

## **Kvantitativní popis diverzifikace, Quantitative Description of Diversification**

Miloslav Malec

Lukáš Malec

Rostislav Tomeš

**Abstrakt:** V článku jsou popsány základní metody kvantitativního a grafického popisu diverzifikace. Jsou uvedeny konstrukce Lorenzovy křivky a odvozeny vzorce pro výpočet Giniho koeficientu v diskrétním a spojitém případě rozdělení standardu. Popsané metody diverzifikace jsou aplikovány zejména na soubory dat z oblasti cestovního ruchu.

**Klíčová slova:** distribuční funkce standardu, Giniho koeficient, Lorenzova křivka, Pareto rozdělení standardu, spojitě a diskrétní rozdělení

**Abstract:** In this article are presented the basics of quantitative and graphical description of diversification phenomena. A Lorenz curve constructions are introduced, together with Gini coefficient calculation formulas in discrete and continuous cases of distribution of standard. The methods of diversification are particularly applied to the tourism data.

**Keywords:** distribution function of standard, Gini coefficient, Lorenz curve, Pareto distribution of standard, continuous and discrete distribution

**JEL Classification:** C18, L83

### Úvod

Lorenzova křivka a Giniho koeficient (příp. jejich modifikace) patří k často používaným matematicko-statistickým metodám kvantitativního popisu ekonomických jevů. V článku, při použití elementárního matematického aparátu, jsou nejen odvozeny, resp. uvedeny, v ekonomické praxi často používané metody diverzifikace standardu v základním souboru (populaci), ale je ukázána i možnost jejich numerického zpracování užitím programu Matlab. Byl studován standardní soubor malých podniků a soubory dat příjzdového cestovního ruchu České republiky a vybraných evropských států.

Pojem diverzifikace (nerovnoměrnosti) se vyskytuje a zkoumá v široké škále přírodních a společenských vědních oborů. Kromě např. ekonometrie je tato charakteristika často užívána v sociologii ve vztahu s nerovnoměrným rozdělením příjmů, majetku, apod. Pro zajímavost uvedme: Jedno horní % světové populace vlastní 40% světového

bohatství, 10% nejbohatších vlastní 85% světového bohatství. Chudší polovina obyvatel vlastní pouze 1% bohatství, viz (Keller, 2010). Uvedená fakta lze názorně vyjádřit Lorenzovou křivkou. Její konstrukce (vzhledem k velkému rozsahu souboru) předpokládá provedení rozsáhlého výběrového šetření; data se pak zpracují metodami statistické indukce v kombinaci s modelováním standardu vhodnými distribučními funkcemi.

## Materiál a metody

### Diskrétní rozdělení standardu

Protože konkrétní datové soubory jsou konečné, jedná se o důležitý případ, zejména pokud rozsah souboru není příliš velký a data nemají náhodný charakter. Škálu veličin, vykazujících nerovnost v základním souboru nazveme souhrnně standardem.

Bud' dán uspořádaný soubor standardu  $\{s_i\}$ ,  $i = 1, 2, \dots, n$ ,  $s_i \leq s_{i+1}$ ,  $s_0 = 0$ . Lorenzovou křivkou (dále stručně L-křivkou) daného souboru  $\{s_i\}$  nazveme po částech lineár-

ní funkci (viz obr. 1), spojující body  $[0, 0]$ ,  $\left[ \frac{i}{n}, \frac{s_1 + s_2 + \dots + s_i}{\bar{s}} = L\left(\frac{i}{n}\right) \right]$ ,  $i = 1, 2, \dots, n$ , kde  $\bar{s} = \frac{1}{n} \sum_{i=1}^n s_i$ . Zřejmě číslo  $100i/n$  udává % podíl populace, číslo  $100L(i/n)$  udává odpovídající % podíl standardu. Čím je L-křivka blíže úhlopříčce, tím je rozdělení rovnoměrnější, čím je blíže odvěsnám, tím je standard více diferencován.

### Giniho koeficient

Giniho koeficient (stručně G-koeficient) vyjadřuje míru diverzifikace. Je dán hodnotou zlomku  $G = \frac{A}{A+B} = 2A = 1 - 2B$ ; čísla  $A$  a  $B$  označují velikosti ploch z obrázku 1.

Zřejmě  $G \in \langle 0, 1 \rangle$ ;  $G = 0$  odpovídá absolutní rovnosti,  $G = 1$  maximální nerovnosti.

Plocha obrázku  $B$  je tvořena lichoběžníky, tedy

$$G = \frac{2 \sum_{i=1}^n i s_i}{n \bar{s}} - \frac{n+1}{n}. \quad (1)$$

1) Tento předpoklad má zásadní význam pro konstrukci L-křivky (viz dále); populace musí být seřazena do neklesající posloupnosti hodnot standardu.

Vzorec (1) lze nazvat výpočtovým vzorcem G-koefficientu pro diskretní případ.

