

http://www.geocities.jp/stat_okamoto/jp/にプレゼン資料をアップ

二重パラメータ化二重パレート対数正規分布 の所得分布への当てはめ

2014年度 統計関連学会連合大会

総務省 岡本政人



所得分布の当てはめに用いられる統計分布モデル

対数正規分布とその一般化モデル

2パラメータ: 対数正規分布(LN)

4パラメータ: 二重パレート対数正規分布(dPLN)

5パラメータ: 一般化二重パレート対数正規分布(GdPLN)

など

ベータ分布の一般化モデル

3パラメータ: Singh-Maddala分布, Dagum分布

4パラメータ: 第一種・第二種一般化ベータ分布(GB1, GB2)

5パラメータ: 一般化ベータ分布(GB)

など

従来, ベータ分布の一般化モデルの適合度が高く, よく使われてきたが, Reed (2003)が提案したdPLNが, それらのモデルよりも優る事例が見出されたことにより対数正規分布の一般化モデルが注目されている



ベータ分布の一般化モデル

第1種一般化ベータ分布(GB1) (McDonald,1984)

$$X \sim GB1(a, b, p, q) \quad \text{if} \quad Z = \left(\frac{X}{b}\right)^a, \quad Z \sim Beta(p, q)$$

$$\text{p.d.f. : } f_{GB1}(x; a, b, p, q) = \frac{ax^{ap-1}[1 - (x/b)^a]^{q-1}}{b^{ap}B(p, q)}$$

第2種一般化ベータ分布(GB2) (McDonald,1984等) Feller-Pareto分布(の特殊型)

$$X \sim GB2(a, b, p, q) \quad \text{if} \quad Z = \frac{\left(\frac{X}{b}\right)^a}{1 + \left(\frac{X}{b}\right)^a}, \quad Z \sim Beta(p, q)$$

$$\text{p.d.f. : } f_{GB2}(x; a, b, p, q) = \frac{ax^{ap-1}}{b^{ap}B(p, q)[1 + (x/b)^a]^{p+q}}$$

$p = 1$: Singh-Maddala (SM)分布; $q = 1$: Dagum分布; $q \rightarrow \infty, b/q^{\frac{1}{a}} \rightarrow \beta$: 一般化ガンマ(GG)分布



二重パレート対数正規分布 (dPLN) とその一般化

対数正規分布 (LN) $\log x \sim N(\mu, \sigma)$

$$\text{p.d.f. : } f_{LN}(x; \mu, \sigma) = \frac{1}{x\sigma} \phi\left(\frac{\log x - \mu}{\sigma}\right) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}$$

二重パレート対数正規分布 (dPLN) (Reed, 2003; Gabaix, 1999)

都市の人口規模
分布の研究

$$\log X = Z + V - W \quad Z, V, W \text{ は互いに独立}$$

$$Z \sim N(\mu, \sigma) \quad V, W \sim \text{スケールパラメータ } \alpha(>0), \beta(>0) \text{ の指数分布}$$

$$\text{p.d.f. : } f_{dPLN}(x|\mu, \sigma, \alpha, \beta) = \frac{\alpha\beta}{\alpha + \beta} x^{\beta-1} e^{-\beta\mu + \beta^2\sigma^2/2} \Phi^c\left(\frac{\log x - \mu + \beta\sigma^2}{\sigma}\right) + \frac{\alpha\beta}{\alpha + \beta} x^{-\alpha-1} e^{\alpha\mu + \alpha^2\sigma^2/2} \Phi\left(\frac{\log x - \mu - \alpha\sigma^2}{\sigma}\right)$$

$$\log X \text{ の c.f. : } \phi_{NL}(s) = \frac{\alpha\beta e^{i\mu s - \frac{1}{2}\sigma^2 s^2}}{(\alpha - is)(\beta + is)}$$

標準正規分布の c.d.f.

