J. Vainu

AEGRIDADE KORRELATSIOON- JA REGRESSIOONANALÜÜS

1983

TARTU RIIKLIK ÜLIKOOL
Majandusküberneetika ja statistika kateeder

J. V a i n u

AEGRIDADE KORRELATSIOON- JA REGRESSIOONANALÜÜS

Õppevahend majandusteaduskonna üliõpilastele

TARTU 1983

Retsenseerinud S. Straž, V. Venael

Kinnitatud majandusteaduskonna nõukogus
15. juunil 1983.

1. AEGRIDADE KORRELATSIOON- JA REGRESSIOONANALÜÜS

1.1. ÜLDISED KÜSIMUSED

Ühiskondlikud nähtused muutuvad ja arenevad pidevalt. Ühe nähtuse muutumisega kaasneb teis(t)e nähtus(t)e muutumine: järelikult on need nähtused mingil määral omavahel seotud. Statistilises kirjanduses esinevad tihti terminid "korrelatiivne sõltuvus" ja "korrelatiivne seos". Terminitel "seos" ja "sõltuvus" on erinev sisu ning seetõttu tuleb eristada ka "korrelatiivset seost" ja "korrelatiivset sõltuvust". Sõna "sõltuvus" viib mõttele põhjuslikkusest. Kui meil on ette teada, et ühe nähtuse (ja ainult selle) muutumine põhjustab teise nähtuse muutumise, siis võib kasutada terminit "korrelatiivne sõltuvus". Kui aga põhjuslikkus pole ilmne, siis on mõistlikum rääkida "korrelatiivsest seosest", et vältida analüüsitulemuste väära interpreteerimist.

Kõige üldisemaks seoste liigiks on stohhastiline seos, mis seisneb selles, et ühe nähtuse (x) muutumisega kaasneb teise nähtuse (y) tingliku jaotusseaduse muutumine:

$$f(y/x) = F(x) .$$

Stohhastiline sõltumatus tähendab seda, et tinglik jaotusseadus $f(y/x)$ jääb muutumatuks x väärtuste muutumise korral.

Korrelatiivne seos on stohhastilise seose üldjuhuks. Stohhastilise seose puudumine tähendab ka korrelatiivse seose puudumist, kuid vastupidine väide pole õige. Stohhastiline seos võib eksisteerida ka korrelatiivse seose puudumise korral.

Korrelatsioonanalüüsi eesmärgiks on hinnata nähtustevahelise seose tugevust. Seoste olemasolu analüüsimiseks on mitmeid meetodeid, kuid ainult korrelatsioonanalüüs annab lihtsa hinnangu nähtustevahelise seose tugevuse kohta.

Regressioonanalüüsi eesmärgiks on nähtustevahelise seose vormi määramine. See on väga vastutusrikas etapp statistilises analüüsis ning mittelineaarsete seoste korral sõltub ka korrelatsioonanalüüsi lõpptulemus regressioonanalüüsi õigsusest. Tihti statistilised kriteeriumid võimaldavad lugeda õigeks mitut erinevat seose vormi, analüüs aga annab igal konkreetsel juhul erinevad tulemused. Siin peab toimuma kompromiss uuritava materjali iseloomu ning analüüsija vahel, mis eeldab uuritavate nähtuste iseloomu põhjalikku tundmist. Liigne lihtsustamine, kuid ka liigne keerukus viivad lõppkokkuvõttes vääradele ja mõttetutele analüüsitulemustele.

Majandusnähtuste vaheliste seoste analüüsimise teeb tihti keerukaks asjaolu, et pole võimalik analüüsiga hõlmata kogu oluliste nähtuste kompleksi. Alati on vaja silmas pidada seda, et iga konkreetne regressioonivõrrand on mingil määral abstraktsioon - regressioonivõrrandi konstrueerimine on nähtustevaheliste seoste vormi selgitamise hüpoteetiliseks eksperimendiks.

Ei tohi unustada asjaolu, et statistiliseks andmestikuks on tavaliselt väljavõtukogum. Seetõttu on analüüsitulemustel tõenäosuslik iseloom ning tulemuste usaldusväärsust tuleb kontrollida spetsiaalsete kriteeriumide abil. Vastasel korral ei tohi analüüsitulemusi laiendada üldkogumile.

Nii korrelatsioon- kui ka regressioonanalüüsile peab eelnema loogilis-teoreetiline analüüs nähtuste iseloomu ning seoste võimalikkuse osas. Ainult sel juhul, kui teoreetiline analüüs võimaldab eeldada seoste olemasolu, võib rakendada korrelatsioon- ja regressioonanalüüsi ning saada tulemusi, millel on reaalne mõte.

Eelnevalt on vaja selgitada, kas seoste analüüsimiseks on tarvis kasutada just korrelatsioonanalüüsi. Võib juhtuda, et piisavaks osutub mõne lihtsama analüüsimeetodi (näiteks dispersioonanalüüsi) kasutamine. Seejuures tuleb hoolikalt jälgida, et statistiline andmestik vastaks korrelatsioonanalüüsi eeldustele.

Tuleb välja selgitada kõik tegurid, mis võivad mõjutada uuritava nähtuse muutumist, samuti seoste võimalikud struktuurid. Analüüsimudeli konstrueerimisel võib vajalike tegurite selgitamiseks kasutada dispersioonanalüüsi.

Väga tähtsal kohal on seose vormi valik. Kuigi on välja töötatud mitmed meetodid seose vormi õigsuse kontrollimiseks, on siiski vajalik professionaalne majanduslik analüüs. Tihti lubavad kriteeriumid õigeks lugeda üheaegselt mitut erinevat seose vormi. Samal ajal tegurite mõjuulatuste kvantitatiivse analüüsi tulemused sõltuvad seose vormist. Siin on vajalik uurija teadmiste, uurimistööde vajalikkuse ja uuritava materjali kompromiss. Ei tule karta mahukaid ja keerukaid arvutustöid, mitmekordseid kontrollimisi ning negatiivseid tulemusi.

Majandusnähtuste omavaheliste seoste keerukus tingib tihti olukorra, kus ei ole võimalik ühel või teisel põhjusel analüüsiga hõlmata kõiki olulisi tegureid. Seetõttu tuleb pidada silmas, et iga konkreetne regressioonivõrrand on mingil määral abstraktsioon - regressioonivõrrandi konstrueerimine on hüpoteetiliseks eksperimendiks nähtustevahelise seose konkreetse vormi selgitamisel.

Ei tohi unustada, et algandmed kujutavad endast praktiliselt alati väljavõtukogumit. Seetõttu on analüüsitulemused õiged ainult teatava tõenäosusega ning vajalik on tulemuste kontrollimine spetsiaalsete kriteeriumide abil. Vastasel korral ei tohi analüüsitulemusi laiendada üldkogumile.

1.2. TSÜKLILINE AUTOKORRELATSIOON

Tsükliline autokorrelatsioon kujutab endast seost aegridade

$y_1, \ldots, y_T$ ja $y_L, y_{L+1}, \ldots, y_T, y_1, y_2, \ldots, y_{L-1}$

vahel. Tsüklilise autokorrelatsiooni olemasolu hinnatakse üldjuhul koefitsiendiga

$$r_L = \frac{E[(y_t - Ey_t)\cdot(y_{t-L} - Ey_t)]}{\sqrt{[E(y_t - Ey_t)^2]\,[E(y_{t-L} - Ey_t)^2]}} . \qquad (1.1)$$

2

Tsüklilise autokorrelatsiooni koefitsiendi jaotused on küllalt põhjalikult läbi uuritud ja on koostatud tabelid, mis võimaldavad kontrollida hüpoteese seose olemasolu või puudumise kohta.

Tähtsaimaks peetakse juhtumit, kus $L = 1$, s. t. eeldatakse, et on tegemist esimest järku autoregressiivse protsessiga. Sel juhul

$$r_1 = \frac{\sum_{t=1}^{T} (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2} = \frac{\sum_{t=1}^{T} y_t y_{t-1} - T\bar{y}^2}{\sum_{t=1}^{T} y_t^2 - T\bar{y}^2}, \quad y_0 \equiv y_T . \qquad (1.2)$$

Koefitsiendi (1.2) jaotuse määras R. Anderson 1942. aastal eeldusel, et $y_1, \ldots, y_T$ on sõltumatud ja jaotunud normaalselt. Koefitsiendi jaotus on toodud tabelis.

T a b e l 1

Tsüklilise autokorrelatsiooni koefitsiendi jaotus[1]

T	0.01	0.05	0.95	0.99
1	2	3	4	5
5	-0,798	-0,753	0,253	0,297
6	-0,863	-0,708	0,345	0,447
7	-0,799	-0,674	0,370	0,510
8	-0,764	-0,625	0,371	0,531
9	-0,737	-0,593	0,366	0,533
10	-0,705	-0,564	0,360	0,525
11	-0,679	-0,539	0,353	0,515
12	-0,655	-0,516	0,348	0,505
13	-0,634	-0,497	0,341	0,495
14	-0,615	-0,479	0,335	0,485
15	-0,597	-0,462	0,328	0,475
20	-0,524	-0,399	0,299	0,432
25	-0,473	-0,356	0,276	0,398
30	-0,433	-0,324	0,257	0,370
(35)	-0,401	-0,300	0,242	0,347
(40)	-0,376	-0,279	0,229	0,329

T a b e l 1 (järg)

1	2	3	4	5
45	-0,356	-0,262	0,218	0,313
(50)	-0,339	-0,248	0,208	0,301
(55)	-0,324	-0,236	0,199	0,289
(60)	-0,310	-0,225	0,191	0,278
(65)	-0,298	-0,216	0,184	0,268
(70)	-0,287	-0,207	0,178	0,259
75	-0,276	-0,201	0,174	0,250

[1] Ümarsulgudes olevatele T-dele vastavad väärtused on saadud graafilisel interpoleerimisel. Tabel on võetud raamatust: Т. Андерсон. Статистический анализ временных рядов. М.: Мир, 1976, с. 354.

Kuivõrd trendist võetud hälvete summa on võrdne nulliga, siis valemi (1.1) võib esitada kujul

$$r_L = \frac{\sum_{t=L+1}^{T} Y_t Y_{t-L}}{\sum_{t=1}^{T} Y_t^2} . \qquad (1.3)$$

Kui L = 1, siis

$$r_1 = \frac{\sum_{t=1}^{T} Y_t Y_{t-1}}{\sum_{t=1}^{T} Y_t^2} , \quad Y_0 \equiv Y_T . \qquad (1.4)$$

Kui on teada uuritava nähtuse üldkeskmine (mitte segi ajada väljavõtukeskmisega, mis on üldkogumi hinnanguks), siis võib tsüklilise autokorrelatsiooni kontrollimiseks kasutada valemit

$$r_1 = \frac{\sum_{t=2}^{T} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y}) + \frac{1}{2}(Y_1 - \bar{Y})^2 + \frac{1}{2}(Y_T - \bar{Y})^2}{\sum_{t=1}^{T} (Y_t - \bar{Y})^2} . \qquad (1.5)$$

Antud juhul võib r_1 kontrollimiseks kasutada K. Pearsoni korrelatsioonikoefitsiendi r jaotust, kusjuures $N = T + 2$.

On võimalik, et aegrida sisaldab perioodilist trendi. Juhul, kui T on paaritu, siis

$$\hat{Y}_t = a_0 + \sum_{k=1}^{q} \left(a(\gamma_k) \cos \frac{2\pi\gamma_k}{T} t + b(\gamma_k) \sin \frac{2\pi\gamma_k}{T} t \right), \tag{1.6}$$

$\gamma_1, \ldots, \gamma_q$ on täisarvude hulga $1, \ldots, (T-1)/2$ alamhulgaks. Kui T on paarisarv, siis

$$\hat{Y}_t = a_0 + \sum_{k=1}^{q} \left(a(\gamma_k) \cos \frac{2\pi\gamma_k}{T} t + b(\gamma_k) \sin \frac{2\pi\gamma_k}{T} t \right) + a_{T/2}(-1)^t. \tag{1.7}$$

$$a(\gamma_k) = \frac{2}{T} \sum_{t=1}^{T} Y_t \cos \frac{2\pi\gamma_k}{T} t, \tag{1.8}$$

$$b(\gamma_k) = \frac{2}{T} \sum_{t=1}^{T} Y_t \sin \frac{2\pi\gamma_k}{T} t, \quad \gamma_k \neq 0, T/2, \tag{1.9}$$

$a_0 = \bar{Y}$.

ja kui T on paarisarv, siis

$$a_{T/2} = \frac{1}{T} \sum_{t=1}^{T} Y_t (-1)^t. \tag{1.10}$$

Tsüklilise autokorrelatsiooni koefitsiendid on järgmised:

$$r_j = \frac{\sum_{t=1}^{T}{}' Y_t Y_{t-j} - T\bar{Y}^2 - \frac{1}{2} T \sum_{k=1}^{q} \cos \frac{2\pi\gamma_k j}{T} \left[a^2(\gamma_k) + b^2(\gamma_k) \right]}{\sum_{t=1}^{T} Y_t^2 - T\bar{Y}^2 - \frac{1}{2} T \sum_{k=1}^{q} \left[a^2(\gamma_k) + b^2(\gamma_k) \right]} \tag{1.11}$$

ja kui T on paarisarv, siis

$$r_j = \frac{\sum_{t=1}^{T} Y_t Y_{t-j} - T\left[\bar{Y}^2 + (-1)^j a^2_{T/2} \right] - \frac{1}{2} T \sum_{k=1}^{q} \left[a^2(\gamma_k) + b^2(\gamma_k) \right]}{\sum_{t=1}^{T} Y_t^2 - T\left[\bar{Y}^2 + a^2_{T/2} \right] - \frac{1}{2} T \sum_{k=1}^{q} \left[a^2(\gamma_k) + b^2(\gamma_k) \right]} \tag{1.12}$$

1.3. MITTETSÜKLILINE AUTOKORRELATSIOON

Mittetsükliline autokorrelatsioon on seos ridade $y_1, \dots, y_{T-L}$ ja $y_{L+1}, \dots, y_T$ vahel. Üldjuhul võib mittetsüklilise autokorrelatsiooni olemasolu kontrollida järgmise koefitsiendiga:

$$R_L = \frac{\sum_{t=1}^{T-L} y_t y_{t+L} - \frac{\sum_{t=1}^{T-L} y_t \sum_{t=L+1}^{T} y_t}{T-L}}{\sqrt{\left[\sum_{t=1}^{T-L} y_t^2 - \frac{\left(\sum_{t=1}^{T-L} y_t\right)^2}{T-L}\right]\left[\sum_{t=L+1}^{T} y_t^2 - \frac{\left(\sum_{t=L+1}^{T} y_t\right)^2}{T-L}\right]}} \qquad (1.13)$$

Juhul, kui $\sum y_t = 0$ (algandmeteks on hälbed trendist, keskmisest jne.), siis võib valemit (1.13) lihtsustada:

$$R_L = \frac{\frac{1}{T-L} \sum_{t=1}^{T-L} y_t y_{t+L}}{\sqrt{\frac{\sum_{t=1}^{T-L} y_t^2 \sum_{t=L+1}^{T} y_t^2}{(T-L)^2}}} . \qquad (1.14)$$

Tuleb aga märkida, et selline lihtsustamine on võimalik ainult sel juhul, kui $T \geqslant 30$, s. t. rida pole lühike.

