

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
UNIVERZITY KOMENSKÉHO V BRATISLAVE

DIPLOMOVÁ PRÁCA

BRATISLAVA 2002

JÁN ĎURAŠ

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
UNIVERZITY KOMENSKÉHO V BRATISLAVE

Ekonomická a finančná matematika



MAXIMUM ENTROPY ECONOMETRICS

Diplomová práca

Diplomant: Ján Ďuraš

Vedúci diplomovej práce: Mgr. Marián Grendár

Bratislava 2002

Prehlasujem, že diplomovú prácu som vypracoval samostatne len na základe vedomostí získaných štúdiom a konzultácií a uvádzam všetky literárne pramene, ktoré som použil.

Ján Ďuraš

Ďakujem vedúcemu diplomovej práce Mgr. Mariánovi Grendárovi za cenné rady a pripomienky pri vedení mojej práce a tiež všetkým, ktorí mali trpezlivosť odpovedať na moje otázky.

Obsah

Úvod	4
1 Maximum Entropy a Cross Entropy	5
1.1 Ilustratívny príklad	5
1.2 Všeobecná ME formulácia	7
1.3 Všeobecná CE formulácia	9
2 Generalized Maximum Entropy a Cross Entropy	12
2.1 Reparametrizácia a zovšeobecnenie ME a CE	12
2.2 Ilustratívny príklad	16
3 Input-Output model a SAM matica	19
3.1 Leontiefov Input-Output model	19
3.2 Social Accounting Matrix	20
4 Odhady IO tabuliek a SAM matice	24
4.1 RAS	24
4.2 ME a GME formulácia	25
4.2.1 Golan, Judge a Miller ME	25
4.2.2 Alternatívna formulácia ME	27
4.2.3 Golan, Judge, Miller GME	28
4.2.4 Alternatívna formulácia GME	29
4.3 CE a GCE formulácia	32
4.4 Porovnanie jednotlivých metód	34
4.4.1 Monte Carlo simulácie	35
4.4.2 Ilustratívna SAM matica	40
4.4.3 IO SR '96-'98	41
4.4.4 SAM SR '95-'96	44
Záver	46
Literatúra	47
Prílohy	48

Úvod

Základnou úlohou, cieľom ekonometrie je získať informácie o zadanom probléme z dostupných dát. V mnohých prípadoch sú nevyhnutné podmienky a predpoklady, za ktorých je možné použiť klasické ekonometrické metódy, porušené v dôsledku toho, akým spôsobom sú dáta generované alebo ako je model špecifikovaný. Problém s dátami spočíva v skutočnosti, že dáta sú často získavané neexperimentálne, pasívnym zozbieraním, preto sú agregované, neúplné, čiastočné... Problém správnej špecifikácie modelu je taktiež pomerne komplikovaný a vyžaduje značné skúsenosti v ekonometrii, pričom ani tieto nemusia byť garantom úspechu, ak samotná formulácia problému a dostupné dáta nedávajú veľké možnosti výberu. V literatúre zaoberajúcej sa ekonometriou sa rozlišujú tieto dva druhy problémov, ktoré môžu vzniknúť pri odhadovaní parametrov

- “ill-posed, underdetermined, indeterminate problem”, teda “zle postavený alebo formulovaný” problém alebo úloha, ak je počet odhadovaných parametrov väčší ako počet pozorovaní, čo vedie k úplnému zlyhaniu konvenčných metód, kedy nie je možné určiť jednoznačné riešenie úlohy
- “ill-conditioned problem”, teda “zle podmieneným” problémom nazývame úlohu, ak nie sú odhady získané klasickými metódami dostatočne stabilné, čo býva hlavne v dôsledku neexperimentálne získaných dát

Tradičný prístup v prípade ill-posed problémov je zavedenie dodatočných predpokladov, ktoré umožnia riešiť úlohu a dostať jednoznačné riešenie. Ak však tieto predpoklady zavádzať nechceme a vychádzame len z toho, čo o probléme vieme, je vhodnou alternatívou štandardných postupov maximum entropy metóda. V prípade ill-conditioned problémov je dôležitá diagnostika, ktorá by umožnila odhalenie príčiny nestability, na základe ktorej môže byť urobená náprava. Problémy multikolinearity, heteroskedasticity či autokorelácie sú v praxi časté a ich náprava nemusí byť vždy jednoduchá. Použitie princípov maximum entropy je možné aj pre tento druh problémov a odhady parametrov pomocou tejto metódy môžu poskytnúť alternatívu výberu.

Táto práca sa zaoberá maximum a cross entropy formuláciou určenou pre riešenie ill-posed a ill-conditioned problémov v ekonometrii. Obsahuje formuláciu autorov tohto princípu, ako aj jej aplikáciu pre odhad SAM matice a IO tabuliek, pričom poukazuje aj na určité nezrovnalosti spojené s touto formuláciou a navrhuje riešenie pre odstránenie týchto nedostatkov.

1 Maximum Entropy a Cross Entropy

V tejto kapitole budú postupne prezentované princípy maximum entropy (ME) a cross entropy (CE) pre tzv. “ill-posed pure inverse problems”, teda problémy v tvare

$$\mathbf{y} = X\mathbf{p}$$

kde \mathbf{y} je vektor $T \times 1$ a X je matica $T \times K$ známych pozorovaných alebo nameraných hodnôt a našim cieľom je nájsť \mathbf{p} , pričom $T < K$ a vieme, že pre \mathbf{p} platí $\mathbf{1}'\mathbf{p} = 1$. Bude uvedená formulácia úlohy ME a CE riešením ktorej, ako uvidíme, prevedieme problém štatistickej inferencie na optimalizačnú úlohu. Následne bude načrtnutý spôsob získania numerického riešenia tejto optimalizačnej úlohy.

1.1 Ilustratívny príklad

Pre lepšie pochopenie maximum entropy princípu je vhodný nasledujúci ilustratívny príklad, ktorý uviedol Jaynes [8]. Predpokladajme že kocka, ktorou hádzame, môže nadobúdať hodnoty $k = 1, 2, \dots, 6$ a našou úlohou je určiť (odhadnúť) rozdelenie pravdepodobnosti pre nasledujúci hod, pričom jedinou informáciou, ktorú máme k dispozícii, je priemer y pre veľký počet hodov N . Ak by sme vedeli počet N_1, N_2, \dots, N_6 , teda koľko krát boli hodené jednotlivé hodnoty $k = 1, 2, \dots, 6$, mohli by sme odhadnúť pravdepodobnosti ako $\hat{p}_i = \frac{N_i}{N}$. Keďže však poznáme iba priemer y , patrí táto úloha medzi ill-posed problémy, pretože počet odhadovaných parametrov je väčší ako počet “informácií”, ktoré poznáme. Konkrétne v tomto prípade potrebujeme odhadnúť šesť parametrov, pričom máme iba dve informácie - jednu o priemere N a druhú o súčte odhadovaných parametrov, ktorý musí byť rovný 1. Ak by bol priemer napríklad $y = 3.5$, bez ďalších dostupných informácií možno predpokladať o kocke, že jej rozdelenie pravdepodobnosti pre nasledujúci hod je rovnomerné. Ak by však priemer bol značne vychýlený od 3.5, napríklad $y = 2$, kocka rovnomerné rozdelenie pravdepodobnosti zrejme mať nebude. Problém môžeme formulovať nasledovne:

Nech N_1, N_2, \dots, N_6 sú počty pokusov s výsledkom $1, 2, \dots, 6$,

$$N = \sum_{k=1}^6 N_k, \quad p_k = \frac{N_k}{N} \text{ pre } k = 1, \dots, 6$$

Keďže pokusov je N , každý so šiestimi možnými výsledkami, celkový počet možností je 6^N . Počet spôsobov, ktorými môže dôjsť k realizácii konkrétnej množiny N_1, \dots, N_6 je

$$W = \frac{N!}{\prod_k N_k!}$$

Ak použijeme monotónnu transformáciu W a Stirlingovu aproximáciu

$$\ln x! \approx x \ln x - x, \quad \text{pre } x \rightarrow \infty$$

postupne úpravami dostávame

$$\begin{aligned} \ln W &\approx N \ln N - N - \sum_{k=1}^6 N_k \ln N_k + \sum_{k=1}^6 N_k \\ \ln W &\approx N \ln N - \sum_{k=1}^6 N_k \ln N_k \\ \ln W &\approx N \ln N - \sum_{k=1}^6 p_k N \ln p_k N \\ \ln W &\approx N \ln N - \sum_{k=1}^6 p_k N \ln p_k - \sum_{k=1}^6 p_k N \ln N \\ \ln W &\approx - \sum_{k=1}^6 p_k N \ln p_k \\ \frac{1}{N} \ln W &\approx - \sum_{k=1}^6 p_k \ln p_k \end{aligned}$$

Keďže nemáme okrem priemeru a aditívnej pomienky žiadne ďalšie informácie, je rozumné spomedzi všetkých možných N_1, N_2, \dots, N_6 , ktorých výsledkom je priemer y vybrať tú množinu N_1, N_2, \dots, N_6 , ktorá mohla byť realizovaná maximálnym možným počtom spôsobov. Ak vychádzame z tejto požiadavky možno problém s kockou formulovať ako

$$\max_{\mathbf{p}} H(\mathbf{p}) = - \sum_{k=1}^6 p_k \ln p_k$$

za podmienok

$$\begin{aligned} \sum_{k=1}^6 p_k x_k &= y \\ \sum_{k=1}^6 p_k &= 1 \end{aligned}$$

kde $x_k = k$, $k=1, \dots, 6$ pričom definujeme $0 \ln 0 = 0$. Toto je formulácia problému s kockou v rámci ME princípu. Nasledujúca tabuľka ukazuje riešenia problému s kockou pre rôzne y . Ako vidno, čím je y ďalej od 3.5, tým je s výsledným rozdelením spojená menšia entropia a rozdelenie sa blíži k degenerovanému, ktoré dostávame pre $y = 1$ alebo $y = 6$, a pre ktoré je entropia nulová. Ak by sme mali okrem priemeru aj informáciu o disperzii alebo o hodnote niektorých pravdepodobností v tvare napríklad $p_i + p_j = z$ alebo $p_l = x$ možno ich zahrnúť do formulácie ME tým, že optimalizačnú úlohu rozšírime o ďalšie ohraničenia.

Tabuľka 1: Výsledné rozdelenia pravdepodobností pre ME odhad

y	\hat{p}_1	\hat{p}_2	\hat{p}_3	\hat{p}_4	\hat{p}_5	\hat{p}_6	$H(\hat{\mathbf{p}})$
1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.478	0.255	0.136	0.072	0.039	0.021	1.367
3	0.247	0.207	0.174	0.146	0.123	0.103	1.749
3.5	0.167	0.167	0.167	0.167	0.167	0.167	1.792
4	0.103	0.123	0.146	0.174	0.207	0.247	1.749
5	0.021	0.039	0.072	0.136	0.255	0.478	1.367
6	0.000	0.000	0.000	0.000	0.000	1.000	0.000

1.2 Všeobecná ME formulácia

Vychádzajme z problému v tvare $\mathbf{y} = X\mathbf{p}$, kde X je matica $T \times K$, \mathbf{y} je vektor $T \times 1$ a \mathbf{p} je $T \times 1$ vektor hľadaných neznámych parametrov s vlastnosťami pravdepodobností, pričom tento problém je ill-posed, teda $T < K$. Úlohou je určiť \mathbf{p} na základe informácie vo forme známych \mathbf{y} a X . Túto úlohu formulujeme v rámci ME princípu nasledovne

$$\max_{\mathbf{p}} H(\mathbf{p}) = - \sum_{k=1}^K p_k \ln p_k$$

za podmienok

$$\begin{aligned} \mathbf{y} &= X\mathbf{p} \\ \mathbf{p}'\mathbf{1} &= 1 \end{aligned}$$

Funkcia $H(\mathbf{p})$ je mierou entropie rozdelenia pravdepodobnosti $\mathbf{p} = (p_1, \dots, p_K)'$, túto mieru neurčitosti pre množinu javov axiomaticky definoval Shannon [11]. Maximálnu hodnotu $-\ln 1/K$ nadobúda $H(\mathbf{p})$ pre $p_1 = p_2 = \dots = p_K = 1/K$, teda keď je rozdelenie pravdepodobnosti rovnomerné. Hodnota $H(\mathbf{p})$ je naopak minimálna $H(\mathbf{p}) = 0$, ak je rozdelenie pravdepodobnosti degenerované, teda ak pre niektoré k je $p_k = 1$. Shannonovu mieru možno brať ako mieru určujúcu “vzdialenosť” medzi rovnomerným rozdelením a rozdelením \mathbf{p} . Ill-posed problem teda riešime použitím ME tak, že spomedzi všetkých riešení, ktoré spĺňajú podmienky určené obmedzeniami (momenty, aditívnosť, ...) vyberieme to, s ktorým je spojená najväčšia miera neurčitosti. V prípade že maximalizujeme funkciu entropie bez ohraničení, iba s podmienkou aditívnosti, dostávame rovnomerné rozdelenie.

Všeobecne, ak sú podmienky v tvare T funkcií $\{f_1(x), \dots, f_T(x)\}$ ME úlohu odhadu parametrov \mathbf{p} formulujeme ako

$$\max_{\mathbf{p}} H(\mathbf{p}) = - \sum_{k=1}^K p_k \ln p_k \quad (1.1)$$

za podmienok

$$\sum_{k=1}^K p_k f_t(x_k) = y_t \quad t = 1, \dots, T \quad (1.2)$$

$$\sum_{k=1}^K p_k = 1 \quad (1.3)$$

(napr. ak $f_t(x_k) = (x_k - \bar{x})^2$ dostávame druhý centrálny moment)

Úloha ME je úlohou na ohraničený extrém, aby sme ju vyriešili zostavme Lagrangeovu funkciu

$$L(\mathbf{p}, \boldsymbol{\lambda}, \mu) = - \sum_{k=1}^K p_k \ln p_k + \sum_{t=1}^T \lambda_t \left(y_t - \sum_{k=1}^K p_k f_t(x_k) \right) + \mu \left(1 - \sum_{k=1}^K p_k \right) \quad (1.4)$$

Podmienky prvého rádu majú tvar

$$\frac{\partial L}{\partial p_k} = - \ln \hat{p}_k - 1 - \sum_{t=1}^T \hat{\lambda}_t f_t(x_k) - \hat{\mu} = 0, \quad k = 1, \dots, K \quad (1.5)$$

$$\frac{\partial L}{\partial \lambda_t} = y_t - \sum_{k=1}^K \hat{p}_k f_t(x_k) = 0, \quad t = 1, \dots, T \quad (1.6)$$

$$\frac{\partial L}{\partial \mu} = 1 - \sum_{k=1}^K \hat{p}_k = 0 \quad (1.7)$$

Vyjadrením \hat{p}_k z (1.5) a dosadením do (1.6) a (1.7) dostávame

$$\begin{aligned} \hat{p}_k &= \exp \left(- \sum_{t=1}^T \hat{\lambda}_t f_t(x_k) - 1 - \hat{\mu} \right), \quad k = 1, 2, \dots, K \\ \sum_{k=1}^K \exp \left(- \sum_{t=1}^T \hat{\lambda}_t f_t(x_k) - 1 - \hat{\mu} \right) f_t(x_k) &= y_t, \quad t = 1, 2, \dots, T \\ \sum_{k=1}^K \exp \left(- \sum_{t=1}^T \hat{\lambda}_t f_t(x_k) - 1 - \hat{\mu} \right) &= 1 \end{aligned}$$

Poslednú rovnicu použijeme na elimináciu $\hat{\mu}$ čím dostávame hľadané parametre \mathbf{p} v tvare funkcie lagrangeových multiplikátorov pre ohraničenia (1.2)

$$\hat{p}_k = \frac{1}{\Omega(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_T)} \exp \left(- \sum_{t=1}^T \hat{\lambda}_t f_t(x_k) \right), \quad k = 1, 2, \dots, K \quad (1.8)$$

kde

$$\Omega(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_T) = \sum_{k=1}^K \exp \left(- \sum_{t=1}^T \hat{\lambda}_t f_t(x_k) \right)$$

je normalizačný faktor. Dosadením (1.8) do (1.1) dostávame mieru entropie H ako funkciu dát

$$H = \ln \Omega(\hat{\boldsymbol{\lambda}}) + \sum_{t=1}^T \hat{\lambda}_t y_t$$

Aby sme určili charakter nájdeného extrému, vytvoríme Hessovu maticu funkcie $H(\mathbf{p})$. Ľubovoľný prvok na diagonále tejto matice má tvar

$$\frac{\partial^2 L}{\partial p_k^2} = -\frac{1}{p_k}, \quad k = 1, \dots, K$$

a prvok mimo diagonály

$$\frac{\partial^2 L}{\partial p_i \partial p_j} = 0$$

Hessova matica

$$Hess = \begin{pmatrix} -\frac{1}{p_1} & & & 0 \\ & -\frac{1}{p_2} & & \\ & & \ddots & \\ 0 & & & -\frac{1}{p_K} \end{pmatrix}$$

je preto záporne definitná pre $p_k > 0$, čo je postačujúca podmienka pre maximum. Riešenie (1.8) vyhovuje podmienke (1.3) a p_k sú kladné, sú však funkciou lagrangeových multiplikátorov λ_t . Jedinou informáciou, ktorú môžeme použiť pre výpočet $\hat{\boldsymbol{\lambda}}$, sú podmienky (1.6). Riešenie nie je možné určiť analyticky, preto musí byť nájdené numericky. Pre uľahčenie tohto numerického výpočtu je vhodné pretransformovať primárnu úlohu s ohraničeniami na duálnu úlohu bez ohraničení. Za týmto účelom označme $\mathbf{f}(\mathbf{x}) = (f_t(x_k))_{tk}$ a dosadíme \mathbf{p} v tvare

$$\mathbf{p}(\boldsymbol{\lambda}) = \frac{1}{\Omega(\boldsymbol{\lambda})} \exp(-\mathbf{f}(\mathbf{x})\boldsymbol{\lambda})$$

do (1.4). Dostávame

$$\begin{aligned} L(\mathbf{p}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) &= -\mathbf{p}(\boldsymbol{\lambda})' \ln \mathbf{p}(\boldsymbol{\lambda}) + \boldsymbol{\lambda}' [\mathbf{y} - \mathbf{f}(\mathbf{x})\mathbf{p}(\boldsymbol{\lambda})] \\ &= -\mathbf{p}(\boldsymbol{\lambda})' [-\mathbf{f}(\mathbf{x})' \boldsymbol{\lambda} - \mathbf{1}_K \ln(\Omega(\boldsymbol{\lambda}))] + [\mathbf{y}' - \mathbf{p}(\boldsymbol{\lambda})' \mathbf{f}(\mathbf{x})'] \boldsymbol{\lambda} \\ &= \mathbf{y}' \boldsymbol{\lambda} + \mathbf{p}(\boldsymbol{\lambda})' \mathbf{1}_K \ln(\Omega(\boldsymbol{\lambda})) + \mathbf{p}(\boldsymbol{\lambda})' \mathbf{f}(\mathbf{x})' \boldsymbol{\lambda} - \mathbf{p}(\boldsymbol{\lambda})' \mathbf{f}(\mathbf{x})' \boldsymbol{\lambda} \\ &= \mathbf{y}' \boldsymbol{\lambda} + \ln(\Omega(\boldsymbol{\lambda})) \equiv M(\boldsymbol{\lambda}) \end{aligned}$$

Z vlastnosti sedlového bod typu minmax riešenia Lagrangeovej úlohy teda vieme, že maximalizačnú primárnu úlohu s ohraničeniami možno zameniť za nasledovnú minimalizačnú duálnu úlohu bez ohraničení.

$$\min_{\boldsymbol{\lambda}} M(\boldsymbol{\lambda}) = \mathbf{y}' \boldsymbol{\lambda} + \ln(\Omega(\boldsymbol{\lambda}))$$

Minimalizovaním $M(\boldsymbol{\lambda})$ určíme $\hat{\boldsymbol{\lambda}}$ a z toho $\hat{\mathbf{p}}$, hodnota $M(\hat{\boldsymbol{\lambda}})$ je hornou hranicou pre $H(\mathbf{p})$, teda $\min_{\boldsymbol{\lambda}} M(\boldsymbol{\lambda}) = \max_{\mathbf{p}} H(\mathbf{p})$.