Jsou-li  $\left[\frac{i}{n}, L\left(\frac{i}{n}\right)\right]$ ,  $\left[\frac{i+j}{n}, L\left(\frac{i+j}{n}\right)\right]$  body L-křivky, potom číslo  $j/n$  udává relativní přírůstek populace a číslo  $\frac{s_{i+1} + s_{i+2} + \dots + s_{i+j}}{s}$  značí odpovídající relativní přírůstek standardu (tedy L-křivka jako důsledek sumace vykazuje aditivitu). Je-li rozsah souboru příliš velký, lze výpočet zjednodušit tak, že vhodně vybereme jen některé body L-křivky a plochu opět (nejsnadněji) spočteme lichoběžníkovou metodou. Označíme-li vybrané body L-křivky  $[x_j, L(x_j)]$ ,  $j = 1, 2, \dots, k$  a přitom  $x_0 = 0$  a  $x_k = 1$ ,  $L(x_0) = 0$ ,  $L(x_k) = 1$ , potom zřejmě

$$G = 1 - \sum_{j=1}^k (x_j - x_{j-1})(L(x_j) + L(x_{j-1})) \quad (2)$$

Uvedme příklad na použití formule (2). Jsou k dispozici počty zaměstnanců  $s_i$  v 72

podnicích,  $\sum s_i = 8850$ ; soubor je uspořádaný ( $s_i \leq s_{i+1}$ ). Provedeme rozdělení souboru zaměstnanců do 8 tříd šířky 200 (viz tabulka 1).

Výpočtem podle vzorce (2) dostaneme  $G \approx 0,165$ . Výsledek získaný užitím vzorce (1) se liší nepatrně.

## Spojité rozdělení standardu

L-křivku pro diskretní soubor dat (a tedy i vzorce pro G-koefficient) lze rovněž získat užitím empirické distribuční funkce. Buď uspořádaný (tj. soubor standardu). Potom funkce definovaná na je po částech konstantní, spojitá zprava v bodech a platí kde je počet hodnot pro které (tj. je největší index souboru, pro který Pokud hodnota je v souboru jedenkrát, má funkce v bodě skok pokud je hodnota obsažena -krát, má skok Více o popisu diverzifikace pro diskretní případ např. viz (Sen, 1973); pro spojitý případ viz (Atkinson, 1970). V případě, že je rozsah souboru příliš velký a je nutné realizovat náhodný výběr (jeho hodnoty už mají náhodný charakter), je účelné tento soubor dostatečně přesně aproximovat spojitou náhodnou veličinou s jistou distribuční funkcí. Při výpočtech se spojitými náhodnými veličinami lze navíc využít rozsáhlý aparát diferenciálního a integrálního počtu.

Tabulka 1. Datový soubor malých podniků

| Intervaly počtu zaměstnanců | Počet podniků | Kumul. relat. četnost podniků $x_j$ | Počet zaměstnanců | Kumul. relat. četnost zaměstnanců $L(x_j)$ |
|-----------------------------|---------------|-------------------------------------|-------------------|--|
| 550 – 750                   | 3             | 0,042                               | 1 830             | 0,021                                      |
| 751 – 950                   | 16            | 0,264                               | 13 824            | 0,176                                      |
| 951 – 1 150                 | 12            | 0,431                               | 12 432            | 0,316                                      |
| 1 151 – 1 350               | 20            | 0,708                               | 24 780            | 0,595                                      |
| 1 351 – 1 550               | 8             | 0,819                               | 11 432            | 0,724                                      |
| 1 551 – 1 750               | 4             | 0,875                               | 6 560             | 0,798                                      |
| 1 751 – 1 950               | 4             | 0,931                               | 7 432             | 0,881                                      |
| 1 951 – 2 150               | 5             | 1,000                               | 10 560            | 1,000                                      |
| $\Sigma$                    | 72            |                                     | 88 850            |  |

Zdroj: Vlastní šetření

Je-li k dispozici uspořádaný náhodný výběr  $\{s_j\}$  ze základního souboru s distribuční funkcí  $F$ , pak empirická distribuční funkce  $\bar{F}$  je aproximací distribuční funkce  $F$ .<sup>2</sup>

#### Pravděpodobnostní model diverzifikace

Buď dána nezáporná (tento předpoklad však není nutný) spojitá náhodná veličina  $s = X(\omega)$ ,  $\omega \in \Omega$  je prostor prvků základního souboru – populace, kde udává hodnotu standardu prvků  $\omega$ . Rozdělení pravděpodobnosti náhodné veličiny  $X$  buď popsáno distribuční funkcí  $F(s) = P\{\omega \in \Omega, X(\omega) \leq s\}$ ; číslo  $F(s)$  je tedy pravděpodobnost jevu, že náhodně vybraný prvek populace má standard  $\leq s$ . Připomeňme: Funkce  $F$  je spojitá, neklesající,  $F(0) = 0$ ,  $F(\infty) = 1$  a dále existuje nezáporná funkce  $f$  (hustota pravděpodobnosti) tak, že platí

$$F(s) = \int_0^s f(t) dt. \quad (3)$$

Níže uvedeme, že např. náhodné veličiny s log-normálním, resp. Paretovým rozdělením pravděpodobnosti jsou vzhledem k jejich šikmosti vhodným nástrojem k modelování rozdělení příjmů, majetku, apod.

2) Aproximaci distribučních funkcí lze popsat užitím pojmu konvergence náhodných veličin.

3) V teorii pravděpodobnosti je nutné pracovat s integrálem založeným na teorii míry. V článku budeme předpokládat, že v integrálech vystupují funkce po částech spojitě, resp. po částech hladké. Potom při výpočtech vystačíme s Riemannovým, resp. Newtonovým integrálem.