$$f_{dPLN}(x; \mu, \sigma, \alpha, \beta) \sim c_1 x^{-\alpha-1} \quad (x \rightarrow \infty), \quad f_{dPLN}(x; \mu, \sigma, \alpha, \beta) \sim c_2 x^{\beta-1} \quad (x \rightarrow 0)$$

一般化二重対数正規分布 (GdPLN) (Reed and Wu, 2008)

$$\log X = Z + V - W \quad Z, V, W \text{ は互いに独立}$$

$$Z \sim N(\mu, \sigma)$$

$V, W \sim$ スケールパラメータ $\alpha(>0), \beta(>0)$, シェープパラメータ $\rho(>0)$ のガンマ分布

$$\log X \text{ の c.f. : } \phi_{GNL}(s) = \left[\frac{\alpha\beta e^{i\mu s - \frac{1}{2}\sigma^2 s^2}}{(\alpha - is)(\beta + is)} \right]^\rho$$

$$f_{GdPLN}(x; \mu, \sigma, \alpha, \beta) \sim c_1 (\log x)^{\rho-1} x^{-\alpha-1} \quad (x \rightarrow \infty)$$

$$f_{GdPLN}(x; \mu, \sigma, \alpha, \beta) \sim c_2 (-\log x)^{\rho-1} x^{\beta-1} \quad (x \rightarrow 0)$$

- 問題点:
- p.d.f. が陽に表現できないため, c.f. を数値解析的に逆変換し実データへの当てはめを行う必要があり, 実用的でない.
 - ローレンツ曲線(格差指標)の推定精度も考慮すると, 必ずしも適合度がdPLNよりも改善するとは言えない.



二重パラメータ化二重パレート対数正規分布 (dP²LN) (Okamoto, 2014)

p.d.f. :
$$f_{dP^2LN}(x|\mu_L, \mu_R, \sigma_L, \sigma_R, \alpha, \beta, r) = r g_L(x|\mu_L, \sigma_L, \alpha, \beta) + (1 - r) g_R(x|\mu_R, \sigma_R, \alpha) =$$

$$r \beta x^{\beta-1} e^{-\beta \mu_L + \beta^2 \sigma_L^2 / 2} \Phi^c \left(\frac{\log x - \mu_L + \beta \sigma_L^2}{\sigma_L} \right) + (1 - r) \alpha x^{-\alpha-1} e^{\alpha \mu_R + \alpha^2 \sigma_R^2 / 2} \Phi \left(\frac{\log x - \mu_R - \alpha \sigma_R^2}{\sigma_R} \right)$$

$$\sigma_L, \sigma_R, \alpha, \beta > 0, 0 < r < 1$$

$\mu_L = \mu_R, \sigma_L = \sigma_R, r = \alpha / (\alpha + \beta)$ のとき dPLN と一致

$g_L(x|\mu_L, \sigma_L, \alpha, \beta) = \beta x^{\beta-1} e^{-\beta \mu_L + \beta^2 \sigma_L^2 / 2} \Phi^c \left(\frac{\log x - \mu_L + \beta \sigma_L^2}{\sigma_L} \right)$ は以下の確率変数 X_L の p.d.f.

$$\log X_L = Z_L - W \quad Z_L, W \text{ は互いに独立}$$

$$Z_L \sim N(\mu_L, \sigma_L) \quad W \sim \text{スケールパラメータ } \beta \text{ の指数分布}$$

$g_R(x|\mu_R, \sigma_R, \alpha) = \alpha x^{-\alpha-1} e^{\alpha \mu_R + \alpha^2 \sigma_R^2 / 2} \Phi \left(\frac{\log x - \mu_R - \alpha \sigma_R^2}{\sigma_R} \right)$ は以下の確率変数 X_R の p.d.f.