1.3 Všeobecná CE formulácia

Okrem informácie, ktorú nám dávajú dáta, existuje v niektorých prípadoch aj určitá a priori informácia o hľadanom odhade, napríklad v ilustratívnom príklade s kockou môžeme mať k dispozícii okrem priemeru y aj prvotný odhad hľadaného rozdelenia pravdepodobnosti \mathbf{q} . Úlohu potom možno formulovať pomocou princípu cross-entropy (CE). Základnou myšlienkou relatívnej entropie je minimalizovanie vzájomnej “pseudo vzdialenosti”

prvotného odhadu \mathbf{q} a odhadu \mathbf{p} . Hľadáme teda odhad, ktorý vyhovuje ohraničeniam a je “najbližšie” ku \mathbf{q} [9].

Aby sme odvodili CE formuláciu úlohy, postupujme ako pri ME. Označme N_1, N_2, \dots, N_k počty pokusov s výsledkom k , ďalej $\mathbf{n} = (N_1, N_2, \dots, N_k)'$ a $N = \sum_k N_k$. Hľadáme

$$\arg \max_{\mathbf{n}} \pi(\mathbf{n}|\mathbf{q})$$

kde

$$\pi(\mathbf{n}|\mathbf{q}) = \frac{N!}{N_1!N_2!\dots N_k!} \prod_{k=1}^K q_k^{N_k}$$

teda spomedzi všetkých k -tic N_1, N_2, \dots, N_k , ktoré vyhovujú aditívnej požiadavke $N = \sum_k N_k$ hľadáme tú, pre ktorú je pravdepodobnosť $\pi(\mathbf{n}|\mathbf{q})$ realizácie daným apriórnym generátorom q_1, \dots, q_k najväčšia. Logaritmicou transformáciou a použitím Stirlingovej aproximácie dostávame

$$\ln \pi(\mathbf{n}|\mathbf{q}) \approx N \ln N - N - \sum_{k=1}^K N_k \ln N_k + \sum_{k=1}^K N_k + \sum_{k=1}^K N_k \ln q_k$$

označme $p_k = \frac{N_k}{N}$ a upravujme ďalej

$$\ln \pi(\mathbf{n}|\mathbf{q}) \approx N \ln N - \sum_{k=1}^K N p_k \ln N p_k + N \sum_{k=1}^K p_k \ln q_k$$

$$\ln \pi(\mathbf{n}|\mathbf{q}) \approx N \ln N - \sum_{k=1}^K N p_k \ln N - N \sum_{k=1}^K p_k \ln p_k + N \sum_{k=1}^K p_k \ln q_k$$

$$\ln \pi(\mathbf{n}|\mathbf{q}) \approx -N \sum_{k=1}^K p_k \ln p_k + N \sum_{k=1}^K p_k \ln q_k$$

$$\frac{1}{N} \ln \pi(\mathbf{n}|\mathbf{q}) \approx -N \sum_{k=1}^K p_k \ln \left(\frac{p_k}{q_k} \right)$$

Úlohu CE preto formulujeme ako

$$\min_{\mathbf{p}} KL(\mathbf{p}, \mathbf{q}) = \sum_{k=1}^K p_k \ln \left(\frac{p_k}{q_k} \right) = \sum_{k=1}^K p_k \ln p_k - \sum_{k=1}^K p_k \ln q_k$$

za podmienok

$$\sum_{k=1}^K p_k f_t(x_k) = y_t \quad t = 1, \dots, T$$

$$\sum_{k=1}^K p_k = 1$$

Tak ako v prípade ME ide o úlohu na viazaný extrém, preto vytvoríme Lagrangeovu funkciu a riešime podmienky prvého rádu $\frac{\partial L}{\partial(\cdot)} = 0$, čím dostávame

$$\hat{p}_k = \frac{q_k}{\Omega(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_T)} \exp\left(\sum_{t=1}^T \hat{\lambda}_t f_t(x_k)\right), \quad k = 1, 2, \dots, K$$

kde

$$\Omega(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_T) = \sum_{k=1}^K q_k \exp\left(\sum_{t=1}^T \hat{\lambda}_t f_t(x_k)\right)$$

V prípade, že $q_k = 1/K$, $k = 1, 2, \dots, K$, teda ak nemáme žiadnu informáciu o hľadanom rozdelení, je riešenie CE totožné s riešením ME. Duála úloha, ktorú odvodíme rovnakým postupom ako v prípade ME, má pre CE formuláciu problému tvar

$$\max_{\boldsymbol{\lambda}} M(\boldsymbol{\lambda}) = \mathbf{y}'\boldsymbol{\lambda} - \ln(\Omega(\boldsymbol{\lambda}))$$

Na riešenie duálnej úlohu pre ME alebo CE je možné použiť ľubovoľný softvér, ktorý umožňuje optimalizáciu, napríklad Matlab alebo GAMS.

2 Generalized Maximum Entropy a Cross Entropy

Použitie princípov maximum a cross-entropy na riešenie rôznych ekonometrických problémov je obmedzené formuláciou ME a CE úlohy. Značne limitujúci je fakt, že odhadované parametre musia mať charakter rozdelenia pravdepodobnosti náhodnej premennej. V tejto kapitole preto bude uvedené riešenie tohto nedostatku ako bolo prezentované v [2]. Ďalším obmedzením je, že táto metóda je určená iba na riešenie deterministických modelov, bolo by vhodné rozšíriť ju aj na riešenie stochastických modelov, ktoré sa vyskytujú v ekonometrickej praxi pri odhadovaní parametrov častejšie.

ME a CE formulované vyššie uvedeným spôsobom určia bodový odhad pre každý hľadaný parameter. S každým odhadovaným parameterom však môže byť spojená iná miera neurčitosti, preto je vhodné rozšíriť tieto formulácie tak, aby zahrňovali túto požiadavku vhodnou reparametrizáciou. Navyše ak chceme odhadovať parametre, ktoré nemajú charakter pravdepodobností a teda ich súčet nie je rovný jednej a nie sú nutne z intervalu $\langle 0, 1 \rangle$, musíme zovšeobecniť uvedený princíp ME a CE, čím dostávame tzv. generalized maximum/cross entropy (GME/GCE).

2.1 Reparametrizácia a zovšeobecnenie ME a CE

Majme teda

$$\mathbf{y} = X\boldsymbol{\beta}$$

kde X je matica $T \times K$, \mathbf{y} je vektor $T \times 1$, pričom X aj \mathbf{y} sú pozorované, známe a $\boldsymbol{\beta}$ je $T \times 1$ vektor hľadaných neznámych parametrov, ktorý môže, ale nemusí mať charakter pravdepodobností, pričom tento problém je opäť ill-posed, teda $T < K$. Predpokladajme najskôr, že $\boldsymbol{\beta}$ má vlastnosti pravdepodobností. Pre každé β_k hľadáme rozdelenie pravdepodobnosti $\mathbf{p}_k = (p_{k1}, p_{k2}, \dots, p_{kM})'$ na supporte $\mathbf{z}_k = (z_{k1}, z_{k2}, \dots, z_{kM})'$, pričom $M \geq 2$ a $z_{k1} = 0$, $z_{kM} = 1$. Ak máme k dispozícii aj apriórny odhad β_k^0 , predpokladajme, že preň existuje diskrétne rozdelenie pravdepodobnosti $\mathbf{q}_k = (q_{k1}, q_{k2}, \dots, q_{kM})'$ na tom istom supporte ako pre β_k . Môžeme teda prepísať $\boldsymbol{\beta}$ ako

$$\boldsymbol{\beta} = Z\mathbf{p}$$

kde Z je matica $K \times KM$ a \mathbf{p} je vektor $KM \times 1$

$$Z\mathbf{p} = \begin{pmatrix} \mathbf{z}'_1 & & & \\ & \mathbf{z}'_2 & & \\ & & \ddots & \\ & & & \mathbf{z}'_K \end{pmatrix} \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_K \end{pmatrix}$$

Reparametrizovaný systém má tvar

$$\mathbf{y} = XZ\mathbf{p}$$

Kedže ME, a teda aj GME formulácia, je špeciálnym prípadom CE resp. GCE, formulujeme a riešime iba všeobecnejší GCE prípad

$$\min_{\mathbf{p}} KL(\mathbf{p}, \mathbf{q}) = \sum_{k=1}^K \sum_{m=1}^M p_{km} \ln \left(\frac{p_{km}}{q_{km}} \right) \quad (2.1)$$

za podmienok

$$\sum_{k=1}^K x_{tk} \sum_{m=1}^M z_{km} p_{km} = y_t \quad t = 1, \dots, T \quad (2.2)$$

$$\sum_{k=1}^K \sum_{m=1}^M z_{km} p_{km} = 1 \quad (2.3)$$

$$\sum_{m=1}^M p_{km} = 1 \quad m = 1, 2, \dots, M \quad (2.4)$$

Postupujeme tak ako v prípade CE a vytvoríme Lagrangeovu funkciu

$$L(\mathbf{p}, \boldsymbol{\lambda}, \mu, \boldsymbol{\gamma}) = \sum_{k=1}^K \sum_{m=1}^M p_{km} \ln \left(\frac{p_{km}}{q_{km}} \right) + \sum_{t=1}^T \lambda_t \left(y_t - \sum_{k=1}^K x_{tk} \sum_{m=1}^M z_{km} p_{km} \right) \quad (2.5)$$

$$+ \mu \left(1 - \sum_{k=1}^K \sum_{m=1}^M z_{km} p_{km} \right) + \sum_{k=1}^K \gamma_k \left(1 - \sum_{m=1}^M p_{km} \right) \quad (2.6)$$

podmienky prvého rádu sú

$$\frac{\partial L}{\partial p_{km}} = \ln \left(\frac{\hat{p}_{km}}{q_{km}} \right) + 1 - \sum_{t=1}^T \hat{\lambda}_t x_{tk} z_{km} - \hat{\mu} z_{km} - \hat{\gamma}_k = 0 \quad k = 1, 2, \dots, K \quad m = 1, 2, \dots, M$$

$$\frac{\partial L}{\partial \lambda_t} = y_t - \sum_{k=1}^K \sum_{m=1}^M x_{tk} z_{km} \hat{p}_{km} = 0 \quad t = 1, 2, \dots, T$$

$$\frac{\partial L}{\partial \mu} = 1 - \sum_{k=1}^K \sum_{m=1}^M z_{km} \hat{p}_{km} = 0$$

$$\frac{\partial L}{\partial \gamma_k} = 1 - \sum_{m=1}^M \hat{p}_{km} = 0 \quad k = 1, 2, \dots, K$$

a ich riešením dostávame pre p_{km} tvar

$$\hat{p}_{km} = q_{km} \exp \left(-1 + \sum_{t=1}^T \hat{\lambda}_t x_{tk} z_{km} + \hat{\mu} z_{km} + \hat{\gamma}_k \right)$$

alebo

$$\hat{p}_{km} = \frac{q_{km}}{\Omega_k(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_T, \hat{\mu})} \exp \left(\sum_{t=1}^T \hat{\lambda}_t x_{tk} z_{km} + \hat{\mu} z_{km} \right) \quad (2.7)$$

kde

$$\Omega_k(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_T, \hat{\mu}) = \sum_{m=1}^M q_{km} \exp \left(\sum_{t=1}^T \hat{\lambda}_t x_{tk} z_{km} + \hat{\mu} z_{km} \right)$$

Pre ľahší numerický výpočet riešenia pretransformujeme primárnu úlohu s ohraničeniami na duálnu úlohu bez ohraničení. Označme preto

$$\boldsymbol{\Omega}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = (\Omega_1(\boldsymbol{\lambda}, \boldsymbol{\mu}), \Omega_2(\boldsymbol{\lambda}, \boldsymbol{\mu}), \dots, \Omega_K(\boldsymbol{\lambda}, \boldsymbol{\mu}))'$$

a dosadíme $\mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ v tvare (2.7) do (2.6) čím dostávame

$$\begin{aligned} L(\mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu}), \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})' \ln \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu}) - \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})' \ln \mathbf{q} \\ &\quad + \boldsymbol{\lambda}' [\mathbf{y} - \mathbf{XZp}(\boldsymbol{\lambda}, \boldsymbol{\mu})] + \boldsymbol{\mu}' [1 - \mathbf{1}'_{\mathbf{K}} \mathbf{Zp}(\boldsymbol{\lambda}, \boldsymbol{\mu})] \\ &= \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})' [\mathbf{Z}' \mathbf{X}' \boldsymbol{\lambda} + \mathbf{Z}' \mathbf{1}_{\mathbf{K}} \boldsymbol{\mu} - (\mathbf{1}_{\mathbf{M}} \otimes \ln(\boldsymbol{\Omega}(\boldsymbol{\lambda}, \boldsymbol{\mu}')))] \\ &\quad + [\mathbf{y}' - \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})' \mathbf{Z}' \mathbf{X}'] \boldsymbol{\lambda} + [1 - \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})' \mathbf{Z}' \mathbf{1}_{\mathbf{K}}] \boldsymbol{\mu} \\ &= \mathbf{y}' \boldsymbol{\lambda} + \boldsymbol{\mu}' - \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})' (\mathbf{1}_{\mathbf{M}} \otimes \ln(\boldsymbol{\Omega}(\boldsymbol{\lambda}, \boldsymbol{\mu}')) \\ &\quad + \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})' \mathbf{Z}' \mathbf{X}' \boldsymbol{\lambda} - \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})' \mathbf{Z}' \mathbf{X}' \boldsymbol{\lambda} \\ &\quad + \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})' \mathbf{Z}' \mathbf{1}_{\mathbf{K}} \boldsymbol{\mu} - \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu})' \mathbf{Z}' \mathbf{1}_{\mathbf{K}} \boldsymbol{\mu} \\ &= \mathbf{y}' \boldsymbol{\lambda} + \boldsymbol{\mu}' - \mathbf{1}'_{\mathbf{K}} \ln(\boldsymbol{\Omega}(\boldsymbol{\lambda}, \boldsymbol{\mu})) \equiv M(\boldsymbol{\lambda}, \boldsymbol{\mu}) \end{aligned}$$

pričom \otimes je Kroneckerov produkt matíc. Duálna úloha bez ohraničení je potom

$$\max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} M(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{y}' \boldsymbol{\lambda} + \boldsymbol{\mu}' - \mathbf{1}'_{\mathbf{K}} \ln(\boldsymbol{\Omega}(\boldsymbol{\lambda}, \boldsymbol{\mu}))$$

Ak riešime $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$, ale $\boldsymbol{\beta}$ nemá charakter pravdepodobností, je často možné ohraničiť odhadované parametre β_k na interval $\langle \beta_{kL}, \beta_{kU} \rangle$ a na tomto intervale pre β_k určiť support $\mathbf{z}_k = (z_{k1}, z_{k2}, \dots, z_{kM})'$, kde $z_{k1} = \beta_{kL}$ a $z_{kM} = \beta_{kU}$. Úlohou je potom nájsť váhy pre jednotlivé prvky supportu p_{km} a za odhad β_k následne berieme konvexnú kombináciu $\sum_m z_{km} p_{km}$. Jedinou zmenou vo formulácii oproti situácii, keď $\boldsymbol{\beta}$ má charakter pravdepodobností, je ten, že vynecháme ohraničenie (2.3), v prípade GCE riešime teda

$$\min_{\mathbf{p}} KL(\mathbf{p}, \mathbf{q}) = \sum_{k=1}^K \sum_{m=1}^M p_{km} \ln \left(\frac{p_{km}}{q_{km}} \right) \quad (2.8)$$

za podmienok

$$\sum_{k=1}^K x_{tk} \sum_{m=1}^M z_{km} p_{km} = y_t \quad t = 1, \dots, T \quad (2.9)$$

$$\sum_{m=1}^M p_{km} = 1 \quad k = 1, 2, \dots, K \quad (2.10)$$

Hľadané parametre majú tvar

$$\hat{p}_{km} = \frac{q_{km}}{\Omega_k(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_T)} \exp \left(\sum_{t=1}^T \hat{\lambda}_t x_{tk} z_{km} \right)$$

kde

$$\Omega_k(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_T) = \sum_{m=1}^M q_{km} \exp \left(\sum_{t=1}^T \hat{\lambda}_t x_{tk} z_{km} \right)$$

a duálna úloha je

$$\max_{\boldsymbol{\lambda}} M(\boldsymbol{\lambda}) = \mathbf{y}'\boldsymbol{\lambda} - \mathbf{1}'_{\mathbf{K}} \ln(\Omega(\boldsymbol{\lambda}))$$

Až doteraz sme sa zaoberali problémami v tvare $\mathbf{y} = X\mathbf{p}$. Ďalším zovšeobecnením, ktoré uvedieme, bude rozšírenie na tzv. “inverse problems with noise”, teda lineárny model v tvare $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Tento model sa bežne používa v ekonometrii pri odhadoch parametrov.

Majme teda $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, kde X je matica $T \times K$, \mathbf{y} je vektor $T \times 1$, pričom X aj \mathbf{y} sú pozorované, známe dáta, $\boldsymbol{\beta}$ je $T \times 1$ vektor hľadaných neznámych parametrov, $\boldsymbol{\epsilon}$ je $T \times 1$ vektor náhodných chýb.