Dále (pro jednoduchost, vyhneme se tak definici zobecněné kvantilové funkce) budeme předpokládat, že funkce  $F(s)$  je rostoucí. Potom funkce  $p = F(s)$  má inverzní funkci  $s = F^{-1}(p)$ ,  $p \in \langle 0, 1 \rangle$ , která se nazývá kvantilová. Tato funkce je rovněž spojitá a rostoucí.

Její interpretaci lze vyjádřit:  $100p\%$  populace (ve smyslu pravděpodobnosti) má standard  $\leq s = F^{-1}(p)$ ; populace je zřejmě uspořádána s neklesajícím standardem. Je-li  $\Delta p = p_2 - p_1$ , potom na  $100\Delta p\%$  populace připadá  $100\Delta s = 100(F^{-1}(p_2) - F^{-1}(p_1))\%$  standardu.

L-křivka, přiřazená distribuční funkci standardu  $F$ , je množina bodů  $[p, L(p)]$   $p \in \langle 0, 1 \rangle$ , kde (viz obr. 2)

$$L(p) = \frac{1}{\mu} \int_0^p F^{-1}(r) dr, \mu = \int_0^{\infty} t f(t) dt \quad (4)$$

je střední hodnota rozdělení standardu. Zřejmě  $\mu > 0$ . Předpokládejme, že  $\mu < \infty$ .

Níže použijeme následující formule. Per partes dává  $\int_0^A t f(t) dt = A F(A) - \int_0^A F(t) dt$ , tj.

$$\frac{\int_0^A t f(t) dt}{F(A)} = \int_0^A \left( 1 - \frac{F(t)}{F(A)} \right) dt; \text{ v limitě pro } A \rightarrow \infty \text{ získáváme}$$

$$\mu = \int_0^{\infty} (1 - F(t)) dt = (\text{substitucí } t = F^{-1}(r) \text{ ve (3)}) = \int_0^1 F^{-1}(r) dr. \quad (5)$$

Interpretace analytického vyjádření L-křivky v (4): Integrál  $\int_0^p F^{-1}(r) dr$  – v Riemannově smyslu – je limita součtů  $\sum_i F^{-1}(r_i) \Delta r_i$ ; každý člen součtu je aproximací přírůstku standardu daný přírůstkem populace  $\Delta r_i$ . Tedy (s využitím vyjádření  $\mu$  ve (4)) číslo  $100L(p)$  udává relativní procenta standardu připadající na  $100p\%$  populace. Dále zřejmě  $100(L(p_2) - L(p_1))/\mu$  udává relativní procentní podíl standardu připadající na  $100(p_2 - p_1)$  procentní podíl populace. V důsledku aditivity integrálu je i L-křivka aditivní.

L-křivku lze dále popsat parametricky: 
$$L(s) = \frac{1}{\mu} \int_0^s t f(t) dt \quad \left. \begin{array}{l} p = F(s) \\ s \in \langle 0, \infty \rangle \end{array} \right\} \text{ Z první rovnice máme } s = F^{-1}(p) \text{ a substitucí } t = F^{-1}(r) \text{ do druhé získáme (4).}$$

Vlastnosti L-křivky:  $L(p)$  je rostoucí, spojitá funkce na  $\langle 0,1 \rangle$  s hodnotami tamtéž.

Z vyjádření  $L(p) = \frac{\int_0^p F^{-1}(r) dr}{\int_0^1 F^{-1}(r) dr}$  a z věty o střední hodnotě integrálu plyne nerovnost  $L(p) \leq p$ ,  $p \in \langle 0,1 \rangle$ .

$L'(p) = \frac{d}{dp} \left( \frac{1}{\mu} \int_0^p F^{-1}(r) dr \right) = \frac{F^{-1}(p)}{\mu} = \frac{s}{\mu}$ ; derivace  $L'(p)$  je tedy rostoucí funkce.

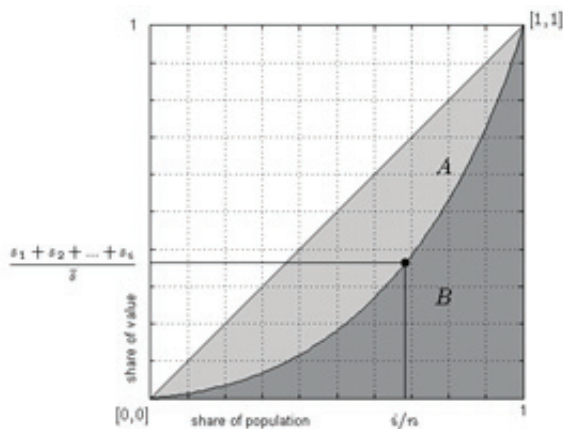
Pokud existuje  $L''(p)$  je funkce  $L(p)$  ryze konvexní.

L-křivka nezávisí na volbě měřítka; náhodné veličiny  $X$  a  $\mathcal{X}$ ,  $c > 0$  mají totožnou L-křivku.

