UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA MATEMATIKU I INFORMATIKU

Nevena Nešić

OTKRIVANJE I POSLEDICE PRISUSTVA HETEROSKEDASTIČNOSTI

-master rad-

Mentor:
prof. dr Zorana Lužanin

Novi Sad, 2014.

Sadržaj

Predgovor

Prevod reči "ekonometrija" znači "merenje u ekonomiji". Ono što se podrazumeva pod pojmom ekonometrija jeste primena statističkih i matematičkih metoda u analizi ekonomskih podataka, čiji je cilj pružanje empirijske podloge ekonomskim teorijama i njihove verifikacije ili odbacivanje. Usko je vezana za discipline kao što su: ekonomska teorija, matematička ekonomija i ekonomska statistika.

Matematička ekonomija izražava ekonomsku teoriju u određenoj matematičkoj formi, bez osvrtanja na empirijsku osnovu u toj primeni. Ekonomska statistika se bavi prikupljanjem, obradom i prezentovanjem ekonomskih pokazatelja u obliku tablica i slika, ali se ne bavi istovremeno i proverom ekonomskih teorija na temelju tako prikupljenih podataka. Upravo ekonometrija utvrđuje povezanost između ekonomskih pojava i potvrđuje odnosno ne potvrđuje ekonomsku teoriju, polazeći od jednačina matematičke ekonomije i gradeći ekonometrijske modele pogodne za testiranje. Regresiona analiza je jedna od najčešće korišćenih alata u ekonometrijskom radu kako bi se opisale veze među ekonomskim pojavama.

Tema rada se odnosi na narušavanje jedne od klasičnih pretpostavki linearne regresije. Reč je o homoskedastičnosti gde je u ekonomskim istraživanjima masovnih uporednih podataka vrlo često prisutna, bilo zbog prisustva netipičnih ekstremnih vrednosti, ili zbog same prirode analizirane pojave.

U prvom delu rada su navedeni osnovni pojmovi linearne regresije, metode ocenjivanja parametara, kao i njihove osobine. Objašnjeni su razlozi i posledice narušavanja osnovne pretpostavke o homoskedastičnosti, gde je najzastupljenija i u kojoj formi se javlja, odnosno kako se generiše varijansa slučajne greške. Navedeni su i objašnjeni oblici multiplikativne i aditivne heteroskedastičnost kao i postupci otklanjanja istih.

Drugi deo rada govori o metodama detektovanja problema heteroskedastičnosti. Heteroskedastičnost je prisutna ako reziduali pokazuju sistematska odstupanja za različite vrednosti nezavisne promenljive. Kao sredstvo preliminarne analize prikazana je grafička metoda, a da bi se dobio precizan odgovor da li je heteroskedastičnost prisutna ili ne, korišćeni su odgovarajući testovi. Najčešće nije poznata prava priroda heteroskedastičnosti, pa izbor odgovarajućeg testa zavisi od prirode podataka. U ekonometrijskoj literaturi postoji veći broj testova, a u radu su detaljnije objašnjeni Goldfeld-Kvantov, Glejzerov, Vajtov i Brojš-Paganov test. Navedeni su postupci primene testova, za koje oblike heteroskedastičnosti se koriste kao i koje su prednosti i koji su nedostaci svakog od njih.

U trećem delu rada, koji se odnosi na četvrto poglavlje, simulacijom podataka su prikazani rezultati prethodnih testova. Simulacije podataka različitih oblika heteroskedastičnosti kao i

navedeni primeri stvarnih podataka ukazuju na sličnosti i razlike testova kao i njihovu značajnost. Na kraju rada su prikazana ispitivanja veličine i moći testova kao i pobojšanja pojedinih testova u smislu njihove egzaktnosti i moći.

Ovom prilikom želela bih da se zahvalim svojoj profesorici i mentoru, dr Zorani Lužanin, za strpljenje i znanje koje mi je pružila tokom studiranja i izradi master rada.

Master rad posvećujem svojim roditeljima i sestri Milici.

1. Osnovni pojmovi linearne regresije

Najjednostavniji oblik linearne regresije, koja pokazuje linearnu zavisnost između dve pojave, jeste jednostruka (prosta) linearna regresija:

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

Y predstavlja zavisnu (stohastičku) promenljivu koja se izražava preko nezavisne (determinističke) promenljive X, koja se još naziva objašnjujuća promenljiva ili regresor.
ε je slučajna greška koju pravimo prilikom linearne regresije, a α i β su nepoznati parametri (regresijski koeficijenti). Da bismo ocenili nepoznate parametre, koristimo uzorak. Za fiksiranih n vrednosti nezavisne promenljive X, određuju se vrednosti promenljive Y. Na taj način se dobija n parova $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ koji čine uzorački model proste linearne regresije:

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad i = 1,2, \dots, n.$$

Za slučajnu promenljivu ε_i važe standardne (klasične) pretpostavke:

1. $E(\varepsilon_i) = 0, \ \ i = 1,2, \dots, n$
2. homoskedastičnost: $Var(\varepsilon_i) = \sigma^2, \ \ i = 1,2, \dots, n$
3. odsustvo autokorelacije: $cov(\varepsilon_i, \varepsilon_j) = 0, \ \ i = 1, \dots, n, \ \ j = 1, \dots, n, \ \ i \neq j$
4. $\varepsilon_i, \ \ i = 1, \dots, n$ imaju normalnu raspodelu
5. X je deterministička promenljiva sa fiksnim vrednostima u ponovljenim uzorcima, takva da je za bilo koji uzorak veličine n, $\sum_{i=1}^{n}(X_i - \bar{X})^2 \neq 0$ i njena granična vrednost je konačna kad $n \to \infty$.

Y_i je slučajna promenljiva čije je očekivanje

$$E(Y_i) = E(\alpha + \beta X_i + \varepsilon_i) = \alpha + \beta X_i,$$

a varijansa

$$Var(Y_i) = Var(\alpha + \beta X_i + \varepsilon_i) = Var(\varepsilon_i) = \sigma^2.$$

Budući da je ε_i normalne raspodele, proizilazi da je i $Y_i : \mathcal{N}(\alpha + \beta X_i, \sigma^2)$.

Jednačina $E(Y_i) = \alpha + \beta X_i$ se naziva regresijska linija populacije, dok njena ocena $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$ predstavlja regresijsku liniju uzorka.

Za koeficijente linearne regresije u slučaju malog uzorka (ispod 30 opažanja), poželjo je da imaju sledeće osobine:

1. Nepristrasnost: Ocena je nepristrrasna ako je očekivana vrednost ocene $\hat{\beta}$, jednaka pravoj vrednosti β tj.
$$E(\hat{\beta}) = \beta.$$
Razlika između ove dve vrednosti predstavlja pristrasnost ocene $\hat{\beta}$.

2. Efikasnost: Ocena je najefikasnija ako je nepristrasna i ima najmanju disperziju među svim ostalim nepristrasnim ocenama istog parametra.

3. BLUE (*best linear unbiased evaluator*): najbolja linearna nepristrasna ocena. Ocena sa ovom osobinom treba da zadovolji uslove da je ocena linearna funkcija opažanja, da je nepristrasna i da ima najmanju disperziju od svih ostalih ocena.

Ukoliko je obim uzorka veliki, poželjne su sledeće asimptotske osobine ocena parametara:

1. Asimptotska nepristrasnost podrazumeva da se povećanjem uzorka dobija što bolja ocena koeficijenta - očekivana vrednost ocene teži pravoj vrednosti koeficijenta kako uzorak raste tj.
$$\lim_{n\to\infty} E(\hat{\beta}) = \beta.$$

2. Konzistentnost: Dovoljan uslov za ovu osobinu je
$$\lim_{n\to\infty} MSE(\hat{\beta}) = 0,$$
gde je $MSE(\hat{\beta}) = E((\hat{\beta} - \beta)^2)$ srednje kvadratna greška ocene $\hat{\beta}$.

3. Asimptotska efikasnost podrazumeva da ocena koeficijenta ima osobinu konzistentnosti, najmanju asimptotsku varijansu i asimptotsku distribuciju sa konačnom sredinom i varijansom.

Pri ocenjivanju parametara α i β koristi se metoda najmanjih kvadrata jer ocene imaju sva poželjna svojstva ukoliko su ispunjene standardne pretpostavke. Ova metoda podrazumeva minimiziranje sume kvadratnih odstupanja registrovanih vrednosti od njihove sredine tj.

$$\min_{\alpha,\beta} \sum_{i=1}^{n} [Y_i - (\alpha + \beta X_i)]^2$$

Prvi izvodi po α i β se izjednače sa nulom i dobija se sistem normalnih jednačina:

$$\sum Y_i = \hat{\alpha} n + \hat{\beta} \sum X_i$$

$$\sum X_i Y_i = \hat{\alpha} \sum X_i + \hat{\beta} \sum X_i^2.$$

Rešenje za $\hat{\beta}$ se može napisati u obliku

$$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2},$$

a ako uvedemo oznake $S_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y})$, $S_{XX} = \sum (X_i - \bar{X})^2$, tada je $\hat{\beta} = \frac{S_{XY}}{S_{XX}}$. Ocenjena vrednost parametra α je $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$.

Ukoliko su ispunjene standardne pretpostavke, ocene parametara α i β dobijene metodom najmanjih kvadrata, BLUE metodom i metodom maksimalne verodostojnosti su ekvivalentne.

BLUE metoda (najbolje linearno nepristrasno ocenjivanje) podrazumeva da ocene dobijene ovom metodom budu linearne po promenljivama, nepristrasne i najmanje disperzije. Ukoliko ocenu β napišemo kao $\tilde{\beta} = \sum a_i Y_i$, koja treba da zadovolji uslove $E(\tilde{\beta}) = \beta$ i $Var(\tilde{\beta}) \to min$, rešavanjem problema

$$\min_{a_1,\dots,a_n} \sigma^2 \sum_{i=1}^{n} a_i^2$$

uz uslove: $\sum a_i = 0$ i $\sum a_i X_i = 1$,

dobijamo ocenu $\tilde{\beta} = \frac{S_{XY}}{S_{XX}}$. Na sličan način dobijamo $\hat{\alpha}$. Varijansa za $\tilde{\beta}$ se može izvesti kao $Var(\tilde{\beta}) = \frac{\sigma^2}{S_{XX}}$.

Metoda maksimalne verodostojnosti se zasniva na ideji da se pomoću izabranog uzorka odabere ona vrednost nepoznatog parametra, koja daje najveću verovatnoću da baš taj uzorak bude odabran. Pošto nezavisna promenljiva Y_i ima normalnu raspodelu, funkciju verodostojnosti možemo definisati kao

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\frac{-\sum [Y_i - (\alpha + \beta X_i)]^2}{2\sigma^2}},$$

gde se često koristi logaritamski oblik ove funkcije

$$L = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2.$$

Izjednačavajući prve izvode po $\alpha,\ \beta$ i σ^2 sa nulom, dobijamo ocene parametara, kao i ocenu za σ^2:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2,$$

gde se $e_i = Y_i - \hat{Y}_i,\ i = 1, \dots n,$ mogu posmatrati kao ocene greške ε_i i nazivaju se reziduali. Ocena disperuije $\hat{\sigma}^2$ je pristrasna, pa je nepristrasna ocena za σ^2 data kao $S^2 = \frac{n}{n-2}\hat{\sigma}^2 = \frac{1}{n-2}\sum e_i^2$. Ocena varijanse parametara $\hat{\beta}$ će tada biti $S_{\hat{\beta}}^2 = \frac{S^2}{S_{XX}}$. Standarna greška ocene $\hat{\beta}$ je označena sa $S_{\hat{\beta}}$.

Ukupno odstupanje jedne registrovane vrednosti promenljive Y_i od srednje vrednosti $\bar{Y}$ se može podeliti na: odstupanje objašnjeno modelom, $\hat{Y}_i - \bar{Y}$ i odstupanje registrovane vrednosti od vrednosti određene modelom (fitovane vrednosti), $Y_i - \hat{Y}_i$. Slično razlaganje važi i za kvadrate ovih odstupanja, odnosno važi

$$\underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}_{SST} = \underbrace{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}_{SSE} + \underbrace{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}_{SSR},$$

gde je sa SST označena ukupna suma kvadrata (ukupna greška koju pravimo regresionim modelom), SSE je suma kvadrata grešaka (odnosi se na deo variranja podataka Y koje pripisujemo greški modela), a SSR je regresijska suma kvadrata (odnosi se na deo variranja podataka Y koje se može objasniti modelom). Kao mera slaganja regresione prave sa eksperimentalnim podacima, koristi se koeficijent determinacije

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

R^2 uzima vrednosti od 0 do 1, a vrednosti bliže jedinici ukazuju na bolje slaganje modela i eksperimentalnih podataka.

Model višestruke regresije se može prikazati kao

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i,\ \ i = 1,2, \dots n,$$

gde je k broj nezavisnih promenljivih, a n broj opažanja (merenja), pa se često koristi matrični zapis višestruke regresije

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon},$$

gde je

$$\boldsymbol{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nk} \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}, \ \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Klasične pretpostavke za model višestruke regresije sadrže pored navedenih još i sledeće:

- broj nezavisnih promenljivih mora biti manji od obima uzorka $(k < n)$,
- ne postoji linearna veza između nezavisnih promenljivih.

Ocenjeni parametri, dobijeni metodom najmanjih kvadrata, u matričnom obliku su

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$

a njihova varijansa se može izraziti kao

$$Var(\widehat{\boldsymbol{\beta}}) = E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \underbrace{E(\boldsymbol{\varepsilon\varepsilon}')}_{\sigma^2 I} \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

Matrica $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ dimenzije $(k+1) \times (k+1)$ se naziva varijansna-kovarijansna matrica ocene parametara dobijenih metodom najmanjih kvadrata.

2. Narušavanje pretpostavke o homoskedastičnosti

Jedna od klasičnih pretpostavki linearne regresije predviđa da je varijansa greške homoskedastična, odnosno da je varijansa greške konstantna i jednaka za sva opažanja, ili za sve vrednosti nezavisnih promenljivih[16]:

$$Var(\varepsilon_i) = E[\varepsilon_i - E(\varepsilon_i)]^2 = E(\varepsilon_i^2) = \sigma^2 \text{ za svako } i.$$

Ukoliko ta pretpostavka nije ispunjena, nego su varijanse različite:

$$Var(\varepsilon_i) = \sigma_i^2, \quad i = 1, \dots n,$$

kaže se da su greške heteroskedastične, ili da postoji heteroskedastičnost u modelu.

Uvek treba delovati na osnovu teorije i podataka. U slučaju heteroskedastičnosti, povremeno postoje precizni teorijski razlozi za pretpostavku da greške imaju različite varijanse za različite vrednosti nezavisne promenljive. Veoma često su tako dobro definisani argumenti za prisustvo heteroskedastičnosti, a ponekad postoji nejasna sumnja da je pretpostavka o homoskedastičnosti suviše jaka[1].

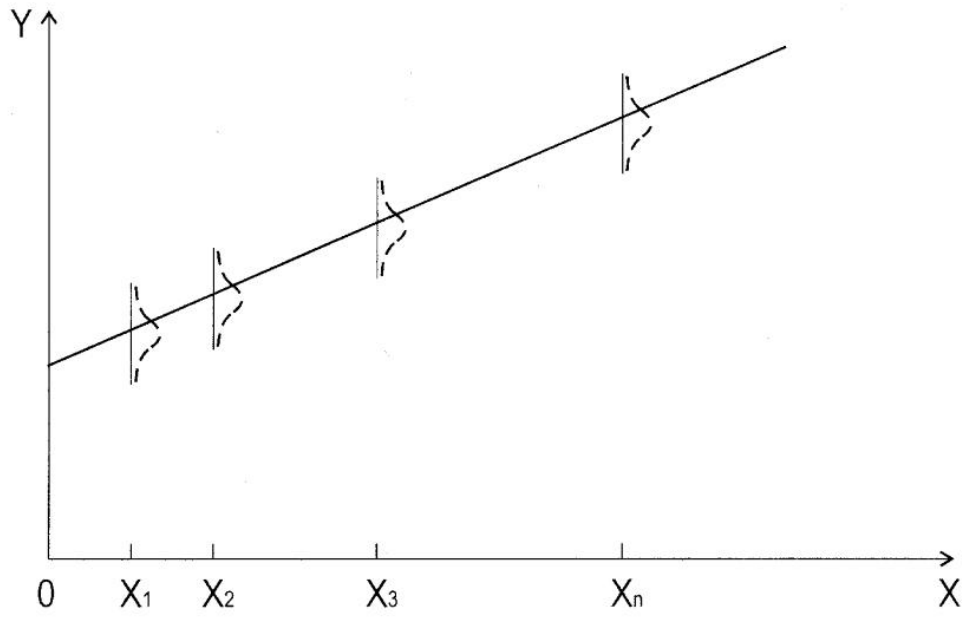
Heteroskedastičnost je pojava koja se najčešće javlja u podacima preseka. Ovi podaci predstavljaju određene vrednosti prikupljene u jednom trenutku vremena za različite jedinice posmatranja (pojedinci, domaćinstva, preduzeća, industrije, geografske oblasti i sl.). Elementi u uzorcima ovog tipa mogu biti različitih veličina kao što su male, srednje ili velike firme ili nizak, srednji i visoki prihod.

Važno je napomenuti, preliminarno, da kada je reč o prostornim uzorcima (posebno sa regionalnim podacima), heteroskedastičnost je uobičajena pojava zbog prirode prikupljanja podataka. Očigledni izvori heteroskedastičnosti su povezani sa različitim dimenzijama za različite regione u istraživanom području, neravnopravne koncentracije stanovništva i privredne aktivnosti u ruralnim i urbanim područjima[2].

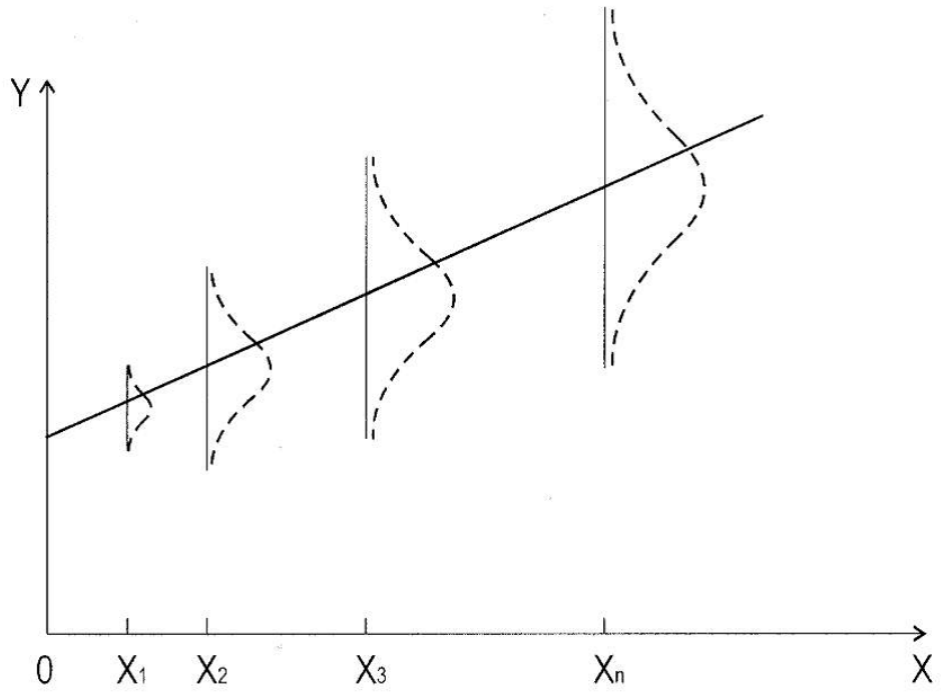
U analizi zavisnosti štednje od visine dohotka, ako su jedinice posmatranja pojedinci, po pravilu se primećuju daleko veće varijanse oko prosečne štednje za ljude sa višim, nego za one sa nižim nivoom dohotka. To je prirodna i uobičajena pojava, jer niži dohodak ne dozvoljava visoku štednju i štedi se obično sa tačno određenim ciljem. Pri višem dohotku, koji prevazilazi najnužnije potrebe, ponašanja mogu biti mnogo različitija, pa je disperzija u ponašanju oko prosečne štednje osetno veća.

Kako prihod raste, tako i ljudi imaju veći izbor raspolaganja njihovim prihodom, σ_i^2 će rasti sa porastom prihoda. Slično, kompanije sa većim profitom generalno očekuju veće variranje dividendi nego kompanije sa manjim profitom. Čak i nakon uzimanja u obzir veličine firme, očekuje se da će veće varijanse biti u profitu velikih firmi nego manjih. Varijansa prihoda može zavisiti od diverzifikacije proizvoda, istraživanja i razvoja potrošnje kao i industrijske karakteristike, pa bi samim tim odstupanja varirala i u firmama sličnih veličina[3].

Na slici 2.1 prikazan je prost linearni regresioni model u kojem je zavisna promenljiva štednja domaćinstva, a nezavisna promenljiva raspoloživi dohodak i pri tom je pretpostavljena homoskedastičnost greške. Dakle, može se primetiti da se rastom dohotka povećava srednji nivo štednje ali varijansa štednje oko srednje vrednosti ostaje nepromenjena za različite nivoe dohotka. Slika 2.2 predstavlja isti model gde je pretpostavljena heteroskedastičnost greške tako da je prikazan rast štednje sa povećanjem dohotka, ali se sada varijansa štednje menja pri različitim nivoima dohotka. Varijansa nije konstantna, već se povećava sa rastom dohotka što i odgovara realnim ekonomskim odnosima.



Slika 2.1: [5] *Homoskedastičnost greške*



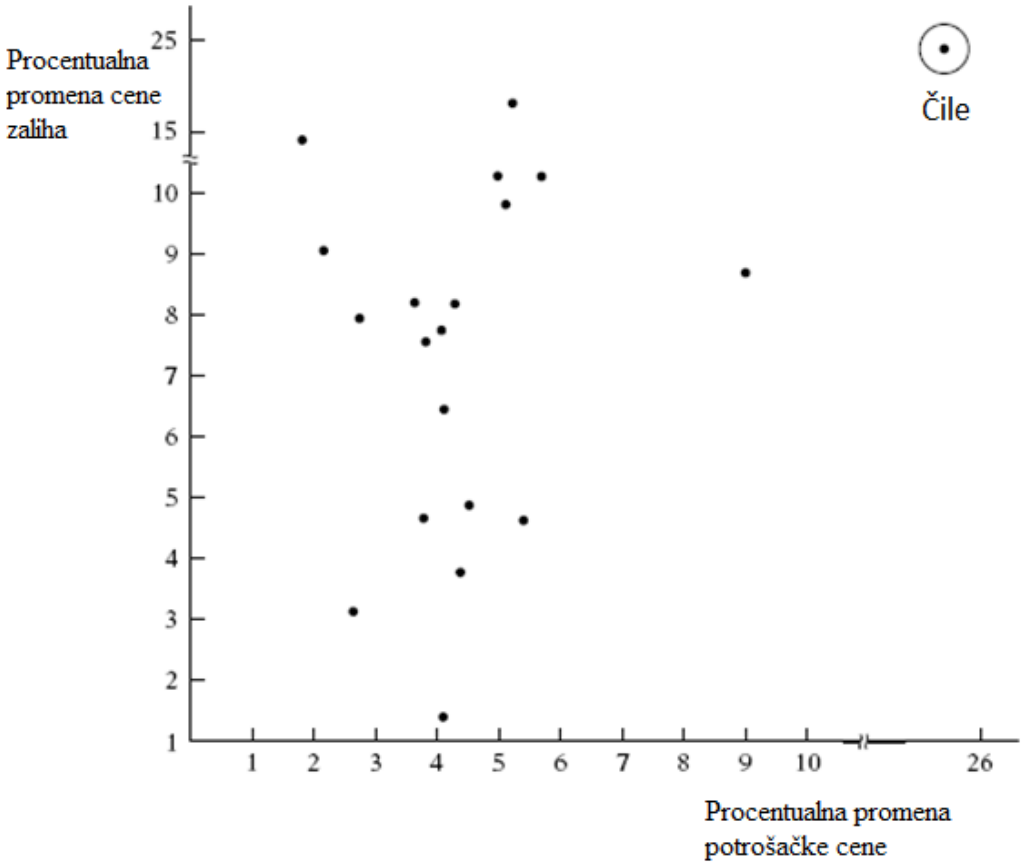
Slika 2.2: [5] *Heteroskedastičnost greške*

Heteroskedastičnost može biti uzrokovana i greškama specifikacije. Recimo, izostavljanjem nekog važnog regresora čiji će uticaj biti obuhvaćen greškom, može se dobiti različito variranje greške za razne opservacije. Slično, pogrešna funkcionalna forma modela takođe može dovesti do heteroskedastičnosti greške. Kako tehnike prikupljanja podataka napreduju, što podrazumeva obezbeđenje reprezantativnih uzoraka za statističku obradu, tako se i greške smanjuju, a samim tim i njihove disperzije. Ovo može biti još jedan razlog nastanka heteroskedastičnosti. Prema tome, banke koje imaju sofisticirane opreme za obradu podataka će verovatnije postići manje greške u mesečnim ili kvartalnim izveštajima svojih klijenata nego banke bez tih objekata.

Ali i kada nema grešaka u specifikaciji, samo postojanje ekstremnih vrednosti u uzorku može dovesti do heteroskedastičnosti. To je takozvano postojanje outlajera, posmatranja koja veoma odstupaju (veoma male ili velike vrednosti) od drugih opservacija u uzorku. Uključivanje ili isključivanje takvih posmatranja, posebno ako je obim uzorka mali, mogu značajno izmeniti rezultate regresione analize. U analizi uticaja ekonomske stabilnosti na stopu privrednog rasta, u raspoloživom skupu zemalja može se pojaviti nekoliko primera izuzetno visoke inflacije.

Kao primer se mogu navesti podaci o stopi promene cene zaliha i cene proizvodnje, prikupljeni tokom 1969.godine za dvadeset zemalja[4]. Kao što je prikazano na slici 2.3, opservacija za Čile

se može smatrati outlajerom zato što je cena zaliha za datu potrošačku cenu mnogo veća u odnosu na druge zemlje. U ovakvoj situaciji veoma je teško pretpostaviti homoskedastičnost.



