Ekonometrija – I deo

Doktorske studije

Predavač: Aleksandra Nojković

Beograd, školska 2024/25

Struktura predavanja

- **Narušavanje pretpostavki KLRM**
  - Heteroskedasticnost
  - Autokorelacija

- **Specifikacija modela**

Pretpostavke KVLRM

1. $E(\varepsilon_i) = 0$

2. $\boldsymbol{Var(\varepsilon_i) = \sigma^2 < \infty}$

3. Cov $\boldsymbol{(\varepsilon_i, \varepsilon_j) = 0}$ **za *i* različito od *j***

4.Objašnjavajuće promenljive nisu određene stohastičkim članom

5. $\varepsilon_i \sim N(0, \sigma^2)$

6. Ne postoji tačna linearna zavisnost između objašnjavajućih promenljivih.

Šta ako su pretpostavke KVLRM narušene?

- Kada dolazi do narušavanja pretpostavki?
- Kako se to odražava na ocene parametara i na standardne greške ocena?
- Kako se ispituje da li su pretpostavke narušene ili ne?
- Šta raditi u slučaju kada su pretpostavke narušene?

Pretpostavka 2: Var$(\varepsilon_i) = \sigma^2 < \infty$
Homoskedastičnost

- Homoskedastičnost: varijansa slučajne greške modela je konstantna za sve opservacije.

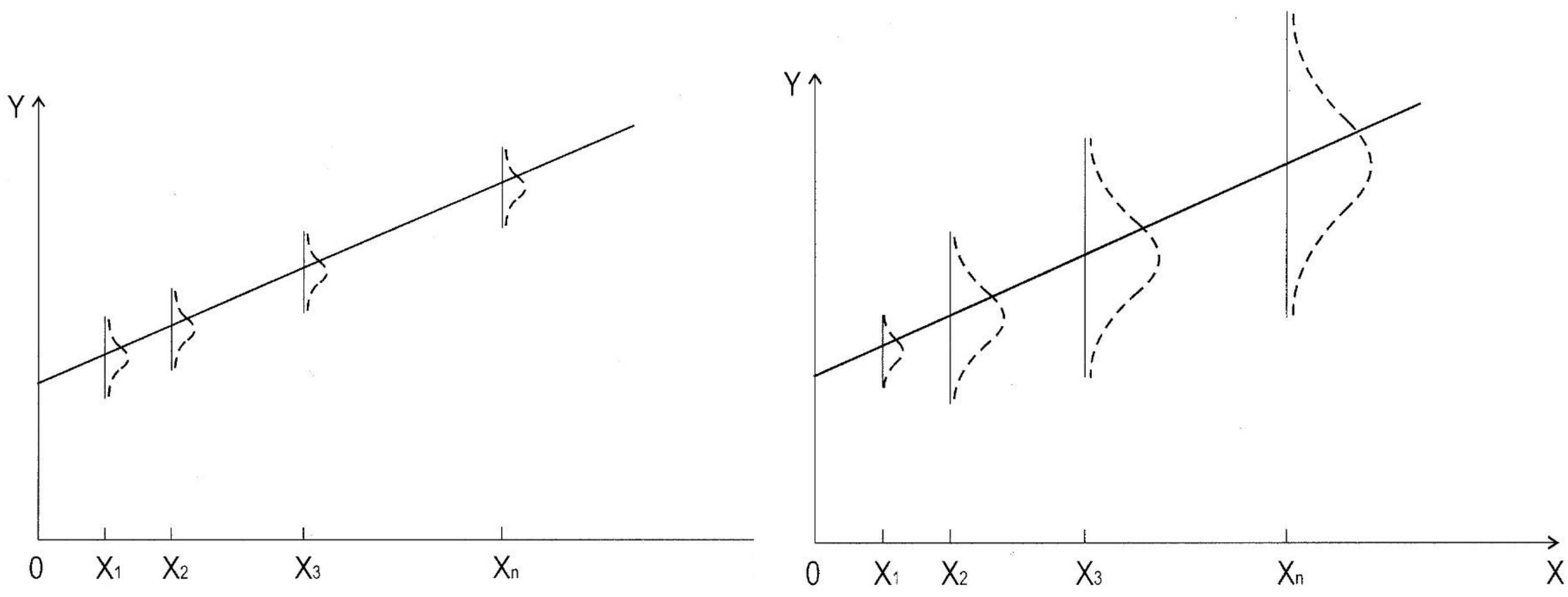
$$\mathrm{var}(\varepsilon_1) = \mathrm{var}(\varepsilon_2) = \ldots = \mathrm{var}(\varepsilon_n) = \sigma^2 = \mathrm{const}$$

- Heteroskedastičnost:

pretpostavka o homoskedastičnosti je narušena, što znači da se varijanse slučajnih greški razlikuju po pojedinim opservacijama:

$$\left.\begin{array}{l} \mathrm{var}(\varepsilon_1) = \sigma_1^2 \\ \mathrm{var}(\varepsilon_2) = \sigma_2^2 \\ \vdots \\ \mathrm{var}(\varepsilon_n) = \sigma_n^2 \end{array}\right\} \sigma_1^2 \neq \sigma_2^2 \neq \ldots \neq \sigma_n^2$$

Homosedastične (levo) i heteroskedastične (desno) greške



Posledice primene metoda ONK u prisustvu heteroskedastičnosti

- Primenom metoda ONK na model sa heteroskedastičnim greškama dobijaju se ocene koje nisu najbolje linearne nepristrasne ocene.
  - Ocene su nepristrasne
  - Ocene nisu efikasne–njihova varijansa nije najmanja moguća **(pokazati...).**

- Posledice:
  - Standardne greške ocena nisu precizna mera varijabiliteta ocena.
  - Standardne greške ocena najčešće potcenjuju stvarnu varijansu ocena parametara modela.
  - *t*-odnosi su nepouzdani.