Mittetsüklilise autokorrelatsiooni jaotus on teadmata.

1.4. DURBINI-WATSONI KRITEERIUM

Autokorrelatsiooni olemasolu võib kontrollida ka Durbini-Watsoni kriteeriumi abil:

$$d = \frac{\sum_{t=1}^{T-1} (y_{t+1} - y_t)^2}{\sum_{t=1}^{T} y_t^2} . \qquad (1.15)$$

Kriteeriumi võimalikud väärtused asuvad intervallis 0 - 4. Autokorrelatsiooni puudumisel kriteeriumi väärtused kõiguvad 2 ümber. Kriteeriumi empiirilist väärtust kontrollitakse selle teoreetilise väärtusega vastavalt L-le.

Kui: 1) $d < d_L$ - reas on autokorrelatsioon;

2) $d > d_U$ - autokorrelatsioon puudub;

3) $d_L < d < d_U$ - probleem jääb lahtiseks.

d_L ja d_U on vastavalt kriteeriumi alumine ja ülemine piir.

Negatiivse autokorrelatsiooni korral asuvad d väärtused vahemikus 2 - 4. Sel juhul tuleb kontrollimiseks võtta suurus $d' = 4 - d$.

1.5. HARMOONILINE AUTOKORRELATSIOON

Eeltoodud autokorrelatsiooni uurimise meetodid eeldasid, et seos aegrea tasemete vahel on lineaarne. Tegelikkuses see nii ei pea olema. On täiesti võimalik, et analüüsitavas aegreas peitub mingi harmooniline võnkumine või harmoonikute summa (Fourier' rida). See aga tähendab, et aegreas on mingi perioodiga autokorrelatsioon. Harmoonilise autokorrelatsiooni uurimine on küllaltki keeruline, kuid võimaldab anda tähtsat informatsiooni prognoosimudelite konstrueerimiseks. Demonstreerime harmoonilise autokorrelatsiooni uurimist kahe lihtsa näite abil.

Sinusoidaalne autokorrelatsioon. Oletame, et aegrida esitub võrrandiga

$$y_1 = \sin 2\pi \frac{t}{T}, \quad t = \ldots, -1, 0, 1, \ldots, \tag{1.16}$$

kus T on võnkeperiood.

Kasutades faasinihet τ, saab reast (1.16) moodustada teise aegrea

$$y_2 = \sin 2\pi \frac{t+\tau}{T}.$$

Võnkeamplituudi võtame võrdseks ühega, sest antud juhul sellel pole põhimõttelist tähtsust. Autokorrelatsiooni uurimiseks võtame aluseks klassikalise valemi

$$r = \frac{\overline{y_1 y_2} - \bar{y}_1 \bar{y}_2}{\sqrt{\sigma_{y_1}^2 \cdot \sigma_{y_2}^2}}. \tag{1.17}$$

Olgu tarvis korreleerida y_1 ja y_2 väärtusi intervallis $u \pm h$. Vajalikud keskmised ja dispersioonid arvutame teise harmooniku põhjal, võttes hiljem $\tau = 0$:

$$\Sigma Y_2 = \int_{u-h}^{u+h} \sin 2\pi \frac{t+\tau}{T} dt = \frac{T}{\pi} \sin 2\pi \frac{u+\tau}{h} \sin 2\pi \frac{h}{T} . \qquad (1.18)$$

$$\bar{Y}_2 = \frac{T}{2\pi h} \sin 2\pi \frac{u+\tau}{T} \sin 2\pi \frac{h}{T} . \qquad (1.19)$$

$$\Sigma Y_2^2 = \int_{u-h}^{u+h} \sin^2 2\pi \frac{t+\tau}{T} dt = h - \frac{T}{4\pi} \cos 4\pi \frac{u+\tau}{T} \sin 4\pi \frac{h}{T} . \qquad (1.20)$$

$$\sigma_2^2 = \frac{1}{2} - \frac{T}{8\pi h} \cos 4\pi \frac{u+\tau}{T} \sin 4\pi \frac{h}{T} - \left[\frac{T}{2\pi h} \sin 2\pi \frac{u+\tau}{T} \sin 2\pi \frac{h}{T}\right]^2 . \qquad (1.21)$$

Võttes valemites (1.18) ja (1.20) $\tau = 0$, saame:

$$\bar{Y}_1 = \frac{T}{2\pi h} \sin 2\pi \frac{u}{T} \sin 2\pi \frac{h}{T} , \qquad (1.22)$$

$$\sigma_1^2 = \frac{1}{2} - \frac{T}{8\pi h} \cos 4\pi \frac{u}{T} \sin 4\pi \frac{h}{T} -$$

$$- \left[\frac{T}{2\pi h} \sin 2\pi \frac{u}{T} \sin 2\pi \frac{h}{T}\right]^2 . \qquad (1.23)$$

Nüüd leiame korrutiste keskmise:

$$\overline{Y_1 Y_2} = \frac{1}{2h} \int_{u-h}^{u+h} \sin 2\pi \frac{t}{T} \sin 2\pi \frac{t+\tau}{T} dt =$$

$$= \frac{1}{2} \cos 2\pi \frac{\tau}{T} - \frac{T}{8\pi h} \cos 2\pi \frac{2u+\tau}{T} \sin 4\pi \frac{h}{T} . \qquad (1.24)$$

Seega on leitud valemis (1.17) vajalikud suurused.

<u>Autokorrelatsioon Fourier' reas</u>. Oletame, et statistiline andmestik on keskmisega null ja esitub Fourier' reana:

$$Y_1 = \sum A_k \sin 2\pi \frac{t+\varphi_k}{kT} , \qquad k = 1, \frac{1}{2}, \frac{1}{3}, \ldots . \qquad (1.25)$$

Dispersiooniks on siin

$$\sigma_1^2 = \frac{1}{2} \sum A_k^2 . \qquad (1.26)$$

Kasutades faasinihet τ, moodustame teise rea

$$Y_2 = \sum A_k \sin 2\pi \frac{t + \varphi_k + \tau}{kT}, \tag{1.27}$$

mille dispersioon on samuti

$$\sigma_2^2 = \frac{1}{2} \sum A_k^2 .$$

Korrutiste keskmiseks saame

$$\overline{Y_1 Y_2} = \frac{1}{2} \sum A_k^2 \cos 2\pi \frac{\tau}{kT} \tag{1.28}$$

ja autokorrelatsiooni koefitsient arvutatakse valemiga

$$r_\tau = \frac{\frac{1}{2} \sum A_k^2 \cos 2\pi \frac{\tau}{kT}}{\frac{1}{2} \sum A_k^2} = \frac{\sum A_k^2 \cos 2\pi \frac{\tau}{kT}}{\sum A_k^2} . \tag{1.29}$$

Igakuulistest andmetest koosnevates majanduslikes aegridades võib tihti sisalduda sinusoid või koosinusoid juba sesoonse komponendi tõttu. Seevastu ridades, mis koosnevad iga-aastastest andmetest, esineb üksainus harmooniline komponent harva. Tavaliselt on mõistlikum seal uurida harmoonikute esitumist Fourier' reana.

1.6. VIITKORRELATSIOON

Aegridade korreleerimisel tuleb silmas pidada, et üks nähtus võib teist mõjutada teatava hilinemisega, s. t. möödub teatud aeg, enne kui ühe nähtuse muutumine toob kaasa teise nähtuse muutumise. Näiteks võib tuua põhifondide ja toodangu mahu vahelise seose. Kui tehases lastakse käiku uued täiendavad seadmed ja tööpingid, siis need ei saavuta oma täisvõimsust kohe, vaid möödub teatud aeg seadmete häälestamiseks, uute töövõtete omandamiseks jne. Alles mingi ajavahemiku möödumisel hakkavad seadmed tööle täie võimsusega, mis lõppkokkuvõttes avaldub toodangu mahu suurenemises.

Sellise ajalise nihke (viitaja) olemasolu selgitamiseks tuleb põhjalikult analüüsida uuritavate nähtuste olemust. Mõnikord on viitaja olemasolu selge ilma pikema arutluseta, vahel see loogilistest arutlustest täpselt ei selgu. Samuti pole võimalik intuitiivselt määrata viitaja pikkust (perioodide arvu).

Viitaja olemasolu ja pikkuse uurimine seisneb tavaliselt viitkorrelatsioonifunktsiooni uurimises, mille moodustavad viitkorrelatsiooni koefitsiendid:

$$r_L = \frac{(T-L)\sum_{t=1}^{T-L} x_t y_{t+L} - \sum_{t=1}^{T-L} x_t \sum_{t=1}^{T-L} y_{t+L}}{\sqrt{\left[(T-L)\sum_{t=1}^{T-L} y_{t+L}^2 - \left(\sum_{t=1}^{T-L} y_{t+L}\right)^2\right]\left[(T-L)\sum_{t=1}^{T-L} x_t^2 - \left(\sum_{t=1}^{T-L} x_t\right)^2\right]}} \qquad (1.30)$$

$L = 1, 2, \ldots$

Juhul, kui algandmeteks on jääkliikmed, siis valem (1.30) lihtsustub:

$$r_L = \frac{\sum_{t=1}^{T-L} x_t y_{t+L}}{\sqrt{\sum_{t=1}^{T-L} x_t^2 \sum_{t=1}^{T-L} y_{t+L}^2}}. \qquad (1.31)$$

Viitaja pikkus (max L) viitkorrelatsioonifunktsiooni arvutamisel sõltub uuritavate nähtuste spetsiifikast. Majanduslikes uurimustes on harva max L > 10. Viitkorrelatsioonifunktsiooni graafik on tavaliselt kiirelt sumbuva võnkumise graafik. Kui graafikul esinevad tipud, kus r(L) väärtus on oluliselt suurem r(L-1) ja r(L+1) väärtustest, siis väidetakse, et uuritavad nähtused on seotud ajalise hilinemisega, mille pikkus on L perioodi. Kui sellised tipud esinevad iga k L perioodi järele, siis on põhjust eeldada perioodilise komponendi olemasolu uuritavates ridades.

Märgime veel, et viitaja olemasolu võimalikkust tuleb kontrollida ka siis, kui teoreetiline analüüs otseselt ei anna põhjust selle olemasolu eeldada. Tihtipeale võivad analüüsitulemused olla uurijale ootamatud ning pole mõtet viitaja olemasolu võimalikkust ignoreerida.

Kõige lihtsamaks juhtumiks seoste analüüsimisel on seos kahe nähtuse vahel. Seose tugevuse hindamiseks kasutatakse üldjuhul korrelatsiooniindeksit

$$R = \sqrt{\frac{\sigma_y^2 - \sigma_{\hat{y}}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sigma_{\hat{y}}^2}{\sigma_y^2}}, \qquad (2.1)$$

kus σ_y^2 - resultaatnähtuse dispersioon;

$\sigma_{\hat{y}}^2$ - jääkdispersioon, mis määratakse resultaatnähtuse tegelike väärtuste ja regressioonijoone punktide väärtuste vaheliste hälvete ruutude keskmisena:

$$\sigma_{\hat{y}}^2 = \frac{\sum (y_t - \hat{y}_t)^2}{T}, \quad t = 1, \ldots, T. \qquad (2.2)$$

Et tavaliselt ei ole võimalik kindlaks määrata, millise märgiga ruutjuur võtta, kasutatakse analüüsis korrelatsiooniindeksi absoluutväärtusi, mis asuvad intervallis $0 \leq |R| \leq 1$.

Kui $R = 1$, siis on tegemist funktsionaalse seosega; kui $R = 0$, siis uuritavad nähtused pole omavahel seotud.

Kui on kindlalt teada, et kahe nähtuse vahelise seose vorm on lineaarne, s. t. esitatav sirgjoone võrrandiga, siis võib valemi (2.1) teisendada lihtsamale kujule:

$$r = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\sigma_x \cdot \sigma_y}. \qquad (2.3)$$

Suurust r nimetatakse lineaarseks korrelatsioonikoefitsiendiks (-kordajaks) ning selle väärtused asuvad intervallis $-1 \leq r \leq 1$.

