{it}</math>: 0.8</i>	0.6381	0.6438
<i>časovna avtokorelacija <math>u_{it}</math>: -0.8</i>	0.8099	0.8013
<i>presečna heteroskedastičnost <math>u_{it}</math></i>	0.7458	0.7453
$N=50$ ( $r=0.3$ )	0.9177	0.9207
$T=20$	0.7170	0.7204
$N=50, T=20$	0.9682	0.9679
<i>2 neodvisni spr.</i>	0.5104	0.5087

\* Ciljana napaka I. vrste:  $\alpha = 0.05$

\*\* Velikost bootstrap testa izenačena z velikostjo ustreznega asimptotičnega testa

Vir: Lastni izračuni

**Tabela 2:** Vpliv ciljane napake I. vrste na velikost in moč testa

	<i>napaka I. vrste</i>	
	<i>asimptotični</i>	<i>bootstrap</i>
$\alpha = 0.05$	0.0622	0.0519
$\alpha = 0.10$	0.1102	0.0967
$\alpha = 0.15$	0.1647	0.1550
$\alpha = 0.20$	0.2083	0.1994

	<i>moč testa</i>	
	<i>asimptotični</i>	<i>bootstrap</i> *
$\alpha = 0.05$	0.7771	0.7782
$\alpha = 0.10$	0.8494	0.8469
$\alpha = 0.15$	0.8960	0.8964
$\alpha = 0.20$	0.9216	0.9212

\* Velikost bootstrap testa izenačena z velikostjo ustreznega asimptotičnega testa

Vir: Lastni izračuni

**Tabela 3:** Vpliv korelacije na moč testa

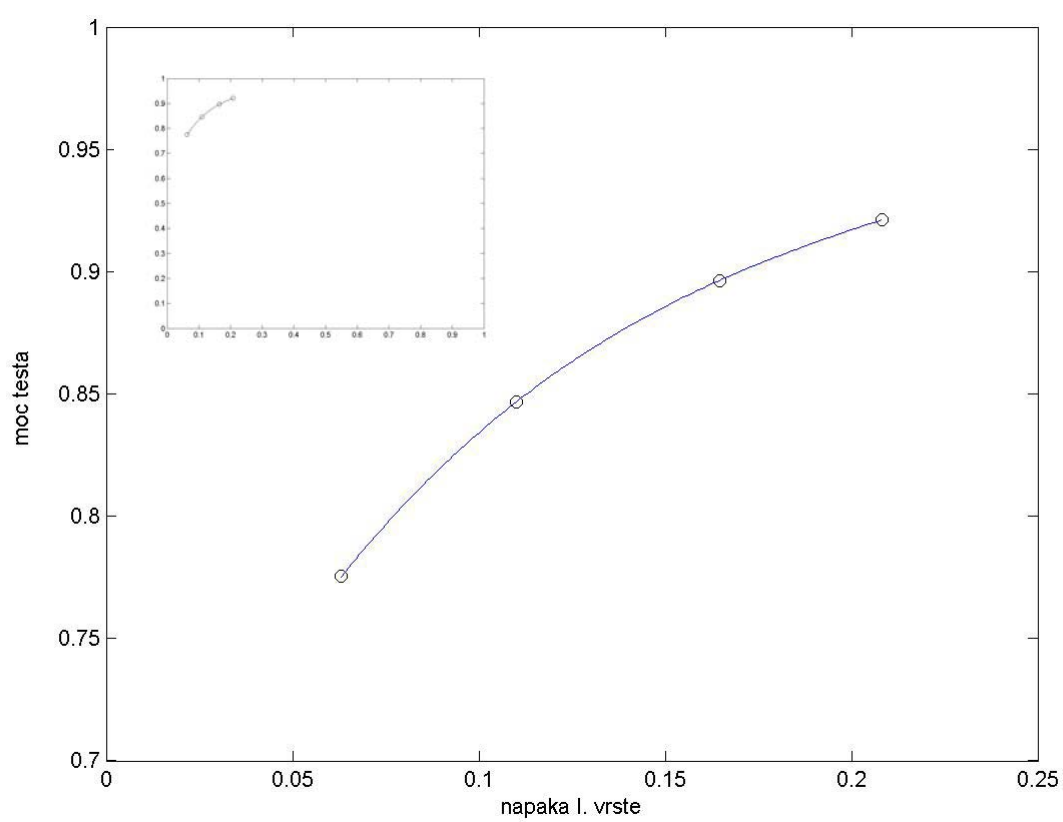
	<i>moč testa</i>	
	<i>asimptotični</i>	<i>bootstrap</i> *
$r = -0.7$	0.9967	0.9969
$r = -0.5$	0.7621	0.7655
$r = -0.3$	0.3392	0.3096
$r = 0$	0.0622	0.0519
$r = 0.3$	0.3361	0.3404
$r = 0.5$	0.7771	0.7782
$r = 0.7$	0.9951	0.9953

\* Velikost bootstrap testa izenačena z velikostjo ustreznega asimptotičnega testa

Pripomba: V čelu tabele je korelacija med časovnim povprečjem neodvisne spremenljivke in slučajno spremenljivko neopazovanega vpliva.

Vir: Lastni izračuni

**Slika 7:** Funkcija moči v odvisnosti od velikosti za bootstrap Hausmanovega testa



Vir: Lastni izračun

Rezultati učinkov na moč testa:

- Popravljen velikost bootstrap testa izenači moč bootstrap in asimptotičnega testa.
- Podvojitve variance šuma rahlo poslabša moč obeh testov.
- Pri negativni avtokorelaciji  $u_{it}$  se moč obeh testov rahlo izboljša, pri pozitivni pa rahlo zmanjša glede na osnovni model.
- Vključitev avtokorelirane neodvisne spremenljivke v matriko strukture zmanjša moč obeh testov.
- S podvojitvijo  $N$  se bistveno poveča moč obeh testov.
- S podvojitvijo  $T$  se rahlo poslabša moč obeh testov.
- S podvojitvijo  $N$  in  $T$  se poveča moč obeh testov.

Vpliv posameznih parametrov na napako I. vrste in moč testa je majhen kar kaže na robustnost testa na spremembe v parametrih.

