Ekonometrický model

Ekonometria vznikla v 30-tych rokoch tohto storočia

Model je určitá zjednodušená matematická reprezentácia skutočného javu alebo procesu, pomocou modelu daný jav analyzujeme, prognózujeme alebo riadime

Fázy pri ekonometrickom skúmaní :

1. konštrukcia samotného ekonometrického modelu - je to kvantifikovanie hypotézy,
2. kvantifikácia samotného ekonometrického modelu - štatistický odhad parametrov,
3. verifikácia modelu - všestranné posúdenie významu závisí od použitých kritérií :

3a. štatistická modifikácia - skúma sa významnosť parametrov, pomocou testovania a robí sa pri štatistickej významnosti (najčastejšie 5%) = parameter je určený s presnosťou 95% a 5% je pravdepodobnosť chyby,

3b. ekonometrická verifikácia - ako sa správa v praxi, ex post - po uplynutí určitého obdobia. Slúžia na preverenie konštrukcie modelu

4. aplikácia - prognostické - získava sa predstava o budúcich hodnotách veličín

Modely : čiastkové (jednorovnicové)
komplexné (viacrovnicové)

Všeobecným typom jednorovnicového modelu je model s viacerými nezávislými náhodnými premennými :

$y = f(X_1, X_2, ..., X_k) + u$

ak je vzťah lineárny : $y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k$

Viacrovnicový model : C_t - osobná spotreba, I_t - investície, Y_t - HDP, G_t - vládne výdavky, R_t - úr.miera

$C_t = \alpha_0 + \alpha_1 Y_t + u_{t1}$

$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \beta_3 R_t + u_{t2}$

$Y_t = C_t + I_t + G_t$

Jednorovnicový model

Jednorovnicový model má tvar : $y = y = f(X_1, X_2, ..., X_k) + u$

Produkčná funkcia : $Q = f(K,L)$, K - kapitál, L - práca

Cobb-Douglasova produkčná funkcia : $Q = AK^{\alpha}L^{\beta}e$, kde e je náhodná porucha

$A>0$, $0<\alpha<1$, $0<\beta<1$, $u = e$

$Q = AK^{\alpha}L^{\beta} + u$ $(u = e^0)$

$\log Q = \log A + \alpha \log K + \beta \log L + u$

zavedieme substitúciu : $y = \log Q$, $A^* = \log A$, $X_1 = \log K$, $X_2 = \log L$

$y = A^* + \alpha X_1 + \beta X_2 + u$ - lineárny model

Lineárny model s dvoma premennými

$y = \beta_0 + \beta_1 X + u$

y - výdavky, X - príjmy, u - náhodný činiteľ, β_0, β_1 - parametre

y, X - pozorovateľné premenné, u - nepozorovateľná náhodná premenná (porucha), β_0, β_1 - nepozorovateľné konštantné parametre

$\beta_1>0$, $\beta_1<1$ - výdavky by nemali prevyšovať príjmy

údaje získané štatisticky

(y_i, X_i), $i = 1, 2, ..., n$ - výber pozorovaní o výdavkoch a príjmoch na n domácnostiach

nakreslíme regresnú priamku

čím väčšie príjmy, tým väčšia variabilita funkcie očakávaní

regresná funkcia základného súboru : $E(y_i/X_i) = \beta_0 + \beta_1 X_i$

keďže nemáme k dispozícii celý základný súbor, ale len určitý výber pozorovaní - konštruujeme výberovú regresnú funkciu : $y_i(s) = \beta_0(s) + \beta_1(s)X_i$ kde $\beta_0(s)$, $\beta_1(s)$ sú odhady skutočných neznámych parametrov β_0, β_1

$y_i = \beta_0(s) + \beta_1(s)X_i + e_i$ kde e_i je reziduál = rozdiel medzi skutočným y_i a vyrovnaným $y_i(s)$

$y_i = E(y_i/X_i) + u_i \rightarrow u_i = y_i - E(y_i/X_i)$

Stochastická špecifikácia modelu

$y_i = E(y_i/X_i) + u_i \rightarrow E(y_i/X_i) = E[E(y_i/X_i)] + E(u_i/X_i) = E(y_i/X_i) + E(u_i/X_i) \rightarrow \mathbf{E(u_i/X_i) = 0}$

$y = f(X_1, X_2, ..., X_k, u)$ kde $k < n$, k-počet premenných, n-počet pozorovaní

špecifikuje súborový efekt náhodných vplyvov (výsledok spracovania, uchovávania inf.)

$y = \beta_0 + \beta_1 X + u$, kde u je súčtom chyby rovnice v a chyby merania w.

Štandardné predpoklady lineárneho modelu s dvoma premennými :

$y_i = \beta_0 + \beta_1 X_i + u_i$, i = 1, 2, ..., n
1). $E(u_i) = 0$, $\forall i$, poruchový člen má vo všetkých pozorovaniach nulovú strednú hodnotu
2). $E(u_i^2) = \sigma^2$ = konstanta - homoskedasticita - rozptyl náhodných porúch je vo všetkých pozorovaniach konštantný
3). $E(u_i/u_j) = 0$, pre $i \neq j$, náhodné poruchy nie sú navzájom korelované
4). vysvetľujúca premenná X je nestochastická (nenáhodná) s fixnými hodnotami X_i v opakovaných výberoch
5). $u_i \sim N(O, \sigma^2)$, náhodné premenné u_i majú normálne rozdelenie

Skúmame, čo spôsobuje narušenie týchto predpokladov :
2). ek.model má tri neznáme : β_0, β_1, σ^2 - nájsť formulu pre odhad σ^2 (pomocou reziduálov, ktoré máme vypočítať)
rozptyl - var $(u_i) = E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2$
ak je tento predpoklad narušený - heteroskedasticita
3). predpoklad o nekorelovanosti - u_i a u_j majú nulovú kovarianciu
$cov(u_i, u_j) = E\{[u_i - E(u_i)] [u_j - E(u_j)]\} = E(u_i u_j) = 0$, pre $i \neq j$
4). X_i je nenáhodná premenná, jej hodnoty sú v opakovaných výberoch y_i fixné
predpoklad : môžem vybrať ďalšie domácnosti s tými istými príjmami X_1, ..., X_n ako pri predchádzajúcom výbere
$\rightarrow cov(X_i, u_i) = 0$
5). implikuje 3.predpoklad, že náhodné premenné sú nezávislé
ak narušené $\rightarrow$ autokorelácia

Odhad parametrov modelu s dvoma premennými

$y_i = \beta_0 + \beta_1 X_i + u_i$ - nepoznáme parametre β_0, β_1
odhadujeme ich na základe výberovej regresnej funkcie $y_i(s) = \beta_0(s) + \beta_1(s) X_i$
hľadáme také $\beta_0(s)$, $\beta_1(s)$ - aby regresná priamka najlepšie vystihovala napozorované hodnoty
celková chyba je determinovaná individuálnymi chybami e_i , kde $e_i = y_i - y(s)_i$
úloha : stanoviť regresnú priamku tak, aby suma všetkých odchýlok skutočných a vyrovnaných hodnôt závisle premennej bola minimálna : $\min \sum(y_i - y(s)_i) = \min \sum[y_i - (\beta_0(s) + \beta_1(s) X_i)]$
$\sum e_i = 0$, toto kritérium nevhodné - kladné a záporné odchýlky rovnakej veľkosti sa rušia, preto :
1. minimalizovať súčet absolútnych hodnôt odchýlok : $\min \sum(i=1,n) |y_i - y(s)_i|$
2. penalizácia odchýlok - metóda najmenších štvorcov : $\min \sum(i=1,n) (y_i - y(s)_i)^2$
penalizácia odchýlok je úmerná štvorcu odchýlky
$\min \sum e_i^2 = \sum(y_i - y(s)_i)^2 = \sum(y_i - \beta_0(s) - \beta_1(s) X_i)^2$
$\sum e_i^2 = f(\beta_0(s), \beta_1(s))$ - treba nájsť extrém tejto funkcie : parciálne derivácie
$\partial f(\beta_0(s), \beta_1(s))/\partial \beta_0(s) = (-2)\sum(y_i - \beta_0(s) - \beta_1(s) X_i) = 0$
$\partial f(\beta_0(s), \beta_1(s))/\partial \beta_1(s) = (-2)\sum X_i(y_i - \beta_0(s) - \beta_1(s) X_i) = 0$
úpravou týchto rovníc dostaneme :
$\sum y_i = n\beta_0(s) + \beta_1(s)\sum X_i$
$\sum X_i y_i = \beta_0(s)\sum X_i + \beta_1(s)\sum X_i^2$
úpravou :
$\sum X_i \sum y_i = n\beta_0(s)\sum X_i + \beta_1(s)(\sum X_i)^2$
$n\sum X_i y_i = n\beta_0(s)\sum X_i + n\beta_1(s)\sum X_i^2$
odpočítame prvú rovnicu od druhej :
$\beta_1(s) = (n\sum X_i y_i - \sum X_i \sum y_i)/(n\sum X_i^2 - (\sum X_i)^2)$
dosadíme za $\beta_1(s)$ a dostaneme :
$\beta_0(s) = (\sum X_i^2 \sum y_i - \sum X_i \sum X_i y_i)/(n\sum X_i^2 - (\sum X_i)^2)$

