



家計調査におけるジニ係数の計測についての一考察

各務, 和彦

(Citation)

國民經濟雜誌, 216(6):47-57

(Issue Date)

2017-12-10

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/E0041343>

(URL)

<https://hdl.handle.net/20.500.14094/E0041343>



家計調査における
ジニ係数の計測についての一考察

各 務 和 彦

国民経済雑誌 第 216 巻 第 6 号 抜刷

平成 29 年 12 月

家計調査における ジニ係数の計測についての一考察

各 務 和 彦

本稿では対数正規分布, Dagum 分布, Singh-Maddala 分布を対象として, 日本の家計調査の勤労者世帯におけるジニ係数の精度を実証的に検討した。所得分布とローレンツ曲線に基づく, 2つの異なるアプローチによる尤度を用いて, マルコフ連鎖モンテカルロ法によるベイズ推定を行った結果, これらの分布を仮定して, ローレンツ曲線による推定を行えば, 概ね良好な結果が得られることが明らかになった。同時に, Dagum 分布と Singh-Maddala 分布のパラメータを直接推定すると, ジニ係数を過大に評価する可能性があることが明らかになった。また, 所得分布のパラメータを直接推定するときには, 対数正規分布を仮定すれば妥当な結果が得られそうであることもわかった。

キーワード ジニ係数, 所得分布, マルコフ連鎖モンテカルロ法,
ローレンツ曲線

1 は じ め に

所得分配の不平等度は古くから研究されてきたテーマの一つであり, 経済学においても今なお関心を集めている研究テーマである。特に, ジニ係数はその解釈が容易なために, 所得分配の不平等度の変化を検討する上で, 広く使われる指標の一つとなっている。ジニ係数は均等分配線とローレンツ曲線で囲まれた面積の2倍として定義されており, ローレンツ曲線と密接な関係がある。そして, Iritani and Kuga (1983) で指摘されているように, ローレンツ曲線は所得分布と対応関係があり, ローレンツ曲線や所得分布の推定は, ジニ係数の計測と関連した, 統計学や計量経済学における重要な課題の一つとなっている。

所得分布の推定においては, 駿河 (1982) でも述べられているように, パラメータの数が少ない分布は推定が簡単なだけでなく, 意味づけも明らかになる一方, 分布のあてはまりは悪い。他方, パラメータの数を増やすと, 分布のあてはまりは改善する一方, 推定が困難になるだけでなく, 意味づけも難しくなる傾向がある。特に, 本稿でも用いる, グループ・データで分析をする場合には, さらに注意すべき点がある。例えば, McDonald and Ransom

(1979a, b) で述べられているように、推定結果は関数型の選択や推定方法に敏感であったりすることが指摘されている。

ローレンツ曲線の推定においても、様々な注意が必要である。例えば、Slottje (1990) では、Gastwirth (1972) で提案されたノンパラメトリックな方法を併用しながら、パラメトリックな方法によるジニ係数の計算を推奨する一方で、Schader and Schmid (1994) は、パラメトリックなローレンツ曲線をあてはめることは、ドイツのデータにおいては、ジニ係数の精度の観点から問題があることを指摘している。ただし、後述するように、パラメトリックにジニ係数を計算する方法は、ローレンツ曲線による方法だけでなく、所得分布の推定からも可能である。しかしながら、これらの方法を比較した研究はないと思われる。そこで、本稿では実証分析で広く使われ、あてはまりがよいと言われている、対数正規分布、Dagum 分布 (Dagum, 1977)、Singh-Maddala 分布 (Singh and Maddala, 1976) を取り上げて、総務省によって報告されている『家計調査報告』(以下、家計調査)の勤労者世帯におけるジニ係数の推定の精度を分析する¹⁾。