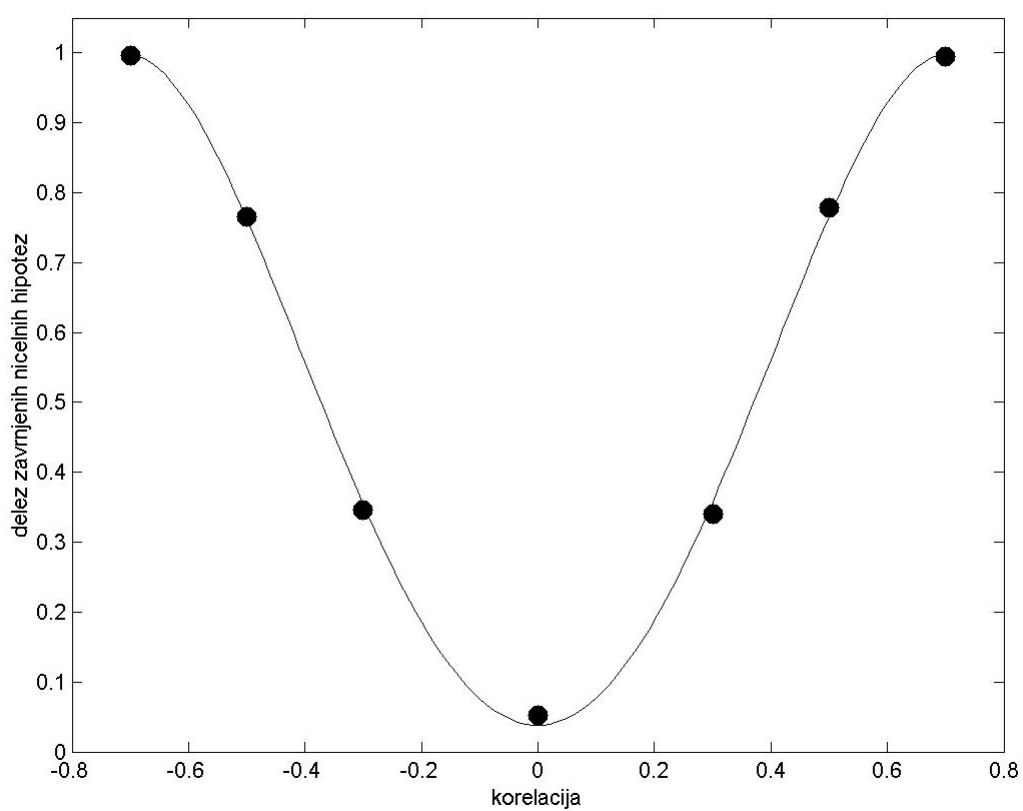
Tabela 2 prikazuje vpliv ciljne velikosti na velikost in moč testa. V čelu tabele so navedene ciljne velikosti testa. Ključne empirične ugotovitve so naslednje:

- Pri vseh štirih ciljanih velikostih asimptotični test precenjuje velikost testa.
- Relativna napaka velikosti asimptotičnega testa je v vseh primerih večja kot od bootstrap testa.
- Odnos med velikostjo in močjo bootstrap testa je prikazan tudi na Sliki 7. Lega krivulje prileganja kaže na dobro karakteristiko moči testa glede na njegovo velikost.

Vpliv korelacije časovnega povprečja neodvisne spremenljivke in slučajne spremenljivke neopazovanih učinkov je prikazan v Tabeli 3. V čelu tabele nastopa korelacijski koeficient. Omeniti velja naslednje rezultate:

- Moč asimptotičnega in bootstrap Hausmanovega testa je simetrična glede na predznak korelacije.
- Moč narašča približno s kvadratom korelacije; odnos je prikazan s krivuljo prileganja na Sliki 8 za bootstrap test.

**Slika 8:** Funkcija moči bootstrap Hausmanovega testa za ciljano velikost 0.05



Pripomba: Korelacija med časovnim povprečjem neodvisne spremenljivke in slučajno spremenljivko neopazovanega učinka

Vir: Lastni izračun



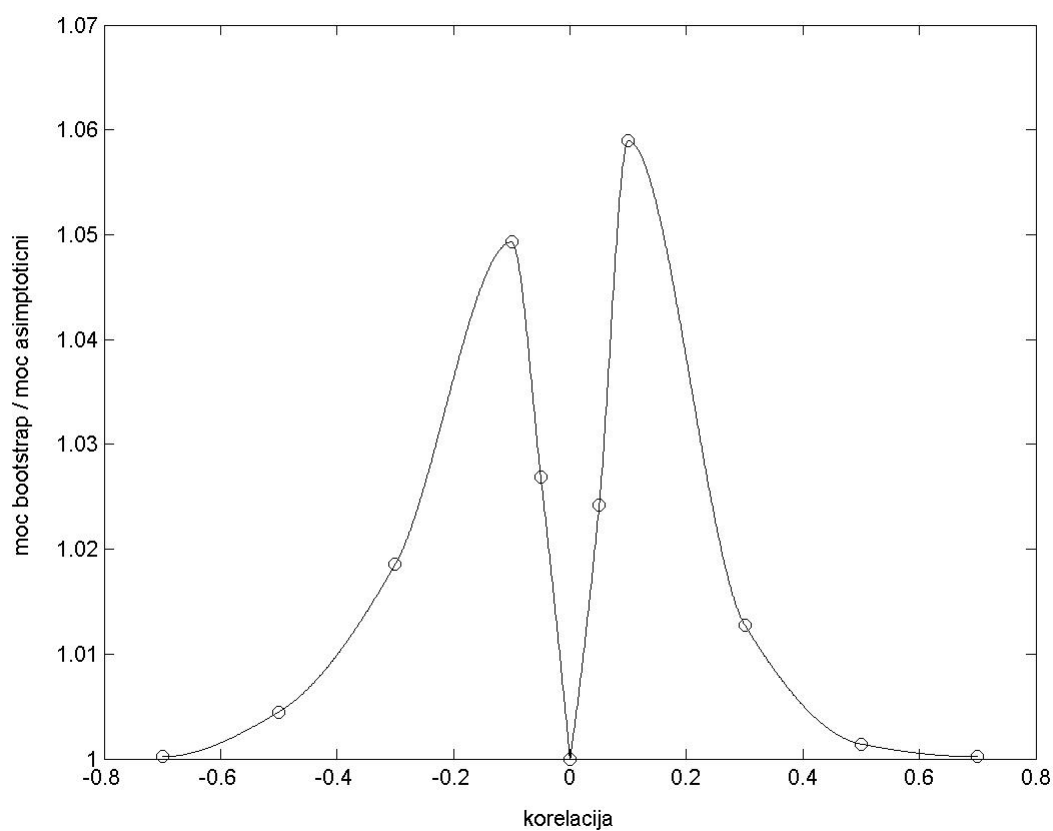
**Tabela 4:** Kvocient moči bootstrap in asimptotičnega Hausmanovega testa

	$k$
$r = -0.7$	1.0002
$r = -0.3$	1.0045
$r = -0.5$	1.0186
$r = -0.1$	1.0493
$r = -0.05$	1.0269
$r = 0$	1
$r = 0.05$	1.0242
$r = 0.1$	1.0590
$r = 0.3$	1.0128
$r = 0.5$	1.0014
$R = 0.7$	1.0002

Pripomba: V glavi tabele je kvocient moči testov, v čelu pa korelacijski koeficient.