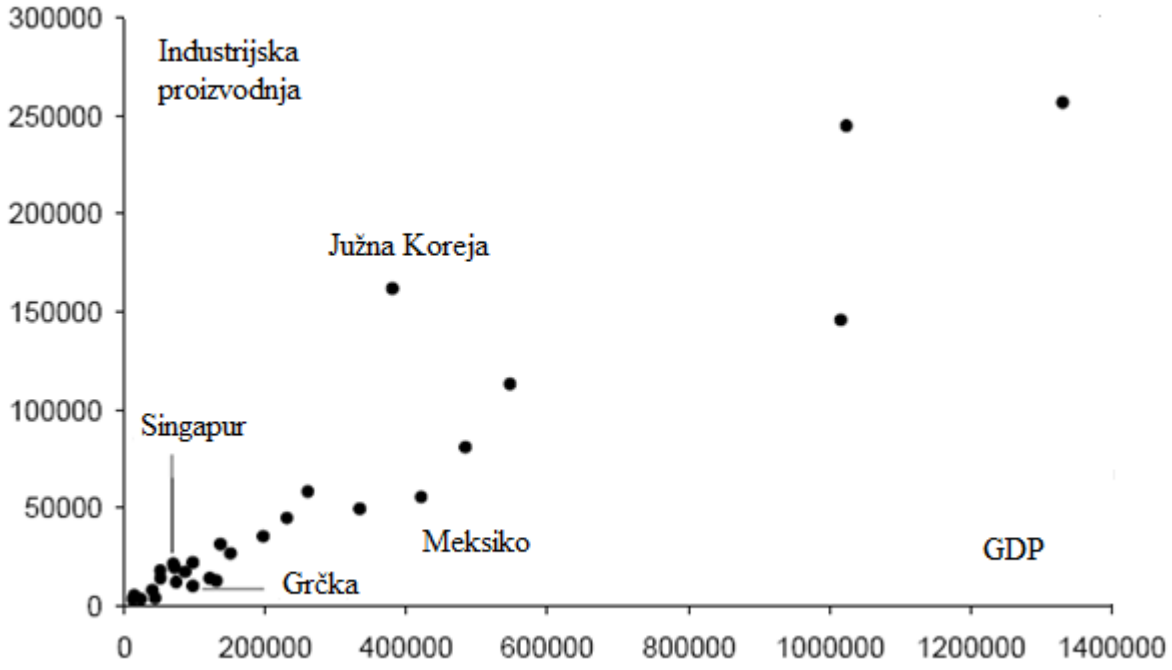
Slika 2.3: [4] *Zavisnost cene zaliha od potrošačke cene i primer outlajera*

Sledeći mogući izvor heteroskedastičnosti je zakrivljenost (nenormalnost) raspodele jedne ili više slučajnih promenljivih koje su uključene u model. Primeri su promenljive u ekonomiji kao što su prihod, bogatstvo, obrazovanje. Poznato je da su raspodele prihoda i bogatstva u društvu neravnomerne, tako da je većina bogatstva i prihoda u svojini nekolicine najbogatijih.

C. Dougherty[6] navodi da će heteroskedastičnost najverovatnije biti problem kada vrednosti promenljive Y značajno variraju za različite vrednosti X, kod modela $Y = \alpha + \beta X + \varepsilon$. Uzrok tome može biti slučaj kada su varijanse u izostavljenim promenljivama i greške merenja zajednički odgovorne za poremećaje koji mogu biti relativno mali, kada Y i X imaju male vrednosti, i veliki u suprotnom. Npr. pretpostavimo da možemo modelovati prostu regresiju kako bismo istražili odnos između vrednosti industrijske proizvodnje i bruto domaćeg proizvoda (GDP) na podacima različitih zemalja. U uzorak se uključuju male ekonomije kao što su Slovenija i Slovačka, i velike kao što su Francuska, Italija i Engleska. Industrijska proizvodnja

teži da bude $15 - 25\%$ GDP-a i kada je GDP veliki, 1% variranja će napraviti mnogo veću razliku u proizvodnji u apsolutnom iznosu, nego kada je GDP mali.

Na slici 2.4 predstavljen je grafik zavisnosti industrijske proizvodnje i bruto domaćeg proizvoda na osnovu uzorka zemalja 1994. godine, a podaci su preuzeti iz UNIDO Yearbook 1997 i izraženi u milionima američkih dolara. Južna Koreja i Meksiko su obe zemlje sa velikim bruto domaćim proizvodom. Sektor proizvodnje je mnogo važniji u Južnoj Koreji, čije su opservacije iznad trenda (*eng.trend line*), nego u Meksiku 1997. godine. Singapur i Grčka su drugi par zemalja sa relativno velikim i malim razvojem industrijske proizvodnje. Međutim, kako je GDP obe zemlje mali, tako je odstupanje od trenda malo.

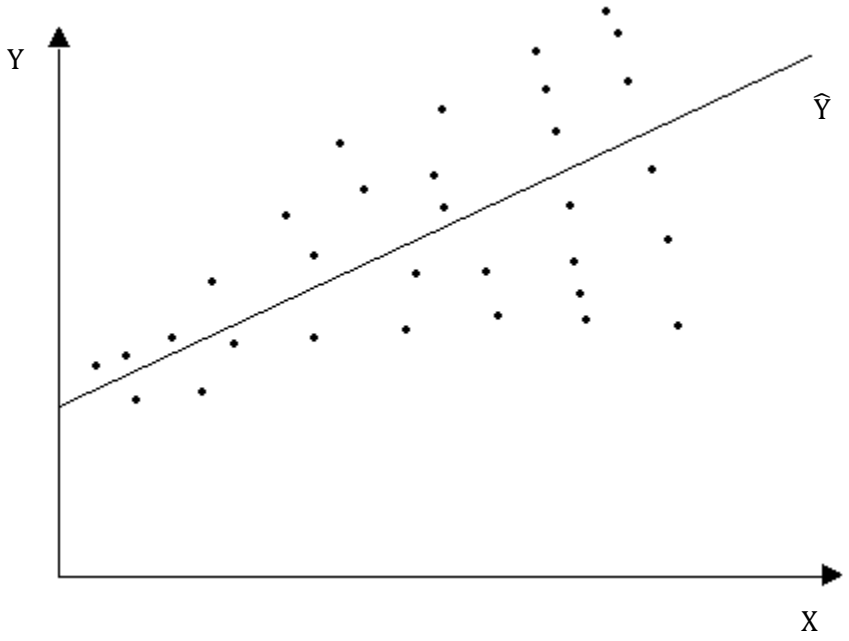


Slika 2.4: [6] *Industrijska proizvodnja i GDP*

Pri modelovanju linearne regresije, postavlja se pitanje da li je pojava heteroskedastičnosti problem? Inače, bilo koji ocenjeni parametri koji nisu BLUE predstavljaju zaista problem. Međutim, u slučaju heteroskedasticiti, to zavisi od specifikacije cilja. Ako je cilj model pouzdanosti, onda je to ozbiljna problem čak i na najmanjem nivou.

2.1 Posledice heteroskedastičnosti

Prisustvo heteroskedastičnosti u modelu zavisnosti štednje od dohotka može se predstaviti na osnovu sledećeg dijagrama rasturanja tačaka:



Slika 2.5 *Dijagram rasturanja tačaka*

Ocene nepoznatih parametara primenom metode najmanjih kvadrata (MNK) određuje se iz uslova da rezidualna suma kvadrata bude minimalna. U tom slučaju svi kvadrati reziduala imaju isti ponder, odnosno daju jednake informacije prilikom obrazovanja potrebnih ocena. Međutim, uslov MNK nije dovoljno precizan za uzorak koji je predstavljen na slici 2.5. Podaci koji su udaljeni od uzoračke regresione prave pružaju manje korisnih informacija o njenom položaju od onih podataka koji su joj bliži. Udaljenijim podacima odgovaraju veće vrednosti reziduala u apsolutnom iznosu. Ovi reziduali dominiraju u ukupnoj rezidualnoj sumi kvadrata. Otuda je realno očekivati da se primenom MNK ne dobijaju ocene sa poželjnim statističkim svojstvima.

Prisustvo heteroskedastičnosti ne dovodi do toga da su ocene dobijene metodom najmanjih kvadrata pristrasne, ali uzrokuje da ocene nemaju minimalnu varijansu, tj. nisu efikasne. Zato, ako je prisutna heteroskedastičnost, ni predviđanja na osnovu ocena originalnog modela neće biti efikasna. Varijansa predviđanja, osim varijanse grešaka, uključuje i varijansu ocena parametara.

Kad su ocene parametara neefikasne, predviđanje dobijenom primenom tih ocena takođe je neefikasno [16].

Posledice se manifestuju i na pouzdanost testiranja hipoteza i intervala poverenja zbog pristrasnosti ocenjenih varijansi. Naime, pristrasnost varijanse može biti pozitivna ili negativna, a samim tim i izračunati intervali širi ili uži od korektnih. Ukoliko je pristrasnost negativna, prava varijansa parametara je veća nego njena ocena. Zato se u takvom slučaju statistički testovi pristrasni, t-odnosi precenjeni, neopravdano visoki, što može uticati na donošenje pogrešnog zaključka o statistički značajnom uticaju promenljive X u situaciji kad ona nije korelisana sa Y. Ocenjeni intervali su uži nego što bi trebalo da budu, a greška prve vrste (verovatnoća odbacivanja hipoteze kad je istinita) potcenjena, odnosno viša nego što se pretpostavlja (češće je odbacujemo, a ne treba). Zbog toga smo skloni da prihvatamo kao statistički značajne i parametre koji to ustvari nisu.

2.1.1 Ispitivanje osobina ocena dobijene metodom najmanjih kvadrata

Pretpostavimo da u modelu oblika

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

postoji heteroskedastičnost: $Var(\varepsilon_i) = \sigma_i^2, i = 1,2,\dots,n,$ a da su sve ostale pretpostavke klasičnog linearnog modela ispunjene.[1]

Tvrđenje: Ocene parametara α i β dobijene metodom najmanjih kvadrata u uslovima heteroskedastičnosti su nepristrasne i konzistentne, ali nemaju osobinu efikasnosti i asimptotske efikasnostu i ocena varijanse je pristrasna.

Dokaz: Izvedena je svaka od osobina za ocenjenu vrednost parametra β [17].

1) Nepristrasnost

Ocenjivač metode najmanjih kvadrata za β je

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{S_{XX}} = \frac{\sum(X_i - \bar{X})[\beta(X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})]}{S_{XX}} = \beta +$$

$$\frac{\sum(X_i - \bar{X})\varepsilon_i - \bar{\varepsilon}\overbrace{\sum(X_i - \bar{X})}^{=0}}{S_{XX}} = \beta + \frac{\sum(X_i - \bar{X})\,\varepsilon_i}{S_{XX}}.$$

Očekivana vrednost za $\hat{\beta}$ je tada

[1] Radi jednostavnosti, ispitivaćemo osobine ocena parametara proste linearne regresije. Takođe, ovo važi i za višestruku linearnu regresiju.

$$E(\hat{\beta}) = \beta + E\left(\frac{\sum(X_i - \bar{X})\,\varepsilon_i}{S_{XX}}\right) = \beta + \frac{\sum(X_i - \bar{X})\,\overbrace{E(\varepsilon_i)}^{=0}}{S_{XX}} = \beta.$$

Slično se dobije i za α,

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = (\alpha + \beta\bar{X} + \bar{\varepsilon}) - \hat{\beta}\bar{X},$$

$$E(\hat{\alpha}) = \alpha + \beta\bar{X} + E(\bar{\varepsilon}) - E(\hat{\beta})\bar{X} = \alpha.$$

2) Konzistentnost

Kako je $\hat{\beta}$ nepristrasan, proizilazi da je $MSE(\hat{\beta}) = Var(\hat{\beta})$. Ispisujući formulu varijanse, imamo

$$\begin{aligned} Var(\hat{\beta}) &= E(\hat{\beta} - \beta)^2 = E\left(\frac{\sum(X_i - \bar{X})\,\varepsilon_i}{S_{XX}}\right)^2 \\ &= \frac{1}{S_{XX}^2} E\left(\sum (X_i - \bar{X})^2\,\varepsilon_i^2 + 2\sum_i \sum_{j>i} (X_i - \bar{X})(X_j - \bar{X})\varepsilon_i\varepsilon_j\right) \\ &= \frac{1}{S_{XX}^2}\left(\sum (X_i - \bar{X})^2\,\overbrace{E(\varepsilon_i^2)}^{=\sigma_i^2} + 2\sum_i \sum_{j>i} (X_i - \bar{X})(X_j - \bar{X})\,\overbrace{E(\varepsilon_i\varepsilon_j)}^{=0}\right) = \frac{\sum(X_i - \bar{X})^2\,\sigma_i^2}{S_{XX}^2}. \end{aligned}$$

Uvodeći oznake: $\theta_i = \sigma_i^2 - \bar{\sigma}^2$ gde je $\bar{\sigma}^2 = \sum \sigma_i^2/n$ dobijamo

$$Var(\hat{\beta}) = \frac{\sum(X_i - \bar{X})^2\,(\bar{\sigma}^2 + \theta_i)}{S_{XX}^2} = \frac{\bar{\sigma}^2}{S_{XX}} + \frac{\sum(X_i - \bar{X})^2\,\theta_i}{S_{XX}^2}.$$

Da bi ocena bila konzistentna, potrebno je da $\lim_{n\to\infty} Var(\hat{\beta}) = 0$.

$$\lim_{n\to\infty} Var(\hat{\beta}) = \lim_{n\to\infty}\left(\frac{\bar{\sigma}^2/n}{S_{XX}/n} + \frac{\sum(X_i - \bar{X})^2\,\theta_i/n^2}{S_{XX}^2/n^2}\right) = 0.$$

3) Efikasnost

Ideja razmatranja se sastoji u tome da se izvedu formule BLUE ocenjivača pod uslovom heteroskedastičnosti i da se uporede sa formulama MNK ocenjivača. Postojanje razlike će ukazati da ocenjivači metode najmanjih kvadrata nisu efikasni.

Podelićemo obe strane linearne regresije sa σ_i:

$$\frac{Y_i}{\sigma_i} = \alpha\left(\frac{1}{\sigma_i}\right) + \beta\left(\frac{X_i}{\sigma_i}\right) + \frac{\varepsilon_i}{\sigma_i}.$$

Ako se uvedu oznake: $Y_i^* = (Y_i/\sigma_i), W_i^* = (1/\sigma_i), X_i^* = (X_i/\sigma_i)$, i $\varepsilon_i^* = (\varepsilon_i/\sigma_i)$, dobije se transformisana regresijska jednačina:

$$Y_i^* = \alpha W_i^* + \beta X_i^* + \varepsilon_i^*.$$

Za ovakav model ispunjene su klasične pretpostavke:

$$E(\varepsilon_i^*) = \frac{E(\varepsilon_i)}{\sigma_i} = 0,$$

$$cov\left(\varepsilon_i^* \varepsilon_j^*\right) = \frac{cov(\varepsilon_i, \varepsilon_j)}{\sigma_i \sigma_j} = 0,$$

$$Var(\varepsilon_i^*) = Var\left(\frac{\varepsilon_i}{\sigma_i}\right) = \frac{Var(e_i)}{\sigma_i^2} = 1,$$

pa su ocene parametara dobijene MNK i BLUE metodom jednake. Primenjujemo metodu najmanjih kvadrata tako što minimiziramo $\sum[Y_i^* - E(Y_i^*)]^2$ i kada prve izvode sume po α i β izjednačimo sa nulom, dobijemo sledeće normalne jednačine:

$$\sum W_i^* Y_i^* = \tilde{\alpha} \sum W_i^{*2} + \tilde{\beta} \sum W_i^* X_i^*,$$

$$\sum X_i^* Y_i^* = \tilde{\alpha} \sum W_i^* X_i^* + \tilde{\beta} \sum X_i^{*2}.$$

Vraćanjem na prethodne oznake i uvođenjem težinskog koeficijenta $w_i = 1/\sigma_i^2$ i težinskih aritmetičkih sredina $\tilde{X} = (\sum w_i X_i)/(\sum w_i)$ i $\tilde{Y} = (\sum w_i Y_i)/(\sum w_i)$, dobija se najbolji linearni nepristrasni ocenjivač za β i α,

$$\tilde{\beta} = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2}$$

$$= \frac{\sum w_i\left(X_i - \tilde{X}\right)(Y_i - \tilde{Y})}{\sum w_i\left(X_i - \tilde{X}\right)^2},$$

$$\tilde{\alpha} = \tilde{Y} - \tilde{\beta}\tilde{X}.$$

Jasno se vidi da $\hat{\beta} \neq \tilde{\beta}$ tj. ocena dobijena metodom najmanjih kvadrata nije BLUE.

4) Asimptotska efikasnost

Ocenjivači dobijeni metodom maksimalne verodostojnosti (MMV) su asimptotski efikasni. Ideja ispitivanje ove osobine u uslovima heteroskedastičnosti se sastoji u tome da se uporede varijanse MMV i MNK ocenjivača. Ukoliko im varijanse nisu asimptotski jednake, tada ocenjivač dobijen metodom najmanjih kvadrata neće biti asimptotski efikasan.

Da bismo dobili maksimalno verodostojni ocenjivač $\check{\beta}$, polazimo od funkcije verodostojnosti u logaritamskom obliku i prvih izvoda te funkcije:

$$L = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log\sigma_i^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{Y_i - \alpha - \beta X_i}{\sigma_i}\right)^2$$

$$\frac{\partial L}{\partial \alpha} = \sum_i \left(\frac{Y_i - \alpha - \beta X_i}{\sigma_i}\right),$$

$$\frac{\partial L}{\partial \beta} = \sum_i \left[\frac{X_i(Y_i - \alpha - \beta X_i)}{\sigma_i^2}\right].$$

Ukoliko ove prve izvode izjednačimo sa nulom i uvedemo oznake kao kod ispitivanja BLUE, dobićemo da su maksimalno verodostojni ocenjivači od α i β isti kao kod BLUE, odnosno

$$\check{\beta} = \frac{\sum w_i\left(X_i - \tilde{X}\right)\left(Y_i - \tilde{Y}\right)}{\sum w_i\left(X_i - \tilde{X}\right)^2} = \tilde{\beta}.$$

Ako je $\tilde{\varepsilon} = (\sum w_i\varepsilon_i)/(\sum w_i)$, zamenjujući Y_i i $\tilde{Y}$ u jednačinu, imamo

$$\tilde{\beta} = \frac{\sum w_i\left(X_i - \tilde{X}\right)\left[\beta\left(X_i - \tilde{X}\right) + (\varepsilon_i - \tilde{\varepsilon})\right]}{\sum w_i\left(X_i - \tilde{X}\right)^2} = \beta + \frac{\sum w_i\left(X_i - \tilde{X}\right)\varepsilon_i}{\sum w_i\left(X_i - \tilde{X}\right)^2}.$$

Varijansa za $\tilde{\beta}$ se može izraziti na sličan način kao i za $\hat{\beta}$,

$$Var(\tilde{\beta}) = E\left(\tilde{\beta} - \beta\right)^2 = E\left(\frac{\sum w_i\left(X_i - \tilde{X}\right)\varepsilon_i}{\sum w_i\left(X_i - \tilde{X}\right)^2}\right)^2 = \frac{\sum w_i^2\left(X_i - \tilde{X}\right)^2\sigma_i^2}{\left[\sum w_i\left(X_i - \tilde{X}\right)^2\right]^2} = \frac{\sum w_i\left(X_i - \tilde{X}\right)^2}{\left[\sum w_i\left(X_i - \tilde{X}\right)^2\right]^2}$$

$$Var(\tilde{\beta}) = \frac{1}{\sum w_i (X_i - \tilde{X})^2}.$$

Može se zaključiti da bez obzira na veličinu uzorka, važi

$$Var(\tilde{\beta}) = \frac{1}{\sum w_i (X_i - \tilde{X})^2} \neq Var(\hat{\beta}) = \frac{\sum (X_i - \bar{X})^2 \sigma_i^2}{S_{XX}^2},$$

tačnije

$$Var(\tilde{\beta}) < Var(\hat{\beta}),$$

tako da ocenjivači dobijeni metodom najmanjih kvadrata nemaju osobinu asimptotske efikasnosti u uslovima heteroskedastičnosti.

5) Varijansa ocenjivača

Pokazujemo da je $S_{\hat{\beta}}^2$ pristrasan ocenjivač od $Var(\hat{\beta})$ tj. $\mathrm{E}(S_{\hat{\beta}}^2) \neq Var(\hat{\beta})$, gde je $S_{\hat{\beta}}^2 = S^2/S_{XX}$, $S^2 = \sum e_i^2/(n-2)$.

$$E\left(S_{\hat{\beta}}^2\right) = \frac{E(S^2)}{S_{XX}},$$

$$E(S^2) = E\left(\frac{\sum e_i^2}{n-2}\right) = E\left(\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}\right) = \frac{1}{n-2} E\left(\sum \left[-(\hat{\alpha} - \alpha) - (\hat{\beta} - \beta) X_i + \varepsilon_i)\right]^2\right).$$

Kada u ovaj izraz uvrstimo $\hat{\alpha} - \alpha = \bar{Y} - \hat{\beta}\bar{X} - \bar{Y} + \beta\bar{X} + \bar{\varepsilon} = -\bar{X}(\hat{\beta} - \beta) + \bar{\varepsilon}$, dobije se

$$\begin{aligned} E(S^2) &= \frac{1}{n-2} E\left(\sum \left[(\hat{\beta} - \beta)(X_i - \bar{X}) + (\varepsilon_i - \bar{\varepsilon})\right]^2\right) \\ &= \frac{1}{n-2}\left(E(\hat{\beta} - \beta)^2 S_{XX} + E\left(\sum (\varepsilon_i - \bar{\varepsilon})^2\right) - 2E(\hat{\beta} - \beta) \sum (X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})\right). \end{aligned}$$

Pojedine izraze u zagradi ćemo računati posebno:

$$\begin{aligned} E(\hat{\beta} - \beta) \sum (X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon}) &= E\left(\frac{\sum (X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})}{S_{XX}}\right) \sum (X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon}) \\ &= E\left(\frac{\sum (X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})}{S_{XX}}\right)^2 S_{XX} = E(\hat{\beta} - \beta)^2 S_{XX} = Var(\hat{\beta}) S_{XX}, \end{aligned}$$

$$E\left(\sum(\varepsilon_i-\bar{\varepsilon})^2\right)=E\left(\sum\varepsilon_i^2\right)+nE(\bar{\varepsilon}^2)-2E\left(\bar{\varepsilon}\sum\varepsilon_i\right)=\ E\left(\sum\varepsilon_i^2\right)-nE(\bar{\varepsilon}^2)$$
$$=\sum\sigma_i^2+\frac{1}{n}\sum\sigma_i^2=\left(\frac{n-1}{n}\right)\sum\sigma_i^2.$$

Vraćanje u izraz, $E(S^2)$ postaje

$$E(S^2)=\frac{1}{n-2}\left[-Var(\hat{\beta})S_{XX}+\left(\frac{n-1}{n}\right)\sum\sigma_i^2\right]=\frac{1}{n-2}\left[-\frac{\sum(X_i-\bar{X})^2\,\sigma_i^2}{S_{XX}}.+\frac{(n-1)\sum\sigma_i^2}{n}\right].$$

Ukoliko upotrebimo oznake $\theta_i=\sigma_i^2-\bar{\sigma}^2$, gde je $\bar{\sigma}^2=\sum\sigma_i^2/n$, dobijamo

$$E(S^2)=\bar{\sigma}^2-\frac{\sum(X_i-\bar{X})^2\theta_i}{(n-2)S_{XX}}.$$

Konačno je

$$E\left(S_{\hat{\beta}}^2\right)=\frac{\bar{\sigma}^2}{S_{XX}}-\frac{\sum(X_i-\bar{X})^2\theta_i}{(n-2){S_{XX}}^2}.$$

Kao što se može videti

$$E\left(S_{\hat{\beta}}^2\right)=\frac{\bar{\sigma}^2}{S_{XX}}-\frac{\sum(X_i-\bar{X})^2\theta_i}{(n-2){S_{XX}}^2}\neq Var(\hat{\beta})=\frac{\bar{\sigma}^2}{S_{XX}}+\frac{\sum(X_i-\bar{X})^2\,\theta_i}{S_{XX}^2}.\qquad\blacksquare$$

2.2 Oblici heteroskedastičnosti

Ponekad smo u stanju da pri formulisanju modela regresije pretpostavimo koji oblik heteroskedastičnosti se pojavljuje u podacima. Najčešći oblici koji se pominju u litaraturi i koji se primenjuju u praksi su multiplikativna i aditivna heteroskedastičnost.

2.2.1 Multiplikativna hetreoskedastičnost

Opšti oblik mutiplikativne heteroskedastičnosti je

$$\log\sigma_i^2=\log\sigma^2+\delta_1\log Z_{i1}+\delta_2\log Z_{i2}+\cdots+\delta_p\log Z_{ip},$$

koji prikazuje zavisnost σ_i^2 od p promenljivih. Prikaz može biti prikladan za model višestruke regresije u kojem Z_s , $s = 1, \dots, p$, mogu biti samo nezavisne promenljive regresijskog modela. Specifikacija se primenjuje samo kada su sve Z_s pozitivne.

Ukoliko imamo multiplikativni oblik heteroskedastičnosti koji sadrži jednu promenljivu Z, dobija se uobičajeni oblik te heteroskedastičnosti

$$\log \sigma_i^2 = \log\sigma^2 + \delta \log Z_i$$

$$\log \sigma_i^2 = \log\sigma^2 Z_i^{\delta},$$

odnosno

$$\sigma_i^2 = \sigma^2 Z_i^{\delta},$$

koji sadrži dva parametra, σ^2 i δ. Posebno je važan parametar δ koji meri jačinu heteroskedastičnosti i što je njegova veličina manja, manje su razlike između pojedinačnih varijansi, a kada je $\delta = 0$, model je homoskedastičan.

Ukoliko su oba parametra σ^2 i δ nepoznata, mora se oceniti zajedno sa regresijskim koeficijentima α i β ili se vrednost barem jednog od njih mora specificirati a priori[17].
Potpuni regresijski model za tu vrstu heteroskedastičnog uslova je

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

$$\varepsilon_i \sim N(0, \sigma_i^2),$$

$$\sigma_i^2 = \sigma^2 Z_i^{\delta} \qquad (\sigma > 0;\ Z_i > 0).$$

Pretpostavlja se da su ε_i nekorelisane, a X_i i Z_i nestohastičke promenljive. Primeni se metoda maksimalne verodostojnsti i funkcija verodostojnosti

$$L = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log\sigma_i^2 - \frac{1}{2}\sum_{i=1}^{n}\left[\frac{Y_i - \alpha - \beta X_i}{\sigma_i}\right]^2$$

postaje

$$L = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(\log\sigma^2 + \delta\log Z_i) - \frac{1}{2}\sum_{i=1}^{n}\left[\frac{Y_i - \alpha - \beta X_i}{\sigma Z_i^{\delta/2}}\right]^2.$$

Prvi izvodi od L su

$$\frac{\partial L}{\partial \alpha} = \frac{1}{\sigma^2} \sum_i \left[\frac{Y_i - \alpha - \beta X_i}{Z_i^\delta} \right]$$

$$\frac{\partial L}{\partial \beta} = \frac{1}{\sigma^2} \sum_i \left[\frac{(Y_i - \alpha - \beta X_i) X_i}{Z_i^\delta} \right]$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_i \left[\frac{Y_i - \alpha - \beta X_i}{Z_i^{\delta/2}} \right]^2$$

$$\frac{\partial L}{\partial \delta} = -\frac{1}{2} \sum_i \log Z_i + \frac{1}{2\sigma^2} \sum_i \left[\frac{(Y_i - \alpha - \beta X_i)^2 \log Z_i}{Z_i^\delta} \right].$$

Svaki od izraza se izjednači sa nulom i dobiju se četiri jednačine sa četiri nepoznate. Međutim, jednačine su izrazito nelinearne zbog prisutnosti δ. Relativno jednostavniji način dobijanja rešenja jeste da se prva tri izraza izjednače sa nulom, i za određene vrednosti δ i odgovarajuće α, β i σ^2, računati L. Za većinu ekonomskih modela δ se nalazi u rasponu od 0 do 3 ili 4, tako da se ocene za parametre mogu dobiti za vrednosti $\delta = 0, \delta = 0.1, \delta = 0.2$ i tako dalje. Potom od svih rešenja se izabere ono koje daje najveću vrednost za L. Ova metoda izračunavanja maksimalno verodostojnih ocena je tzv. metoda traganja. Za veću preciznost, mogu se izabrati finiji intervali za δ. Postoje i druge metode izračunavanja, ali metoda traganja ima prednost zbog jednostavnosti i manjeg rizika dobijanja rešenja u lokalnom umesto globalnom maksimumu funkcije verodostojnosti. Maksimalno verodostojni ocenjivači od α, β, δ i σ^2 imaju sva poželjna asimptptska svojstva.