Štatistické vlastnosti estimátorov

skúmame náhodnú premennú Z, ktorej rozdelenie pravdepodobnosti je charakterizované parametrom -0-.
Estimátor parametra -0- je funkciou výberu Z_1, Z_2, ..., Z_n a označujeme ho -0- : -0-(s) = -0-(s)(Z_1, Z_2, ..., Z_n)
Základné štatistické charakteristiky : 1.rozptyl - meria disperziu estimátora okolo strednej hodnoty : var(-0-(s)) = $E[-0-(s) - E(-0-(s))]^2$ 2.výberová chyba - rozdiel medzi odhadom a skutočnou hodnotou : -0-(s) - -0-, mení sa od výberu k výberu, 3.skreslenie - rozdiel medzi strednou hodnotou estimátora a skutočnou hodnotou parametra : E(-0-) - -0-, 4.stredná štvorcová chyba MSE - meria disperziu estimátora okolo skutočnej hodnoty parametra : MSE(-0-(s)) = $E(-0-(s) - -0-)^2$, MSE = rozptyl + $(skreslenie)^2$

Vlastnosti estimátorov - želateľné

1). Neskreslenosť(nestrannosť) - estimátor -0- sa nazýva neskresleným, ak platí : E(-0-(s)) = -0-, v kontexte opakovaných výberov je estimátor s touto vlastnosťou v priemere správny

môže sa stať, že estimátor je nestranný, ale rozptyl je veľký
2). Efektívnosť(výdatnosť) - neskreslený estimátor -0- sa nazýva efektívnym, ak je jeho rozptyl menší alebo rovný ako rozptyl ľubovoľnéhoiného neskresleného estimátora -0-(~) : $E[-0-(s) - E(-0-(s))]^2 \leq E[-0-(\sim) - E(-0-(\sim)]^2$ - nazýva sa tiež najlepším neskresleným estimátorom
čím väčšia je výdatnosť estimátorov, tým silnejšie štatistické názory možno vysloviť o odhadnutom parametre
najlepší lineárny neskreslený estimátor NLNE spĺňa : a) -0-(s) je lineárnou funkciou výberových pozorovaní, b) -0-(s) je neskreslený estimátor, c) $var(-0-(s)) \leq var(-0-(\sim))$
3). Asymptotická neskreslenosť - estimátor -0-(s) sa nazýva asymptoticky neskresleným, ak $\lim E(-0-(s)) = -0-$, $n \to \infty$, n je veľkosť výberu
4). Konzistentnosť - estimátor -0-(s) by sa mal s rastom veľkosti výberu blížiť k skutočnému -0- v tom zmysle, že ak $n \to \infty$ potom pravdepodobnosť, že -0-(s) sa bude líšiť od skutočného -0-, konverguje k nule
pre konzistentný estimátor platí : $\lim P\{|-0-(s) - -0-| < \delta\} = 1$, $n \to \infty$,
konzistentnosť sa uprednostňuje pred neskreslenosťou, zisťuje sa skúmaním správania sa rozptylu pri zvyšovaní n(ako sa to odzrkadľuje na skreslení), ak s rastom n sa skreslenie zmenšuje a rozptyl konverguje k 0, tak -0-(s) je konzistentný : $\lim MSE(-0-(s)) \leq 0$, $n \to \infty$

Štatistické vlastnosti estimátorov najmenších štvorcov lineárneho modelu s dvoma premennými

$y_i = \beta_0 + \beta_1 X_i + u_i$
$\beta_0(s), \beta_1(s)$ - odhady pomocou metódy najmenších štvorcov

Neskreslenosť

$\beta_0(s) = \sum((1/n) - Xc_i)y_i$, $c_i = x_i/\sum x_i^2$
dosadíme za $y_i \rightarrow \beta_0(s) = \beta_0 + \sum((1/n) - Xc_i)u_i$ (podmienky : $\sum c_i = 0$, $\sum c_i X_i = 1$)
$E(\beta_0(s)) = \beta_0 + \sum((1/n) - Xc_i)E(u_i) = \beta_0$ (podmienka : $E(u_i) = 0$) - $\beta_0(s)$ je neskreslený
$\beta_1(s) = \sum c_i y_i = \sum c_i(\beta_0 + \beta_1 X_i + u_i) = \beta_1 + \sum c_i u_i$
$E(\beta_1(s)) = \beta_1 + \sum c_i E(u_i) = \beta_1$ - je neskreslený

Efektívnosť(výdatnosť)

$var(\beta_0(s)) = E[\beta_0(s) - \beta_0]^2$, $var(\beta_1(s)) = [\beta_1(s) - \beta_1]^2$
$\beta_0(s) - \beta_0 = \sum((1/n) - Xc_i)u_i$
$var(\beta_0(s)) = E[\beta_0(s) - \beta_0]^2 = \sum((1/n) - Xc_i)^2 E(u_i^2) = \sigma^2 \sum((1/n) - Xc_i)^2 = \sigma^2[(1/n) + (X^2/\sum X_i^2]$
$var(\beta_0(s)) = \sigma^2[\sum X_i^2/n\sum(X_i - X)^2]$
$\beta_1(s) - \beta_1 = \sum c_i u_i$
$var(\beta_1(s)) = E[\beta_1(s) - \beta_1]^2 = E[\sum c_i u_i]^2 = E(c_1^2 u_1^2) + ... + E(c_n^2 u_n^2) + 2c_1c_2E(u_1u_2) + ... + 2c_{n-1}c_nE(u_{n-1}u_n)$
podľa predpokladu o nekorelovanosti porúch $\rightarrow E(u_iu_j) = 0$ pre $i \neq j$
$var(\beta_1(s)) = \sum[E(c_i^2 u_i^2)] = \sum c_i^2 E(u_i^2)$
podľa predpokladu o homoskedasticite $\rightarrow E(u_i^2) = \sigma^2$
$var(\beta_1(s)) = \sigma^2 \sum c_i^2 = \sigma^2/\sum(X_i - X)^2$
estimátory $\beta_0(s), \beta_1(s)$ sú priamo úmerné rozptylu náhodných porúch σ^2