本稿では、推定方法を統一するために、マルコフ連鎖モンテカルロ (Markov chain Monte Carlo: MCMC) 法を用いたベイズ推定を行う。そして、ジニ係数の推定の精度を家計調査のデータに基づいて実証的に検討する。Gelfand *et al.* (1990) で述べられているように、MCMC 法ではパラメータの非線形な関数に対しての統計的推測が可能であり、パラメータの関数として表されるジニ係数の推測に適していると考えられるからである。つまり、事後平均や事後信用区間を求めることで、ジニ係数に対する統計的推測が可能になるという利点がある。データに関しては、2009年と2012年のデータを採用し、勤労者世帯の5分位と10分位の両方のデータから分析を行った。分析の結果、取り上げた3つの分布で、ローレンツ曲線を推定すれば、概ね良好な結果が得られることがわかった。同時に、Dagum 分布と Singh-Maddala 分布のパラメータを直接推定すると、ジニ係数が過大に推定される可能性があることが明らかになった。結論としては、対数正規分布を仮定すれば、推定方法にかかわらず、概ね良好な推測が可能であると言えるであろう。

本稿の構成は以下の通りである。次節では、所得分布とローレンツ曲線に基づくパラメータの推定方法と、推定されたパラメータからジニ係数を計測する方法を解説する。第3節では、日本の家計調査のデータを用いて、計測されたジニ係数の精度について議論する。第4節では、この分析において得られた結論と、今後の課題について述べる。

2 分析手法

$k+1$ の階級があるグループ・データにおいては、表1に示されているような情報が利用可能である。言い換えれば、 $n = \sum_{j=1}^{k+1} n_j^*$ 世帯の標本を $k+1$ 個の階級 ($x_0=0, x_1, x_2, \dots, x_k$,

表 1 グループ・データの例

所得階級	階級平均所得	世帯数
0 — x_1	$\bar{x}_1$	n_1^*
x_1 — x_2	$\bar{x}_2$	n_2^*
$\vdots$	$\vdots$	$\vdots$
x_{k-1} — x_k	$\bar{x}_k$	n_k^*
x_k — ∞	$\bar{x}_{k+1}$	n_{k+1}^*

$x_{k+1}=\infty$), その階級に属する世帯数 ($n_1^*, n_2^*, \dots, n_{k+1}^*$), 階級平均所得 ($\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{k+1}$) に集約したデータが提供されている。また, $n_i = \sum_{j=1}^i n_j^*$ と $s_i = \sum_{j=1}^i \bar{x}_j n_j^*$ が, それぞれ, 第 n_i 家計までの累積家計数と累積所得を表すものとする。このとき, $s = \sum_{j=1}^{k+1} \bar{x}_j n_j^*$ を総所得として, $z_i = \frac{n_i}{n}$ と $\eta_i = \frac{s_i}{s}$ を計算することで, 累積家計比率 ($z_0=0, z_1, \dots, z_k, z_{k+1}=1$) と累積所得比率 ($\eta_0=0, \eta_1, \dots, \eta_k, \eta_{k+1}=1$) という, 異なる形式のデータ・セットが利用可能となる。このデータに基づいて, ヒストグラムを描いたり, ジニ係数を計算したりすることは可能であるが, n 世帯の観測値がパラメトリックな所得分布からのランダムな標本であると考えて統計的推測を行うことも可能である。そして, このデータを用いて, パラメトリックにジニ係数を計測することができる。

ジニ係数の計測方法を説明するに当たっては, ローレンツ曲線を定義する必要がある。そこで, 以下の分位点関数を定義する。

$$F^{-1}(y) = \inf \{x : F(x) \geq y\} \quad (1)$$

すると, (1) を用いて, ローレンツ曲線が以下のように定義される。

$$L(z) = \frac{1}{\mu} \int_0^z F^{-1}(y) dy, \quad 0 \leq z \leq 1 \quad (2)$$

そして, $F(x)$ を累積分布関数, $f(x)$ を確率密度関数とすると, ジニ係数は以下の式を用いて計算できることが知られている。

$$G = -1 + \frac{2}{\mu} \int_0^\infty x F(x) f(x) dx \quad (3)$$

$$= 1 - 2 \int_0^1 L(z) dz \quad (4)$$

ただし, (3) では, 期待値 μ が存在することを仮定している。これまでに, 様々な分布からのジニ係数が陽表的に導出されている (例えば, Sarabia (2008) 等を参照)。しかしながら, Hajargasht *et al.* (2012) でも述べられているように, ジニ係数の陽表的な表現が利用可