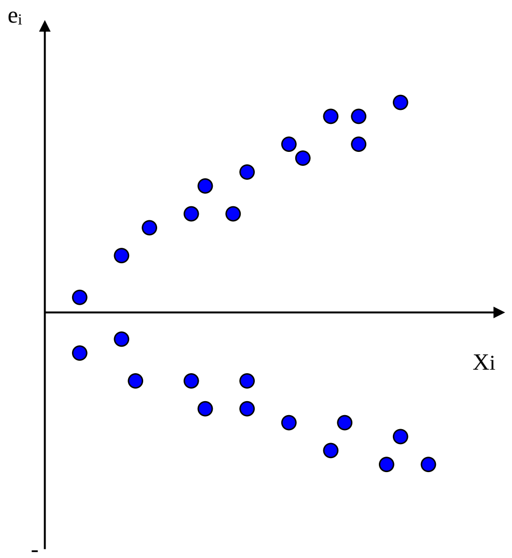
Kako se otkriva prisustvo heteroskedastičnosti u modelu?

1. Neformalni (grafički) metodi

2. Formalni metodi (testiranje)

Neformalni (preliminarni) metodi

Grafički prikazi: dijagram rasturanja reziduala (apsolutne vrednosti reziduala iili njihovih kvadrata) u odnosu na neku od objašnjavajućih promenljivih ili prema ocenjenoj vrednosti Yi (lin. kombinaciija svih objašnjavajućih promenljivih).



Testiranje postojanja heteroskedastičnosti (formalni testovi)

- Goldfeld-Kvant -ov (engl. Goldfeld-Quandt) test
- Glejzeov -ov (engl. Glejser) test
- Vajt-ov (engl. White) test

Goldfeld-Quandtov test

- Algoritam:

1. Pretpostavimo da je polazni model oblika:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i.$$

2. Opservacije poređati prema rastućem redosledu nezavisne promenljive.
3. Izostaviti jedna broj (c) centralnih opservacija (oko četvrtina).
4. Obaviti odvojeno regresije za prvih i poslednjih (n-c)/2 opsrevacija.
5. Statistika testa je: $$F = \frac{\Sigma e_2^2}{\Sigma e_1^2} \sim F_{(n-c-2k)/2}^{(n-c-2k)/2},$$

pri čemu se indeks 1 odnosi na reziduale dobijene za niže vrednosti regresora, a indeks 2 za više.

- Pogodan za modele sa malim brojem param. i velike uzorke.

Glejser-ov test

- Ne zahteva se unapred poznavanje prirode heteroskedastičnosti.
- Algoritam:

1. Iz polazne regresije računaju se reziduali e_i ($\hat{u}_i$)
2. Ocenjiju se sledeće regresije:

$$|e_i| = \delta_0 + \delta_1 X_i^h + \text{greška}.$$

(parametar *h* najčešće:1,-1,1/2 i 2).

3. Testa se statistička značajnost ocene parametra δ_1 primenom t-testa.
4. Upoređuju se koef. determinacije dobijeni za različite vrednosti *h*, a sam karakter heteroskedastičnosti određuje se prema regresiji sa najvećim R^2.

White - ov test

- <u>Osnove testa:</u>

Nulta hipoteza: slucajne greške imaju stabilnu varijansu

Alternativna hipoteza: varijansa slucajne greške je zavisna od objašnjavajucih promenljivih, njihovih kvadrata i međuproizvoda.

- <u>Algoritam:</u>

1. Pretpostavimo da je polazni model oblika:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i.$$

2. Ocenjujemo model iz 1., dobijamo reziduale i potom ocenjujemo pomocnu regresiju:

White-ov test (nastavak)

$$e_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{1i}^2 + \alpha_4 X_{2i}^2 + \alpha_5 X_{1i} X_{2i} + v_i$$

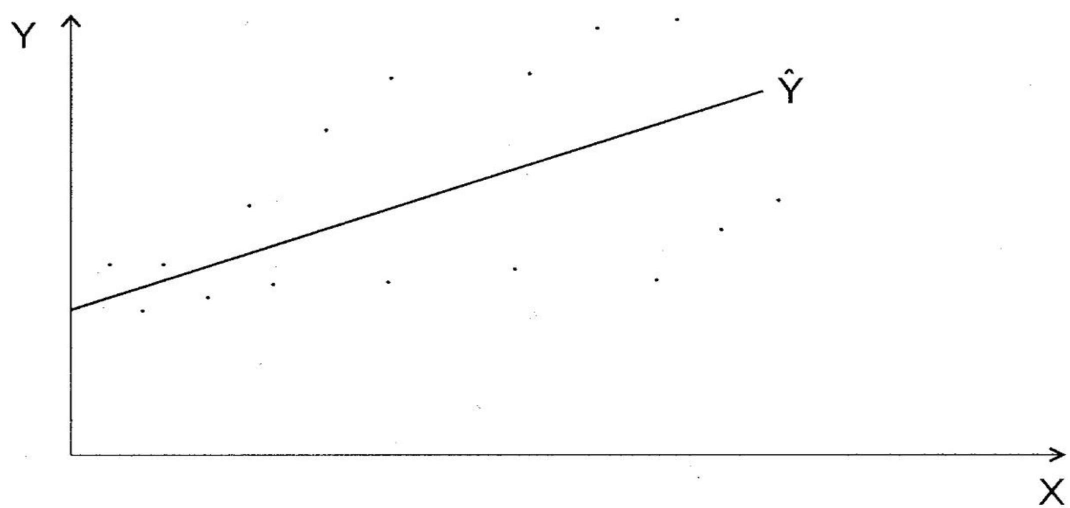
3. Fakticki, nulta hipoteza se svodi na:

$$\mathrm{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_5 = 0$$

4. Odredujemo koeficijent determinacije *R^2 iz pomocne regresije i potom* ga množimo obimom uzorka *n. To je (nR^2) Whiteova test-statistika.* Može se pokazati da pri istinitosti nulte hipoteze važi: $nR^2 \sim \chi^2$ sa *m stepeni slobode* i *m* je broj objašnjavajućih promenljivih pomoćne regresije bez slobodnog člana (*m=5*).
5. Ako je izracunata vrednost test-statistike veca od korespondirajuce kriticne vrednosti χ^2 testa na datom nivou znacajnosti tada se odbacuje nulta hipoteza o odsustvu heteroskedasticnosti.