Praktično, najveći problem je problem izbora Z. Mala je verovatnoća da istraživač zna promenljivu koja je povezana sa varijansom odstupanja, a da nije uključena u jednačinu regresije. Obično je nezavisna promenljiva X izbor za Z.

Specijalni slučaj multiplikativne heteroskedastičnosti se dobija kada je $Z_i = X_i$ i $\delta = 2$, tj.

$$\sigma_i^2 = \sigma^2 X_i^2.$$

Ovaj oblik se obično koristi u primenjenom radu. Ponekad ovo može biti data razumna pretpostavka da je standardna devijacija regresijskog odstupanja proporcionalna vrednosti nezavisne promenljive. Kao primer se može navesti zavisnost štednje domaćinstva od raspoloživog dohotka gde možemo pretpostaviti ovaj oblik heteroskedastičnosti.

2.2.2 Aditivna heteroskedastičnost

Opšti oblik aditivne heteroskedastičnosti je

$$\sigma_i^2 = a_0 + a_1 Z_{i1} + a_2 Z_{i2} + \cdots + a_p Z_{ip}.$$

Ukoliko je $p = 2, Z_{i1} = X_i$ i $Z_{i2} = X_i^2$, dobija se uobičajeni oblik

$$\sigma_i^2 = a + bX_i + cX_i^2,$$

gde su a, b i c konstante čije se vrednosti ocenjuju ili pretpostavljaju[17]. Kada su b i c jednaki nuli, model je homoskedastičan, a ako su a i b jednaki nuli, dobija se specijalni slučaj multiplikativne heteroskedastičnosti, istaknut u prethodnom razmatranju. Ovaj oblik omogućava homoskedastičnost kao i dve vrste heteroskedastičnosti što je znatno manje ograničeno od oblika $\sigma_i^2 = \sigma^2 Z_i^\delta$.

Specijalni slučaj aditivne heteroskedastičnosti, koji se zove heteroskedastičnost zavisne promenljive, uključuje pretpostavku da je varijansa odstupanja proporcionalna kvadratnoj sredini od Y_i, odnosno $\sigma_i^2 = \sigma^2[E(Y_i)]^2 = \sigma^2(\alpha + \beta X_i)^2$. Tako je u jednoj empirijskoj analizi potrošnje čaja u zavisnosti od dohotka i veličine domaćinstva utvrđen ovaj oblik varijanse [5]. Do rezultata se došlo tako što su sva domaćinstva bila grupisana po veličini i dohotku, pa se variranje potrošnje unutar grupe (za datu veličinu dohotka i grupe) prepisuje dejstvu ostalih faktora. Stoga, standardna devijacija potrošnje čaja unutar grupe i odgovara σ_i. Primećeno je da su one približno proporcionalne aritmetičkoj sredini potrošnje čaja u svakoj grupi. Kako aritmetičke sredine odgovaraju očekivanim vrednostima zavisne promenljive, došlo se do pretpostavke o proporcionalnosti standardne devijacije i očekivane vrednosti zavisne promenljive.

2.3 Otklanjanje heteroskedastičnosti

Predložene su različite metode otklanjanja heteroskedastičnosti u zavisnosti od forme σ_i^2.

Ukoliko je σ_i^2 poznato, koristi se metoda ponderisanih najmanjih kvadrata, odnosno, polazna regresija $Y_i = \alpha + \beta X_i + \varepsilon_i$, se podeli sa σ_i^2 i na taj način se dobije transformisan model

$$\frac{Y_i}{\sigma_i} = \alpha\left(\frac{1}{\sigma_i}\right) + \beta\left(\frac{X_i}{\sigma_i}\right) + \frac{\varepsilon_i}{\sigma_i}$$

na koji se primeni MNK.

Međutim, σ_i^2 najčešće nije poznato, već se mogu pretpostaviti različite specifikacije varijanse kao što su:

- multiplikativna heteroskedastičnost: $\sigma_i^2 = \sigma^2 X_i^2$

Jednačina regresije se pretvara u homoskedastičnu deljenjem obe strane sa X_i pa se tako dobije:

$$\frac{Y_i}{X_i} = \alpha\left(\frac{1}{X_i}\right) + \beta + \frac{\varepsilon_i}{X_i}$$

odnosno

$$Y_i^* = \beta + \alpha X_i^* + \varepsilon_i^*,$$

gde je $Y_i^* = Y_i/X_i$,$X_i^* = 1/X_i$ i $\varepsilon_i^* = \varepsilon_i/X_i$. Može se zaključiti da ε_i^* zadovoljava sve osnovne pretpostavke klasičnog modela:

$$Var(\varepsilon_i^*) = Var\left(\frac{\varepsilon_i}{X_i}\right) = \frac{1}{X_i^2} Var(\varepsilon_i) = \frac{1}{X_i^2}\sigma^2 X_i^2 = \sigma^2.$$

Sa regresijskom jednačinom se može baratati kao i sa klasičnim modelom, osim što se uloge α i β menjaju. Ocenjivači od α i β dobijaju se metodom najmanjih kvadrata i imaju sva poželjna svojstva i svi će relevantni intervali poverenja i testovi značajnosti vredeti za sve veličine uzorka ukoliko je specifikacija heteroskedastičnosti korektna. Jednostavnost ovakvog oblika heteroskedastičnosti je njena prednost što objašnjava njenu popularnost. Nedostatak je krutost jer ta specifikacija isključuje mogućnost da odstupanje bude homoskedastično ili da heteroskedastičnost bude različita od stepena koji odgovara $\delta = 2$.

- varijansa proporcionalna sa X_i: $\sigma_i^2 = \sigma^2 X_i$

Na sličan način kao i kod multiplikativne heteroskedastičnosti, jednačina regresije se transformiše u

$$\frac{Y_i}{\sqrt{X_i}} = \alpha\left(\frac{1}{\sqrt{X_i}}\right) + \beta\sqrt{X_i} + \frac{\varepsilon_i}{\sqrt{X_i}}, \quad X_i > 0$$

i primeni se metoda najmanjih kvadrata.

- aditivna heteroskedastičnost: $\sigma_i^2 = a + bX_i + cX_i^2$

Postupci metode ocenjivanja modela sa ovim oblikom heteroskedastičnosti je sledeća:

1. Primenjuje se metoda najmanjih kvadrata na

$$e_i^2 = a + bX_i + cX_i^2 + u_i,$$

gde su e_i reziduali regresije Y na X dobijeni metodom najmanjih kvadrata i gde je $u_i = e_i^2 - \sigma_i^2$. Na taj način se dobijaju ocenjivači „prve runde“ od a, b i c i oni su označeni sa $\hat{a}, \hat{b}$ i $\hat{c}$. Odgovarajući ocenjivač prve runde za σ_i^2 je

$$\hat{\sigma}_i^2 = \hat{a} + \hat{b}X_i + \hat{c}X_i^2.$$

2. Ocenjivači prve runde od a, b i c nisu asimptotski efikasni jer je u_i heteroskedastična. Stoga, primenom metode najmanjih kvadrata na

$$\frac{e_i^2}{\hat{\sigma}_i^2} = a\frac{1}{\hat{\sigma}_i^2} + b\frac{X_i}{\hat{\sigma}_i^2} + c\frac{X_i^2}{\hat{\sigma}_i^2} + u_i^*$$

dobijaju se ocenjivači „druge runde“ od a, b i c, označeni sa $\tilde{a}$,$\tilde{b}$, i $\tilde{c}$ koji su asimptotski efikasni. Odgovarajući ocenjivač druge runde za σ_i^2 je tada

$$\tilde{\sigma}_i^2 = \tilde{a} + \tilde{b}X_i + \tilde{c}X_i^2.$$

3. Primenjuje se metoda najmanjih kvadrata na

$$\frac{Y_i}{\tilde{\sigma}_i^2} = \alpha\frac{1}{\tilde{\sigma}_i^2} + \beta\frac{X_i}{\tilde{\sigma}_i^2} + \varepsilon_i^*.$$

Problem ovog iterativnog postupka je u tome što se ne garantuje da će ocene $\tilde{\sigma}_i^2$ biti pozitivne, što znači da $\tilde{\sigma}_i$ možda ne postoje[12].

- heteroskedastičnost zavisne promenljive: $\sigma_i^2 = \sigma^2[E(Y_i)]^2$

Funkcija verodostojnosti u logaritamskom obliku sadrži tri nepoznata parametra, α, β i σ^2

$$L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n}\log(\alpha + \beta X_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\frac{Y_i - \alpha - \beta X_i}{(\alpha + \beta X_i)}\right]^2.$$

Dobijanje vrednosti tih parametara nije jednostavno za izvođenje. Izračunavanje se može pojednostaviti ako se primeni metoda najmanjih kvadrata na

$$\frac{Y_i}{\hat{Y}_i} = \alpha \frac{1}{\hat{Y}_i} + \beta \frac{X_i}{\hat{Y}_i} + \varepsilon_i^*,$$

gde su $\hat{Y}_i$ predviđene vrednosti od Y iz regresije Y na X dobijene MNK. Taj postupak potiče od toga što je $\hat{Y}_i$ i nepristrasan i konzistentan ocenjivač od $E(Y_i)$.

Kod transformacija koje se zasnivaju na metodi ponderisanih najmanjih kvadrata postoji problem mogućeg pojavljivanja korelacije npr. između promenljivih Y_i/X_i i $1/X_i$ iako u polaznom modelu promenljive Y_i i X_i ne moraju biti korelisane.

- Logaritamska transformacija $\ln Y_i = \alpha + \beta \ln X_i + \varepsilon_i$ veoma često redukuje heteroskedastičnost u odnosu na linearnu regresiju $Y_i = \alpha + \beta X_i + \varepsilon_i$.

Ova transformacija je popularna u ekonometrijskim istraživanjima prvenstveno zbog bolje ekonomske interpretacije. Naime, koeficijent pravca β će meriti elastičnost Y u odnosu na X, tj. kolika je procentualna promena Y za procentualnu promenu X. Međutim, ukoliko je neko od Y ili X negativno, tada logaritamska forma neće biti odgovarajuća.

Ukoliko se ne može pretpostaviti oblik heteroskedastičnosti, predlaže se korišćenje tzv. Vajtove korekcije u izračunavanju standardnih grešaka ocena parametara, koja obezbeđuje njihovu efikasnost i konzistentnost[5]. Korekcija se zasniva na ocenama varijansi svake slučajne greške, $\sigma_i^2, i = 1, \dots, n$ u

$$Var(\hat{\beta}) = \frac{\sum (X_i - \bar{X})^2 \, \sigma_i^2}{S_{XX}^2}.$$

Kao ocene koriste se kvadrati reziduala svake opservacije, $e_i^2, i = 1, \dots, n$. Na taj način se dobija korigovana ocena varijanse ocene β:

$$S_{K\hat{\beta}}^2 = \frac{\sum (X_i - \bar{X})^2 \, e_i^2}{S_{XX}^2}.$$

Kvadratni koren nove ocene, $S_{K\hat{\beta}}$, označava korigovanu standardnu grešku ocene β. Takođe se koristi termin - robusna standardna greška. Ukoliko se koriguju standardne greške ocena modela, neće biti potrebe za prethodno objašnjenim transformacijama polaznog modela. Međutim, ova korekcija se može upotrebiti da bismo, barem asimptotski, izbegli problem pristrasnog ocenjivanja varijansi regresijskih koeficijenta dobijenih MNK. Ipak, za uzorke male i srednje veličine, ocene koeficijenata mogu biti pristrasne i imati tendanciju potcenjivanja stvarne nepoznate varijanse. Ukoliko je obim uzorka manji od 250, nije preporučljivo da se koristi ova metoda.

Ukoliko imamo heteroskedastični model višestruke regresije

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i, \;\; i = 1,2, \dots n,$$

metoda otklanjanja heteroskedastičnosti je slična i polazni model se transformiše kao i kod proste linearne regresije, kada je σ_i^2 poznato. Međutim, često se ne zna a priori koja od slučajnih promenljivih treba izabrati za transformaciju modela. Ispitivanjem veze između svake od promenljivih i kvadriranih reziduala polazne višestruke regresije, e_i^2, može se izabrati jedna od promenljivih.

Ukoliko oblik heteroskedastičnosti nije poznat, tada se koristi metoda heteroskedastične konzistentne kovarijansne matrice (HCCM) koja računa konzistentne ocene koeficijenata u prisustvu heteroskedastičnosti[10].

Radi jednostavnosti, koristićemo matrični zapis višestruke regresije

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

gde je $\boldsymbol{y}$ vektor n zavisnih promenljivih, $\boldsymbol{X}$ matrica nezavisnih promenljivih dimenzije $n \times (k + 1)$ i $\boldsymbol{\beta}$ vektor $k + 1$ koeficijenata. Varijansa ocenjene vrednosti $\boldsymbol{\beta}$ se može predstaviti na sledeći način:

$$Var(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

gde je $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, a $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \dots, \omega_n)$, $\omega_i = Var(\varepsilon_i), \; i = 1,2, \dots, n.$

Vajtova korekcija kod višestruke regresije se još naziva HC0 i glasi:

$$\mathrm{HC0} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

gde je $\widehat{\boldsymbol{\Omega}} = \mathrm{diag}(e_1^2, \dots, e_n^2)$.

Kako bi bile unapređene osobine ocenjenih parametara u manjim uzorcima, predloženi su alternativni ocenjivači HC1, HC2 i HC3 kao korekcije HC0.

Prvu korekciju predložio je Hinkli tako što je HC0 pomnožio sa $\frac{n}{n-k}$, odnosno:

$$\mathrm{HC1} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{diag}\left[\frac{ne_i^2}{n-k}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

Pod pretpostavkom homoskedastičnosti važi da je $E(\boldsymbol{e}'\boldsymbol{e}) = \sigma^2\boldsymbol{M}$, gde je $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{H}$, $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Tada je $E(e_i^2) = \sigma^2 m_{ii}$, gde su m_{ii} dijagonalni elementi matrice $\boldsymbol{M}$. Kako je e_i^2 pristrasna ocena za σ_i^2, druga ideja korekcije HC0 se zasniva na $e_i^2/1 - h_{ii}$ kao nepristrasnoj ocena za σ_i^2:

$$\mathrm{HC2} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{diag}\left[\frac{e_i^2}{1 - h_{ii}}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

gde su h_{ii} dijagonalni elementi matrice $\boldsymbol{H}$.

Treća verzija korekcije HC0, koju autori predlažu da se koristi ukoliko je $n < 250$ je

$$\mathrm{HC3} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{diag}\left[\frac{e_i^2}{(1 - h_{ii})^2}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

Novija predložena korekcija koja podrazumeva različite efekte $\boldsymbol{H}$ matrice je

$$\mathrm{HC4} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{diag}\left[\frac{e_i^2}{(1 - h_{ii})^{\delta_i}}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1},$$

gde je $\delta_i = \min\{4, h_{ii}/\bar{h}\}$, a $\bar{h}$ srednja vrednost $h_{ii}, i = 1,2, \dots n$. [7]

Svi ovi HCCM ocenjivači, bazirani na metodi najmanjih kvadrata, su konzistentni u uslovima heteroskedastičnosti, ali se ova osobina može izgubiti ukoliko imamo outlajere. Kao što je već napomenuto, ova motada otklanjanja heteroskedastičnosti je preporučljiva kada imamo velike uzorke.

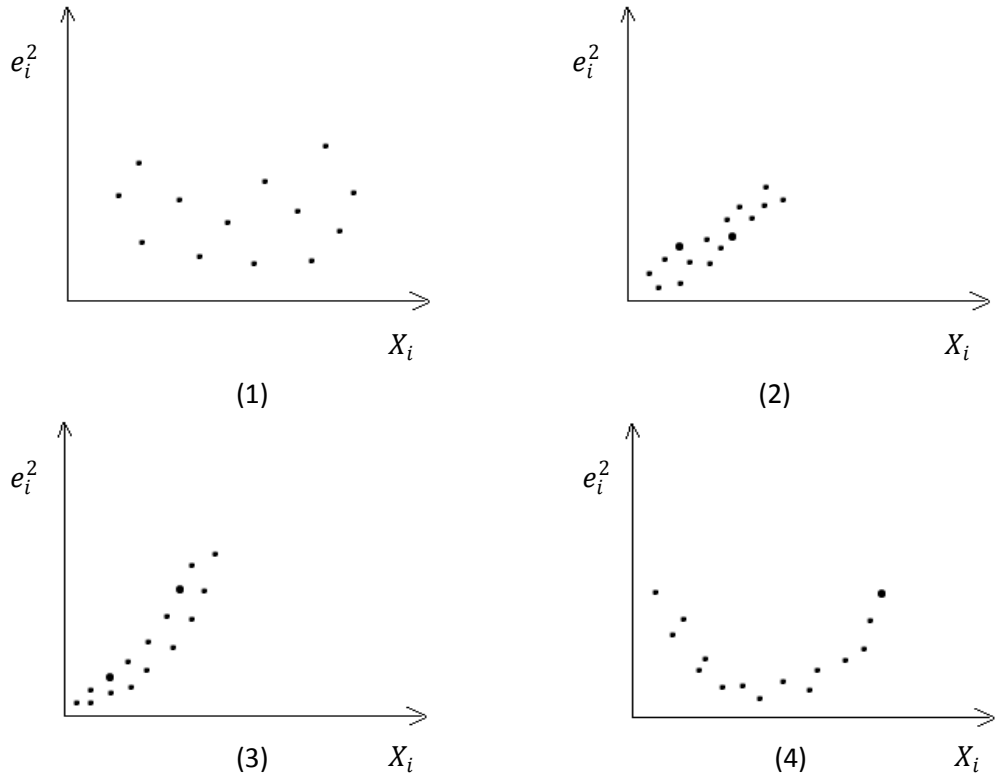
3. Testiranje heteroskedastičnosti

Najčešće nije poznata prava priroda heteroskedastičnosti, pa izbor odgovarajućeg testa zavisi od prirode podataka. Ali kako visina variranja grešaka oko srednje vrednosti po pravilu zavisi od visine nezavisnih promenljivih, svi testovi se oslanjaju na ispitivanje da li je varijansa greške neka funkcija od regresora.

3.1 Grafička metoda

Jedan od najjednostavnijih metoda za ispitivanje postojanja heteroskedastičnosti sastoji se u vizuelnom pogledu reziduala ocenjenog modela. Uobičajeno je da se formiraju dijagrami rasturanja tačaka reziduala e_i, ili njihove apsolutne vrednosti, $|e_i|$, i nezavisne promenljive X_i. Budući da je varijansa slučajne greške $E(\varepsilon_i^2)$, postoji mišljenje da na dijagramu rasturanja tačaka vrednosti reziduala treba zameniti njihovim kvadratom, e_i^2.

Na bazi dijagrama rasturanja tačaka donosimo zaključak o tome da li heteroskedastičnost postoji, i ako postoji u kojoj formi se javlja, tj.kako se generiše varijansa slučajne greške. Na slici 3.1. predstavljeni su grafici nekih od mogućih dijagrama rasturanja tačaka [5].



Slika 3.1 *Dijagrami rasturanja tačaka*

Prvi grafik odgovara modelu u kojem ne postoji sistematska zavisnost između varijansi slučajnih grešaka i nezavisne promenljive X_i. Može se smatrati da su u takvom modelu slučajne greške homoskedastične. Ostali grafici pokazuju pravilnost u položaju tačaka na dijagramu rasturanja, sugerišući moguću heteroskedastičnost. Drugi grafik ukazuje na linearnu zavisnost, dok treći i četvrti grafici predastavljaju zavisnost izraženu kvadratnom formom, u smislu da je varijansa slučajne greške korelisana sa X_i^2.

U slučaju analize višestrukog regresionog modela dati grafici mogu se obrazovati za svaku od nezavisnih promenljivih, kako bi se ustanovilo sa kojom od njih je varijansa slučajne greške u najvećem stepenu slaganja. Alternativno, moguće je vrednost e_i^2 staviti u odnos $\hat{Y}_i$, gde $\hat{Y}_i$ možemo posmatrati kao linearnu kombinaciju svih nezavisnih promenljivih. Ispoljena pravilnost položaja tačaka na dijagramu rasturanja u ovom slučaju ukazuje na to da je heteroskedastičnost posledica prirode podataka iz uzorka.

Tsung-Chi Cheng [7] je predložio korišćenje *Jigsaw* grafika za istovremeno identifikovanje outlajera i heteroskedastičnosti bez specificirane forme. Problem je u prepoznavanju zapažanja koji povećavaju varijansu i pripadaju outlajerima od onih koji se prepisuju heteroskedastičnosti, kao posledica same strukture podataka. Grafik je konstruisan tako da sadrži koverte (eng. *envelopes*) koje predstavljaju granice za identifikovanje outlajera, pristupom zasnovanom na metodi ponderisanih najmanjih apsolutnih odstupanja (PNAO).

Za višestruku linearnu regresiju u matričnom obliku, $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon},$ gde se matrica $\boldsymbol{X}$ može zapisati kao $\boldsymbol{X} = (\boldsymbol{x}_1', \dots, \boldsymbol{x}_n')$ i za koju važe sve standardne pretpostavke osim homoskedastičnosti, definišemo

$$RD(\boldsymbol{x}_i) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{t})'\boldsymbol{C}^{-1}(\boldsymbol{x}_i - \boldsymbol{t})} \quad i = 1, \dots n.$$

$\boldsymbol{C}$ predstavlja kovarijansnu matricu dobijenu metodom minimalne kovarijansne determinante (MCD), a $\boldsymbol{t}$ je vektor opservacija koje odstupaju od ostalih u uzorku, tzv. robusne lokacije. MCD je najčešće korišćen robusni matrični ocenjivač. Za matricu $\boldsymbol{X}$ podrazumeva traženje podskupova podataka date veličine i minimizira determinantu odgovarajuće kovarijansne matrice. Kao rezultat, dobija se matrica $\boldsymbol{C}$.

Metod PNAO podrazumeva ocenjivanje $\boldsymbol{\beta}$:

$$\min_{\beta} \sum_{i=1}^{n} \omega_i |r_i(\boldsymbol{\beta})|,$$

gde je $r_i(\boldsymbol{\beta}) = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$, $\omega_i = \min\left\{1, \frac{k}{(RD(\boldsymbol{x}_i))^2}\right\}$ $i = 1, \dots, n$, i k je broj nezavisnih promenljivih. Standardizovani reziduali se mogu definisati kao

$$t_i = \frac{r_i}{\hat{\sigma}} \tag{3.1}$$

gde je $\hat{\sigma} = 1.4826\,\text{median}|r_i|$. Konstanta 1.4826 dovodi do konzistentne ocene standardnog odstupanja, pod pretpostavkom normalnosti greške [18]. Predloženo je korišćenje

$$\hat{\sigma}^* = 1.4826 s^* \tag{3.2}$$

zbog mogućeg postojanja nula rezidula ($r_i = 0$), gde je s^* mediana apsolutnih nenula reziduala, tako da je

$$t_i^* = \frac{r_i}{\hat{\sigma}^*}. \tag{3.3}$$

Ukoliko je $|t_i^*| > 2.5$ tada je opažanje outlajer. Međutim, te vrednosti nisu verodostojne ako imamo heteroskedastičnost, tako da *Jigsaw* grafik sadrži simulirane granice, koje formiraju tzv. koverte, za identifikaciju outlajera i heteroskedastičnosti. Odgovarajuća promenljiva se generiše iz normalne raspodele sredine 0 i varijanse

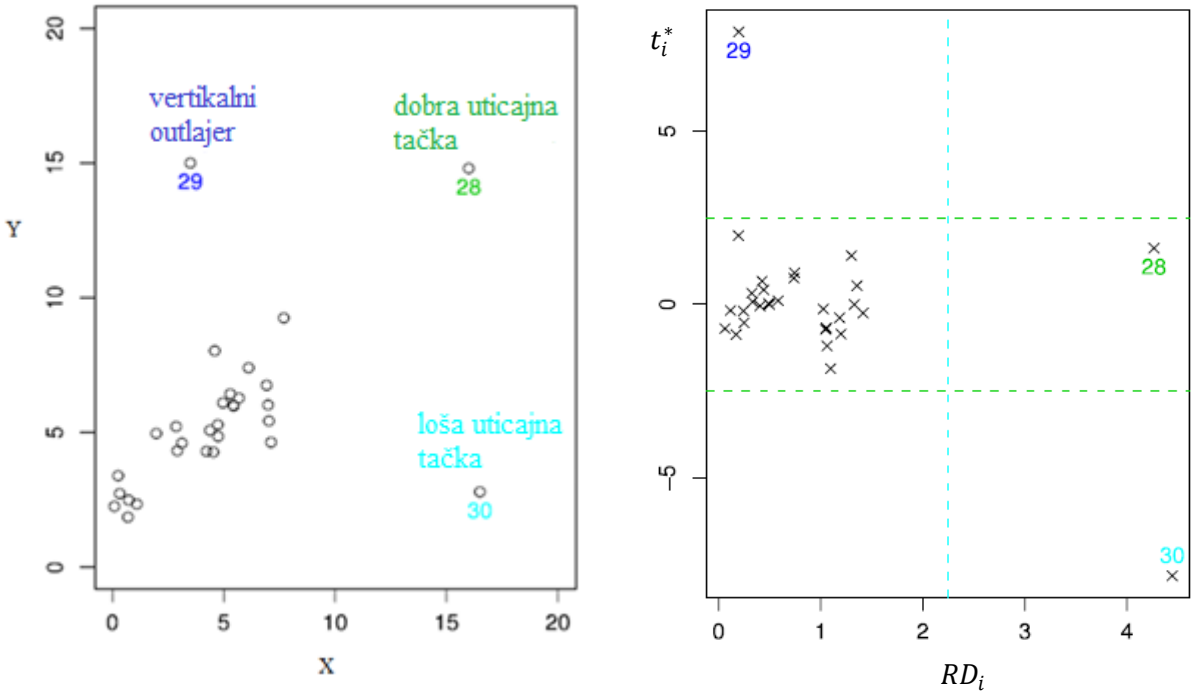
$$\sigma^2\left\{(1 - h_{ii})\omega_i + \sum_{k \neq i} \frac{\omega_i h_{ik}}{(1 - h_{ii})}\right\}, \tag{3.4}$$

gde su h_{ik} elemeni $\boldsymbol{H}$ matrice. Ako sa $t_{m(i)}^*$, $i = 1, \dots, n$ označimo vrednost standardizovanog reziduala m-te simulacije i-tog opažanja, tada se simulacija može ponavljati fiksiran broj puta. Simulirane granice definisane kao

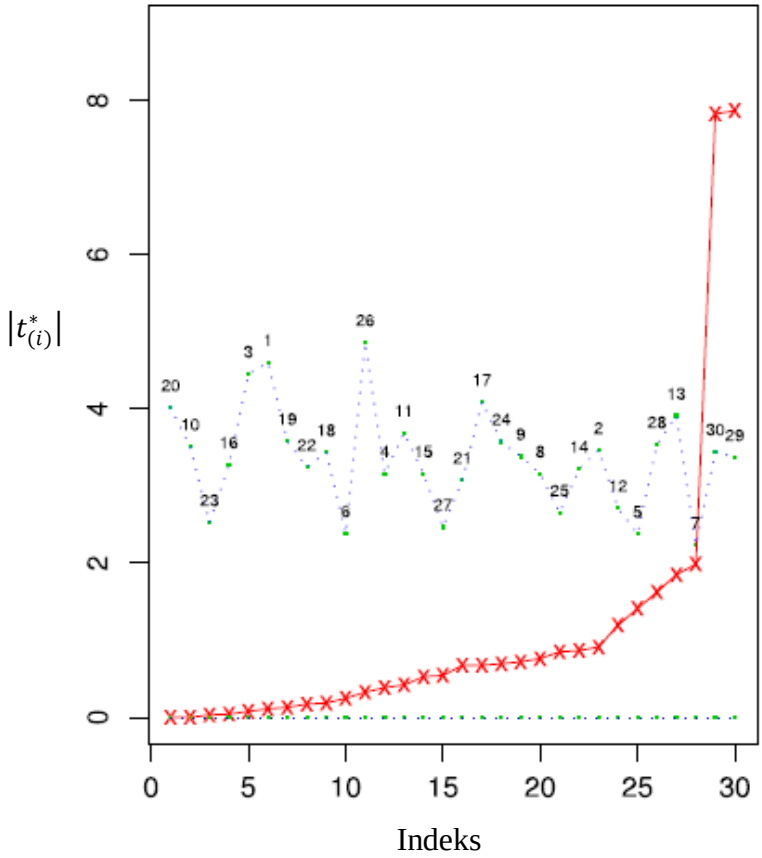
$$t_{l(i)}^* = \min_m \left|t_{m(i)}^*\right|,$$

$$t_{u(i)}^* = \max_m \left|t_{m(i)}^*\right|,$$

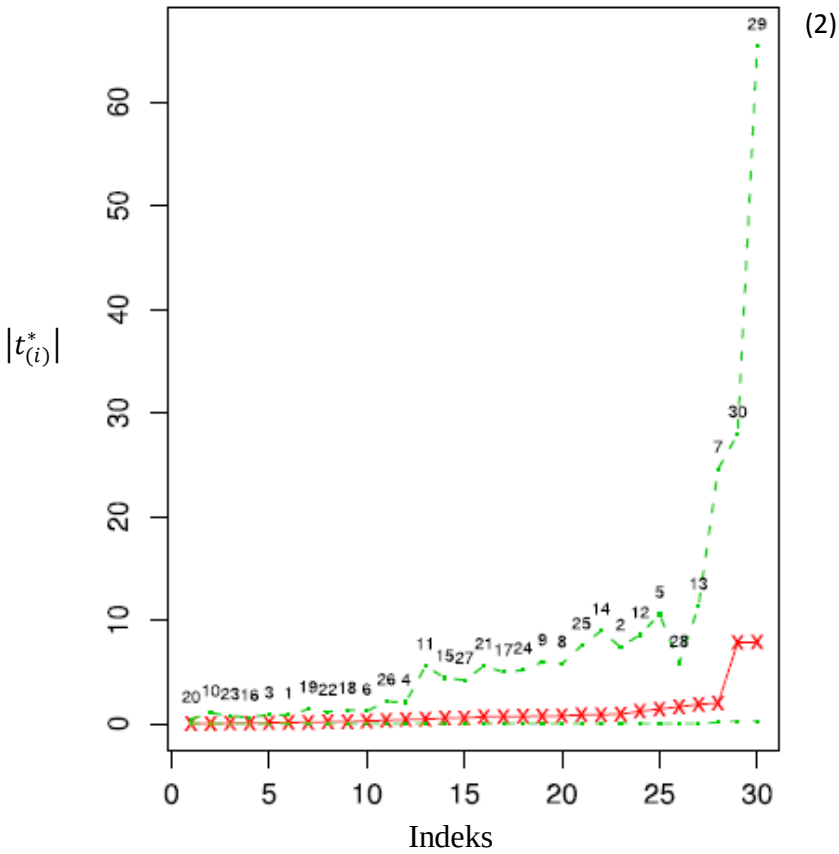
formiraju najnižu i najvišu kovertu, respektivno. Dve vrste ocenjivanja se koriste za σ u izrazu (3.4) da bi se utvrdile koverte. Prva je $\hat{\sigma}^*$ iz (3.2), a druga je standardizovani rezidual iz (3.3) za i-to posmatranje. Prva daje granicu za identifikovanje outlajera, a druga reflektuje heteroskedastičnost greške za svaku opservaciju.