能でなかったとしても、(3)や(4)の積分をすることで、ジニ係数を計算することができる。従って、 $F(x)$ と $f(x)$ あるいは $L(z)$ のパラメータを推定することは、パラメトリックにジニ係数を計測する上で重要であることがわかる。本稿では、これらを陽表的に表すことのできる分布に限定し、分析の対象とした分布を中心に、代表的な分布についての詳細を表2にまとめている（さらに詳しくは、Kleiber and Kotz (2003) を参照）。

(3)と(4)からわかるように、 $F(x)$ と $f(x)$ または $L(z)$ のパラメータが推定できれば、ジニ係数を計測することが可能である。しかしながら、これら2つのアプローチに対する研究は独立に行われており、 $F(x)$ と $f(x)$ のパラメータを推定する方法は、例えば、McDonald (1984) や McDonald and Xu (1995) で行われてきた。他方、 $L(z)$ のパラメータを推定する方法は、例えば、Kakwani and Podder (1973) や Chotikapanich and Griffiths (2002) で行われてきた。このような現状において、2つのアプローチに基づくジニ係数の比較を行った研究は見当たらないと思われる。そこで、本稿では、推定方法を統一して、ベイズMCMC法による推定に基づいて検証していく。Gelfand *et al.* (1990) では、MCMC法はパラメータの非線形な関数として表されたものに対しての統計的推測を可能にするということが述べられている。表2に示されているように、ジニ係数はパラメータの非線形な関数なので、本稿の目的に適した推定方法と考えられるからである。

そこで、最初に、 $F(x)$ や $f(x)$ のパラメータを推定する方法について説明する。グループ・データから $F(x)$ や $f(x)$ のパラメータを推定する方法はいくつも存在する。それらは、最尤法や最小2乗法、ベイズMCMC法 (Chotikapanich and Griffiths, 2000) 等である。本稿では、MCMC法を採用するので、尤度を定義する必要がある。尤度に基づく推定では、McDonald and Ransom (1979a, b) で用いられているような、多項分布に基づく尤度が広く使われている。しかしながら、Nishino and Kakamu (2011) で示されているように、順序統計量に基づく尤度の方がより厳密であることから、Nishino and Kakamu (2011) に従って、以下の尤度を用いる。

$$\mathcal{L}_D(x|\theta) = n! \frac{F(x_1)^{n_1^*-1}}{(n_1-1)!} f(x_1) \left\{ \prod_{i=2}^k \frac{(F(x_i) - F(x_{i-1}))^{n_i^*-1}}{(n_i^*-1)!} f(x_i) \right\} \frac{(1 - F(x_k))^{n_{k+1}^*}}{(n_{k+1}^*)!} \quad (5)$$

ただし、 $x = (x_1, \dots, x_k)'$ で、 θ は関心のある分布のパラメータを表す。注意すべき点は、階級と世帯数、階級平均所得のデータが利用可能であるにもかかわらず、階級と世帯数のみを利用している点である。

次に、 $L(z)$ のパラメータを推定する方法について述べる。累積家計比率と累積所得比率から、Kakwani and Podder (1973) は最初に、最小2乗法によるローレンツ曲線のパラメータを推定する方法を提案した。その後、 $L(z)$ のパラメータの推定においても、一般化最小

2乗法 (Kakwani and Podder, 1976), 最尤法 (Chotikapanich and Griffiths, 2002), MCMC 法 (Chotikapanich and Griffiths, 2005) を含む, 様々な推定方法が提案されてきた。本稿では, Chotikapanich and Griffiths (2005) に則った, バイズ・アプローチを採用する。そこで, Chotikapanich and Griffiths (2002) で提案された尤度を用いることとすると, 尤度は以下のように定義される。

$$\mathcal{L}_{LC}(\eta|\theta^*) = \Gamma(\lambda) \prod_{i=1}^{k+1} \frac{(\eta_i - \eta_{i-1})^{\lambda[L(z_i|\theta) - L(z_{i-1}|\theta)]}}{\Gamma(\lambda[L(z_i|\theta) - L(z_{i-1}|\theta)])} \quad (6)$$