Kui korrelatsioonikoefitsient on positiivne, siis uuritavate nähtuste vahel valitseb võrdeline seos (nähtused muutuvad samas suunas); negatiivne korrelatsioonikoefitsient tähendab pöördvõrdelist seost.

Praktilisteks arvutusteks on sobivamad valemi (2.3) teisendused

$$r = \frac{T\sum xy - \sum x \sum y}{\sqrt{[T\sum x^2 - (\sum x)^2][T\sum y^2 - (\sum y)^2]}} \qquad (2.4)$$

ja

$$r = \frac{\sum xy - \sum x \sum y / T}{\sqrt{[\sum x^2 - (\sum x)^2/T][\sum y^2 - (\sum y)^2/T]}} \qquad (2.5)$$

Mônikord kasutatakse kôrvuti korrelatsiooniindeksiga ka determinatsioonikoefitsienti

$$d = R^2, \qquad (2.6)$$

mis näitab, kui suur osa resultaatnähtuse dispersioonist on pôhjustatud uuritava teguri varieerumisest. Sellesse näitajasse tuleb suhtuda ettevaatusega, sest majandusnähtuste analüüsimisel pole kunagi küllaldast alust väita, et just uuritav tegur on ainuke, mis pôhjustab resultaatnähtuse varieerumist.

Seose vormi määramisel kasutatakse järgmisi pôhilisi funktsioone:

sirge $\hat{y} = a_0 + a_1x$, (2.7)

parabool $\hat{y} = a_0 + a_1x + a_2x^2$, (2.8)

hüperbool $\hat{y} = a_0 + a_1\frac{1}{x}$, (2.9)

eksponentfunktsioon $\hat{y} = a_0a_1^x$, (2.10)

astmefunktsioon $\hat{y} = a_0x^{a_1}$, (2.11)

poollogaritmiline funktsioon $\hat{y} = a_0 + a_1\ln x$, (2.12)

Funktsioonide parameetrid leitakse tavaliselt vähimruutude meetodil.

2.1. LINEAARSE TRENDIGA AEGRIDADE KORRELEERIMINE

Aegridade korrelatsioon- ja regressioonanalüüs esitab statistilisele andmestikule terve rea nôudeid: aegread ei tohi sisaldada trendi, autokorrelatsiooni, peavad olema sôltumatud jm. Tavaliselt trendi väärtused lahutatakse aegrea tegelikust väärtusest ning analüüsimiseks kasutatakse hälbeid trendist. Sellisel juhul jääb uurijal saamata osa informatsiooni, mis mõnel juhul võiks analüüsi kvaliteeti tõsta.

Aegridade korrelatsioon võib toimuda ka nõnda, et trendi aegreast ei eemaldata. Sisuliselt kujutab aegridade korrelatsioonanalüüs endast aegridade paralleelse varieerumise astme määramist. Seejuures on kasutusel kaks põhilist näitajat - kovariatsioon ja dispersioon. Esimene neist iseloomustab kahe aegrea paralleelset variatsiooni, teine aga ühe aegrea individuaalset variatsiooni.

Tähistame: x, y - uuritavad aegread;

X, Y - jääkliikmetest koosnevad aegread (näiteks $X_t = x_t - a_0 - a_1 t$) ;

t - aeg;

C - kovariatsioon;

V - variatsioon;

a_j, b_j - trendide parameetrid;

α_j - regressioonijoone parameetrid.

Eeldame, et $\sum t = 0$.

$$C(x,y) = \overline{xy} - \bar{x}\bar{y}, \qquad (2.13)$$

$$V(t) = \overline{t^2} - \bar{t}^2 = \overline{t^2}. \qquad (2.14)$$

Oletame, et kaks aegrida x, y sisaldavad mõlemad lineaarseid trende. Seose tugevuse hindamiseks peab siis kasutama valemit

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} = \frac{\overline{XY}}{\sqrt{\overline{X^2}\cdot\overline{Y^2}}}. \qquad (2.15)$$

Kui aegridadest trendi mitte eemaldada, siis tuleb valem (2.3) teisendada kujule, kus lugejas ja nimetajas olevatest näitajatest oleksid eemaldatud trendiga seotud kovariatsioonid ja variatsioonid.

$$\overline{XY} = \frac{1}{T}\sum (x - a_0 - a_1 t)(y - b_0 - b_1 t). \qquad (2.16)$$

Kuna aegridade silumise teooriast on teada, et

$a_0 = \bar{x}$, $b_0 = \bar{y}$,

$$a_1 = \frac{\sum tx}{\sum t^2} = \frac{\frac{1}{T}\sum tx}{\frac{1}{T}\sum t^2} = \frac{C(t,x)}{V(t)}, \qquad (2.17)$$

$$b_1 = \frac{\Sigma ty}{\Sigma t^2} = \frac{\frac{1}{T}\Sigma ty}{\frac{1}{T}\Sigma t^2} = \frac{C(t,y)}{V(t)}, \tag{2.18}$$

siis (2.16) saab kuju

$$\overline{\mathfrak{x}\mathfrak{y}} = \frac{1}{T}\Sigma(x-\bar{x}-a_1t)(y-\bar{y}-b_1t) =$$

$$= \frac{1}{T}\Sigma(xy - x\bar{y} - b_1tx - y\bar{x} - \bar{x}\bar{y} - b_1t\bar{x} - a_1ty +$$

$$+ a_1t\bar{y} + a_1b_1t^2) =$$

$$= C(x,y) - \frac{C(t,x)C(t,y)}{V(t)}. \tag{2.19}$$

Siin $\frac{C(t,x)C(t,y)}{V(t)}$ mõõdab lineaarsetest tendentsidest tingitud kovariatsiooni, mis on uuritavate ridade kovariatsiooni üheks osaks.

$$\sigma_{\mathfrak{x}}^2 = \frac{1}{T}\Sigma(x-\bar{x}-a_1t)^2 = \frac{1}{T}\Sigma(x^2 - 2\bar{x}x + \bar{x}^2 -$$

$$- 2a_1tx + 2a_1t\bar{x} + a_1^2t^2 =$$

$$= V(x) - \frac{C^2(t,x)}{V(t)}; \tag{2.20}$$

$\frac{C^2(t,x)}{V(t)}$ iseloomustab seda osa aegrea x variatsioonist, mis on omane tendentsile. Analoogiliselt saame

$$\sigma_{\mathfrak{y}}^2 = V(y) - \frac{C^2(t,y)}{V(t)}. \tag{2.21}$$

Valem (2.15) saab nüüd kuju

$$r = \frac{C(x,y) - \frac{C(t,x)C(t,y)}{V(t)}}{\sqrt{\left[V(x) - \frac{C^2(t,x)}{V(t)}\right]\left[V(y) - \frac{C^2(t,y)}{V(t)}\right]}}. \tag{2.22}$$

Lineaarse regressiooni koefitsient α_1 leitakse vähimruutude meetodil saadud võrrandisüsteemist

$$T\alpha_0 + \alpha_1 \sum (x - a_0 - a_1 t) = \sum (y - b_0 - b_1 t),$$
$$\alpha_0 \sum (x - a_0 - a_1 t) + \alpha_1 \sum (x - a_0 - a_1 t)^2 =$$
$$= \sum (x - a_0 - a_1 t)(y - b_0 - b_1 t) . \qquad (2.23)$$

Siit $\alpha_0 = 0$ ja

$$\alpha_1 = \frac{C(x,y) - \frac{C(t,x)C(t,y)}{V(t)}}{V(x) - \frac{C^2(t,x)}{V(t)}} . \qquad (2.24)$$

Lineaarse regressiooni jääkdispersiooni leiame valemiga

$$\sigma^2_{\hat{y}} = \frac{1}{T} \sum \left[y - \bar{y} - b_1 t - \alpha_1 (x - \bar{x} - a_1 t) \right]^2 =$$

$$= V(y) - \frac{C^2(t,y)}{V(t)} - \frac{\left[C(x,y) - \frac{C(t,x)C(t,y)}{V(t)} \right]^2}{V(x) - \frac{C^2(t,x)}{V(t)}} \qquad (2.25)$$

Näide 1. Olgu teada tehase kogutoodangu maht (tuh. rbl.) ja töötajate arv aastail 1969 - 1976. Silume mõlemad read sirgega ning leiame variatsioonid ja kovariatsioonid. Arvutuste tulemused on esitatud tabelis 2.

Vastavalt (2.13)

$$C(x,y) = \overline{xy} - \bar{x}\bar{y} = \frac{29644039}{8} - \frac{34816 \cdot 6855}{64} =$$

$$= 3705504,8 - 3729120,0 = -23615,2 .$$

(2.14) annab

$$V(t) = \overline{t^2} = \frac{168}{8} = 21 .$$

Edasi, $C(t,x) = \overline{tx} = \frac{-1207}{8} = -150,875 ;$

$$C(t,y) = \overline{ty} = \frac{26518}{8} = 3314,75 ;$$

$$V(y) = \overline{y^2} - \bar{y}^2 = \frac{155751116}{8} - 18939904 = 528985,5 ;$$

$$V(x) = \overline{x^2} - \bar{x}^2 = \frac{5885089}{8} - 734234,76 = 1401,36 .$$

Tabel 2

Lineaarse trendiga aegridade korreleerimine

Aasta	Toodang y	Töötajad x	t	t_y	t_x	xy	t^2	y^2	x^2
1969	3341	909	-7	-23387	-6363	3036969	49	11162281	826281
1970	3526	893	-5	-17630	-4465	3148718	25	12432676	797449
1971	3799	881	-3	-11397	-2643	3346919	9	14432461	776161
1972	4136	866	-1	-4136	- 866	3581776	1	17106496	749956
1973	4601	866	1	4601	866	3984466	1	21169201	749956
1974	4838	791	3	14514	2373	3826858	9	23406244	625681
1975	5036	826	5	25180	4130	4159736	25	25361296	682276
1976	5539	823	7	38773	5761	4558597	49	30680521	677329
	34816	6855	0	26518	-1207	29644039	168	155751116	5885089

Nüüd leiame lineaarse korrelatsioonikoefitsiendi valemiga (2.22):

$$r = \frac{-23615,2 - \dfrac{-150,875 \cdot 3314,75}{21}}{\sqrt{\left[1401,36 - \dfrac{22763,265}{21}\right]\left[528985,5 - \dfrac{10987566,5625}{21}\right]}} =$$

$$= \frac{199,7}{\sqrt{317,395 \cdot 5755,1875}} = \frac{199,7}{\sqrt{1826700,4}}$$

$$= \frac{199,7}{1351,5548} = 0,1478 .$$

Esimesel pilgul tulemus erineb oodatust. Kuivõrd uuritaval perioodil kogutoodangu maht on suurenenud ja töötajate arv vähenenud, siis võiks oodata negatiivset seost uuritavate nähtuste vahel (saaksimegi r = -0,8674). Peale seda aga, kui aegridadest on eemaldatud lineaarne trend, näeme, et tegelik seos on siiski positiivne, kuid väga nõrk. Järelikult uuritavas ettevõttes kogutoodangu maht pole töötajate arvuga märkimisväärselt seotud.

2.2. LINEAARNE JA PARABOOLNE TREND

Oletame, et resultaatnähtuse aegreas on lineaarne trend ja faktornähtuse aegreas - paraboolne. Jääkliikmete regressioon peaks olema lineaarne, sest trendi eemaldamist aegreast võib samastada mastaabi muutumisega, mistõttu uue aegrea (jääkliikmetest koosneva) tasemed varieeruvad abstsisstelje ümber.