Kovariancia

$cov(\beta_0(s)\beta_1(s)) = E[(\beta_0(s) - \beta_0)(\beta_1(s) - \beta_1)] = (-X\sigma^2)/\sum(X_i - X)^2$
$\beta(\sim)$ - neskreslený estimátor = $\sum b_i y_i$, $b_i = c_i + r_i$, $c_i = x_i/\sum x_i^2$
$\beta_1(\sim) = \sum b_i(\beta_0 + \beta_1 X_i + u_i)$
$E(\beta_1(\sim)) = \beta_0 \sum b_i + \beta_1 \sum b_i X_i + \sum b_i E(u_i)$, podmienky : $E(u_i) = 0$, $\sum b_i = 0$, $\sum b_i X_i = 1$
$\beta_1(\sim) = \beta_1 + \sum b_i u_i \rightarrow \beta_1(\sim) - \beta_1 = \sum b_i u_i$
$var\beta_1(\sim) = E(\beta_1(\sim) - \beta_1)^2 = E(\sum b_i u_i)^2 = \sigma^2 \sum b_i^2$
$var\beta_1(\sim) = \sigma^2[1/n\sum(X_i - X)^2] + \sigma^2 \sum r_i^2$
$var\beta_1(\sim) = var\beta_1(s) + \sigma^2 \sum r_i^2$
$\beta_1(\sim) = \beta_1(s)$ vtedy, ak to je estimátor najmenších štvorcov
$\min \sigma^2 \sum b_i^2$, podmienky : $\sum b_i = 0$, $\sum b_i X_i = 1$
$e_i = y_i - \beta_1(s)x_i$
$e_i = (\beta_1(s) - \beta_1)x_i + (u_i - u)$
$\sum e_i^2 = (\beta_1(s) - \beta_1)^2 \sum x_i^2 + \sum(u_i - u)^2 - 2(\beta(s) - \beta)\sum x_i(u_i - u)$
$E(\sum e_i^2) = E(\beta_1(s) - \beta_1)^2 \sum x_i^2 + E[(u_i - u)^2] - 2E(\beta(s) - \beta)\sum x_i(u_i - u)$
$E(\sum e_i^2) = \sigma^2 + (n-1)\sigma^2 - 2\sigma^2$
$E(\sum e_i^2) = (n-2)\sigma^2 \rightarrow \sigma^2 = E(\sum e_i^2/(n-2))$
$s^2 = \sum e_i^2/(n-2)$, s^2 je neskresleným estimátorom σ^2
$E(s^2) = \sigma^2$
namiesto rozptylov sa často používajú štandardné odchýlky :

$s(\beta_0(s)) = s\sqrt{(\sum X_i^2/n\sum(X_i - X(p))^2)}$
$s(\beta_1(s)) = s\sqrt{(1/\sum(X_i - X(p))^2)}$
Metódou najmenších štvorcov možno získať estimátory parametrov β_0 ,β_1 v tvare :
$\beta_0(s) = [(\sum X_i^2 \sum y_i - \sum X_i \sum X_i y_i)/(n\sum X_i^2 - (\sum X_i)^2]$
$\beta_1(s) = [(\sum X_i y_i - \sum X_i \sum y_i)/(n\sum X_i^2 - (\sum X_i)^2)]$

Všeobecný lineárny model

$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + u_i$, $i = 1, 2, ..., n$
y_i - závisle premenná, X_i - nezávislé premenné(vysvetlujúce), u_i - nepozorovateľná náhodná porucha, β_0, ..., β_k - neznáme parametre
alternatívny zápis : $y_i = \beta_0 + \sum(j=1,k)\beta_j X_{ij} + u_i$, $i = 1, 2, ..., n$
Model zapíšeme v maticovom tvare takto : $y = X\beta + u$, kde **dorobiť**
Štatistické predpoklady :
1). $E(u) = 0$
2). $E(uu^T) = \sigma^2 I_n$, kde I_n je jednotková matica rozmeru n*n
3). X_i sú nestochastické, t.j. v opakovaných výberoch sú prvky matice X rovnaké
$X_i \rightarrow E(X^T u) = 0$
4). $h(X) = k + 1 \leq n$
$X_0, X_1, ..., X_k \rightarrow$ ak toto platí, tak vektory sú LNezávislé
5). u_i majú normálne rozdelenie, t.j. $u \sim N(O,\sigma^2 I_n)$

Odhad parametrov všeobecného lineárneho modelu metódou najmenších štvorcov

pri odvodzovaní vyjdeme z maticového zápisu : $y = X\beta + u$
$e = y - y(s) = y - X\beta(s) \rightarrow y = X\beta(s) + e$
metóda najmenších štvorcov :
$\sum(i=1,n)e_i^2 = e^T e = (y - X\beta(s))^T(y - X\beta(s)) = y^T y - y^T X\beta(s) - \beta(s)^T X^T y + \beta(s)^T X^T X^T \beta(s)$
$e^T e = y^T y - 2\beta(s)^T X^T y + \beta(s)^T X^T X^T \beta(s)$
$\partial e^T e/\partial\beta(s) = (-2)X^T y + 2X^T X\beta(s)$
$X^T X\beta(s) = X^T y$
vynásobením maticou $(X^T X)^{-1}$ zľava :
$\beta(s) = (X^T X)^{-1} X^T y$
Vlastnosti :
1). $\beta(s)$ je neskreslený, $E(\beta(s)) = \beta$
2). $\beta(s)$ je efektívny
3). $\beta(s)$ je konzistentný
$var(\beta(s)) = \sigma^2(X^T X)^{-1}$
$E(e^T e) = \sigma^2[n - (k + 1)] \rightarrow \sigma^2 = E[e^T e/(n - (k+1))]$
$s^2 = e^T e/(n - (k+1))$
$s^2 = [(y - X\beta(s))^T(y - X\beta)/(n - (k + 1))] = [y^T[I_n - X(X^T X)^{-1}X]y/(n - (k + 1))]$
$E[s^2] = E[e^T e/(n - (k+1))] = \sigma^2$
$var(\beta(s)) = s^2(X^T X)^{-1}$
$s^2 = \sqrt{(var(\beta(s)))}$

Meranie kvality vyrovnania - koeficient determinácie

variabilitu možno merať ako vzdialenosť pozorovaných hodnôt y od ich priemeru y(p)
túto variabilitu budeme označovať ako celkový súčet štvorcov (CSŠ) :
$CSŠ = \sum(i=1,n)(y_i - y(p))^2$
dokončiť obrázok
z obrázku vidíme, že $y = y(s) + e$
úpravou (-y(p)) dostaneme :
$(y_i - y(p)) = (y_i(s) - y(p)) + e_i$, kde :
$(y_i - y(p))$ je celková vzdialenosť y od y(p),
$(y_i(s) - y(p))$ je vzdialenosť regresnej priamky y(s) od y(p),
e_i je vzdialenosť y_i od y(s) (reziduál)
$\sum(y_i - y(p))^2 = \sum(y_i(s) - y(p))^2 + \sum e_i^2$
CSŠ = VSŠ + RSŠ, kde :
VSŠ je regresný súčet štvorcov(vysvetlený súčet štvorcov),

RSŠ je súčet štvorcov reziduálov(nevysvetlený súčet štvorcov)
príslušné súčty štvorcov môžeme nazvať variabilitami :
celková variabilita = vysvetlená + nevysvetlená variabilita
Miera vyrovnania :
1 = VSŠ/CSŠ + RSŠ/CSŠ
Podiel VSŠ/CSŠ udáva, akú časť CSŠ náš model vysvetľuje pomocou premenných X_i = aká časť celkovej variability závisle premennej y je determinovaná nezávislými premennými
Koeficient determinácie : R^2 = VSŠ/CSŠ = (CSŠ - RSŠ)/CSŠ
$R^2 = 1 - RSŠ/CSŠ = 1 - (\sum e_i^2/\sum(y_i - y(p))^2)$
$0 \leq R^2 \leq 1$: ak $R^2=0 \rightarrow$model nevysvetľuje žiadnu časť variability y, ak $R^2=1 \rightarrow$všetky body y ležia na regresnej priamke y(s), R^2=0.942 - model vysvetľuje 94.2%variability a 5.8%nevysvetľuje(vplyv u)
Rozptyly sú variability podelené príslušnými stupňami voľnosti :
$var(e) = s^2 = \sum e_i^2/(n - (k+1))$
$var(y) = \sum(y_i - y(p))/(n - 1)$
Korigovaný koeficient determinácie :
$R(p)^2 = 1 - var(e)/var(y) = 1 - [\sum e_i^2/(n - (k+1))]/[\sum(y_i - y(p))/(n - 1)]$
$R(p)^2 = 1 - \sum e_i^2(n -1)/\sum(y_i - y(p))(n - (k+1)$
$R^2 \leq R(p)^2$

Intervalový odhad a testovanie hypotéz o parametroch modelu

$y = X\beta + u$
$y(s) = X\beta(s)$
Vektor $\beta_i(s)$ - bodové odhady, treba určiť interval presnosti s príslušnou pravdepodobnosťou - intervalový odhad
$y_i = \beta_0 + \beta_1 X_i + u_i$, kde $\beta_1(s)$ je sklon k spotrebe
$\beta_1(s) \rightarrow \beta_1$ (ako blízko je $\beta_1(s)$ k β_1) : nájdeme číslo ξ - dostatočne malé, že pri zadanom α : $0<\alpha<1$ bude pravdepodobnosť, že interval $(\beta_1(s) - \xi, \beta_1(s) + \xi)$ pokrýva skutočné β_1, rovná $1-\alpha$
$P\{\beta_1(s) - \xi \leq \beta_1 \leq \beta_1(s) + \xi\} = 1-\alpha$ je konfidenčný interval
$1-\alpha$ je konfidenčný koeficient, α je hladina významnosti