$$\log X_R = Z_R + V \quad Z_R, V \text{ は互いに独立}$$

$$Z_R \sim N(\mu_R, \sigma_R) \quad V \sim \text{スケールパラメータ } \alpha \text{ の指数分布}$$

dP²LN は,
混合比 $r:1-r$
の混合分布

$$f_{dP^2LN}(x; \mu_L, \mu_R, \sigma_L, \sigma_R, \alpha, \beta, r) \sim d_1 x^{-\alpha-1} \quad (x \rightarrow \infty)$$

$$f_{dP^2LN}(x; \mu_L, \mu_R, \sigma_L, \sigma_R, \alpha, \beta, r) \sim d_2 x^{\beta-1} \quad (x \rightarrow 0)$$

◎ dP²LNの $h(< \alpha)$ 次モーメント

$$\mu_h = E(X^h) = r \frac{\beta}{\beta + h} e^{\mu_L h + \frac{1}{2} h^2 \sigma_L^2} + (1 - r) \frac{\alpha}{\alpha - h} e^{\mu_R h + \frac{1}{2} h^2 \sigma_R^2}$$

◎ dP²LNの平均値(1次モーメント)

$$\mu = E(X) = r \frac{\beta}{\beta + 1} e^{\mu_L + \frac{1}{2} \sigma_L^2} + (1 - r) \frac{\alpha}{\alpha - 1} e^{\mu_R + \frac{1}{2} \sigma_R^2}$$



$(\alpha, \beta \rightarrow \infty)$

$$r e^{\mu_L + \frac{1}{2} \sigma_L^2} + (1 - r) e^{\mu_R + \frac{1}{2} \sigma_R^2}$$

◎ dP²LNのc.d.f.

$$F(x) = r \left[x^{\beta} e^{-\beta\mu_L + \frac{1}{2}\beta^2\sigma_L^2} \Phi^c \left(\frac{\log x - \mu_L + \beta\sigma_L^2}{\sigma_L} \right) + \Phi \left(\frac{\log x - \mu_L}{\sigma_L} \right) \right] + \\ (1-r) \left[-x^{-\alpha} e^{\alpha\mu_R + \frac{1}{2}\alpha^2\sigma_R^2} \Phi \left(\frac{\log x - \mu_R - \alpha\sigma_R^2}{\sigma_R} \right) + \Phi \left(\frac{\log x - \mu_R}{\sigma_R} \right) \right]$$

◎ dP²LNの正規化1次不完全モーメント

$$F_{(1)}(x) = \frac{\int_0^x t f(t) dt}{\mu} = \\ \frac{r}{\mu} \frac{\beta}{\beta+1} \left[x^{\beta+1} e^{-\beta\mu_L + \frac{1}{2}\beta^2\sigma_L^2} \Phi^c \left(\frac{\log x - \mu_L + \beta\sigma_L^2}{\sigma_L} \right) + e^{\mu_L + \frac{1}{2}\sigma_L^2} \Phi \left(\frac{\log x - \mu_L - \sigma_L^2}{\sigma_L} \right) \right] + \\ \frac{1-r}{\mu} \frac{\alpha}{\alpha-1} \left[-x^{-\alpha+1} e^{\alpha\mu_R + \frac{1}{2}\alpha^2\sigma_R^2} \Phi \left(\frac{\log x - \mu_R - \alpha\sigma_R^2}{\sigma_R} \right) + e^{\mu_R + \frac{1}{2}\sigma_R^2} \Phi \left(\frac{\log x - \mu_R - \sigma_R^2}{\sigma_R} \right) \right]$$

◎ dP²LNのローレンツ曲線

$$L(\theta) = F_{(1)}(F^{-1}(\theta)) \quad 0 < \theta < 1$$

● dP²LNの平均対数偏差

$$MLD = \log(\mu) - r\mu_L - (1-r)\mu_R + \frac{r}{\beta} - \frac{1-r}{\alpha}$$

● dP²LNのタイル係数

$$Theil = \frac{r\beta \left[-\frac{1}{(\beta+1)^2} e^{\mu_L + \sigma_L^2/2} + \frac{1}{\beta+1} \sigma_L^2 e^{\mu_L + \sigma_L^2/2} \right] + (1-r)\alpha \left[\frac{1}{(\alpha-1)^2} e^{\mu_R + \sigma_R^2/2} + \frac{1}{\alpha-1} \sigma_R^2 e^{\mu_R + \sigma_R^2/2} \right]}{\mu} - \log(\mu)$$