Slika 3.2: [7] *Grafik rasturanja tačaka i dijagnostički grafik*



(1)



Slika 3.3: [7] *Jigsaw grafik koristeći* (3.2) *i Jigsaw grafik koristeći* (3.3)

Na slici 3.2, prvi grafik je grafik rasturanja tačaka koji sadrži 30 simululiranih podataka, koje su klasifikovane kao regularne tačke, vertikalni outlajeri, dobre uticajne tačke (*eng. good leverage point*) i loše uticajne tačke (*eng. bad leverage point*). Drugi grafik je dijagnostički grafik koji koristi PNAO ocene i raspoređuje opažanja u odgovarajuće oblasti. Linije koje razdvajaju te oblasti su ± 2.5 i $\sqrt{\chi^2_{p,0.975}}$ (ovde je $p = 1$) horizontalne i vertikalne linije, respektivno. Grafici na slici 3.3 odgovaraju *Jigsaw* graficima apsolutnih vrednosti reziduala označenih sa $\times$ sa kovertama koje su generisane koristeći (3.2) i (3.3) za σ, respektivno. Prvi grafik obezbeđuje odgovarajuće granice za svako posmatranje, identifikujući 29. i 30. opažanje kao slučajeve outlajera, dok drugi grafik prikazuje opažanja koja prepisujemo heteroskedastičnosti, odnosno regularne tačke koje dovode do većih vrednosti gornje koverate ukazuju na izvor heteroskedastičnosti.

.

3.2 Sprovođenje testova

Grafički metodi predstavljaju samo sredstvo preliminarne analize. Da bi se dobio precizniji odgovor na pitanje da li je heteroskedastičnost prisutna ili ne, potrebno je koristiti odgovarajuće testove.

3.2.1 Goldfeld-Kvantov (Goldfeld-Quandt) test

Jedan od najranijih, koji je vrlo jednostavan i intuitivan i koji se često koristi je Goldfeld-Kvantov test. Ovim testom proverava se nulta hipoteza o konstantnosti slučajne greške protiv alternativne da je varijansa slučajne greške linearna funkcija nezavisne promenljive. Pretpostavlja se da je slučajna greška neautokorelisana i normalne raspodele. Postupak sprovođenja testa je sledeći:

- Opažanja iz uzorka se uređuju prema rastućim vrednostima nezavisne promenljive
- Iz skupa od n opažanja izostavlja se c centralnih opažanja, tako da se dalja analiza zasniva na dva skupa opažanja: prvih $\frac{n-c}{2}$ i poslednjih $\frac{n-c}{2}$ opažanja gde je potrebno obezbediti da je $\frac{n-c}{2} > k$, a k je broj ocenjenih parametara
- Pojedinačno ocenjujemo dve regresije na osnovu prvih $\frac{n-c}{2}$ i poslednjih $\frac{n-c}{2}$ opažanja. Dobijene sume kvadrata reziduala su označene sa $\sum e_1^2$ i $\sum e_2^2$ ($\sum e_1^2$ odgovara regresiji sa nižim, $\sum e_2^2$ a regresiji sa višim vrednostima nezavisne promenljive.)

Homoskedastičnost slučajne greške podrazumeva isti stepen variranja u dva podskupa opservacija, što se manifestuje približno istim vrednostima varijabilnih suma $\sum e_1^2$ i $\sum e_2^2$. U tom slučaju količnik ove dve sume je blizak vrednosti jedan. Suprotno, postojanje heteroskedastičnosti, ima za posledicu veću vrednost rezidualne sume $\sum e_2^2$. Smisao testa je da proveri da li je $\sum e_2^2 / \sum e_1^2$ statistički značajno različit od jedan.

Pod pretpostavkom da je tačna nulta hipoteza o konstantnoj varijansi, važi sledeće:

$$\frac{\sum e_1^2}{\sigma^2} : \chi^2_{\frac{n-c-2k}{2}}$$

$$\frac{\sum e_2^2}{\sigma^2} : \chi^2_{\frac{n-c-2k}{2}},$$

gde je k broj parametara za ocenjivanje u polaznom modelu. Odnos:

$$\frac{\sum e_1^2}{\sum e_2^2},$$

poseduje F- raspodelu sa $\frac{n-c-2k}{2}$ i $\frac{n-c-2k}{2}$ stepeni slobode.
Prema tome, Goldfeld-Kvantova test statistika je oblika:

$$F = \frac{\sum e_1^2}{\sum e_2^2}.$$

Ukoliko je izračunata vrednost F-statistike veća od odgovarajuće kritične vrednosti na datom nivou značajnosti, zaključujemo da u modelu postoji heteroskedastičnost.

Ovaj test se primenjuje na velike uzorke, sa malim brojem parametara za ocenjivanje. Problem koji se javlja u primeni ovog testa je u tome što treba odlučiti koliko centralnih opservacija eliminisati iz analize. Goldfeld-Kvantov test je egzaktan ali ne i jako moćan, (odnosno ima veliku verovatnoću prihvatanja H_0 kada je pogrešna), kad su odstupanja heteroskedastična i kad njihova varijansa iz prvog dela uzorka nije mnogo različita od varijanse iz drugog dela uzorka. Kako jačina testa, ili moć razgraničenja između dve alternativne hipoteze, zavisi od broja c, treba ga pažljivo izabrati. Njihovim izostavljanjem razlika između dve rezidualne sume kvadrata postaje izraženija. Za visoke vrednosti c moć testa je manja, ali za niske vrednosti c dve sume rezidualnih kvadrata imaju tendenciju izjednačenja. Ne postoji jedinstveno pravilo o načinu određivanja vrednosti c. Eksperimentalni rezultati pokazuju da je razumno izostaviti oko jednu šestinu srednjih opažanja. Takođe varijansa mora biti monotona funkcija nezavisne promenljive. Npr. kada smo suočeni sa kvadratnom funkcijom koja povezuje nezavisnu promenljivu sa varijansom greške, Goldfeld-Kvantov test može nepropisno prihvatiti nultu hipotezu homoskedastičnosti grešaka.

U slučaju da postoji više regresora u modelu, kao prvi korak ovog testa, može da se obavi na bazi bilo kog regresora [4]. Ako nismo sigurni a priori koja je nezavisna promenljiva odgovarajuća, test se može izvesti za svaku od njih. Ipak, kako bi se moglo ocenjivati da li variranje reziduala zavisi od svih regresora u modelu, pogodnije je koristiti neki drugi test.

3.2.2 Brojš-Pagan- Godfri (Breusch-Pagan-Godfrey) test[2]

Savremeniji test, polazi od višestruke linearne regresije i zasniva se na testiranju zavisnosti variranja reziduala od visine svih regresora. Temelji se na zamisli da se ocene regresijskih

[2] T.S.Breusch i A.R.Pagan su o ovom testu objavili rad 1979. god., a L.G.Godfrey 1978. god., pa se zbog sličnosti ovi testovi često zovu prema sva tri autora.

koeficijenata dobijene metodom najmanjih kvadrata ne bi trebalo značajno razlikovati od maksimalno verodostojnih ocena, ako je hipoteza o homoskedastičnosti istinita. Konkretno, kad je L funkcija verodostojnosti koja uzima u obzir heteroskedastičnost, prvi izvodi funkcije L su jednaki nuli, kada se nepoznati parametri zamene njihovim maksimalno verodostojnim ocenama. Ako se, umesto toga, nepoznati parametri zamene ocenama dobijene metodom najmanjih kvadrata i ako su odstupanja uistinu homoskedastična, tada se prvi izvodi funkcije L ne bi trebalo značajno razlikovati od nule.

Neka je dat model linearne regresije:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i \qquad i = 1,2 \ldots, n \tag{3.5}$$

U Brojš-Paganovoj formulaciji testa, testira se hipoteza o homoskedastičnosti nasuprot alternativnoj:

$$H_A\colon\ \sigma_i^2 = g\left(\gamma_0 + \gamma_1 Z_{i1} + \gamma_2 Z_{i2} + \cdots + \gamma_p Z_{ip}\right) \qquad i = 1,2, \ldots, n, \tag{3.6}$$

gde je g neprekidan funkcija koja ima neprekidne prve izvode po svim promenljivama. Promenljive Z su neke poznate promenljive koje utiču na varijansu greške. One će obično biti jednake nezavisnim promenljivama regresijske jednačine ili neke poznate funkcije nezavisnih promenljivih kao što su x, x^2, e^x ... Test ne zavisi od funkcionalne forme. Funkcija g je dovoljno uopštena tako da uključuje multiplikativnu heteroskedastičnost

$$\log \sigma_i^2 = \log\gamma_0 + \gamma_1 \log Z_{i1} + \gamma_2 \log Z_{i2} + \cdots + \gamma_p \log Z_{ip},$$

i aditivnu heteroskedastičnost

$$\sigma_i^2 = \gamma_0 + \gamma_1 Z_{i1} + \gamma_2 Z_{i2} + \cdots + \gamma_p Z_{ip}$$

kao specijalne slučajeve.

Test se sprovodi u nekoliko koraka:

- Model (3.5) oceniti metodom najmanjih kvadrata da se dobije serija reziduala
- Izračunati ocenjenu vrednost varijanse na sledeći način:

$$\hat{\sigma}^2 = \sum \frac{e_i^2}{n},$$

  a to je ocena varijanse dobijena metodom maksimalne verodostojnosti.
- Kvadrirati reziduale i podeliti ocenjenom varijansom:

$$p_i = \frac{e_i^2}{\hat{\sigma}^2}$$

- Primeniti metodu najmanjih kvadrata na novu jednačinu linearne regresije:

$$p_i = \gamma_0 + \gamma_1 Z_{i1} + \gamma_2 Z_{i2} + \cdots + \gamma_p Z_{ip} + v_i, \tag{3.7}$$

  gde su Z_{ij} promenljive iz (3.6).
- Formirati statistiku testa kao polovinu regresijske sume kvadrata modela (3.7), označenu sa SSR_{BP}, koja pod uslovima normalne distribucije gršaka ima χ^2 raspodelu:

$$\frac{SSR_{BP}}{2} \sim \chi_p^2,$$

  sa brojem stepeni slobode jednakim broju ocenjenih parametara. Veće vrednosti test statistike od tablične, vode odbacivanju nulte hipoteze.

Da bismo pokazali da Brojš-Paganova test statistika ima χ^2 raspodelu sa p stepeni slobode, pokazaćemo da je to ustvari LM test statistika, gde je $LM = nR^2$, koja ima χ^2 raspodelu [8]. Neka je R^2 koeficijent determinacije u regresije od e_i^2 i $Z_{i1}, Z_{i2}, .., Z_{ip}$. Tada važi:

$$nR^2 = \frac{S}{Var(e_i^2)},$$

gde je S regresijska suma kvadrata u regresiji od e_i^2 i $Z_{i1}, Z_{i2}, .., Z_{ip}$. Pod pretpostavkom da je tačna nulta hipoteza o homoskedastičnosti, ε_i^2/σ^2 ima χ^2 raspodelu sa jednim stepenom slobode. Takođe, $Var(\varepsilon_i^2/\sigma^2) = 2$ (varijansa sličajne promenljive koja ima χ_m^2 raspodelu je $2m$). Onda je

$$Var(\varepsilon_i^2) = 2\sigma^4$$

Za velike uzorke važi da je $Var(e_i^2) = Var(\varepsilon_i^2)$ tj. $\hat{\sigma}^4 = \sigma^4$, tako da je $nR^2 = S/2\hat{\sigma}^4$ i ima χ_p^2. Slično se može izvesti za $SSR_{BP}/2$, jer je $nR^2 = SS/Var(\varepsilon_i^2/\sigma^2) = SSR_{BP}/2$, gde je SS regresijska suma kvadrata u regresiji od $e_i^2/\hat{\sigma}^2$ i $Z_{i1}, Z_{i2}, .., Z_{ip}$.

Čini se da bi Brojš-Paganov bio razumno moćan test kad je heteroskedastičnost prisutna. Međutim, kod malih uzoraka je postavljen nivo značajnosti samo gruba naznaka stvarnog nivoa. Test je kritikovan zbog toga što je vrlo osetljiv na neznatna narušavanja pretpostavke o normalnosti regresijskog odstupanja. Ta se zavisnost o normalnosti može otkloniti malom

promenom test veličine. Doduše, proizilazi da nakon te prepravke i odgovarajuće specifikacije promenljivih Z, Brojš-Paganov test postaje jednak testu koji je preporučio H.Vajt.

3.2.3 Vajtov (White) test

Test se bazire na poređenju varijanse ocenjivača dobijenih metodom najmanjih kvadrata u uslovima homoskedastičnosti i heteroskedastičnosti. Ako je nulta hipoteza tačna, dve bi se ocenjene varijanse razlikovale samo zbog kolebanja u uzorku. Nulta hipoteza o homoskedastičnosti slučajne greške testira se protiv široko postavljene alternativne hipoteze da je varijansa slučajne greške zavisna od objašnjenih promenljivih, njihovih kvadrata i međuproizvoda tj. ispituje se variranje reziduala pod kombinovanim dejstvom regresora.

Vajtov test se sastoji od sledećih koraka:

- Model $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i$ oceniti da se dobije serija reziduala e_i odnosno njihove kvadrirane vrednosti
- Oceniti pomoćnu regresiju u kojoj su kvadrati reziduala funkcije svih regresora modela, njihovih kvadrata i međuproizvoda, odnosno primeniti metod najmanjih kvadrata na

$$e_i^2 = \delta_0 + \delta_1 Z_{i1} + \delta_2 Z_{i2} + \cdots + \delta_p Z_{ip} + u_i, \quad i = 1,2, \dots n$$

  gde je za prostu regresiju $p = 2$, $Z_{i1} = X_i$ i $Z_{i2} = X_i^2$, pa se test zasniva na analizi modela

$$e_i^2 = \delta_0 + \delta_1 X_i + \delta_2 X_i^2 + u_i, \quad i = 1,2, \dots n.$$

  Značajan uticaj nezavisnih promenljivih X_i i X_i^2 na e_i^2 ima za rezultat visoku vrednost koeficijenta determinacije R^2.

  Za model višestruke regresije sa dve nezavisne promenljive X_{i1} i X_{i2}, specifikacija je $p = 5$, $Z_{i1} = X_{i1}$, $Z_{i2} = X_{i2}$, $Z_{i3} = X_{i1}X_{i2}$, $Z_{i4} = X_{i1}^2$ i $Z_{i5} = X_{i2}^2$. Zbog mogućeg velikog gubitka stepeni slobode, moguće je koristiti umesto individualnih vrednosti regresora, njihove linearne kombinacije: $\hat{Y}_i, \hat{Y}_i^2$.
- Na bazi vrednosti koeficijenta determinacije iz pomoćne regresije, R_W^2, formirati Vajtov test nR_W^2, gde je n obim uzorka. Asimptotski, u uslovima nulte hipoteze o homoskedastičnosti, test statistika nR_W^2 teži χ^2 raspodeli sa brojem stepeni slobode jednakom broju regresora u pomoćnoj regresiji:

$$nR_W^2 \sim \chi_p^2$$

- Ako je izračunata vrednost statistike testa veća od tablične vrednosti, dakle ako je dovoljno visok koeficijent determinacije u pomoćnoj funkciji kvadrata reziduala, odbacuje se hipoteza o homoskedastičnosti.

Kao ni u Brojš-Paganovom testu, ni u Vajtovom se testu ne zahteva specifikacija oblika heteroskedastičnosti. Za razliku od prethodnog testa, Vajtov test nije osetljiv na odstupanje grešaka od normalnosti i jednostavniji je pa se i češće upotrebljava u testiranju heteroskedastičnosti. U slučaju da postoji više regresora, uvođenje kvadrata i svih međuproizvoda u pomoćnu regresiju može značiti veliki gubitak u broju stepeni slobode. Zato se često Vajtov test izvodi bez međuproizvoda.

Ako rezltat testa pokaže značajno visoku statistiku, razlog ne mora biti heteroskedastičnost, već greška u specifikaciji modela (izostavljeni regresor). Dakle, Vajtov test se može koristiti u otkrivanju heteroskedastičnosti ili greške specifikacije.

Treba napomenuti i to da je Vajtov test često neuspešan za male obime uzoraka. Takođe, ako rezultat testa govori da treba da odbacimo nultu hipotezu, nemamo informaciju o tome šta je prouzrokovalo heteroskedastičnost. To može dovesti do problema pri pokušaju konstruisanja generalizovane metode najmanjih kvadrata za ocenjivanje parametara, za koji je potrebno znati oblik greške.

3.2.4 Glejzerov (Glejser) test

Primena ovog testa ne zahteva apriorno poznavanje prirode heteroskedastičnosti, već se do nje dolazi u toku samog testiranja. Postupak sprovođenja testa je sledeći:

- Ocenjuje se polazna regresija $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i$ metodom najmanjih kvadrata i računaju se reziduali e_i
- Ocenjuje se sledeća regresija:

$$|e_i| = \delta_0 + \delta_1 X_i^h + u_i.$$

Parametru h najčešće se dodeljuju vrednosti: $1, -1$, i $1/2$, tako da se ocenjuju regresije:

$$|e_i| = \delta_0 + \delta_1 X_i + u_i,$$

$$|e_i| = \delta_0 + \delta_1 / X_i + u_i,$$

$$|e_i| = \delta_0 + \delta_1 \sqrt{X_i} + u_i.$$

- Testira se statistička značajnost ocene parametra δ_1 primenom *t*-testa.
- Upoređuju se koeficijenti determinacije dobijeni za različite vrednosti parametra h.

Statistička značajnost ocene δ_1 povlači za sobom zaključak o postojanju heteroskedastičnosti. Sam karakter heteroskedastičnosti se određuje prema regresiji sa najvećom vrednošću koeficijenta determinacije. Moć testa zavisi od prave forme heteroskedastičnosti. Međutim, ukoliko je pogrešno h izabrano, to neće ozbiljno umanjiti moć testa[3].

U svom istraživanju, Glejzer [9] je osim pomenutih regresija, ispitivao i

$$|e_i| = \delta_0 + \delta_1 \frac{1}{\sqrt{X_i}} + u_i,$$

$$|e_i| = \sqrt{\delta_0 + \delta_1 X_i} + u_i,$$

$$|e_i| = \sqrt{\delta_0 + \delta_1 X_i^2} + u_i.$$

Međutim, dodatni problem Glejzerove metode je nelinearnost parametara kao što je slučaj sa poslednje dve regresije, tako da se ne može primeniti metoda najmanjih kvadrata. Glejzer je zaključio da ukoliko imamo veliki obim uzorka, prve četiri navedene regresije daju generalno zadovoljavajuće rezultate u detektovanju heteroskedastičnosti.

3.2.5 Ostali testovi heteroskedastičnosti

Goldfeld-Kvantov, Brojš-Paganov, Vajtov i Glejzerov test su najviše razmatrani i upoređivani testovi heteroskedastičnosti. Navešćemo još neke testove koji nisu toliko popularni, ali koji mogu biti primenljivi u praksi.

1) Parkov (Park) test

Ovaj test je određeniji od Goldfeld-Kvantovog testa, odnosno, on povezuje reziduale specifične forme sa nezavisnom promenljivom za koju se pretpostavlja da je prouzrokovala heteroskedastičnost.

Za model višestruke linearne regresije, $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i,\ i = 1, \dots n,$ Parkov test podrazumeva sledeće korake:

[3] Ali i Giaccotto su 1984. godine potvrdili ovaj rezultat koristeći obimne Monte Carlo eksperimente.

- Oceniti linearnu regresiju u kojoj je zavisna promenljiva logaritam kvadriranih reziduala (e_i^2), a nezavisna promenljiva logaritam X_i tj.

$$\ln e_i^2 = \ln r_0 + r_1 \ln X_i + u_i,$$

  gde je X_i izabrana nezavisna promenljiva.
- U prethodnoj jednačini $\ln r_0$ je konstanta. Sprovesti t-test koeficijenta r_1 i ako je statistički značajan, heteroskedastičnost je prisutna.

Međutim, u slučaju da je rezidual e_i, povezan sa X_i u drugoj formi (drugačijoj od prethodne jednačine), test može pokazati da ne treba odbaciti nultu hipotezu, iako heteroskedastičnost postoji, tako da smo suočeni sa malom moći Parkovog testa.

2) Bartletov (Bartlett) test

Ovaj test se može izvesti ukoliko imamo ponovljenih opservacija i pogodan je za testiranje heteroskedastičnosti za koju nemamo specificiran oblik. Test se bazira na tome da se vrednost maksimalne funkcije verodostojnosti, koja se dobija uz pretpostavku homoskedastičnosti, ne bi značajno razlikovala od vrednosti koja se dobija uz pretpostavku heteroskedastičnosti.
Nulta hipoteza homoskedastičnosti za n opažanja i m različitih vrednosti za X je

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2.$$

Pod pretpostavkom nulte hipoteze, varijansa σ^2 može biti ocenjena grupnom varijansom

$$S^2 = \sum_{i=1}^{m} v_i s_i^2 / v,$$

gde je $v = \sum_{i=1}^{m} v_i$ i $v_i = n_i - 1$, $\sum_{i=1}^{m} n_i = n$, $s_i^2 = \sum (Y_{ij} - \bar{Y}_i)^2 / v_i$, $j = 1,2, \dots n_i$. Pod alternativnom hipotezom postoji m različitih varijansi ocenjenih sa s_i^2 za $i = 1,2, \dots, m$. Test verodostojnim odnosom (LR), koji računa odnos verovatnoća pod nultom i alternativnom hipotezom se svodi na

$$B = \left[v \log S^2 - \sum_{i=1}^{m} v_i \log s_i^2 \right] / c,$$

gde je $c = 1 + [\sum_{i=1}^{m} (1/v_i) - 1/v]/3(m-1)$. Pod nultom hipotezom B ima χ^2 raspodelu sa $m-1$ stepenom slobode. Velika p vrednost B statistike znači da ne odbacujemo homoskedastičnost, dok male p vrednosti obacuje H_0 u korist heteroskedastičnosti. U slučaju da ne postoje ponovljene opservacije, koriste se drugi testovi.

3) Spirmanov (Spearman) test

Zasniva se na rangiranju vrednosti nezavisne promenljive, X_i i apsolutnih vrednosti reziduala dobijenih metodom najmanjih kvadrata, $|e_i|$. Onda se računa razlika između tih rangova kao

$$d_i = rank(|e_i|) - rank(X_i).$$

Spirmanov korelacijski koeficijent je

$$r = 1 - \left[6\sum_{i=1}^{n} d_i^2 \Big/ (n^3 - n)\right].$$

Konačno, testira se hipoteza H_0 koja pretpostavlja da je koeficijent korelacije između rangova nula. Statistika ovog testa je

$$t = [r^2(n-2)/(1-r^2)]^{1/2}$$

koja ima studentovu raspodelu sa $n-2$ stepena slobode. Na osnovu p vrednosti odlučujemo da li odbacujemo ili ne odbacujemo homoskedastičnost.

4) Harvijev (Harvey) test

U slučaju da imamo multiplikativnu heteroskeastičnost, Harvey je predložio sledeći test koji oblik

$$\log \sigma_i^2 = \log\sigma^2 + \delta_1 \log Z_{i1} + \delta_2 \log Z_{i2} + \cdots + \delta_p \log Z_{ip}$$

modifikuje u

$$\log e_i^2 = \log\sigma^2 + \delta_1 \log Z_{i1} + \delta_2 \log Z_{i2} + \cdots + \delta_p \log Z_{ip} + v_i, \tag{3.8}$$

gde je $v_i = \log(e_i^2/\sigma_i^2)$. Ovaj izraz ima asimptotsku $\log\chi_1^2$ raspodelu. Na osnovu očekivane vrednosti i disperzije slučajne promenljive, koja je izražena kao logaritam količnika χ^2 raspodele i stepeni slobode, Harvi [19] je izračunao srednju vrednost i disperziju za v_i, koje redom iznose -1.2704 i 4.9348. Ocenio je regresiju (3.8) i testirao nultu hipotezu:

$$H_0\colon\ \delta_1 = \delta_2 = \cdots = \delta_p = 0.$$

Ako sa SSR_H označimo regresijsku sumu kvadrata, tada će Harvijeva test statistika

$$H = \frac{SSR_H}{4.9348}$$

imati asimptotsku χ^2 raspodelu sa p stepeni slobode.