V zmysle rozšírenia pre parametre, ktoré nemajú charakter pravdepodobností, predpokladajme aj teraz, že β_k možno ohraničiť na interval $\langle \beta_{kL}, \beta_{kU} \rangle$ a že na tomto intervale pre β_k určíť support $\mathbf{z}_k = (z_{k1}, z_{k2}, \dots, z_{kM})'$, kde $z_{k1} = \beta_{kL}$ a $z_{kM} = \beta_{kU}$. Rovnako pre každé ϵ_t určíme ohraničenia ϵ_{tL} a ϵ_{tU} a J -prvkový support, $2 \leq J \leq \infty$, $\mathbf{v}_t = (v_{t1}, v_{t2}, \dots, v_{tJ})'$ tak, že $v_{t1} = \epsilon_{tL}$ a $v_{tJ} = \epsilon_{tU}$. Úlohou je potom nájsť váhy pre jednotlivé prvky supportu p_{km} a w_{tj} , za odhad β_k resp. ϵ_t následne berieme konvexné kombinácie bodov supportu

$$\begin{aligned} \beta_k &= \sum_{m=1}^M z_{km} p_{km} \\ \epsilon_t &= \sum_{j=1}^J v_{tj} w_{tj} \end{aligned}$$

V maticovej forme teda reparametrizáciu môžeme zapísať ako

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \mathbf{XZ}\mathbf{p} + \boldsymbol{\epsilon}$$

kde

$$\boldsymbol{\beta} = \mathbf{Z}\mathbf{p} = \begin{pmatrix} \mathbf{z}'_1 & 0 & \cdot & 0 \\ 0 & \mathbf{z}'_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \mathbf{z}'_K \end{pmatrix} \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \cdot \\ \mathbf{p}_K \end{pmatrix}$$

a

$$\boldsymbol{\epsilon} = \mathbf{V}\mathbf{w} = \begin{pmatrix} \mathbf{v}'_1 & 0 & \cdot & 0 \\ 0 & \mathbf{v}'_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \mathbf{v}'_T \end{pmatrix} \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \cdot \\ \mathbf{w}_T \end{pmatrix}$$

Pre tento problém bola navrhnutá v práci [2] formulácia GME v tvare

$$\max_{\mathbf{p}, \mathbf{w}} H(\mathbf{p}, \mathbf{w}) = -\mathbf{p}' \ln \mathbf{p} - \mathbf{w}' \ln \mathbf{w}$$

za podmienok

$$\begin{aligned}\mathbf{XZp} + \mathbf{Vw} &= \mathbf{y} \\ (\mathbf{I}_K \otimes \mathbf{1}'_M)\mathbf{p} &= \mathbf{1}_K \\ (\mathbf{I}_T \otimes \mathbf{1}'_J)\mathbf{w} &= \mathbf{1}_T\end{aligned}$$

a v prípade, že sú známe aj apriórne odhady vo forme váh \mathbf{q}_k a \mathbf{u}_t pre prvky supportu \mathbf{z}_k a \mathbf{v}_t , je GCE formulácia podľa [2]

$$\min_{\mathbf{p}, \mathbf{w}} KL(\mathbf{p}, \mathbf{w}) = \mathbf{p}' \ln(\mathbf{p}/\mathbf{q}) + \mathbf{w}' \ln(\mathbf{w}/\mathbf{u})$$

za podmienok

$$\begin{aligned}\mathbf{XZp} + \mathbf{Vw} &= \mathbf{y} \\ (\mathbf{I}_K \otimes \mathbf{1}'_M)\mathbf{p} &= \mathbf{1}_K \\ (\mathbf{I}_T \otimes \mathbf{1}'_J)\mathbf{w} &= \mathbf{1}_T\end{aligned}$$

pričom \otimes je Kroneckerov produkt matíc.

2.2 Ilustratívny príklad

Vráťme sa teraz k ilustratívne príkladu z časti 1.1. Označme jednotlivé hľadané pravdepodobnosti v súlade s vyššie uvedeným zovšeobecnením $\beta_1, \beta_2, \dots, \beta_6$. Ak pre každú pravdepodobnosť β_i chceme určiť vlastnú mieru neurčitosti, môžeme preformulovať problém z ME na GME tak, že pre každé β_i zvolíme napríklad M-prvkový support $\mathbf{z} = (z_1, z_2, \dots, z_M)'$ kde $z_1 = 0$ a $z_2 = 1$, čím dostávame úlohu

$$\max_{\mathbf{p}} H(\mathbf{p}) = - \sum_{k=1}^6 \sum_{m=1}^M p_{km} \ln p_{km}$$

za podmienok

$$\begin{aligned}\sum_{k=1}^6 x_k \sum_{m=1}^M z_m p_{km} &= y \\ \sum_{k=1}^6 \sum_{m=1}^M z_m p_{km} &= 1 \\ \sum_{m=1}^M p_{km} &= 1 \quad k = 1, \dots, K\end{aligned}$$

Problém s takouto formuláciou je, že pre ME autori [2] ešte vychádzajú z maximalizácie W , pre GME formuláciu však už nie. Rozšírenie z ME na GME je urobené "hrubou silou" tým, že je zmenená účelová funkcia bez korektného odvodu. Ak by sme totiž chceli reparametrizovať úlohu z časti 1.1 tak, že označíme $\beta_k = z_1 p_{k1} + z_2 p_{k2} + \dots + z_M p_{kM}$,

potom $N_k = Nz_1p_{k1} + Nz_2p_{k2} + \dots + Nz_Mp_{kM} = N_{k1} + N_{k2} + \dots + N_{kM}$ a počet spôsobov, ktorými je možné realizovať konkrétnu množinu N_{km} , je preto

$$W = \frac{N!}{\prod_k \prod_m N_{km}!}$$

Ak maximalizujeme W , dostávame použitím Stirlingovej aproximácie ako vyššie

$$\begin{aligned} \ln W &\approx N^1 \ln N - N - \sum_{k=1}^K \sum_{m=1}^M (N_{km} \ln N_{km} - N_{km}) \\ \ln W &\approx N \ln N - \sum_{k=1}^K \sum_{m=1}^M N_{km} \ln z_m p_{km} N \\ \ln W &\approx N \ln N - \sum_{k=1}^K \sum_{m=1}^M N_{km} \ln z_m p_{km} - \sum_{k=1}^K \sum_{m=1}^M N_{km} \ln N \\ \ln W &\approx -N \sum_{k=1}^K \sum_{m=1}^M z_m p_{km} \ln z_m p_{km} \end{aligned}$$

Úloha by teda mala byť formulovaná nasledovne

$$\max_{\mathbf{p}} - \sum_{k=1}^6 \sum_{m=1}^M z_m p_{km} \ln z_m p_{km}$$

pri podmienkach

$$\begin{aligned} \sum_{k=1}^6 x_k \sum_{m=1}^M z_m p_{km} &= y \\ \sum_{k=1}^6 \sum_{m=1}^M z_m p_{km} &= 1 \\ \sum_{m=1}^M p_{km} &= 1 \quad k = 1, \dots, K \end{aligned}$$

Kritika formulácie GME úlohy, ako bola navrhnutá v [2], je ďalej v skutočnosti, že táto úloha nedáva vždy riešenie, ktoré je konzistentné s riešením ME v zmysle $\hat{\beta}_k = z_1 \hat{p}_{k1} + z_2 \hat{p}_{k2} + \dots + z_M \hat{p}_{kM}$, kde $\hat{\beta}_k$ je ME riešenie a \hat{p}_{km} je riešenie GME.

Tabuľka 2: Rozdelenia pravdepodobností pre ME odhad

y	\hat{p}_1	\hat{p}_2	\hat{p}_3	\hat{p}_4	\hat{p}_5	\hat{p}_6
1	1.000	0.000	0.000	0.000	0.000	0.000
1.5	0.664	0.224	0.075	0.025	0.009	0.003
2	0.478	0.255	0.136	0.072	0.039	0.021
2.5	0.347	0.240	0.165	0.114	0.079	0.054
3	0.247	0.207	0.174	0.146	0.123	0.103
3.5	0.167	0.167	0.167	0.167	0.167	0.167
4	0.103	0.123	0.146	0.174	0.207	0.247
4.5	0.054	0.079	0.114	0.165	0.240	0.347
5	0.021	0.039	0.072	0.136	0.255	0.478
5.5	0.003	0.009	0.025	0.075	0.224	0.664
6	0.000	0.000	0.000	0.000	0.000	1.000

Tabuľka 3: Rozdelenia pravdepodobností pre GME odhad

y	\hat{p}_1	\hat{p}_2	\hat{p}_3	\hat{p}_4	\hat{p}_5	\hat{p}_6
1	1.000	0.000	0.000	0.000	0.000	0.000
1.5	0.628	0.272	0.077	0.018	0.004	0.001
2	0.456	0.275	0.147	0.072	0.034	0.016
2.5	0.339	0.246	0.172	0.116	0.077	0.051
3	0.245	0.208	0.176	0.147	0.123	0.102
3.5	0.167	0.167	0.167	0.167	0.167	0.167
4	0.102	0.123	0.147	0.176	0.208	0.245
4.5	0.051	0.077	0.116	0.172	0.246	0.339
5	0.016	0.034	0.072	0.147	0.275	0.456
5.5	0.001	0.004	0.018	0.077	0.272	0.628
6	0.000	0.000	0.000	0.000	0.000	1.000

V prípade, že support je dvojprvkový $\mathbf{z} = (0, 1)'$, kedy by sme intuitívne čakali, že riešenia ME a GME budú ekvivalentné, dáva GME formulácia autorov pre niektoré y riešenie, ktoré sa líši od ME riešenia na druhom desatinnom mieste. Rozdiel medzi ME a GME sa zväčšuje s tým, ako sa y vzdaluje od 3.5. Riešenie pre alternatívnu formuláciu GME, ktorá maximalizuje W , je totožné s riešením ME.

3 Input-Output model a SAM matica

3.1 Leontiefov Input-Output model

Uvažujme o ekonomike s K odvetviami národného hospodárstva, pričom každý sektor produkuje jeden statok, a ako vstup výroby používa okrem primárnych faktorov kapitálu a práce aj statky ostatných sektorov (tzv. medzispotreba). Produkcia sektora, ktorá sa nepoužije na medzispotrebu, tvorí konečnú spotrebu. Do konečnej spotreby radíme osobnú spotrebu, spotrebu verejného sektora, investície, zásoby a export. Údaje o tejto ekonomike za daný rok (jednotlivé nominálne toky) možno potom usporiadať nasledovne

	1 ... K	konečná spotreba					hrubá produkcia
		osobná	verejná	investície	vývoz	zásoby	
1	$x_{11} \dots x_{1K}$	y_{1o}	y_{1s}	y_{1i}	y_{1e}	y_{1z}	x_1
\vdots							
K	$x_{K1} \dots x_{KK}$	y_{Ko}	y_{Ks}	y_{Ki}	y_{Ke}	y_{Kz}	x_K
dovoz	$d_1 \dots d_K$						
primárne faktory	$l_1 \dots l_K$ $k_1 \dots k_K$						
	$x_1 \dots x_K$						

kde x_{ij} je časť produkcie i -teho odvetvia, použitá ako vstup j -teho odvetvia,
 l_j a k_j sú odmeny výrobným faktorom,
 d_j je dovoz j -teho statku
 y_{ih} sú zložky konečnej spotreby podľa sektorov a odvetví
 x_k je celková produkcia k -teho odvetvia.

Leontiefov input-output model popisuje medziodvetvové vzťahy v ekonomike, nákup statku v rámci medzispotreby jedného odvetvia je súčasne predajom iného odvetvia, teda stĺpec j vyjadruje vstupy j -tého odvetvia. Zachytené toky sú v nominálnych veličinách nie vo fyzických jednotkách, vstupy sú teda nákladmi, peňažnými výdavkami, výstupy sú zdrojmi peňažných príjmov.

Ak označíme $y_k = y_{io} + y_{ks} + y_{ki} + y_{ke} + y_{kz}$ celkovú konečnú spotrebu k -teho statku a $w_k = l_k + k_k$ hodnotu pridanú spracovaním v k -tom sektore môžeme vyjadriť vzťahy z IO tabuľky ako

$$\sum_{j=1}^K x_{ij} + y_i = x_i \quad (3.1)$$

$$\sum_{i=1}^K x_{ij} + d_i + w_i = x_j \quad (3.2)$$

Podiel sporeby i -teho výrobku na jednotku výroby j -teho výrobku označme $a_{ij} = \frac{x_{ij}}{x_j}$. Tieto input koeficienty vyjadrujú priame väzby v produkcii ekonomiky, ich použitím môžeme prepísať (3.1) do maticového tvaru $\mathbf{Ax} + \mathbf{y} = \mathbf{x}$. V ďalšom sa budeme zaoberať možnosťami odhadu matice medzispotreby a input koeficientov na základe neúplnej informácie.

3.2 Social Accounting Matrix

Koncept Social Accounting Matrix (SAM) je rozšírením princípu Leontiefovho IO modelu o platby medzi sektormi ekonomiky (vládou, domácnosťami, podnikmi a zahraničím). Zobrazuje teda všetky nominálne toky v ekonomike, zachytáva výrobnú oblasť, odmeny výrobným faktorom, rozdelenie a použitie dôchodkov a ich akumuláciu. SAM predstavuje štvorcovú maticu, v ktorej sa uplatňujú dva princípy

- prvý je input-output princíp pochádzajúci z input-output tabuliek, všetky toky v SAM matici sú opísané ako toky peňažných prostriedkov od jedného subjektu (v stĺpci) k druhému (v riadku), príjmy určitého subjektu sú zároveň výdavkami iného subjektu
- v SAM sa uplatňuje aj princíp národného účtovníctva, súčet príjmov subjektu je rovný súčtu jeho výdavkov, pričom príjmy sú zachytené v riadkoch a výdavky v stĺpcoch

Napríklad: v ilustratívnej SAM matici celkové príjmy domácností z výrobných faktorov K,L (suma prvkov v riadku D) sa rovnajú celkovým výdavkom na nákup statkov produkovaných odvetviami X,Y (suma prvkov v stĺpci D). Položka v riadku X a stĺpci D je jednak výdavkom domácností na statok X a súčasne aj príjmom sektora X z predaja domácnostiam.

Tabuľka 4: Príklad SAM matice

	X	Y	K	L	D	G	Total
X					100		100
Y					100		100
K	60	40					100
L	40	60					100
D			100	100			200
G							0
Total	100	100	100	100	200	0	

X,Y odvetvia hospodárstva (produkčné sektory)
 K,L primárne faktory produkcie (kapitál a práca)
 D sektor domácností
 G vládny sektor

V riadku a stĺpci G sú iba nuly - vláda nevyberá žiadne dane a neposkytuje žiadne transfery domácnostiam.

Tabuľky SAM sa využívajú na rôzne účely: slúžia na sledovanie vývoja hlavných ukazovateľov ekonomiky, resp. na medzinárodné porovnanie, tvoria databázu pre modely všeobecnej rovnováhy a pre rôzne typy ekonomických analýz. Podľa účelu použitia sa konštruujú na rôznom stupni agregácie. Na Slovenku sa konštrukciou SAM matíc zaoberá INFOSTAT, v ďalšom použijeme ich SAM matice pre SR z rokov 1995 a 1996.

Tabuľka 5: SAM SR 1995 (agregovaná. bežné ceny. mld.SKK)

Výdavky			Výrobky a služby	Produkcia	Tvorba, rozdelenie a použitie dôchodkov						Kapitál	Zahraničie	Spolu
					Prev. prebytok	Odmeny zamestnancov	Čisté dane	Podniky	Štátna správa	Domácnosti			
Prijmy			1	2	3	4	5	6	7	8	9	10	
Výrobky a služby		1		801.8					108.1	252.7	146.6	325.8	1 635.0
Produkcia		2	1 253.5										1 253.5
Tvorba dôchodkov	Prev. prebytok	3		220.7									220.7
	Odmeny zamestnancov	4		239.6								0.7	240.3
	Čisté dane	5		-8.6									-8.6
Rozdelenie a použitie dôchodkov	Podniky	6			138.7	1.0			16.9	10.7		7.1	174.5
	Štátna správa	7	65.0		8.2	63.3	-8.6	51.9		48.1		0.6	228.6
	Domácnosti	8			73.8	175.9		28.8	70.1			4.4	353.2
Kapitál		9						85.6	31.2	40.7		-10.8	146.6
Zahraničie		10	316.4			0.07		8.1	2.2	0.8			327.8
Spolu			1 635.0	1 253.5	220.7	240.4	-8.6	174.4	228.6	353.6	146.6	327.8	

Zdroj [6].

Tabuľka 6: SAM SR 1996 (agregovaná. bežné ceny. mld.SKK)

Výdavky			Výrobky a služby	Produkcia	Tvorba, rozdelenie a použitie dôchodkov						Kapitál	Zahraničie	Spolu
					Prev. prebytok	Odmeny zamestnancov	Čisté dane	Podniky	Štátna správa	Domácnosti			
Prijmy			1	2	3	4	5	6	7	8	9	10	
Výrobky a služby		1		938.2					135.3	315.8	225.1	334.7	1949.1
Produkcia		2	1 478.4										1 478.4
Tvorba dôchodkov	Prev. prebytok	3		272.1									272.1
	Odmeny zamestnancov	4		266.9								0.4	267.3
	Čisté dane	5		1.3									1.3
Rozdelenie a použitie dôchodkov	Podniky	6			153.2	0.9			18.6	15.8		15.9	204.4
	Štátna správa	7	65.9		10.9	63.3	1.3	49.6		59.8		1	251.8
	Domácnosti	8			108	202.9		36.1	81.5			4.3	432.6
Kapitál		9						104.8	15.2	39.9		65.2	225.1
Zahraničie		10	404.8			0.3		13.9	1.2	1.3			421.5
Spolu			1 949.1	1 478.4	272.1	267.3	1.3	204.4	251.8	432.6	225.1	421.5	

Zdroj [7].