Veta 1: Nech $Z_1, ..., Z_n$ sú normálne a nezávisle rozdelené náhodné premenné také, že $Z_i \sim N(\mu_i, \sigma_i^2)$, potom súčet $Z = \sum c_i Z_i$, kde c_i = konst. je $Z \sim N(\sum c_i\mu_i, \sum c_i^2\sigma_i^2)$
Veta 2: Ak $Z_1, ..., Z_n \sim N(0,1)$, t.j. sú normované normálne rozdelené premenné, potom $Z = \sum(i=1,n)Z_i^2$ má χ^2 - rozdelenie s n stupňami voľnosti : $Z \sim \chi_n^2$
Veta 3: Ak $Z_1, ..., Z_n$ sú nezávislé náhodné premenné také, že $Z_i \sim \chi_{ki}^2$, potom súčet $\sum Z_i$ má tiež χ^2 - rozdelenie s $\sum ki$ stupňami voľnosti
Veta 4: Ak $Z_1 \sim N(0,1)$ a Z_2 má χ^2 - rozdelenie s k stupňami voľnosti a je nezávislá od Z_1, potom premenná t má Studentovo rozdelenie s k stupňami voľnosti :
$t = Z_1/\sqrt{(Z_2/k)} = Z_1\sqrt{(k)}/\sqrt{(Z_2)}$
Veta 5: Ak $Z_1 \sim \chi_{k1}$, $Z_2 \sim \chi_{k2}$, Z_1 a Z_2 sú nezávisle premenné, potom premenná F má Fisherovo rozdelenie s k_1 a k_2 stupňami voľnosti :
$F = (Z_1/k_1)/(Z_2/k_2)$

Konštrukcia konfidenčného intervalu pre $\beta(s)$:
$\beta(s) = (X^TX)^{-1} X^Ty$
$\beta(s) \sim N(\beta, \sigma^2(X^TX)^{-1})$
$\beta_i(s) \sim N(\beta_i, \sigma^2 a_{ii})$, kde a_{ii} je i-ty diagonálny prvok matice $(X^TX)^{-1}$
β_i normujeme : $(\beta_i(s) - \beta_i)/\sigma\sqrt{(a_{ii})} \sim N(0,1)$
pre zvolené α v tabuľke kritických hodnôt normovaného rozdelenia nájdeme $n_{\alpha/2}$ pre ktorú platí :
$P\{-n_{\alpha/2} \leq (\beta_i(s) - \beta_i)/\sigma\sqrt{(a_{ii})} \leq n_{\alpha/2}\} = 1-\alpha$
rozptyl σ^2 náhodných porúch u nepoznáme, preto :
$\sum e_i^2\sigma^2 \sim \chi_d^2$, kde $d=[n-(k+1)] \rightarrow (n-k-1)s^2/\sigma^2 \sim \chi^2_{n-k-1}$
$t = [(\beta_i(s) - \beta_i)/\sigma\sqrt{(a_{ii})}]/\sqrt{[(n-k-1)(s^2/\sigma^2)/(n-k-1)]}$
úpravou dostaneme :
$t = (\beta_i(s) - \beta_i)/s\sqrt{(a_{ii})}$, menovateľ je druhou odmocninou i-teho diagonálneho prvku odhadu variačno-kovariačnej matice $s^2(X^TX)^{-1}$
$t = (\beta_i(s) - \beta_i)/s(\beta_i(s))$
v tabuľke Studentovho rozdelenia najdeme kritickú hodnotu t_c, že :

$P\{-t_c \leq t \leq t_c\} = 1-\alpha$
dosadíme za t : $P\{-t_c \leq (\beta_i(s) - \beta_i)/s(\beta_i(s)) \leq t_c\} = 1-\alpha \rightarrow P\{\beta_i(s)-t_c s(\beta_i(s)) \leq \beta_i \leq \beta_i(s)+t_c s(\beta_i(s))\} = 1-\alpha$
Interval spoľahlivosti sa stručne označuje $\beta_i(s) \pm t_c s(\beta_i(s))$

Konfidenčný interval pre rozptyl σ^2 :
$\sum e_i^2(n-k-1)/\sigma^2 \sim \chi^2_{n-k-1}$
$P\{\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}\} = 1-\alpha$
$P\{\chi^2_{1-\alpha/2} \leq s^2(n-k-1)/\sigma^2 \leq \chi^2_{\alpha/2}\} = 1-\alpha$
$P\{ s^2(n-k-1)/\chi^2_{1-\alpha/2} \leq \sigma^2 \leq s^2(n-k-1)/\chi^2_{\alpha/2}\} = 1-\alpha$ čo je $100(1-\alpha)\%$ konfidenčný interval pre σ^2

Testovanie hypotéz o parametroch lineárneho modelu

$H_0 : \beta_i = \beta_i^*$, kde β_i^* je špecifikovaná numerická hodnota
$H_1 : \beta_i \neq \beta_i^*$
$t = (\beta_i(s) - \beta_i^*)/s(\beta_i(s))$
$P\{\beta_i^* - t_{\alpha/2} s(\beta_i(s)) \leq \beta_i(s) \leq \beta_i^* + t_{\alpha/2} s(\beta_i(s))\} = 1-\alpha$

Nulová hypotéza sa formuluje takto :
$H_0 : \beta_i = 0$
$H_1 : \beta_i \neq 0$
$t = \beta_i(s)/s(\beta_i(s)$ = odhad parametra/odhad jeho štandardnej odchýlky, ak je tento pomer v absolútnej hodnote väčší ako kritická hodnota, zamietame H_0 v prospech H_1 :
test je štatisticky významný, ak hodnota t leží v kritickej oblasti a štatisticky nevýznamný, ak leží v oblasti akceptovania
Ak $(n-k-1)>30 \rightarrow N(\mu_i, \sigma_i^2)$ - tu sa dá dokázať, že pre náš bod :
$P\{(-2) \leq (\beta_i(s) - \beta_i)/s\sqrt{(a_{ii})} \leq 2\} = 1-\alpha = 0.95$
pre $(\beta_i(s) - \beta_i)/s\sqrt{(a_{ii})}$ je kritická hodnota rovná 2 (ak pri $H_0 : \beta_i = 0$ je $\beta_i(s)/s(\beta_i) \geq 2$ potom $\beta_i(s)$ je štatisticky významný a naopak)

Testovanie modelu ako celku

$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$
$H_0 : \beta_1 = \beta_2 = 0, \text{cov}(\beta_1, \beta_2) = 0$
$P\{\beta_1 \in (a_1,b_1), \beta_2 \in (a_2,b_2)\} = (1-\alpha)^2$ - oba konfidenčné intervaly (a_1,b_1), (a_2,b_2) súčasne pokrývajú parametre β_1, β_2 s pravdepodobnosťou $(1-\alpha)(1-\alpha)=(0.95)^2$ - združená hypotéza

Pre všeobecný model $y = X\beta + u$, resp. $y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + u$ bude združená hypotéza :
$H_0 : \beta_0 = \beta_1 = ... = \beta_k = 0$
H_1 : niektoré $\beta_i \neq 0$
$(y-y(p))^2 = (y(s)-y(p))^2 + e^T e$, resp. CSŠ=VSŠ+RSŠ, resp. $(y^T y - ny(p)^2) = (\beta(s)^T X^T y - ny(p)^2) + e^T e$
tieto premenné majú χ^2 -rozdelenie a majú Fisherovo rozdelenie :
$F = (VSŠ/k)/(RSŠ/(n-k-1))$
$R^2 = VSŠ/CSŠ = (CSŠ-RSŠ)/CSŠ$
upravíme vzťah F :
$F = (VSŠ/k)/[(CSŠ-VSŠ)/(n-k-1)] = [(VSŠ/CSŠ)/((CSŠ-VSŠ)/CSŠ)][(n-k-1)/k] = R^2(n-k-1)/(1-R^2)k$
$F = (R^2/k)/((1-R^2)/(n-k-1))$
Medzi R a F je tesná súvislosť : ak $R^2 \rightarrow 0$ potom $F \rightarrow 0$, ak $R^2 \rightarrow 0$ potom $F \rightarrow \infty$