Määrame paraboolse trendi parameetrid. Vähimruutude meetodil saame võrrandisüsteemi

$$\begin{cases} T a_0 + a_2 \Sigma t^2 = \Sigma x, \\ a_0 \Sigma t^2 + a_2 \Sigma t^4 = \Sigma t^2 x, \\ a_1 \Sigma t^2 \qquad\qquad = \Sigma t x . \end{cases}$$

Siit

$$a_0 = \frac{\bar{x}\,\overline{t^4} - \overline{t^2}\,\overline{t^2 x}}{\overline{t^4} - \overline{t^2}^2} = \bar{x} - V(t)\,\frac{C(t^2, x)}{V(t^2)}; \qquad (2.26)$$

$$a_1 = \frac{\overline{t x}}{\overline{t^2}} = \frac{C(t, x)}{V(t)}; \qquad (2.27)$$

$$a_2 = \frac{\overline{t^2 y} - \overline{t^2}\,\bar{y}}{\overline{t^4} - \overline{t^2}^2} = \frac{C(t^2, x)}{V(t^2)} . \qquad (2.28)$$

Edasi,

$$\overline{\mathfrak{x}\mathfrak{y}} = \frac{1}{T}\Sigma(x - a_0 - a_1 t - a_2 t^2)(y - b_0 - b_1 t) =$$

$$= \frac{1}{T}\Sigma\,(xy - a_0 y - a_1 t y - a_2 t^2 y - b_0 x + a_0 b_0 + a_1 b_0 t$$

$$+ a_2 b_0 t^2 - b_1 t x + a_0 b_1 t + a_1 b_1 t^2 + a_2 b_1 t^3) =$$

$$= \overline{xy} - \bar{x}\bar{y} + \bar{y}\,\overline{t^2}\,\frac{C(t^2, x)}{V(t^2)} - \frac{C(t,x)\,C(t,y)}{V(t)} -$$

$$- \overline{t^2 y}\,\frac{C(t^2, x)}{V(t^2)} - \bar{x}\bar{y} - \bar{x}\bar{y} - \bar{y}\,\overline{t^2} \cdot \frac{C(t^2, x)}{V(t^2)} +$$

$$+ \bar{y}\,\overline{t^2}\,\frac{C(t^2, x)}{V(t^2)} - \frac{C(t,x)\,C(t,y)}{V(t)} + \frac{C(t,x)\,C(t,y)}{V^2(t)} \cdot V(t) =$$

$$= C(x, y) - \frac{C(t,x)\,C(t,y)}{V(t)} - \frac{C(t^2,x)\,C(t^2,y)}{V(t^2)} .$$

(2.29)

$$\sigma_x^2 = V(x) - \frac{C^2(t,x)}{V(t)} - \frac{C^2(t^2,x)}{V(t^2)} \qquad (2.30)$$

Nüüd saame määrata korrelatsioonikoefitsiendi:

$$r = \frac{C(x,y) - \frac{C(t,x)\,C(t,y)}{V(t)} \quad \frac{C(t^2,x)\,C(t^2,y)}{V(t^2)}}{\sqrt{\left[V(x) - \frac{C^2(t,x)}{V(t)} - \frac{C^2(t^2,x)}{V(t^2)}\right]\left[V(y) - \frac{C^2(t,y)}{V(t)}\right]}} \qquad (2.31)$$

Regressioonikoefitsiendi saame valemiga

$$\alpha_1 = \frac{C(t,y) - \frac{C(t,x)C(t,y)}{V(t)} - \frac{C(t^2,x)C(t^2,y)}{V(t^2)}}{V(x) - \frac{C(t^2,x)}{V(t)} - \frac{C^2(t^2,x)}{V(t^2)}} \qquad (2.32)$$

Jääkdispersioon määratakse valemiga

$$\sigma_y^2 = \frac{V(x) - \frac{C^2(t,x)}{V(t)} - C(x,y) - \frac{C(t,x)C(t,y)}{V(t)} - \frac{C(t^2,x)C(t^2,y)}{V(t^2)}}{V(x) - \frac{C(t^2,x)}{V(t)} - \frac{C^2(t^2,x)}{V(t^2)}} \qquad (2.33)$$

Oletame nüüd, et paraboolset trendi omab resultaatnähtus, faktornähtuse trend on aga lineaarne.

$$\overline{xy} = \frac{1}{T}\sum (y - b_0 - b_1 t - b_2 t^2)(x - a_0 - a_1 t) =$$

$$= \frac{1}{T}\sum (xy - a_0 y - a_1 t y - b_0 x + a_0 b_0 + a_1 b_0 t + a_0 b_1 t -$$

$$- b_1 t x + a_1 b_1 t^2 - b_2 t^2 x + a_0 b_2 t^2 + a_1 b_2 t^3 =$$

$$= \overline{xy} - a_0 \bar{y} - a_1 \overline{ty} - b_0 \bar{x} + a_0 b_0 - b_1 \overline{tx} + a_1 b_1 \overline{t^2} - b_2 \overline{t^2 x} + a_0 b_2 \overline{t^2} =$$

$$= C(x,y) - \frac{C(t,x)\,C(t,y)}{V(t)} - \frac{C(t^2,x)\,C(t^2,y)}{V(t^2)} \qquad (2.34)$$

Valem (2.34) ühtib täielikult valemiga (2.29). Järelikult, sõltumatult sellest, millise nähtuse trend on mittelineaarne, trendi mõju avaldub ühtemoodi:

$$r = \frac{e(x,y) - \frac{e(t,x)e(t,y)}{V(t)} - \frac{e(t^2,x)e(t^2,y)}{V(t^2)}}{\sqrt{\left[V(x) - \frac{e^2(t,x)}{V(t)}\right]\left[V(y) - \frac{e^2(t,y)}{V(t)} - \frac{e^2(t^2,y)}{V(t^2)}\right]}} \quad (2.35)$$

$$\alpha_1 = \frac{e(x,y) - \frac{e(t,x)e(t,y)}{V(t)} - \frac{e(t^2,x)e(t^2,y)}{V(t^2)}}{V(x) - \frac{e^2(t,x)}{V(t)}} \quad (2.36)$$

$$\sigma_{\Delta}^2 = \frac{1}{T}\sum\left[y - b_0 - b_1 t - b_2 t^2 - \alpha_1(x - a_0 - a_1 t)\right]^2 =$$

$$= V(y) - \frac{e^2(t,y)}{V(t)} - \frac{e^2(t^2,y)}{V(t^2)} -$$

$$- \frac{\left[e(x,y) - \frac{e(t,x)e(t,y)}{V(t)} - \frac{e(t^2,x)e(t^2,y)}{V(t^2)}\right]^2}{V(x) - \frac{e^2(t,x)}{V(t)} - \frac{e^2(t^2,x)}{V(t^2)}} \quad (2.37)$$

2.3. PARABOOLSED TRENDID

Vaatleme lõpuks juhtumit, kus mõlemad aegread sisaldavad paraboolset trendi.

$$\overline{\mathfrak{x}\mathfrak{y}} = \frac{1}{T}\sum(x - a_0 - a_1 t - a_2 t^2)(y - b_0 - b_1 t - b_2 t^2) =$$

$$= e(x,y) - \frac{e(t,x)e(t,y)}{V(t)} - \frac{e(t^2,x)e(t^2,y)}{V(t^2)} \quad (2.38)$$

Valem (2.38) ühtib valemitega (2.34) ja (2.29). Seetõttu võime väita, et paraboolne (mittelineaarne) trend põhjustab alati ühesuguse aegridade kovariatsiooni struktuuri. Korrelatsioonikoefitsient

$$r = \frac{e(x,y) - \frac{e(t,x)e(t,y)}{V(t)} - \frac{e(t^2,x)e(t^2,y)}{V(t^2)}}{\sqrt{\left[V(y) - \frac{e^2(t,y)}{V(t)} - \frac{e^2(t^2,y)}{V(t^2)}\right]\left[V(x) - \frac{e^2(t,x)}{V(t)} - \frac{e^2(t^2,x)}{V(t^2)}\right]}} \cdot \quad (2.39)$$

Regressioonikoefitsient

$$\alpha_1 = \frac{C(x,y) - \dfrac{C(t,x)C(t,y)}{V(t)} - \dfrac{C(t^2,x)C(t^2,y)}{V(t^2)}}{V(x) - \dfrac{C^2(t,x)}{V(t)} - \dfrac{C^2(t^2,x)}{V(t^2)}} . \tag{2.40}$$

Jääkdispersioon

$$\sigma^2_{\hat{y}} = \frac{1}{T}\sum\left[y - b_0 - b_1 t - b_2 t^2 - \alpha_1(x - a_0 - a_1 t - a_1 t^2)\right]^2 =$$

$$= V(x) - \frac{C^2(t,y)}{V(t)} - \frac{C^2(t^2,y)}{V(t^2)} -$$

$$- \frac{\left[C(x,y) - \dfrac{C(t,x)C(t,y)}{V(t)} - \dfrac{C(t^2,x)C(t^2,y)}{V(t^2)}\right]^2}{V(x) - \dfrac{C^2(t,x)}{V(t)} - \dfrac{C^2(t^2,x)}{V(t^2)}} . \tag{2.41}$$

Trendidega aegridade regressioon.

Olgu meil kaks aegrida, mis sisaldavad lineaarseid trende:

$$y_t = y(t) + \eta_t = a_0 + a_1 t + \eta_t , \quad t = 1, \dots, T ; \tag{2.42}$$

$$x_t = x(t) + \xi_t = b_0 + b_1 t + \xi_t , \quad t = 1, \dots, T . \tag{2.43}$$

Regressioonivõrrand on lineaarne

$$\hat{\eta}_t = \alpha_0 + \alpha_1 x_t = \alpha_0 + \alpha_1 (b_0 + b_1 t + \xi_t) . \tag{2.44}$$

Parameetrid α_0 ja α_1 määrame vähimruutude meetodil saadava normaalvõrrandite süsteemi abil:

$$\begin{cases} Ta_0 + a_1\sum t - \sum\eta_t - T\alpha_0 - \alpha_1 T b_0 - \alpha_1 b_1\sum t - \alpha_1\sum\xi_t = 0 ; \\ Ta_0 b_0 + a_1 b_0\sum t + b_0\sum\eta_t - T\alpha_0 b_0 - T\alpha_1 b_0^2 - \alpha_1 b_0 b_1\sum t - \\ - \alpha_1 b_0\sum\xi_t + a_0 b_1\sum t + a_1 b_1\sum t^2 + b_1\sum\eta_t t - \alpha_0 b_1\sum t - \\ - \alpha_1 b_0 b_1\sum t - \alpha_1 b_1^2\sum t^2 - \alpha_1 b_1\sum t\xi_t + a_0\sum\xi_t + \\ + a_1\sum t\xi_t + \sum\eta_t\xi_t - \alpha_0\sum\xi_t - \alpha_1 b_0\sum\xi_t - \alpha_1 b_1\sum t\xi_t - \\ - \alpha_1\sum\xi_t^2 = 0 . \end{cases} \tag{2.45}$$

Aja mastaapi võime muuta nii, et $\sum t = 0$. Peale selle, vähimruutude meetodil määratud trendide parameetrite a_0, a_1, b_0 ja b_1 korral $\sum \eta_t = \sum \xi_t = 0$.

Näitame nüüd, et $\sum t\eta_t = \sum t\xi_t = 0$.

Kui $\sum t = 0$, siis näiteks parameetri b_1 saame määrata valemiga

$$b_1 = \frac{\sum x_t t}{\sum t^2}. \tag{2.46}$$

Asendades valemis (2.46) x_t võrrandi (2.45) parema poolega, saame:

$$b_1 = \frac{\sum (b_0 + b_1 t + \xi_t) t}{\sum t^2} = \frac{b_0 \sum t + b_1 \sum t^2 + \sum t \xi_t}{\sum t^2} =$$

$$= b_1 + \frac{\sum \xi_t t}{\sum t^2}, \tag{2.47}$$

millest

$$\frac{\sum t \xi_t}{\sum t^2} = 0. \tag{2.48}$$

Kuivõrd $\sum t^2 \neq 0$, siis $\sum t \xi_t = 0$. Analoogiliselt võib näidata, et ka $\sum t \eta_t = 0$.

Arvestades eeltoodut, võime süsteemi (2.45) teisendada järgmisele kujule:

$$T\alpha_0 + T\alpha_1 b_0 = Ta_0; \tag{2.49}$$

$$T\alpha_0 b_0 + T\alpha_1 b_0^2 + \alpha_1 b_1^2 \sum t^2 + \alpha_1 \sum \xi_t^2 =$$

$$= Ta_0 b_0 + a_1 b_1 \sum t^2 + \sum \xi_t \eta_t. \tag{2.50}$$

Jagades võrrandid (2.49) ja (2.50) T-ga, saame

$$\begin{cases} \alpha_0 + \alpha_1 b_0 = a_0; \\ \alpha_0 b_0 + \alpha_1 \left(b_0^2 + b_1^2 \frac{\sum t^2}{T} + \frac{\sum \xi_t^2}{T} \right) = \\ = a_0 b_0 + a_1 b_1 \frac{\sum t^2}{T} + \frac{\sum \xi_t \eta_t}{T}. \end{cases} \tag{2.51}$$

Lahendame saadud süsteemi Krameri meetodil:

$$\Delta = \begin{vmatrix} 1 & b_0 \\ b_0 & b_0^2 + b_1^2 \frac{\Sigma t^2}{T} + \frac{\Sigma \xi_t^2}{T} \end{vmatrix} =$$

$$= b_0^2 + b_1^2 \frac{\Sigma t^2}{T} + \frac{\Sigma \xi_t^2}{T} - b_0^2 = b_1^2 \frac{\Sigma t^2}{T} + \frac{\Sigma \xi_t^2}{T} . \qquad (2.52)$$

$$\Delta_{\alpha_0} = \begin{vmatrix} a_0 & b_0 \\ a_0 b_0 + a_1 b_1 \frac{\Sigma t^2}{T} + \frac{\Sigma \eta_t \xi_t}{T} & b_0^2 + b_1^2 \frac{\Sigma t^2}{T} + \frac{\Sigma \xi_t^2}{T} \end{vmatrix} =$$

$$= a_0 b_0^2 + a_0 b_1^2 \frac{\Sigma t^2}{T} + a_0 \frac{\Sigma \xi_t^2}{T} - a_0 b_0^2 - a_1 b_0 b_1 \frac{\Sigma t^2}{T} -$$

$$- b_0 \frac{\Sigma \eta_t \xi_t}{T} =$$

$$= a_0 b_1^2 \frac{\Sigma t^2}{T} + a_0 \frac{\Sigma \xi_t^2}{T} - a_1 b_0 b_1 \frac{\Sigma t^2}{T} - b_0 \frac{\Sigma \eta_t \xi_t}{T} . \qquad (2.53)$$

$$\Delta_{\alpha_1} = \begin{vmatrix} 1 & a_0 \\ b_0 & a_0 b_0 + a_1 b_1 \frac{\Sigma t^2}{T} + \frac{\Sigma \eta_t \xi_t}{T} \end{vmatrix} =$$

$$= a_0 b_0 + a_1 b_1 \frac{\Sigma t^2}{T} + \frac{\Sigma \eta_t \xi_t}{T} - a_0 b_0 =$$

$$= a_1 b_1 \frac{\Sigma t^2}{T} + \frac{\Sigma \eta_t \xi_t}{T} . \qquad (2.54)$$

$$\alpha_0 = \frac{\Delta_{\alpha_0}}{\Delta} = \frac{(a_0 b_1^2 - a_1 b_0 b_1) \frac{\Sigma t^2}{T} + a_0 \frac{\Sigma \xi_t^2}{T} - b_0 \frac{\Sigma \xi_t \eta_t}{T}}{b_1^2 \frac{\Sigma t^2}{T} + \frac{\Sigma \xi_t^2}{T}} =$$

$$= \frac{(a_0 b_1^2 - a_1 b_0 b_1) \Sigma t^2 + a_0 \Sigma \xi_t^2 - b_0 \Sigma \xi_t \eta_t}{b_1^2 \Sigma t^2 + \Sigma \xi_t^2} \qquad (2.55)$$

$$\alpha_1 = \frac{\Delta_{\alpha_1}}{\Delta} = \frac{a_1 b_1 \Sigma t^2 + \Sigma \eta_t \xi_t}{b_1^2 \Sigma t^2 + \Sigma \xi_t^2} . \qquad (2.56)$$

