Praktični aspekti modeliranja stacionarnih vremenskih serija

Zorica Mladenović

BOKS-DŽENKINSOVA STRATEGIJA MODELIRANJA

- Autori: Box i Jenkins, 1976
- Polazna osnova: ARMA model predstavlja odgovarajući okvir modeliranja

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}$$

- Cilj modeliranja: izbor ARMA modela koji na zadovoljavajući način opisuje kretanje konkretnog skupa podataka vremenske serije.
- 'Sofisticirana ekstrapolacija'
- Box: „Svi modeli su pogrešni, samo su neki korisni”

BOKS-DŽENKINSOVA STRATEGIJA MODELIRANJA II

- Pristup se sastoji od tri faze:
  - identifikacija modela
  - ocena parametara modela i
  - provera adekvatnosti modela.
- Iterativna procedura

I Faza identifikacije modela

- Potrebno je izabrati užu klasu ARMA modela za koju pretpostavljamo da predstavlja potencijalni generator skupa podataka.
- Ključno pitanje:
  Koliki je red autoregresione i
  komponente pokretnih proseka?
- Ključni princip:
  Jednostavnost (ekonomičnost)
- Ključni metodološki okvir:
  Analiza obične i parcijalne
  autokorelacione funkcije

Model	Obična autokorelaciona funkcija	Parcijalna autokorelaciona funkcija
AR(p)	Vrednosti opadaju tokom vremena po eksponencijalnoj, eksponencijalno oscilatornoj ili sinusoidnoj putanji	$\phi_{11}\neq 0, \phi_{22}\neq 0,..., \phi_{pp}=\phi_p$ $\phi_{kk}=0$ za k>p.
MA(q)	$\rho_1\neq 0, \rho_2\neq 0,..., \rho_q\neq 0$, $\rho_k=0$ za k>q.	Vrednosti opadaju tokom vremena po eksponencijalnoj, eksponencijalno oscilatornoj ili sinusoidnoj putanji
ARMA(p,q)	Vrednosti opadaju tokom vremena. Prvih q koeficijenata je određeno parametrima AR i MA komponente, dok se za docnje veće od q koeficijenti ponašaju kao kod AR modela.	Vrednosti opadaju tokom vremena. Prvih p koeficijenata je određeno parametrima AR i MA komponente. Za docnje veće od p koeficijenti slede putanju sličnu kao kod MA modela.

II Faza ocene parametara modela

- Metod običnih najmanjih kvadrata se može koristiti u oceni parametara AR modela.
- Za ocenu parametara MA i ARMA modela koristi se metod nelinearnih najmanjih kvadrata koji se zasniva na upotrebi metoda numeričke optimizacije.

III Faza provere adekvatnosti modela

1. Da li je model saglasan sa podacima?
   Da li su reziduali normalno raspodeljeni i neautokorelisani?

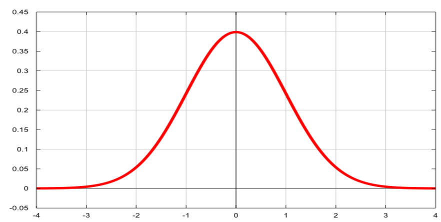
2. Da li je izbor AR i MA komponente optimalan?
   Da li je model istovremeno ekonomičan i dovoljno precizan?

III-1. Analiza reziduala

- Normalnost
- Autokorelacija

Testiranje normalnosti reziduala u ocenjenom modelu

- Uobičajena pretpostavka: slučajna greška ima normalnu raspodelu, $e \sim N(0,\sigma^2)$
- Zbirno dejstvo velikog broja sporadičnih i nesistematičnih faktora opravdano je modelirati normalnom raspodelom. Otuda i pretpostavka.



Šta ako slučajna greška nema normalnu raspodelu?

- Ukoliko je samo ova pretpostavka narušena, tada se dobijaju pouzdane ocene.
- Međutim,
  - Testiranje hipoteza je nepouzdano.
  - Vrednosti t-odnosa i F-odnosa su netačne.
  - Verovatno postoji greška u specifikaciji modela.
- Zaključak:
  - Postupak statističkog zaključivanja je pogrešan.
  - Pretpostavka je vitalni deo specifikacije modela.