Porušenie základných predpokladov lineárneho modelu

Predpoklad 1 : Náhodné poruchy majú vo všetkých pozorovaniach nulovú strednú hodnotu, t.j. $E(u)=0$
Predpoklad 2 : Rozptyly náhodných porúch u_i sú vo všetkých pozorovaniach rovnaké : $E(u_i^2)=\sigma^2$, t.j. náhodné poruchy sú homoskedastické
Predpoklad 3 : Náhodné poruchy u_i sú navzájom nekorelované, t.j. $E(u_i u_j)=0$ pre $i \neq j$
Predpoklad 4 : Vysvetľujúce premenné sú alebo nestochastické(nenáhodné), t.j. v opakovaných výberoch sú ich hodnoty fixné, alebo ak sú stochastické, nie sú korelované s náhodnými poruchami :
$E(u_i X_j)=0$, $i=1,...,n$ a $j=1,...,k$
Predpoklad 5 : Medzi vysvetľujúcimi premennými neexistuje lineárny vzťah, t.j. sú navzájom nezávislé - hodnosť matice vysvetľujúcich premenných je rovná počtu jej stĺpcov :
$h(X)=k+1$

Predpoklad 6 : Náhodné poruchy sú normálne rozdelené, pričom stredná hodnota $E(u_i)$ a rozptyl $E(u_i^2)$ vyplývajú z predpokladu 1 a 2 : $u \sim N(O,\sigma^2 I)$

Heteroskedasticita

Heteroskedasticita je porušenie platnosti 2.predpokladu
$y = \beta_0 + \beta_1 X_1 + u_i$
Klasický predpoklad homoskedasticity :
$E(u_i^2) = \sigma^2$, i=1,2,...,n, resp. v maticovom tvare $E(uu^T) = \sigma^2 I$
Heteroskedasticita - rozptyly náhodných porúch závisia od pozorovania :
$E(u_i^2) = \sigma_i^2$, i=1,2,...,n, resp. v maticovom tvare $E(uu^T) = \phi$, kde ϕ je diagonálna matica s σ_1^2, ..., σ_n^2 na diagonále $= \sigma^2\Omega$
Dôsledky heteroskedasticity :
nevplýva na neskreslenosť ani na konzistentnosť
vplýva na efektívnosť - pri dôkaze efektívnosti vychádzame z toho, že variačno-kovariačná matica $\beta(s)$:
$var(\beta(s)) = \sigma^2(X^TX)^{-1}$, kde σ^2 je konstanta - ak neplatí $\rightarrow E(uu^T) = \phi$, $var(\beta(s)) = (X^TX)^{-1}X^TE(uu^T)X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}X^T\Omega X(X^TX)^{-1}$
skreslený odhad parametrov - zložka $s^2(X^TX)^{-1}$
t-test ani F-test nebudú presné

Zisťovanie heteroskedasticity :
Graficky : odhadneme parametre modelu za predpokladu homoskedasticity a potom analyzujeme reziduály e_i , či je y(s) v nejakom vzťahu s e^2 - neexistencia vzťahu naznačuje neprítomnosť heteroskedasticity
dokonči obrázky
Parkov test : presne testuje
Bartletov test, Goldfeld-Quandtov test

Riešenie problému (odstraňovanie) heteroskedasticity :
1). rozptyly σ_i^2 sú známe
$y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_k X_{ik} + u_i$, i=1,...,n
$y_i/\sigma_i = \beta_0(1/\sigma_i) + \beta_1(X_{i1}/\sigma_i) + ... + \beta_k(X_{ik}/\sigma_i) + u_i/\sigma_i$
$y_i^* = \beta_0 X_{i0}^* + \beta_1 X_{i1}^* + ... + \beta_k X_{ik}^* + u_i^*$ transformovaný model, kde $y_i^* = y_i/\sigma_i$, $X_{ij}^* = X_{ij}/\sigma_i$
stanovíme rozptyl transformovanej náhodnej poruchy u_i^* :
$var(u_i^*) = var(u_i/\sigma_i) = (1/\sigma_i^2)var(u_i) = (\sigma_i^2/\sigma_i^2) = 1$
náhodné poruchy v transformovanom modeli sú už homoskedastické a parametre β_i môžeme odhadnúť metódou najmenších štvorcov (ide o tzv.váženú metódu naj.štv.)
2). rozptyly σ_i^2 sú neznáme, ale možno ich odhadnúť z výberu
máme prierezové dáta, že každej množine hodnôt nezávislej premennej X_{ij} zodpovedá niekoľko rôznych hodnôt y_i , potom možno získať odhady s_i^2 rozptylov σ_i^2 z výberu, tieto odhady použijeme na konštrukciu váh w_i pre váženú metódu najmenších čtvorcov, $w_i = 1/s_i$

Autokorelácia

Autokorelácia je porušenie 3.predpokladu
u_i sú navzájom sériovo korelované, u_i je ovplyvnená náhodnou poruchou z iného pozorovania, vyskytuje sa v modeloch založených na časových radoch : $E(u_t u_{t+p}) \neq 0$, $p \neq 0$, kde t je čas

Príčiny autokorelácie :
1). zotrvačnosť - hodnota pozorovaní v čase t (GNP) závisí od hodnoty v čase t-1, až kým sa neobjaví faktor (úr.miera), ktorý tento rast znižuje
2). vplyv vynechaných premenných - náhodná porucha v modeli s vynechanou premennou je súčtom náhodnej poruchy a vynechanej premennej s pôvodného modelu
3). vplyv nesprávnej špecifikácie funkčného vzťahu :
skutočný model - $y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + u_t$ je nelineárny
odhadujeme - $y_t = \beta_0 + \beta_1 X_t + v_t$ lineárny model
$v_t = \beta_2 X_t^2 + u_t$, na grafe vidíme dôsledky tejto chyby (nadhodnotené náklady) **dokonči obrázok**

Dôsledky autokorelácie :
$y = X\beta + u$
$E(u_t u_{t+p}) \neq 0$, $p \neq 0$

$u_t = \rho u_{t-1} + \varepsilon_t$ (autoregresná schéma t-teho rádu), kde ρ je koeficient autokorelácie, je symetrický ($r_{ij}=r_{ji}$) a -1(dokonale záporná autokorelácia)$\leq\rho\leq$1(dokonale kladná autokorelácia)
$u_t u_{t-1} \rightarrow$variačno-kovariačná matica $E(uu^T)$ nemá nulové mimodiagonálne prvky, preto nemožno zapísať $\sigma^2 I$
Dôsledky :
1). $\beta(s) = (X^TX)^{-1} X^Ty$ - neskreslený estimátor, konzistentný, ale nie je efektívny
2). $s^2(X^TX)^{-1}$, kde $s^2 = e^Te/(n-k-1)$ - skreslený estimátor variačno-kovariačnej matice
3). t-test a F-test nie sú adekvátne, výsledky týchto testov by boli chybné