Kako se eliminiše uticaj heteroskedastičnosti (I) ?

- Primenjuje se metod ponderisanih najmanjih kvadrata (metod uopštenih najmanjih kvadrata).
- Ideja: u postupku minimiziranja sume kvadrata reziduala, onim rezidualima koji su po apsolutnoj vrednosti veći daje se manji ponder i obratno.



Kako se eliminiše uticaj heteroskedastičnosti (II)?

- Pretpostavimo da postoji zavisnost varijanse slučajne greške od objašnjavajuće promenljive x_t

$$\operatorname{var}(\varepsilon_i) = kx_i^2,\ \mathrm{k} = \mathrm{const}$$

- Sve promenljive modela delimo sa merom varijabiliteta, x_t

$$\frac{y_i}{x_i} = b_0 \frac{1}{x_i} + b_1 \frac{x_i}{x_i} + \frac{\varepsilon_i}{x_i}$$

- U ovom modelu nova slučajna greška je $\frac{\varepsilon_i}{x_i}$.

  Njena varijansa je stabilna:

$$\operatorname{var}\left(\frac{\varepsilon_i}{x_i}\right) = \frac{\operatorname{var}(\varepsilon_i)}{x_i^2} = \frac{kx_i^2}{x_i^2} = k = \mathrm{const.}$$

Alternativni pristupi eliminisanja efekata heteroskedastičnosti

1. Koristimo **logaritmovane vrednosti** podataka.

2. Prilikom računanja standardnih grešaka ocena **pravimo korekciju** koju je predložio Vajt (engl. White). Na ovaj nacin dobijaju se standardne greške ocena koje su vece od standardnih grešaka ocena po metodu ONK. <u>Ovo je najzastupljeniji pristup u empirijskoj analizi poslednjih godina.</u>

Pretpostavka 3: Cov $(\varepsilon_i , \varepsilon_j) = 0$ za $i \neq j$ Odsustvo autokorelacije

- Odsustvo autokorelacije: slučajne greške su nekorelisane
  - Cov $(\varepsilon_i , \varepsilon_j) = 0$ za $i \neq j$

  Nema pravilnosti u korelacionoj strukturi slučajnih greški.
- Postoji autokorelacija: slučajne greške za različite opservacije (a najčešće uređene tokom vremena su korelisane):
  - Cov $(\varepsilon_i , \varepsilon_j) \neq 0$ za $i \neq j$

  Slučajne greške slede prepoznatljiv obrazac u kretanju.
- Najčešća se javlja u analizi vremenskih serija:
  - Cov $(\varepsilon_t , \varepsilon_{t-s}) \neq 0$ za $s=1,2,\ldots$

Zašto se javlja autokorelacija?

1. Trajni efekat egzogenih šokova na kretanje ekonomskih vremenskih serija
   - Primer: obustava rada i ocenjivanje zavisnosti ostvarene proizvodnje od količine uloženog rada.
2. Inercija u kretanju ekonomskih veličina.
3. Modifikacija polaznih podataka
   - Neki kvartalni podaci se dobijaju kao prosek tromesečnih vrednosti.

- Autokorelacija može biti "prava" i "lažna"
  - "Prava": posledica prirode podataka
  - "Lažna": model je pogrešno postavljen.
- Autokorelacija može biti pozitivna ili negativna (koef. korelacije između sukcesivnih vrednosti = autokor. koef. prvog reda, AR(1) šema - **pokazati...**).

Posledice autokorelacije

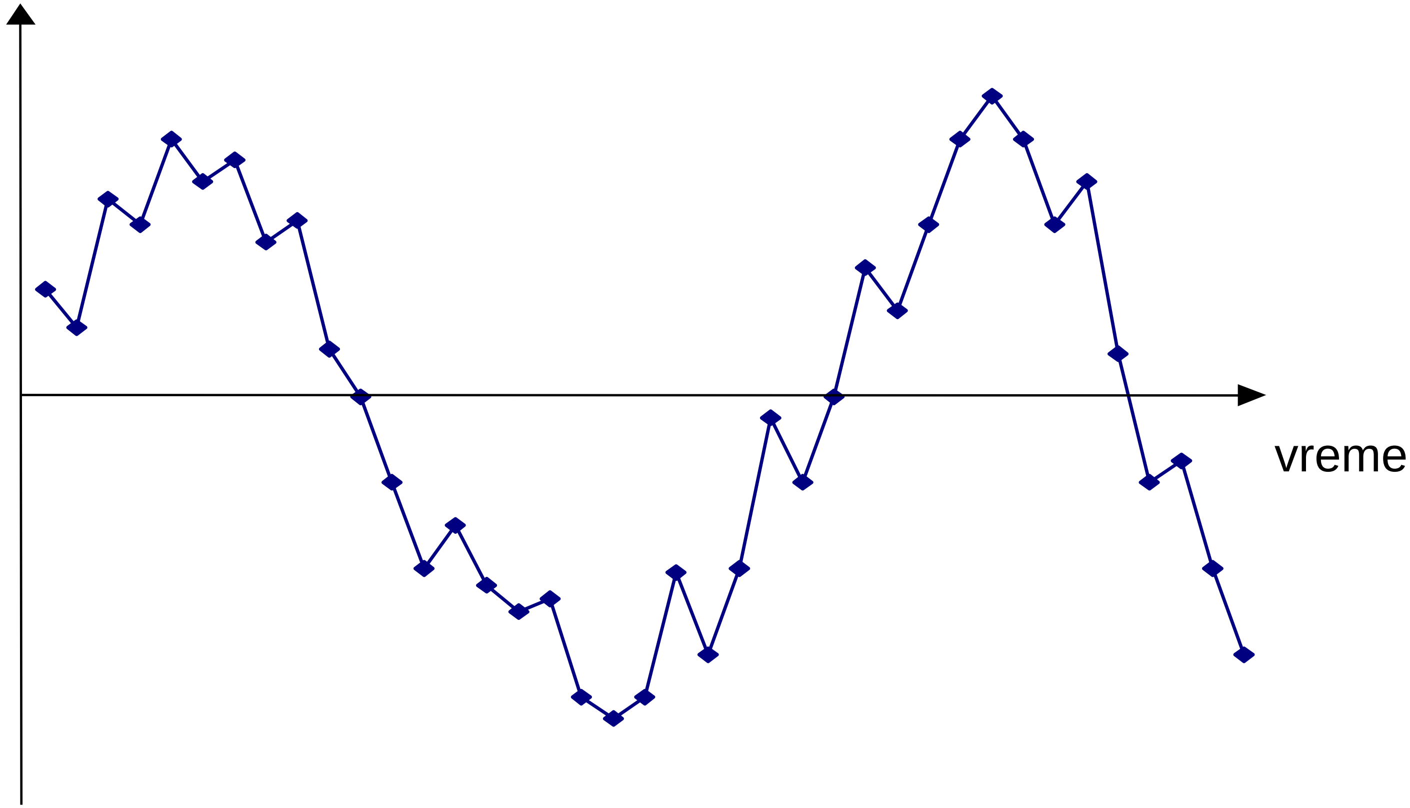
- Ocene ONK su nepristrasne, ali neefikasne.
- Ocena varijanse slučajne greške je pristrasna.
- R^2 nije valjan pokazatelj kvaliteta regresije.
- Rezultati t i F testa su pristrasni i nepouzdani.
- Intervali poverenja su neprecizni.
- Predviđanje je nepouzdano.
- **Pokazati...**