G-koefficient lze opět vyjádřit užitím L-křivky (viz obr. 2)

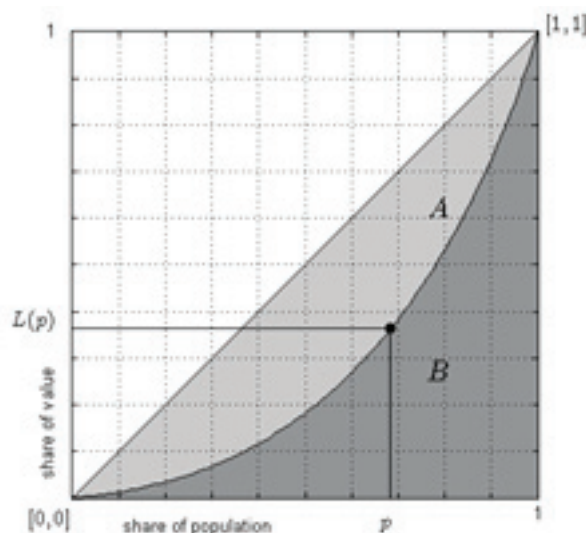
$$G = 1 - 2 \int_0^1 L(p) dp. \quad (6)$$

### Obrázek 1: Lorenzova křivka – diskrétní



Zdroj: vlastní zpracování

Obrázek 2: Lorenzova křivka – spojitý



Zdroj: vlastní zpracování

Existují jiné alternativní definice G-koefficientu, které uvádí např. Xu (2003). Jeho význam je stejný jako u diskrétního případu. Vzorec (4) udává analytické vyjádření L-křivky pomocí distribuční funkce rozdělení standardu. Totéž provedeme pro G-koefficient daný vzorcem (6). Užitím metody per partes v (6) postupně dostaneme

$$G = -1 + 2 \int_0^1 pL'(p) \mathbf{p} = (\text{subst. } p = F(s) = \frac{1}{\mu} \left[ \int_0^\infty 2sF(s)f(s) \mathbf{d} - \mu \right])$$

Nejdříve užitím per partes upravíme (vzhledem ke konvergenci nevlastních integrálů)

$$\int_0^A 2sF(s)f(s) \mathbf{d} = [F^2(s)s]_0^A - \int_0^A F^2(s) \mathbf{d}, \text{ tj.}$$

$$\frac{\int_0^A 2sF(s)f(s) \mathbf{d}}{F^2(A)} = \int_0^A \left( 1 - \frac{F^2(s)}{F^2(A)} \right) \mathbf{d}$$

a tedy v limitě pro  $A \rightarrow \infty$  máme  $\int_0^\infty 2sF(s)f(s) \mathbf{d} = \int_0^\infty (1 - F^2(s)) \mathbf{d}$ . Odtud podle (4)

$$G = \frac{1}{\mu} \left( \int_0^\infty (1 - F^2(s)) \mathbf{d} - \int_0^\infty (1 - F(s)) \mathbf{d} \right). \text{ A nakonec}$$

$$G = \frac{1}{\mu} \int_0^{\infty} F(s)(1-F(s)) \mathbf{d}s. \quad (7)$$

Vzorec (6) lze vyjádřit ve tvaru  $G = 1 - \frac{1}{\mu} \int_0^{\infty} (1-F(s))^2 \mathbf{d}s$ , protože pravá strana v (7) je rovna  $1 - \frac{1}{\mu} \left( \int_0^{\infty} (1-F(s)) \mathbf{d}s + \int_0^{\infty} F(s)(F(s)-1) \mathbf{d}s \right)$ .

Význam vzorce (7) je evidentní v případě, jsou-li k dispozici distribuční funkce standardu.

V modelech popisujících např. příjem v populaci, se často užívá log-normální rozdělení (má větší citlivost v nižších hodnotách příjmů) a Paretovo rozdělení (citlivější ve vyšších hodnotách příjmů). Uvedme, že pro log-normální rozdělení, je-li známa jeho směrodatná odchylka  $\sigma$ , platí  $G = 2\Phi(\sigma/\sqrt{2}) - 1$ , kde  $\Phi$  je distribuční funkce normovaného normálního rozdělení  $N(0,1)$ .

Aplikace log-normálního a Paretova rozdělení na příjmy v České republice jsou uvedeny např. v publikacích (Bílková, 2008; Bílková, 2009).

V tomto článku popíšeme Paretovo rozdělení pro jeho jednoduchost při výpočtech a jeho relativní důležitost v aplikacích. Paretovo rozdělení je dáno distribuční funkcí

$$F(s) = \begin{cases} 0 & \text{pro } s < s_0, \\ 1 - \left(\frac{s}{s_0}\right)^{-\alpha} & \text{pro } s \geq s_0, s_0 > 0, \alpha > 0. \end{cases} \quad (8)$$

Čísla  $s_0$  a  $\alpha$  jsou parametry rozdělení. Hustota pravděpodobnosti je tedy dána vzorcem  $f(s) = \alpha s_0^{-\alpha} s^{-\alpha-1}$  pro  $s \geq s_0$ , jinak nula.

Při popisu příjmů je parametr  $s_0$  dán hodnotou standardu, na který aplikujeme Paretovo rozdělení. Jeho hodnota se buď zvolí, nebo odhadne ze zadaných dat. Za  $s_0$  lze zvolit buď nejmenší hodnotu příjmů podléhajících zdanění, nebo hodnotu mediánu příjmů.