Testovanie autokorelácie :
1). Grafická analýza : $e_t = y_t - y_t(s)$, $e = [I_n - X(X^TX)^{-1} X^T]u$, resp. $e = Mu$
2). Durbin-Watsonov test :
$H_0 : \rho=0$
$H_1 : \rho\neq0$
Ak zamietame H_0 - estimátor $\beta(s)$ nie je adekvátny
Testovacia charakteristika $d = \sum(t=2,n)(e_t - e_{t-1})^2 / \sum(t=1,n)e_t^2$ je to pomer súčtu štvorcov diferencií za sebou nasledujúcich reziduálov k celkovému súčtu štvorcov reziduálov
V maticovom tvare $d = e^TAe/e^Te$, $e=Mu$, $d = u^TMAMu/u^TMu$, kde A=**dokonči**
Medzi d a u priama závislosť
Stopa matice A je 2(n-1), t.j. tr(A)=2(n-1), potom stredná hodnota koeficientu d :
$E(d) = (tr(A) - tr[X^TAX(X^TX)^{-1}])/(n-k-1)$
Vzťah medzi d a ρ :
$d = (\sum e_t^2 + \sum e_{t-1}^2 - 2\sum e_t e_{t-1})/\sum e_t^2$, $\sum e_t^2$ a $\sum e_{t-1}^2$ sa líšia nepatrne=približne rovné
$d \cong (2\sum e_t^2 - 2\sum e_t e_{t-1})/\sum e_t^2 \cong 2 - (2\sum e_t e_{t-1})/\sum e_t^2 \cong 2[1 - \sum e_t e_{t-1}/\sum e_t^2]$
$d = e^TAe/e^Te$, $\rho(s) = \sum e_t e_{t-1}/\sum e_t^2$, $d \cong 2(1- \rho(s))$, keďže platí $-1\leq\rho\leq1$ máme : $0\leq d\leq4$
Ak $\rho(s)=0$, t.j.reziduály a teda aj náhodné poruchy nie sú autokorelované, t.j. $\rho=0$, potom $d\cong2(1-0)=2$
Ak $\rho\in(0,1\rangle$ pozitívna autokorelácia, $d\rightarrow0$, ak $\rho\in\langle-1,0\rangle$ negatívna autokorelácia, $d\rightarrow4$
Durbin a Watson odvodili dolnú (d_L) a hornú hranicu (d_U) : závisia iba od matice A, odvodili rozdelenia pre d_L,d_U a tabelovali ich kritické hodnoty d_L^*,d_U^* , že pre pre zvolenú hladinu $\alpha=0.05$ platí :
$P\{d_U<d_U^*\}=0.05$, $P\{d_L<d_L^*\}=0.05$ tieto kritické hodnoty závisia len od n(počet pozorovaní) a k(počet vysvetľujúcich premenných)
D-W test sa líši od t a F-testu : 2 kritické body, 2 neznáme rozdelenia
D-W test : odhadneme parametre modelu, vypočítame e a d, n a k+1 máme, v tabuľkách nájdeme kritické hodnoty d_U^*,d_L^*
1). ak $d <d_L^*$ - zamietame H_0 v prospech $H_1 : \rho>0$, t.j. ide o pozitívnu autokoreláciu
2). ak $d >4 - d_L^*$ - zamietame H_0 v prospech $H_1 : \rho <0$, t.j. ide o negatívnu autokoreláciu
3). ak $d_U^* <d_U <4 - d_U^*$ - akceptujeme H_0 , neprítomnosť autokorelácie
4). ak $d_L^* <d <d_U^*$ alebo $4 - d_U^* <d <4 - d_L^*$, test neumožňuje rozhodnúť o H_0 a tým o prítomnosti autokorelácie
dokonči **obrázok**
Riešenie problému (odstránenie) autokorelácie :
$y_t = \beta_0 + \beta_1 X_{t1} +...+ \beta_k X_{tk} + u_t$
oneskoríme ho o jedno obdobie :
$y_{t-1} = \beta_0 + \beta_1 X_{t-1,1} +...+ \beta_k X_{t-1,k} + u_{t-1}$
oneskorený vzťah vynásobíme ρ a odpočítame od pôvodného :
$y_t - \rho y_{t-1} = \beta_0 - \rho\beta_0 + \beta_1(X_{t1} - \rho X_{t-1,1}) +...+ \beta_k(X_{tk} - \rho X_{t-1,k}) + u_t - \rho u_{t-1}$
to možno zapísať $y_t^* = \beta_0^* + \beta_1 X_{t1}^* +...+ \beta_k X_{tk}^* + u_t^*$ - tzv.zovšeobecnené diferencie
Metóda Cochrane-Orcuttova :
iteračná metóda , v každej iterácii sa pomocou odhadu $\rho(s)$ z predchádzajúcej iterácie vypočítajú zovšeobecnené diferencie - $y_t - \rho(s)y_{t-1}$, $X_t - \rho(s)X_{t-1,j}$, j=1,...,n a na základe toho sa získajú nové odhady parametrov β a nový odhad $\rho(s)$ z reziduálov
Postup : odhadnú sa parametre metódou najmenších štvorcov, z reziduálov sa sformuje model zodpovedajúci autoregresnej schéme - $e_t = \rho_1 e_{t-1} + \varepsilon_t$, odhadne sa $\rho_1(s)$, vypočítajú sa zovš.diferencie a vytvorí sa transformovaný model - $y_t^* = \beta_0(1-\rho_1(s)) + \beta_1 X_{t1}^* +...+ \beta_k X_{tk}^* + u_t^*$, metódou najmenších štvorcov sa odhadnú parametre a vypočítajú nové reziduály - $e_t = y_t - \beta_0(s) - \beta_1(s)X_{t1} -...-\beta_k(s)X_{tk}$, z nich sa opäť vytvorí regresný model a získa sa nový odhad ρ_2 , ... proces sa opakuje kým sa nesplní požadované kritérium konvergenice - $|\rho_r(s) - \rho_{r-1}(s)|\leq\delta$
Metóda Hildreth-Luova :

ρ sa získava delením intervalu $\langle -1,1\rangle$, ak sme už urobili D-W test, vyšetrujeme len príslušnú polovicu intervalu, t.j. $\langle -1,0)$ pri negatívnej a $(0,1\rangle$ pri pozitívnej aoutokorelácii
Postup : rozdelíme interval $(0,1\rangle$ napr. s krokom 0.1, dostaneme $\{0,0.1,...,1\}$ ako možné $\rho(s)$, pre každú z nich vypočítame zovšeobecnené diferencie y_t^*, X_{tj}^*, j=0,...,k , odhadneme parametre modelu a vypočítame sumu štvorcov reziduálov, najlepší odhad=najmenší súčet štvorcov reziduálov, v okolí najlepšieho odhadu $\rho(s)$ urobíme nové delenie intervalu s menším krokom

Multikolinearita

Multikolinearita je porušenie 5.predpokladu, t.j. predpokladu $h(X) = k+1$ - hovoríme o neprítomnosti multikolinearity
Ak $h(X)<k+1 \rightarrow$v matici X je aspoň jeden vzťah LZ$\rightarrow$matica X^TX je singulárna$\rightarrow\beta(s)=(X^TX)^{-1}X^Ty$ sa nedá vypočítať$\rightarrow$úplná multikolinearita
vyjadríme LZ - lineárna kombinácia medzi dvoma a viac X_i :
$c_1X_1 + c_2X_2 +...+ c_kX_k = 0$, c sú konštanty, aspoň jedna z nich $\neq 0$
$c_1X_1 + c_2X_2 +...+ c_kX_k + \varepsilon = 0$, kde ε je stochastický člen
Ak napr. $c_2 \neq 0$: $X_2 = -(c_1/c_2)X_1 - (c_3/c_2)X_3 - ... - (c_k/c_2)X_k$
$X_2 = -(c_1/c_2)X_1 - (c_3/c_2)X_3 - ... - (c_k/c_2)X_k - (1/c_2)\varepsilon$ - ide o približnú LK, medzi $X_1,...,X_k$ je korelačný vzťah - jeho intenzita od veľkosti stochastického člena ε

Štatistické dôsledky multikolinearity :
Ak úplná kolinearita : $r_{1,2} = 1$, $r_{1,2}^2 = 1 \rightarrow \beta_1(s)$ a $var(\beta_1(s))$ nie sú definované
Ak sa $r_{1,2}^2$ blíži k $1\rightarrow$výberový rozptyl je veľmi veľký, $cov(\beta_1(s),\beta_2(s))$ ide k ∞
$var(\beta_1(s)) = \sigma^2/[\sum(X_{t1} - X_1(p))^2][1 - r_{1,2}^2]$
$cov(\beta_1(s),\beta_2(s)) = (-\sigma^2 r_{1,2})/[(1 - r_{1,2}^2)\sqrt{(\sum(X_{t1} - X_1(p))^2(X_{t2} - X_2(p))^2}]$
S rastom korelačného koeficientu rýchlo rastie rozptyl estimátora $\beta_1(s)$:
$r_{1,2}$ - $var(\beta_1(s))$: 0 - $\sigma^2/\sum(X_{i1} - X_1(p))^2$, 0.5 - $1.33\sigma^2/\sum(X_{i1} - X_1(p))^2$, 0.7 - $1.95\sigma^2/\sum(X_{i1} - X_1(p))^2$, 0.8 - $2.78\sigma^2/\sum(X_{i1} - X_1(p))^2$, 0.9 - $5.26\sigma^2/\sum(X_{i1} - X_1(p))^2$, 0.95 - $10.26\sigma^2/\sum(X_{i1} - X_1(p))^2$, 1 - ∞
$y = X\beta + u$ - výhodnejšie je skúmať charakteristické korene matice (X^TX) :
čím menší je charakteristický koreň matice, tým väčší je rozptyl
ak sa niektorý charakteristický koreň=0 nastáva úplná kolinearita $h(X)<k+1$