4 Odhady IO tabuliek a SAM matice

Pre rôzne ekonomické analýzy sú potrebné IO tabuľky a SAM matica, nie vždy však sú k dispozícii IO tabuľky alebo SAM matica pre aktuálne obdobie. Často trvá dlhšiu dobu kým sú údaje o ekonomike zozbierané štatistickým úradom spracované a publikované. K dispozícii môžu byť iba čiastočné údaje vo forme celkových súčtov príjmov/výdavkov pre jednotlivé sektory. Odhadnúť IO tabuľku alebo SAM maticu v prípade, ak sú známe iba stĺpcové a riadkové súčty daného roka, poprípade SAM matica za predchádzajúce obdobie, možno jednak bežne používanou RAS metódou, jednak využitím princípu ME alebo CE, keďže ide o problém odhadu K^2 parametrov na základe $2K$ údajov, a teda problém je ill-posed. Tieto úlohy sú blízke úlohe o voľbách, ktorú sformulujeme, a na jej základe poukážeme na nedostatky prístupu autorov [2], (ďalej GJM) k týmto problémom. Postupne uvedieme formulácie odhadov pomocou jednotlivých metód, pre ME a CE budú uvedené jednak odhady pre formuláciu GJM, jednak pre pozmenenú formuláciu vychádzajúcu z maximalizácie počtu možných spôsobov realizácie matice. Metódy budú porovnané na základe presnosti odhadu pre Monte Carlo simulácie a pre reálne dáta o slovenskej ekonomike.

Označme N_{ij}^1 prvky SAM matice rozmeru $K \times K$, ktoré chceme odhadnúť, ďalej označme

$$N^1 = \sum_{i=1}^K \sum_{j=1}^K N_{ij}^1 \quad N_{i\cdot}^1 = \sum_{j=1}^K N_{ij}^1 \quad N_{\cdot j}^1 = \sum_{i=1}^K N_{ij}^1$$

riadkové a stĺpcové súčty pre daný rok a obdobie N_{ij}^0 , N^0 , $N_{i\cdot}^0$, $N_{\cdot j}^0$ súčty pre predchádzajúce obdobie. Tieto označenia budeme používať v celej kapitole. Poznamenajme že pre SAM maticu platí $N_{k\cdot} = N_{\cdot k}$.

4.1 RAS

RAS procedúru postupných biproporcionálnych zmien pôvodne vytvorili Stone a kol. [12] pre odhad prvkov aktuálnej IO tabuľky na základe informácie z predchádzajúceho obdobia a aktuálnych riadkových a stĺpcových súčtov. Túto procedúru je ale možné použiť na ľubovoľnú maticu, pričom nemusí byť ani štvorcová. Pre RAS procedúru je nevyhnutné poznať nielen aktuálne súčty po riadkoch a stĺpcoch, ale aj celú SAM maticu z predchádzajúceho obdobia. Označme

$$a_{ij}^1 = \frac{N_{ij}^1}{N_{\cdot j}^1}, \quad a_{ij}^0 = \frac{N_{ij}^0}{N_{\cdot j}^0}, \quad \mathbf{A}^0 = (a_{ij}^0), \quad \mathbf{A}^1 = (a_{ij}^1)$$

$$\mathbf{U}^1 = (N_{1\cdot}^1 \dots N_{K\cdot}^1)' \quad \mathbf{V}^1 = (N_{\cdot 1}^1 \dots N_{\cdot K}^1) = \mathbf{U}^{1'}$$

Aby sme z dostupnej informácie o SAM z predchádzajúceho obdobia dostali odhad novej SAM matice, postupujeme nasledovne:

- i) Za prvotný odhad \mathbf{A}^1 vezmeme \mathbf{A}^0 . Najskôr overme, či sú splnené riadkové súčty, teda či $\mathbf{A}^0 \mathbf{V}^1 = \mathbf{U}^1$ je rovné \mathbf{U}^1 .
- ii) Ak táto rovnosť neplatí, zostrojíme maticu $\mathbf{R}^1 = \text{diag}\{r_1^1 \cdots r_K^1\}$, kde $r_i^1 = \frac{U^1(i)}{U^1(i)}$ a touto maticou vynásobíme SAM maticu z predchádzajúceho obdobia: $\mathbf{R}^1 \mathbf{A}^0$.
- iii) Teraz sú už riadkové súčty splnené, preto overíme ešte stĺpcové, teda či $\mathbf{R}^1 \mathbf{A}^0$ sumované po riadkoch dáva jednotkový vektor: $\mathbf{V}^1 = \mathbf{1}$. Zrejme to však platiť nebude, preto vynásobíme $\mathbf{R}^1 \mathbf{A}^0$ sprava maticou $\mathbf{S}^1 = \text{diag}\{s_1^1 \cdots s_K^1\}$, kde $s_i^1 = \frac{1}{V^1(i)}$, čím upravíme stĺpcové súčty, opäť však nebudú splnené riadkové.
- iv) Nový odhad matice \mathbf{A}^1 je $\mathbf{R}^1 \mathbf{A}^0 \mathbf{S}^1$.

Takýmito iteráciami postupne dostávame stále lepší odhad, až kým nedosiahneme požadovanú presnosť pre odchylku riadkových a stĺpcových súčtov odhadu od skutočných hodnôt, teda kým odchylka každého súčtu, či už po riadku alebo po stĺpci, bude menšia ako nami zvolené ϵ . Konvergencia postupu bola ukázaná [12]. Neskôr bolo ukázané v práci [1], že RAS procedúru dostávame ako výsledok

$$\min_{N_{ij}} \sum_{i=1}^K \sum_{j=1}^K N_{ij} \ln \left(\frac{N_{ij}}{N_{ij}^0} \right) \quad (4.1)$$

za podmienok

$$\sum_{i=1}^K N_{ij} = N_{.j} \quad (4.2)$$

$$\sum_{j=1}^K N_{ij} = N_{i.} \quad (4.3)$$

4.2 ME a GME formulácia

4.2.1 Golan, Judge a Miller ME

GJM formulujú úlohu odhadu SAM nasledovne - ME úloha, I. postup.

$$\max_{\mathbf{a}} H(\mathbf{a}) = - \sum_{i=1}^K \sum_{j=1}^K a_{ij} \ln a_{ij} \quad (4.4)$$

pri podmienkach

$$\sum_{j=1}^K a_{ij} N_{.j}^1 = N_{i.}^1 \quad \forall i$$

$$\sum_{i=1}^K a_{ij} = 1 \quad \forall j$$

Odhad prvkov SAM matice v prípade takejto formulácie je $\hat{N}_{ij}^1 = \hat{a}_{ij} N_{.j}^1$. Táto formulácia je, tak ako v prípade s kockou, urobená iba hrubou silou, pretože nevychádza z maximalizácie možných spôsobov realizácie. Autori ju zaviedli iba na základe nájdenia podobnosti s predchádzajúci úlohami, ktorá spočíva v tom, že pre neznáme parametre $a_{ij} = \frac{N_{ij}}{N_{.j}}$ má byť $\sum_i a_{ij} = 1$ a ďalej $a_{ij} \in \langle 0, 1 \rangle$. Toto je v súlade s ME formuláciou pre parametre s charakterom pravdepodobností. Úloha ME by však mohla byť formulovaná aj nasledovne. Označme

$$b_{ij}^1 = \frac{N_{ij}^1}{N_{.i}^1}, \quad b_{ij}^0 = \frac{N_{ij}^0}{N_{.i}^0}, \quad \mathbf{B}^0 = (b_{ij}^0), \quad \mathbf{B}^1 = (b_{ij}^1)$$

Keďže aj v tomto prípade je $\sum_j b_{ij} = 1$ a $b_{ij} \in \langle 0, 1 \rangle$, možno použiť princíp maximum entropy a formulácia úlohy, II. postup je potom

$$\max_{\mathbf{b}} H(\mathbf{b}) = - \sum_{i=1}^K \sum_{j=1}^K b_{ij} \ln b_{ij}$$

pri podmienkach

$$\sum_{i=1}^K b_{ij} N_{.i}^1 = N_{.j}^1 \quad \forall j$$

$$\sum_{j=1}^K b_{ij} = 1 \quad \forall i$$

Odhad prvkov SAM matice je $\hat{N}_{ij}^1 = \hat{b}_{ij} N_{.i}^1$. V prípade, že matica nie je vybilancovaná a súčty po riadkoch a po stĺpcoch sú teda rôzne, dávajú tieto rozdielne formulácie aj rôzne riešenia. Toto síce nie je prípad SAM matice, ale takýto problém má tiež svoje odôvodnenie, ide o nasledovný volebný problém (VP). Predpokladajme, že sa konajú voľby, pričom sú dvaja kandidáti a voličov delíme na dve skupiny (napríklad na mužov a ženy). Ak sú známe výsledky volieb - koľko dostal ktorý kandidát hlasov, a známa je aj informácia o počte voličov v obidvoch skupinách, môže nás zaujímať koľko voličov z danej skupiny volilo prvého a koľko druhého kandidáta. Chceme teda odhadnúť jednotlivé prvky, pričom poznáme súčty po riadkoch a stĺpcoch matice. Nasledujúce tabuľky ukazujú zadanie ilustratívneho príkladu VP a riešenie ME na základe oboch možných postupov.

		4
		12
10	6	

Tabuľka 7: Zadanie

2.12279	1.87721	4
7.87721	4.12279	12
10	6	

Tabuľka 8: Riešenie I.

2.20501	1.79499	4
7.79499	4.20501	12
10	6	

Tabuľka 9: Riešenie II.

Pre úlohu so SAM maticou, GJM pri hľadaní odhadu prvkov tabuľky používajú ME v tvare I. postupu, pre zadanie VP v tvare II. postupu. V obidvoch prípadoch by však mohli byť použité obidve formulácie, keďže sú v súlade s princípom ME pre problémy, kde odhadované parametre majú charakter pravdepodobností, teda maximalizujú mieru entropie pre tieto parametre. Riešenia sú rôzne, aby sme sa vedeli rozhodnúť pre jedno z nich, postupujme ako pri odvodení princípu ME, vychádzajme teda z maximalizáciu počtu spôsobov, ktorými mohla byť daná matica realizovaná.

4.2.2 Alternatívna formulácia ME

ME formuláciu dostávame maximalizovaním

$$W = \prod_{j=1}^K W_j = \prod_{j=1}^K \frac{N_{\cdot j}^1!}{\prod_{i=1}^K N_{ij}^1!}$$

pričom W vyjadruje počet spôsobov, ktorými mohlo dôjsť k realizácii danej SAM matice.

$$\ln W = \sum_{j=1}^K \ln W_j = \sum_{j=1}^K \ln N_{\cdot j}^1! - \sum_{i=1}^K \ln N_{ij}^1!$$

Použitím Stirlingovej aproximácie dostávame

$$\ln W_j \approx N_{\cdot j}^1 \ln N_{\cdot j}^1 - N_{\cdot j}^1 - \sum_{i=1}^K (N_{ij}^1 \ln N_{ij}^1 - N_{ij}^1)$$

Ak označíme $\beta_{ij} = \frac{N_{ij}^1}{N^1}$, potom

$$\ln W_j \approx N_{\cdot j}^1 \ln N_{\cdot j}^1 - N_{\cdot j}^1 - \sum_{i=1}^K \beta_{ij} N^1 \ln \beta_{ij} N^1 + \sum_{i=1}^K N_{ij}^1$$

$$\ln W_j \approx N_{\cdot j}^1 \ln N_{\cdot j}^1 - \sum_{i=1}^K \beta_{ij} N^1 \ln N^1 - \sum_{i=1}^K \beta_{ij} N^1 \ln \beta_{ij}$$

alebo

$$\ln W \approx \sum_{j=1}^K N_{\cdot j}^1 \ln N_{\cdot j}^1 - N^1 \ln N^1 - N^1 \sum_{i=1}^K \sum_{j=1}^K \beta_{ij} \ln \beta_{ij}$$

keďže $\sum_j N_{\cdot j}^1 \ln N_{\cdot j}^1$ a $N^1 \ln N^1$ sú známe konštanty, dostávame úlohu

$$\max_{\boldsymbol{\beta}} H(\boldsymbol{\beta}) = - \sum_{i=1}^K \sum_{j=1}^K \beta_{ij} \ln \beta_{ij} \quad (4.5)$$

pri podmienkach

$$\sum_{j=1}^K N \beta_{ij} = N_i^1 \quad \forall i$$

$$\sum_{i=1}^K N \beta_{ij} = N_{\cdot j}^1 \quad \forall j$$

$$\sum_{i=1}^K \sum_{j=1}^K \beta_{ij} = 1$$

Odhad prvkov SAM matice resp. prvkov tabuľky VP úlohy je $\hat{N}_{ij}^1 = \hat{\beta}_{ij} N^1$. Táto formulácia je odlišná od oboch predchádzajúcich, rovnako ako je odlišný odhad pre ilustratívny problém s voľbami.

2.5	1.5	4
7.5	4.5	12
10	6	

Tabuľka 10: Riešenie, III. postup

Dôvod prečo nie je formulácia uvedená autormi úplne korektná, ak predpokladáme, že princíp entropie vychádza z maximalizácie W , je nasledovný. Ak maximalizujeme

$$W = \prod_{j=1}^K W_j = \prod_{j=1}^K \frac{N_{.j}^1!}{\prod_{i=1}^K N_{ij}^1!}$$

a namiesto $\beta_{ij} = \frac{N_{ij}^1}{N^1}$ zavedieme $a_{ij} = \frac{N_{ij}^1}{N_{.j}^1}$, dostávame po úpravách podobných ako vyššie

$$\begin{aligned} \ln W_j &\approx N_{.j}^1 \ln N_{.j}^1 - N_{.j}^1 - \sum_{i=1}^K a_{ij} N_{.j}^1 \ln a_{ij} N_{.j}^1 + \sum_{i=1}^K N_{ij}^1 \\ \ln W_j &\approx N_{.j}^1 \ln N_{.j}^1 - \sum_{i=1}^K a_{ij} N_{.j}^1 \ln N_{.j}^1 - \sum_{i=1}^K a_{ij} N_{.j}^1 \ln a_{ij} \end{aligned}$$

alebo

$$\ln W \approx - \sum_{i=1}^K \sum_{j=1}^K N_{.j}^1 a_{ij} \ln a_{ij}$$

čiže po predelení N dostávame účelovú úlohu

$$\max_{\mathbf{a}} - \sum_{j=1}^K w_j \sum_{i=1}^K a_{ij} \ln a_{ij}$$

kde $w_j = \frac{N_{.j}^1}{N}$ sú váhy pre jednotlivé entropie $\sum_{i=1}^K a_{ij} \ln a_{ij}$. Táto účelová funkcia na rozdiel od účelovej funkcie s rovnakými váhami pre každú entropiu, ktorú používajú GJM, dáva riešenie rovnaké ako alternatívny III. postup.

4.2.3 Golan, Judge, Miller GME

Aby rozšírili ME na GME navrhli GJM reparametrizáciu $a_{ij} = z_{ij1} p_{ij1} + \dots + z_{ijM} p_{ijM}$, pričom z_{ijm} sú prvky M rozmerného supportu pre a_{ij} a platí $z_{ij1} = 0$ a $z_{ijM} = 1 \forall i, j$. GME formulácia úlohy, ako ju predkladajú v [2], je potom

$$\max_{\mathbf{p}} H(\mathbf{p}) - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M p_{ijm} \ln p_{ijm} \quad (4.6)$$

pri podmienkach

$$\begin{aligned} \sum_{j=1}^K N_{.j} \sum_{m=1}^M z_{ijm} p_{ijm} &= N_i^1 \quad \forall i \\ \sum_{i=1}^K \sum_{m=1}^M z_{ijm} p_{ijm} &= 1 \quad \forall j \\ \sum_{m=1}^M p_{ijm} &= 1 \quad \forall i, j \end{aligned}$$

Opakuje sa situácia ako v príklade s kockou a riešenie takto formulovanej úlohy je opäť závislé od počtu prvkov supportu a ich konkrétnej voľby. V prípade, že support z_{ijm} je zvolený tak, že $z_{ij1}=0$, $z_{ijM}=1$, nie je riešenie vždy ekvivalentné s riešením ME v zmysle $\hat{p}_{ij} = \sum_m z_{ijm} \hat{v}_{ijm}$, kde \hat{p}_{ij} je ME riešenie a \hat{v}_{ijm} je GME riešenie. Keďže v prípade ME sa ukázalo, že formulácia daná autormi nie je úplne v súlade s maximalizáciou počtu spôsobov, ktorými mohla byť SAM matica realizovaná, je pravdepodobné, že aj pre GME je možné formulovať úlohu inak.