● dP²LNの平方変動係数

$$SCV = \frac{1}{2} \left[\frac{r \frac{\beta}{\beta+2} e^{2\mu_L + 2\sigma_L^2} + (1-r) \frac{\alpha}{\alpha-2} e^{2\mu_R + 2\sigma_R^2}}{\mu^2} - 1 \right]$$

● dP²LNのジニ係数

$$G = 2 \frac{r^2 I_{LL} + (1-r)^2 I_{RR} + r(1-r) I_{RL} + r(1-r) I_{LR}}{\mu} - 1$$

ここで,

$$I_{LL} = \frac{1}{\beta(\beta+1)} e^{\mu_L + \frac{1}{2}\sigma_L^2} \Phi\left(\frac{1}{\sqrt{2}}\sigma_L\right) + \frac{1}{\beta} e^{\mu_L + (\beta^2 + \beta + \frac{1}{2})\sigma_L^2} \left[\frac{2}{2\beta+1} - \frac{1}{\beta+1}\right] \Phi\left(-\frac{2\beta+1}{\sqrt{2}}\sigma_L\right)$$

$$I_{RR} = \frac{1}{\alpha(\alpha-1)} e^{\mu_R + \frac{1}{2}\sigma_R^2} \Phi\left(\frac{1}{\sqrt{2}}\sigma_R\right) + \frac{1}{\alpha} e^{\mu_R + (\alpha^2 - \alpha + \frac{1}{2})\sigma_R^2} \left[\frac{1}{\alpha-1} - \frac{2}{2\alpha-1}\right] \Phi\left(-\frac{2\alpha-1}{\sqrt{2}}\sigma_R\right)$$

$$\text{if } \alpha \neq \beta + 1 \left[\begin{aligned} I_{RL} &= \frac{1}{\alpha} \frac{1}{\beta+1} \exp\left(\frac{1}{2}\sigma_L^2 + \mu_L\right) \Phi\left(\frac{-\mu_R + \mu_L + \sigma_L^2}{\sqrt{2}\sigma_{RL}}\right) + \frac{1}{\alpha} \frac{1}{\alpha-\beta-1} \exp\left(\alpha^2\sigma_{RL}^2 - \frac{2\alpha-1}{2}\sigma_L^2 + \alpha\mu_R - (\alpha-1)\mu_L\right) \Phi\left(\frac{-\mu_R + \mu_L - 2\alpha\sigma_{RL}^2 + \sigma_L^2}{\sqrt{2}\sigma_{RL}}\right) \\ &\quad - \frac{1}{(\beta+1)(\alpha-\beta-1)} \exp\left(\beta^2\sigma_{RL}^2 + \frac{2\beta+1}{2}\sigma_R^2 + (\beta+1)\mu_R - \beta\mu_L\right) \Phi\left(\frac{-\mu_R + \mu_L - 2\beta\sigma_{RL}^2 - \sigma_R^2}{\sqrt{2}\sigma_{RL}}\right) \\ I_{LR} &= \frac{1}{\beta} \frac{1}{\alpha-1} \exp\left(\frac{1}{2}\sigma_R^2 + \mu_R\right) \Phi\left(\frac{-\mu_L + \mu_R + \sigma_R^2}{\sqrt{2}\sigma_{LR}}\right) + \frac{1}{\beta} \frac{1}{\alpha-\beta-1} \exp\left(\beta^2\sigma_{LR}^2 + \frac{2\beta+1}{2}\sigma_R^2 - \beta\mu_L + (\beta+1)\mu_R\right) \Phi\left(-\frac{-\mu_L + \mu_R + 2\beta\sigma_{LR}^2 + \sigma_R^2}{\sqrt{2}\sigma_{LR}}\right) \\ &\quad - \frac{1}{(\alpha-1)(\alpha-\beta-1)} \exp\left(\alpha^2\sigma_{LR}^2 - \frac{2\alpha-1}{2}\sigma_L^2 - (\alpha-1)\mu_L + \alpha\mu_R\right) \Phi\left(-\frac{-\mu_L + \mu_R + 2\alpha\sigma_{LR}^2 - \sigma_L^2}{\sqrt{2}\sigma_{LR}}\right) \end{aligned} \right]$$