Zisťovanie prítomnosti multikolinearity :
1). analýza korelácií medzi vysvetľujúcimi premennými, medzi každou dvojicou X_i a X_j vypočítame korelačný koeficient r_{ij} a všetky r_{ij} tvoria tzv. korelačnú maticu $R=\{r_{ij}\}$
ak nejaké $r_{ij} > 0.8$ - prítomnosť multikolinearity
ak nejaké $r_{ij} > R^2$(koeficient determinácie) - prítomnosť multikolinearity a odhad $\beta_i(s)$ bude nepresný
2). výpočet determinantu matice (X^TX), štandardizujeme X_i ,t.j. od každého pozorovania odpočítame X(p) a rozdiel podelíme jej štandardnou odchýlkou : $(X - X(p))/\sigma \rightarrow$prvky matice (X^TX) sú korelačné koeficienty $r_{i,j}$ a determinant matice leží v intervale (0,1) : ak $|X^TX| = 0$ - jedna alebo viac LZ, ak $|X^TX| = 1$ - stĺpce matice sú ortogonálne, t.j. multikolinearita nie je prítomná
test Farrara-Glaubera : skúma sa odchýlka od ortogonálnosti stĺpcov matice X, $\chi^2 = -[(n-1) - (2k+5)/6]\ln|X^TX|$, má χ^2 rozdelenie s (1/2)k(k-1) stupňami voľnosti, ak $\chi^2 < \chi_\alpha^2$(kritické)$\rightarrow$stĺpce matice X sú približne ortogonálne, t.j. multikolinearita nie je prítomná
3). pomocné regresie, každú z vysvetľujúcich premenných dáme do regresnej závislosti na ostatných :
$X_j = \beta_0 + \beta_1X_1 +...+ \beta_{j-1}X_{j-1} + \beta_{j+1}X_{j+1} +...+ \beta_kX_k + u$, j=1,2,...,k
ak príslušný koeficient determinácie R^2 je vysoký$\rightarrow$vysoký stupeň závislosti medzi vysvetľujúcimi premennými nevýhody - samotné pomocné regresie môžu byť ovplyvnené multikolinearitou

Odstránenie dôsledkov multikolinearity :
1). vynechanie premenných - ak zistíme medzi X_i a X_j závislosť, tak jednu z nich vynecháme a odhadujeme len parametre redukovaného modelu
nevýhody - môže spôsobovať chyby špecifikácie, výsledky budú chudobnejšie
2). využitie apriórnej informácie - vieme apriori veľkosť jedného z parametrov ($\beta_2 = 0.1\beta_1$) :
$y_i = \beta_0 + \beta_1X_{i1} + 0.1\beta_1X_{i2} + u_i = \beta_0 + \beta_1X_i + u_i$, kde $X_i = X_{i1} + 0.1X_{i2}$
metódou najmenších štvorcov získame $\beta_1(s)$ a z neho $\beta_2(s) = 0.1\beta_1(s)$
3). kombinácia prierezových údajov a údajov z časových radov - analógia 2)., apriórnu informáciu získavame z prierezových údajov, jeden parameterodhadneme z prierezových údajov, dosadíme, metódou najmenších štvorcov získame $\beta_i(s)$
4). transformácie premenných

$y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + u_t$
oneskoríme ho o jedno obdobie
$y_{t-1} = \beta_0 + \beta_1 X_{t-1,1} + \beta_2 X_{t-1,2} + u_{t-1}$
odpočítame od pôvodného
$y_t - y_{t-1} = \beta_1(X_{t1} - X_{t-1,1}) + \beta_2(X_{t2} - X_{t-1,2}) + v_t$, kde $v_t = u_t - u_{t-1}$
riešime transformovaný model, ktorý redukuje multikolinearitu
5). získanie dodatočných dát - keďže multikolinearita je výberový problém, treba nájsť iný výber tých istých premenných (rozšíriť výber)
6). metódy skresleného odhadu parametrov - najznámejšia je metóda hrebeňová regresia(ridge regresia) - perturbácia matice (X^TX) pred jej inverziou, následok - ak by bola prítomná multikolinearita, potom determinant matice by bol nízky a prvky inverznej matice by boli vysoké
hrebeňová regresia spočíva vo vzdialení matice od singularity, ridge estimátor $\beta_k(s)=(X^TX+kI)^{-1}X^Ty$, kde $k>0$
ak $k=0$, potom $\beta_k(s)=\beta(s)$,t.j. estimátor najmenších štvorcov, veľkosť skreslenia ridge estimátora je $Bias\beta_k(s) = E(\beta_k(s)) - \beta = -k(X^TX+kI)^{-1}\beta$

Prognostická aplikácia jednorovnicového lineárneho modelu

$y_i = \beta_0 + \beta_1 X_i + u_i$
Prognóza je výrok vyslovený pomocou odhadnutého modelu $y_i(s) = \beta_0(s) + \beta_1(s)X_i$ o hodnotách závislej premennej y mimo rozsah pozorovaní
Klasifikácia prognózy :
1a). bodová prognóza - jedna numerická hodnota y pre každé pozorované obdobie
1b). intervalová prognóza - pre každé obdobie vypočétame interval, v ktorom bude ležať budúca hodnota y s určitou pravdepodobnosťou
2a). ex post prognóza - robí sa za obdobie, za ktoré sú známe hodnoty X aj y, slúži na otestovanie prognostických schopností modelu a to tak, že sa vynechajú určité údaje (1rok) a tak získaný model sa aplikuje na vynechaných údajoch, ak je model správnyprognóza→ex ante
2b). ex ante prognóza - na budúce obdobie
Chyba prognózy :
$y_p(s) = \beta_0(s) + \beta_1(s)X_p(s)$, kde $X_p(s)$ je predpokladaná budúca hodnota, $y_p(s)$ je prediktor
skutočná hodnota je generovaná skutočným modelom $y_p = \beta_0 + \beta_1 X_p + u_p$ (X_p nepoznáme, len jeho odhad)
Absolútna chyba prognózy je rozdiel medzi $y_p - y_p(s)$: $y_p - y_p(s) = \beta_0 + \beta_1 X_p + u_p - \beta_0(s) - \beta_1(s)X_p(s)$
pripočítame a odpočítame na ľavej strane $\beta_0 + \beta_1 X_p(s)$:
$y_p - y_p(s) = \beta_0 + \beta_1 X_p + (\beta_0 + \beta_1 X_p(s)) - (\beta_0 + \beta_1 X_p(s)) - \beta_0(s) - \beta_1(s)X_p(s) + u_p$
úpravou dostaneme : $y_p - y_p(s) = [(\beta_0 - \beta_0(s)) + (\beta_1 - \beta_1(s))X_p(s)] + \beta_1(X_p - X_p(s)) + u_p$
Absolútna chyba je rozdelená do troch častí :
1). náhodná porucha u_p , ktorú nevieme prognózovať, vždy je prítomná→absolútna chyba nikdy≠0
2). $\beta_1(X_p - X_p(s)) = 0$, ak $X_p = X_p(s)$, $\beta_1(X_p - X_p(s)) = 0$, ak $X_p \neq X_p(s)$ - ide o chybu v prognóze vysvetľujúcej premennej
3). $[(\beta_0 - \beta_0(s)) + (\beta_1 - \beta_1(s))X_p(s)] = 0$, ak $\beta_0 = \beta_0(s)$ a $\beta_1 = \beta_1(s)$, chyba spočíva v samotnom odhade parametrov
Chybu prognózy nemožno nikdy presne určiť, len odhadnúť, snaha čo najmenej znížiť celkovú chybu