Vir: Lastni izračuni

**Slika 9:** Relativna moč bootstrap Hausmanovega testa



Pripomba: Korelacija med časovnim povprečjem neodvisne spremenljivke in slučajno spremenljivko neopazovanega učinka

Vir: Lastni izračun

Relativna prednost bootstrap Hausmanovega testa glede na asimptotični v odvisnosti od korelacije je prikazana v Tabeli 4. Ocene kvocienta moči obeh verzij testa, ki niso razvidne iz predhodnih tabel, so bile zaradi časovne intenzivnosti eksperimenta ocenjene s 5000 Monte Carlo simulacijami. V glavi tabele nastopa kvocient moči, v čelu pa korelacijski koeficient. Eksperiment je pokazal:

- Prednost bootstrap Hausmanovega testa hitro naraste s povečevanjem korelacije in nato začne upadati; največja je pri nizkih korelacijah, ki jih asimptotični test ne uspe identificirati, vendar je pri zelo nizkih korelacijah relativna prednost bootstrap testa zopet manjša.
- Približno simetričnost glede na predznak korelacije kar je razvidno tudi iz Slike 9.

V osnovnem modelu in v vseh alternativah, razen pri spremenjeni matriki strukture in spremenjeni dimenziji panela, asimptotični test precej napako I. vrste bootstrap testa za približno 20 odstotkov oz. med 0.0078 in 0.0103 odstotne točke. S pomočjo Izreka D2.2 (glej dodatek D2) lahko pokažemo, da napaka I. vrste asimptotičnega testa tudi teoretično presega napako I. vrste bootstrap testa in sicer za najmanj 0.004 in največ 0.009 odstotne točke.

Empirična razlika med močjo za velikost popravljenega bootstrap in asimptotičnega testa je zanemarljiva in potrjuje teoretično napoved (Davidson in MacKinnon, 2002).

## 6. Zaključek

### Cilj in značilnosti eksperimenta

V nalogi smo analizirali bootstrap Hausmanovega specifikacijskega testa za regresijski model s strukturo motnje. Obravnavali smo vpliv različnih parametrov procesa generiranja podatkov. S primerjavo osnovnega modela in več alternativ smo analizirali vpliv različne jakosti koreliranosti časovnega povprečja neodvisne spremenljivke s slučajno spremenljivko neopazovanega učinka, ter vpliv dimenzije panela na Hausmanov test. Poleg tega smo analizirali še vpliv heteroskedastičnosti in časovne avtokorelacije šuma, različne relativne variance obeh komponent motnje in primer z dvema neodvisnima spremenljivkama, kjer je bila ena izmed njiju časovno avtokorelirana.

### Povzetek najpomembnejših ugotovitev, ki jih nismo zasledili v dostopni literaturi:

- Povečanje števila časovnih opazovanj vpliva na zmanjšanje moči testa.
- Nasproten vpliv pozitivne in negativne avtokorelacije šuma na moč testa.
- Vpliv matrike strukture na moč testa: Vključitev avtokorelirane neodvisne spremenljivke zmanjša moč testa.
- Odstopanje moči bootstrap testa s popravljeno velikostjo od asimptotičnega je majhno.

Asimptotični Hausmanov test v vseh analiziranih primerih precenjuje ciljno velikost testa. Relativna napaka velikosti asimptotičnega testa je bila v vseh primerih večja od bootstrap testa in povsem v skladu s teoretično razliko. Ocenjena razlika med močjo za velikost popavljenega bootstrap in asimptotičnega testa je zanemarljiva in potrjuje teoretično napoved.

## Dodatek D1

POMOŽNA TRDITEV D1.1: Naj bo  $\{\mathbf{Z}_N : N = 1, 2, \dots\}$  zaporedje slučajnih matrik  $K \times K$ , in naj bo  $\mathbf{A}$  neslučajna, obrnljiva matrika  $K \times K$ . Če velja  $\mathbf{Z}_N \xrightarrow{p} \mathbf{A}$ , potem sledi:

- (1)  $\mathbf{Z}_N^{-1}$  obstaja z verjetnostjo, ki se približuje 1;
- (2)  $\text{plim} \mathbf{Z}_N^{-1} = \mathbf{A}^{-1}$ .

LEMA (Slutsky) D1.2: Naj bo  $\mathbf{g} : \mathbb{R}^K \rightarrow \mathbb{R}^l$  zvezna v točki  $\mathbf{c} \in \mathbb{R}^K$  in naj bo  $\{\mathbf{x}_N : N = 1, 2, \dots\}$  zaporedje slučajnih vektorjev  $K \times 1$  tako, da  $\mathbf{x}_N \xrightarrow{p} \mathbf{c}$ . Potem  $\mathbf{g}(\mathbf{x}_N) \xrightarrow{p} \mathbf{g}(\mathbf{c})$ , ko  $N \rightarrow \infty$  ( $\text{plim} \mathbf{g}(\mathbf{x}_N) = \mathbf{g}(\text{plim} \mathbf{x}_N)$ ), če je  $\mathbf{g}(\cdot)$  zvezna v  $\text{plim} \mathbf{x}_N$ .

### DEFINICIJA D1.3

- (1) Zaporedje  $\{a_N\}$  je  $O(N^\lambda)$  (največ reda  $N^\lambda$ ), če je  $N^{-\lambda} a_N$  omejeno. Če  $\lambda = 0$  je  $\{a_N\}$  omejeno, kar zapišemo kot  $a_N = O(1)$ .
- (2)  $\{a_N\}$  je  $o(N^\lambda)$ , če  $N^{-\lambda} a_N \rightarrow 0$ . Če  $\lambda = 0$ ,  $a_N$  konvergira k 0, kar zapišemo kot  $a_N = o(1)$ .