Regressioonivõrrand omandab nüüd kuju

$$\hat{y}_t = \alpha_0 + \alpha_1(b_0 + b_1 t + \xi_t) = \frac{(a_0 b_1^2 - a_1 b_0 b_1)\sum t^2 + a_0 \sum \xi_t^2 - b_0 \sum \xi_t \eta_t}{b_1^2 \sum t^2 + \sum \xi_t^2} +$$

$$+ \frac{a_1 b_1 \sum t^2 + \sum \eta_t \xi_t}{b_1^2 \sum t^2 + \sum \xi_t^2} (b_0 + b_1 t + \xi_t) =$$

$$= \frac{a_0 b_1 \sum t^2 + a_0 \sum \xi_t^2}{b_1^2 \sum t^2 + \sum \xi_t^2} + \frac{(a_1 b_1 \sum t^2 + \sum \xi_t \eta_t) b_1 t}{b_1^2 \sum t^2 + \sum \xi_t^2} +$$

$$+ \frac{a_1 b_1 \sum t^2 + \sum \eta_t \xi_t}{b_1^2 \sum t^2 + \sum \xi_t^2} \xi_t = a_0 + \alpha_1 (b_1 t + \xi_t). \qquad (2.58)$$

Nüüd võime teha järelduse, et kui aegridade korreleerimisel lineaarseid trende ridadest ei eemaldate, siis regressioonivõrrandi liikmed sõltuvad trendide parameetritest a_0, b_1, ajast t, jääkliikmest ξ_t, kovariatsioonist $C(\eta, \xi)$ ja ξ_t dispersioonist.

Vaatame nüüd, milline on tulemus, kui koostada regressioonivõrrand ainult trendide põhjal. Kui trendid on lineaarsed, siis ka regressioonivõrrand peab olema lineaarne:

$$\hat{y}(t) = \beta_0 + \beta_1 x(t) = \beta_0 + \beta_1 (b_0 + b_1 t). \qquad (2.59)$$

Parameetrid β_0 ja β_1 leiame vähimruutude meetodil. Sihifunktsioon on

$$\min S = \sum [y(t) - \hat{y}(t)]^2 = \sum [a_0 + a_1 t - \beta_0 - \beta_1 (b_0 + b_1 t)]^2. \qquad (2.60)$$

Normaalvõrrandi süsteem on

$$\begin{cases} T a_0 - T\beta_0 - T b_0 \beta_1 = 0; \\ T a_0 b_0 - T b_0 \beta_0 - T \beta_1 b_0^2 + a_1 b_1 \sum t^2 - \beta_1 b_1^2 \sum t^2 = 0, \end{cases}$$

millest

$$\begin{cases} \beta_0 + \beta_1 b_0 = a_0, \\ \beta_0 b_0 + \beta_1 \left(b_0^2 + b_1^2 \frac{\sum t^2}{T}\right) = a_0 b_0 + a_1 b_1 \frac{\sum t^2}{T}. \end{cases} \qquad (2.61)$$

Lahendame süsteemi (2.61) Krameri meetodil:

$$\Delta = \begin{vmatrix} 1 & b_0 \\ b_0 & b_0^2 + b_1^2 \frac{\sum t^2}{T} \end{vmatrix} = b_1^2 \frac{\sum t^2}{T}. \tag{2.62}$$

$$\Delta_{\beta_0} = \begin{vmatrix} a_0 & b_0 \\ a_0 b_0 + a_1 b_1 \frac{\sum t^2}{T} & b_0^2 + b_1^2 \frac{\sum t^2}{T} \end{vmatrix} =$$

$$= a_0 b_0^2 + a_0 b_1^2 \frac{\sum t^2}{T} - a_0 b_0^2 - a_1 b_0 b_1 \frac{\sum t^2}{T} =$$

$$= (a_0 b_1^2 - a_1 b_0 b_1) \frac{\sum t^2}{T}. \tag{2.63}$$

$$\Delta_{\beta_1} = \begin{vmatrix} 1 & a_0 \\ b_0 & a_0 b_0 + a_1 b_1 \frac{\sum t^2}{T} \end{vmatrix} = a_1 b_1 \frac{\sum t^2}{T}. \tag{2.64}$$

$$\beta_0 = \frac{\Delta_{\beta_0}}{\Delta} = \frac{(a_0 b_1^2 - a_1 b_0 b_1) \frac{\sum t^2}{T}}{b_1^2 \sum t^2 / T} = \frac{a_0 b_1 - a_1 b_0}{b_1}. \tag{2.65}$$

$$\beta_1 = \frac{\Delta_{\beta_1}}{\Delta} = \frac{a_1 b_1 \sum t^2 / T}{b_1^2 \sum t^2 / T} = \frac{a_1}{b_1}. \tag{2.66}$$

Nüüd konstrueerime regressioonivõrrandi jääkliikmete põhjal:

$$\hat{\eta}_t = \gamma_0 + \gamma_1 \xi_t. \tag{2.67}$$

Vähimruutude meetodil saame normaalvõrrandite süsteemi

$$\begin{cases} \gamma_0 = 0, \\ \gamma_1 = \dfrac{\sum \eta_t \xi_t}{\sum \xi_t^2}, \end{cases} \tag{2.68}$$

sest $\sum \eta_t = \sum \xi_t = 0.$

Summeerides nüüd trendide ja jääkliikmete regresseerimise tulemused, saame

$$\hat{Y}_t = \hat{y}(t) + \hat{\eta}_t = \beta_0 + \beta_1 x(t) + \gamma_1 \xi_t =$$

$$= a_0 - b_0 \frac{a_1}{b_1} + b_0 \frac{a_1}{b_1} + a_1 t + \gamma_1 \xi_t =$$

$$= a_0 + a_1 t + \gamma_1 \xi_t = y(t) + \hat{\eta}_t . \tag{2.69}$$

Seega saime tulemuseks resultaatnähtuse trendi ja jääkliikmete regressiooni summa.

Näide 2. Vaatleme aegridade regressiooni näite 1 andmete põhjal.

Resultaatnähtuse lineaarne trend avaldub võrrandiga

$$\hat{y}(t) = 4352{,}0 + 157{,}85t \qquad t = -7, -5, \ldots, 5, 7.$$

Nüüd leiame jääkliikmete regressioonivõrrandi. Arvutused on toodud tabelis 3.

$$\hat{x}(t) = 856{,}875 - 7{,}185t \ ;$$

$$\gamma_1 = \frac{1581{,}7}{2534{,}7464} = 0{,}624 \ .$$

Seega saame analüüsitavate aegridade regressioonivõrrandiks

$$\hat{Y}_t = 4352{,}0 + 157{,}85t + 0{,}624\xi_t \ .$$

2.4. PAARISKORRELATSIOONI JA -REGRESSIOONI TULEMUSTE KONTROLLIMINE DISPERSIOONANALÜÜSIGA

Lineaarset trendi sisaldavate aegridade regressioonivõrrandiks saime

$$\hat{Y}_t = a_0 + \alpha_1 (b_1 t + \xi_t), \tag{2.70}$$

kus α_1 on regressioonikoefitsiendi α_1' hinnanguks võrrandis

$$Y_t = a_0 + \alpha_1' (b_1 t + \xi_t) + \varepsilon_t \ . \tag{2.71}$$

Traditsiooniline kontroll toimub nõnda.

Eeldades ε_t normaaljaotust, võib kontrollida nullhüpoteesi ($H_0 : \alpha_1' = 0$) F-kriteeriumi abil, kasutades dispersioonanalüüsi tabelit 4.

T a b e l 3

Regressioonanalüüsi arvutused

Aasta	t	y_t	$\hat{y}(t)$	η_t	x_t	$\hat{x}(t)$	ξ_t	$\xi_t \eta_t$	ξ_t^2
1969	-7	3341	3247,05	93,95	909	907,11	1,89	177,5655	3,5721
1970	-5	3526	3562,75	-36,75	893	892,75	0,25	- 9,1875	0,0625
1971	-3	3799	3878,45	-79,45	881	878,43	2,57	-204,1865	6,6049
1972	-1	4136	4194,15	-58,15	866	864,06	1,94	-112,8110	3,7636
1973	1	4601	4509,85	91,15	866	849,90	16,10	1467,5150	259,2100
1974	3	4838	4825,55	12,45	791	835,32	-44,32	-551,7840	1964,2624
1975	5	5036	5141,25	-105,25	826	820,90	5,10	-536,7750	26,0100
1976	7	5539	5456,95	82,05	823	806,53	16,47	1351,3635	271,2669
		34816	34816,00	0	6855	6855,00		1581,70	2534,7464

T a b e l 4

Dispersioonanalüüsi tabel

Variatsiooni allikas	Ruutude summa	Vabadus-astmed	Keskmine ruut	F
Regressioon (α_1)	$Q_1=\sum(\hat{y}_t-\bar{y})^2= =\alpha_1^2\sum(x_t-\bar{x})^2$	1	$\sum(\hat{y}_t-\bar{y})^2$	$\frac{\sum(\hat{y}_t-\bar{y})^2}{\sum\varepsilon_t^2/(T-2)}$
Jääkliige (regressioonist)	$Q_2=\sum(y_t-\hat{y}_t)^2= =\sum\varepsilon_t^2$	T-2	$\frac{\sum\varepsilon_t^2}{T-2}$	
Kõik	$Q=Q_1+Q_2= =\sum(y_t-\bar{y})^2$	T-1		

Hälvete ruutude summad tabelis (4) võib teisendada järgmiselt:

$$Q_1=\sum(\hat{y}_t-\bar{y})^2=\sum[a_0+\alpha_1(b_1t+\xi_t)-a_0]^2=$$
$$=\alpha_1^2(b_1^2\sum t^2+\sum\xi_t^2). \qquad (2.72)$$

$$Q_2=\sum\varepsilon_t^2=\sum(y_t-\hat{y}_t)^2=$$
$$=\sum[a_0+a_1t+\eta_t-a_0-\alpha_1(b_1t+\xi_t)]^2=$$
$$=a_1^2\sum t^2+\sum\eta_t^2-\alpha_1^2(b_1^2\sum t^2+\xi_t^2). \qquad (2.73)$$

$$Q=\sum(y_t-\bar{y})^2=\sum(a_0+a_1t+\eta_t-a_0)^2=a_1^2\sum t^2+\sum\eta_t^2. \qquad (2.74)$$

On teada, et F-jaotus kehtib sõltumatute χ^2-muutujate suhte korral, kusjuures kumbki χ^2 jagatakse läbi oma vabadusastmetega. χ^2-jaotus kehtib omakorda standardse normaaljaotusega muutujate ruutude korral. Seega, eeldades (2.71)-s jääkliikmete normaaljaotust, jääkliikmete ε_t ruutude summa, jagatud nende dispersiooni hinnanguga σ_ε^2, on jaotunud nagu χ^2 vabadusastmetega T-2:

$$\chi^2_{\alpha,T-2}=\frac{\sum\varepsilon_t^2}{\sigma_\varepsilon^2}. \qquad (2.75)$$

Jääkliikmete normaaljaotusest johtub samuti

$$\frac{\alpha_1-\alpha_1'}{\sigma_{\alpha_1}}=\frac{(\alpha_1-\alpha_1')\sqrt{\sum(x_t-\bar{x})^2}}{\sigma_\varepsilon}\sim N(0,1). \qquad (2.76)$$

Järelikult

$$\frac{(\alpha_1-\alpha_1')^2 \sum (x_t-\bar{x})^2}{\sigma_\varepsilon} \tag{2.77}$$

on jaotunud nagu χ^2 vabadusastmega 1. Kui lähtuda asjaolust, et (2.75) ja (2.77) on sõltumatud, siis võib nullhüpoteesi ($H_0 : \alpha_1' = 0$) kontrollida suuruse

$$F = \frac{Q_1}{Q_2/(T-2)} \tag{2.78}$$

abil, vastavalt vabadusastmeile 1 ja T-2. Kasutades valemeid (2.72) ja (2.73), saame

$$F = \frac{Q_1}{Q_2/(T-2)} = \frac{\alpha_1^2(b_1\Sigma t^2+\Sigma \xi_t^2)(T-2)}{(a_1^2\Sigma t^2+\Sigma \eta_t^2)-\alpha_1^2(b_1^2\Sigma t^2+\Sigma \xi_t^2)} \tag{2.79}$$

Kuigi formaalselt valemi (2.79) lugeja ja nimetaja võivad olla sõltumatud (lugejas on ainult ühe juhusliku suuruse dispersioon $\Sigma \xi_t^2$, nimetajas kahe juhusliku suuruse dispersioonid $\Sigma \eta_t^2$ ja $\Sigma \xi_t^2$), suhe annab ikkagi väära tulemuse, sest peale parameetrite a_1 , b_1 on valemis ka veel aja dispersioon Σt^2 . Järelikult esialgu loogilisena paistnud tulemus pole õige ning lineaarseid trende sisaldavate aegridade korrelatsioon- ja regressioonanalüüsi tulemuste kontrollimine tavalise F-kriteeriumi abil pole võimalik.