$$\text{if } \alpha = \beta + 1 \left[\begin{aligned} I_{RL} &= \frac{1}{(\beta+1)^2} \exp\left(\frac{1}{2}\sigma_L^2 + \mu_L\right) \Phi\left(\frac{-\mu_R + \mu_L + \sigma_L^2}{\sqrt{2}\sigma_{RL}}\right) - \frac{1}{\beta+1} \sqrt{2}\sigma_{RL} \exp\left(\frac{1}{2}\sigma_R^2 + \mu_R\right) \phi\left(\frac{-\mu_R + \mu_L - \sigma_R^2}{\sqrt{2}\sigma_{RL}}\right) + \\ &\quad \frac{1}{\beta+1} \left[(2\beta\sigma_{RL}^2 + \sigma_R^2 + \mu_R - \mu_L) - \frac{1}{\beta+1}\right] \exp\left(\beta^2\sigma_{RL}^2 + \frac{2\beta+1}{2}\sigma_R^2 + (\beta+1)\mu_R - \beta\mu_L\right) \Phi\left(\frac{-\mu_R + \mu_L - 2\beta\sigma_{RL}^2 - \sigma_R^2}{\sqrt{2}\sigma_{RL}}\right) \\ I_{LR} &= \frac{1}{(\alpha-1)^2} \exp\left(\frac{1}{2}\sigma_R^2 + \mu_R\right) \Phi\left(\frac{-\mu_L + \mu_R + \sigma_R^2}{\sqrt{2}\sigma_{LR}}\right) + \frac{1}{\alpha-1} \sqrt{2}\sigma_{LR} \exp\left(\frac{1}{2}\sigma_L^2 + \mu_L\right) \phi\left(-\frac{-\mu_L + \mu_R - \sigma_L^2}{\sqrt{2}\sigma_{LR}}\right) - \\ &\quad \frac{1}{\alpha-1} \left[(2\alpha\sigma_{LR}^2 - \sigma_L^2 - \mu_L + \mu_R) - \frac{1}{\alpha-1}\right] \exp\left(\alpha^2\sigma_{LR}^2 - \frac{2\alpha-1}{2}\sigma_L^2 - (\alpha-1)\mu_L + \alpha\mu_R\right) \Phi\left(-\frac{-\mu_L + \mu_R + 2\alpha\sigma_{LR}^2 - \sigma_L^2}{\sqrt{2}\sigma_{LR}}\right) \end{aligned} \right]$$



統計分布モデルの所得分布への適合度の比較

■ 使用データ（いずれも世帯年収データ，等価調整せず）

- 1984～2009年全国消費実態調査，2人以上の世帯の粗収入，集計値
- 1992～2010年US Survey of Consumer Finances (SCF)，総世帯の可処分所得，マイクロデータ
- 2000～2012年Italian Survey of Household Income and Wealth (SHIW)，総世帯の粗収入，マイクロデータ