ただし, $\eta = (\eta_0, \dots, \eta_{k+1})'$, $z = (z_0, \dots, z_{k+1})'$, $\theta^* = (\theta', \lambda)'$ で, $\Gamma(\cdot)$ はガンマ関数である。 λ は分布とは独立した追加的なパラメータである (Chotikapanich and Griffiths (2002) を参照)。ここで注意すべき点は, $F(x)$ や $f(x)$ のパラメータの推定とは異なり, 用いるデータは階級平均所得と世帯数から計算された, 累積家計比率と累積所得比率の情報であるという点である。

表2 所得分布

分布	確率密度関数	累積分布関数	ローレンツ曲線	ジニ係数
パレート分布	$\alpha \theta^\alpha x^{-\alpha-1}$	$1 - \left(\frac{x}{\theta}\right)^{-\alpha}$	$1 - (1-z)^{1-1/\alpha}$	$\frac{1}{2\alpha-1}$
対数正規分布	$\frac{1}{\sqrt{2\pi\sigma^2}x} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}$	$\Phi\left(\frac{\ln x - \mu}{\sigma}\right)$	$\Phi(\Phi^{-1}(z) - \sigma)$	$2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$
Dagum 分布	$\frac{apx^{ap-1}}{b^{ap}\left[1 + \left(\frac{x}{b}\right)^{ap}\right]^{p+1}}$	$\left[1 + \left(\frac{x}{b}\right)^{ap}\right]^{-p}$	$I_t(p+1/a, 1-1/a)$ ただし, $t = z^{1/p}$	$\frac{\Gamma(p)\Gamma(2p+1/a)}{\Gamma(2p)\Gamma(p+1/a)} - 1$
Singh-Maddala 分布	$\frac{aqx^{a-1}}{b^a\left[1 + \left(\frac{x}{b}\right)^a\right]^{1+q}}$	$1 - \left[1 + \left(\frac{x}{b}\right)^a\right]^{-q}$	$I_t(1+1/a, q-1/a)$ ただし, $t = 1 - (1-z)^{1/q}$	$1 - \frac{\Gamma(q)\Gamma(2q-1/a)}{\Gamma(q-1/a)\Gamma(2q)}$

$\Phi(\cdot)$ は標準正規分布の累積分布関数, $\Phi^{-1}(\cdot)$ は標準正規分布の分位点関数, $I_x(a, b)$ は正規化された不完全ベータ関数を表す。

バイズ推定を行うに当たっては, 尤度が定義されるだけでなく, 推定対象となるパラメータに対して事前分布 $\pi(\theta)$ または $\pi(\theta^*)$ を割り当て²⁾, 同時事後分布

$$\pi(\theta|x) \propto \pi(\theta)\mathcal{L}_D(x|\theta) \quad (7)$$

$$\pi(\theta^*|\eta) \propto \pi(\theta^*)\mathcal{L}_{LC}(\eta|\theta^*) \quad (8)$$

が必要になる。(7)と(8)が与えられると, MCMC 法によるパラメータ推定とジニ係数の計測が可能となる。(7)に基づく $F(x)$ や $f(x)$ のパラメータを推定する方法については, Kakamu (2016) で詳細に述べられているので, ここでは, (8)に基づく $L(z)$ を用いた MCMC 法について説明することにする。Kakamu (2016) では, Chotikapanich and Griffiths

(2000) の酔歩連鎖による MCMC 法を修正しており、以下の手順で乱数を生成する。

Step 1. 初期値 $\theta^{*(0)}$ を設定する。

Step 2. $\mathcal{N}(\theta^{*(m-1)}, c^2 \Sigma)$ から候補値 θ^{*new} を生成する。ただし、 c はチューニング・パラメータであり、 Σ は θ^* の最尤推定値の分散共分散行列である。³⁾

Step 3. 以下の採択確率を計算する。

$$\alpha(\theta^{*(m-1)}, \theta^{*new}) = \min \left\{ 1, \frac{\pi(\theta^{*new} | \eta)}{\pi(\theta^{*(m-1)} | \eta)} \right\}$$