Označme  $x(s) = \left(\frac{s}{s_0}\right)^{-\alpha}$ ,  $s \geq s_0$ ; odtud  $\ln x = \ln s_0^{-\alpha} - \alpha \ln s$ , což je lineární závislost. Sledovanou veličinu lze tedy modelovat Paretovým rozdělením, pokud transformovaná data (tj. logaritmus standardu a kumulativní relativní četnosti seřazené v obráceném pořadí) leží přibližně na přímce. Parametr má význam opačné hodnoty směrnice této



přímky. K posouzení kvality modelu lze užít standardních metod statistické indukce.

Další charakteristiku Paretova rozdělení dostaneme následovně: Označme  $p = F(s)$   $s \geq s_0$ ,  $p \in \langle 0, 1 \rangle$ , tedy  $s = F^{-1}(p)$  je kvantilová funkce rozdělení. Výpočtem dostáváme

$$1 - p = \left( \frac{s}{s_0} \right)^{-\alpha}, \text{ kde } F_0 = 0, \text{ odtud (9)}$$

$$\frac{s_1}{s_2} = \left( \frac{1 - p_2}{1 - p_1} \right)^{1/\alpha}, \text{ (10)}$$

tedy podíl standardů Paretova rozdělení je dán výrazem (10). Vzorce (10) lze užít k výpočtu odhadu parametrů Paretova rozdělení, viz např. (Bílková, 2009).

Střední hodnota Paretova rozdělení je  $\mu = \int_0^{\infty} s f(s) ds = \int_{s_0}^{\infty} \alpha s_0^{\alpha} s s^{-\alpha-1} ds = \alpha \left[ \frac{s^{-\alpha+1}}{-\alpha+1} \right]_{s_0}^{\infty} = s_0 \frac{\alpha}{1-\alpha}$ ; nevlastní integrál konverguje pro  $\alpha > 1$ .

Hodnotu G-koefficientu Paretova rozdělení získáme ze vzorce  $G = \frac{1}{\mu_0} \int_0^{\infty} F(s)(1 - F(s)) ds = \frac{1}{2\alpha - 1}$ ; integrál konverguje pro  $\alpha > \frac{1}{2}$ .

Nakonec při zvolené hodnotě parametru  $s_0$  uvedeme výpočet odhadu parametru  $\alpha$ . Předpokládejme, že jsou k dispozici hodnoty kvantilů souboru dat  $x_p$  pro zvolené hodnoty  $p_i, i = 1, 2, \dots, n$ . Označme příslušnou hodnotu standardu Paretova rozdělení.

Podle (9) platí  $s_p = \frac{s_0}{(1 - p_i)^{1/\alpha}}$ . Minimalizace kvadratické odchylky  $\sum_{i=1}^n (x_p - s_p)^2$  v proměnné  $\alpha$  nemá analytické řešení. Minimalizace výrazu  $\sum_{i=1}^n (\ln x_p - \ln s_p)^2$  je dána řešením rovnice

$$\sum_i \left( (\ln x_p - \ln s_0 + \frac{1}{\alpha} \ln(1 - p_i)) \left( -\frac{1}{\alpha^2} \right) \ln(1 - p_i) \right) = 0. \text{ Odtud } \alpha = \frac{\sum_i \ln^2(1 - p_i)}{\sum_i (\ln s_0 - \ln x_p)}.$$

Kvalitnější model lze vytvořit užitím Paretova rozdělení o více parametrech (Arnold, 2008).

Na závěr (bez odvození) uvedme dvě důležitá vyjádření G-koefficientu. Jsou-li známy pravděpodobnosti  $P(s_i)$  uspořádaného souboru  $\{s_i\}_{i=1}^n, s_i < s_{i+1}$  standardu, potom

$$G = 1 - \frac{\sum_{i=1}^n P(s_i)(F_{i-1} + F_i)}{F_n}, \text{ kde } F_0 = 0,$$

$$F_i = \sum_{j=1}^i P(s_j)s_j.$$

Je-li ze základního souboru proveden náhodný výběr  $S = \{s_i\}_{i=1}^n$  a je označen tak, že je uspořádan  $(s_i \leq s_{i+1})$  potom statistika

$$G(s) = \frac{1}{n-1} \left( n+1 - 2 \frac{\sum_{i=1}^n (n+1-i)s_i}{\sum_{i=1}^n s_i} \right) = 1 - \frac{2}{n-1} \left( n - \frac{\sum_{i=1}^n i s_i}{\sum_{i=1}^n s_i} \right) \quad (11)$$

je konzistentní odhad G-koefficientu základního souboru, více viz (Deaton, 1997). Vzor (11) je modifikací vzorce (1).

## Výsledky a diskuze

Pro sledování nerovnoměrnosti byla vybrána data cestovního ruchu v ročních časových řadách. Jedná se o data týkající se výjezdů českých návštěvníků (rezidentů) do nejnavštěvovanějších evropských států. Nerovnoměrnost je porovnána s celkovou návštěvností těchto států všemi nerezidenty (uvažováni jsou evropští i mimoevropští návštěvníci). Tento soubor dat označme číslicí I. Dále byl prošetřen soubor dat popisující příjezdy zahraničních návštěvníků do různých krajů České republiky (tento soubor označme II). Pro výběr datových souborů cestovního ruchu bylo využito několika databází.