Kako se proverava postojanje autokorelacije?

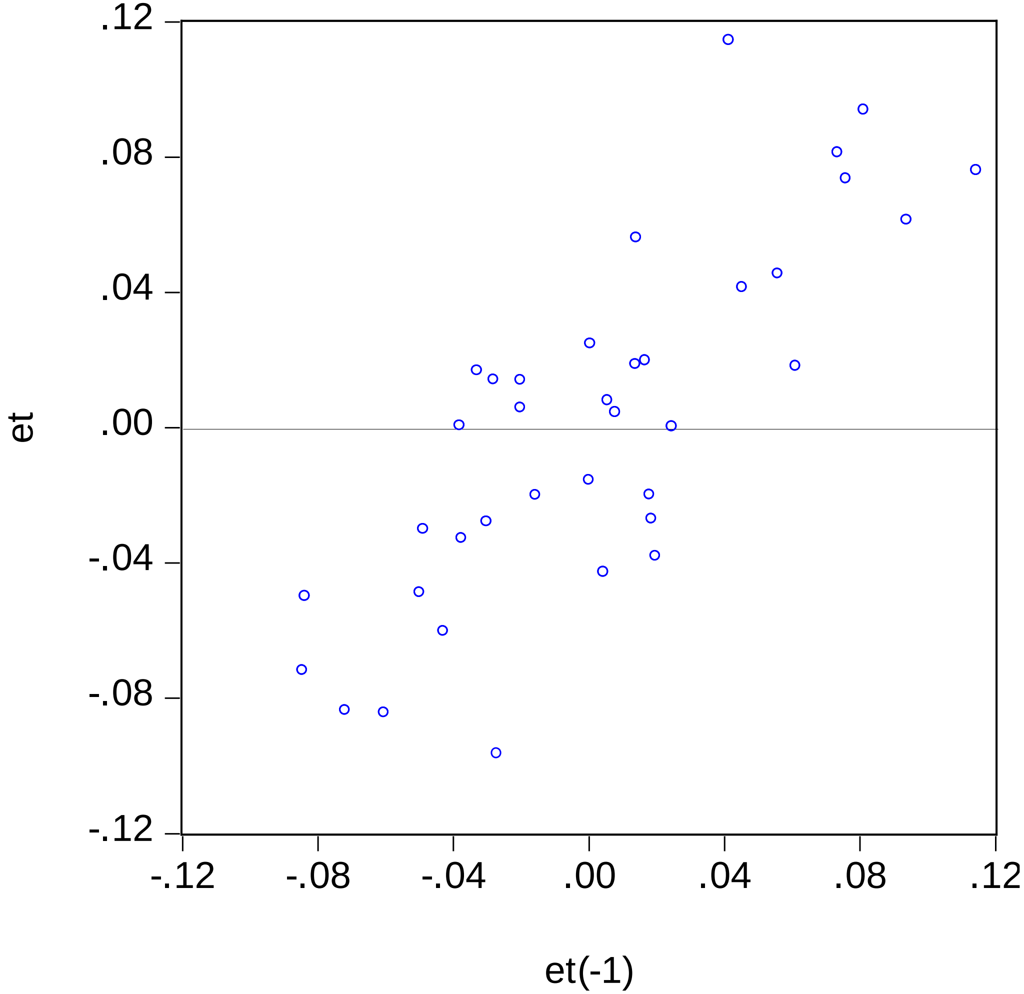
1. Neformalni (grafički) metodi

2. Formalni metodi (testiranje)

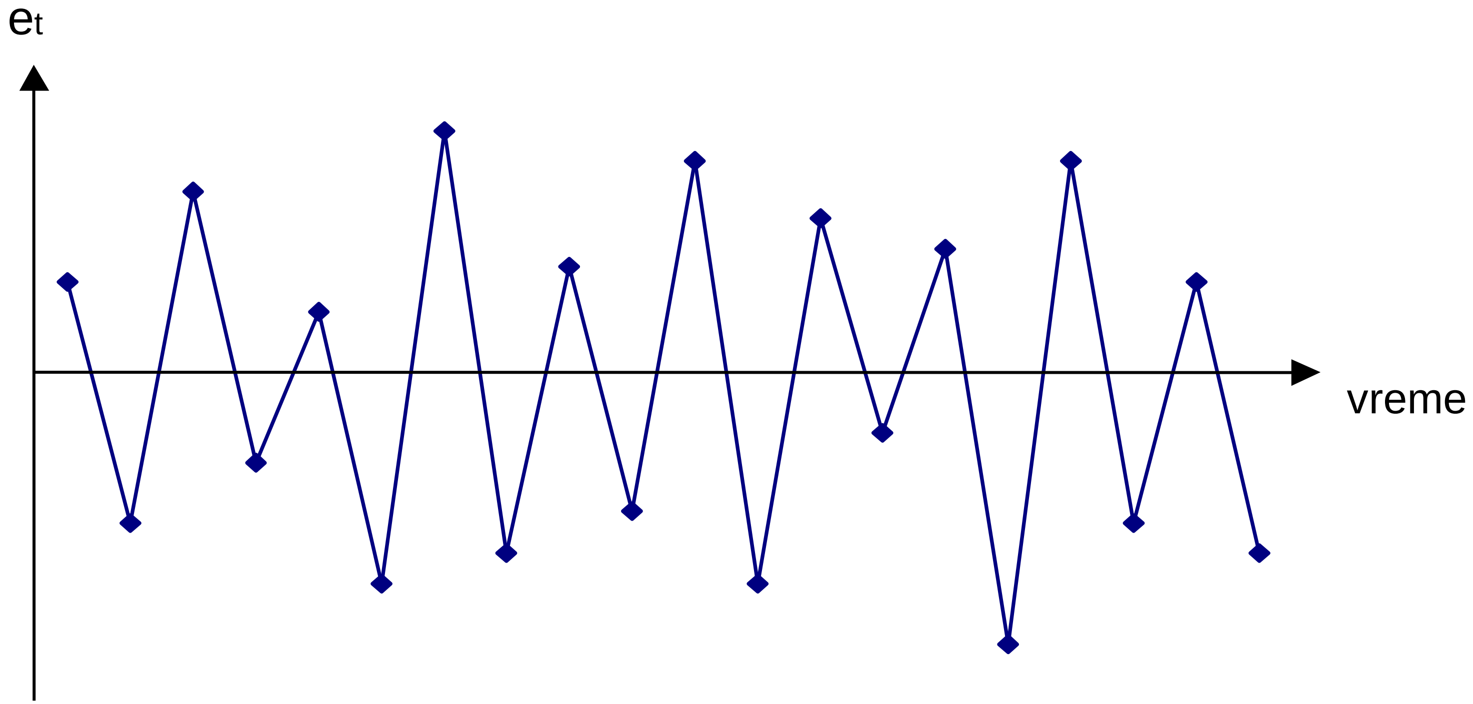
Pozitivna autokorelacija