そして、任意の θ^{*new} が制約の外でサンプリングされたら、 $\alpha(\theta^{*(m-1)}, \theta^{*new}) = 0$ と置く。

Step 4. $\mathcal{U}(a, b)$ を (a, b) 区間の一様分布を表すとして、 $U(0, 1)$ から u を生成する。

Step 5. もし $u \leq \alpha(\theta^{*(m-1)}, \theta^{*new})$ ならば、 $\theta^{*(m)} = \theta^{*new}$ として、そうでなければ、 $\theta^{*(m)} = \theta^{*(m-1)}$ とする。

Step 6. m を $m+1$ として、Step 2. に戻る。

以上の方法を用いて、以下では実証的にジニ係数の精度について計測することにする。

3 実 証 分 析

分析に先立って、本稿で用いたデータを説明する。2009年と2012年の家計調査の勤労者世帯の5分位と10分位のデータを用いた。データの詳細については、豊田・和合(1977)やKakamu and Nishino (2016), 各務(2013)を参照されたい。そして、MCMC法に対しては、200,000回の反復を行い、MCMC法による事後分布からの確率標本を得た。そして、稼働検査期間として最初の50,000回を切り捨て、残りの150,000回のサンプルを事後分布からサンプリングされたものとみなして推定に用いた。本稿の推定結果はOx Version 7.10 (OS_X_64/U)を用いて得られたものである(Doornik (2009)を参照)。

まず、表3には、2009年の推定から得られた、ジニ係数の結果が示されている。ジニ係数の精度に関して言及するときに使われる、Gastwirth (1972) のノンパラメトリックに推定したときの、ジニ係数の上限と下限もあわせて報告されている。まず、5分位データについて見てみると、事後平均はDagum分布とSingh-Maddala分布の分布のパラメータを直接推定するときを除き、Gastwirth (1972) の上限と下限の間で推定されていることがわかる。特に、対数正規分布では、精度よく推定されており、ローレンツ曲線を推定したときの上限を除いて、Gastwirth (1972) の上限と下限の間で推定されている。それに対して、Dagum分布とSingh-Maddala分布の分布のパラメータを直接推定した結果を見てみると、95%信用区間の下限が、Gastwirth (1972) の上限よりも大きく推定されており、ジニ係数を過大に推定して

表 3 2009年の推定結果

データ	分布	推定	下限	平均	上限
5 分位	Gastwirth (1972)		0.238		0.261
	対数正規分布	分布	0.251 *	0.256 *	0.261 *
		ローレンツ曲線	0.243 *	0.253 *	0.263 +
	Dagum 分布	分布	0.270 +	0.277 +	0.284 +
		ローレンツ曲線	0.232 -	0.254 *	0.275 +
	Singh-Maddala 分布	分布	0.262 +	0.271 +	0.281 +
		ローレンツ曲線	0.237 -	0.253 *	0.268 +
10分位	Gastwirth (1972)		0.249		0.256
	対数正規分布	分布	0.250 *	0.253 *	0.258 +
		ローレンツ曲線	0.248 -	0.253 *	0.258 +
	Dagum 分布	分布	0.264 +	0.270 +	0.275 +
		ローレンツ曲線	0.240 -	0.251 *	0.262 +
	Singh-Maddala 分布	分布	0.258 +	0.264 +	0.270 +
		ローレンツ曲線	0.244 -	0.252 *	0.260 +

推定結果の上限と下限は95%信用区間を表す。また、-は Gastwirth (1972) の下限よりも小さいことを、*は下限と上限の間にあることを、+は上限よりも大きいことを表す。

いる可能性があることがわかる。

同様の結果が10分位データからも見て取ることができるが、注目すべき点は、データの区切りが多くなった分、Gastwirth (1972) のノンパラメトリックに推定したときの、ジニ係数の上限と下限は5分位と比べて狭く推定されるが、それにもかかわらず、5分位の結果と大きく変わらない点である。事後平均は Dagum 分布と Singh-Maddala 分布の分布のパラメータを直接推定するときを除き、Gastwirth (1972) の上限と下限の間で推定されているだけでなく、95%信用区間はどの推定結果も5分位の結果と比べて狭く推定されていることがわかる。そして、5分位の結果と比べてみると、いずれの値も Gastwirth (1972) の上限に近づいており、区切りを増やしたことによる結果の改善も確認できる。