Analoogilisele tulemusele jõuame ka Studenti t-kriteeriumi kasutamise suhtes.

Kontrollsuurus t nullhüpoteesi kontrollimise korral on

$$t = \frac{\alpha_1\sqrt{\Sigma[(x_t-\bar{x})-\overline{(x_t-\bar{x})}]^2}}{\sigma_\varepsilon} = \frac{\alpha_1\sqrt{(b_1\Sigma t^2+\Sigma \xi_t^2)(T-2)}}{\sqrt{(a_1^2\Sigma t^2+\Sigma \eta_t^2)-(b_1^2\Sigma t^2+\Sigma \xi_t^2)\alpha_1^2}} \tag{2.80}$$

sest sõltumatuks muutujaks on $b_1 t + \xi_t$ ja regressioonivõrrandi jääkliikme standardhälbe hinnang on

$$\sigma_{\varepsilon_t} = \sqrt{\frac{\Sigma(\varepsilon_t-\bar{\varepsilon})^2}{T-2}} = \sqrt{(a_1^2\Sigma t^2+\Sigma \eta_t^2)-(b_1^2\Sigma t^2+\Sigma \xi_t^2)\alpha_1^2} \tag{2.81}$$

Võrreldes kontrollsuurusi (2.79) ja (2.80), näeme, et $F=t^2$, mida oligi oodata.

Analoogilistele tulemustele jõuame ka korrelatsioonikoefitsiendi kontrollimisel.

2.5. AUTOREGRESSIIVSED PROTSESSID

Autoregressiivseteks protsessideks nimetatakse selliseid aegridu, kus aegrea tase mingil ajamomendil t sõltub sellele eelnevatest tasemetest:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + u_t . \qquad (2.82)$$

Võrrand (2.82) esitatakse tavaliselt kujul

$$Y_t + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} = u_t , \quad t = p+1, \ldots , \qquad (2.83)$$

kus u_{p+1}, u_{p+2}, ... on sõltumatud ühtlaselt jaotunud juhuslikud suurused, kusjuures $E u_t = 0$ ja $E u_t^2 = \sigma^2$, $E u_t u_s = 0$, $t \neq s$. Välismaises kirjanduses nimetatakse võrrandit (2.83) tihti stohhastiliseks vahevõrrandiks (*stochastic difference equations*).

Iga juhusliku suuruse Y_t võib esitada temale eelnevate ja juhusliku suuruse Y_τ ja u_τ lineaarse kombinatsioonina.

Lihtsaimaks juhtumiks on esimest järku võrrand

$$Y_t = \rho Y_{t-1} + u_t . \qquad (2.84)$$

Kui võrrandisse (2.84) asendada Y_{t-1} temale vastava võrrandiga $Y_{t-1} = \rho Y_{t-2} + u_{t-1}$, saame

$$Y_t = u_t + \rho u_{t-1} + \rho^2 Y_{t-2} . \qquad (2.85)$$

Korrates asendamisi, saame lõpuks

$$Y_t = u_t + \rho u_{t-1} + \ldots + \rho^s u_{t-s} + \rho^{s+1} Y_{t-(s+1)} , \qquad (2.86)$$

millest

$$Y_t - (u_t + \rho u_{t-1} + \ldots + \rho^s u_{t-s}) = \rho^{s+1} Y_{t-(s+1)} . \qquad (2.87)$$

Võrrandi (2.87) matemaatiline ootus ei sõltu t-st ja läheneb nullile s kasvamisel. Seetõttu

$$Y_t = \sum_{\tau=0}^{\infty} \rho^\tau u_{t-\tau} . \qquad (2.88)$$

Võrrandi (2.83) võime esitada kujul

$$Y_t = u_t - \beta_1 Y_{t-1} - \beta_2 Y_{t-2} - \ldots - \beta_p Y_{t-p} , \qquad (2.89)$$

kus

$$Y_{t-1} = u_t - \beta_1 Y_{t-2} - \ldots - \beta_p Y_{t-1-p} . \qquad (2.90)$$

Asendades (2.90) (2.89)-sse, saame

$$Y_t = u_t - \beta_1 u_{t-1} - (\beta_2 - \beta_1)^2 Y_{t-2} - \ldots + \beta_1 \beta_p Y_{t-1-p} . \tag{2.91}$$

Korrates seda protseduuri s korda, saame

$$\begin{aligned} Y_t = u_t + \delta_1 u_{t-1} + \ldots + \delta_s u_{t-s} + \alpha_{s1} Y_{t-s-1} + \\ + \alpha_{s2} Y_{t-s-2} + \ldots + \alpha_{sp} Y_{t-s-p} . \end{aligned} \tag{2.92}$$

Asendades (2.92)-s

$$Y_{t-s-1} = u_{t-s-1} - \beta_1 Y_{t-s-2} - \ldots - \beta_p Y_{t-s-p-1} , \tag{2.93}$$

saame

$$\begin{aligned} Y_t = u_t + \delta_1 u_{t-1} + \ldots + \delta_s u_{t-s} + \alpha_{s1} u_{t-s-1} + \\ + (\alpha_{s2} - \alpha_{s1}\beta_1) Y_{t-s-2} + \ldots + (\alpha_{sp} - \alpha_{s1}\beta_{p-1}) Y_{t-s-p} - \\ - \alpha_{s1}\beta_p Y_{t-s-p+1} . \end{aligned} \tag{2.94}$$

Koefitsiendid võrrandis (2.94) leitakse rekurrentsete võrranditega

$$\begin{cases} \delta_{s+1} = \alpha_{s1} ; \\ \alpha_{s+1,j} = \alpha_{s,j+1} - \alpha_{s1}\beta_j , \quad j = 1, \ldots, p-1 ; \\ \alpha_{s-1,p} = - \alpha_{s1}\beta_p . \end{cases} \tag{2.95}$$

Eeltoodud protseduuri kordamine annab lõpuks

$$Y_t = \sum_{\tau=0}^{\infty} \delta_\tau u_{t-\tau} , \quad \delta_0 = 1 . \tag{2.96}$$

On tõestatud, et küllalt suure τ korral võrrandite (2.88) ja (2.96) paremad pooled korduvad. See tähendab, et autoregressiivsel protsessil on alati olemas järk p $(0 \leq p < \infty)$.

Eeltoodust on selgesti näha, et autoregressiivsetel protsessidel on prognoosimisel väga tähtis osa. Kui on määratud autoregressiivse protsessi järk p ning parameetrid β_i, võime leida protsessi väärtused momentidel $t+1$, $t+2$, ... Autoregressiivse protsessina võib esitada ükskõik milliseid aegridu; majandusliku analüüsis uuritakse tavaliselt jääkliikmete η_t (hälbed trendist, perioodilisest ja sesoonsest komponendist) esitamise võimalust autoregressiivse protsessina.

Põhiliseks tööks autoregressiooni uurimisel on parameetrite β_i ja protsessi järgu p määramine. Parameetreid on võimalik määrata nii autokorrelatsiooni koefitsientide kaudu kui ka tavalisel vähimruutude meetodil.

Kui u_t on jaotunud normaalselt (üldreeglina on) ja kasutame mudelit (2.83), siis saame maksimaalse tõepära meetodil, kasutades autokorrelatsiooni koefitsiente, parameetrite β määramiseks järgmised rekurrentsed võrrandid:

$$\beta_{p+1}(p+1) = -\frac{r_{p+1} + \sum_{j=1}^{p} r_{p+1-j}\beta_j(p)}{1 + \sum_{j=1}^{p} r_j \beta_j(p)}, \qquad (2.97)$$

$$\beta_j(p+1) = \beta_j(p) + \beta_{p+1-j}(p)\, \beta_{p+1}(p+1), \quad j=1,\dots,p. \qquad (2.98)$$

<u>Nullindat järku</u> protsessi korral $\beta_0 = 0$ ja $r_0 = 1$. <u>Esimest järku</u> protsessi parameetri β_1 saame (2.97)-st:

$$\beta_1(1) = -\frac{r_1 + r_0\beta_0(0)}{1 + r_0\beta_0(0)} = -r_1. \qquad (2.99)$$

<u>Teist järku</u> protsessi parameetrid β_1 ja β_2 on

$$\beta_2(2) = -\frac{r_2 + r_1\beta_1(1)}{1 + r_1\beta_1(1)} = -\frac{r_2 - r_1^2}{1 - r_1^2}; \qquad (2.100)$$

$$\beta_1(2) = \beta_1(1) + \beta_1(1)\beta_2(2) = -\frac{r_1(1 + r_2)}{1 - r_1^2}. \qquad (2.101)$$

Analoogiliselt saame ka kõrgemat järku protsesside parameetrid.

Normaaljaotuse korral annavad vähimruutude meetod ja maksimaalse tõepära meetod ühesugused tulemused. Demonstreerime seda esimest järku protsessi abil.

Mudeliks on

$$y_t + \beta y_{t-1} = u_t.$$

Sihifunktsiooniks saame

$$S = \sum (y_t - u_t + \beta y_{t-1})^2 \to \min,$$

mille diferentseerimine annab

$$\frac{\partial S}{\partial \beta} = \sum y_{t-1}(y_t - u_t + \beta y_{t-1}).$$

Kuivõrd

$$\sum y_{t-1} u_t = 0,$$

siis

$$\sum y_t y_{t-1} + \beta \sum y^2_{t-1} = 0$$

ja

$$\beta = - \frac{\sum y_t y_{t-1}}{\sum y^2_{t-1}} = - r_1 . \qquad (2.102)$$

Autoregressiooni uurimisel võib kasutada ka mudelit, mis sisaldab mõningaid sõltumatuid eksogeenseid muutujaid:

$$y_t + \sum_{i=1}^{p} b_i y_{t-i} + \sum_{j=1}^{q} c_j z_{jt} = u_t . \qquad (2.103)$$

Vähimruutude meetodi sihifunktsioon on

$$S = \sum_{t=1}^{T} (y_t + \sum_{i=1}^{p} b_i y_{t-i} + \sum_{j=1}^{q} c_j z_{jt}) \to min . \qquad (2.104)$$

Tavaliselt eeldatakse, et $z_{1t} = 1$ ning c_1 jäetakse parameetrite arvutamisel kõrvale. Kui $q = 1$, siis parameetrid b_i saame võrrandist

$$\sum_{j=1}^{p} a_{ij} b_j = - \alpha_{i0} , \quad i = 1, \dots, p , \qquad (2.105)$$

kus

$$a_{ij} = \sum_{t=1}^{T} y_{t-i} y_{t-j} - T \bar{y}_{(i)} \bar{y}_{(j)} , \quad i,j = 1, \dots, p ; \qquad (2.106)$$

$$a_{i0} = \sum_{t=1}^{T} y_{t-i} y_t - T \bar{y}_{(i)} \bar{y} , \quad i = 1, \dots, p \qquad (2.107)$$

$$\bar{y} = y_{(0)} , \quad \bar{y}_{(i)} = \frac{1}{T} \sum_{t=1}^{T} y_{t-i} , \quad i = 0, 1, \dots, p \qquad (2.108)$$

Autoregressiivse protsessi parameetrite usaldatavuse kontrollimiseks võib kasutada standardiseeritud normaaljaotust N (0,1), arvutades kriteeriumi empiirilise väärtuse valemiga

$$N(\beta_i) = \frac{\beta_i - N(T) s \sqrt{\bar{a}^{ii}}}{\sqrt{T}} , \qquad (2.109)$$

kus $N(T)$ määratakse standardiseeritud normaaljaotuse tabelist vastavalt T väärtusele,

$$s^2 = \frac{T \sigma^2}{T-p-q} \quad \frac{\sum_{t=1}^{T} u^2_t}{T-p-q} , \qquad (2.110)$$

$\bar{a}^{ii}$ on maatriksi A diagonaali element;

$$\begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 & \cdots & \frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_{t-p} \\ \vdots & & \vdots \\ \frac{1}{T}\sum_{t=1}^{T} y_{t-p}y_{t-1} & \cdots & \frac{1}{T}\sum_{t=1}^{T} y_{t-p}^2 \end{bmatrix} \qquad (2.111)$$

<u>Autoregressiivse protsessi järgu määramine</u> seisneb valikus, kas järk on p või p-1, ehk teisiti öelduna kontrollimises, kas $\beta_p = 0$. Püstitatakse nullhüpotees $H_0: \beta_p = 0$. Nullhüpotees lükatakse ümber, kui

$$t_{\beta_p} = \frac{|\beta_p|}{s\sqrt{\dfrac{1}{\sum_{t=p+1}^{T} y_{t-p}^2 - T\bar{y}_{(p)}^2}}} > t(\varepsilon), \qquad (2.112)$$

kus s on määratud valemiga (2.110), $t(\varepsilon)$ vastab 100 -protsendilisele kahepoolsele punktile standardiseeritud normaaljaotuses. Kui nullhüpotees osutub õigeks, siis tuleb kontrollida nullhüpoteesi $H_0: \beta_{p-1} = 0$ jne., kuni protsessi järk on määratud. Kuivõrd kõrgemat järku protsessi parameetrid sõltuvad madalama järgu parameetritest (vt. võrrandid (2.98) ja (2.99)), siis on otstarbekas paralleelselt $\beta_{p+1}^{(p+1)}$ leidmisega kontrollida ka nullhüpoteesi $H_0: \beta_{p+1} = 0$. Protsessi järk on määratud, kui madalamatelt järkudelt kõrgematele üle minnes on leitud selline β_p, mille korral nullhüpotees ei kehti. Muidugi ei saa võtta $\max p = T$. Küllalt suurte T väärtuste korral (T > 100) võib võtta $\max p = \frac{T}{2}$, kuid praktikas on väga harvad juhud, kus $p \geq \frac{T}{4}$ või $p > \frac{T}{5}$.