### DEFINICIJA D1.4

- (1) Zaporedje slučajnih spremenljivk  $\{x_N : N = 1, 2, \dots\}$  konvergira v verjetnosti h konstanti  $a$ , če  $\forall \varepsilon > 0$  velja  $P(|x_N - a| > \varepsilon) \rightarrow 0$ , ko  $N \rightarrow \infty$ . To lahko zapišemo tudi kot  $\text{plim} x_N = a$ .
- (2) Če je  $a = 0$ , potem  $x_N = o_p(1)$ .
- (3) Zaporedje slučajnih spremenljivk  $\{x_N\}$  je omejeno v verjetnosti, če in samo če  $\forall \varepsilon > 0 \exists b_\varepsilon < \infty$  in naravno število  $N_\varepsilon$  tako, da  $P(|x_N| \geq b_\varepsilon) < \varepsilon$  za  $\forall N \geq N_\varepsilon$  ( $x_N = O_p(1)$ ).

DEFINICIJA D1.5

(1) Slučajno zaporedje  $\{x_N : N=1,2,\dots\}$  je  $o_p(a_N)$ , kjer je  $\{a_N\}$  neslučajno, pozitivno zaporedje, če  $\frac{x_N}{a_N} = o_p(1)$ .

(2) Slučajno zaporedje  $\{x_N : N=1,2,\dots\}$  je  $O_p(a_N)$ , kjer je  $\{a_N\}$  neslučajno, pozitivno zaporedje, če  $\frac{x_N}{a_N} = O_p(1)$ .

IZREK D1.6: Naj bo  $\{\mathbf{w}_i : i=1,2,\dots\}$  zaporedje neodvisnih, identično porazdeljenih slučajnih vektorjev  $G \times 1$  tako, da  $E(|w_{ig}|) < \infty$ ,  $g=1,2,\dots,G$ . Potem zaporedje zadošča šibkemu zakonu velikih števil:  $N^{-1} \sum_{i=1}^N \mathbf{w}_i \xrightarrow{p} \boldsymbol{\mu}_w$ , kjer je  $\boldsymbol{\mu}_w \equiv E(\mathbf{w}_i)$ .

IZREK (Lindeberg-Levy) D1.7: Naj bo  $\{\mathbf{w}_i : i=1,2,\dots\}$  zaporedje neodvisnih, identično porazdeljenih slučajnih vektorjev  $G \times 1$  tako, da  $E(w_{ig}^2) < \infty$ ,  $g=1,2,\dots,G$ , in  $E(\mathbf{w}_i) = 0$ . Potem  $\{\mathbf{w}_i : i=1,2,\dots\}$  zadošča centralnemu limitnemu izreku (CLT):

$$N^{-1/2} \sum_{i=1}^N \mathbf{w}_i \xrightarrow{d} N(0, \mathbf{B})$$

kjer je  $\mathbf{B} = \text{Var}(\mathbf{w}_i) = E(\mathbf{w}_i \mathbf{w}_i^T)$  pozitivno semidefinitna.

LEMA (Asimptotična ekvivalenca) D1.8: Naj bosta  $\{\mathbf{x}_N\}$  in  $\{\mathbf{z}_N\}$  zaporedji slučajnih vektorjev  $K \times 1$ . Če  $\mathbf{z}_N \xrightarrow{d} \mathbf{z}$  in  $\mathbf{x}_N - \mathbf{z}_N \xrightarrow{p} 0$ , potem  $\mathbf{x}_N \xrightarrow{d} \mathbf{z}$ .

POMOŽNA TRDITEV D1.9: Če je  $\{\mathbf{z}_N\}$  zaporedje slučajnih vektorjev  $K \times 1$  tako, da  $\mathbf{z}_N \xrightarrow{d} N(0, \mathbf{V})$ , potem

(1) za vsako  $K \times M$  neslučajno matriko  $\mathbf{A}$ ,  $\mathbf{A}^T \mathbf{z}_N \xrightarrow{d} N(0, \mathbf{A}^T \mathbf{V} \mathbf{A})$ ,

(2)  $\mathbf{z}_N^T \mathbf{V}^{-1} \mathbf{z}_N \xrightarrow{d} \chi_K^2$ .

## Dodatek D2: Edgeworthov razvoj porazdelitve Hausmanove statistike

V nadaljevanju povzemamo Edgeworthov razvoj porazdelitve Hausmanove statistike po Boletu (2003).

Naj bo model s strukturo motnje

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}, \quad (\text{D2.1})$$

kjer je vektor ostankov  $\mathbf{v}$  porazdeljen  $\mathbf{v} \sim N(0, \boldsymbol{\Sigma})$ ,  $\boldsymbol{\Sigma} = \sigma_u^2 \left( \frac{1}{\psi^2} \mathbf{P} + \mathbf{Q} \right)$ , pri tem sta  $\mathbf{P}$  in  $\mathbf{Q}$  "between" oziroma "within" projektorja ( $\mathbf{P} = \mathbf{I}_N \otimes \bar{\mathbf{J}}_T$ , oziroma  $\mathbf{Q} = \mathbf{I} - \mathbf{P}$ ),  $\mathbf{y}$  je vektor  $NT \times 1$ ,  $\mathbf{X}$  matrika neodvisnih spremenljivk  $NT \times K$  in  $\boldsymbol{\beta}$  vektor  $K \times 1$ . Testiranje fiksnih učinkov glede na ničelno hipotezo slučajnih učinkov s pomočjo Hausmanove statistike (HT) v takšnem modelu je ekvivalentno testiranju ničelnosti parametra  $\gamma$  v modelu

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \tilde{\mathbf{X}} \boldsymbol{\gamma} + \mathbf{w}, \quad (\text{D2.2})$$

kjer je  $\psi = \frac{\sigma_u}{(T\sigma_c^2 + \sigma_u^2)^{1/2}}$ ,  $\mathbf{y}^* = \sigma_u \boldsymbol{\Sigma}^{-1/2} \mathbf{y}$  vektor  $n \times 1$ ,  $\mathbf{X}^* = \sigma_u \boldsymbol{\Sigma}^{-1/2} \mathbf{X}$  matrika  $n \times K$ ,  $\tilde{\mathbf{X}} = \mathbf{Q}\mathbf{X}$  matrika  $n \times K$  in  $\mathbf{w} \sim N(0, \sigma_u^2)$  vektor  $n \times 1$  in seveda  $n = NT$  (Baltagi, 2002, str.69). Zaradi Frisch-Waugh-Lovellovega izreka (glej, naprimer, Davidson-Mackinnon, 1994, str. 19) je testiranje ničelnosti parametra ekvivalentno testiranju ničelnosti parametra v modelu