(reziduali zadržavaju isti znak u nizovima)



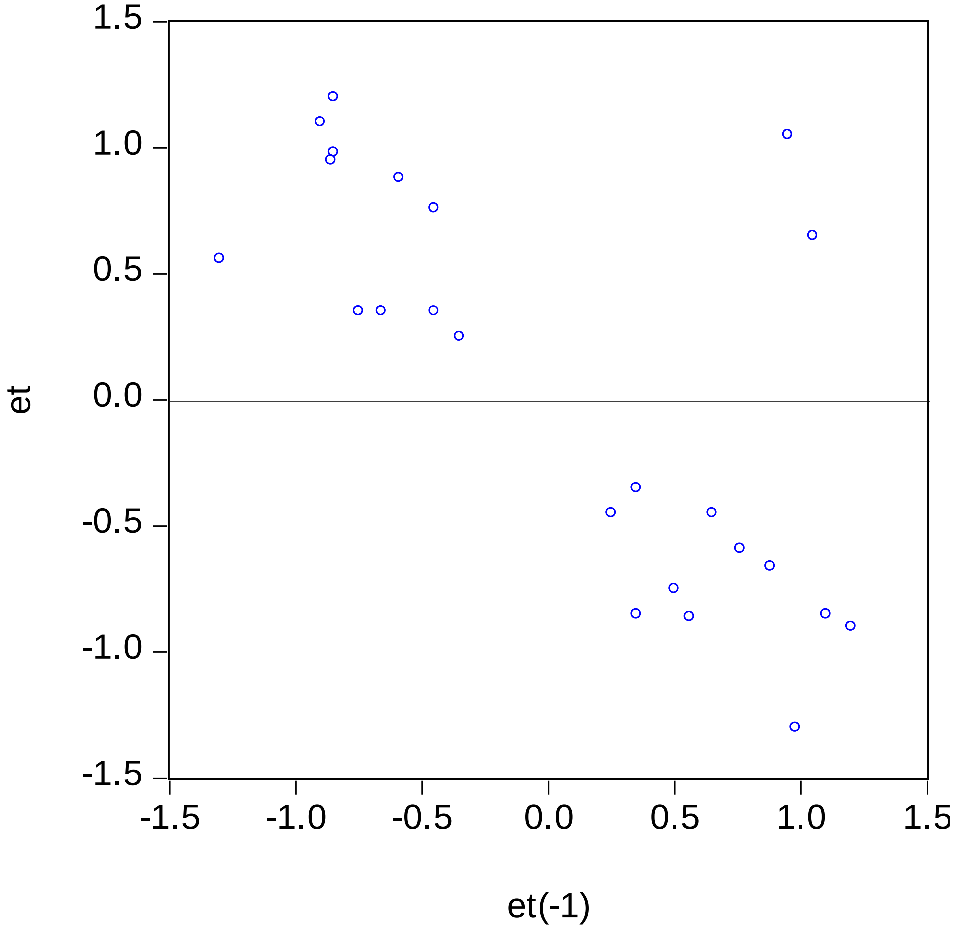
Pozitivna autokorelacija
(reziduali u funkciji sopstvenih prethodnih vrednosti grupisani u I i III kvadrantu)