次に、表4には、2012年の推定から得られた、ジニ係数の結果が示されている。まず、5分位の結果に注目すると、2009年の結果と異なり、ローレンツ曲線の推定では事後平均が Gastwirth (1972) の上限と下限の間で推定されている一方、分布のパラメータの推定ではどれも上限と下限の外で推定されている。Dagum 分布と Singh-Maddala 分布の分布のパラメータを直接推定した結果を見てみると、95%信用区間の下限が、Gastwirth (1972) の上限よりも大きく推定されており、2009年の結果と同様に、ジニ係数を過大に推定している可能性があることがわかる。一方、対数正規分布では、分布のパラメータを推定すると、いずれの値も Gastwirth (1972) の下限より小さく推定されており、対数正規分布が必ずしもあてはま

表 4 2012年の推定結果

データ	分布	推定	下限	平均	上限
5 分位	Gastwirth (1972)		0.231		0.254
	対数正規分布	分布	0.181 —	0.188 —	0.196 —
		ローレンツ曲線	0.242 *	0.246 *	0.250 *
	Dagum 分布	分布	0.259 +	0.265 +	0.272 +
		ローレンツ曲線	0.230 —	0.246 *	0.261 +
	Singh-Maddala 分布	分布	0.255 +	0.264 +	0.274 +
		ローレンツ曲線	0.235 *	0.247 *	0.259 +
10分位	Gastwirth (1972)		0.242		0.248
	対数正規分布	分布	0.239 —	0.243 *	0.247 *
		ローレンツ曲線	0.243 *	0.247 *	0.251 +
	Dagum 分布	分布	0.255 +	0.260 +	0.265 +
		ローレンツ曲線	0.236 —	0.245 *	0.252 +
	Singh-Maddala 分布	分布	0.251 +	0.257 +	0.263 +
		ローレンツ曲線	0.239 —	0.245 *	0.251 +

推定結果の上限と下限は95%信用区間を表す。また、—は Gastwirth (1972) の下限よりも小さいことを、*は下限と上限の間にあることを、+は上限よりも大きいことを表す。

りがいいというわけではないことがわかる。

10分位の結果に目を移すと、2009年と同様の結果が得られていることがわかる。データの区切りが多くなった分、Gastwirth (1972) の上限と下限は狭く推定されているが、事後平均は Dagum 分布と Singh-Maddala 分布の分布のパラメータを直接推定するときを除き、Gastwirth (1972) の上限と下限の間で推定されている。事後平均が Gastwirth (1972) の上限と下限の間で推定されている結果だけに注目すると、95%信用区間は、どれも Gastwirth (1972) の上限と下限よりも広いが、対数正規分布が最も精度よく推定できていそうである。

以上の結果を要約すると、本稿で取り上げた、対数正規分布、Dagum 分布、Singh-Maddala 分布を仮定して、ローレンツ曲線に基づくパラメータ推定を行えば、家計調査の勤労者世帯のデータを使う限りにおいては、ジニ係数の計測という観点から見れば、概ね良好な結果が得られそうであるということがわかった。特に、対数正規分布を採用すれば、分布を推定しても概ね良好な結果が得られ、Nishino and Kakamu (2011) と同様の結果が得られているように見える。しかしながら、分布のあてはまりとジニ係数の精度には必ずしも一貫性がないということに言及しておかなければならない。例えば、Dagum 分布や Singh-Maddala 分布は多くの国で、あてはまりがいいということが報告されているが、本稿の結果からはジニ係数の過大評価が見られた。従って、分布のあてはまりとジニ係数の精度に関する関係は更なる分析が必要であるが、家計調査の勤労者世帯の分析においては、対数正規分