3. MITMENE KORRELATSIOON JA REGRESSIOON

3.1. ÜLDINE LINEAARNE MUDEL

Mingi majandusnähtuse muutumine on harva seotud ainult ühe teguri muutumisega. Enamasti moodustavad majandusnähtused kompleksi, kus üht nähtust mõjutab terve rida tegureid, mis omakorda võivad olla resultaatnähtusteks ja sõltuda mingist kolmandast tegurite ringist.

Mitme teguri mõju resultaatnähtusele uurib mitmene korrelatsioon- ja regressioonanalüüs. Mitmeste seoste analüüsimisel on alati primaarseks regressioonanalüüs. Alles peale seose vormi määramist saab uurida seose tugevust. Mitmese regressiooni võrrandid võib konstrueerida kas tavalises või normeeritud mastaabis, viimasel puhul on erinevad nähtused esitatud võrreldavates ühikutes.

Kõige sagedamini on mitmese regressiooni võrrandina kasutatav lineaarne funktsioon

$$\hat{y} = a_0 + a_1 x_1 + \ldots + a_n x_n . \tag{3.1}$$

Kasutatakse ka veel järgmisi funktsioone:

$$\hat{y} = a_0 x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} , \tag{3.2}$$

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_{11} x_1^2 + \\ + a_{12} x_1 x_2 + \ldots \quad . \tag{3.3}$$

Funktsioonide (3.1) - (3.3) parameetrid võib leida vähimruutude meetodil. Normaalvõrrandite süsteemi üldkuju on järgmine:

$$\begin{cases} T a_0 + a_1 \Sigma x_1 + a_2 \Sigma x_2 + \ldots + a_n \Sigma x_n = \Sigma y , \\ a_0 \Sigma x_1 + a_1 \Sigma x_1^2 + a_2 \Sigma x_1 x_2 + \ldots + a_n \Sigma x_1 x_n = \Sigma x_1 y , \\ \vdots \qquad \vdots \qquad \vdots \qquad \ldots \qquad \vdots \qquad \vdots \\ a_0 \Sigma x_n + a_1 \Sigma x_1 x_n + a_2 \Sigma x_2 x_n + \ldots + a_n \Sigma x_n^2 = \Sigma x_n y . \end{cases} \tag{3.4}$$

Võrrandi (3.2) parameetrite leidmiseks tuleb võrrand kõigepealt logaritmimisega teisendada lineaarsele kujule:

$$\ln \hat{y} = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2 + \ldots + a_n \ln x_n . \qquad (3.5)$$

Võrrandis (3.3) käsitletakse tegurite astmeid ja korrutisi iseseisvate teguritena.

Mitmese regressiooni võrrandeid, parameetrite hindamist ja mitmesuguseid kontrollimisi on mugav kirjeldada maatriksvormis. Lineaarse mudeli (3.1) (ka teised mudelid mõningate teisendustega) võime kirjutada kujul

$$\mathcal{Y} = \mathcal{X}A + u, \qquad (3.6)$$

kus

$$\mathcal{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix}; \quad \mathcal{X} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{n1} \\ 1 & x_{12} & \ldots & x_{n2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1T} & \ldots & x_{nT} \end{bmatrix}; \quad A = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}; \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}$$

Maatriksis $\mathcal{X}$ ühtedest koosnev veerg on tingitud vabaliikme olemasolust võrrandis (3.6). Vektor u kujutab endast resultaatnähtuse tegelike ning regressioonivõrrandi kaudu leitud väärtuste vaheliste hälvete vektorit. Neid hälbeid nimetatakse tavaliselt kas "juhuslikeks liikmeteks" või "jääkliikmeteks". Edaspidi kasutame viimast terminit.

Regressioonivõrrandilt (3.6) nõutakse, et jääkliikmed u_t oleksid juhuslikud, keskmisega 0 ning dispersiooniga $\sigma^2 \neq 0$, s. t.:

$$E(u) = 0; \qquad (3.7)$$

$$E(uu') = \sigma^2 I. \qquad (3.8)$$

Siit järeldub, et jääkliikmete u_t ja u_{t+k} kovariatsioon peab olema null

$$E(u_t, u_{t+k}) = 0, \quad k \neq 0, \qquad (3.9)$$

kuivõrd

$$E(u_t, u_t) = \sigma^2. \qquad (3.10)$$

Vähimruutude meetodil on vaja leida vektori A elementide hinnangud $\hat{a}_i$ (mittetäieliku informatsiooni korral) nii, et oleks minimiseeritud jääkliikmete hinnangu vektori e pikkus. Siin muudame tähistusi, sest vektor u on tegelike jääkliikmete vektor, vektor e aga jääkliikmete hinnangu vektor, mis saadakse peale parameetrite hindamist:

$$e = \mathcal{Y} - \mathcal{X}\hat{A}. \qquad (3.11)$$

Kuna vektori e pikkus

$$\sum_{t=1}^{T} e_t^2 = e'e = (\mathcal{Y} - \mathcal{X}\hat{A})'(\mathcal{Y} - \mathcal{X}\hat{A}) = \mathcal{Y}'\mathcal{Y} - 2\mathcal{Y}'\mathcal{X}\hat{A} + \hat{A}'\mathcal{X}'\mathcal{X}A,$$

siis, võttes osatuletised $\hat{A}$ järgi ning võrdsustades need nulliga, saame

$$\frac{\partial}{\partial \hat{A}}(e'e) = -2\mathcal{X}'\mathcal{Y} + 2\mathcal{X}'\mathcal{X}\hat{A} = 0,$$

millest

$$\mathcal{X}'\mathcal{X}\hat{A} = \mathcal{X}'\mathcal{Y}$$

ja

$$\hat{A} = (\mathcal{X}'\mathcal{X})^{-1}\mathcal{X}'\mathcal{Y}, \qquad (3.12)$$

mis ongi otsitavaks parameetrite hinnangute vektoriks.

3.2. LINEAARSE MUDELI PARAMEETRITE USALDUSPIIRID

Mudeli (3.6) parameetrite hinnangute usaldusväärsuse kontrollimiseks eeldame, et jääkliikmed jaotuvad normaalselt:

$$u \sim N(0, \sigma^2 I). \qquad (3.13)$$

Võib näidata, et ka vektori $\hat{A}$ elemendid vastavad mitmemõõtmelisele normaaljaotusele:[1]

$$\hat{A} \sim N[A, \sigma^2(\mathcal{X}'\mathcal{X})^{-1}]. \qquad (3.14)$$

Kuivõrd ruutude summa $e'e$ jaotus ei sõltu $\hat{A}$ jaotusest, siis parameetrite usaldusväärsuse kontrollimiseks võib kasutada t-kriteeriumi.

Vastavalt Studenti jaotusele

$$t = \frac{\hat{a}_i - a_i}{\sqrt{\left[\sum_{t=1}^{T} e_t^2 / (T-n-1)\right] c_{ii}}}. \qquad (3.15)$$

[1] Vt. näiteks Феллер В. Введение в теорию вероятностей и ее приложения. М.: Мир, 1964, т. I, с. 259.

vabadusastmetega (T-n-1), kus c_{ii} on maatriksi $(\mathcal{X}'\mathcal{X})^{-1}$ diagonaalielement. Kontrollimisel pannakse valemisse (3.15) hinnang $\hat{a}_i$ ning a_i hüpoteetiline väärtus, leitakse t-kriteeriumi empiiriline väärtus ning võrreldakse seda t-kriteeriumi teoreetilise väärtusega $t_{T-n-1,\alpha}$, mis vastab vabadusastmete arvule (T-n-1) ning vea tõenäosusele α. Kui $t_{emp.} > t_{T-n-1,\alpha}$, siis hinnang $\hat{a}_i$ erineb oluliselt hüpoteetilisest väärtusest a_i ja ütleme, et hinnang $\hat{a}_i$ on usaldusväärne (tavaliselt kontrollitakse hüpoteesi $H_0 : a_i = 0$).

(3.15)-st järeldub, et 100(1 - α)-%-lised usalduspiirid $\hat{a}_i$ jaoks saadakse järgmiselt:

$$\hat{a}_i \pm t_{\alpha/2} \sqrt{\frac{\sum e_i^2}{T-n-1}} \sqrt{c_{ii}} \quad (3.16)$$

3.3. FIKTIIVSED MUUTUJAD

Mõnikord on tegemist seoste uurimisega selliste nähtuste vahel, kus mingil põhjusel on teatud momendil toimunud nihked nähtuste struktuuris või mastaabis. Näiteks osa andmeid on sõjaaegse, osa rahuaegse perioodi kohta või on mingil momendil toimunud hindade muutumine, mis mõjustab nähtuste struktuuri, või midagi muud. Lihtne on seda situatsiooni illustreerida tarbimise kaudu. Olgu tarbimise maht sõltuv tuludest sõjaajal

$$y = a_1 + bx + u \quad (3.17)$$

ja rahuajal

$$y = a_2 + bx + u, \quad a_2 > a_1 \quad (3.18)$$

kus koefitsiendi b sisu on mõlemal perioodil ühesugune - näitab kalduvust tarbimisele (siin peitubki üks tähtis moment fiktiivsete muutujate meetodi juures - regressioonikoefitsiendi sisu peab jääma eri perioodidel samaks; vastasel korral ei jää muud üle, kui kumbki periood esitada eri funktsiooniga). See võimaldab esitada võrrandid (3.17) ja (3.18) kompaktsel kujul

$$y = a_1 z_1 + a_2 z_2 + bx + u, \quad (3.19)$$

kus z_1 ja z_2 on fiktiivsed muutujad:

$$z_1 = \begin{cases} 1 \text{ igal sõja-aastal,} \\ 0 \text{ igal rahuaastal;} \end{cases}$$

$$z_2 = \begin{cases} 0 \text{ igal sõja-aastal,} \\ 1 \text{ igal rahuaastal .} \end{cases}$$

Siinjuures tuleb silmas pidada asjaolu, et mudel (3.19) ei sisalda vabaliiget selle traditsioonilises mõttes. Tegelikult on vabaliikmeteks koefitsiendid a_1 ja a_2, neist esimene vastab sõjaajale ning teine rahuajale. Vabaliiget n.-ö. "puhtal kujul" leida pole võimalik. Kui tahetakse saada võrrandit vabaliikmega (arvutustehnika võimaldab ainult selliste võrrandite parameetrite leidmist), siis tuleb regressioonivõrrand esitada kujul

$$y = c_0 + c_1 z + bx + u, \tag{3.20}$$

kus

$$z = \begin{cases} 0 \text{ igal sõja-aastal,} \\ 1 \text{ igal rahuaastal .} \end{cases}$$

Nüüd c_0 on vabaliige sõja-aastate jaoks ning $c_0 + c_1$ - vabaliige rahuaastate jaoks.

x x
x

Aegridade korrelatsioon- ja regressioonanalüüsi juures võib kokku puutuda ka mõningate spetsiaalküsimustega, nagu jaotatud viitajad, peakomponendid, kanoonilised korrelatsioonid, diskriminantanalüüs jm., kuid nende käsitlemist ei võimalda õppevahendi maht ning taolistel juhtudel on vaja kasutada spetsiaalsete monograafiate abi.

4. SPEKTRAALANALÜÜS

Aegridade spektraalanalüüsi teooria baseerub eeldusel, et protsessi dispersiooni võib jaotada lõigul $-\pi, \pi$ asuvate sageduste λ vahel. Kui dispersioon jaotub nii, et $f(\lambda-\Theta) \approx f(\lambda) \approx f(\lambda+\Theta)$, s. t. ühelegi sagedusele ei vasta oluliselt suurem osa dispersioonist, siis on tegemist puhtjuhusliku protsessiga, mis ei sisalda mingit informatsiooni. Selliseid protsesse nimetatakse valgeks müraks. Kui mingile sagedusele vastab oluliselt suurem osa protsessi dispersioonist kui naabersagedustele, siis protsess sisaldab vastava sagedusega perioodilist võnkumist. Järelikult, spektraalanalüüsi tulemused võimaldavad oluliselt täpsustada prognoosivaid võrrandeid.

4.1. STATSIONAARSED PROTSESSID

Eeldame, et meil on antud aegrea tasemed võrdse pikkusega ajavahemike järgi, s. t. diskreetse ajaga protsess. Üldreeglina eeldatakse, et protsessi tasemed on jaotunud normaalselt.

Olgu

$$E y_t = m(t), \tag{4.1}$$

$$E[y_t - m(t)][y_s - m(s)] = C(y_t, y_s) = \sigma(t,s). \tag{4.2}$$

Kui tasemete $y_{t_1}, \dots y_{t_n}$ jaotus on samasugune kui tasemete $y_{t_1+t}, \dots, y_{t_n+t}$ jaotus, siis öeldakse, et protsess on <u>statsionaarne kitsas mõttes</u>. Statsionaarsus ise tähendab aga seda, et protsessi statistilised karakteristikud ei sõltu momentide t ja s asendist ajateljel, vaid sõltuvad ainult ajavahemiku t - s pikkusest.