4.2.4 Alternatívna formulácia GME

Aby sme ME formuláciu rozšírili na GME, hľadáme neznáme β_{ij} v tvare $\beta_{ij} = z_{ij1} p_{ij1} + \dots + z_{ijM} p_{ijM}$, kde $(z_{ij1}, \dots, z_{ijM})'$ je M-prvkový support pre p_{ij} a $(p_{ij1}, \dots, p_{ijM})'$ sú váhy pre prvky supportu, pričom $\sum_{m=1}^M p_{ijm} = 1 \quad \forall i, j$. V súlade s touto reparametrizáciou úlohy môžeme prvky SAM napísať v tvare $N_{ij}^1 = \sum_{m=1}^M z_{ijm} p_{ijm} N^1 = \sum_{m=1}^M N_{ijm}^1$. Počet spôsobov, ktorými sa mohla daná SAM matica realizovať, má byť maximálny, preto opäť maximalizujeme W

$$W = \frac{N^1!}{\prod_i \prod_j \prod_m N_{ijm}^1!}$$

Logaritmicou transformáciou a použitím Stirlingovej aproximácie dostávame

$$\begin{aligned} \ln W &\approx N^1 \ln N^1 - N^1 - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M (N_{ijm}^1 \ln N_{ijm}^1 - N_{ijm}^1) \\ \ln W &\approx N^1 \ln N^1 - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M N_{ijm}^1 \ln z_{ijm} p_{ijm} N^1 \\ \ln W &\approx N^1 \ln N^1 - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M N_{ijm}^1 \ln z_{ijm} p_{ijm} - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M N_{ijm}^1 \ln N^1 \\ \ln W &\approx -N^1 \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} p_{ijm} \ln z_{ijm} p_{ijm} \end{aligned}$$

a úloha GME má potom tvar

$$\max_{\mathbf{p}} H(\mathbf{p}) = - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} p_{ijm} \ln z_{ijm} p_{ijm}$$

pri podmienkach

$$\begin{aligned} \sum_{j=1}^K N \sum_{m=1}^M z_{ijm} p_{ijm} &= N_i^1 \quad \forall i \\ \sum_{i=1}^K N \sum_{m=1}^M z_{ijm} p_{ijm} &= N_{\cdot j}^1 \quad \forall j \\ \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} p_{ijm} &= 1 \\ \sum_{m=1}^M p_{ijm} &= 1 \quad \forall i, j \end{aligned}$$

Čiže účelová funkcia je na rozdiel od formulácie GJM mierou entropie pre parametre $\tilde{p}_{ijm} = z_{ijm} p_{ijm}$. V prípade, že support z_{ijm} je zvolený tak, že $z_{ij1}=0$, $z_{ijM}=1$ a prvky supportu sú od seba ekvidistantne vzdialené, takáto reparametrizácia dáva riešenie totožné s ME riešením (opäť v zmysle $\hat{\beta}_{ij} = \sum_m z_{ijm} \hat{p}_{ijm}$). V účelovej funkcii vystupuje okrem hľadaných p_{ijm} aj z_{ijm} , preto aj riešenie a duálna úloha budú mať mierne odlišný tvar od úloh, v ktorých v účelovej funkcii vystupuje iba p_{ijm} . Lagrangeova funkcia pre alternatívnu formuláciu GME úlohy je

$$L(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \eta, \boldsymbol{\gamma}) = - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} p_{ijm} \ln p_{ijm} z_{ijm} \quad (4.7)$$

$$+ \sum_{i=1}^K \lambda_i \left(N_i - \sum_{j=1}^K N \sum_{m=1}^M z_{ijm} p_{ijm} \right) + \sum_{j=1}^K \mu_j \left(N_{\cdot j} - \sum_{i=1}^K N \sum_{m=1}^M z_{ijm} p_{ijm} \right) \quad (4.8)$$

$$+ \eta \left(1 - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} p_{ijm} \right) + \sum_{i=1}^K \sum_{j=1}^K \gamma_{ij} \left(1 - \sum_{m=1}^M p_{ijm} \right) \quad (4.9)$$

podmienky prvého rádu sú

$$\frac{\partial L}{\partial p_{ijm}} = -z_{ijm} \ln z_{ijm} \hat{p}_{ijm} - z_{ijm} - \hat{\lambda}_i z_{ijm} N - \hat{\mu}_j z_{ijm} N - \hat{\eta} z_{ijm} - \hat{\gamma}_{ij} = 0 \quad (4.10)$$

$i, j = 1, 2, \dots, K \quad m = 1, 2, \dots, M$

$$\frac{\partial L}{\partial \lambda_i} = N_i^1 - \sum_{j=1}^K N \sum_{m=1}^M z_{ijm} \hat{p}_{ijm} = 0 \quad \forall i \quad (4.11)$$

$$\frac{\partial L}{\partial \mu_j} = N_{\cdot j}^1 - \sum_{i=1}^K N \sum_{m=1}^M z_{ijm} \hat{p}_{ijm} = 0 \quad \forall j \quad (4.12)$$

$$\frac{\partial L}{\partial \eta} = 1 - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} \hat{p}_{ijm} = 0 \quad (4.13)$$

$$\frac{\partial L}{\partial \gamma_{ij}} = 1 - \sum_{m=1}^M \hat{p}_{ijm} = 0 \quad \forall i, j \quad (4.14)$$

Keďže nie je vylúčené že $z_{ijm} = 0$, vyjadríme p_{ijm} zvlášť pre túto situáciu a zvlášť pre ostatné prípady. Najskôr nájdime \hat{p}_{ijm} pre tie i, j, m , že $z_{ijm} \neq 0$. Z podmienky (4.10) po predelení z_{ijm} úpravami dostávame

$$\hat{p}_{ijm} = \exp \left(-\ln z_{ijm} - 1 - \lambda_i N - \mu_j N - \eta - \frac{\gamma_{ij}}{z_{ijm}} \right)$$

Ak použijeme podmienku (4.14) a uvedomíme si, že

$$\sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} p_{ijm} = \sum_{i=1}^K \sum_{j=1}^K \sum_{\substack{m: \\ z_{ijm} \neq 0}} z_{ijm} p_{ijm}$$

môžeme \hat{p}_{ijm} vyjadriť ako

$$\hat{p}_{ijm} = \frac{1}{\Omega(\hat{\lambda}, \hat{\mu}, \hat{\gamma})} \exp \left(-\ln z_{ijm} - \lambda_i N - \mu_j N - \frac{\gamma_{ij}}{z_{ijm}} \right) \quad (4.15)$$

kde

$$\Omega(\hat{\lambda}, \hat{\mu}, \hat{\gamma}) = \sum_{i=1}^K \sum_{j=1}^K \sum_{\substack{m: \\ z_{ijm} \neq 0}} \exp \left(-\hat{\lambda}_i N - \hat{\mu}_j N - \frac{\hat{\gamma}_{ij}}{z_{ijm}} \right)$$

Vieme, že $z_{ijm} = 0$ pre najviac jedno m , pre každú pevne zvolenú dvojicu i, j . Preto pre tie i, j, m , pre ktoré je $z_{ijm} = 0$, vyjadríme \hat{p}_{ijm} ako

$$\hat{p}_{ijm} = 1 - \sum_{\substack{m: \\ z_{ijm} \neq 0}} \hat{p}_{ijm} \quad (4.16)$$

Transformujme teraz primárnu úlohu s ohraničeniami na duálnu úlohu bez ohraničení tým, že dosadíme p_{ijm} v tvare (4.15) a (4.16) do (4.7). Dostávame

$$\begin{aligned} L(\mathbf{p}(\lambda, \mu, \gamma), \lambda, \mu, \eta, \gamma) &= \\ &= - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} p_{ijm}(\lambda, \mu, \gamma) \ln(z_{ijm} p_{ijm}(\lambda, \mu, \gamma)) \\ &\quad + \sum_{i=1}^K \lambda_i \left[N_i - \sum_{j=1}^K N \sum_{m=1}^M z_{ijm} p_{ijm}(\lambda, \mu, \gamma) \right] + \sum_{j=1}^K \mu_j \left[N_{\cdot j} - \sum_{i=1}^K N \sum_{m=1}^M z_{ijm} p_{ijm}(\lambda, \mu, \gamma) \right] \\ &\quad + \eta \left[1 - \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} p_{ijm}(\lambda, \mu, \gamma) \right] + \sum_{i=1}^K \sum_{j=1}^K \gamma_{ij} \left[1 - \sum_{m=1}^M p_{ijm}(\lambda, \mu, \gamma) \right] \\ &= - \sum_{i=1}^K \sum_{j=1}^K \sum_{\substack{m: \\ z_{ijm} \neq 0}} z_{ijm} p_{ijm}(\lambda, \mu, \gamma) \left[-\lambda_i N - \mu_j N - \frac{\gamma_{ij}}{z_{ijm}} - \ln(\Omega(\lambda, \mu, \gamma)) \right] \\ &\quad + \sum_{i=1}^K \lambda_i \left[N_i - \sum_{j=1}^K N \sum_{\substack{m: \\ z_{ijm} \neq 0}} z_{ijm} p_{ijm}(\lambda, \mu, \gamma) \right] + \sum_{j=1}^K \mu_j \left[N_{\cdot j} - \sum_{i=1}^K N \sum_{\substack{m: \\ z_{ijm} \neq 0}} z_{ijm} p_{ijm}(\lambda, \mu, \gamma) \right] \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^K \sum_{j=1}^K \gamma_{ij} \left[1 - \sum_{\substack{m: \\ z_{ijm} \neq 0}} p_{ijm}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}) - \sum_{\substack{m: \\ z_{ijm} = 0}} p_{ijm}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}) \right] \\
= & \sum_{i=1}^K \lambda_i N_i + \sum_{j=1}^K \mu_j N_{.j} + \sum_{i=1}^K \sum_{j=1}^K \gamma_{ij} + \ln(\Omega(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma})) - \sum_{i=1}^K \sum_{j=1}^K \gamma_{ij} \sum_{\substack{m: \\ z_{ijm} = 0}} p_{ijm}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}) \\
& + \sum_{i=1}^K \sum_{j=1}^K \sum_{\substack{m: \\ z_{ijm} \neq 0}} z_{ijm} p_{ijm}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}) \left[\lambda_i N + \mu_j N + \frac{\gamma_{ij}}{z_{ijm}} \right] - \sum_{i=1}^K \sum_{j=1}^K \gamma_{ij} \sum_{\substack{m: \\ z_{ijm} \neq 0}} p_{ijm}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}) \\
& - \sum_{i=1}^K \lambda_i \left[\sum_{j=1}^K N \sum_{\substack{m: \\ z_{ijm} \neq 0}} z_{ijm} p_{ijm}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}) \right] - \sum_{j=1}^K \mu_j \left[\sum_{i=1}^K N \sum_{\substack{m: \\ z_{ijm} \neq 0}} z_{ijm} p_{ijm}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}) \right] \\
= & \sum_{i=1}^K \lambda_i N_i + \sum_{j=1}^K \mu_j N_{.j} + \sum_{i=1}^K \sum_{j=1}^K \gamma_{ij} + \ln(\Omega(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma})) - \sum_{i=1}^K \sum_{j=1}^K \gamma_{ij} \sum_{\substack{m: \\ z_{ijm} = 0}} p_{ijm}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma})
\end{aligned}$$

Z podmienky (4.10) dostávame, že pre i, j, m také, že $z_{ijm} = 0$ je aj $\gamma_{ij} = 0$. Preto

$$L(\mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}), \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\eta}, \boldsymbol{\gamma}) = \sum_{i=1}^K \lambda_i N_i + \sum_{j=1}^K \mu_j N_{.j} + \sum_{\substack{(i,j): \\ \forall m z_{ijm} \neq 0}} \gamma_{ij} + \ln(\Omega(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma})) \equiv M(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma})$$

Duálna úloha bez ohraničení má teda tvar

$$\min_{\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}} M(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}) = \sum_{i=1}^K \lambda_i N_i + \sum_{j=1}^K \mu_j N_{.j} + \sum_{\substack{(i,j): \\ \forall m z_{ijm} \neq 0}} \gamma_{ij} + \ln(\Omega(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}))$$

4.3 CE a GCE formulácia

Úlohu ME a GME možno rozšíriť na CE/GCE aby zahŕňali aj informáciu o SAM matici z predchádzajúceho obdobia. Rovnako možno zahrnúť aj apriórny odhad rozdelenia hlasov podľa kandidátov vo forme prvkov tabuľky N_{ij}^0 v úlohe VP. Toto rozšírenie je založené podľa GJM na zmene účelovej funkcie (4.4) alebo (4.6) na jeden z nasledujúcich tvarov

$$\begin{aligned}
& \min_{\mathbf{a}} \sum_{i=1}^K \sum_{j=1}^K a_{ij} \ln \left(\frac{a_{ij}}{a_{ij}^0} \right) \\
& \min_{\mathbf{P}} \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M p_{ijm} \ln \left(\frac{p_{ijm}}{q_{ijm}} \right)
\end{aligned}$$

v závislosti na tom, či ide o prechod z ME na CE alebo z GME na GCE, pričom ohraničenia ostanú rovnaké ako pre ME/GME. Ak však chceme korektné odvodiť tvar CE/GCE úlohy pre VP alebo pre odhad SAM matice, musíme vychádzať od začiatku a maximalizovať $\pi(\mathbf{n}|\mathbf{q})$ ako v časti 1.3. Uvažujme najskôr o CE. Aby sme modifikovali $\pi(\mathbf{n}|\mathbf{q})$ špecifikám

zadania VP a SAM úlohy, označme $\mathbf{n} = (N_{11}, N_{12}, \dots, N_{ij}, \dots, N_{KK})'$ $K^2 \times 1$ vektor prvkov SAM matice alebo hľadanej tabuľky pre VP, ďalej $\beta_{ij}^0 = \frac{N_{ij}^0}{N^0}$. Naším cieľom je nájsť

$$\arg \max_{\mathbf{n}} \pi(\mathbf{n} | \beta^0)$$

kde

$$\pi(\mathbf{n} | \beta^0) = \frac{N!}{\prod_{i=1}^K \prod_{j=1}^K} \prod_{i=1}^K \prod_{j=1}^K (\beta_{ij}^0)^{N_{ij}}$$

logaritmickej transformáciou, Stirlingovou aproximáciou a ďalšími úpravami podobnými ako v prípade ME pre alternatívny postup dostávame

$$\ln \pi(\mathbf{n} | \beta^0) \approx -N \sum_{i=1}^K \sum_{j=1}^K \beta_{ij} \ln \left(\frac{\beta_{ij}}{\beta_{ij}^0} \right)$$

Pre odvodenie GCE označme $\mathbf{n} = (N_{111}, N_{112}, \dots, N_{ijm}, \dots, N_{KKM})'$ a reparametrizujeme $\beta^0 = \sum_m z_{ijm} q_{ijm} = \sum_m \tilde{q}_{ijm}$, kde z_{ijm} je support pre p_{ijm} a q_{ijm} . Hľadáme

$$\arg \max_{\mathbf{n}} \pi(\mathbf{n} | \tilde{\mathbf{q}}) = \arg \max_{\mathbf{n}} \frac{N!}{\prod_{i=1}^K \prod_{j=1}^K \prod_{m=1}^M N_{ijm}!} \prod_{i=1}^K \prod_{j=1}^K \prod_{m=1}^M \tilde{q}_{ijm}^{N_{ijm}}$$

postupom ako vyššie dostávame

$$\ln \pi(\mathbf{n} | \tilde{\mathbf{q}}) \approx -N \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} p_{ijm} \ln \left(\frac{z_{ijm} p_{ijm}}{\tilde{q}_{ijm}} \right)$$

Na základe týchto výsledkov by korektná formulácia účelovej funkcie pre CE mala mať tvar

$$\min_{\beta} KL(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^K \sum_{j=1}^K \beta_{ij} \ln \left(\frac{\beta_{ij}}{\beta_{ij}^0} \right)$$

respektíve pre GCE

$$\min_{\mathbf{p}} KL(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^K \sum_{j=1}^K \sum_{m=1}^M z_{ijm} p_{ijm} \ln \left(\frac{p_{ijm}}{q_{ijm}} \right)$$

Ohraničenia, pri ktorých hľadáme extrém, sú rovnaké ako v prípade alternatívneho postupu pre ME/GME. V GCE úlohe máme účelovú funkciu ako mieru relatívnej entropie pre parametre $\tilde{p}_{ijm} = z_{ijm} p_{ijm}$. Zostavením Lagrangeovej funkcie a úpravami ako v prípade GME dostávame riešenie

$$\hat{p}_{ijm} = \frac{q_{ijm}}{\Omega(\hat{\lambda}, \hat{\mu}, \hat{\gamma})} \exp \left(\lambda_i N + \mu_j N + \frac{\gamma_{ij}}{z_{ijm}} \right) \quad (4.17)$$

kde

$$\Omega(\hat{\lambda}, \hat{\mu}, \hat{\gamma}) = \sum_{i=1}^K \sum_{j=1}^K \sum_{\substack{m: \\ z_{ijm} \neq 0}} z_{ijm} q_{ijm} \exp \left(\hat{\lambda}_i N + \hat{\mu}_j N + \frac{\hat{\gamma}_{ij}}{z_{ijm}} \right)$$

pre tie i, j, m že $z_{ijm} \neq 0$ a

$$\hat{p}_{ijm} = 1 - \sum_{\substack{m: \\ z_{ijm} \neq 0}} \hat{p}_{ijm} \quad (4.18)$$

pre tie i, j, m kde $z_{ijm} = 0$. Dosadením p_{ijm} v tvare (4.17) a (4.18) a úpravami dostávame duálnu úlohu GCE bez ohraničení

$$\max_{\lambda, \mu, \gamma} M(\lambda, \mu, \gamma) = \sum_{i=1}^K \lambda_i N_{i.} + \sum_{j=1}^K \mu_j N_{.j} + \sum_{\substack{(i,j): \\ \forall m z_{ijm} \neq 0}} \gamma_{ij} - \ln(\Omega(\lambda, \mu, \gamma))$$

4.4 Porovnanie jednotlivých metód

Z formulácie úlohy RAS v tvare (4.1)-(4.3) možno jednoduchými úpravami dostať úlohu CE v alternatívnej formulácii. Tieto dva postupy sú preto ekvivalentné pre dostatočne malé ϵ zvolené ako presnosť pre RAS. V období vzniku RAS neboli k dispozícii rýchle počítače, preto sa úloha riešila iteračnou procedúrou a nie optimalizačnou úlohou (4.1)-(4.3), z ktorej iteračnú procedúru RAS dostávame. V ďalšom budeme používať RAS ako iteračnú procedúru.

Našou požiadavkou je, aby boli odhady čím najpresnejšie, teda aby čo najviac zodpovedali skutočným parametrom, budeme preto porovnávať odhady pomocou jednotlivých metód so skutočnými hodnotami. Zaujímať nás bude presnosť odhadu jednak pre N_{ij} , jednak pre $a_{ij}^1 = \frac{N_{ij}^1}{N_{i.}^1}$. Pre porovnanie budú použité tieto ukazovatele:

- i) korelačné koeficienty CORR 1 medzi a_{ij} a \hat{a}_{ij} a CORR 2 medzi N_{ij} a \hat{N}_{ij}
- ii) mean absolute deviation (priemerná absolútna odchylka) pre a_{ij} a N_{ij}

$$\text{MAD 1} = \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K |\hat{a}_{ij}^1 - a_{ij}^1|$$

$$\text{MAD 2} = \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K |\hat{N}_{ij}^1 - N_{ij}^1|$$

- iii) squared error measure (priemerná kvadratická chyba) pre a_{ij} a N_{ij}

$$\text{SEM 1} = \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K (\hat{a}_{ij}^1 - a_{ij}^1)^2$$

$$\text{SEM 2} = \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K (\hat{N}_{ij}^1 - N_{ij}^1)^2$$

iv) maximum proportionate error (maximálna proporčná chyba)

$$\text{MaxPE} = \max_{i,j} \frac{|\hat{a}_{ij}^1 - a_{ij}^1|}{a_{ij}^1}$$

v) mean absolute proportionate error (priemerná proporčná chyba)

$$\text{MAPE} = \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K \frac{|\hat{a}_{ij}^1 - a_{ij}^1|}{a_{ij}^1}$$

vi) goodness of fit pre a_{ij} a pre N_{ij}

$$\text{GOF 1} = \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K \left| \frac{(\hat{a}_{ij}^1 - a_{ij}^1)^2}{\hat{a}_{ij}^1} \right|$$

$$\text{GOF 2} = \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K \left| \frac{(\hat{N}_{ij}^1 - N_{ij}^1)^2}{\hat{N}_{ij}^1} \right|$$

Presnosť jednotlivých odhadov nemožno porovnať len na základe jedného ukazovateľa, ako to urobili GJM v prácach [2] a [3], ktorí zvolili ako nástroj na porovnanie SEM. Je totiž možné, že niektorá metóda bude vykazovať lepšie výsledky podľa jedného ukazovateľa, avšak horšie podľa iného z ukazovateľov. Tento fakt zrejme z nasledovného príkladu: ak máme napríklad preložiť danými bodmi v rovine priamku, môžeme použiť rôzne metódy - metódu najmenších štvorcov, minimalizovanie priemernej absolútnej vzdialenosti, ... každá metóda teda minimalizuje inú mieru a je teda zrejme, že ak použijeme ako ukazovateľ SEM bude dávať lepšie výsledky metóda najmenších štvorcov, v prípade, že ukazovateľ bude MAD, bude lepšia metóda minimalizovania priemernej absolútnej vzdialenosti. . . Ak teda zvolíme ukazovateľov viac, každý poskytne čiastkový pohľad na presnosť metódy.