Zabývejme se nejprve datovým souborem příjezdového cestovního ruchu vybraných evropských států. Výjezdy českých návštěvníků jsou definovány jako delší cesty, tj. s délkou trvání přesahující 3 noci. Tyto údaje jsou přístupné na stránkách Českého statistického úřadu ([www.czso.cz](http://www.czso.cz)). Vzhledem ke skutečnosti, že byla v roce 2009 změněna metodika šetření, je tato studie zaměřena na data zjištěná do roku 2008 včetně. Ke zjištění počtu všech nerezidentů navštěvujících tyto destinace byla využita databáze Eurostat s volbou počtu příjezdů do hromadných ubytovacích zařízení ([epp.eurostat.ec.europa.eu](http://epp.eurostat.ec.europa.eu)). Tato data jsou s údaji o výjezdech českých rezidentů v absolutních číslech porovnatelná jen do určité míry. Mohou však sloužit jako užitečný podklad pro

sledování změn trendů návštěvnosti. Při sběru všech těchto informací je pro členské státy Evropské unie platná směrnice 95/57/ES o sběru statistických informací z oblasti cestovního ruchu.

Při zpracování matic datového souboru I, tedy údajů o výjezdech českých návštěvníků a celkovém počtu nerezidentů evropských států, je využito rozmezí let 2003 – 2008. Tento soubor vstupních dat byl pro vhodnou interpretaci transformován na jednotkovou návštěvnost; viz Přílohy 1 a 2. <sup>4</sup> Pomocí vybraných ukazatelů byla studována nerovnoměrnost. Výpočty byly provedeny v prostředí Matlab. Při grafickém zpracování Lorenzovy křivky je využito modifikace programu gini (Lengwiler, 2010).

Hodnoty Giniho koeficientů zjištěné podle vztahu (1) jsou uvedeny v následující tabulce.

**Tabulka 2: Giniho koeficienty (datový soubor I)**

|                   | 2003  | 2004  | 2005  | 2006  | 2007  | 2008  |
|-------------------|-------|-------|-------|-------|-------|-------|
| Čeští návštěvníci | 0,508 | 0,436 | 0,482 | 0,474 | 0,453 | 0,420 |
| Zahr. návštěvníci | 0,513 | 0,507 | 0,500 | 0,502 | 0,497 | 0,493 |

*Zdroj: vlastní výpočty*

Odpovídající Lorenzovy křivky jsou uvedeny na obrázku 3.

Lze konstatovat, že v průběhu sledovaných let dochází v obou uvažovaných datových souborech k mírnému snižování nerovnoměrnosti, což je z hlediska cestovního ruchu žádoucí jev. Téměř výhradně kontinuální pokles nerovnoměrnosti je zřejmý u celkového počtu zahraničních návštěvníků v uvažovaných destinacích. Z hlediska výjezdů českých návštěvníků je tento pokles přerušen rokem 2005. Analýzou transformovaných vstupních dat (viz Příloha 1) můžeme říci, že rok 2005 je poznamenán zejména výrazným snížením návštěvnosti Chorvatska a současně vzrůstem návštěvnosti Slovenska, českými turisty nejnavštěvovanějších destinací. Chorvatsko vykazuje v rozmezí let 2004 – 2008 významný pokles počtu českých návštěvníků, který se výrazně projevuje ve tvaru Lorenzovy křivky.

<sup>4</sup>) Ve výčtu států je užito kódů Evropské unie.

Druhý datový soubor popisuje návštěvnost krajů České republiky ([www.czso.cz](http://www.czso.cz)) v letech 2001 až 2011. Byl zvolen počet přenocování nerezidentů v hromadných ubytovacích zařízeních. Odpovídající hodnoty Giniho koeficientů jsou uvedeny v tabulce 3.

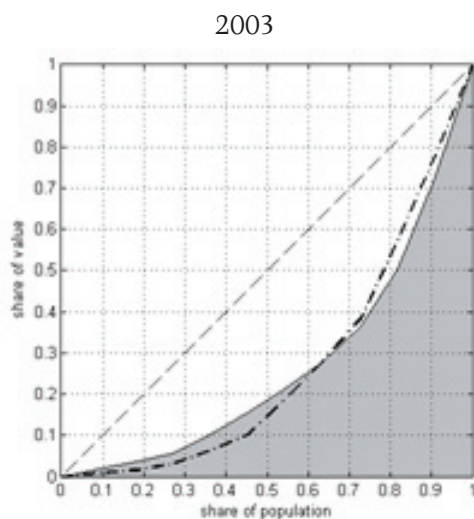
**Tabulka 3: Giniho koeficienty (datový soubor II)**

| 2001  | 2002  | 2003  | 2004  | 2005  | 2006  | 2007  | 2008  | 2009  | 2010  | 2011  |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0,616 | 0,585 | 0,628 | 0,677 | 0,694 | 0,695 | 0,713 | 0,738 | 0,759 | 0,773 | 0,773 |