■ 最尤法による当てはめ

マイクロデータの場合： $\ell\ell = \sum_{i=1}^n w_i \log f(x_i; \text{para.}) \rightarrow \text{最大化}$

w_i ：世帯 i の集計ウェイト

$$\sum_{i=1}^n w_i = n$$

集計値の場合：
(固定金額階級の場合)

$$\ell\ell = c + \sum_{k=1}^K w_k \log[F(x_k; \text{para.}) - F(x_{k-1}; \text{para.})] \rightarrow \text{最大化}$$

$$c = \log \frac{n!}{\prod_{k=1}^K w_k!}$$

$$x_0 = 0, x_K = \infty$$

$$\sum_{k=1}^K w_k = n$$

w_k ：所得階級区分 k の集計ウェイト

適合度の評価指標

赤池情報量基準

$$AIC = -2 \cdot \ell\ell + 2 \cdot \#para.$$

ローレンツ曲線の平方誤差の平方根

$$LRSSE = \sqrt{\sum_{i=1}^n \left(\hat{L}_i - L(\theta_i; para.) \right)^2}$$

LRSSEは、パラメータ数が多い一般化したモデルを最尤法で当てはめた場合、尤度や χ^2 統計量などとは異なり、むしろ悪化する傾向があるため、パラメータ数に応じたペナルティは不要

累積世帯割合: $\theta_i = \sum_{j \leq i} w_j / \sum_{j=1}^n w_j$

累積金額割合: $\hat{L}_i = \sum_{j \leq i} w_j x_j / \sum_{j=1}^n w_j x_j$

x_j は昇順



比較結果 — 全国消費実態調査, 2人以上の世帯の粗収入



統計分布モデルの適合度

	パラ メー タ数	AIC (dPLNのAICとの差を表章)							LRSSE						
		1984	1989	1994	1999	2004	2009	平均	1984	1989	1994	1999	2004	2009	RMSE
GG	3	458.2	317.6	215.8	79.8	135.2	116.4	220.5	0.043	0.040	0.028	0.022	0.028	0.022	0.032
SM	3	20.2	27.6	39.8	172.8	163.4	164.6	98.1	0.005	0.017	0.012	0.010	0.010	0.007	0.011
GB2	4	12.0	14.2	2.6	19.0	23.0	19.0	15.0	0.003	0.021	0.018	0.018	0.020	0.014	0.017
E _κ G1	4	20.6	8.0	12.4	97.6	73.0	70.2	47.0	0.002	0.024	0.017	0.013	0.014	0.008	0.015
E _κ G2	4	37.0	28.6	5.4	25.2	31.2	25.6	25.5	0.008	0.019	0.016	0.017	0.019	0.014	0.016
IE _κ G1	4	13.2	42.0	37.6	186.0	151.8	155.2	97.6	0.007	0.013	0.009	0.007	0.007	0.010	0.009
IE _κ G2	4	-1.0	9.0	3.6	16.6	18.8	15.4	10.4	0.004	0.023	0.019	0.018	0.020	0.014	0.017
dPLN	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.003	0.021	0.017	0.019	0.019	0.013	0.017
GB	5	-20.4	13.6	4.6	15.6	24.8	21.0	9.9	0.051	0.031	0.018	0.021	0.021	0.014	0.029
GdPLN	5	-0.4	-8.2	2.0	-34.4	-55.4	-28.2	-20.8	0.003	0.021	0.017	0.091	0.030	0.009	0.041
dP ² LN _{μr}	6	-28.4	-8.6	-4.6	-55.0	-140.8	-132.6	-61.7	0.015	0.003	0.014	0.003	0.009	0.014	0.011
dP ² LN	7	-29.8	-9.0	-4.6	-81.4	-170.2	-168.6	-77.3	0.012	0.468	0.011	0.002	0.028	0.024	0.192

dP²LN_{μr}: dP²LNで $\sigma_L = \sigma_R$
という制約を課した場合

E_κG1, E_κG2: 第1種, 第2種拡張κG分布

(Okamoto, 2013)