4.4.1 Monte Carlo simulácie

Pre porovnanie jednotlivých postupov boli vykonané Monte Carlo simulácie IO modelu: z rovnomerného rozdelenia na $\langle 0, 10 \rangle$ bola generovaná matica \mathbf{X} rozmerov 10×10 predstavujúca medzispotrebu pre rok 0 a vektor konečnej spotreby \mathbf{y} 10×1 , vektor pridanej hodnoty \mathbf{w} (vektor odmien primárnym faktorom) bol dopočítaný reziduálne, aby bola IO tabuľka vybilancovaná). Údaje pre rok 1 boli vytvorené vynásobením jednotlivých prvkov matice a vektora konečnej spotreby náhodným číslom z $N(1.3, \sigma^2)$ normálneho rozdelenia, kde $\sigma^2 \in \{0.05, 0.15, 0.25, \dots, 0.65\}$ a následným vybilancovaním pomocou vektora pridanej hodnoty. Pre každú σ^2 bolo vykonaných 5000 Monte Carlo simulácií. Úlohou je odhadnúť maticu medzispotreby a input koeficienty pre rok 1 na základe riadkových a stĺpcových súčtov matice medzispotreby z roka 1 a matice z roka 0. Odhady boli urobené pomocou RAS, ME, CE, GME, GCE a to jednak postupom pre maximalizovanie W ,

jednak postupom GJM. Pri voľbe supportu pre GME a GCE máme dve základné možnosti. Prvou je voľba supportu $\mathbf{z}=(0,1)'$, druhá vyplýva z nasledujúcej úvahy. Ak chceme dodržať riadkové a stĺpcové súčty, môže každý prvok N_{ij} nadobúdať iba niektoré hodnoty. Ak by danom v stĺpci boli všetky ostatné prvky nulové môže byť N_{ij}^1 maximálne $\min(N_{.j}^1, N_{i.}^1)$. Minimálnu hodnotu pre N_{ij}^1 dostávame ak sú všetky prvky v danom stĺpci maximálne. Vtedy je N_{ij}^1 rovné $N_{.j}^1 - (N^1 - N_{i.}^1)$. Táto hodnota môže byť záporná, preto support zvolíme pre GJM formuláciu pre parametre a_{ij} ako $(l_{ij}, u_{ij})'$, kde

$$l_{ij} = \max \left\{ 0, \frac{N_{.j}^1 - (N^1 - N_{i.}^1)}{N_{.j}^1} \right\} \quad u_{ij} = \min \left\{ 1, \frac{N_{i.}^1}{N_{.j}^1} \right\}$$

Pre ALT formuláciu pre hľadané parametre β_{ij} je support $(l_{ij}, u_{ij})'$, kde

$$l_{ij} = \max \left\{ 0, \frac{N_{.j}^1 - (N^1 - N_{i.}^1)}{N^1} \right\} \quad u_{ij} = \min \left\{ \frac{N_{.j}^1}{N^1}, \frac{N_{i.}^1}{N^1} \right\}$$

Pre GCE musíme tieto hodnoty ešte porovnať s a_{ij}^0 resp. β_{ij}^0 , support je potom

$$l_{ij} = \min \left\{ a_{ij}^0, \max \left\{ 0, \frac{N_{.j}^1 - (N^1 - N_{i.}^1)}{N_{.j}^1} \right\} \right\} \quad u_{ij} = \max \left\{ a_{ij}^0, \min \left\{ 1, \frac{N_{i.}^1}{N_{.j}^1} \right\} \right\}$$

pre GJM GCE a

$$l_{ij} = \min \left\{ \beta_{ij}^0, \max \left\{ 0, \frac{N_{.j}^1 - (N^1 - N_{i.}^1)}{N^1} \right\} \right\} \quad u_{ij} = \max \left\{ \beta_{ij}^0, \min \left\{ \frac{N_{.j}^1}{N^1}, \frac{N_{i.}^1}{N^1} \right\} \right\}$$

pre ALT GCE. V ďalšom budeme označovať voľbu $\mathbf{z} = (0, 1)'$ ako A a $\mathbf{z} = (l_{ij}, u_{ij})'$, ako B.

Nasledujúce tabuľky uvádzajú priemerné hodnoty jednotlivých ukazovateľov pre použité metódy.

Tabuľka 11: Výsledky Monte-Carlo simulácii ME

	GJM	0.05	0.15	0.25	0.35	0.45	0.55	0.65
MAD 1		0.03746	0.03754	0.03817	0.03946	0.04128	0.04335	0.04529
SEM 1		0.00204	0.00207	0.00219	0.00238	0.00265	0.00296	0.00328
MAD 2		2.85346	2.89463	2.98175	3.11984	3.29876	3.50537	3.71140
SEM 2		11.58654	12.09358	13.12221	14.67606	16.72436	19.19265	21.88200
MaxPE		481.08275	478.95504	479.09421	485.00206	515.35001	606.48269	958.85456
MAPE		6.36293	6.36363	6.41499	6.57580	7.12763	8.54938	12.66159
GOF 1		0.02498	0.02557	0.02703	0.02941	0.03269	0.03655	0.04043
GOF 2		1.87984	1.94697	2.08437	2.29543	2.57810	2.91596	3.26797
CORR 1		0.37606	0.37435	0.37456	0.37612	0.37793	0.37954	0.38050
CORR 2		0.41048	0.41311	0.41713	0.42162	0.42570	0.42923	0.43247
	ALT	0.05	0.15	0.25	0.35	0.45	0.55	0.65
MAD 1		0.03730	0.03738	0.03803	0.03935	0.04124	0.04333	0.04537
SEM 1		0.00202	0.00206	0.00217	0.00237	0.00264	0.00296	0.00328
MAD 2		2.84090	2.88243	2.97096	3.11187	3.29621	3.50461	3.72010
SEM 2		11.48351	11.99521	13.03199	14.59453	16.66507	19.13879	21.87917
MaxPE		483.80883	479.54426	478.30592	483.62278	514.30306	606.88514	960.64037
MAPE		6.38239	6.36201	6.39998	6.55503	7.11134	8.54909	12.68688
GOF 1		0.02477	0.02537	0.02685	0.02925	0.03258	0.03646	0.04044
GOF 2		1.86228	1.92995	2.06846	2.28078	2.56683	2.90563	3.26539
CORR 1		0.38417	0.38194	0.38098	0.38128	0.38259	0.38344	0.38443
CORR 2		0.41931	0.42103	0.42372	0.42683	0.42973	0.43219	0.43445

Tabuľka 12: Výsledky Monte-Carlo simulácii CE

	GJM	0.05	0.15	0.25	0.35	0.45	0.55	0.65
MAD 1		0.00222	0.00659	0.01088	0.01511	0.01928	0.02319	0.02666
SEM 1		0.00001	0.00009	0.00023	0.00045	0.00074	0.00106	0.00139
MAD 2		0.17212	0.51661	0.86175	1.20784	1.55271	1.88455	2.19058
SEM 2		0.05836	0.52570	1.46273	2.87341	4.73918	6.93182	9.28402
MaxPE		0.10356	0.37651	0.86563	2.94487	18.12551	61.55876	158.69725
MAPE		0.02835	0.08695	0.15244	0.24520	0.50138	1.12260	2.35175
GOF 1		0.00009	0.00083	0.00231	0.00455	0.00754	0.01102	0.01457
GOF 2		0.00722	0.06524	0.18261	0.36200	0.60325	0.88690	1.18415
CORR 1		0.99793	0.98195	0.95275	0.91467	0.87266	0.83267	0.79905
CORR 2		0.99803	0.98289	0.95563	0.92076	0.88313	0.84798	0.81864
	ALT	0.05	0.15	0.25	0.35	0.45	0.55	0.65
MAD 1		0.00223	0.00660	0.01088	0.01509	0.01924	0.02313	0.02657
SEM 1		0.00001	0.00009	0.00023	0.00045	0.00073	0.00106	0.00138
MAD 2		0.17229	0.51693	0.86173	1.20676	1.54972	1.87924	2.18310
SEM 2		0.05841	0.52583	1.46127	2.86546	4.71683	6.88916	9.21964
MaxPE		0.10394	0.37698	0.86494	2.93082	17.97589	60.88918	158.60517
MAPE		0.02842	0.08715	0.15268	0.24524	0.49966	1.11462	2.34795
GOF 1		0.00009	0.00083	0.00231	0.00454	0.00753	0.01098	0.01451
GOF 2		0.00723	0.06530	0.18257	0.36137	0.60118	0.88276	1.17759
CORR 1		0.99793	0.98196	0.95282	0.91497	0.87338	0.83389	0.80070
CORR 2		0.99802	0.98289	0.95571	0.92108	0.88388	0.84916	0.82022

Tabuľka 13: Výsledky Monte-Carlo simulácii GME GJM

	GJM A	0.05	0.15	0.25	0.35	0.45	0.55	0.65
MAD 1		0.03746	0.03754	0.03817	0.03945	0.04132	0.04339	0.04541
SEM 1		0.00204	0.00207	0.00219	0.00238	0.00265	0.00297	0.00329
MAD 2		2.85339	2.89457	2.98168	3.11977	3.30203	3.50883	3.72287
SEM 2		11.58619	12.09319	13.12172	14.67537	16.73949	19.20976	21.94744
MaxPE		481.10390	478.96080	479.09611	485.00908	515.45161	606.80188	962.90788
MAPE		6.36314	6.36369	6.41501	6.57587	7.13066	8.55500	12.71759
GOF 1		0.02498	0.02557	0.02703	0.02941	0.03272	0.03658	0.04054
GOF 2		1.87988	1.94701	2.08440	2.29545	2.58048	2.91863	3.27769
CORR 1		0.37608	0.37438	0.37458	0.37616	0.37846	0.38008	0.38176
CORR 2		0.41051	0.41314	0.41716	0.42165	0.42574	0.42927	0.43251
	GJM B	0.05	0.15	0.25	0.35	0.45	0.55	0.65
MAD 1		0.03728	0.03737	0.03804	0.03937	0.04127	0.04336	0.04541
SEM 1		0.00202	0.00206	0.00218	0.00237	0.00265	0.00296	0.00329
MAD 2		2.83990	2.88158	2.97147	3.11353	3.29849	3.50734	3.72318
SEM 2		11.47988	11.99662	13.04166	14.61300	16.68900	19.16686	21.91146
MaxPE		486.37657	479.57675	475.45602	477.37040	509.83034	605.58684	959.93771
MAPE		6.40461	6.35946	6.36925	6.49114	7.06517	8.53660	12.68283
GOF 1		0.02480	0.02542	0.02691	0.02933	0.03268	0.03657	0.04056
GOF 2		1.86419	1.93249	2.07229	2.28621	2.57357	2.91345	3.27437
CORR 1		0.38420	0.38161	0.38000	0.37974	0.38078	0.38155	0.38251
CORR 2		0.41931	0.42103	0.42372	0.42683	0.42973	0.43219	0.43445

Tabuľka 14: Výsledky Monte-Carlo simulácii GCE GJM

	GJM A	0.05	0.15	0.25	0.35	0.45	0.55	0.65
MAD 1		0.00223	0.00661	0.01090	0.01514	0.01932	0.02325	0.02673
SEM 1		0.00001	0.00009	0.00023	0.00045	0.00074	0.00107	0.00140
MAD 2		0.17253	0.51783	0.86378	1.21068	1.55638	1.88908	2.19596
SEM 2		0.05860	0.52787	1.46875	2.88515	4.75846	6.96011	9.32263
MaxPE		0.10417	0.37812	0.86830	2.95002	18.15581	61.71934	158.42408
MAPE		0.02847	0.08730	0.15304	0.24609	0.50283	1.12566	2.35104
GOF 1		0.00009	0.00084	0.00232	0.00457	0.00758	0.01107	0.01464
GOF 2		0.00726	0.06556	0.18351	0.36378	0.60621	0.89130	1.19015
CORR 1		0.99792	0.98188	0.95254	0.91430	0.87210	0.83193	0.79812
CORR 2		0.99790	0.98180	0.95281	0.91581	0.87605	0.83919	0.80862
	GJM B	0.05	0.15	0.25	0.35	0.45	0.55	0.65
MAD 1		0.00223	0.00661	0.01090	0.01513	0.01930	0.02321	0.02668
SEM 1		0.00001	0.00009	0.00023	0.00045	0.00074	0.00106	0.00139
MAD 2		0.17263	0.51802	0.86375	1.21001	1.55458	1.88588	2.19156
SEM 2		0.05863	0.52790	1.46766	2.88004	4.74482	6.93414	9.28387
MaxPE		0.10447	0.37873	0.86822	2.94070	18.11079	61.33296	158.42666
MAPE		0.02853	0.08747	0.15326	0.24623	0.50247	1.12126	2.34942
GOF 1		0.00009	0.00084	0.00232	0.00457	0.00758	0.01106	0.01463
GOF 2		0.00726	0.06561	0.18354	0.36355	0.60531	0.88935	1.18702
CORR 1		0.99792	0.98189	0.95260	0.91449	0.87252	0.83263	0.79907
CORR 2		0.99790	0.98180	0.95284	0.91596	0.87643	0.83984	0.80952

Pre alternatívny postup (v tabuľkách označený ako ALT), ktorý maximalizuje W , dáva GME rovnaký výsledok ako ME a GCE ako CE. Ani GME ani GCE nezávisia pri tomto postupe od toho, či je použitá A alebo B verzia. Celá informácia o probléme je teda obsiahnutá v ohraničeniach týkajúcich sa riadkových a stĺpcových súčtov. Pre $\epsilon = 0.0001$ dáva RAS takmer rovnaké výsledky ako ALT CE/GCE. ME/GME odhad je od skutočných hodnôt vzdialený viac ako ostatné, treba si však uvedomiť, že na rozdiel od ME/GME RAS procedúra používa aj informáciu o A^0 .

Vzájomné porovnanie GJM a ALT prístupu je možné z grafov uvedených v prílohe, ktoré zobrazujú rozdiel medzi GJM a ALT pre jednotlivé ukazovatele s rastúcou disperziou σ^2 . Ukazujú teda o koľko sú výsledky pre formuláciu GJM horšie ako ALT.

Z grafov je zrejmé, že s rastúcou σ^2 dáva GJM v prípade CE, GME a GCE stále horšie výsledky v porovnaní s ALT, a to pre všetky ukazovatele s výnimkou MaxPE a MAPE pre GME, kde priebeh nie je úplne jednoznačný. V prípade ME rozdiel a výhodnosť ALT postupne klesá k 0 pre všetky ukazovatele s výnimkou MaxPE a MAPE kde priebeh opäť nie je úplne jednoznačný. Čím je teda matica pre rok 1 odlišnejšia od matice z roka 0, tým je výhodnejší postup ALT.

4.4.2 Ilustratívna SAM matica

Pre RAS aj CE/GCE v prípade $N_{ij}^0=0$ platí, že aj $\hat{N}_{ij}^1=0$. Preto ak je v matici pre rok 0 niektorý stĺpec celý nulový a v matici pre rok 1 v tomto stĺpci existuje nenulový prvok obidve metódy zlyhávajú. Jediná procedúra, ktorú možno použiť, je ME/GME. Takto to je pri ilustratívnej SAM matici z časti 3.2. Táto popisuje jednoduchú ekonomiku s dvoma výrobnými sektormi, okrem ktorých obsahuje ešte domácnosti a vládu. Vláda nevyberá najskôr žiadne dane. Ak sa rozhodne zdaňovať výrobok X 10% DPH aplikovaním CGE modelu (model všeobecnej rovnováhy), dostávame novú SAM maticu

	X	Y	K	L	D	G	Total
X					103.774		103.774
Y					103.774		103.774
K	56.604	41.509					98.113
L	37.736	62.264					100
D			98.113	100		9.434	207.547
G	9.434						9.434
Total	103.774	103.774	98.113	100	207.478	9.434	

Keďže odhad pomocou ME/GME by priradil nenulové hodnoty všetkým prvkom matice teda aj tým, o ktorých vieme že sú nulové, musíme upraviť formuláciu týchto metód. Veľkou výhodou metód založených na maximalizácii entropie je možnosť zahrnúť ľubovoľnú dostupnú informáciu o probléme. Odhad pomocou ME/GME možno ľahko modifikovať tak, že pridáme ďalšie ohraničenia pre tie N_{ij}^1 , ktoré majú byť rovné nule. Všetky ukazovatele sú pre ALT postup maximalizujúci W lepšie ako pre postup navrhnutý GJM, aj keď rozdiely sú iba mierne.

	GME ALT	GME GJM
MAD 1	0.01058	0.01060
SEM 1	0.00101	0.00101
MAD 2	1.09817	1.10051
SEM 2	10.85305	10.89813
MaxPE	0.26190	0.26245
MAPE	0.02315	0.02320
GOF 1	0.00212	0.00213
GOF 2	0.21966	0.22058
CORR 1	0.99474	0.99472
CORR 2	0.99520	0.99518

4.4.3 IO SR '96-'98

Posledné input-output tabuľky pre SR, ktoré boli zverejnené, sú z rokov 1996 a 1998. Tieto sú v takzvanej use-make forme, je v nich teda zachytených 46 produkčných sektorov a 46 statkov, pričom každý sektor môže produkovať každý statok. Aby sme mohli tieto IO tabuľky použiť pre porovnanie jednotlivých metód odhadu, musíme ich transformovať na tvar sektor-sektor, ktorý bol opísaný v časti 3.1. Táto transformácia je založená na predpoklade tzv. pevných technologických koeficientov, jej popis je náročný na priestor a z hľadiska zámeru tejto práce nie je nevyhnutné ho uviesť. Možno ho nájsť v práci [10]. Matica medzispotreby a input koeficienty boli odhadnuté podobne ako pri Monte-Carlo simuláciách jednotlivými metódami, presnosť jednotlivých odhadov zachytávajú nasledujúce tabuľky.