布は分布のあてはまりも、ジニ係数の精度も良さそうであるという結果は注目に値するであろう。

4 結 論

本稿では、2009年と2012年の家計調査の勤労者世帯の5分位と10分位のデータを用いて、パラメトリックにジニ係数を計測したときのジニ係数の精度について考察をした。ジニ係数の計測に際しては、対数正規分布、Dagum 分布、Singh-Maddala 分布を仮定して、分布のパラメータを直接推定する方法とローレンツ曲線を推定する方法の比較も行った。分析の結果、これらの分布を仮定して、ローレンツ曲線による推定を行えば、概ね良好な結果が得られることが明らかになった。同時に、Dagum 分布と Singh-Maddala 分布のパラメータを直接推定すると、ジニ係数には上方バイアスが存在することが明らかになった。また、所得分布のパラメータを直接推定するならば、対数正規分布を仮定すれば妥当な結果が得られそうであることもわかった。これらの結果を総合すると、Slottje (1990) の言葉を借りれば、家計調査の勤労者世帯のデータを使う限りにおいては、「ワープロを切って、眠れそうである」(“turn off the word processor and go to bed”)。しかしながら、所得分配の分析で用いるデータは家計調査だけでないことに注意しなければならない。Slottje (1990) や Schader and Schmid (1994) で指摘されているように、国や年度が変われば、支持される結果が変わるように、用いるデータが変わっても支持される結果が変わることが予想される。本稿の結果は近年の家計調査における傾向を示したに過ぎず、一般的な結論ではない。この問題を解決するには、Kakamu and Nishino (2016) や Kobayashi and Kakamu (2016) で考えられているような、柔軟性の高い分布の推定が鍵になるかもしれない。また、所得分布のパラメータの推定に当たっては、積率条件を用いた Hitomi *et al.* (2008) の推定方法が推定精度の改善に対する鍵になるかもしれない。これらの問題については今後の課題としたい。

注

本稿はウィーン経済大学滞在中に執筆されたものである。なお、本稿は科学研究費補助金（#16KK0081, #16K03592, #25245035）の助成を受けている。記して感謝したい。

- 1) 西埜他 (2009) でも検討されているように、パレート分布も所得分布を考える上で重要な分布の一つであるが、ローレンツ曲線による推定で、あてはまりが悪いことに起因する推定上の問題が生じたことから、分析の対象から外した。
- 2) 本稿では、対数正規分布、Dagum 分布、Singh-Maddala 分布を分析の対象とした。事前分布 $\pi(\theta)$ に対して、対数正規分布では $1/\sigma$ としたのに対し、Dagum 分布と Singh-Maddala 分布では $1/b$ とした。他方、事前分布 $\pi(\theta^*)$ に対しては、一律で $1/\lambda$ とした。
- 3) 最尤法によってパラメータのモードを求めることが困難なことがあるので、Goffe *et al.* (1994)

による simulated annealing を MCMC サンプリング・スキームに入る前に 1 度実行し、 Σ を求める。そして、 Σ のコレスキー分解ができないときには、Nocedal and Wright (2000) の一般化コレスキー分解を用いる。 c は、Holloway *et al.* (2002) の調整法を用いて、稼働検査期間中に調整される。