$$\mathbf{M}_{\mathbf{X}^*} \mathbf{y}^* = \mathbf{M}_{\mathbf{X}^*} \tilde{\mathbf{X}} \boldsymbol{\gamma} + \mathbf{M}_{\mathbf{X}^*} \mathbf{w}, \quad (\text{D2.3})$$

ki ima število pojasnjevalnih spremenljivk manjše za število motečih (»nuisance«) parametrov (D2.2) in  $\mathbf{M}_{\mathbf{X}^*} = \mathbf{I} - \mathbf{X}^* (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T}$ . Če obe strani relacije (D2.3) pomnožimo z ortogonalno matriko  $\mathbf{P}$ , ki diagonalizira  $\mathbf{M}_{\mathbf{X}^*}$  imajo motnje v novem modelu  $\mathbf{v}2 = \mathbf{P}\mathbf{M}_{\mathbf{X}^*} \mathbf{w}$  enako varianco  $\sigma_u^2$  ter pričakovano vrednost 0. Če so prvotne motnje ( $\mathbf{v}$ ) porazdeljene normalno (in neodvisno), so transformirane motnje ( $\mathbf{v}2$ ) porazdeljene normalno in neodvisno. Če je alternativna hipoteza v modelu (D2.1) običajne oblike  $E(v_{it} | \mathbf{X}) = \bar{\mathbf{x}}_i^T \boldsymbol{\delta}$  (glej (4.9))

lahko brez zmanjšanja splošnosti predpostavimo, da je  $\gamma = \tilde{\gamma}e$ , pri čemer je  $e$  znan vektor (naprimer enic) in  $\tilde{\gamma}$  neznan parameter. Ker je v modelu (D2.3) le ena pojasnjevalna spremenljivka  $z = \mathbf{PM}_X \cdot \mathbf{QXe}$  se lahko, če so motnje  $v_2$  enako porazdeljene (naprimer, v normalnem primeru) neposredno uporabijo relacije Edgeworth-ovega razvoja (ER) (glej, Hall, 1992) za naklonski koeficient v enostavni regresiji v obliki

$$P\left(n^{-1/2}(\hat{\gamma} - \tilde{\gamma}_0) \sigma(z) / \hat{\sigma}_u < \xi\right) = \Phi(\xi) + n^{-1/2} q_1(\xi) \phi(\xi) + n^{-1} q_2(\xi) \phi(\xi) + \dots, \quad (\text{D2.4})$$

kjer je

$$\tilde{\gamma}_0 = 0,$$

$$\sigma(z) = \sigma(\mathbf{PM}_X \cdot \mathbf{QXe}) = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z}_i)^2,$$

$$q_1(\xi) = -\frac{1}{6} \lambda \lambda_z (\xi^2 - 1),$$

$$q_2(\xi) = -\xi \left(2 + \frac{1}{24} (\kappa \kappa_z + 6) (\xi^2 - 3) + \frac{1}{72} \lambda^2 \lambda_z^2 (\xi^4 - 10\xi^2 + 15)\right),$$

in sta  $\lambda$ ,  $\kappa$  alternativna indikatorja asimetričnosti in sploščenosti standarizirane motnje  $(\frac{v_2^k}{\sigma_u})$ , torej  $\lambda = E\left(\frac{v_2^k}{\sigma_u}\right)^3$  oziroma  $\kappa = E\left(\frac{v_2^k}{\sigma_u}\right)^4 - 3$ . Podobno sta  $\lambda_z$  in  $\kappa_z$  alternativna empirična indikatorja asimetričnosti in sploščenosti za pojasnjevalno spremenljivko  $z$  formalno izračunana na danih vzorčnih vrednostih.

Relacija (D2.4) omogoča ER za Hausmanovo statistiko saj je  $P(HT \leq \xi^2) =$   
 $= P\left(n^{-1/2}(\hat{\gamma} - \tilde{\gamma}_0) \sigma(z) / \hat{\sigma}_u \leq \xi\right)$ .

V primeru normalnih motenj ( $v$ ), so tudi transformirane motnje normalne, neodvisno in enako porazdeljene in zato oba parametra  $\lambda$ ,  $\kappa$  enaka 0. Ciljana velikost asimptotičnega testa je torej v normalnem primeru, pri znanem  $\psi$ , točna do vključno reda  $n^{-1}$  poleg tega členi reda  $n^{-1}$  niso odvisni niti od parametrov modela niti od strukturne matrike, saj je očitno  $q_1(\xi)$  enak 0, v  $q_2(\xi)$  pa so od nič različni le členi, ki niso odvisni niti od parametrov

modela niti od matrike strukture; zato je tudi velikost bootstrap testa napačna največ za člene reda  $O(n^{-2})$ . Formulirajmo naslednji izrek

IZREK D2.1: V modelu s strukturo motnje (D2.1) z danim razmerjem obeh spektralnih vrednosti variančno kovariančne matrike motnje  $v$  ( $\psi = \frac{\sigma_u}{(T\sigma_c^2 + \sigma_u^2)^{1/2}}$ ) se velikost testa HT razlikuje od ciljne velikosti za člene, ki niso večji od reda velikosti  $O(n^{-1})$ . Velikost bootstrap testa je napačna največ za člene reda  $O(n^{-2})$ . Če je  $\xi_\alpha^2$  kvantil  $\chi^2$  porazdelitve za ciljano velikost (testa)  $\alpha$  potem je dejanska velikost testa enaka

$$\alpha_{eff} = \alpha + \frac{2}{n} \left( \xi_\alpha \left( 2 + \frac{1}{4} (\xi_\alpha^2 - 3) \right) \phi(\xi_\alpha) \right) + O(n^{-3/2})$$

in je  $\phi(\xi_\alpha)$  vrednost funkcije verjetnostne gostote standardizirane normalne spremenljivke pri  $\xi_\alpha$ .

Edgeworthov razvoj za studentizirano statistiko, ki je ekvivalentna HT statistiki v modelu (D2.1), torej Edgeworthov razvoj HT statistike v primeru, če je parameter  $\psi$  neznan (vendar dosledno ocenjen), motnja  $v$  v modelu (D2.1) pa normalno porazdeljena, se lahko dobi s pomočjo Edgeworthovega razvoja »gladke funkcije« vektorjev slučajnih spremenljivk (glej definicijo družine »gladkih funkcij« pri ER v Hall, 1992). Izpeljava je enostavna posplošitev ER porazdelitve naklonskega koeficienta v regresijskem modelu (glej Hall, 1992), vendar je potrebno model (D2.1) drugače transformirati kakor v primeru znanega parametra  $\psi$ .