5) Koanker-Basetov (Koenker–Bassett) test

Ovaj test je veoma jednostavan za primenu. Kao i neki prethodni testovi heteroskedastičnosti, i Koanker-Basetov je baziran na kvadriranim rezidualima e_i^2, u kome je slučajna promenljiva, ocenjena vrednost Y_i. Naime, neka imamo model regresije

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i \quad i = 1, \dots n. \tag{3.9}$$

Ocenimo regresiju metodom najmanjih kvadrata i ocenimo pomoćnu regresiju

$$e_i^2 = \gamma_1 + \gamma_2 \left(\hat{Y}_i\right)^2 + u_i,$$

gde je $\hat{Y}_i$ ocenjena vrednost Y_i iz modela (3.9). Tada je nulta hipoteza homoskedastičnosti $\gamma_2 = 0$. Ako se nulta hipoteza ne odbacuje, može se zaključiti da nije prisutna heteroskedastičnost. Hipoteza se može testirati uobičajenim t ili F testom. Druga prednost ovog testa je ta što je primenljiv čak i kada greške ε_i nemaju normalnu raspodelu i može se koristiti ako imamo jednu ili više nezavisnih promenljivih.

4. Simulacije

4.1 Multiplikativna heteroskedastičnost

U sledećoj tabeli prikazani su podaci, tako da simuliraju odstupanje oblika $\sigma_i^2 = 0.01X_i^2$. Populaciona prava je $Y = 2 + 3X$. U koloni $Y_i, i = 1, \dots, 30$, nalaze se vrednosti Y kojima su dodate greške ε_i.

br.	X_i	ε_i	Y_i	Y	br.	X_i	ε_i	Y_i	Y
1	10	-0,13677	31,86323	32	16	20	-6,47024	55,52976	62
2	10	1,045263	33,04526	32	17	20	9,51661	71,51661	62
3	10	0,324248	32,32425	32	18	20	2,045394	64,04539	62
4	10	-1,80589	30,19411	32	19	20	5,286107	67,28611	62
5	10	0,568473	32,56847	32	20	20	7,451416	69,45142	62
6	10	-0,17024	31,82976	32	21	30	19,59112	111,5911	92
7	10	0,676169	32,67617	32	22	30	33,2486	125,2486	92
8	10	-0,57257	31,42743	32	23	30	-23,2211	68,7789	92
9	10	-1,53944	30,46056	32	24	30	-28,8606	63,13944	92
10	10	-0,38377	31,61623	32	25	30	38,95497	130,955	92
11	20	-4,85783	57,14217	62	26	30	-1,97921	90,02079	92
12	20	-1,66701	60,33299	62	27	30	-36,9439	55,05615	92
13	20	9,513881	71,51388	62	28	30	12,09004	104,09	92
14	20	0,817791	62,81779	62	29	30	41,06767	133,0677	92
15	20	-11,1762	50,82381	62	30	30	-38,0374	53,96263	92

Tabela 4.1 *Prikaz simuliranih podataka*

Ocenjujemo model linearne regresije

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

$$\varepsilon_i : \mathcal{N}(0, 0.01X_i^2),$$

$$i = 1,2, \dots, 30$$

na osnovu vrednosti Y_i i X_i datih u tabeli 4.1. pomoću programskog paketa SPSS-a.

Dobijeni su sledeći rezultati:

Koeficijenti

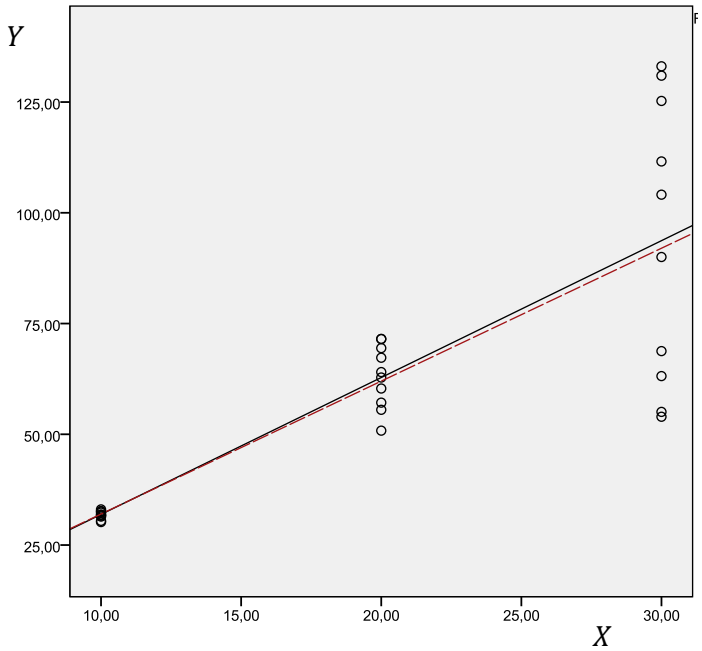
model	ocenjena vrednost	standardna greška	p-vrednost
$\hat{\alpha}$	**1.022**	8.880	0.909
$\hat{\beta}$	**3.090**	0.411	0.000

Analiza varijanse

	suma kvadrata	br. stepeni slobode	srednja vrednost sume	p-vrednost
regresijska (SSR)	19090.319	1	19090.319	0.000
rezidualna (SSE)	**9462.096**	28	337.932	
totalna (SST)	28552.414	29		

gde je koeficijent determinacije, $R^2 = 0.669$.

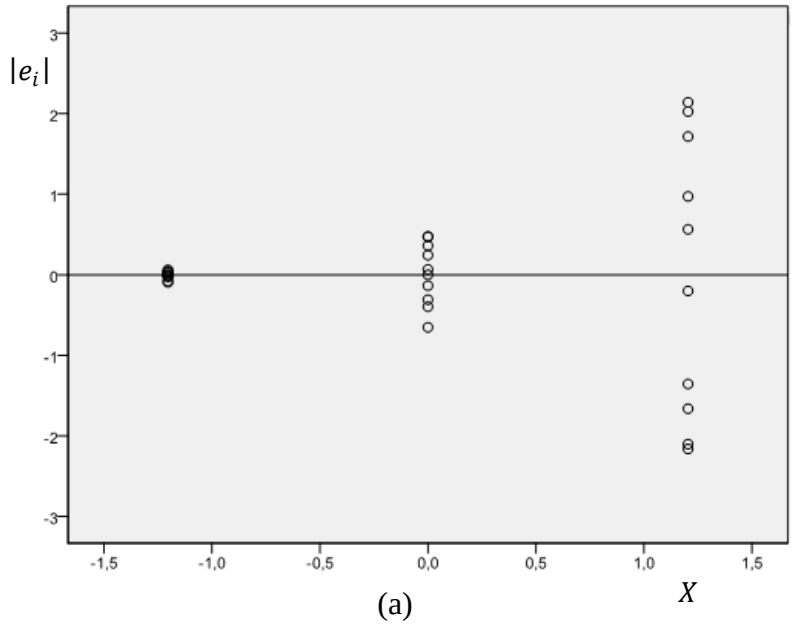
Na grafiku 4.1 je isprekidanom pravom prikazana populaciona linija $Y = 2 + 3X$, dok crna prava označava uzoračku liniju $\hat{Y} = 1.022 + 3.090X$. Može se zaključiti da su rasipanja veća za veće vrednosti nezavisne promenljive X i da uzoračka linija, dobijena na osnovu simulitranih podataka, malo odstupa od prave $Y = 2 + 3X$.



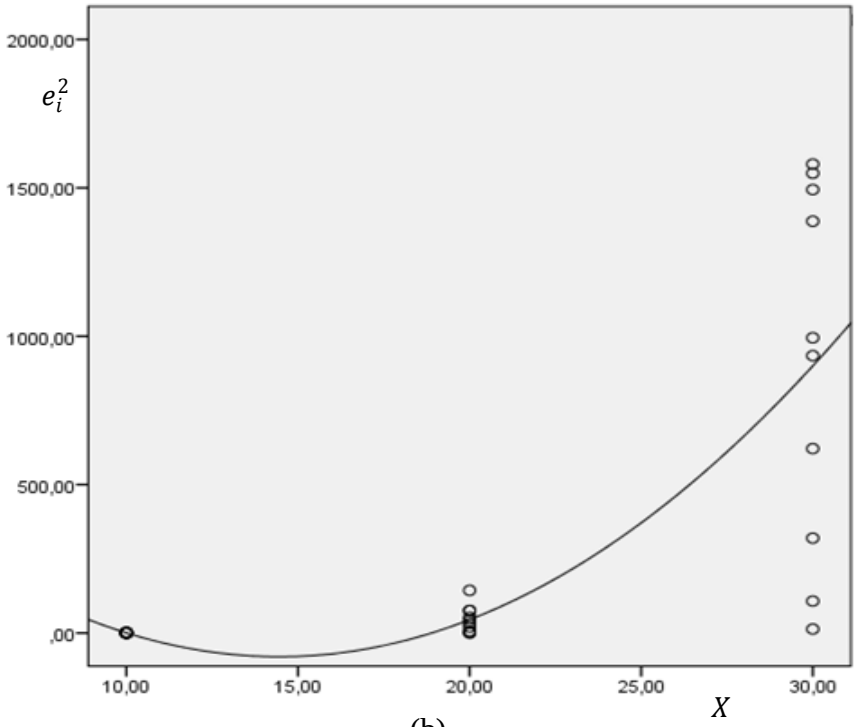
Grafik 4.1

Sada će se testirati heteroskedastičnost uz pomoć opisanih procedura i tako se prikazati koji su testovi uspešno detektovali ovaj oblik heteroskedastičnosti.

Grafička metoda:



(a)



(b)

Grafik 4.2 *Dijagrami rasturanja reziduala*

Na grafiku 4.2 (a) predstavljen je dijagram rasturanja tačaka na osnovu koga se može zaključiti da postoji zavisnost između reziduala i nezavisne promenljive X- što je veće X, veći su i reziduali. Dijagram (b) prikazuje zavisnost kvadriranih reziduala u odnosu na X gde se može primetiti zavisnost u kvadratmoj formi.

Sprovođenje testova:

1) Goldfeld-Kvantov test

Neće se izostaviti centralna opazanja zbog male veličine uzorka. Posle poretka opservacija po rastućem redosledu veličine X, ocenjuju se dva modela linearne regresije $Y_i = \alpha + \beta X_i + \varepsilon_i$, za prvih 15 i poslednjih 15 opservacija.

Poduzorak 1:

$$\hat{Y}_i = \underset{(3.324)}{3.075} + \underset{(0.235)}{2.873} X_i \qquad R^2 = 0.92$$

Poduzorak 2:

$$\hat{Y}_i = \underset{(39.369)}{9.516} + \underset{(1.454)}{2.803} X_i \qquad R^2 = 0.222.$$

Rezidualne sume poduzoraka su redom 18,411 i 704.495, pa je vrednost test statistike

$$F = \frac{704.495}{18.411} = 38.26.$$

Kako kritična vrednost F-raspodele sa 13 i 13 stepeni slobode i nivo značajnosti 0.05 iznosi 2.58, ovaj test pokazuje da je heteroskedastičnost prisutna.

2) Brojš-Paganov test

Na osnovu jednačine linearne regresije $\hat{Y}_i = 1.022 + 3.090 X_i$ dobija se ocenjena vrednost varijanse greške

$$\hat{\sigma}^2 = \frac{9462.10}{30} = 315.403.$$

Prikazana je nova jednačina regresije

$$\hat{p}_i = \underset{(0.609)}{-1.852} + \underset{(0.028)}{0.143} X_i$$

gde je $p_i = e_i^2 / \hat{\sigma}^2 = e_i^2 / 315.403$.

Test statistika je

$$\frac{SSR_{BP}}{2} = \frac{40.660}{2} = 20.33$$

gde je vrednost SSR_{BP} regresijska suma kvadrata. Kako kritična vrednost χ^2 raspodele sa jednim stepenom slobode i nivoom značajnosti 0.05 iznosi 3.841, takođe se zaključuje da je prisutna heteroskedastičnost.

3) Vajtov test

Ocenjena je pomoćna linearna regresija

$$e_i^2 = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

čije su vrednosti prikazane u narednoj tabeli,

Koeficijenti

model	ocenjena vrednost	standardna greška	p-vrednost
$\hat{\beta}_0$	**766.923**	484.031	0.125
$\hat{\beta}_1$	**-117.142**	54.964	0.042
$\hat{\beta}_2$	**4.053**	1.360	0.006

gde je koeficijent determinacije $R_W^2 = 0.607$. Može se primetiti da su koeficijenti uz X_i i X_i^2 statistički značajni dok konstanta nije. Vajtova test statistika je $nR_W^2 = 30 \cdot 0.607 = 18.21$ što je veće od tablična vrednosti χ^2 raspodele sa dva stepena slobode, 5.991. Isti je zaključak kao i do sada, da je prisutna heteroskedastičnost.

4) Glejzerov test

Testiraju se tri modela linearne regresije:

Model1:

$$|e_i| = \alpha + \beta X_i + u_i$$

Model2:

$$|e_i| = \alpha - \beta / X_i + u_i$$

Model3:

$$|e_i| = \alpha + \beta\sqrt{X_i} + u_i.$$

Rezultati su prikazani u sledećoj tabeli:

Koeficijenti

	ocenjeni parametri	ocenjene vrednosti	standardna greška	p-vrednost	R^2
Model1	$\hat{\alpha}$ $\hat{\beta}$	**-15.419** **1.335**	4.169 0.193	0.001 0.000	0.631
Model2	$\hat{\alpha}$ $\hat{\beta}$	**31.510** **-331.137**	4.500 66.813	0.000 0.000	0.467
Model3	$\hat{\alpha}$ $\hat{\beta}$	**-37.469** **11.151**	7.804 1.745	0.000 0.000	0.593

Ocenjeni parametri koji stoje uz regresore su statistički značajni. Svi su pogodni za testiranje hipoteze o heteroskedastičnosti, a na osnovu koeficijenta determinacije prvi se preferira jer je najveći. I ovaj test pokazuje prisustvo heteroskedastičnosti, mada se ne može sa sigurnošću tvrditi koji je oblik zastupljen.

4.2 Aditivna heteroskedastičnost

Podaci su uzeti iz tabele 4.1, ali tako da je simuliran oblik odstupanja greške $\sigma_i^2 = 2 + 0.5X_i + 0.01X_i^2$. Populaciona prava je $Y = 5 + 2X$, pa ocenjujemo model linearne regresije

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

$$\varepsilon_i : \mathcal{N}(0, 2 + 0.5X_i + 0.01X_i^2), \qquad i = 1,2,\dots,30.$$

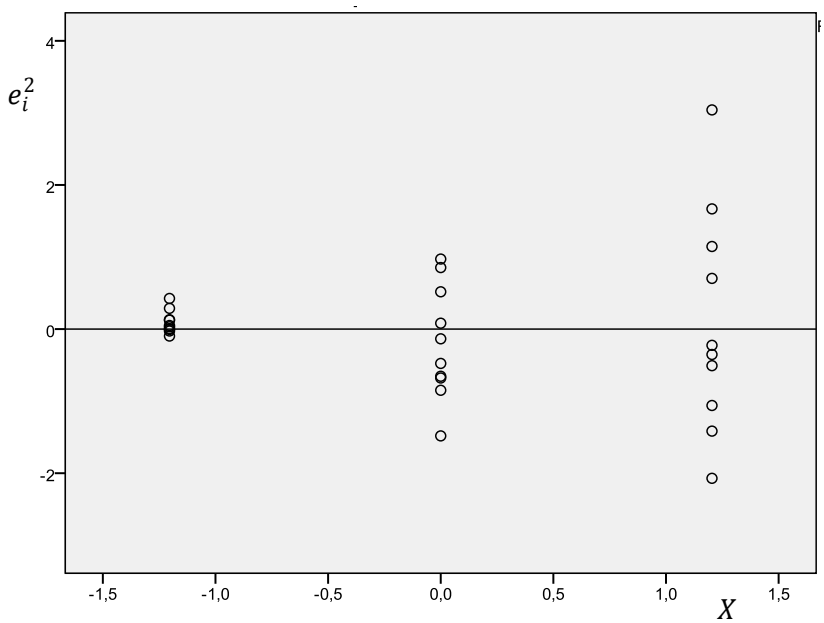
Ocenjena jednačina linearne regresije je

$$\hat{Y}_i = \underset{(1.913)}{6.272} + \underset{(0.089)}{2.036}X_i$$

gde je $R^2 = 0.95$, a rezidualna suma kvadrata $SSE = 438.946$.

Kao i u prethodnom slučaju, testiraće se prisustvo heteroskedastičnosti uz pomoć opisanih procedura.

Grafička metoda ukazuje na moguće postojanje heteroskedastičnosti.



Grafik 4.3 *Dijagram rasturanja reziduala*

1) Goldfeld-Kvantovim testom su ocenjene dve linearne regresije prvih i poslednjih 15 opservacija.

Poduzorak 1:

$$\hat{Y}_i = \underset{(1.913)}{6.499} + \underset{(0.077)}{2.050}X_i \qquad R^2 = 0.982$$

Poduzorak 2:

$$\hat{Y}_i = \underset{(8.125)}{-0.378} + \underset{(0.300)}{2.270}X_i \qquad R^2 = 0.815$$

sa rezidualnim sumama redom 25.450 i 390.076. Vrednost test statistike je 15.327 što je veće od tablične $F_{13,13}$ vrednosti, pa ovaj test detektuje heteroskedastičnost.

2) Brojš-Paganov test

Ocenjena vrednost varijanse greške je

$$\hat{\sigma}^2 = \frac{438.946}{30} = 14.632.$$

Pomoćna jednačina regresije

$$\hat{p}_i = \underset{(0.860)}{-1.291} + \underset{(0.040)}{0.115} X_i$$

gde je $p_i = e_i^2/\hat{\sigma}^2 = e_i^2/14.632$.

Test statistika je

$$\frac{SSR_{BP}}{2} = \frac{26.240}{2} = 13.12$$

gde je SSR_{BP} vrednost regresijske sume kvadrata. Kako kritična vrednost χ^2 raspodele sa jednim stepenom slobode i nivoom značajnosti 0.05 iznosi 3.841, takođe se zaključuje da je prisutna heteroskedastičnost.

3) Vajtov test

Ocenjena je pomoćna linearna regresija

$$e_i^2 = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

sa vrednostima prikazanim u narednoj tabeli:

Koeficijenti

model	ocenjena vrednost	standardna greška	p-vrednost
$\hat{\beta}_0$	**7.243**	36.183	0.843
$\hat{\beta}_1$	**-1.460**	4.109	0.725
$\hat{\beta}_2$	**0.078**	0.102	0.447

Koeficijent determinacije iznosi $R^2 = 0.245$.

Međutim, primećuje se da ni jedan ocenjeni koeficijent nije statistički značajan, što se moze videti na osnovu p-vrednosti. Vajtova test statistika je $nR_W^2 = 30 \cdot 0.245 = 7.35$. Ako je nivo poverenja 95%, kritična vrednost χ^2 raspodele sa 2 stepena slobode je 5.991 i tada se detektuje heteroskedastičnost. Ali, ako je nivo poverenja 99%, tada je kritična tablična vrednost 9.210 što je veće od vrednosti Vajtove test statistike, pa hipotezu o homoskedastičnosti ne odbacujemo, odnosno, zaključuje se da heteroskedastičnost nije prisutna. Ovo ukazuje na manji stepen pouzdanosti Vajtovog testa što se može opravdati malim obimom uzorka.

4) Glejzerov test

Testiraju se tri modela linearne regresije i rezultati su prikazani u sledećoj tabeli:

Koeficijenti

	ocenjeni parametri	ocenjene vrednosti	standardna greška	p-vrednost	R^2
Model1	$\hat{\alpha}$ $\hat{\beta}$	**-1.710** **0.218**	1.054 0.049	0.116 0.000	0.416
Model2	$\hat{\alpha}$ $\hat{\beta}$	**6.339** **-60.38**	0.973 14.442	0.000 0.000	0.384
Model3	$\hat{\alpha}$ $\hat{\beta}$	**-5.534** **1.872**	1.884 0.421	0.007 0.000	0.414

Ocenjeni parametri koji stoje uz regresore su statistički značajni. Iako koeficijenti determinacije nisu visoki, prvi ili poslednji se uzima za testiranje heteroskedastičnosti, koji se detektuje, mada se ovim testom ne prepoznaje aditivni oblik heteroskedastičnosti.

4.3 Heteroskedastičnost zavisne promenljive

Na osnovu simuliranih podataka, testiraće se heteroskedastičnost oblika

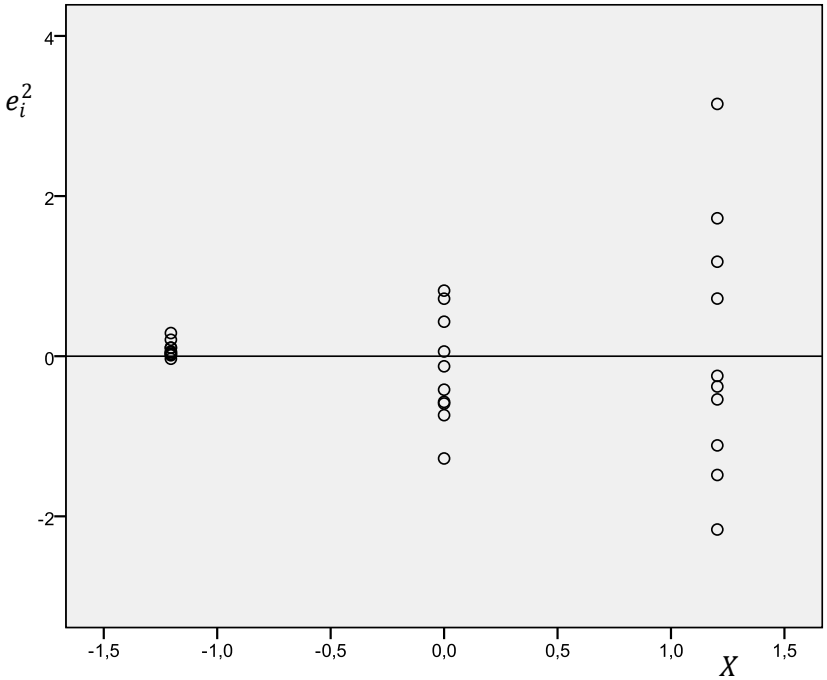
$$\sigma_i^2 = 0.01(5 + 2X_i)^2.$$

Ocenjena jednačina linearne regresije je

$$\hat{Y}_i = \underset{(3.238)}{3.8} + \underset{(0.150)}{2.063}X_i$$

gde je $R^2 = 0.871$, a rezidualna suma kvadrata $SSE = 1258.322$.

Grafička metoda ukazuje na moguće postojanje heteroskedastičnosti, što se može videti na grafiku 4.4.



Grafik 4.4 *Dijagram rasturanja reziduala*

Sprovođenjem testova dobija se sledeće:

1) Goldfeld-Kvantov test

Poduzorak 1:

$$\hat{Y}_i = \underset{(1.515)}{4.278} + \underset{(0.107)}{2.072}X_i \qquad R^2 = 0.966$$

Poduzorak 2:

$$\hat{Y}_i = \underset{(13.994)}{-6.119} + \underset{(0.517)}{2.412}X_i \qquad R^2 = 0.626.$$

Deljenjem rezidualnih suma poduzoraka izračunata je vrednost test statistike

$$F = \frac{1157.14}{49.716} = 23.275.$$

Kako kritična vrednost F-raspodele sa 13 i 13 stepeni slobode i nivo značajnosti 0.05 iznosi 2.58, ovaj test potvrđuje prisutnost heteroskedastičnosti.

2) Brojš-Paganov test

Na osnovu jednačine linearne regresije $\hat{Y}_i = 3.8 + 2.063X_i$, dobija se ocenjena vrednost varijanse

$$\hat{\sigma}^2 = \frac{1258.14}{30} = 41.9441.$$

Ako se napravi nova promenljiva $p_i = e_i^2/\hat{\sigma}^2$, pomoćna jednačina linearne regresije je

$$\hat{p}_i = \underset{(0.922)}{-1.493} + \underset{(0.043)}{0.125} X_i.$$

Test statistika je

$$\frac{SSR_{BP}}{2} = \frac{31.086}{2} = 15.543,$$

gde je SSR_{BP} vrednost regresijske sume kvadrata. Kako kritična vrednost χ^2 raspodele sa jednim stepenom slobode i nivoom značajnosti 0.05 iznosi 3.841, takođe se zaključuje da je prisutna heteroskedastičnost.

3) Vajtov test

Ocenjuje se linearna regresija čije su vrednosti zavisne promenljive e_i^2, a nezavisne promenljive su X_i i X_i^2 i ocenjene vrednosti parametra su prikazane u sledećoj tabeli:

Koeficijenti

model	ocenjena vrednost	standardna greška	p-vrednost
$\hat{\beta}_0$	**47.869**	110.005	0.667
$\hat{\beta}_1$	**-8.032**	12.492	0.526
$\hat{\beta}_2$	**0.332**	0.309	0.293

Do istog zaključka se može doći kao i u prethodnom slučaju: Nijedan ocenjeni koeficijent nije statistički značajan, Vajtova test statistika je $nR_W^2 = 30 \cdot 0.265 = 7.95$. Ako je nivo poverenja 95%, kritična vrednost χ^2 raspodele sa 2 stepena slobode je 5.991 i tada se detektuje heteroskedastičnost. Ali, ako je nivo poverenja 99%, tada je kritična tablična vrednost 9.21 što je veće od vrednosti Vajtove test statistike, pa hipotezu o homoskedastičnosti ne odbacujemo, odnosno, zaključuje se da heteroskedastičnost nije prisutna. Još jednom je potvrđena manja pouzdanost Vajtovog testa.

4) Glejzerov test

Testiraju se tri modela linearne regresije i rezultati su prikazani u sledećoj tabeli:

Koeficijenti

	ocenjeni parametri	ocenjene vrednosti	standardna greška	p-vrednost	R^2
Model1	$\hat{\alpha}$ $\hat{\beta}$	**-3.578** **0.395**	1.791 0.083	0.055 0.000	0.448
Model2	$\hat{\alpha}$ $\hat{\beta}$	**10.813** **-106.155**	1.693 25.132	0.000 0.000	0.389
Model3	$\hat{\alpha}$ $\hat{\beta}$	**-10.393** **3.368**	3.221 0.720	0.003 0.000	0.438

Ocenjeni parametri koji stoje uz regresore su statistički značajni, pa su pogodni za testiranje hipoteze o heteroskedastičnosti, a na osnovu koeficijenta determinacije prvi se preferira jer je najveći.