Tabuľka 17: Výsledky Input Output, GJM a ALT '96-'98

	GJM	ME	CE	GME A	GME B	GCE A	GCE B
MAD 1		0.01223	0.00420	0.01223	0.01186	0.00418	0.00410
SEM 1		0.00114	0.00026	0.00114	0.00116	0.00026	0.00024
MAD 2		571.593	149.199	571.584	556.782	148.676	146.221
SEM 2		4 046 697	336 359	4 044 830	4 017 864	325 043	307 846
MaxPE		1 290.686	388.681	1 288.744	1 139.317	378.894	398.994
MAPE		13.83311	1.48036	13.83355	12.17510	1.46996	1.48006
GOF 1		0.09175	0.00520	0.09707	0.08609	0.00524	0.00497
GOF 2		7376.451	215.652	8039.617	2832.792	220.934	201.879
CORR 1		0.56143	0.91740	0.56148	0.55390	0.91827	0.92378
CORR 2		0.44618	0.96603	0.44651	0.44811	0.96717	0.96890
	ALT	ME	CE	GME A	GME B	GCE A	GCE B
MAD 1		0.01166	0.00407	0.01166	0.01166	0.00407	0.00407
SEM 1		0.00114	0.00024	0.00114	0.00114	0.00024	0.00024
MAD 2		552.221	141.041	552.221	552.221	141.041	141.041
SEM 2		3 988 866	281 747	3 988 866	3 988 866	281 747	281 747
MaxPE		1 562.758	419.585	1 562.758	1 562.758	419.585	419.585
MAPE		11.46274	1.42971	11.46274	11.46274	1.42971	1.42971
GOF 1		0.09298	0.00471	0.09298	0.09298	0.00471	0.00471
GOF 2		2734.410	185.144	2734.410	2734.410	185.144	185.144
CORR 1		0.56197	0.92336	0.56197	0.56197	0.92336	0.92336
CORR 2		0.45579	0.97157	0.45579	0.45579	0.97157	0.97157

Tabuľka 18: Rozdiel medzi GJM a ALT (GJM-ALT)

		ME	CE	GME A	GME B	GCE A	GCE B
MAD 1		0.000570	0.000130	0.000570	0.000200	0.000110	0.000030
SEM 1		-0.000001	0.000019	-0.000001	0.000016	0.000016	-0.000001
MAD 2		19.37220	8.15800	19.36290	4.56160	7.63470	5.17970
SEM 2		57 830.813	54 612.004	55 964.083	28 998.110	43 295.841	26 099.244
MaxPE		-272.072	-30.903	-274.013	-423.441	-40.691	-20.591
MAPE		2.370370	0.050650	2.370810	0.712360	0.040250	0.050350
GOF 1		-0.001230	0.000490	0.004090	-0.006890	0.000530	0.000260
GOF 2		4 642.041	30.508	5305.207	98.382	35.790	16.735
CORR 1		-0.000540	-0.005960	-0.000490	-0.008070	-0.005090	0.000420
CORR 2		-0.009610	-0.005540	-0.009280	-0.007680	-0.004400	-0.002670

Tabuľka 19: Výsledky Input Output, RAS a ALT CE '96-'98

	RAS	0.01	0.005	0.001	0.0005	0.0001	0.00005
MAD 1		0.00408	0.00407	0.00407	0.00407	0.00407	0.00407
SEM 1		0.00025	0.00025	0.00025	0.00024	0.00024	0.00024
MAD 2		141.085	141.040	141.038	141.040	141.041	141.041
SEM 2		283 330	282 549	281 870	281 815	281 759	281 754
MaxPE		401.977	409.658	417.781	418.560	419.397	419.478
MAPE		1.41961	1.42403	1.42868	1.42912	1.42960	1.42965
GOF 1		0.00472	0.00471	0.00471	0.00471	0.00471	0.00471
GOF 2		185.330	185.281	185.178	185.164	185.148	185.146
CORR 1		0.92219	0.92274	0.92326	0.92330	0.92335	0.92336
CORR 2		0.97141	0.97149	0.97156	0.97157	0.97157	0.97157
	ALT	CE	CE	CE	CE	CE	CE
MAD 1		0.00407	0.00407	0.00407	0.00407	0.00407	0.00407
SEM 1		0.00024	0.00024	0.00024	0.00024	0.00024	0.00024
MAD 2		141.041	141.041	141.041	141.041	141.041	141.041
SEM 2		281 747	281 747	281 747	281 747	281 747	281 747
MaxPE		419.585	419.585	419.585	419.585	419.585	419.585
MAPE		1.42971	1.42971	1.42971	1.42971	1.42971	1.42971
GOF 1		0.00471	0.00471	0.00471	0.00471	0.00471	0.00471
GOF 2		185.144	185.144	185.144	185.144	185.144	185.144
CORR 1		0.92336	0.92336	0.92336	0.92336	0.92336	0.92336
CORR 2		0.97157	0.97157	0.97157	0.97157	0.97157	0.97157

Tabuľka 20: Rozdiel medzi RAS a ALT (RAS-ALT)

		0.01	0.005	0.001	0.0005	0.0001	0.00005
MAD 1		0.00001	0.00000	0.00000	0.00000	0.00000	0.00000
SEM 1		0.00001	0.00001	0.00001	0.00000	0.00000	0.00000
MAD 2		0.04380	-0.00110	-0.00340	-0.00190	-0.00030	-0.00020
SEM 2		1 583.308	802.308	123.308	68.308	12.308	7.308
MaxPE		-17.608	-9.927	-1.804	-1.024	-0.188	-0.107
MAPE		-0.01010	-0.00568	-0.00103	-0.00059	-0.00011	-0.00006
GOF 1		0.00001	0.00000	0.00000	0.00000	0.00000	0.00000
GOF 2		0.18550	0.13640	0.03370	0.01990	0.00380	0.00220
CORR 1		-0.00117	-0.00062	-0.00010	-0.00006	-0.00001	0.00000
CORR 2		-0.00016	-0.00008	-0.00001	0.00000	0.00000	0.00000

Hodnoty všetkých ukazovateľov okrem MaxPE sú pre alternatívne formulácie až na pár výnimiek lepšie ako pre formulácie GJM a to pre všetky tvary úlohy, či už ME, CE, GME A, GME B, GCE A alebo GCE B. Priemerne iba dva ukazovatele z desiatich sú lepšie pre GJM ako pre ALT. Porovnanie RAS s ALT CE ukazuje, že s klesajúcou povolenou odchylkou ϵ pre splnenie riadkových a stĺpcových súčtov sa odhady podľa očakávania približujú, pričom pre SEM 2 a GOF 2 je RAS mierne horšie ako ALT CE a pre MaxPE a MAPE naopak.

4.4.4 SAM SR '95-'96

Pri odhadoch SAM matice, založených na maximalizovaní entropie, sa stretávame pri reálnych dátach s nasledovným problémom. Až doteraz sme pri odhadovaní vychádzali z predpokladu, že prvky SAM matice sú nezáporné, čo nám umožnilo po zavedení $a_{ij} = N_{ij}/N_{.j}$ alebo $\beta = N_{ij}/N$ definovať mieru entropie. Ak by však bol niektorý prvok SAM matice záporný, nie je možné zaviesť tieto nové parametre a definovať entropiu.

Ako východisko sa ponúka reparametrizácia a prechod na GME a GCE, čo bol pôvodne dôvod vzniku týchto formulácií [2]. V tvare, ako ho navrhujú GJM, majú však tieto formulácie ten nedostatok, že nevychádzajú z maximalizácie počtu možných realizácií. Treba si uvedomiť, že ani použitie alternatívnej formulácie nie je korektné, keďže pre $N_{ij} < 0$ sú $N_{ijm} < 0$ a pri odvádzaní formulácie sme vychádzali z maximalizácie výrazu, v ktorom vystupovali faktoriály N_{ijm} , a tieto nie sú definované pre záporné čísla. Navyše pre GME v prípade že N_{ijm} máme $z_{ijm} < 0$ a účelová funkcia nie je definovaná. Otázne ostáva, ako vôbec formulovať počet spôsobov realizácie matice, ak je niektorý jej prvok záporný.

Problémy sú v takýchto prípadoch aj s RAS procedúrou, ako ukázali autori v článku [5]. Vychádzajúc faktu že RAS dáva ekvivaletný výsledok ako optimalizačná úloha (4.1)-(4.3), navrhli modifikáciu tejto úlohy a následne aj RAS tak, že účelovú funkciu (4.1) nahradili funkciou

$$\min_{N_{ij}} \sum_{i=1}^K \sum_{j=1}^K |N_{ij}| \ln \left(\frac{N_{ij}}{N_{ij}^0} \right)$$

Vychádzajúc z blízkeho vzťahu medzi RAS a alternatívnou formuláciou CE a GCE, ponúka sa táto zmena účelovej funkcie ako jedna z možností ako sa vysporiadať so zápornými prvkami.

Ak dokážeme určiť interval, z ktorého je daný záporný prvok matice, je ďalšou možnosťou nasledujúci postup. Ak by napríklad v nasledovnom zadaní

N_{11}	N_{12}	4
N_{21}	N_{22}	12
4	12	

bol prvok N_{12} záporný, pričom by sme vedeli, že nie je menší ako -1, pripočítame k N_{12} 1. Týmto sa súčasne zväčší príslušný riadkový aj stĺpcový súčet, aby boli stĺpcové a riadkové

súčty totožné pripočítame 1 aj k prvku N_{21} . Teraz sú už všetky prvky matice nezáporne, stĺpcové a riadkové súčty sa rovnajú a nové zadanie je najšť prvky matice

N'_{11}	N'_{12}	5
N'_{21}	N'_{22}	13
5	13	

Po odhadnutí prvkov tejto matice odpočítame 1 od prvkov N'_{12} a N'_{21} , čím dostávame odhad pre pôvodnú maticu.

Vychádzajme z predpokladu, že poznáme SAM maticu pre SR z roku 1995 a riadkové/stĺpcové súčty z roku 1996. SAM matica z roku 1995 má celkovo tri prvky menšie ako nula. Dva však stoja v stĺpci respektíve riadku osamostatnene a príslušný stĺpcový/riadkový súčet je pre rok 1996 kladný. Jediný záporný prvok teda môže byť tok kapitálu zo zahraničia (prvok [9,10]). Majúc na pamäti, že tento postup takisto nie je úplne korektný, môžeme modifikovať GCE formuláciu vyššie uvedeným spôsobom. Ako dolné ohraničenie zvolíme $-N_{10}^1$. Pripočítaním hodnoty N_{10}^1 k daným záporným prvkom a prvkom ležiacim symetricky podľa diagonály, dostávame všetky prvky kladné, pričom riadkové a stĺpcové prvky sú zhodné. Formuláciu ďalej doplníme o podmienky pre prvky, ktoré boli pôvodne nulové a po posune majú mať hodnotu N_{10}^1 .

Pre porovnanie boli urobené odhady týmito metódami: ME GJM bez modifikácie, ME ALT bez modifikácie, GCE GJM zo supportom $\mathbf{z}=(-1,1)'$ pre prvky, ktoré boli v 1995 záporné, GCE ALT modifikované, RAS 0.0001 bez modifikácie a RAS 0.0001 modifikované.

Tabuľka 21: Výsledky, SAM 1995-1996

	ME GJM	ME ALT	GCE GJM	GCE ALT	RAS	RAS mod.
MAD 1	0.03434	0.03257	0.01022	0.00560	0.01230	0.00794
SEM 1	0.00612	0.00542	0.00122	0.00027	0.00179	0.00059
MAD 2	13.55740	11.94563	3.67301	1.73926	4.29872	2.45966
SEM 2	1004.69700	679.16150	176.82050	26.60198	252.00540	54.43888
MaxPE	232.15189	183.70924	1.80344	1.41495	2.37951	1.08982
MAPE	4.85646	4.07080	0.09311	0.07032	0.11605	0.07260
GOF 1	0.03263	0.03178	0.02439	0.00136	0.01090	0.00235
GOF 2	12.44291	10.40385	10.02250	0.43658	4.22182	0.75317
CORR 1	0.93901	0.94617	0.98847	0.99740	0.98331	0.99433
CORR 2	0.98572	0.99018	0.99748	0.99962	0.99641	0.99922

Takmer všetky ukazovatele sú najlepšie pre modifikovanú úlohu ALT GCE. Oba korelačné koeficienty sú veľmi blízke jednej, SEM 1 a SEM 2 sú polovičné v porovnaní s modifikovanou RAS. Ostatné ukazovatele s výnimkou MaxPE sú nezanedbatelne lepšie ako pre ostatné metódy. Druhou najlepšou metódou je modifikované RAS, tretie v poradí je GCE GJM. ME ALT a ME GJM značne zaostávajú, pričom hlavne SEM 2 a MaxPE sú pre ME GJM vysoké.

Záver

Diplomová práca je venovaná problematike použitia princípu maximum entropy v ekonometrii.

Cieľom tejto práce bolo nájsť uplatnenie princípu entropie v ekonometrii pre odhad IO tabuliek a SAM matice na základe neúplnej informácie. Ako sa však ukázalo s formuláciou ako je prezentovaná v prácach [2] a [3] sú spojené určité nezrovnalosti. Vzhľadom na existenciu týchto nezrovnalostí je opodstatnenosť použitia princípov entropie v tvare ako ho predkladajú autormi spochybnená. Pre odstránenie nedostatkov bola navrhnutá alternatívna formulácia.

Výhodou metód, ktoré sú založené na maximalizácii entropie, je možnosť zahrnúť ľubovoľnú dostupnú informáciu o probléme, formuláciu je možné jednoducho rozšíriť pridaním ďalších ohraničení v optimalizačnej úlohe. V konečnom dôsledku je jedinou zmenou pridaním ďalších lagrangeových multiplikátorov v duálnej úlohe, charakter úlohy sa teda nemení.

Použitie GME a GCE pre stochastické modely nebola bližšie skúmaná, autori [2] dokázali, že odhady použitím týchto metód sú konzistentné, rovnako preukázali výhody použitia GME a GCE v prípade heteroskedasticity, autokorelácie a multikolinearity. Chýba však určité prepojenie s východiskami teórie maximálnej entropie, kedy vychádzame z maximalizácie počtu možných spôsobov realizácie experimentu. Ukázalo sa tiež, že zovšeobecnenie z ME/CE na GME/GCE, prevedené jednoduchou zámenou účelovej funkcie, nie je úplne korektné.

Cieľom ďalšieho skúmania ostáva predovšetkým korektná formulácia úlohy GME a GCE pre prípady SAM matice so zápornými prvkami, ktorú sa nepodarilo nájsť v tejto práci. Nájdienie tejto formulácie je nevyhnutné pre riešenie úloh s reálnymi dátami, keďže práve pri reálnych SAM maticiach je výskyt záporných prvkov bežný.

Literatúra

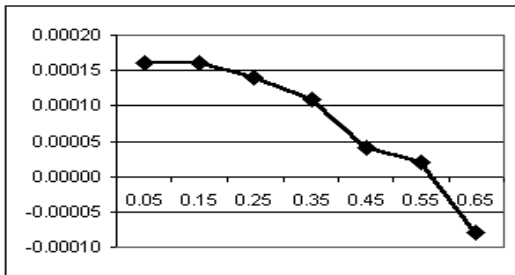
- [1] Bacharach M. (1965): Biproportional matrices and input-output change *PhD. Thesis* Department of Economics, Cambridge University
- [2] Golan, A., Judge, G. Miller, D. (1996): *Maximum Entropy Econometrics. Robust Estimation with Limited Data*, New York: John Wiley and Sons
- [3] Golan, A., Judge, G., Robinson, S. (1994): Recovering Information from Incomplete or Partial Multisectoral Data, *The Review of Economics and Statistics*, **76**, 541-549.
- [4] Grendar, M. Jr., Grendar, M. (2001): What is the question that MaxEnt answers? A probabilistic interpretation, *Maximum Entropy and Bayesian methods*, A. Mohammad-Djafari ed. AIP
- [5] Gunluk-Senesen, G., Bates, J. M. (1994): Some Experiments with Methods of Adjusting Unbalanced Data Matrices, *Journal of the Royal Statistical Society*, A **151**, 473-490
- [6] Hajnovičová, V. (1998): Formovanie dôchodkových tokov v období transformácie ekonomiky Slovenskej republiky, *Ekonomický časopis* **46**, 241-263
- [7] Hajnovičová, V., Lapišáková, J. (1998): Analýza štrukturálnych zmien v ekonomike Slovenska v období 1991-96, *Pracovné materiály* Prognostický ústav SAV Bratislava
- [8] Jaynes, E. T. (1963): Information Theory And Statistical Mechanics. In K. W. Ford (Ed.) *Statistical Physics* pp. 181-218, W.A. Benjamin, New York
- [9] Kullback, S. (1959): *Information Theory and Statistics* New York: John Wiley and Sons
- [10] van der Ploeg, F. (1963): *Mathematical Methods In Economics London School of Economics and Political Science* New York: John Wiley and Sons
- [11] Shannon, C. (1948): A Mathematical Theory Of Communication, *Bell System Technical Journal*, **27**, 379-423
- [12] Stone R., Bates, J. M., Bacharach, M. (1963): *A Program for Growth*, vol. 3 *Input-Output Relationships 1954-1966* London: Chapman and Hall

Prílohy

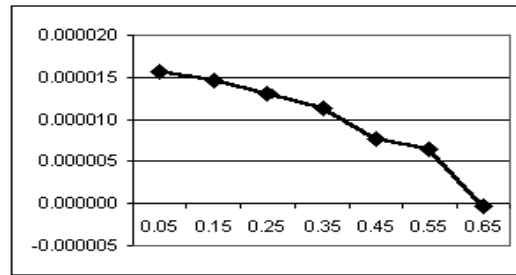
Diplomová práca obsahuje nasledovné prílohy na diskete:

readme.txt	podrobný popis obsahu diskety
adresár DATA	
IO-96-98.xls	input-output tabuľky SR v rokoch 1996 a 1998
SAM-95-96.xls	SAM matice SR v rokoch 1995 a 1996
MC IO.xls	výsledky Monte-Carlo simulácií IO modelu
adresár GAMS	funkcie použité pre jednotlivé úlohy

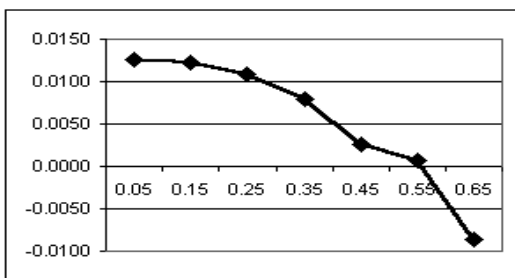
Nasledujúce grafy zobrazujú rozdiel hodnôt jednotlivých ukazovateľov pre GJM a ALT pre formulácie ME, CE, GME B, GCE B.