IZREK D2.2: V modelu s strukturo motnje (D2.1) z normalno porazdeljeno motnjo in konstatnim razmerjem  $N$  in  $T$  se velikost testa HT razlikuje od ciljne velikosti za člene, ki so reda velikosti  $O(n^{-1})$ . Členi reda  $O(n^{-1})$  so večji od ustreznih členov pri HT testu z znanim razmerjem obeh spektralnih vrednosti variančno kovariančne matrike motnje  $v$ , torej večji od

$$\frac{2}{n} \xi_\alpha \left( 2 + \frac{1}{4} (\xi_\alpha^2 - 3) \right) \phi(\xi_\alpha).$$

Velikost bootstrap testa je napačna za člene največ reda  $O(n^{-3/2})$ .



Dokaz izreka je daljši kakor v primeru znanega parametra  $\psi$  vendar vsebinsko prav tako sledi dokaz za ER porazdelitve naklonskega koeficienta v navadni regresiji (glej Hall, 1992). Ker je  $\mathbf{Q} + \mathbf{P} = \mathbf{I}$ , obstoja ortogonalna matrika  $\mathbf{R}$ , ki diagonalizira oba projektorja istočasno, lastne vrednosti (diagonalni elementi) tako  $\mathbf{P}$  in  $\mathbf{Q}$  pa so 1 ali 0. Transformirani vektor motenj  $\mathbf{R}\mathbf{v}$  je vsota dveh vektorjev motenj

$$\mathbf{v}_{2_B} = \mathbf{R}\mathbf{v}_B \quad \text{in} \quad \mathbf{v}_{2_W} = \mathbf{R}\mathbf{v}_W, \quad \text{kjer je} \quad \mathbf{v}_B = \mathbf{P}\mathbf{v} \quad \text{in} \quad \mathbf{v}_W = \mathbf{Q}\mathbf{v}. \quad (\text{D2.5})$$

Za transformirane motnje  $\mathbf{v}_{2_B}$  in  $\mathbf{v}_{2_W}$  velja  $\text{Cov}(\mathbf{v}_{2_B}, \mathbf{v}_{2_W}) = 0$ ; njune komponente pa so neodvisno in enako porazdeljene po normalnih porazdelitvah

$$\mathbf{v}_{2_B} \sim N(0, \text{diag}(\sigma_1^2)_T) \quad \text{in} \quad \mathbf{v}_{2_W} \sim N(0, \text{diag}(\sigma_u^2)_{T(N-1)}), \quad (\text{D2.6})$$

kjer je  $\sigma_1^2 = T\sigma_c^2 + \sigma_u^2$ , z  $\text{diag}(\alpha)_m$  pa je označena diagonalna matrika, ki ima na diagonali natanko  $m$  neničelnih elementov (ki so enaki  $\alpha$ ), vsi ostali pa so enaki 0. Vektor motenj  $\mathbf{v}_{2_B}$  potemtakem obsega le  $T$  neodvisnih, enako (normalno) porazdeljenih transformiranih motenj s pričakovano vrednostjo 0 in varianco  $\sigma_1^2$ ,

vsil ostali elementi vektorja  $\mathbf{v}_{2_B}$  so identično enaki 0. Analogno velja za vektor transformiranih motenj  $\mathbf{v}_{2_W}$ .

Označimo z

$$\frac{\mathbf{X}^T \mathbf{Q} \mathbf{X}}{n} = \mathbf{W}_{XX} \quad \text{in} \quad \frac{\mathbf{X}^T \mathbf{P} \mathbf{X}}{n} = \mathbf{B}_{XX}, \quad (\text{D2.7})$$

kjer sta  $\mathbf{Q}$  in  $\mathbf{P}$  »between« in »within« projektorja in  $n = NT$ . V ekvivalentnem modelu (D2.2) sta ocenjevani parameter  $\hat{\gamma}$  in njegova varianca  $\hat{\text{var}}(\hat{\gamma})$  v splošnem enaka

$$\hat{\gamma} = n^{-1} (\mathbf{W}_{XX}^{-1} \mathbf{X}^T \mathbf{Q} - \mathbf{B}_{XX}^{-1} \mathbf{X}^T \mathbf{P}) \mathbf{v}$$

in (D2.8)

$$\text{vâr}(\boldsymbol{\gamma}) = n^{-1}(\sigma_u^2 \mathbf{W}_{\text{xx}}^{-1} + \sigma_1^2 \mathbf{B}_{\text{xx}}^{-1}).$$

Če se upošteva, da je v modelu, ki je uporabljen v Monte Carlo eksperimentu matrika  $\mathbf{X}$   $NT \times 1$ , se lahko nadaljnje izvajanje poenostavi, saj so  $\mathbf{B}_{\text{xx}}$ ,  $\mathbf{W}_{\text{xx}}$  in  $\boldsymbol{\gamma}$  skalarji. Če upoštevamo to poenostavitev, sta kvantilna in  $t$ -kvantilna (studentizirana) statistika enaki

$$S = \frac{n^{1/2} (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) (\mathbf{B}_{\text{xx}} \mathbf{W}_{\text{xx}})^{1/2}}{(\sigma_u^2 \mathbf{B}_{\text{xx}} + \sigma_1^2 \mathbf{W}_{\text{xx}})^{1/2}}, \quad (\text{D2.9})$$

$$T = \frac{n^{1/2} (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) (\mathbf{B}_{\text{xx}} \mathbf{W}_{\text{xx}})^{1/2}}{(\hat{\sigma}_u^2 \mathbf{B}_{\text{xx}} + \hat{\sigma}_1^2 \mathbf{W}_{\text{xx}})^{1/2}}. \quad (\text{D2.10})$$

Seveda je pri testiranju HT testa  $\boldsymbol{\gamma}_0 = 0$ . Obe statistiki sta seveda, pri določanju kritičnih območij enakih repov, ekvivalentni statistiki HT (saj je njun kvadrat enak statistiki HT).

Pri oceni hitrosti približevanja (v verjetnosti)  $T$  k  $S$  lahko priredimo postopek, ki je uporabljen v Hall(1992) na strani 172-175.