Na osnovu navedenih postupaka Parkovog, Bartletovog, Spirmanovog, Harvijevog i Basetovog testa izvršena su testiranja simuliranih oblika heteroskedastičnosti uz pomoć programa SPSS-a i Excel-a. U tabeli 4.2 su prikazane vrednosti test statistika i tabelarnih vrednosti, kako bi se uporedila uspešnost detektovanja. Nivo poverenja je 95%, a u zagradama su prikazane p-vrednosti testova.

testovi	multiplikativna heteroskedastičnost	aditivna heteroskedastičnost	het.zavisne promenljive	tablične vrednosti
Park	$P = 5.519$ $\langle 0.000 \rangle$	$P = 5.924$ $\langle 0.000 \rangle$	$P = 7.115$ $\langle 0.000 \rangle$	$t_{28} = 1.701$
Bartlet	$B = 25.7206$ $\langle 0.000 \rangle$	$B = 12.431$ $\langle 0.002 \rangle$	$B = 16.9665$ $\langle 0.000 \rangle$	$\chi_2^2 = 5.991$
Spirman	$S = 8.7639$ $\langle 0.000 \rangle$	$S = 6.06875$ $\langle 0.000 \rangle$	$S = 6.75465$ $\langle 0.000 \rangle$	$t_{28} = 1.701$
Harvi	$H = 11.2163$ $\langle 0.000 \rangle$	$H = 6.626$ $\langle 0.010 \rangle$	$H = 6.35$ $\langle 0.012 \rangle$	$\chi_1^2 = 3.841$

Koenker-Baset	$KB = 5.732$ $\langle 0.000 \rangle$	$KB = 2.973$ $\langle 0.006 \rangle$	$KB = 3.075$ $\langle 0.005 \rangle$	$t_{28} = 1.701$

Tabela 4.2 *Prikaz vrednosti test statistika različitih testova heteroskedastičnosti*

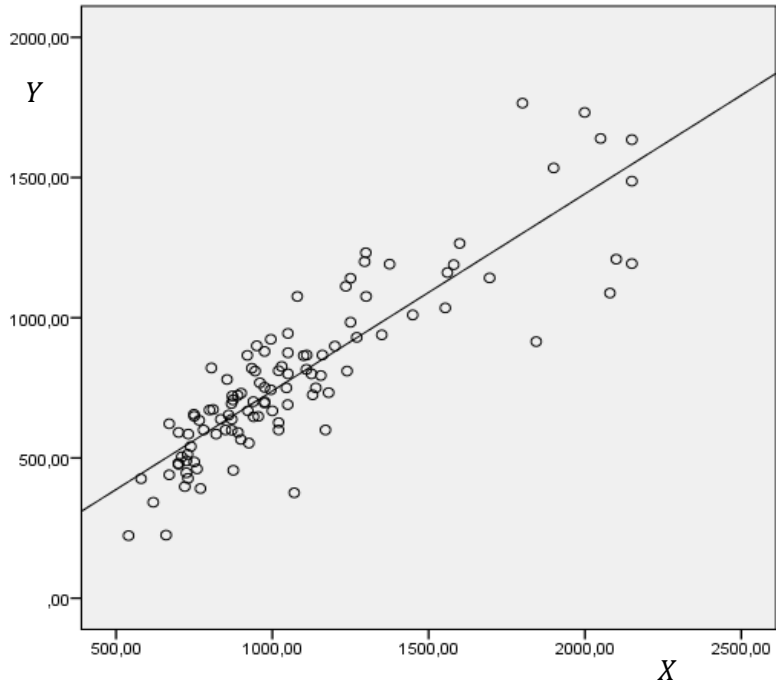
Primećuje se da su svi testovi navedeni u tabeli uspešno detektovali heteroskedastičnost. Međutim, kod Harvijevog testa ukoliko povećamo nivo poverenja na 99%, kritična vrednost će iznositi 6.635 pa se neće otkriti aditivni oblici heteroskedastičnosti.

Na osnovu simuliranih oblika heteroskedastičnosti, Goldfeld-Kvantov, Brojš-Pagan-Godfri i Glejzerov test uspešno detektuju heteroskedastičnost. Problem nastaje kod Vajtovog testa koji pokazuje različite rezultate za različite nivoe poverenja. Ovde smo simulirali 30 podataka. Međutim, pojedini testovi neće uvek uspešno detektovati heteroskedastičnost. Rezultati će zavisiti i od veličine obima uzoraka kao i od prave prirode stvarnih podataka. Neke studije su poredile ove testove na osnovu različitih uzoraka i simulacija i došle do pojedinih zaključaka.

4.4 Numerički primeri

Sada ćemo koristiti stvarne podatke, sa većim obimom uzorka, kako bismo sproveli testove heteroskedastičnosti. U tabeli 4.3 su prikazani podaci iz baze Albakerkijevog odbora nekretninama. Naime, podaci predstavljaju slučajan uzorak iz evidencije prodaje kuća u Albakerkiju (Novi Meksiko) 1993. godine. Y_i predstavlja godišnji porez nekretnine (u dolarima), dok X_i predstavlja prodajnu vrednost nekretnine (u hiljadama dolara). Veličina uzorka, tj. broj kuća je 107.

Na osnovu dijagrama rasturanja tačaka prikazanog na grafiku 4.5, primećuje se linearna zavisnost između vrednosti poreza kuća i njene prodajne vrednosti.



Grafik 4.5 *Zavisnost poreza od cena kuća*

Dakle, rezultati ocene linearne regresije oblika

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \qquad i = 1,2, \dots 107,$$

dati su u narednoj tabeli:

Koeficijenti

model	ocenjena vrednost	standardna greška	t-vrednost	p-vrednost
$\hat{\alpha}$	**36.344**	43.237	0.841	0.402
$\hat{\beta}$	**0.703**	0.038	18.581	0.000

Sa ovakvim podacima preseka se očekuje postojanje heteroskedastičnosti. Vidimo da je t-vrednost nagiba visoka, pa postoji mogućnost precenjivanja t-testa, a samim tim i pocenjivanje standardne grešake. Sa druge strane, standardna greška odsečka ima visoku vrednost. Koeficijent β je statistički značajno različit od nule što pokazuje p-vrednost.

U tabeli analize varijanse su prikazane vrednosti SSR, SSE, SST,

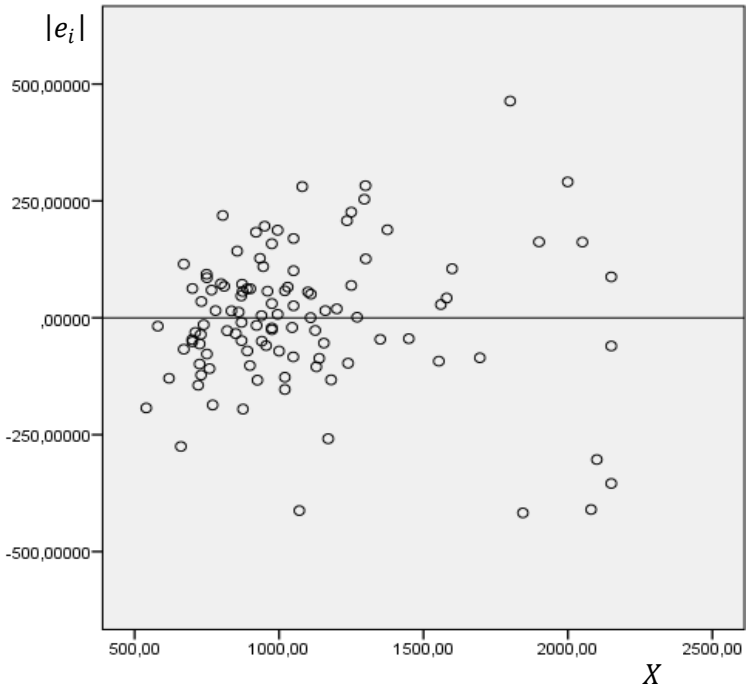
Analiza varijanse

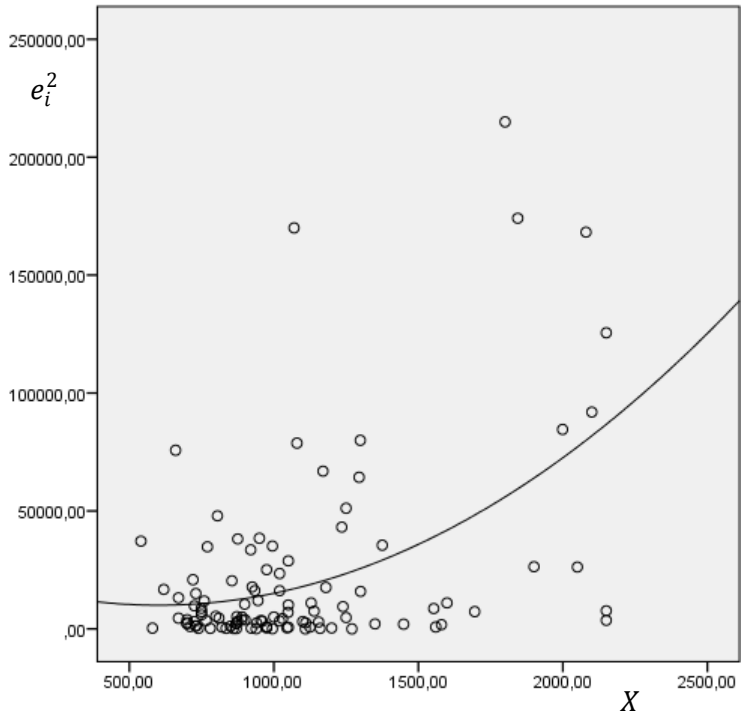
	suma kvadrata	br. stepeni slobode	srednja vrednost sume	p-vrednost
regresijska (SSR)	7 719 537.265	1	7 719 537.265	0.000
rezidualna (SSE)	**2 347 821.464**	105	22 360.204	
totalna (SST)	10 067 358.73	106		

gde je koeficijent determinacije, $R^2 = 0.767$.

Testiranje heteroskedastičnosti počećemo sa grafičkom metodom i prikazom reziduala $|e_i|$ i e_i^2 u zavisnosti od nezavisne promenljive X_i na dijagramu rasturanja tačaka koji su predstavljeni na grafiku 4.6. Očigledno postoji veza između reziduala i vrednosti nezavisne promenljive: što su vrednosti X_i veće, to su veća i odstupanja.

Testove heteroskedastičnosti ćemo sprovesti uz pomoć programa SPSS-a i Excela-a.





Grafik 4.6 *Dijagrami rasturanja reziduala*

Postupci sprovođenja testova su prikazani kroz simulacije. Rezultati Goldfeld-Kvantovog, Brojš-Paganovog i Vajtovog testa su prikazani u tabeli 4.4.

testovi	vrednost test statistike	tablična vrednost	p-vrednost
Goldfeld-Kvant	$F = 3.86$	$F_{43,43} = 1.66$	0.000
Brojš-Pagan	$\frac{SSR_{BP}}{2} = 34.86$	$\chi_1^2 = 3.84$	0.000
Vajt	$nR_W^2 = 24.824$	$\chi_2^2 = 5.991$	0.000

Tabela 4.4 *Prikaz pojedinih vrednosti test statistika*

Sva tri testa detektuju heteroskedastičnost. Kod Goldfeld-Kvantovog testa smo izostavili 17 centralnih opažanja. Naredni testovi su detaljnije prikazani:

Glejzerov test podrazumeva ocenjivanje tri modela linearne regresije

Model1:

$$|e_i| = \alpha + \beta X_i + u_i$$

Model2:

$$|e_i| = \alpha - \beta / X_i + u_i$$

Model3:

$$|e_i| = \alpha + \beta \sqrt{X_i} + u_i.$$

Rezultati su prikazani u sledećoj tabeli:

Koeficijenti

	ocenjeni parametri	ocenjene vrednosti	standardna greška	p-vrednost	R^2
Model1	$\hat{\alpha}$ $\hat{\beta}$	**-9.547** **0.111**	26.108 0.023	0.715 0.000	0.184
Model2	$\hat{\alpha}$ $\hat{\beta}$	**225.432** **-112 405.706**	32.909 30 841.022	0.000 0.000	0.112
Model3	$\hat{\alpha}$ $\hat{\beta}$	**-134.378** **7.554**	53.432 1.628	0.013 0.000	0.170

Ocenjeni parametri koji stoje uz regresore su statistički značajni. I ovaj test pokazuje prisustvo heteroskedastičnosti, mada se ne može odrediti koji je oblik zastupljen. Zbog niskih koeficijenata determinacije u sva tri slučaja, ni jedan oblik nije pogodan za testiranje hipoteze o heteroskedastičnosti.

Parkov test

Rezultati ocene linearne regresije oblika

$$\ln e_i^2 = \ln r_0 + r_1 \ln X_i + u_i, \qquad i = 1,2,\dots,107$$

dati su u narednoj tabeli:

Koeficijenti

model	ocenjena vrednost	standardna greška	t-vrednost	p-vrednost
$\widehat{\ln r_0}$ $\widehat{r_1}$	**-2.552** **1.591**	4.844 0.698	-0.527 2.279	0.599 0.025

Ako je nivo poverenja 95%, p-vrednost koeficijenta r_1 je 0.025 i možemo zaključiti da je prisutna heteroskedastičnost. Međutim, ukoliko je nivo poverenja 99%, tada koeficijent neće biti statistički značajan, pa mozemo prihvatiti nultu hipotezu, tj.da je model homoskedastičan. Ovo ukazuje na malu moć Parkovog testa jer je moguće da je veza između reziduala i nezavisne promenljive drugačije forme.

Bartletov test ne možemo sprovesti jer nemamo ponovljenih opservacija.

Spirmanov test

Posle rangiranja vrednosti nezavisne promenljive i apsolutnih vrednosti reziduala, izračunat je Spirmanov korelacijski koeficijent i iznosi $r = 0.23559$. Statistika ovog testa je

$$t = [r^2 105/(1 - r^2)]^{1/2}$$

i njena vrednost je $t = 2.48$. Studentova raspodela sa 105 stepeni slobode je oko 1.658 čiji je nivo poverenja 95%, pa se zaključuje da i on pokazuje prisutnost heteroskedastičnosti.

Harvijev test

Ocenjujemo linearnu regresiju oblika

$$\log e_i^2 = \log\sigma^2 + \delta \log X_i + v_i, \qquad i = 1,2, \dots ,107$$

i dobijamo regresijsku sumu kvadrata koja iznosi 5.089. Tada će vrednost test statistike biti $H = 5.089/4.9348 = 1.0312$. Kako je kritična vrednost χ^2 raspodela sa jednim stepenom slobode 3.841, može se zaključiti da se nulta hipoteza prihvata, odnosno da ne postoji heteroskedastičnost. Međutim drugi testovi detektuju heteroskedastičnost, što može značiti da ona ipak postoji ali da njen oblik nije multiplikativan.

Koanker-Basetov test

Rezultati ocene pomoćne linearne regresije

$$e_i^2 = \gamma_1 + \gamma_2 \left(\hat{Y}_i\right)^2 + u_i, \qquad i = 1,2, \dots ,107$$

su prikazani u tabeli

Koeficijenti

model	ocenjena vrednost	standardna greška	t-vrednost	p-vrednost
$\hat{\gamma}_1$	**-2712.565**	5566.139	-0.487	0.627
$\hat{\gamma}_2$	**0.035**	0.006	5.559	0.000

Zaključujemo da se nulta hipoteza o homoskedastičnosti ($\gamma_2 = 0$) odbacuje na osnovu ovog testa, dakle prisutna je heteroskedastičnost.

Kao što je i očekivano, većina testova detektuje heteroskedastičnost. Glejzerov test nam ne daje odgovore o obliku heteroskedastičnosti, Parkov test u zavisnosti od nivoa poverenja pokazuje da je prisutna heteroskedastičnost, dok Harvijev govori da nije prisutna multiplikativna heteroskedastičnost. Tada, da bismo otklonili uticaj heteroskedastičnosti, ne možemo koristiti transformaciju metodom ponderisanih najmanjih kvadrata. Za ovakav model linearne regresije, gde su nam sve vrednosti nezavisne i zavisne promenljive pozitivne, koristićemo logaritamsku transformaciju.

Ocenjujemo model linarne regresije

$$\ln Y_i = \alpha + \beta \ln X_i + \varepsilon_i, \quad i = 1,2,\dots,107.$$

Rezultati su prikazani u tabeli:

Koeficijenti

model	ocenjena vrednost	standardna greška	t-vrednost	p-vrednost
$\hat{\alpha}$	**-0.650**	0.419	-1.552	0.124
$\hat{\beta}$	**1.047**	0.060	17.332	0.000

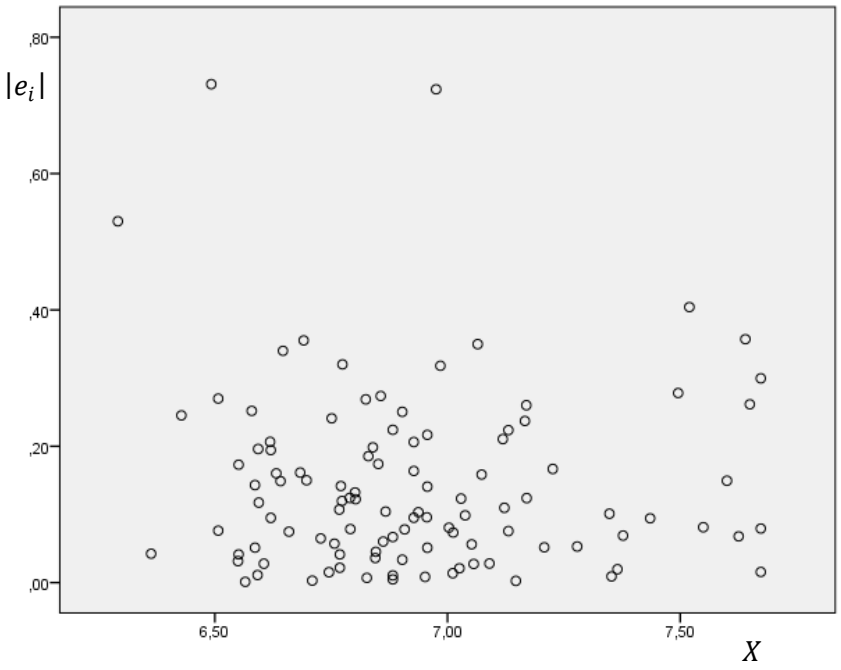
Analiza varijanse

	suma kvadrata	br. stepeni slobode	srednja vrednost sume	p-vrednost
regresijska (SSR)	11.682	1	11.682	0.000
rezidualna (SSE)	**4.083**	105	0.039	
totalna (SST)	15.765	106		

gde je koeficijent determinacije, $R^2 = 0.741$.

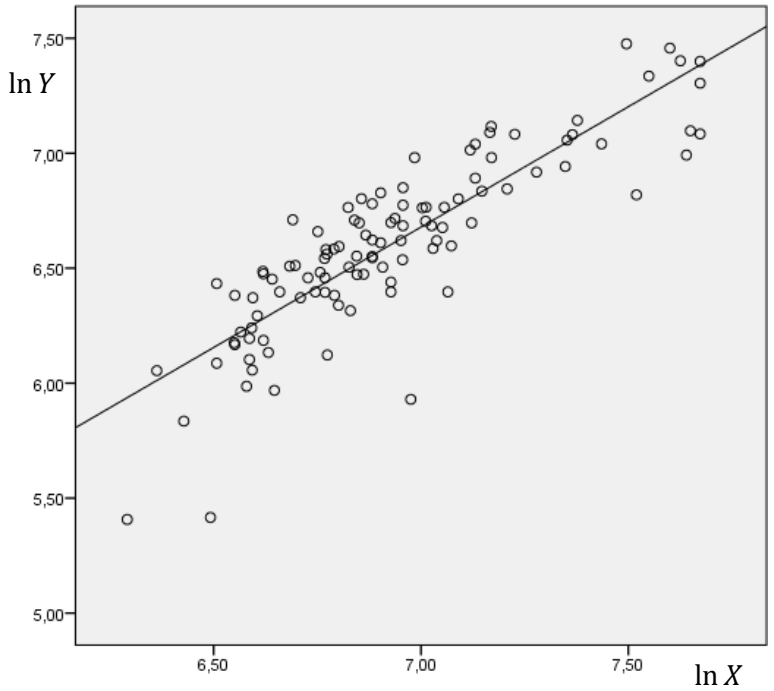
Koeficijent determinacije logaritamske transformacije linearne regresije je približno jednak koeficijentu determinacije originalne linearne regresije. Vidimo da je t-vrednost nagiba manja, ali ne puno u odnosu na početni model. Vrednost odsečka je smanjena, kao i njena standardna greška. Koeficijent β je i dalje statistički značajno različit od nule što pokazuje p-vrednost.

Na osnovu dijagrama rasturanja reziduala prikazanog na grafiku 4.7, ne može se zaključiti da postoji zavisnost reziduala i nezavisne promenljive, što potvrđuju i testovi koji su sprovedeni na isti način kao i za originalni model. Svi sprovedeni testovi na transformisan model su pokazali da treba prihvatiti hipotezu o homoskedastičnosti varijanse greške.



Grafik 4.7 *Dijagram rasturanja reziduala*

Na grafiku 4.8 su predstavljene vrednosti zavisne promenljive, $\log Y$, tj. logaritmovane vrednosti poreza u zavisnosti od logaritmovanih vrednosti cena kuća.



Grafik 4.8 *Zavisnost logaritmovanih vrednosti poreza i cena kuća*

red.br.	cena	porez	red.br	cena	porez	red.br	cena	porez
1	2050	1639	38	720	398	75	820	585
2	2080	1088	39	749	656	76	780	600
3	2150	1193	40	731	585	77	770	391
4	2150	1635	41	725	490	78	700	591
5	1999	1732	42	670	440	79	540	223
6	1900	1534	43	2150	1487	80	1070	376
7	1800	1765	44	1599	1265	81	2100	1209
8	1560	1161	45	1350	939	82	725	447
9	1449	1010	46	1299	1232	83	660	225
10	1375	1191	47	1250	1141	84	580	426
11	1270	930	48	1239	810	85	1844	915
12	1250	984	49	1200	899	86	1580	1189
13	1235	1112	50	1125	800	87	699	481
14	1170	600	51	1100	865	88	1160	867
15	1180	733	52	1080	1076	89	1109	816
16	1155	794	53	1050	875	90	1129	725
17	1110	867	54	1049	690	91	1050	800
18	1139	750	55	955	648	92	1045	750
19	995	923	56	934	820	93	1050	944

20	995	743	57	875	456	94	1020	811
21	975	752	58	889	723	95	1000	668
22	975	696	59	855	780	96	1030	826
23	900	731	60	835	638	97	975	880
24	960	768	61	810	673	98	950	900
25	860	653	62	805	821	99	940	647
26	1695	1142	63	799	671	100	920	866
27	1553	1035	64	750	649	101	945	810
28	1300	1076	65	759	461	102	874	707
29	1020	626	66	750	486	103	872	721
30	1020	600	67	730	427	104	870	638
31	922	668	68	729	513	105	869	694
32	925	553	69	710	504	106	766	634
33	899	566	70	670	622	107	739	541
34	850	600	71	619	342			
35	890	591	72	1295	1200			
36	870	599	73	975	700			
37	700	477	74	939	701			

Tabela 4.3 *Prikaz stvarnih podataka*

Kao drugi primer, navešćemo grafički pristup analize stavrnih podataka koji je predložio Tsung-Chi Cheng [7]. Kako bi ispitao slučajeve outlajera i njihov uticaj na ocene parametara u prisustvu heteroskedastičnosti, Green je koristio podatke prihoda-rashoda, za koje zavisna promenljiva Y predstavlja troškove za državno školovanje po glavi stanovnika, a nezavisna promenljiva X, (skalirana za 10^{-4}) je prihod po glavi stanovnika iz 1979.god u SAD, kao i kvadrat ovih vrednosti, X^2.

Ove podatke je analizirao Cribari-Neto i pokazao u svojoj analizi da je Aljaska netipičan slučaj te je ocenjivao parametre sa i bez Aljaske. Takođe su razmatrana tri slučaja: Aljaska, Misisipi i Vašingtong DC kao potencijalne uticajne tačke, a zatim su predstavljeni efekti uticajnih tačaka na ocenjene standardne greške i t-statistike. Međutim, jedini načini da se identifikuju outlajeri su zasnovani na H-matrici. Takva analiza predstavlja opasnost od pojave maskiranja (*eng. masking*- mogućnost pretvaranja outlajera u uticajne tačke) i potapanja (*eng. swamping*- mogućnost pretvaranja uticajnih tačaka u outlajere) zbog većeg broja outlajera u podacima. Model koji je prvi predložen kako bi se analizirali dati podaci je

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i.$$

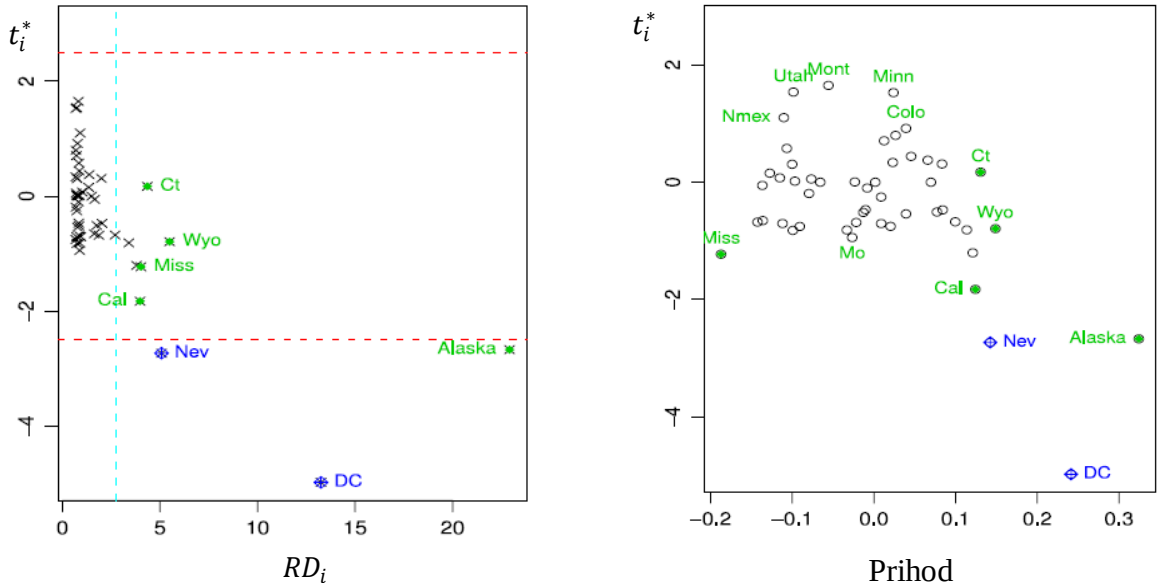
Međutim, kvadrirani model rezultrira velikom korelacijom između X i X^2, što dovodi do nestabilnih ocena $HCCM$. Zato je predložen drugi model

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 Z_i^2 + \varepsilon_i,$$

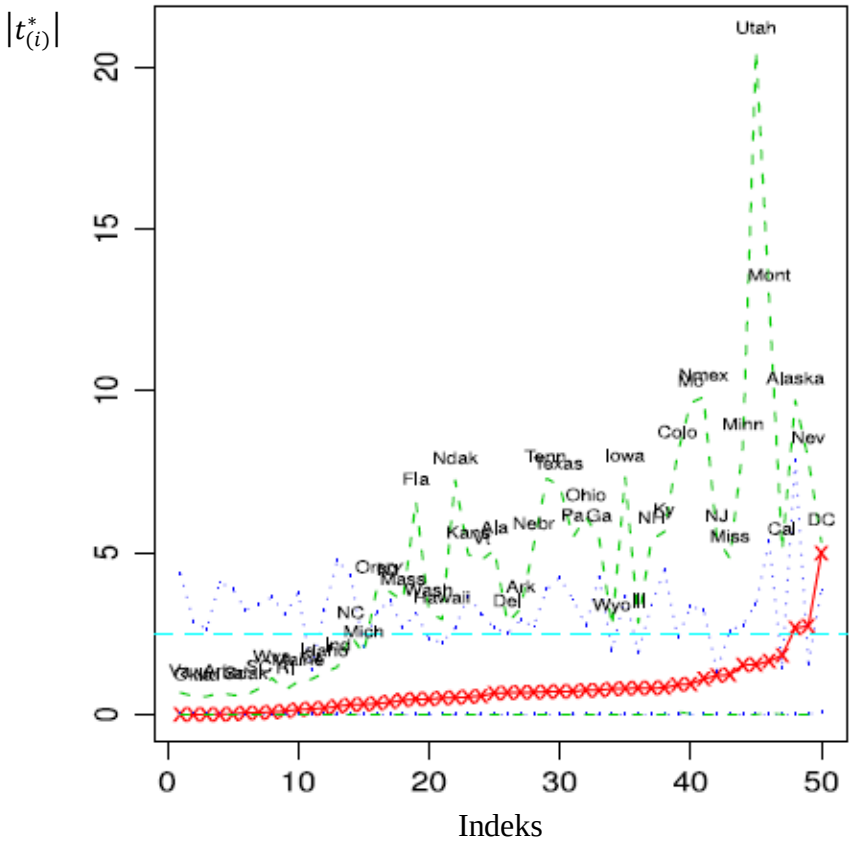
gde je $Z_i = X_i - \bar{X}$ [7].