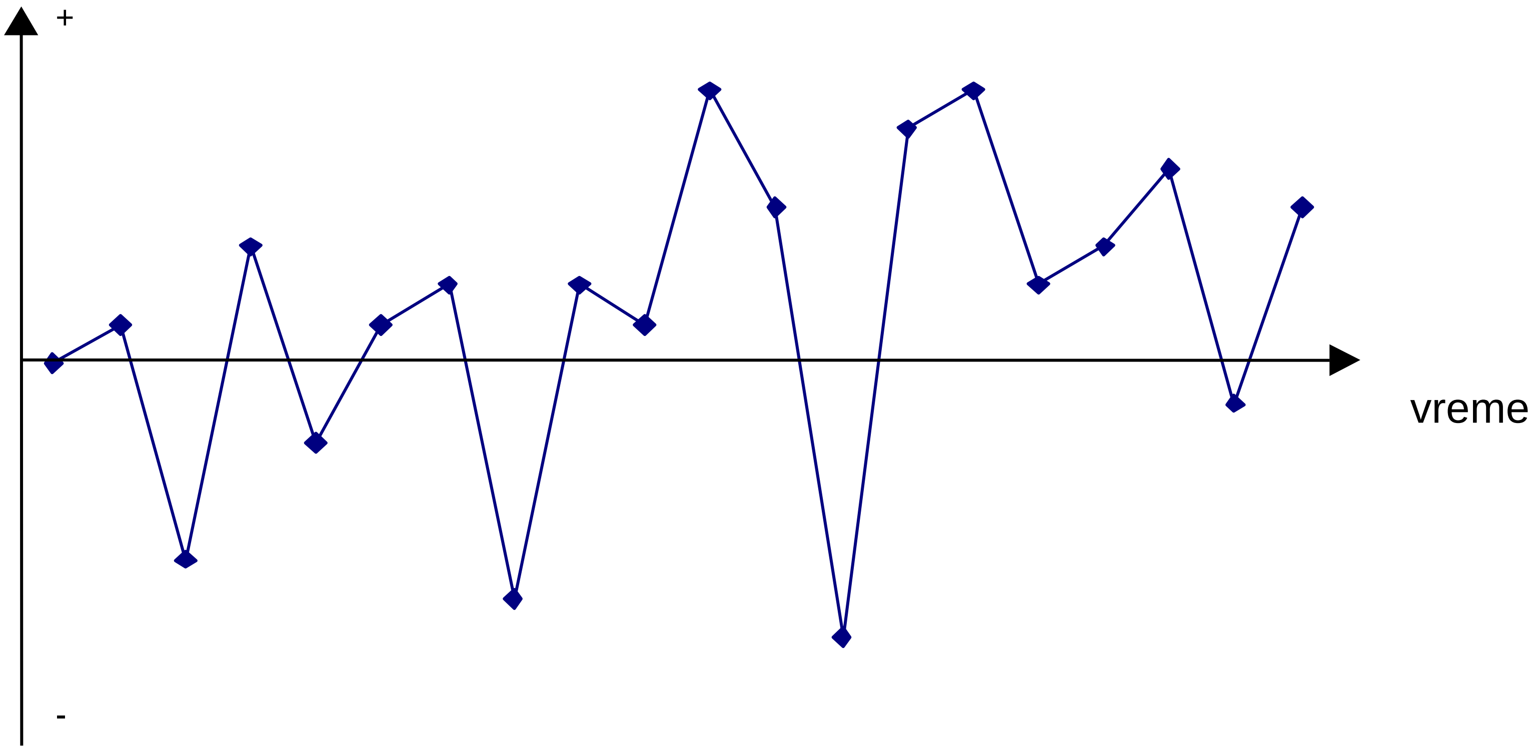
Negativna autokorelacija (reziduali naizmenično menjaju znak)



Negativna autokorelacija (reziduali u funkciji sopstvenih prethodnih vrednosti grupisani u II i IV kvadrantu)



Ne postoji autokorelacija (reziduali ne pokazuju pravilnost promene tokom vremena)



Ispitivanje postojanja autokorelacije: Darbin-Votsonov (engl. Durbin-Watson) test

- Darbin-Votsonov test (oznaka: DW ili d) se koristi za proveru postojanja autokorelacije prvog reda:

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + v_t$$

gde je $v_t \sim N(0, \sigma_v^2)$ i ρ je autokorelacioni koeficijent prvog reda, koji se nalazi u intervalu (-1,+1).

ρ = 0 ne postoji autokorelacija,

ρ = 1, ekstremna pozitivna autokorelacija

ρ = -1, ekstremna negativna autokorelacija

0< *ρ<1, pozitivna autokorelacija*

-1< ρ<0, negativna autokorelacija

Relevantne hipoteze:

H_0 *: ρ*=0 *(nema autokorelacije)*

H_1 *: $\rho \neq$*0 *(postoji autokorelacija prvog reda)*

DW test (II)

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

$$DW \approx 2\left(1 - \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\right)$$

$$DW \approx 2(1 - \hat{\rho})$$

DW test (III)