Ker so transformirane motnje  $\mathbf{Rv}$  neodvisne, lahko ocenjen izraz  $\hat{\Omega}$  (pod korenem v imenovalcu (D2.10)) preoblikujemo v obliko »gladke funkcije« (momentov neodvisnih slučajnih spremenljivk-motenj); takšna funkcija je namreč zadostni pogoj za Edgeworth-ov razvoj studentizirane statistike  $T$  (glej Hall, 1992). Kot rečeno,  $T$  statistika določa ekvivalentna kritična območja enakih repov kot statistika HT. Če je  $N = cn^{1/2}$ , velja za poljubni dosledni cenilki varianc  $\sigma_u^2$ ,  $\sigma_1^2$

$$\hat{\Omega} = (\sigma_u^2 \mathbf{B}_{\text{xx}} + \sigma_1^2 \mathbf{W}_{\text{xx}}) + \mathbf{B}_{\text{xx}} \left( \frac{1}{N(T-1)} \mathbf{v}_{\mathbf{w}}^T \mathbf{Q} \mathbf{v}_{\mathbf{w}} - \sigma_u^2 \right) + \mathbf{W}_{\text{xx}} \left( \frac{1}{N} \mathbf{v}_{\mathbf{B}}^T \mathbf{P} \mathbf{v}_{\mathbf{B}} - \sigma_1^2 \right) + O_p(n^{-1}) \quad (\text{D2.11})$$

Očitno je za običajni cenilki parametrov motnje

$$\hat{\sigma}_1^2 = \frac{1}{N} \mathbf{v}_B^T \mathbf{P} \mathbf{v}_B \quad \text{ter} \quad \hat{\sigma}_u^2 = \frac{1}{N(T-1)} \mathbf{v}_W^T \mathbf{Q} \mathbf{v}_W$$

v relaciji (D2.11) ostanek kar enak 0!

Če v relacijo (D2.11) uvrstimo transformirano motnjo  $\mathbf{v}_2 = \mathbf{R} \mathbf{v}$  (ki je vsota dveh vektorjev, katerih komponente so neodvisne in enako porazdeljene slučajne spremenljivke), se relacija (D2.11) spremeni v obliko, ki ustreza defniciji »gladke funkcije« (ustrezen izraz je torej primeren za ER)

$$\hat{\Omega} = (\sigma_u^2 \mathbf{B}_{XX} + \sigma_1^2 \mathbf{W}_{XX}) + \left[ \mathbf{B}_{XX} \frac{1}{N(T-1)} \sum (v_2^2_{w_i} - \sigma_u^2) + \mathbf{W}_{XX} \frac{1}{N} \sum (v_2^2_{B_i} - \sigma_1^2) \right] + O_p(n^{-1}). \quad (\text{D2.12})$$

Norma vektorjev  $\mathbf{v}_B$  oziroma  $\mathbf{v}_W$  (v metriki  $\mathbf{P}$  oziroma  $\mathbf{Q}$ ) je seveda enaka evklidski normi vektorjev  $\mathbf{v}_{2B}$  oziroma  $\mathbf{v}_{2W}$  (neničelnih elementov v zadnjih dveh vektorjih je le  $N$  oziroma  $N(T-1)$ , toliko kot je rank ustreznih matrik  $\mathbf{P}$  oziroma  $\mathbf{Q}$ ).

Definirajmo slučajno spremenljivko

$$\Delta = \frac{1}{2} \Omega^{-1} \left[ \mathbf{B}_{XX} \frac{1}{N(T-1)} \sum (v_2^2_{w_i} - \sigma_u^2) + \mathbf{W}_{XX} \frac{1}{N} \sum (v_2^2_{B_i} - \sigma_1^2) \right] \quad (\text{D2.13})$$

S tako definirano slučajno spremenljivko lahko izrazimo relativno napako na naslednji način

$$\frac{\hat{\Omega}}{\Omega} = 1 + 2\Delta + O_p(n^{-1}). \quad (\text{D2.14})$$

V primeru običajnih cenilk je očitno relacija enaka kar  $\frac{\hat{\Omega}}{\Omega} = 1 + 2\Delta$ .

Pri predpostavki  $N = \alpha n^{1/2}$ , velja naslednja zveza med statistiko  $S$  in studentizirano statistiko  $T$

$$T = S(1 - \Delta) + O_p(n^{-1}). \quad (\text{D2.15})$$

Statistiko  $S$  lahko zapišemo

$$S = n^{-1/2} \frac{(\mathbf{B}_{XX} \mathbf{W}_{XX})^{1/2}}{(\sigma_u^2 \mathbf{B}_{XX} + \sigma_1^2 \mathbf{W}_{XX})^{1/2}} [(\mathbf{W}_{XX}^{-1} \mathbf{X}^T \mathbf{Q} - \mathbf{B}_{XX}^{-1} \mathbf{X}^T \mathbf{P}) \mathbf{v}]$$

oziroma s transformiranimi motnjami  $\mathbf{v2} = \mathbf{Rv}$  kot

$$S = n^{-1/2} \frac{(\mathbf{B}_{XX} \mathbf{W}_{XX})^{1/2}}{(\sigma_u^2 \mathbf{B}_{XX} + \sigma_1^2 \mathbf{W}_{XX})^{1/2}} [(\mathbf{W}_{XX}^{-1} \mathbf{X}^T \mathbf{Q} - \mathbf{B}_{XX}^{-1} \mathbf{X}^T \mathbf{P}) \mathbf{R}^T \mathbf{v2}]. \quad (\text{D2.16})$$

Velja opozoriti, da je v prvi relaciji  $S$  linearna forma motenj, ki so »delno« med seboj korelirane (odvisne) (za isto medsektorsko enoto  $i$ ), medtem ko je v drugi relaciji  $S$  linearna forma med seboj nekoreliranih (in neodvisnih) motenj.

Če se upošteva, da so slučajne spremenljivke  $v2_{B_i}$  in  $v2_{W_i}$  medseboj neodvisne in da je v (D2.13) pričakovana vrednost izrazov  $(v2_{W_i}^2 - \sigma_u^2)$  oziroma  $(v2_{B_i}^2 - \sigma_1^2)$  za vsak  $i$  enaka 0, potem lahko s pomočjo zapisa (D2.16) za statistiko  $S$  oziroma zapisa (D2.13) za statistiko  $\Delta$  neposredno preverimo, da je pričakovana vrednost spremenljivk  $S^2 \Delta$ ,  $S \Delta^2$ ,  $\Delta^3$  reda  $O(n^{-1})!$  Potem takem se prve tri kumulante  $S$  in  $S(1-\Delta)$  razlikujejo za člene, ki so največ reda  $O(n^{-1})!$  Če upoštevamo še (D2.15) lahko trdimo, da se prve tri kumulante statistik  $S$ , in  $T$  razlikujejo za člene, ki so največ reda  $O(n^{-1})!$ . Potemtakem so členi pri katerih se ER obeh statistik razlikuje največ reda  $O(n^{-1})!$  Torej je

$$P(T \leq \xi) = P(S \leq \xi) + O(n^{-1}). \quad (\text{D2.17})$$

Potemtakem se tudi porazdelitev statistike HT pri znanem parametru  $\psi$  razlikuje od porazdelitve statistike HT pri neznanih vendar dosledno ocenjenih parametrih motnje  $\sigma_u^2$  in  $\sigma_1^2$  za člene, ki so največ reda  $O(n^{-1})!$  Ker smo že pokazali, da so členi  $O(n^{-1/2})$  (torej v oznakah ER polinomi  $p_1(\xi)$ ) pri statistiki  $S$  enaki 0, lahko zaključimo, da se tudi vrednost porazdelitvene funkcije statistike HT in torej ustrezno pokritje razlikujeta od ciljanega za največ člene reda  $O(n^{-1})!$