Provera validnosti pretpostavke da slučajna greška ima normalnu raspodelu

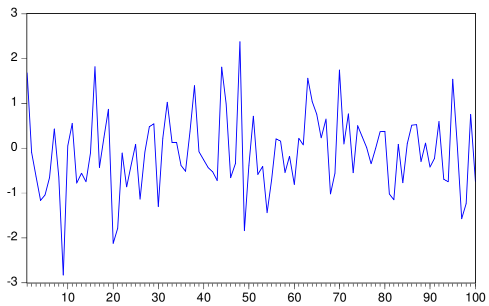
- *Neformalni* (grafički) pristup
- *Formalni* pristup

Neformalni (grafički) pristup

- ***Histogram***
- Grafički prikaz učestalosti pojavljivanja podataka (reziduala) u pojedinim grupnim intervalima
  - x-osa: celokupni raspon vrednosti date promenljive deli se na određeni broj podintervala jednake širine
  - y-osa: broj pojavljivanja podataka u svakom od podintervala

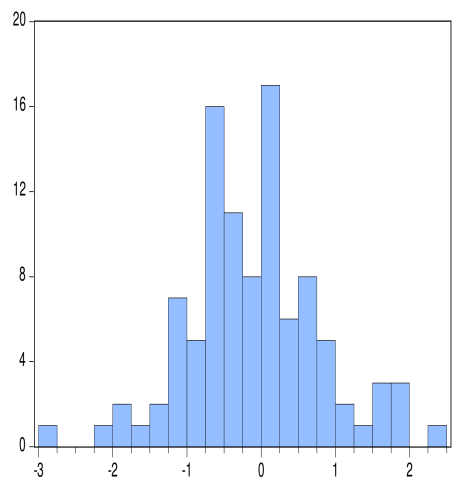
Histogram: primer za 100 podataka koji su predstavljeni tabelarno i grafički

1	2	3	4	5
1.68	-0.09	-0.63	-1.16	-1.04
-0.66	0.43	-0.64	-2.83	0.06
0.55	-0.78	-0.55	-0.75	-0.10
1.82	-0.43	0.24	0.87	-2.12
-1.79	-0.10	-0.86	-0.37	0.09
-1.14	-0.08	0.49	0.54	-1.30
0.22	1.03	0.13	0.13	-0.38
-0.51	0.37	1.39	-0.08	-0.26
-0.43	-0.52	-0.72	1.81	0.99
-0.66	-0.34	2.38	-1.84	-0.38
0.72	-0.58	-0.40	-1.44	-0.70
0.21	0.16	-0.54	-0.17	-0.80
0.22	0.08	1.56	1.04	0.76
0.23	0.65	-1.02	-0.55	1.75
0.10	0.77	-0.54	0.51	0.26
0.02	-0.35	-0.00	0.37	0.38
-1.02	-1.15	0.09	-0.77	0.11
0.52	0.53	-0.29	0.12	-0.42
-0.23	0.60	-0.69	-0.75	1.54
0.10	-1.57	-1.23	0.75	-0.75



Histogram: vizuelni i tabelarni pregled frekvencije podataka

Interval		Učestalost
-3.000000	-2.750000	1.000000
-2.750000	-2.500000	0.000000
-2.500000	-2.250000	0.000000
-2.250000	-2.000000	1.000000
-2.000000	-1.750000	2.000000
-1.750000	-1.500000	1.000000
-1.500000	-1.250000	2.000000
-1.250000	-1.000000	7.000000
-1.000000	-0.750000	5.000000
-0.750000	-0.500000	16.00000
-0.500000	-0.250000	11.00000
-0.250000	0.000000	8.000000
0.000000	0.250000	17.00000
0.250000	0.500000	6.000000
0.500000	0.750000	8.000000
0.750000	1.000000	5.000000
1.000000	1.250000	2.000000
1.250000	1.500000	1.000000
1.500000	1.750000	3.000000
1.750000	2.000000	3.000000
2.000000	2.250000	0.000000
2.250000	2.500000	1.000000



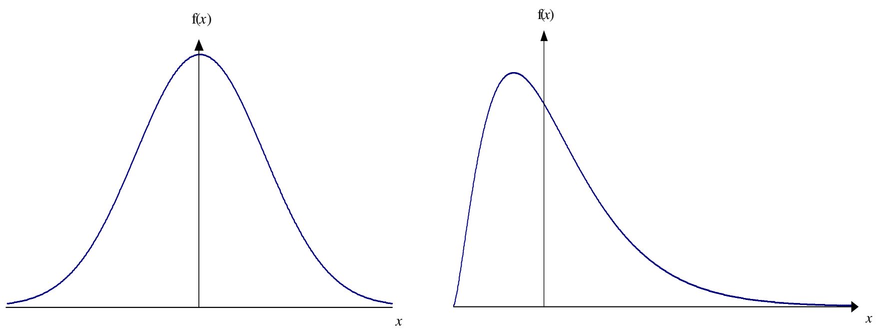
Formalni pristup

- Primena različitih test-statistika
- Najpopularniji test normalnosti je **Žark-Bera** (engl. Jarque-Bera) test
  - Oznaka: JB
  - Zasniva se na ocenama koeficijenata kojima se opisuju svojstva raspodele