Slike 4.1 i 4.2 predstavljaju grafičku analizu ovih podataka fitujući predloženi model. Prvi grafik na slici 4.1 je dijagnostički grafik zasnovan na PNAO ocenjivanju koja identifikuje Vašingtong DC, Nevadu i Aljasku kao loše uticajne tačke. Ipak *Jigsaw* grafik pokazuje da su Vašingtong, Nevada i Kalifornija outlajeri dok je Aljaska dobra uticajna tačka pre nego loša ukoliko se koristi (3.2) za σ. *Jigsaw* grafik koristeći (3.3) za σ, reflektuje heteroskedastičnost u kojoj se Juta ističe kao vrh. Dalji rezultati su prikazani na drugom grafiku na slici 4.1 koji predstavlja standardizovane PNAO reziduale u odnosu na nezavisnu promenljivu X. Vašington, Nevada, Kalifornija i Aljaska se nalaze u desnom uglu ovog grafika. Juta, Montana i Novi Meksiko su najviše heteroskedastične zemlje u smislu da se razlikuju od drugih "regularnih" zemalja.

Prema navedenim *Jigsaw* graficima, Aljaska i Misisipi su dobre uticajne tačke. Vašingtong DC je očigledno loša uticajna tačka potvrđena različitim metodama. Kalifornija je loša uticajna tačka ali u ranijim analizama istraživača nije bila primećena. Juta se pojavljuje kao "najviše heteroskedastična" država kada analiza uzima u obzir outlajere.



Slika 4.1: [7] *Dijagnostički grafik i grafik rasturanja reziduala zasnovanih na PNAO metodi*



Slika 4.2: [7] *Jigsaw grafik*

Sada će se prikazati kako izostavljanje nekih outlajera utiče na ocenjivanje parametara. Neka je $S_1 = \{Aljaska, Misisipi, Vašington\ DC\}$ skup koji sadrži outlajere koji su otkriveni u drugim literaturama, a $S_2 = \{\ Nevada,\ Kalifornija,\ Vašington\ DC\ \}$ skup koji sadrži outlajere identikovanim *Jigsaw* grafikom.

Tabela 4.5 sadrži rezultate ocena koeficijenata i zaključke koristeći PNAO metodu i HC3 ocene bazirane na različitim podskupovima podataka. Predstavljene su i standardne greške i p-vrednosti. Uključivanje svih slučajeva i isključivanje različitih podskupova S_1 i S_2, dovodi do različitih rezultata za klasični HCCM. Rezultati za HC3, kada se isključuje S_2, potvrđuju da su opservacije u tom podskupu outlajeri. Ipak, PNAO pokazuje dosta drugačije ocene posebno ocenu standardne greške za X^2. Razlika je u tome što sva posmatranja imaju iste pondere kada se koristi HC3 ocenjivač, što ignoriše heteroskedastičnost u podacima.

	HC3			PNAO
	ceo skup	S_1 izostavljen	S_2 izostavljen	ceo skup
$\hat{\beta}_0$	356.09	362.61	359.37	354.79
stan.greška	16.55	11.25	8.29	11.77
p-vrednost	0.001	0.001	0.001	0.001
$\hat{\beta}_1$	580.82	564.85	702.82	774.77
stan.greška	108.01	98.95	69.97	198.38
p-vrednost	0.001	0.001	0.001	0.001
$\hat{\beta}_2$	1587.04	952.64	2138.08	3735.33
stan.greška	1995.21	1297.36	**501.18**	**2282.64**
p-vrednost	0.430	0.467	0.001	0.108

Tabela 4.5: [7] *Rezultati ocena za prihod-rashod podatke*

5. Ispitivanje veličine i moći testova

Mnogi Monte Carlo eksperimenti su izvođeni za izučavanje performansi testova heteroskedastičnosti. Jedna takva studija je od Ali i Giaccotto koji su razmatrali šest tipova specifikacija heteroskedastičnosti [12]:

i. $\sigma_i^2 = \sigma^2$
ii. $\sigma_i^2 = \sigma^2 |X_i|$
iii. $\sigma_i^2 = \sigma^2 |E(Y_i)|$
iv. $\sigma_i^2 = \sigma^2 X_i^2$
v. $\sigma_i^2 = \sigma^2 [E(Y_i)]^2$
vi. $\sigma_i^2 = \sigma^2$ na $i \leq n/2$ i $\sigma_i^2 = 2\sigma^2$ za $i > n/2$

Razmatrano je šest različitih skupova podataka i ispitivana su pet modela počevši sa jednim regresorom i zavrsno sa pet promenljivih u linearnoj regresiji. Korišćena su četiri tipa raspodela: normalna, studentova, Košijeva i log normalna. Prve tri su simetrične, dok poslednja nije. Tri obima uzoraka je uzeto u obzir: $n = 10, 25$ i 40. Testovi koji su se primenjivali su bili: Glejzerov, Goldfeld-Kvantov, Brojš-Paganov i Vajtov. Rezultati su brojni, ali neki od glavnih zaključaka su sledeći:

1) Moć ovih testova raste sa rastom obima uzorka i variranjem regresora. Takođe, moć testova slabi ukoliko ima više regresora i ukoliko postoje odstupanja od normalne raspodele. Rezultati su uglavnom nepravilni kada su greške autokorelisane.
2) Nijedan od ovih testova nema značajnu moć detektovanja heteroskedastičnosti koja značajno odstupa od pravilne heteroskedastičnosti, tj. njihovih osnovnih oblika. Npr. nijedan od ovih testova nije mogao da detektuje šestu vrstu, odnosno $\sigma_i^2 = \sigma^2$ na $i \leq n/2$ i $\sigma_i^2 = 2\sigma^2$ za $i > n/2$. Zapravo, maksimalno otkrivanje ovog oblika bila je u 9% slučajeva.
3) Testovi iste forme koji koriste apsolutne vrednosti radije od kvadrata reziduala, verovatnije će biti nerobusni i manje moći.

Iako je malo pažnje posvećeno testiranju heteroskedastičnosti u regresiji kada postoje greške merenja, mogu se uporediti različiti testovi heteroskedastičnosti u ovim situacijama. Cilj je ispitivanje njihove veličine i moći kroz simulacije [11].

Kada se ispituju specifikacija i greške merenja, radi jednostavnosti se koristi jedan regresor u modelu:

$$Y_i = \beta_0 + \beta_1 x_i + u_i$$

gde je Y_i zavisna promenljiva, a x_i stohastička promenljiva, koji se još naziva pravi regresor, a greška merenja je označena sa ε_i, pa je

$$X_i = x_i + \varepsilon_i.$$

Pretpostavlja se da greška u linearnoj regresiji, u_i i greška merenja, ε_i, imaju srednju vrednost nula, varijansu σ_u^2 i σ_ε^2 i nisu u korelaciji, takođe greške nisu u korelaciji sa pravim regresorom x_i. Da bismo definisali različite vrste reziduala, koristi se oblik jednačine

$$Y_i = \beta_0 + \beta_1 X_i - \beta_1 \varepsilon_i + u_i$$

sa očekivanjem

$$E(Y_i) = \beta_0 + \beta_1 X_i - \beta_1 E(\varepsilon_i). \tag{5.1}$$

Dakle, reziduali najmanjih kvadrata, u oznaci LS, oceniće razliku ova dva izraza:

$$Y_i - E(Y_i) = -\beta_1(\varepsilon_i - E(\varepsilon_i)) + u_i$$

koji će imati nula očekivanje

$$E\big(-\beta_1\big(\varepsilon_i - E(\varepsilon_i)\big) + u_i\big) = -\beta_1 E\big(\varepsilon_i - E(\varepsilon_i)\big) + E(u_i) = 0.$$

Definišemo IV1 greške, čiji reziduali ocenjuju razliku između izraza (5.1) i $\beta_0 + \beta_1 X_i$, odnosno

$$Y_i - (\beta_0 + \beta_1 X_i) = -\beta_2 \varepsilon_i + u_i \tag{5.2}$$

Ako se ocena parametra dobija metodom ocene instrumentalne regresije u kojoj se koriste jedna ili više instrumentalnih promenljivih kako bi se dobile konzistentne ocene koeficijenata, može se izračunati treća grupa reziduala. Koristi se samo jedan instrument z_i specificiran kao

$$x_i = \alpha_0 + \alpha_1 z_i + r_i$$

gde z_i i r_i nisu u korelaciji sa u_i i ε_i. Takođe se pretpostavlja da r_i ima srednju vrednost nula, konstantnu varijansu σ_r^2 i nije u korelaciji sa z_i, varijanse σ_z^2. Tada je regresor X_i u obliku

$$X_i = \alpha_0 + \alpha_1 z_i + r_i + \varepsilon_i.$$

U praksi se napravi regresija od X_i na z_i i dobije fitovana vrednost $\hat{X}_i$, na osnovu koje možemo napraviti treći skup IV2 grešaka kao

$$Y_i - (\beta_0 + \beta_1 \hat{X}_i) = -\beta_1 \left(x_i - \hat{X}_i\right) + u_i = -\beta_1 \left(\alpha_0 + \alpha_1 z_i + r_i - (\alpha_0 + \alpha_1 z_i)\right) + u_i$$
$$= -\beta_1 r_i + u_i. \tag{5.3}$$

Ocene grešaka u (5.2) i (5.3) nazivaju se IV1 i IV2 reziduali.

Bo Wallentin i Anders Agren [11] su prikazali ispitivanje moći i veličine Glejzerovog, Goldfeld-Kvantovog i Vajtovog testa tako što su test statistike izračunate na osnovu LS, IV1 i IV2 reziduala.

U Glejzerovom i Goldfeld-Kvantovom testu imamo jednu promenljivu, dok je Vajtov test uopšteniji pristup. Pretpostavlja se da je varijansa greške povezana sa pravim regresorom kao

$$Var(u_i) = \sigma_u^2 x_i^\delta. \tag{5.4}$$

U Glejzerovom testu koristi se *t*-test za nagib u regresiji

$$|e_i| = \delta_0 + \delta_1 W_i + u_i,$$

gde je e_i jedan od tri skupa reziduala koji su prethodno objašnjeni, a W_i je posmatrano X_i ili njegova fitovana vrednost u instrumentalnoj regresiji $X_i = \alpha_1 + \alpha_2 z_i + r_i + \varepsilon_i$. Goldfeld-Kvantov test se samo primenjuje na reziduale najmanjih kvadrata. Izostavljenje opservacija iz sredine se u ovom slučaju ne radi. U Vajtovom testu se računa nR^2 iz sledeće regresije

$$e_i^2 = \delta_0 + \delta_1 W_i + \delta_2 W_i^2 + \eta_i$$

gde je n broj opservacija i pretpostavlja se da ova test statistika ima χ^2 raspodelu sa dva stepena slobode.

Simulacioni eksperiment podrazumeva da je regresor x_i funkcija od z_i gde se bira da je $\alpha_0 = 6, \alpha_1 = 1$ pa je $x_i = 6 + z_i + r_i$. Za z_i se pretpostavlja uniformna raspodela od 0 do $\sqrt{12}$ i r_i ima standardizovanu normalnu raspodelu što implicira da je $Var(x_i) = 2$. Drugi parametri su fiksirani i iznose $\beta_0 = 0$ i $\beta_1 = 1$. Kada nema heteroskedastičnosti, varijansa od u_i je podešena da se dobije R^2-vrednost $2/3$, a kada je prisutna heteroskedastičnost, simulacije su izvedene tako da je srednja vrednost n varijansi od u_i jednaka jedinici. Broj replikacija je 10 000.

Rezultati eksperimenta su prikazani u tabeli 5.1 u sličaju bez prisustva heteroskedastičnosti ($\delta = 0$), bez i sa greškama merenja ($Var(\varepsilon) = 0,\ n = 200$ i $Var(\varepsilon) = 2/3\ n = 50$ i $n = 200$). Glejzerova test statistika je označena sa GL, Goldfeld-Kvantova sa GQ, a Vajtova sa WH. Kao što se može videti, sve empirijske vrednosti su razumno bliske nominalnoj vrednosti osim kod

Vajtovog testa kada se primene IV1 i IV2 reziduali u situaciji postojanja greške merenja. Druga situacija kod Vajtovog testa koja je loša je kada nemamo grešku merenja i koriste se IV2 reziduali.

	n	$Var(\varepsilon)$	GL	GQ	WH
LS	200	0	5.06	4.94	4.46
LS	200	2/3	4.56	5.03	3.89
LS	50	2/3	5.20	5.17	3.65
IV1	200	0	5.22	-	4.61
IV1	200	2/3	5.26	-	29.56
IV1	50	2/3	5.34	-	15.02
IV2	200	0	5.11	-	25.45
IV2	200	2/3	5.44	-	16.88
IV2	50	2/3	5.57	-	9.38

Tabela 5.1: [11] *Empirijske vrednosti različitih testova*

Kada se kod testova napravi pomoćna regresija od kvadrata IV1 reziduala, mora se pretpostaviti da su regresori asimptotski nezavisno distribuirani u odnosu na grešku u_i. Za Vajtov test koji je baziran na X, ova pretpostavka nije ispunjena kada se koriste reziduali IV1 i IV2 u slučaju greške merenja i za IV2 reziduale kada nema greške merenja. Razlog toga je kvadrirani regresor koji je u Vajtovom testu u korelaciji sa kvadratom reziduala.

Sada se uvodi heteroskedastičnost: $\delta = 1$ i $\delta = 2$ u (5.4). Ispituju se uzorci obima $n = 50, 200$ i 800. U tabeli 5.2 prikazani su rezultati odbacivanja hipoteze o homoskedastičnosti. Niži nivo heteroskedastičnosti dovodi do male stope odbacivanja hipoteze za $n = 200$. Za $n = 800$, stepen odbacivanja hipoteze je između 34 i 48%. Za visoku heteroskedastičnost stopa odbacivanja hipoteze je i dalje veoma mala za najmanji obim uzorka. Za $n = 200$ varira od 29 do 45%, a za $n = 800$ skoro svi su iznad 90%.

Generalno, najviša stopa odbacivanja homoskedastičnosti je izračunata za Glejzerov test gde je X regresor i koriste se LS reziduali. Interesantno je što ovaj test ima veću stopu odbacivanja kada su uključeni reziduali najmanjih kvadrata nego IV1 reziduali. Relativno loši rezultati Vajtovog testa mogu biti zbog činjenice da je zasnovan na χ^2 raspodeli sa dva stepena slobode.

	GL	GQ	WH
$\delta = 2;\ n = 50$			
LS	13.42	11.92	8.14
IV1	12.55	-	7.96
IV2	12.78	-	9.52
$\delta = 2;\ n = 200$			
LS	45.30	36.55	34.55
IV1	36.52	-	29.41
IV2	36.88	-	35.58
$\delta = 2;\ n = 800$			
LS	96.35	90.83	95.46
IV1	90.20	-	87.72
IV2	91.61	-	93.96
$\delta = 1;\ n = 200$			
LS	14.80	12.99	10.75
IV1	13.26	-	10.57
IV2	13.60	-	12.47
$\delta = 1;\ n = 800$			
LS	48.17	39.66	45.61
IV1	39.22	-	34.46
IV2	40.74	-	41.29

Tabela 5.2: [11] *Procenat odbacivanja nulte hipoteze*

Zaključuje se da su testovi, sem jednog izuzetka, dobri pokazatelji postojanja heteroskedastičnosti bez obzira koja vrsta reziduala se koristi. Taj izuzetak je Vajtov test čija je vrednost previsoka kada se koriste IV1 i IV2 reziduali i kada su opažajne, a ne fitovane vrednosti regresora korišćene u pomoćnoj regresiji. Dakle, korišćenje Vajtovog testa u ovom slučaju, kao i ponekad u primeni u radu je neprikladno. Glejzerov test je bolji kada se primene LS nego IV1 reziduali. Ovi rezultati se možda promene ako su regresori veoma loše izmereni u kombinaciji sa dobrim instrumentima, ali u praksi to nije česta situacija.

5.1 Poboljšan Vajtov test

Vajtov test upoređuje ocenjenu varijansu regresijskih koeficijenata u uslovima homoskedastičnosti, sa varijansom u uslovima heteroskedastičnosti, ima asimptotski χ^2

raspodelu i dobar je test za velike obime uzorka. Međutim, ukoliko imamo male obime uzoraka, veličina i moć Vajtove statistike je nezadovoljavajuća.

Za višestruku regresiju sa k nezavisnih promenljivih, Vajtov test je definisan kao

$$W = \left(\frac{1}{n}\sum_{i=1}^{n}(e_i^2 - \hat{\sigma}^2)\psi_i'\right)\widehat{D}_n^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}(e_i^2 - \hat{\sigma}^2)\psi_i\right),$$

gde je ψ_i vektor svih elemenata $\boldsymbol{X}\otimes\boldsymbol{X}$ (Kroneckerov proizvod matrice $\boldsymbol{X}$) isključujući konstantu, $\bar{\psi}$ je uzoračka sredina ψ_i, $\boldsymbol{X}$ je matrica nezavisnih promenljivih dimenzije $n \times k$ i

$$\widehat{D}_i = \frac{1}{n}\sum_{i=1}^{n}(e_i^2 - \hat{\sigma}^2)^2\,(\psi_i - \bar{\psi})(\psi_i - \bar{\psi})'.$$

Greška $\boldsymbol{\varepsilon}$, ima očekivanje nula vektor i varijansu $\sigma^2\boldsymbol{I}$. Vajt je pokazao da W asimtotski teži ka χ^2 raspodeli sa $\frac{k(k-1)}{2}$ stepeni slobode. Pošto je $\widehat{D}_i$ teško izračunati, Vajt je predložio asimptotski ekvivalentnu verziju ovog testa kao

$$W^* = nR^2$$

gde je R^2 koeficijent determinacije pomoćne regresije e_i^2 na konstantu i sve elemente u ψ_i[13].

Dok Vajtova statistika, W^*, asimptotski teži ka χ^2 raspodeli pod nultom hipotezom homoskedastičnosti, raspodela konačnog uzorka W^* to ne sledi. Tako, kod malih uzoraka, raspodela može znatno odstupati od χ^2, pa zaključci, bazirani na asimptotskom Vajtovom testu mogu biti zbunjujući i pogrešni. Veoma je bitno empirijsko pitanje koliko dobro χ^2 raspodela aproksimira raspodelu konačnog uzorka Vajtove statistike kada je obim uzorka mali.

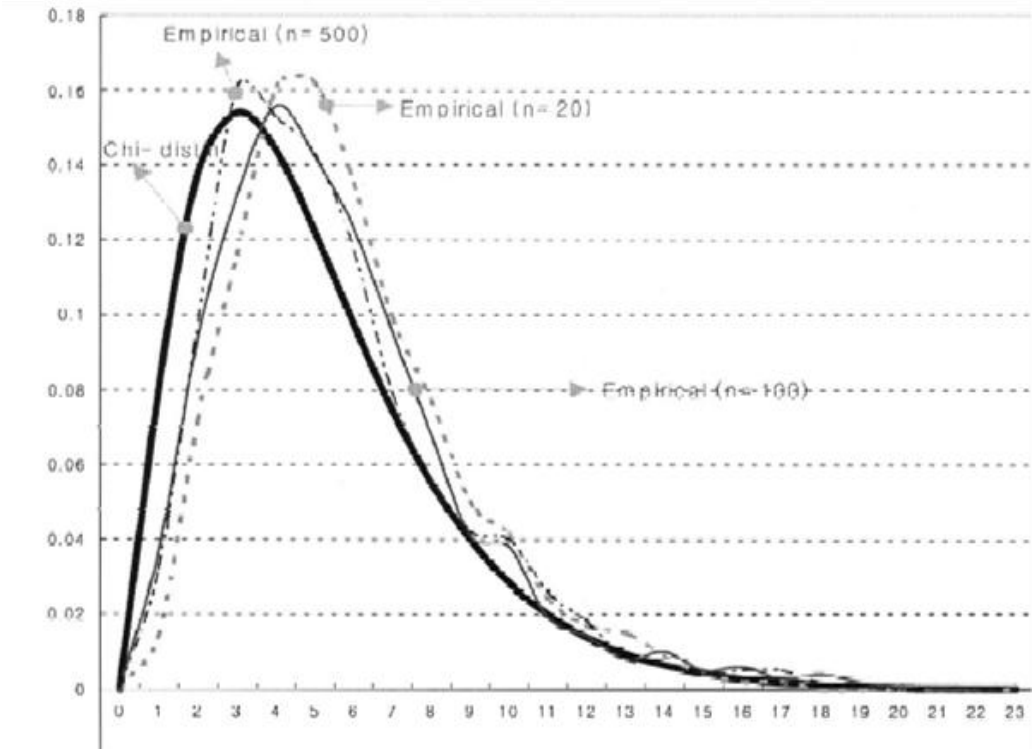
Radi jednostavnosti, posmatra se sledeći regresijski model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \varepsilon_i \quad i = 1,2,\dots n$$

$$\varepsilon_i = Z_i v_i + u_i \quad i = 1,2,\dots,n$$

gde je u_i iid sredine nula i varijanse σ_u^2, v_i je iid sredine nula i varijanse σ_v^2. Da bi se model pojednostavio, pretpostavlja se da je $cov(u_i, v_i) = 0$ što implicira $Var(\varepsilon_i) = Z_i^2\sigma_v^2 + \sigma_u^2$ tako da model postaje više heteroskedastičan kako σ_v^2 raste. Naravno, ako je σ_v^2 nula, model neće biti heteroskedastičan.

Slika 5.1 prikazuje razliku između asimptotske χ^2 raspodele i stvarne empirijske raspodele Vajtove statistike, izračunate iz prethodnog modele za obime uzorka $20, 100$ i 500. Jasno je da χ^2 raspodela zadovoljavajuće aproksimira empirijsku raspodelu Vajtove statistike kada su obimi uzorka dovoljno veliki (500). Međutim, kada su uzorci manji, stvarna empirijska raspodela W^* se dosta razlikuje od teoretske χ^2 raspodele. Odstupanja će prirodno oslabiti veličinu i moć Vajtovog testa.



Slika 5.1: [13] *Empirijska i asimptotska raspodela Vajtove statistike*

Potrebno je poboljšati osobine Vajtovog testa kod konačnog uzorka boljom aproksimacijom prave raspodele Vajtove statistike. Za to se koristi neparametarski pristup, tzv. *bootstrap* (eng.) metoda koja je bazirana na ideji reprezentativnog uzorka koji dobro predstavlja osnovnu raspodelu populacije. Ona može biti aproksimirana izvlačenjem uzorka više puta (ponovnim izvlačenjem) iz originalnog uzorka. Ključno pitanje u definisanju *bootstrap* test procedure koristeći Vajtovu statistiku je kako konstruisati raspodelu Vajtove statistike pod pretpostavkom da je tačna nulta hipoteza o homoskedastičnosti.

Predlog procedure Vajtovog testa za regresiju oblika $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ je sledeći:

i. Metodom najmanjih kvadrata izračunati ocenjenu vrednost $\widehat{\boldsymbol{\beta}}$ i reziduale e_i.

ii. Kreirati *bootstrap* grešku $\boldsymbol{\varepsilon}^*$ kao $\varepsilon_i^* = \hat{\sigma}^2 t_i^*$, $(i = 1,2,\ldots,n)$ gde je $\hat{\sigma}^2 = (\boldsymbol{e}'\boldsymbol{e})/(n-k)$ i t_i^* je slučajno izabran iz raspodele koja zadovoljava $E(\boldsymbol{t}) = 0$ i $Var(\boldsymbol{t}) = \boldsymbol{I}$.

iii. "Lažne" podatke kreirati koristeći $\boldsymbol{y}^* = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}^*$

iv. Ponovo oceniti $\boldsymbol{\beta}^* = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y}^*)$ i u skladu sa tim izračunati W^*

v. Ponoviti postupke (i) – (iv) m puta da bi se konstruisala empirijska raspodela W^* i množiti sa $1/m$ svako od $W_1^*, W_2^*, \ldots, W_m^*$.

U drugom koraku, greška t_i^* je homoskedastična, sledi da će i ε^* takođe biti homoskedastična sa varijansom $\hat{\sigma}^2$. Tada je nulta hipoteza o homoskedastičnosti veštački nametnuta i *bootstrap* metoda se može primeniti za *bootstrap* Vajtov test. Interesantno je pitanje kako kreirati t^*. Najjednostavniji način je da se slučajno izabere iz standardizovane normalne raspodele.

J.Jeong i K.Lee [13] su uz pomoć Monte Carlo simulacija ispitivali performanse *bootstrap* metode koristeći prethodno razmatran regresijski model: Promenljive X i Z su slučajno i nezavisno izabrane iz uniformne raspodele $(0,3)$, a koeficijenti β_0, β_1 i β_2 su jedinice. Pretpostavka za Vajtov šum u_i je normalna raspodela, sredine 0 i varijanse jedan. Izvor heteroskedastičnosti v_i takođe je konstruisan iz normalne raspodele sa sredinom 0, dok varijansa varira od 0 (bez heteroskedastičnosti) do 10 (visoka heteroskedastičnost), a t^* je konstruisan na sledeći način:

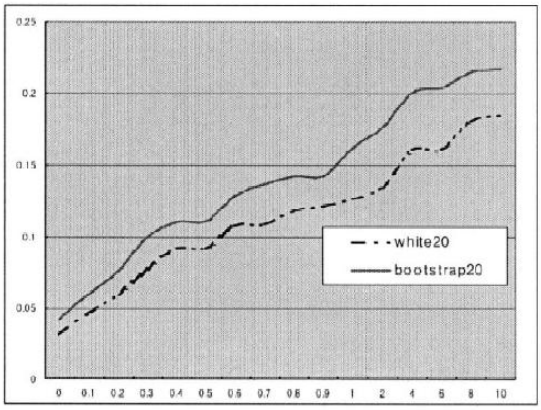
$$t_i^* = \frac{(e_i - \bar{e})}{\sqrt{\sum (e_j - \bar{e})^2}}$$

gde je $\bar{e}$ uzoračka sredina od $\boldsymbol{e}$.