Ker se pri precej splošnih pogojih odstopanje velikosti bootstrap testa od velikosti asimptotičnega testa zmanjšuje enako hitro (z velikostjo vzorca) kakor ustrezne moči pri vseh alternativnih hipotezah (glej, Davidson in MacKinnon, 2002) mora biti moč bootstrap HT testa pri alternativni hipotezi prav tako točna do členov  $O(n^{-2})$  pri modelu kjer je znano razmerje obeh spektralnih vrednosti variančno kovariančne matrike motnje  $u$  ( $\psi = \frac{\sigma_u}{(T\sigma_c^2 + \sigma_u^2)^{1/2}}$ ), oziroma do členov  $O(n^{-3/2})$  če so vsi ustrezni parametri motnje dosledno ocenjeni ! Če torej, naprimer, empirični rezultati kažejo, da je moč bootstrap testa, ki je popravljen na velikost asimptotičnega testa, enaka moči asimptotičnega testa, je torej tudi moč asimptotičnega testa točna do členov  $O(n^{-2})$  v primeru znanega razmerja obeh spektralnih vrednosti variančno kovariančne matrike motnje  $v$  ( $\psi = \frac{\sigma_u}{(T\sigma_c^2 + \sigma_u^2)^{1/2}}$ ), oziroma do členov  $O(n^{-3/2})$  v splošnem primeru!

## Literatura

1. Amemiya T.: The Estimation of the variances in a variance-components model. *International Economic Review* 12, 1971, str. 1-13.
2. Amemiya, T., MaCurdy T.E.: Instrumental-variable estimation of an error components model. *Econometrica* 54, 1986, str. 869-881.
3. Anderson T.W.: *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons Ltd, 1974. 374 str.
4. Andersson M.K.: Karlsson, S., *Bootstrapping Error Component Models*. *Computational Statistics*, vol. 16, 2001, str. 221-231.
5. Arellano Manuel: On the testing of correlated effects with panel data. *Journal of Econometrics* 59, 1993, str. 87-97.
6. Balestra P., Nerlove M.: Pooling cross-section and time-series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica* 34, 1966, str. 585-612.
7. Balestra P.: Best quadratic unbiased estimators of the variance-covariance matrix in normal regression. *Journal of Econometrics* 2, 1973, str. 17-28.
8. Baltagi B.H.: *Econometric Analysis of Panel Data*. John Wiley & Sons Ltd, 2002. 257 str.
9. Bergström Pal: *Bootstrap Methods and Applications in Econometrics – a Brief Survey*. Uppsala University, 1999, 26 str.
10. Bole Velimir: *Edgeworthov razvoj porazdelitve Hausmanove statistike*. Neobjavljeno, 2003.
11. Davidson R., MacKinnon J.G.: *Estimation and inference in econometrics*, Oxford University Press, 1993. 874 str.
12. Davidson R., MacKinnon J.G.: *Bootstrap tests: How many bootstraps?* Queens Institute of Economic Research, 2000, 14 str.
13. Davidson R., MacKinnon J.G.: *The Power of Bootstrap and Asymptotic Tests*. Queens Institute of Economic Research, 2002, 20 str.
14. Efron B., Tibshirani R.J.: *An Introduction to the Bootstrap*. Chapman & Hall, 1993. 436 str.
15. Hall Peter: *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, 1992. 352 str.
16. Hausman J.A.: Specification tests in econometrics. *Econometrica* 46, 1978, str. 1251-1270.
17. Hsiao C.: *Analysis of panel data*. Cambridge University Press, 1986. 246 str.

18. Hsiao C.: Statistical inference for a model with both random cross-sectional and time effects. *International Economic Review* 15, 1974, str. 12-30.
19. Hsiao C.: Some estimation methods for a random coefficient model. *Econometrica* 43, 1975, str. 305-325.
20. Kotz S., Johnson N.L.: *Encyclopedia of Statistical Sciences*, vol. 4. John Wiley & Sons Ltd, 1983.
21. Lepage Raoul, edt.: *Exploring the Limits of Bootstrap*. John Wiley & Sons Ltd, 1992.
22. Lillard L.A., Weiss Y.: Components of variation in panel earnings data. *Econometrica* 47, 1979, str. 437-454.
23. Maddala G.S.: The Use of Variance Components Models in Pooling Cross Section and Time Series Data. *Econometrica* 39, 1971, str. 341-358.
24. Maddala G.S.: To pool or not to pool: That is the question. *Journal of Quantitative Economics* 7, 1991, str. 255-264.
25. Mittelhammer R.C. et al.: *Econometric Foundations*. Cambridge University Press, 2000. 756 str.
26. Mundlak Y.: On the pooling of time series and cross-sectional data. *Econometrica* 46, 1978, str. 69-86.
27. Rao, C.R.: Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis*, 1971, str. 445-456.
28. Rao, C.R.: Estimation variance and covariance components in linear models. *Journal of the American Statistical Association*, 1972, str. 112-115.
29. Rice J.: *Mathematical Statistics and Data Analysis*. Duxbury Press, 1995. 602 str.
30. Swamy P.A.V.B., Arora S.S.: The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models. *Econometrica* 40, 1972, str. 261-275.
31. Wallace T.D., Hussain A.: The Use of Error Components Models in Combining Cross Section with Time Series Data. *Econometrica* 37, 1969, str. 55-72.
32. Wilks S.S.: *Mathematical Statistics*. John Wiley & Sons Ltd, 1962. 644 str.
33. Wong K.: Bootstrapping Hausman's exogeneity test. *Economic letters* 53, 1996, str. 139-143.
34. Wooldridge J.M.: *Econometric analysis of cross section and panel data*. MIT Press, 2002. 752 str.