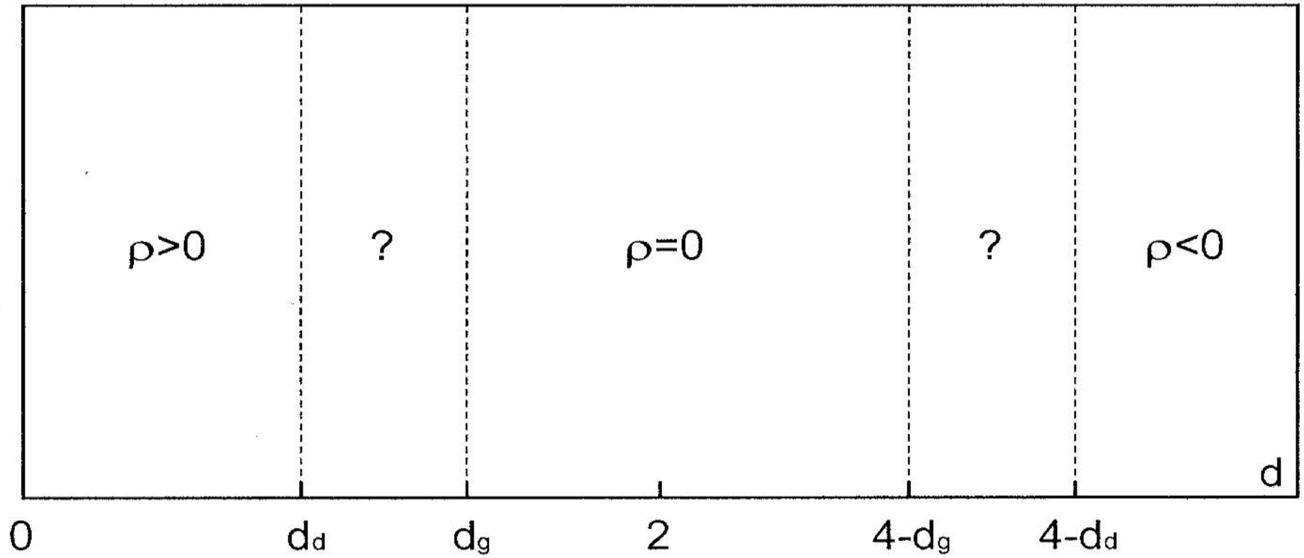
- U postupku testiranja koriste se kritične vrednosti koje su autori testa označili kao donja i gornja kritična vrednost (koje, kao i sama raspodela *sl. prom. d* zavise od podataka nezavisnih promenljivih u uzorku).

- Donja kritična vrednost: *dd*,
- Gornja kritična vrednost: *dg*.

- Kritične vrednosti zavise od obima uzorka i broja objašnjavajućih promenljivih.

- **Objasniti postupak testiranja** …

Primena DW testa



Ograničenja u primeni DW testa

- Ograničenja u primeni:

1. Postoje situacije kada se primenom testa ne može doneti precizan zaključak.
2. Test je definisan samo za model sa slobodnim članom.
3. Testom se <u>ne može proveriti postojanje autokorelacije većeg reda.</u>
4. Test nije pouzdan u situaciji kada se kao objašnjavajuća promenljiva javlja zavisna sa docnjom:

$$y_t = \beta_0 + \beta_1 x_{1t} + \gamma_1 y_{t-1} + \varepsilon_t$$

Opšti test autokorelacije: Brojš-Godfrijev (engl. Breusch-Godfrey) test

- U opštem slucaju autokorelacija može biti reda *m:*

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \ldots + \rho_m \varepsilon_{t-m} + v_t, \quad v_t \sim N(0, \sigma_v^2).$$

- <u>Nulta i alternativna hipoteza</u>

H_0 : $\rho_1 = \rho_2 = \ldots = \rho_m = 0$ *(ne postoji autokorelacija)*

H_1 : bar jedan od parametara je razlicit od nule (postoji autokorelacija do reda m)

- Algoritam testiranja:

1. Pretpostavimo da je polazni model oblika:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t$$

2. Ocenjujemo model iz 1., dobijamo reziduale i potom ocenjujemo pomocnu regresiju:

$$e_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \ldots + \rho_m e_{t-m} + v_t,$$

3. Odredujemo koeficijent determinacije *R^2 iz pomocne regresije i potom ga* množimo obimom uzorka *T. To je ($T R^2$) Brojš-Godfrijeva test-statistika. Može* se pokazati da važi: $T R^2 \sim \chi^2$ sa *m stepeni* slobode, pri uslovu istinitosti nulte hipoteze.

Kako se eliminiše uticaj autokorelacije?

- Korekcija polaznog modela u pravcu transformisanja promenljivih (**pokazati...).**
- Korekcija polaznog modela u pravcu eksplicitnog ukljucivanja dinamike – **dinamicki modeli**.
- **Korekcija standardnih grešaka ocena** kako bi odražavale stvarni varijabilitet ocena parametara: Njui-Vestova korekcija (engl. Newey-West).

Dinamički modeli

- KLRM model je statički:

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{k,\,t} + \varepsilon_t$$

- Model postaje dinamički ako se kao objašnjavajuće promenljive javljaju promenljive sa docnjama prvog reda, kako zavisne tako i objašnjavajućih promenljivih:

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + \gamma_1 \boldsymbol{y_{t-1}} + \gamma_2 x_{1t-1} + \ldots + \gamma_k x_{kt-1} + \varepsilon_t$$

- Mogu se dodati promenljive sa docnjama višeg reda: x_{1t-2}, y_{t-3}, itd.
  - Ovo može biti problematično ako se kao objašnjavajuća javlja zavisna promenljiva sa docnjom. Ona je slučajna promenljiva, pa se na taj način narušava pretpostavka KLRM da **objašnjavajuće promenljive nisu slučajne**.

Newey-West-ova procedura

- Obezbeđuje robusne standardne grešaka u prisustvu autokorelacije ili (i) heteroskedastičnosti (*engl. HAC standrd errors*).
- Podrazumeva direktno ocenjivanje asim. kovarijantne matrice, uz korišćenje izraza koji uključuje informaciju o heteroskedastičnosti i autokorelaciji, koja linearno opada do docnje L. Treba odgovoriti na dva pitanja:

1) Koliko docnji uzeti u obzir?
2) Kako modelirati smanjenje autokorel. tokom vremena?