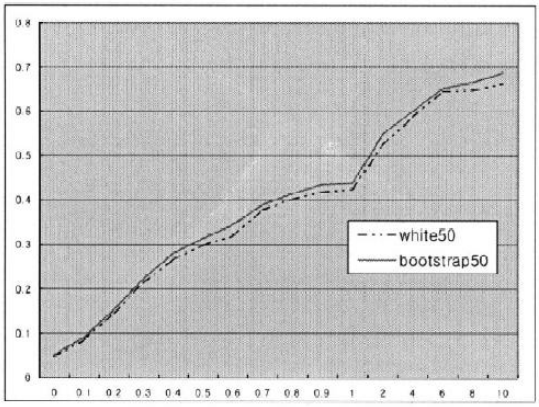
Tabela 5.3 i slika 5.2 upoređuju kritične vrednosti Vajtovog testa za varijansu σ_v^2 koristeći χ^2 raspodelu, i one vrednosti testa koristeći *bootstrap* raspodelu. Kao što se vidi iz tabele 5.1 i slike 5.2, *bootstrap* korekcija Vajtovog testa nudi značajnije pobojšanje veličine i moći Vajtovog testa kada je obim uzorka mali, kao što je npr. 20, za srednju veličinu uzorka ($n = 50$) bootstrap test pokazuje skroman ali konzistentan napredak u veličini i moći. Kada je obim uzorka veliki, kao što je 100, oba testa pokazuju sličnu moć ali je i dalje *bootstrap* test bolji od asimptotskog. Saznanje ukazuje na to da istraživači mogu izbeći potencijalnu pristrasnost ili neefikasnost ocena zbog pogrešnog zaključka o heteroskedastičnosti.

	$n = 20$		$n = 50$		$n = 100$	
σ_v^2	Asimptotski Vajtov test	Bootstrap test	Asimptotski Vajtov test	Bootstrap test	Asimptotski Vajtov test	Bootstrap test
0	0.0315	0.0420	0.048	0.052	0.045	0.050
0.1	0.0460	0.0600	0.085	0.092	0.134	0.138
0.2	0.0590	0.0760	0.146	0.154	0.313	0.322
1	0.1260	0.1620	0.423	0.436	0.857	0.863
2	0.1340	0.1765	0.523	0.549	0.941	0.946
4	0.1605	0.2005	0.587	0.601	0.968	0.971
6	0.1610	0.2040	0.644	0.653	0.975	0.977
10	0.1850	0.2170	0.662	0.686	0.984	0.985

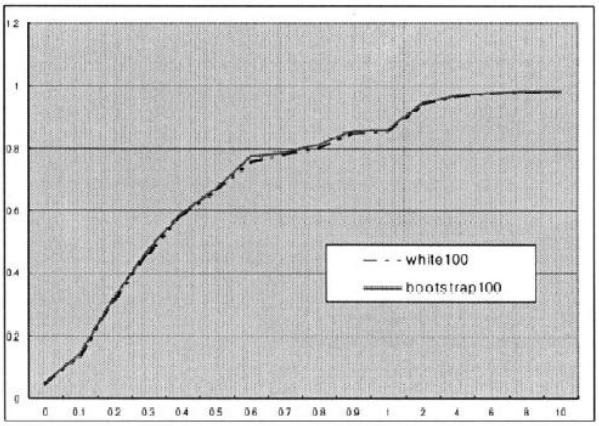
Tabela 5.3: [13] *Kritične vrednosti Vajtovog testa*



$n = 20$



$n = 50$



$n = 100$

Slika 5.2: [13] *Simulirana moć Bootstrap i asimptotskog Vajtovog testa za različite obime uzoraka*

5.2 Poboljšan Glejzerov test

U analizi metode najmanjih kvadrata, najčešće široko upotrebljavani testovi za prisutnost heteroskedastičnosti ispituju da li su kvadrati reziduala u vezi sa nekom nezavisnom promenljivom. Ovakva osnova testova se nametnula kao standardna procedura u puno ispitivanja. Mnogo veća briga se međutim fokusirala na potencijalni gubitak snage testova kada gustina poremećaja ima debele repove, što je dovelo do drugih alternativnih pristupa. Kasnije se obratila pažnja na test koji je predložio Glejzer, dizajniran da testira da li su apsolutne vrednosti reziduala u korelaciji sa nekom promenljivom.

U literaturi se postavljalo pitanje da li je Glejzerov test validan kada gustina nije simetrična. Nekoliko studija je pokazalo: Barone-Adesi i Talvar su primetili da je empirijska veličina Glejzerovog testa dosta drugačija od nominalne veličine kada greške imaju χ^2 raspodelu sa 4 stepena slobode. Međutim, Ali i Giaccotto su dobili drugačije rezultate. Naime, simulirana veličina Glejzerovog testa je mnogo bliža nominalnoj veličini kada je gustina greške lognormalna, mnogo više zakrivljena od χ_4^2.

Neka je dat linearni model

$$Y_i = x_i\beta + \varepsilon_i, \;\; i = 1,2, \dots n,$$

gde je x_i ,$1 \times k$ vektor nezavisnih promenljuvih i β, $k \times 1$ vektor parametara. Mnogi testovi heteroskedastičnosti, bazirani na redukovanom broju regresora za koje se smatra da su u korelaciji sa rezidualima, pokazuju da je njihova moć značajno veća. Tada, neka je z_i, $1 \times h$ vektor odabranih regresora od strane istraživača.

Uopšteni Glejzerov test homoskedastičnosti zasnovan je na regresiji:

$$|e_i| = \delta_0 + z_i\delta + u_i$$

i sastoji se u uočavanju veze između apsolutnih vrednosti reziduala $|e_i|$ i vektora z_i, odnosno nulta hipoze homoskedastičnosti je:

$$H_0 = E\left[z_i^{'}(|\varepsilon_i| - \underline{\mu})\right] = 0,$$

gde je $\underline{\mu} = E(|\varepsilon_i|)$. Neka je $\underline{\hat{\mu}} = \frac{1}{n}\sum|e_i|$ i $\ddot{z}_i = z_i - \frac{1}{n}\sum z_i$, Glejzerova test statistika definisana kao

$$GL = n\frac{\sum(|e_i| - \underline{\hat{\mu}})\, \ddot{z}_i(\sum \ddot{z}_i^{'}\ddot{z}_i)^{-1} \sum \ddot{z}_i^{'}(|e_i| - \underline{\hat{\mu}})}{\sum(|e_i| - \underline{\hat{\mu}})^2}$$

ima asimptotsku χ^2 raspodelu sa h stepeni slobode.

Kada je gustina greške simetrična, reziduali uspešno zamenjuju poremećaje u smislu da zamena ne menja asimptotsku raspodelu. Međutim, kako Glejzerova tast statistika zahteva da gustina greške bude simetrična, ova nerobusnost može biti prevaziđena jednostavnom modifikacijom ovog testa koju je 2000. godine otkrio Kyung So Im [14] razmatrajući sledeću regresiju:

$$|e_i| - \widehat{m}_n e_i - \underline{\hat{\mu}} = \delta_0 + z_i\delta + u_i$$

gde je $\widehat{m}_n = \frac{n_1 - n_2}{n}$, n_1 predstavlja broj pozitivnih, a n_2 broj negativnih reziduala. Nova, modifikovana test statistika će tada biti

$$MGL = n\frac{\sum(|e_i| - \widehat{m}_n e_i - \underline{\hat{\mu}})\, \ddot{z}_i (\sum \ddot{z}_i^{'} \ddot{z}_i)^{-1} \sum \ddot{z}_i^{'} (|e_i| - \widehat{m}_n e_i - \underline{\hat{\mu}})}{\sum(|e_i| - \widehat{m}_n e_i - \underline{\hat{\mu}})^2},$$

čija je asimptotska raspodela χ^2 sa h stepeni slobode ako su zadovoljeni i dodatni uslovi:

1. $E(x_i^{'} x_i \varepsilon_i^2 | \varepsilon_i > 0) = E(x_i^{'} x_i \varepsilon_i^2 | \varepsilon_i < 0) = E(x_i^{'} x_i)\sigma^2$
2. $E(x_i^{'} x_i |\varepsilon_i|) = E(x_i^{'} x_i)\underline{\mu}$
3. $E(x_i^{'} x_i \varepsilon_i) = 0.$

Iako ovi dodatni uslovi nisu previše restriktivni pod nultm hipotezom bez heteroskedastičnosti, korisno je da postoji i test koji ne zavisi od ovih uslova. Takav test je i robusan test modifikovanog Glejzerovog testa definisan kao

$$RGL = \sum(|e_i| - \widehat{m}_n e_i - \underline{\hat{\mu}})\, \ddot{z}_i \left[\sum \ddot{z}_i^{'} \ddot{z}_i\, (|e_i| - \widehat{m}_n e_i - \underline{\hat{\mu}})^2\right]^{-1} \sum \ddot{z}_i^{'} (|e_i| - \widehat{m}_n e_i - \underline{\hat{\mu}})$$

i koji teži ka χ^2 raspodeli sa h stepeni slobode ukoliko postoji šesti moment od (Y_i, x_i).

Asimptotska moć testa ne bi trebalo biti poremećena modifikacijom Glejzerovog testa jer prilagođava samo ocene asimptotske varijanse da one budu konzistentne i za nesimetrične (zakrivljene) greške. Kako bi se poredila moć Glejzerovog testa i njegove modifikacije, ispitivana je veličina testova na osnovu sledeće simulacije:

Neka su generisani podaci

$$Y_i = \alpha + x_i\beta + \varepsilon_i, \qquad i = 1,2, \dots n$$

gde je $\alpha = \beta = 1$, x_i, $i = 1,2, \dots, n$ je nezavisno generisana iz χ^2 raspodele sa jednim stepenom slobode u svakoj replikaciji. Greške su međusobno nezavisne i nezavisne u odnosu na x_i. Razmatraju su tri različite gustine grešaka: i) normalna raspodela, ii) χ^2 raspodela sa četiri stepena slobode, iii) lognormalna raspodela. Za svaku gustinu greške razmatraju se slučajevi za

$n = 100$ i $n = 500$. Birajući da je $z_i = x_i$, porede se testovi heteroskedastičnosti koji se zasnivaju na kvadriranim rezidualima (KV), Glejzerov test (GL), modifikovan Glejzerov test (MGL) i robusna verzija modifikovanog Glejzerovog testa (RGL). Broj replikacija je 5000.

Tabela 5.4 sadrži empirijske veličine računate kao stopa odbacivanja nulte hipoteze čiji je nivo značajnosti 95%. Asimptotski 95%-ni interval poverenja stope odbacivanja je $(0.044, 0.056)$.

	$n = 100$				$n = 500$			
raspodele	KV	GL	MGL	RGL	KV	GL	MGL	RGL
normalna	0.028	0.045	**0.044**	0.055	0.036	0.049	**0.048**	0.054
χ_4^2	0.033	0.110	**0.048**	0.068	0.038	0.093	**0.047**	0.063
lognormalna	0.043	0.151	**0.036**	0.072	0.041	0.155	**0.044**	0.068

Tabela 5.4: [14] *Veličina testova za različite raspodele grešaka*

Kao što se i očekivalo, empirijska veličina GL testa je dosta drugačija od nominalne veličine kada je gustina greške zakrivljena. Teži ka odbijanju nulte hipoteze veoma često i ta tendencija raste sa porastom zakrivljenja. Ipak, empirijska veličina MGL testa je blizu nominalne za svaki slučaj. Postoji blaga tendencija više odbacivanja hipoteze kod RGL testa. Iznenađujuće je to što ta tendencija ostaje gotovo na istom nivou čak i kada je $n = 500$. Slaba konvergencija ka asimptotskim raspodelama možda proizilazi iz ocene viših momenata koji su uključeni u računanje RGL testa. Iako postoji blaga tendencija većeg prihvatanja nulte hipoteze, empirijska veličina KV testova je odgovarajuća za svaki slučaj, kao što bi i trebalo da bude.

Kada se grafički ispita mogućnost prisutnosti heteroskedastičnosti, praktičnije je prikazati reziduale u zavisnosti od z_i na grafiku, nego kvadrirane reziduale u zavisnosti od z_i. Glejzerov test se može posmatrati kao formalizacija ove česte grafičke procedure. Iako srednja apsolutna devijacija nije popularna kao varijansa, svakako može bolje poslužiti kada je gustina populacije zakrivljena. Varijansa je osetljiva na outlajere koja je česta u takvoj raspodeli. U tom smislu ne iznenađuje jača moć Glejzerovog testa od standardnih testova heteroskedastičnosti zasnovanih na kvadriranim rezidualima kada gustina ima debele repove. Asimetričnost je drugi tipični simptom nenormalnosti raspodele, i često se detektuje u radu. Rezultati simulacije pokazuju da je veličina modifikovanog testa tačna za obe simetrične i nesimetrične raspodele grešaka. Dakle, Glejzerov test se čini kao privlačan test heteroskedastičnosti primenljiv u praksi.

5.3 Nedostatak Brojš-Paganovog testa

Postoji jaka zavisnost Brojš-Paganovog testa od normalnosti raspodele grešaka. Ako je ε homoskedastična ali ne i normalne raspodele, $Var(\varepsilon_i^2/\sigma^2)$ biće različita od 2 što dovodi do netačnih rezultata testa kada se koristi $SSR_{BP}/2$. U praksi, Brojš-Paganov test često odbacuje istinitu homoskedastičnost kada je raspodela ε zakrivljena.

Kroz Monte Carlo simulacije Germà Coenders and Marc Saez [15] su pokazali slabost ovog testa. Simulira se model proste regresije i sprovodi se standardni Brojš-Paganov test i modifikacija ovog testa koristeći nR^2, predložen od Koenkera i Evansa gde je n obim uzorka, a R^2 predstavlja koeficijent korelacije pomoćne regresije. Isti regresor X_1 se koristi u glavnoj

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i,$$

i pomoćnoje regresiji

$$p_i = \gamma_0 + \gamma_1 X_{1i} + v_i.$$

X_1 ima uniformnu raspodelu uzimajući pet uzastopnih celobrojnih vrednosti. R^2 glavne regresije je 0.8, a broj replikacija je 500. Simulacija je sprovedene pod pretpostavkom sledećih uslova:

1) Greške glavne regresije mogu biti homoskedastične, umereno heteroskedastične (varijansa je proporcionalna sa X_1 gde X_1 uzima vredosti od 11 do 15) i jako heteroskedastične (varijansa je proporcionalna sa X_1 gde X_1 uzima vredosti od 4 do 8).
2) Raspodela grešaka može biti platikurtična, *eng. platykurtic* (zaravnjena, u ovom slučaju uniformna), normalna, umereno zakrivljena (studentova sa 5 stepeni slobode) i jako zakrivljena, *eng.leptokurtic* (studentova sa 3 stepeni slobode).
3) Broj opservacija je $n = 500$ i $n = 100$.

Zavisne promenljive u eksperimentu predstavljaju procenat replikacija $SSR_{BP}/2$ i nR^2 statistike koje prelaze 5% kritične vrednosti χ_1^2. Rezultati su prikazani u tabeli 5.5. Prve dve kolone u tabeli sadrže stepen odbacivanja nulte hipoteze kada postoji homoskedastičnost. nR^2 statistika uvek ima stopu odbacivanja blizu teoretske od 5% čak i kada ε nije normalne raspodele ili kada mali broj opservacija možda može proizvesti netačnu asimptotsku aproksimaciju. Suprotno, originalni Brojš-Pagan test je blizu teoretske stope ako je ε normalne raspodele. Stopa odbacivanja je približna nuli ako je gustina raspodele greške *platykurtic* i postaje velika ukoliko je gustina greške *leptokurtic*, posebno kada je zakrivljenost velika.

Slučaj kada je istinita alternativna hipoteza prikazan je ostalim rezultatima u tabeli 5.5, koji predstavljaju različitu moć testova dobijenu iz različitog obima uzoraka i stepena heteroskedastičnosti. Moć je veća kod nR^2 kada je ε *platykurtic* i veća za $SSR_{BP}/2$ kada je ε

leptokurtic. Činjenica je da moć obe statistike izgleda otprilike isto kada je ε normalne raspodele, što je argumenta za korišćenje nR^2 statistike.

		homoskedastičnost		umerena heteroskedastičnost		jaka heteroskedastičnost	
n	raspodela ε	nR^2	$SSR_{BP}/2$	nR^2	$SSR_{BP}/2$	nR^2	$SSR_{BP}/2$
500	uniformna	0.042	0.022	0.797	0.360	1	0.992
500	normalna	0.034	0.032	0.388	0.388	0.950	0.946
500	t_5	0.042	0.200	0.224	0.468	0.678	0.894
500	t_3	0.044	0.470	0.104	0.526	0.296	0.738
100	uniformna	0.072	0.006	0.208	0.028	0.668	0.264
100	normalna	0.046	0.040	0.100	0.086	0.330	0.326
100	t_5	0.046	0.162	0.082	0.236	0.220	0.430
100	t_3	0.038	0.314	0.058	0.326	0.138	0.432

Tabela 5.5: [15] *Rezultati Monte Carlo eksperimenta*

Originalni Brojš-Paganov test će posebno dovoditi u zabludu kada je ε *leptokuric* i homoskedastična. Ako je veliki broj replikacija dostupan za neke kombinacije vrednosti regresora, poželjno je da se normalnost testira pre sprovođenja testa. Inače, predlaže se da se generalno koristi nR^2 test čime je Brojš-Paganov test blizak Vajtovom bez problema Vajtovog testa sa velikim stepenom slobode, tj.kada je broj ocenjenih parametara veliki koji dovodi do značajnog smanjenja moći testa.

Zaključak

Problem heteroskedastičnosti se uglavnom javlja u podacima preseka čiji su primeri navedeni u radu. Ukazali smo na posledice koje nastaju metodom najmanjih kvadrata u vidu neefikasnosti ocena i pristrasnosti ocenjene varijanse što može prouzrokovati netačnost t i F testova. Ukoliko se nastavi korišćenje uobičajenih statističkih testova i procedura bez obzira na činjenicu da je prisutna heteroskedastičnost u modelu, koji god zaključak da se dobije, može biti pogrešan.

Za otkrivanje heteroskedastičnosti se koriste grafičke metode i testovi heteroskedastičnosti, mada se ne može sa sigurnošću tvrditi koji je test najbolji. Najčešće korišćen test jeste Vajtov test, međutim, on može biti često neuspešan za male obime uzorka i ukoliko postoje greške merenja. *Bootstrap* Vajtov test pokazuje znatna pobojšanja moći ovog testa za male obime uzorka. Sa druge strane, ne tako popularan, Glejzerov test i njegova korekcija mogu biti korisniji ako greške nemaju normalnu raspodelu. Ukoliko imamo multiplikativnu heteroskedastičnost, kroz simulacije se može zaključiti da su svi testovi pogodni za njeno detektovanje. U praksi nemamo definisanu formu heteroskedastičnosti kod fitovanog regresijskog modela. Tada je najoptimalnije rešenje, kombinacija više testova. Ocenjivanje i testovi u takvoj situaciji su mnogo nesigurniji ukoliko postoje i potencijalni outlajeri. Moguće je tada uz pomoć *Jigsaw* grafika prikazati dve vrste potencijalnih ekstremnih vrednosti kako bi se testiralo postojanje outlajera i izvora heteroskedastičnosti.

Čak i kada se detektuje heteroskedastičnost, nije lako korigovati problem. Ukoliko je poznat oblik heteroskedastičnosti, može se koristiti transformacija jednačine linearne regresije i primeniti MNK. Popularni način transformacije je logaritamski oblik što je prikazano na primeru stvarnih podataka. Za velike obime uzorka može se korigovati varijansa predloženom Vajtovom metodom.

Pojedini autori smatraju da heteroskedastičnost ne predstavlja veliki problem u ekonometrijskim istraživanjima. Jedan od njih je i N. Gregori Mankiv koji tvrdi da "heteroskedastičnost nikada nije bio razlog da se odbaci inače dobar model."[4] Sa druge strane Džon Foks ističe da "nejednake varijanse grešaka treba popraviti jedino ako je problem ozbiljan. Uticaj nekonstantnih varijansi grešaka na efikasnost ocenjivanja metodom najmanjih kvadrata kao i validnost njenih zaključaka zavisi od nekoliko faktora uključujući obim uzorka, stepen variranja σ_i^2, vrednosti regresora i veze između varijanse greške i nezavisnih promenljivih. Zbog toga nije moguće doći do potpunih opštih zaključaka koji se odnosi na posledice koje proizvodi heteroskedastičnost."[5]

[4] N. Gregory Mankiw, "*A Quick Refresher Course in Macroeconomics*," Journal of Economic Literature, br. 28, 1990, p. 1648.
[5] John Fox, *Applied Regression Analysis, Linear Models, and Related Methods,* Sage Publications, California, 1997, p. 306.

Literatura

[1] H.Barreto, F.M.Howland, *Introductory to econometrics using Monte Karlo Simulation with Microsoft Excel*, Cambridge University Press, 2006

[2] G.Arbia, *Spatial Econometrics*, Springer, 2006

[3] W.H.Green, *Econometric analysis*, fifth edition, Prentice Hall, 2003

[4] D.Gujarati, *Basic econometrics*, fourth edition, The McGraw−Hill Companies, 2004

[5] Z.Mladenović, *Uvod u ekonometriju*, Centar za izdavačku delatnost Ekonomskog fakulteta, Beograd, 2011

[6] C.Dougherty, *Introduction of econometrics*, third edition, Oxford University Press, 2001

[7] Tsung-Chi Cheng, *On simultaneously identifying outliers and heteroscedasticity without specific form*, Computational Statistics and Data Analysis 56, (p.2258–2272), 2012

[8] G.S.Maddala, *Introduction to econometrics*, third edition, John Wiley & sons, LTD, 2001

[9] H.Glejser, *A new test for heteroscedasticity*, Journal of the American Statistical Association 64, 1969

[10] J. Scott Long and Laurie H. Ervin, *Correcting for Heteroscedasticity with Heteroscedasticity Consistent Standard Errors in the Linear Regression Model: Small SampleConsiderations*, Indiana University, Bloomington, IN 47405, 1998

[11] Bo Wallentin, Anders Agren *Test of heteroscedasticity in a regression model in the presence of measurement errors*, Economics Letters 76, (p.205–211), 2002

[12] B.H.Baltagi, *Econometrics*, forth edition, Springer, 2008

[13] Jinook Jeong, Kyoungwoo Lee, *Bootstrapped White's test for heteroskedasticity in regression models*, Economics Letters 63, (p.261-267), 1999

[14] Kyung So Im, *Robustifying Glejser test of heteroskedasticity*, Journal of Econometrics 97 (p.179-188), 2000

[15] Germà Coenders and Marc Saez, *Collinearity, Heteroscedasticity and Outlier Diagnostics in Regression. Do They Always Offer What They Claim?*, Metodološki zvezki, 16, Ljubljana: FDV, 2000

[16] M. Jovičić, *Ekonometrijski metodi i modeli*, Centar za izdavačku delatnost Ekonomskog fakulteta, Beograd, 2011

[17] Jan Kmenta, *Počela ekonometrije*, Zagreb: Mate, 1997

[18] D.Ruppert, *Statistics and Data Analysis for Financial Engineering*, Springer, 2010

[19] A.C.Harvey, *Estimating regression models with multiplicative heteroscedasticity*, Econometrica 44, (p.461-465), 1976

Kratka biografija

Nevena Nešić je rođena 1. aprila 1989. godine u Kruševcu. Pohađala je osnovnu školu „Sveti Sava“ u Trsteniku završivši je kao đak generacije. Takođe, završila je nižu muzičku školu, smer violina. Nakon toga upisuje Gimnaziju „Vuk Karadžić“ u Trsteniku, prirodno-matematički smer, gde je maturirala 2008. godine kao nosilac Vukove diplome.

Iste godine se upisuje na Prirodno-matematički fakultet Univerziteta u Novom Sadu, smer primenjena matematika-matematika finansija. Zaključno sa junskim ispitnim rokom završava osnovne akademske studije 2011. godine, sa prosečnom ocenom 9,88 i odmah potom, upisuje master studije primenjene matematike. U aprilu 2013.godine, položila je sve ispite predviđene planom i programom sa prosečnom ocenom 9,64 i tako stekla pravo na odbranu master rada.

Stipendista je Fonda za mlade talente Republike Srbije, kako na osnovnim tako i na master studijama.

Nevena Nešić
Novi Sad, 2014. godina

UNIVERZITET U NOVOM SADU
PRIRODNO - MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:
RBR

Identifikacioni broj:
IBR

Tip dokumentacije: Monografska dokumentacija
TD

Tip zapisa: Tekstualni štampani materijal
TZ

Vrsta rada: Master rad
VR

Autor: Nevena Nešić
AU

Mentor: dr Zorana Lužanin
MN

Naslov rada: Otkrivanje i posledice prisustva heteroskedastičnosti
NR

Jezik publikacije: srpski (latinica)
JP

Jezik izvoda: s / e
JI

Zemlja publikovanja: Republika Srbija
ZP

Uže geografsko područje: Vojvodina
UGP

Godina: 2014.
GO

Izdavač: Autorski reprint
IZ

Mesto i adresa: Novi Sad, Departman za matematiku i informatiku, Prirodno-matematički fakultet, Univerzitet u Novom Sadu, Trg Dositeja Obradovića 4
MA

Fizički opis rada: (4,92,33,19,0,17,5)
FO

Naučna oblast: Matematika
NO

Naučna disciplina: Ekonometrija
ND

Ključne reči: heteroskedastičnost, grafička metoda, testovi heteroskedastičnosti, varijansa
PO

UDK:
Čuva se: Biblioteka Departmana za matematiku i informatiku, Prirodno matematički fakultet, Univerzitet u Novom Sadu.
ČU

Važna napomena:
VN

Izvod: U prvom delu rada su navedeni osnovni pojmovi linearne regresije, metode ocenjivanja parametara i njihove osobine, oblici heteroskedastičnosti, kao i posledice njenog prisustva. U drugom delu rada se govori o metodama detektovanja problema heteroskedastičnosti gde su opisane grafičke metode i odgovarajući testovi. U poslednjem delu su simulacijom prikazani rezultati testova kao i primeri stvarnih podataka. Takođe, navedeni su neki primeri istraživanja koji prikazuju pobojšanje pojedinih testova u smislu njihove egzaktnosti i moći.
IZ

Datum prihvatanja teme od strane NN veća: 11.06.2013.
DP

Datum odbrane: februar 2014.
DO

Članovi komisije**:**
KO

Predsednik: drDanijela Rajter-Ćirić, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu
Mentor: dr Zorana Lužanin, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu
Član: dr Dora Seleši, vanredni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sadu

UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCE
KEY WORDS DOCUMENTATION

Accession number:
ANO

Identificatio number:
INO

Document type: Monograph type
DT

Type of record: Printed text
TR

Contents Code: Master's thesis
CC

Author: Nevena Nešić
AU

Mentor: Zorana Lužanin, Ph.D.
MN

Title: Detection and consequences of heteroskedasticity
TI

Language of text: Serbian
LT

Language of abstract: English
LA

Country of publication: Serbia
CP

Locality of publication: Vojvodina
LP

Publication year: 2014.
PY

Publisher: Author's reprint
PU

Publ. place: Novi Sad, Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad, Trg Dositeja Obradovića 4
PP

Physical description: (4,92,33,19,0,17,5)
PD

Scientific field: Mathematics
SF

Scientific discipline: Econometrics
SD

Subject/Key words: heteroskedasticity, graphical methods, tests of heteroskedasticity, variance
SKW

UC:
Holding data: The Library of the Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad
HD

Note:
N

Abstract: The first part of the paper presents the basic concepts of linear regression, methods of estimation parameters and their characteristics, types of heteroscedasticity as well as the consequences of its presence. In the second part of the paper discusses the methods for detecting problems of heteroscedasticity where graphical methods are described and appropriate tests. The final section presents the results of the simulation tests as well as

examples of actual data. It also gives some examples of studies that show the improvement of the tests in terms of their exactness and power.
AB

Accepted by the Scientific Board on: 11.06.2013.
ASB

Defended: 2014.
DE

Thesis defend board:
DB

President: Danijela Rajter-Ćirić, Ph.D., full professor, Faculty of Science, University of Novi Sad
Member: Zorana Lužanin, Ph.D., full professor, Faculty of Science, University of Novi Sad
Member: Dora Seleši, Ph.D., associate professor, Faculty of Science, University of Novi Sad