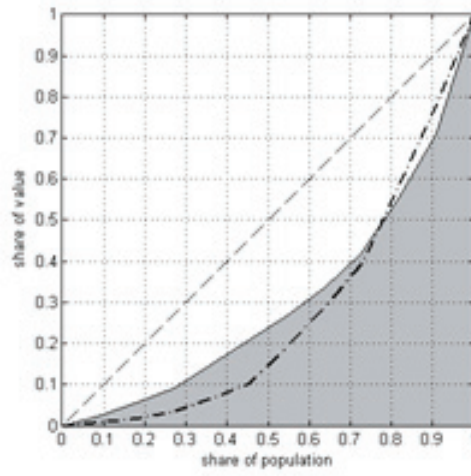
*Zdroj: vlastní výpočty*

**Obrázek 3: Lorenzovy křivky (pro období 2003 – 2008)**

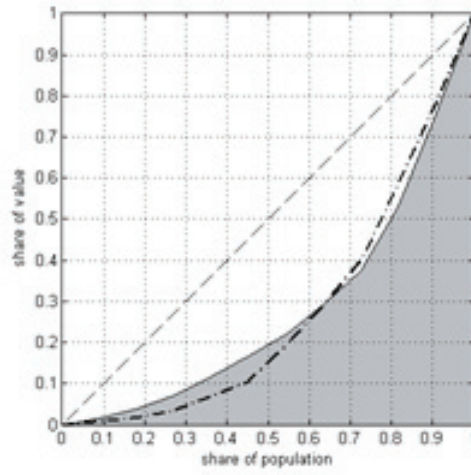
*(Plná čára s výplní v grafech odpovídá zastoupení českých návštěvníků; čerchovaná čára značí celkový počet nerezidentů.)*



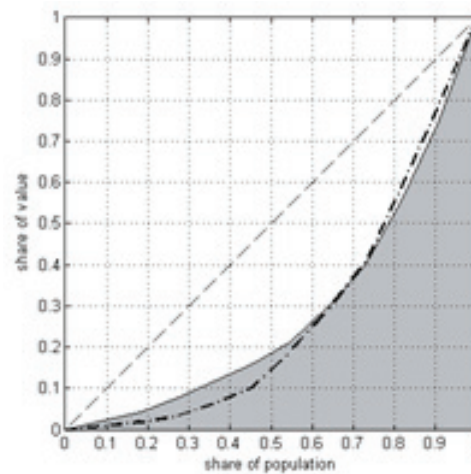
2004



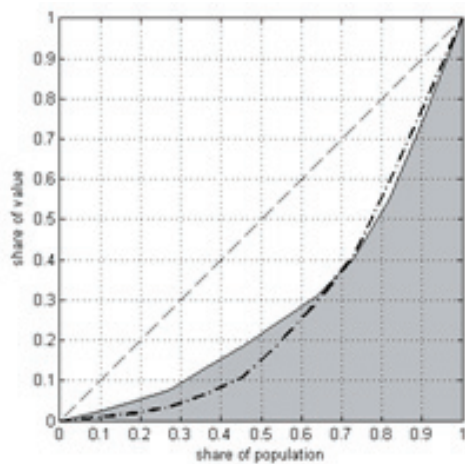
2005



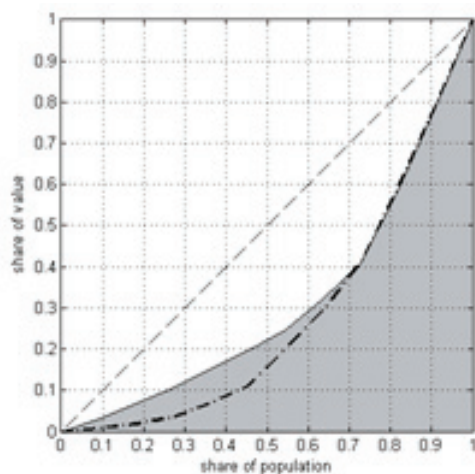
2006



2007



2008

*Zdroj: vlastní zpracování*

Datový soubor II ukazuje výrazné zvyšování nerovnoměrnosti. Pozornost zahraničních návštěvníků je tedy zřejmě více soustředěna na kraje Praha a Karlovarský, přičemž trend této návštěvnosti v průběhu sledovaných let roste.

## Závěr

V článku jsou metodami diverzifikace vyhodnoceny výběrové datové matice z oblasti cestovního ruchu a malých podniků. Tyto metody lze rovněž aplikovat na data z mnoha jiných oblastí ekonomie. Dále má článek poukázat na nezastupitelnou roli (zde elementárního) matematického aparátu při formulaci ekonomických zákonitostí.