Kui tingimused

$$E y_s = E y_{s+t} = const., \tag{4.3}$$

$$\sigma(t_1, t_2) = \sigma(t_1+t, t_2+t), \tag{4.4}$$

millest

$$\sigma(t_1, t_2) = \sigma(t_1 - t_2, 0) = \sigma(t_1 - t_2), \qquad (4.5)$$

kehtivad, kuid ei seata tingimust jaotuse kohta, siis öeldakse, et protsess on <u>statsionaarne laias mõttes</u> ehk <u>nõrgalt statsionaarne</u>. Normaaljaotuse korral statsionaarsus laiemas mõttes samastub statsionaarsusega kitsamas mõttes.

4.2. SPEKTRAALTIHEDUS JA SPEKTRAALFUNKTSIOON

Olgu antud statsionaarne laias mõttes juhuslik protsess. Selle põhjal saame määrata kovariatsioonifunktsiooni väärtused $\sigma(0), \sigma(1)$. Fourier' teisendus kovariatsioonifunktsiooni suhtes annab

$$f(\lambda) = \frac{1}{2\pi}\sigma(0) + \frac{1}{\pi}\sum_{h=1}^{\infty}\sigma(h)\cos\lambda h =$$

$$= \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\sigma(h)\cos\lambda h, \quad -\pi \le \lambda \le \pi. \qquad (4.6)$$

Kui see rida koondub ühtlaselt, s. t.

$$\sigma(0) + 2\sum_{h=1}^{\infty}|\sigma(h)| = \sum_{h=-\infty}^{\infty}|\sigma(h)| < \infty,$$

siis, peale (4.6) korrutamist $\cos\lambda k$ -ga ja integreerimist saame

$$\sigma(k) = \int_{-\pi}^{\pi}\cos\lambda k\, f(\lambda)\,d\lambda. \qquad (4.7)$$

Seega kovariatsioonifunktsioon määrab funktsiooni ja vastupidi. Funktsiooni $f(\lambda)$ nimetatakse protsessi <u>spektraaltiheduseks</u>.

Spektraalfunktsiooniks nimetatakse funktsiooni

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\nu)\,d\nu, \qquad (4.8)$$

kusjuures

$$\sigma(k) = \int_{-\pi}^{\pi}\cos\lambda k\, dF(\lambda). \qquad (4.9)$$

Kuivõrd $f(\lambda)$ ja $\cos\lambda h$ on paarisfunktsioonid, $\sin\lambda h$ aga paaritu funktsioon, siis (4.6) ja (4.7) võib esitada järgmiselt:

$$f(\lambda)=\frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\sigma(h)e^{i\lambda h} \tag{4.10}$$

ja

$$\sigma(k)=\int_{-\pi}^{\pi}e^{i\lambda k}f(\lambda)d\lambda. \tag{4.11}$$

Normeeritud spektraaltiheduse saame, kui koavariatsioonifunktsiooni normeerime ja kasutame $\sigma(h)$ asemel $\rho_h = \sigma(h)/\sigma(0)$. Siis $\bar{f}(\lambda)=f(\lambda)/\sigma(0)$.

4.3. SPEKTRAALTIHEDUSE MÄÄRAMINE

Olgu antud statsionaarne laias mõttes juhuslik protsess statistiliste karakteristikutega

$$E y_t = \mu, \tag{4.12}$$

$$C(y_t, y_{t+h}) = E(y_t-\mu)(y_{t+h}-\mu) = \sigma(h),$$

$$t=\ldots,-1,0,1,\ldots;\ h=0,\pm1,\pm2,\ldots, \tag{4.13}$$

kusjuures $\sigma(h)=\sigma(-h)$. Oletame, et spektraalfunktsioon on absoluutselt pidev (vajalik tingimus integreerimiseks), tihedusega $f(\lambda)$.

Tavaliselt uurija käsutuses olev andmestik moodustab väljavõtukogumi $y_1,\ldots,y_T$, mis on võetud üldkogumist $\{y_t\}$, $t = \ldots,-1,0,1,\ldots$. Seega tuleb üldkogumi parameetrid hinnata väljavõtukogumi põhjal.

Keskmine: $$\mu=\bar{y}=\frac{1}{T}\sum_{t=1}^{T}y_T. \tag{4.14}$$

Kovariatsioonifunktsiooni võib määrata kolmel erineval kujul:

$$c_h^* = c_{-h}^* = \frac{1}{T-h}\sum_{t=1}^{T-h}(y_t-\bar{y})(y_{t+h}-\bar{y}),\ h=0,1,\ldots,T-1, \tag{4.15}$$

$$\tilde{c}_h = \tilde{c}_{-h} = \frac{1}{T-h}\sum_{t=1}^{T-h}(y_t-\bar{y})(y_{t+h}-\bar{y}_{h+}) =$$

$$=\frac{1}{T-h}\left[\sum_{t=1}^{T-h}y_t y_{t+h}-(T-h)\bar{y}_h\bar{y}_{h+}\right],\ h=0,1,\ldots,T-2, \tag{4.16}$$

kus

$$\bar{y}_h = \frac{1}{T-h} \sum_{t=1}^{T-h} y_t, \quad h = 0,1,\ldots,T-2, \tag{4.17}$$

$$\bar{y}_{h+} = \frac{1}{T-h} \sum_{t=1}^{T-h} y_{t+h}, \; h = 0,1,\ldots,T-2, \tag{4.18}$$

$$\begin{aligned} c_h^* = c_{-h}^* &= \frac{1}{T} \sum_{t=1}^{T-h} (y_t - \bar{y})(y_{t+h} - \bar{y}) = \\ &= \frac{T-h}{T} C_h^*, \quad h = 0, 1, \ldots, T-1. \end{aligned} \tag{4.19}$$

Olgu

$$\begin{aligned} A_j^*(\lambda) &= \frac{2}{T} \sum_{t=1}^{T} (y_t - \bar{y}) \cos \lambda_j t = \\ &= \frac{2}{T} \sum_{t=1}^{T} y_t \cos \lambda_j t - \bar{y} \frac{2}{T} \sum_{t=1}^{T} \cos \lambda_j t, \quad -\pi \le \lambda \le \pi; \end{aligned} \tag{4.20}$$

$$\begin{aligned} B_j^*(\lambda) &= \frac{2}{T} \sum_{t=1}^{T} (y_t - \bar{y}) \sin \lambda_j t = \\ &= \frac{2}{T} \sum_{t=1}^{T} y_t \sin \lambda_j t - \frac{2}{T} \bar{y} \sum_{t=1}^{T} \sin \lambda_j t, \quad -\pi \le \lambda \le \pi; \end{aligned} \tag{4.21}$$

$$R_j^{*2} = A_j^{*2}(\lambda) + B_j^{*2}(\lambda), \quad -\pi \le \lambda \le \pi; \tag{4.22}$$

$$\lambda = \frac{2\pi}{T}. \tag{4.23}$$

R_j^{*2} graafikut nimetatakse <u>spektrogrammiks</u>.

Spektraaltiheduse määrame järgmiselt:

$$J^*(\lambda) = \frac{T}{8\pi} R^{*2}(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} (y_t - \bar{y}) e^{i\lambda t} \right|^2, \tag{4.24}$$

kusjuures

$$J^*(\lambda) = J^*(-\lambda).$$

Võib näidata, et $E J(\lambda) = f(\lambda)$, kui $T \to \infty$, s.t. küllalt pika aegrea korral väljavõtukogumi põhjal saadud spektraaltiheduse hinnangud lähenevad spektraaltiheduse teoreetilistele väärtustele. Seejuures aga $J(\lambda)$ dispersioon ei lähene nullile, mistõttu väljavõttelist spektraaltihedust ei saa lugeda rahuldavaks hinnanguks. Paremaid tulemusi võib saada siis, kui spektraaltiheduse määramisel võtta appi kaalufunktsioon $\omega(\lambda)$. Kui on vaja hinnata $f(\nu)$ väärtust, siis on soovitav, et $\omega(\lambda)$ graafikul oleks punktis $\lambda = \nu$ "tipp".

Kui $f(0)$ väärtust hinnata avaldisega

$$\hat{f}(0) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \omega_\tau C_\tau, \quad (\omega_\tau = \omega_{-\tau}), \tag{4.25}$$

siis

$$\hat{f}(\nu) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \omega_\tau \cos \nu\tau \, C_\tau. \tag{4.26}$$

Kaalufunktsioonid on

$$\omega(\lambda|0) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \omega_\tau \cos \lambda\tau, \tag{4.27}$$

$$\omega(\lambda|\nu) = \frac{1}{2\pi} = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \omega_\tau \cos \lambda\tau \cos \nu\tau =$$

$$= \frac{1}{2}\left[\omega(\lambda-\nu|0) + \omega(\lambda+\nu|0)\right]. \tag{4.28}$$

Tavaliselt nõutakse, et

$$\int_{-\pi}^{\pi} \omega(\lambda|\nu) d\lambda = 1, \tag{4.29}$$

millest $\omega_0 = 1$.

Kaalufunktsiooni $\omega(\lambda|\nu)$ nimetatakse tihti "aknaks".

4.4. MÕNINGAID NÄITEID KAALUFUNKTSIOONI KOHTA

Paljudel juhtudel võetakse kaalumisel appi funktsioonid

$$h_T(\lambda) = \frac{\sin^2 \frac{1}{2}\lambda T}{2\pi T \sin^2 \frac{1}{2}\lambda}, \tag{4.30}$$

$$h_{2T-1}(\lambda) = \frac{\sin \frac{1}{2}\lambda(2T-1)}{2\pi \sin \frac{1}{2}\lambda}, \tag{4.31}$$

$$\omega_\tau = 1 - \frac{|\tau|}{T}, \quad \tau = 0, \pm 1, \ldots, \pm(T-1) \tag{4.32}$$

ja

$$\omega_r^* = 1, \qquad r = 0, \pm 1, \ldots, \pm (T-1). \tag{4.33}$$

1. $$\hat{f}(\nu) = \frac{1}{2\pi} \sum_{r=-k}^{k} \cos \nu r\, C_r =$$

$$= \frac{1}{2\pi} \sum_{r=-k}^{k} \left(1 - \frac{|r|}{T}\right) \cos \nu r\, C_r, \tag{4.34}$$

$k \leq T-1$ on täisarv.

Aknad on siinjuures järgmised:

$$\omega(\lambda|0) = \frac{k+1}{T} h_{k+1}(\lambda) + \left(1 - \frac{k+1}{T}\right) h_{2k+1}(\lambda), \tag{4.35}$$

$$\omega^*(\lambda|0) = h_{2k+1}(\lambda). \tag{4.36}$$

Akna (4.35) maksimum asub punktis $\lambda = 0$ **ja**

$$\max \omega(\lambda|0) = \frac{2k+1}{2\pi} - \frac{k(k+1)}{2\pi T}. \tag{4.37}$$

Akna (4.36) maksimum on samuti punktis $\lambda = 0$ **:**

$$\max \omega^*(\lambda|0) = \frac{2k+1}{2\pi}. \tag{4.38}$$

2. Bartlett'i aken

$$\hat{f}(\nu) = \frac{1}{2\pi} \sum_{r=-k}^{k} \left(1 - \frac{|r|}{k}\right) \cos \nu r\, C_r =$$

$$= \frac{1}{2\pi} \sum_{r=-k}^{k} \frac{1 - \frac{|r|}{k}}{1 - \frac{|r|}{T}} \cos \nu r\, c_r. \tag{4.39}$$

Siin $\omega_r = 1 - \frac{|r|}{k}$, kui $r = 0, \pm 1, \ldots, \pm k$ ja $\omega_r = 0$ ülejäänud juhtudel.

$$\omega_r^* = \frac{1 - |r|/k}{1 - |r|/T},$$

kui

Aknad on vastavalt -

$$\omega(\lambda|0) = \frac{\sin^2(k\lambda/2)}{2\pi k \sin^2(\lambda/2)}, \tag{4.40}$$

$$\omega^*(\lambda|0) = \frac{T}{k}\left\{h_{2k+1}(\lambda) - \frac{T-k}{2\pi}\sum_{r=-k}^{k}\frac{\cos\lambda r}{T-|r|}\right\}. \tag{4.41}$$

3. Parzeni hinnang

$$\hat{f}(\nu) = \frac{1}{2\pi}\sum_{r=-k}^{k}\left(1-\frac{r^2}{k^2}\right)\left(1-\frac{|r|}{T}\right)\cos\nu r\, C_r =$$

$$= \frac{1}{2\pi}\sum_{r=-k}^{k}\left(1-\frac{r^2}{k^2}\right)\cos\nu r\, c_r.$$

$$\omega^*(\lambda|0) = \frac{1}{2\pi}\left\{\frac{\sin\lambda k\cos(\lambda/2)}{2k^2\sin^3(\lambda/2)} - \frac{\cos\lambda k}{k\sin^2(\lambda/2)}\right\}. \tag{4.42}$$

Selle akna maksimum on punktis $\lambda = 0$,

$$\max\omega^*(\lambda|0) = \frac{1}{2\pi}\left(4\frac{k}{3} - \frac{1}{3k}\right) \tag{4.43}$$

S i s u k o r d

Яан В а й н у.
КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ
ВРЕМЕННЫХ РЯДОВ.
Учебное пособие для студентов экономического
факультета.
На эстонском языке.
Тартуский государственный университет.
ЭССР, 202400, г.Тарту, ул.Юликооли, 18.
Vastutav toimetaja V. Tamm.
Paljundamisele antud 17.08.1983.
Formaat 60x84/16.
Kirjutuspaber.
Masinakiri. Rotaprint.
Tingtrükipoognaid 3,02.
Arvestuspoognaid 2,86. Trükipoognaid 3,25.
Trükiarv 300.
Tell. nr. 821.
Hind 10 kop.
TRÜ trükikoda. ENSV, 202400 Tartu, Pälsoni t. 14.
MB 08407.