Koeficijenti za deskripciju raspodele

- Empirijska raspodela često se opisuje na osnovu dva koeficijenta: *asimetrije* i *spljoštenosti.*
- Koeficijent asimetrije meri stepen u kojem raspodela nije simetrična u odnosu na srednju vrednost.
- Raspodela može biti
  - Simetrična
  - Asimetrična
    - Ulevo (negativna)
    - Udesno (pozitivna)

Simetrična i asimetrična raspodela



Koeficijent asimetrije (engl. *skewness*)
Oznaka koeficijenta: α_3

Vrednost koeficijenta α_3	Tip raspodele
Nula	Simetrična
Veća od nule	Asimetrična udesno
Manja od nule	Asimetrična ulevo

Koeficijenti za deskripciju raspodele II

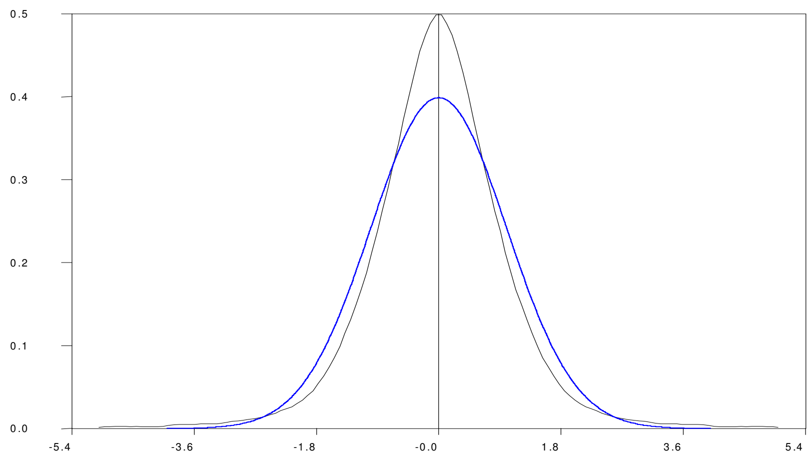
- Koeficijent spljoštenosti meri težinu (debljinu) repova raspodele

  Svojstva repova raspodele iskazuju se u odnosu na normalnu raspodelu
- Repovi raspodele mogu biti
  - *Iste težine* kao kod normalne raspodele
  - *Teži* od repova normalne raspodele
  - *Lakši* od repova normalne raspodele

Koeficijent spljoštenosti (engl. *kurtosis*) Oznaka koeficijenta: α_4

Vrednost koeficijenta α_4	**Težina repova**
Tri	Odgovara normalnoj raspodeli
Veća od tri	Veći deo jedinične verovatnoće je pod repovima nego kod repova *N* raspodele (prisustvo ekstremnih opservacija)
Manja od tri	Manji deo jedinične verovatnoće je pod repovima nego kod repova *N* raspodele

Normalno spljoštena raspodela (plavi grafik) $\alpha_4 = 3$
i raspodela koja ima teže repove od normalne $\alpha_4 > 3$



Koeficijent asimetrije	Koeficijent spljoštenosti
$\alpha_3 = 0$ *za N raspodelu*	$\alpha_4 = 3$ *za N raspodelu*
Ocena $\hat{\alpha}_3 = \dfrac{\frac{\sum e_t^3}{T}}{\hat{\sigma}^3}$	Ocena $\hat{\alpha}_4 = \dfrac{\frac{\sum e_t^4}{T}}{\hat{\sigma}^4}$
$\hat{\alpha}_3 : N\left(0, \frac{6}{T}\right)$	$\hat{\alpha}_4 : N\left(3, \frac{24}{T}\right)$
$\sqrt{\frac{T}{6}}\hat{\alpha}_3 : N(0,1)$	$\sqrt{\frac{T}{24}}(\hat{\alpha}_4 - 3) : N(0,1)$

Nulta hipoteza: data raspodela je normalna, odnosno vrednosti koeficijenata asimetrije i spljoštenosti su redom 0 i 3.