## References

- [1] Atkinson, A. B.: On the measurement of inequality. *Journal of Economic Theory* 2, 1970, s. 244-263. ISSN: 0022-0531.
- [2] Arnold, B.C.: Pareto and generalized Pareto distributions. In: *Modelling Income Distributions and Lorenz Curves*, ed. D. Chotikapanich, vol. 5 of *Economic Studies in Equality, Social Exclusion and Well-Being*, kap. 7, s. 119-145. Springer, New York 2008. ISBN: 978-03-8772-756-1.
- [3] Bílková, D.: Application of lognormal curves in modelling of wage distributions. *Journal of Applied Mathematics* 1, 2008, s. 341-352. ISSN: 1337-6365.
- [4] Bílková, D.: Pareto distribution and wage models. *Journal of Applied Mathematics* 2, 2009, s. 37-46. ISSN: 1337-6365.
- [5] Deaton, A.: *The analysis of household surveys*. The John Hopkins University Press. Baltimore 1997. ISBN: 978-0-8018-5254-1.
- [6] Keller, J. *Tři sociální světy*. Sociologické nakladatelství. Praha 2010. ISBN: 978-80-7419-031-5.
- [7] Lengwiler Y.: Gini. Matlab fileexchange, 2010. Dostupné na: <http://www.mathworks.com/matlabcentral/fileexchange/28080-gini-coefficient-and-the-lorentz-curve>, staženo 14. 2. 2012.
- [8] Sen, A.: *On economic inequality*. Clarendon Press. Oxford 1973. ISBN: 978-0-1982-8182-5.
- [9] Směrnice Rady 95/57/ES ze dne 23. listopadu 1995 o sběru statistických informací v oblasti cestovního ruchu. (Council Directive 95/57/EC of 23 November 1995 on the collection of statistical information in the field of tourism.)
- [10] URL: [http://www.czso.cz/csu/redakce.nsf/i/cru40\\_cr](http://www.czso.cz/csu/redakce.nsf/i/cru40_cr), staženo 11. 4. 2012.
- [11] URL: [http://www.czso.cz/csu/redakce.nsf/i/cru\\_cr](http://www.czso.cz/csu/redakce.nsf/i/cru_cr), staženo 15.11.2012.
- [12] URL: [http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search\\_database](http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database), staženo 11. 4. 2012.
- [13] Xu, K.: How has the literature on Gini's index evolved in the past 80 years? *China Economic Quarterly* 2, 2003, s. 757-778.

**Příloha 1: Výjezdový cestovní ruch českých návštěvníků**

|      | SK     | HR     | IT     | EL     | AT     | UK     | ES     | FR     | BG     | HU     | DE     |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 2003 | 0,2220 | 0,2748 | 0,1398 | 0,0855 | 0,0455 | 0,0186 | 0,0630 | 0,0580 | 0,0211 | 0,0188 | 0,0529 |
| 2004 | 0,1546 | 0,2907 | 0,1353 | 0,0579 | 0,0848 | 0,0258 | 0,0607 | 0,0659 | 0,0345 | 0,0319 | 0,0579 |
| 2005 | 0,2229 | 0,2444 | 0,1578 | 0,0806 | 0,0723 | 0,0229 | 0,0490 | 0,0462 | 0,0325 | 0,0164 | 0,0548 |
| 2006 | 0,1926 | 0,2497 | 0,1603 | 0,1010 | 0,0845 | 0,0341 | 0,0522 | 0,0418 | 0,0214 | 0,0215 | 0,0409 |
| 2007 | 0,2135 | 0,2369 | 0,1508 | 0,0913 | 0,0605 | 0,0546 | 0,0600 | 0,0545 | 0,0285 | 0,0269 | 0,0224 |
| 2008 | 0,2151 | 0,2024 | 0,1735 | 0,0850 | 0,0781 | 0,0483 | 0,0450 | 0,0449 | 0,0403 | 0,0368 | 0,0305 |

*Zdroj: ČSÚ a vlastní výpočty*

**Příloha 2: Příjezdový cestovní ruch nerezidentů**

|      | SK     | HR     | IT     | EL     | AT     | UK     | ES     | FR     | BG     | HU     | DE     |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 2003 | 0,0076 | 0,0326 | 0,1946 | 0,0376 | 0,0905 | 0,0883 | 0,2037 | 0,2188 | 0,0078 | 0,0164 | 0,1022 |
| 2004 | 0,0074 | 0,0335 | 0,1961 | 0,0345 | 0,0889 | 0,0939 | 0,1939 | 0,2176 | 0,0091 | 0,0175 | 0,1076 |
| 2005 | 0,0076 | 0,0338 | 0,1942 | 0,0375 | 0,0874 | 0,0972 | 0,1909 | 0,2141 | 0,0099 | 0,0176 | 0,1097 |
| 2006 | 0,0077 | 0,0321 | 0,1988 | 0,0374 | 0,0847 | 0,1011 | 0,2082 | 0,1905 | 0,0099 | 0,0160 | 0,1137 |
| 2007 | 0,0078 | 0,0328 | 0,2002 | 0,0418 | 0,0846 | 0,0984 | 0,2053 | 0,1887 | 0,0104 | 0,0161 | 0,1139 |
| 2008 | 0,0082 | 0,0335 | 0,1976 | 0,0420 | 0,0902 | 0,0924 | 0,2066 | 0,1849 | 0,0105 | 0,0166 | 0,1175 |

*Zdroj: Eurostat a vlastní výpočty*

*Miloslav Malec, Katedra ekonomie a ekonomiky, Vysoká škola hotelová v Praze, Svědnická 506, 181 00 Praha 8, Česká republika, E-mail: malec@vsh.cz*

*Lukáš Malec, Katedra matematiky a statistiky, Vysoká škola obchodní v Praze, Spálená 76/14, 110 00 Praha 1, Česká republika, E-mail: malec@vso-praha.eu*

*Rostislav Tomeš, Katedra výpočetní techniky, Vysoká škola obchodní v Praze, Spálená 76/14, 110 00 Praha 1, Česká republika, E-mail: tomes@vso-praha.eu*