Pri výpočte prognózy $y_p(s)$ potrebujeme odhadnúť parametre $\beta_0(s),\beta_1(s)$, potom $y_p(s)$ vypočítame :
$y_p(s) = E(y_p) = \beta_0(s) + \beta_1(s)X_p$
chybu prognózy označíme e_p : $e_p = y_p - y_p(s) = \beta_0 - \beta_0(s) + (\beta_1 - \beta_1(s))X_p + u_p$
e_p má normálne rozdelenie
$E(e_p) = E[\beta_0 - \beta_0(s)] = E[\beta_1 - \beta_1(s)]X_p + E(u_p)$
keďže $\beta_0(s),\beta_1(s)$ sú neskreslené, X_p je dané a $E(u_p)=0 \rightarrow E(e_p)=0$
rozptyl $\sigma_p^2 = E(e_p^2) = var(\beta_0(s)) + X_p^2 var(\beta_1(s)) + 2X_p cov(\beta_0(s),\beta_1(s)) + \sigma^2$
$var(\beta_0(s)) = \sigma^2 \sum X_i^2 / n\sum(X_i - X(p))^2$
$var(\beta_1(s)) = \sigma^2 / \sum(X_i - X(p))^2$
$cov(\beta_0(s),\beta_1(s)) = -X(p)\sigma^2 / \sum(X_i - X(p))^2$
$\sigma_p^2 = \sigma^2[(\sum X_i^2 / n\sum(X_i - X(p))^2) - (2XX_p(p)/\sum(X_i - X(p))^2) + (X_p^2/\sum(X_i - X(p))^2) + 1]$
$\sigma_p^2 = \sigma^2[1 + (1/n) + (X(p)^2 - 2XX_p(p) + X_p^2)/\sum(X_i - X(p))^2]$
$\sigma_p^2 = \sigma^2[1 + (1/n) + (X(p) - X(p))^2/\sum(X_i - X(p))^2]$
Rozptyl chyby prognózy bude tým menší, čím :
a) väčší je výber n
b) väčšia je variabilita nezávislej premennej
c) menší je rozdiel medzi X_p a X(p)

Výpočet rozptylu chyby prognózy :
σ_p^2 odhadneme tak, že σ^2 nahradíme jeho odhadom - $s^2=\sum e_i^2/(n-k-1)$
$s_p^2 = s^2[1 + (1/n) + (X(p) - X(p))^2/\sum(X_i - X(p))^2]$
$y_p - y_p(s) \sim N(O,\sigma_p^2)$, potom $((y_p - y_p(s))/\sigma_p) \sim N(0,1)$
dá sa dokázať, že má studentovo rozdelenie
$((y_p - y_p(s))/s_p) \sim t_{n-2}$ - to umožňuje intervalový odhad chyby
ak zadáme príslušnú hladinu α, potom v tabuľke kritických hodnôt T-rozdelenia nájdeme $t_{\alpha/2}$ a ak platí :
$P\{-t_{\alpha/2} \leq (y_p - y_p(s))/s_p \leq t_{\alpha/2}\} = 1-\alpha$
konfidenčný interval $(1-\alpha)\%$ pre y_p je :
$y_p(s) - t_{\alpha/2}s_p \leq y_p \leq y_p(s) + t_{\alpha/2}s_p$
dokonči obrázok

Prognóza v prípade autokorelovaných náhodných porúch :
$y_t = \beta_0 + \beta_1 X_t + u_t$
$u_t = \sigma u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(O,\sigma_i^2)$, $|\rho|<1$, t=1,2,...,n
tento vzťah poskytuje dodatočné informácie o korelovanosti porúch
$y_p(s) = \beta_0(s) + \beta_1(s)X_p$ - pre obdobie p neobsahuje predikciu náhodnej poruchy u_p (u_p =0)
pre n+1 obdobie : $y_{n+1}(s) = \beta_0 + \beta_1 X_{n+1} + \rho u_n$
$u_{n+1}(s)$ prognózujeme na základe autoregresnej schémy - $u_{n+1} = \rho u_n + \varepsilon_{n+1} \rightarrow u_{n+1}(s) = \rho u_n$ lebo $E(\varepsilon_{n+1})=0$
pre n +2 obdobie : $u_{n+2}(s) = \rho u_{n+1}(s) = \rho^2 u_n$
pre n +3 obdobie : $u_{n+3}(s) = \rho u_{n+2}(s) = \rho^3 u_n$
pre n +s obdobie : $u_{n+s}(s) = \rho u_{n+s}(s) = \rho^s u_n$
s postupujúcim horizontom prognózy sú informácie o autokorelácii stále menej užitočné, lebo ρ konverguje k 0

Prognózu pre obdobie n +1 pomocou modelu v tvare zovšeobecnených diferencií :
$y_{n+1}^* = \beta_0(1 - \rho) + \beta_1 X_{n+1}^*$, kde $y_{n+1}^* = y_{n+1}(s) - \rho y_n$, $X_{n+1}^* = X_{n+1} - \rho X_n$
takto vypočítaná prognóza je rovnaká ako pri autoregresnej schéme
$y_{n+1}(s) = \beta_0 + \beta_1 X_{n+1} + \rho u_n$
chyba prognózy je : $e_{n+1} = y_{n+1} - y_{n+1}(s) = u_{n+1} - \rho u_n = \varepsilon_{n+1}$
chyba prognózy má normálne rozdelenie s nulovou strednou hodnotou a rozptylom :
$\sigma_p^2 = E[(u_{n+1} - \rho u_n)^2] = E(u_{n+1}^2) + \rho^2 E(u_n^2) - 2\rho E(u_{n+1}u_n) = \sigma^2 + \rho^2\sigma^2 - 2\rho^2\sigma^2 = \sigma^2(1 - \rho^2)$
Ak neberieme autokoreláciu pri prognóze do úvahy, rozptyl chyby prognózy = rozptylu náhodných porúch σ^2
Ak berieme autokoreláciu pri prognóze do úvahy, rozptyl chyby prognózy = $\sigma^2(1 - \rho^2)$
V praxi nepoznáme β_0,β_1 - získame ich odhadom pomocou metód Cochrane-Orcutt alebo Hildreth-Lu
prognózu potom vypočítame : $y_{n+1}(s) = \rho(s)y_n + \beta_0(s)(1 - \rho(s)) + \beta_1(s)(X_{n+1} - \rho(s)X_n)$

Prognostická aplikácia všeobecného modelu s k vysvetľujúcimi premennými :
$y = X\beta + u$
$\beta(s) = (X^TX)^{-1}X^Ty$
pre obdobie n+1 : $X_{n+1} = (1, X_{n+1,1}, X_{n+1,2}, ..., X_{n+1,k})$ - riadkový vektor
prognóza : $y_{n+1}(s) = X_{n+1}\beta(s)$, pre ďalšie obdobia : $y_{n+i}(s) = X_{n+i}\beta(s)$
vektor X pre ľubovoľné obdobie je X(~), potom $y(s) = X(\sim)\beta(s)$, resp. $y(\sim) = X(\sim)\beta + u(\sim)$
chyba prognózy : $e = y(\sim) - y(s) = X(\sim)\beta + u(\sim) - X(\sim)\beta(s) = u(\sim) + X(\sim)\beta - X(\sim)(X^TX)^{-1}X^Ty$
$e = u(\sim) - X(\sim)(X^TX)^{-1}X^Tu$
$\sigma_p^2 = E(e^2)$
predpokladáme, že náhodné poruchy u a u(~) sú nekorelované
$E(u^Tu(\sim))=0$ a $u_1, ..., u_n$ sú homoskedastické : $E(uu^T)= \sigma^2 I$
potom rozptyl chyby prognózy je $\sigma_p^2 = \sigma^2 + \sigma^2 X(\sim)(X^TX)^{-1}X^TX(X^TX)^{-1}X(\sim)^T = \sigma^2(1 + X(\sim)(X^TX)^{-1}X(\sim)^T)$
závisí od konkrétnych hodnôt X(~)
najmenšia možná chyba prognózy - ak rozptyl minimálny - $\min X(\sim)(X^TX)^{-1}X(\sim)^T$, ak $X_0(\sim)=1 \rightarrow$ úlohy na viazaný extrém
dokonči