参 考 文 献

- Chotikapanich, D. and Griffiths, W. E. (2000) "Posterior distributions for the Gini coefficient using grouped data," *Australian and New Zealand Journal of Statistics*, **42**, 383-392.
- Chotikapanich, D. and Griffiths, W. E. (2002) "Estimating Lorenz curves using a Dirichlet distribution," *Journal of Business & Economic Statistics*, **20**, 290-295.
- Chotikapanich, D. and Griffiths, W. E. (2005) "Averaging Lorenz curves," *Journal of Economic Inequality*, **3**, 1-19.
- Dagum, C. (1977) "A new model of personal income distribution: Specification and estimation," *Économie Appliquée*, **30**, 413-437.
- Doornik, J. A. (2009) *Ox: An Object Oriented Matrix Programming Language*, Timberlake, London.
- Gastwirth, J. L. (1972) "The estimation of the Lorenz curve and Gini index," *Review of Economics and Statistics*, **54**, 306-316.
- Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990) "Illustration of Bayesian inference in normal data models using Gibbs sampling," *Journal of the American Statistical Association*, **85**, 972-985.
- Goffe, W. L., Ferrier, G. D. and Rogers, J. (1994) "Global optimization of statistical functions with simulated annealing," *Journal of Econometrics*, **60**, 65-99.
- Hajargasht, G., Griffiths, W. E., Brice, J., Rao, D. S. P. and Chotikapanich, D. (2012) "Inference for income distributions using grouped data," *Journal of Business & Economic Statistics*, **30**, 563-575.
- Hitomi, K., Liu, Q.-F., Nishiyama, Y. and Sueishi, N. (2008) "Efficient estimation and model selection for grouped data with local moments," *Journal of the Japan Statistical Society*, **38**, 131-143.
- Holloway, G., Shankar, B. and Rahman, S. (2002) "Bayesian spatial probit estimation: A primer and an application to HYV rice adoption," *Agricultural Economics*, **27**, 384-402.
- Iritani, J. and Kuga, K. (1983) "Duality between the Lorenz curves and the income distribution functions," *The Economic Studies Quarterly*, **34**, 9-21.
- Kakamu, K. (2016) "Simulation studies comparing Dagum and Singh-Maddala income distributions," *Computational Economics*, **48**, 593-605.
- Kakamu, K. and Nishino, H. (2016) "Bayesian estimation of beta-type distribution parameters from grouped data," Discussion Paper No. 2016-08, Graduate School of Business Administration, Kobe University.
- Kakwani, N. C. and Podder, N. (1973) "On the estimation of Lorenz curves from grouped observations," *International Economic Review*, **14**, 278-292.
- Kakwani, N. C. and Podder, N. (1976) "Efficient estimation of the Lorenz curve and associated inequality measures from grouped observations," *Econometrica*, **44**, 137-148.

- Kleiber, C. and Kotz, S. (2003) *Statistical Size Distributions in Economics and Actuarial Science*, Wiley, New York.
- Kobayashi, G. and Kakamu, K. (2016) “Approximate Bayesian computation for Lorenz curves from grouped data,” arXiv: 1607.02735.
- McDonald, J. B. (1984) “Some generalized functions for the size distribution of income,” *Econometrica*, **52**, 647–663.
- McDonald, J. B. and Ransom, M. R. (1979a) “Functional forms, estimation techniques and the distribution of income,” *Econometrica*, **47**, 1513–1525.
- McDonald, J. B. and Ransom, M. R. (1979b) “Alternative parameter estimators based upon grouped data,” *Communications in Statistics*, **A8**, 899–917.
- McDonald, J. B. and Xu, Y. J. (1995) “A generalization of the beta distribution with applications,” *Journal of Econometrics*, **66**, 133–152.
- Nishino, H. and Kakamu, K. (2011) “Grouped data estimation and testing of Gini coefficients using lognormal distributions,” *Sankhya Series B*, **73**, 193–210.
- Nocedal, J. and Wright, S. J. (2000) *Numerical Optimization*, Second Edition, Springer, New York.
- Sarabia, J. M. (2008) “Parametric Lorenz curves: Models and applications,” in Chotikapanich, D. (ed.), *Modelling Income Distributions and Lorenz Curves*, Springer, New York, 167–190.
- Schader, M. and Schmid, F. (1994) “Fitting parametric Lorenz curves to grouped income distributions — A critical note,” *Empirical Economics*, **19**, 361–370.
- Singh, S. K. and Maddala, G. S. (1976) “A function for size distribution of income,” *Econometrica*, **44**, 963–970.
- Slottje, D. J. (1990) “Using grouped data for constructing inequality indices: Parametric vs. non-parametric methods,” *Economics Letters*, **32**, 193–197.
- 各務和彦 (2013) 「家計調査からみた日本の所得分配：Singh-Maddala 分布による検討」, 『千葉大学経済研究』, 第28巻, 83–99.
- 駿河輝和 (1982) 「所得分布の関数型：『家計調査年報』年間収入の場合」, 『季刊理論経済学』, 第33巻, 79–85.
- 豊田敬・和合肇 (1977) 「昭和40年代の職業別所得不平等度とその計測」, 『国民経済』, No. 137, 41–60.
- 西埜晴久・各務和彦・大鋸崇 (2009) 「パレート分布を用いた経済格差の検定」, 『日本統計学会誌』, 第38巻, 151–164.