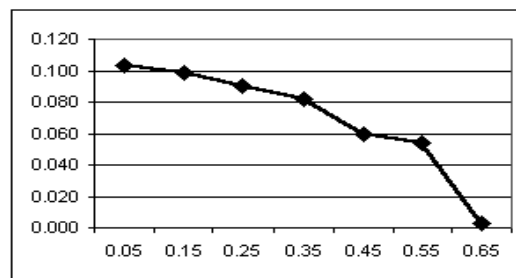
Obr. 1: MAD 1 ME



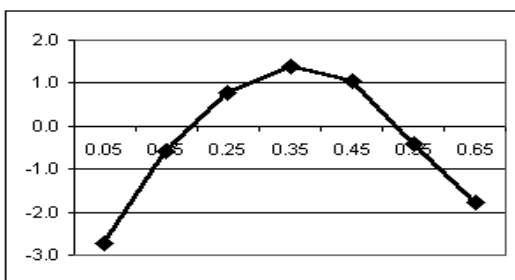
Obr. 2: SEM 1 ME



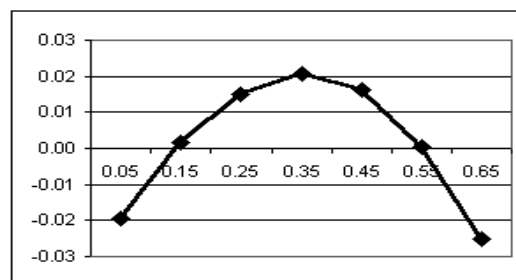
Obr. 3: MAD 2 ME



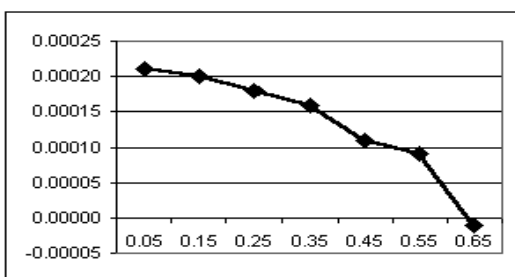
Obr. 4: SEM 2 ME



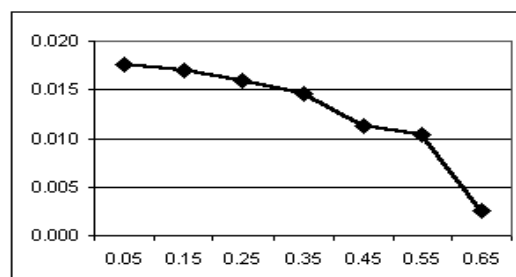
Obr. 5: MaxPE ME



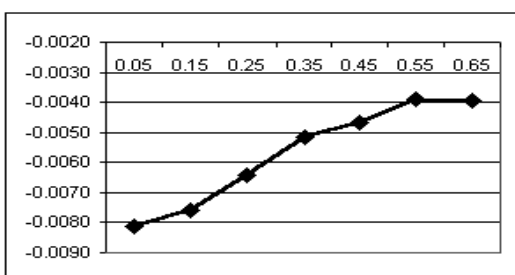
Obr. 6: MAPE ME



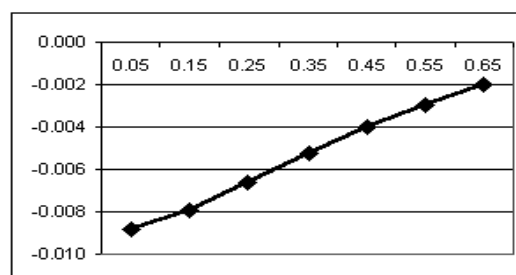
Obr. 7: GOF 1 ME



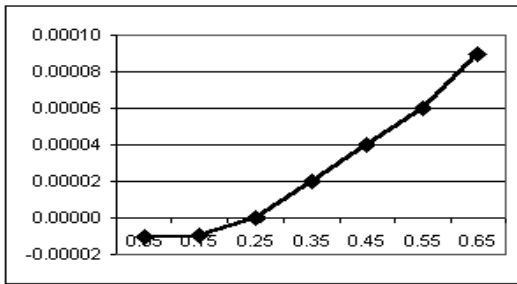
Obr. 8: GOF 2 ME



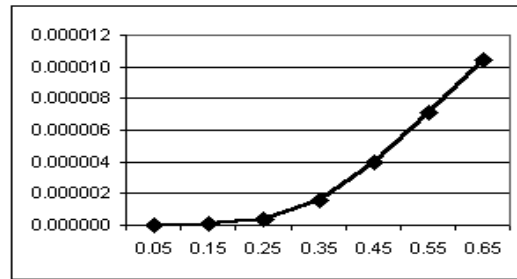
Obr. 9: CORR 1 ME



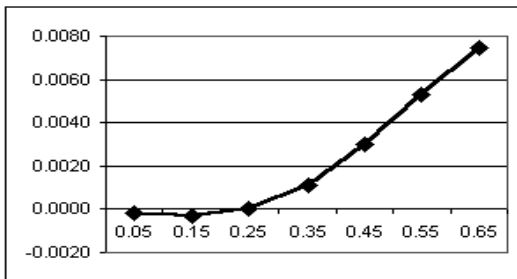
Obr. 10: CORR 2 ME



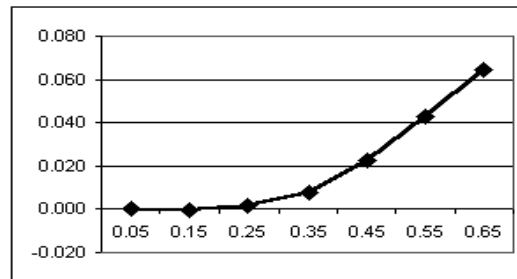
Obr. 11: MAD 1 CE



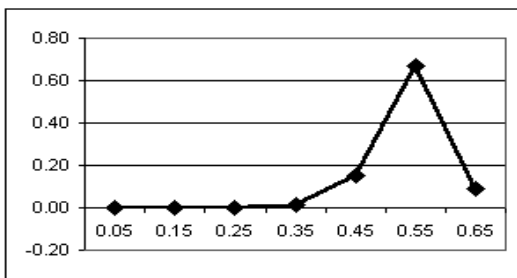
Obr. 12: SEM 1 CE



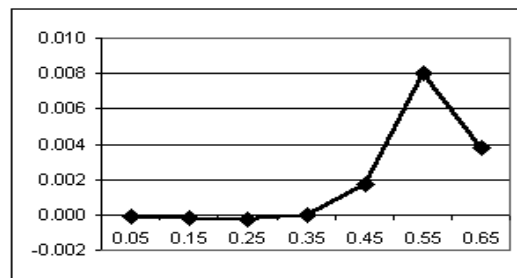
Obr. 13: MAD 2 CE



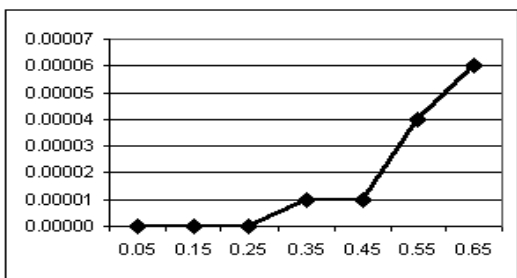
Obr. 14: SEM 2 CE



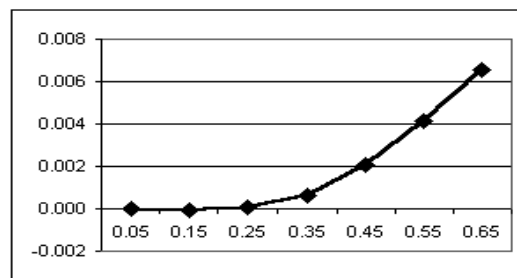
Obr. 15: MaxPE CE



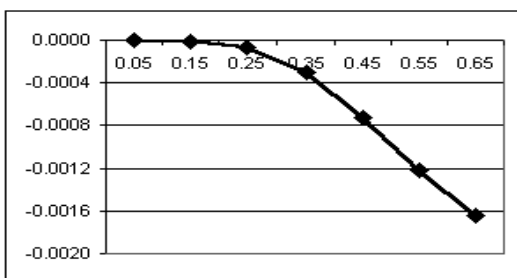
Obr. 16: MAPE CE



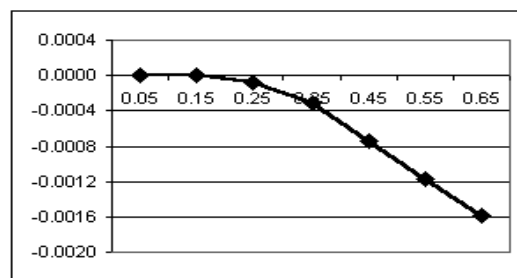
Obr. 17: GOF 1 CE



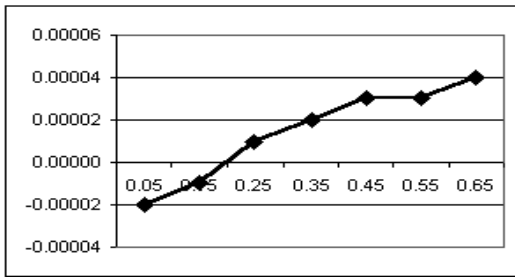
Obr. 18: GOF 2 CE



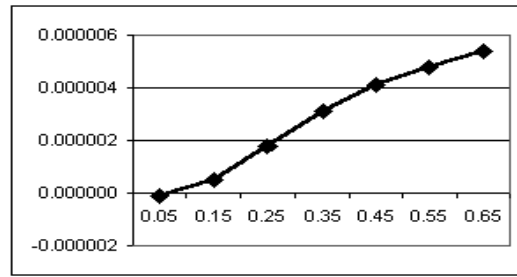
Obr. 19: CORR 1 CE



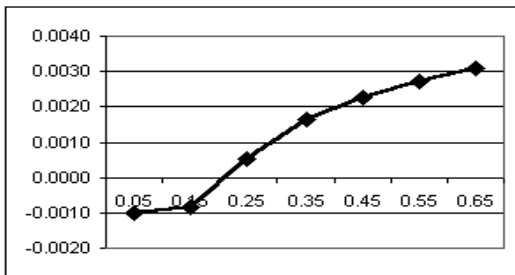
Obr. 20: CORR 2 CE



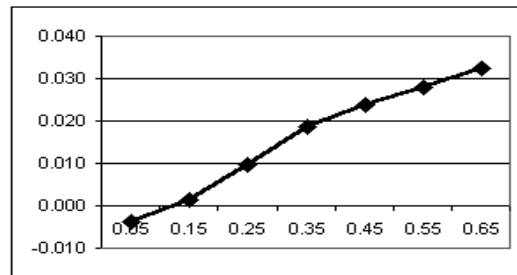
Obr. 21: MAD 1 GME



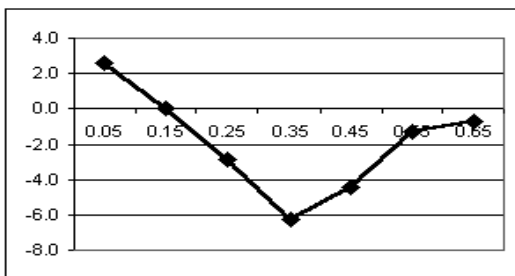
Obr. 22: SEM 1 GME



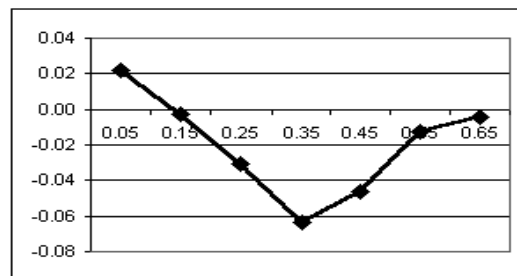
Obr. 23: MAD 2 GME



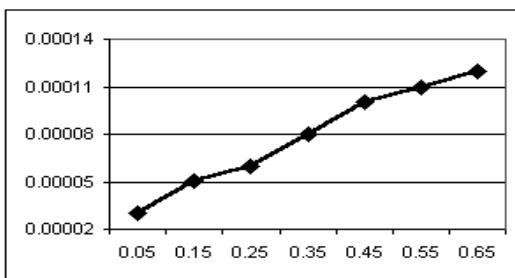
Obr. 24: SEM 2 GME



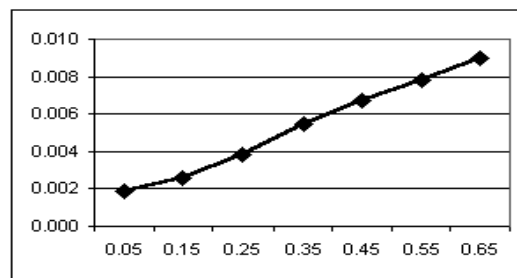
Obr. 25: MaxPE GME



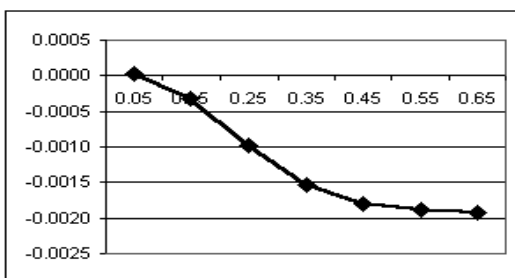
Obr. 26: MAPE GME



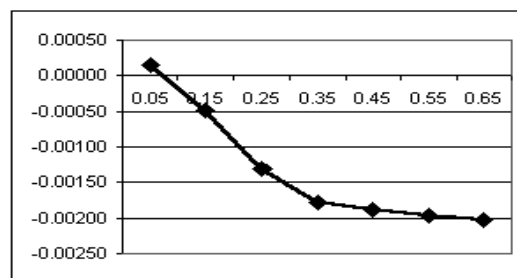
Obr. 27: GOF 1 GME



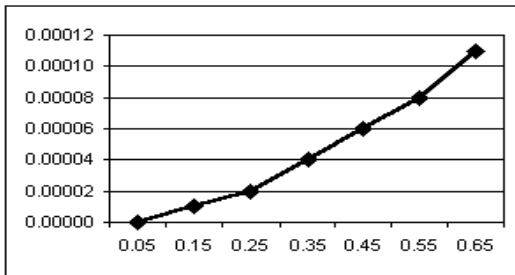
Obr. 28: GOF 2 GME



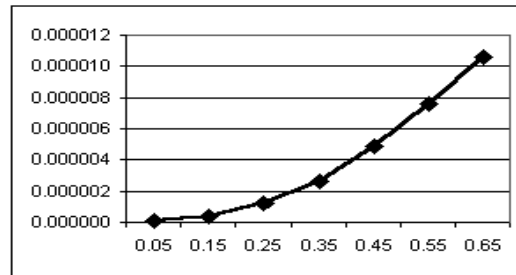
Obr. 29: CORR 1 GME



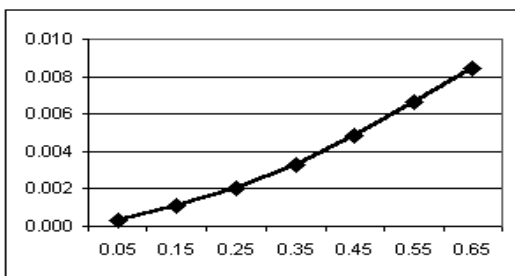
Obr. 30: CORR 2 GME



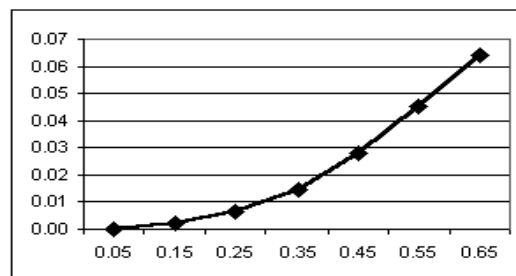
Obr. 31: MAD 1 GCE



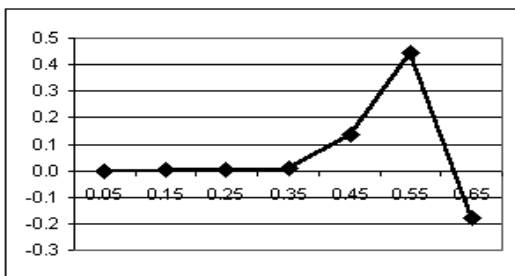
Obr. 32: SEM 1 GCE



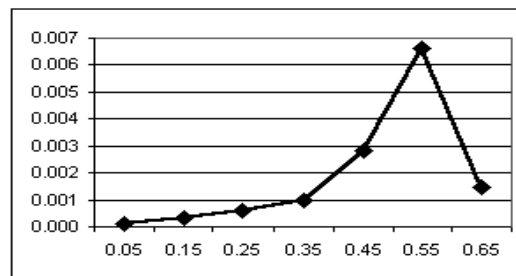
Obr. 33: MAD 2 GCE



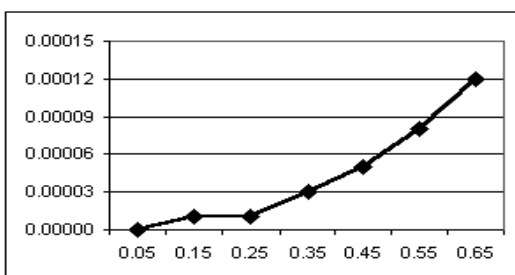
Obr. 34: SEM 2 GCE



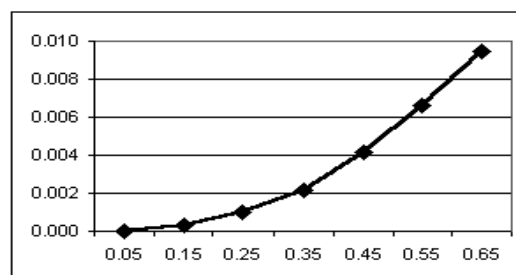
Obr. 35: MaxPE GCE



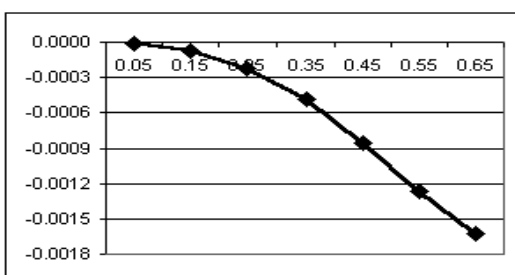
Obr. 36: MAPE GCE



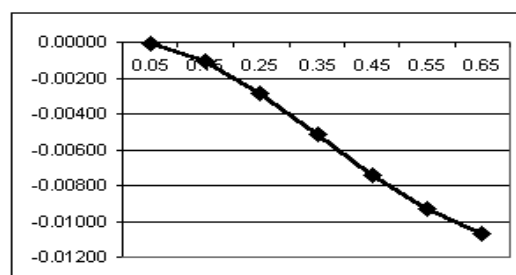
Obr. 37: GOF 1 GCE



Obr. 38: GOF 2 GCE



Obr. 39: CORR 1 GCE



Obr. 40: CORR 2 GCE