Alternativna hipoteza: data raspodela nije normalna.

Statistika:

$$JB = \frac{T}{6}\left[\hat{\alpha}_3^2 + \frac{(\hat{\alpha}_4 - 3)^2}{4}\right] : \chi_2^2$$

Odgovarajuća kritična vrednost na nivou značajnosti 5% je 5.99.

Primer za 100 podataka:

1. Ocena koeficijenta asimetrije 0.1

2. Ocena koeficijenta spljoštenosti 3.65

3. Vrednost JB testa 1.93(0.38)

Mean	-0.101697
Maximum	2.378484
Minimum	-2.830084
Skewness	0.100453
Kurtosis	3.651004
Jarque-Bera	1.934038
Probability	0.380215
Observations	100

Šta raditi u slučaju da raspodela odstupa od normalne?

- Ne postoji jedinstveno rešenje zato što je odstupanje od normalnosti posledica pogrešne specifikacije modela.
- Često se modifikuje polazna specifikacija uključivanjem promenljivih kojima se eksplicitno modeliraju ekstremni događaji.
  - Takve promenljive se nazivaju veštačke promenljive.

Testiranje autokorelacije reziduala u ocenjenom modelu

- Da li postoji autokorelacije na određenoj, k-toj, docnji?
  (H_0: $\rho_k=0$)
- Da li postoji autokorelacija na svim docnjama do **K**-te?
  (H_0: $\rho_1= \rho_2 =...= \rho_{\mathbf{K}} =0$).

Da li postoji autokorelacije na određenoj, k-toj, docnji? (H_0: $\rho_k=0$)

Validnost nulte hipoteze H_0: $\rho_k=0$ protiv alternativne H_1: $\rho_k\neq0$ se testira tako što se proverava da li ocena autokorelacionog koeficijenta na docnji k serije reziduala, pripada intervalu $(-1.96/\sqrt{T}, 1.96/\sqrt{T})$.

Da li postoji autokorelacija
na svim docnjama do **K**-te?
(H_0: $\rho_1 = \rho_2 = ... = \rho_K = 0$).

- Validnost nulte hipoteze H_0: $\rho_1 = \rho_1 = ... = \rho_K = 0$ se testira protiv alternativne da je bar jedan od prvih **K** autokorelacionih koeficijenata serije reziduala različit od nule.
- Boks-Pirsova i Boks-Ljungova statistika.
- Boks-Ljungova statistika:

$$Q(K) = BLj(K) = T(T+2)\sum_{i=1}^{K} \frac{\hat{\rho}_i^{\,2}}{T-i} : \chi^2_{K-p-q}$$

Šta raditi u slučaju da autokorelacija postoji u rezidualima?

- To je znak da ARMA model nije dobro postavljen, odnosno da odabir objašnjavajućih promenljivih nije adekvatan.
- Neoophodno je redefinisati izbor AR i MA komponenti u skladu sa evidentiranim tipom (redom) autokorelacije i potom oceniti novi model.

III-2. Optimalan izbor parametara modela (Informacioni kriterijum)

Informacioni kriterijum predstavlja zbir dva elementa

1. Element koji je funkcija neobjašnjenog varijabiliteta modela
2. Element kojim se sankcioniše gubitak u broju stepeni slobode zbog povećanja broja parametara za ocenjivanje

$$IC(p,q) = ln\, s^2 + g\,\frac{p+q}{T}$$

g je nenegativna kaznena funkcija

s^2 je ocena varijanse slučajne greške modela

Informacioni kriterijum

- Sabirci u informacionoj funkciji različito reaguju na povećanje *p* i *q:*
  - Ocena varijanse sl. greške modela se smanjuje
  - Kaznena komponenta se povećava
- Cilj: izbor kombinacije *p* i *q* koja minimizira vrednost informacionog kriterijuma

$$IC(p,q) = ln\, s^2 + g\,\frac{p+q}{T}$$

Informacioni kriterijum (II)

Funkcija g	**Kaznena komponenta**	**Naziv**	**Oznaka**
2	2(p+q)/T	Akaikeov	AIC
lnT	(lnT)(p+q)/T	Švarcov	SC / SIC
2ln(lnT)	2(ln(lnT))* (p+q)/T	Hana-Kvinov	HQC / HQIC