Uobičajeni postupci u praksi:

1) Preporučuje se da je L = $T^{1/4}$ ili $(4(T/100)^{1/4})$.
2) Bartlett-ov prozor docnje $w_j=1-j/(L+1)$, j=1,2,.., L. (za $w_j=0$, svodi se na White-ovu korekciju, tj. otklanja samo heteroskedastičnost).

Specifikacija modela

1. Formulacija matematičke forme regresione jednačine
2. Izbor skupa objašnjavajućih promenljivih
3. Postavka pretpostavki o slučajnoj grešci

- Do sada smo razmatrali 3. pod pretpostavkom da je 1. i 2. korektno
- Greške specifikacije:
  - Pogrešna funkcionalna forma
  - Pogrešan skup objašnjavajućih promenljivih
  - Pogrešno postavljene pretpostavke o svojstvima slučajne greške

Greške specifikacije (u užem smislu)

- Izostavljanje relevantnih promenljivih
- Uključivanje irelevantnih promenljivih
- Netačna matematička forma regresione jednačine.
- Netačna specifikacija uticaja slučajnog člana (greške) jednačine

1. Pogrešan skup objašnjavajućih promenljivih

1. Izostavljanje relevantne objašnjavajuće promenljive

Posledice: ocene su pristrasne, sa varijansom koja nije najmanja moguća

(ocene nagiba će biti nepristrasne jedino ako izostavljena promenljiva nije korelisana sa onom koja je u modelu, ali ocena slobodnog člana ostaje pristrasna).

2. Uključivanje irelevantne objašnjavajuće promenljive

Posledice: ocene su nepristrasne, ali neefikasne

(ocene će biti efikasne jedino ako uključena promenljiva nije korelisana sa onom koja figuriše u modelu).

2. Pogrešna funkcionalna forma

- Najcešce se pretpostavlja da je specifikacija linearna, što ne mora uvek biti slučaj.
- Da bi se proverila opravdanost upotrebe linearne specifikacije koristi se Ramezejev (engl. Ramsey) RESET test.

RESET: *Regression equation specification error test*

- Nulta hipoteza: model ima korektnu specifikaciju
- Alternativna hipoteza: nulta hipoteza nije tacna.

- Primenom RESET testa proverava se prisustvo različitih grešaka specifikacije modela (pogrešne fun. forme, izostavljanje relev. promenljive, korelacija između regresora i greške), a koja je od njih zaista prisutna pokazuje dalja analiza.

RESET test

- Testira se ispravnost specifikacije modela.
- Algoritam testiranja:

1. Na osnovu ocenjenog polaznog modela: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i},$

dobijaju se ocenjene vrednosti $\hat{Y}_i^s, s = 1, 2, \ldots$

2. Potom ocenjujemo istu regresiju proširenu regresorima koji predstavljaju ocenjene vrednosti zavisne prom. polaznog modela dignute na stepen po izboru (predstavljaju zamenu, aproksimaciju za izostavljene prom.)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \alpha_1 \hat{Y}_i^2 + \alpha_2 \hat{Y}_i^3 + \varepsilon_i.$$

- Test se zasniva na F-statistici ili log. odnosa verodostojnosti (LR testu), u testu hipoteze da su koeficijenti svih dodatih regresora jednaki nuli.

Zadovoljavajući model

1. Regresija je statistički značajna (prema *F*-testu).
2. Svi ocenjeni parametri su statistički značajni (na osnovu *t*-testa) i odgovarajućeg su znaka.
3. U modelu nema autokorelacije.
4. U modelu ne postoji heteroskedastičnost.
5. Reziduali su normalno raspodeljeni.
6. Ne postoje indikacije o pogrešnoj specifikaciji modela.

Optimalan skup objašnjavajućih promenljivih

- Kako izabrati optimalan skup objašnjavajucih promenljivih?
- Kriterijumi:

- Najvece vrednosti korigovanog koeficijenta determinacije (min. s^2). Navedeno je posledica relacije:

$$s^2 = \left(1 - \bar{R}^2\right) \frac{\sum y_i^2}{n-1}.$$

- Najmanje vrednosti informacionog kriterijuma (uobičajeno za modele vremenskih serija).

Kriterijumi za izbor optimalnog skupa objašnjavajućih promenljivih (nastavak)

- **Informacioni kriterijum** je zbir dve komponente koje različito reaguju na promenu broja parametara modela (K):

$$IC(K) = \ln(s^2) + g(K/n).$$

  - Model sa najmanjom vrednošću IC je optimalan uz uslov da su valjane sve pretpostavke KLRM
  - AIC – Akaikeov informacioni kriterijum (g=2)
  - SC – Švarcov informacioni kriterijum (g=ln(n))
  - HQC – Hana-Kvinov kriterijum (g=2lnln(n)).

Alternativne strategije u postupku izbora modela

- Pored tradiconalnog pristupa (formulacija najjednostavnije jednačine koja je konzistentna sa ekonomskom teorijom, čiji kvalitet proveramo primenom statistikih i ekonometrijskih testova) – **induktivno istraživanje**.
- Novi pristup se vezuje za LSE i Dejvida Hendrija (engl. David Hendry) i danas predstavlja dominantan okvir ekonometrijskog modeliranja.
- Osnova je u **deduktivnom načinu razmišljanja**, gde prema strategiji modeliranja **<u>od opšteg ka posebnom</u>**, polazni model treba da bude što opštiji (obuhvata kao svoje specijalne slučajeve više jednostavnijih modela).

Pristup od opšteg ka posebnom (nastavak)

- Osnovna ideja ovog pristupa sastoji se u zahtevu da polazni model treba da:

1. Obuhvata sve modele mogućih alternativnih strategija.
2. Poseduje svojstva tačne specifikacije.

- Ova dva kriterijuma se međusobno ne isključuju (npr. izostavljanje neke relevantne ekonomske promenljive iz analize znači zanemarivanje određene postavke ekonomske teorije, što se primenom odgovarajućih testova može otkriti kao pogrešna specifikacija modela).