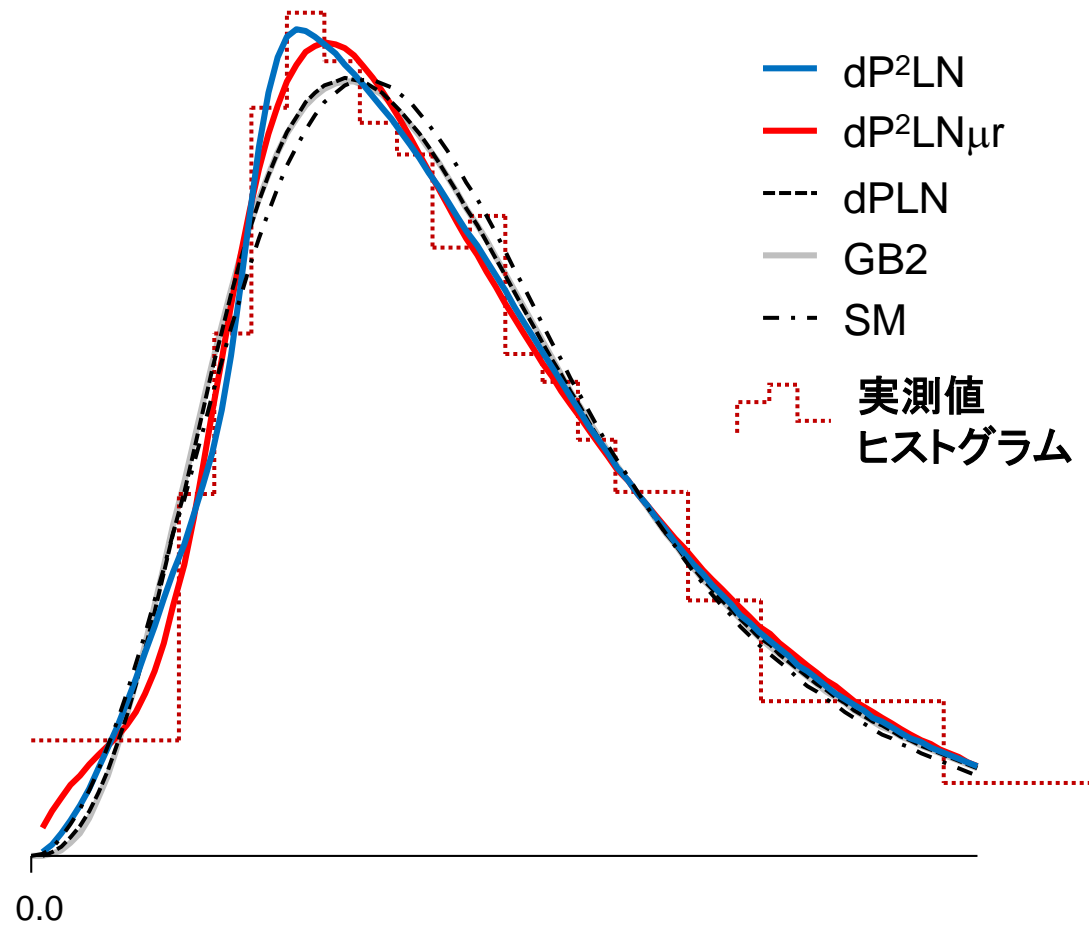
IE_κG1, IE_κG2: 第1種, 第2種拡張κG分布に従う確率変数の逆数の分布

● 当てはめた統計分布モデルから推定したジニ係数

	パラ メー タ数	ジニ係数						推定* RMSE
		1984	1989	1994	1999	2004	2009	
GG	3	0.268	0.281	0.289	0.294	0.300	0.305	0.010
SM	3	0.280	0.288	0.294	0.299	0.306	0.312	0.004
GB2	4	0.282	0.286	0.292	0.295	0.302	0.307	0.006
EκG1	4	0.281	0.285	0.292	0.297	0.304	0.309	0.005
EκG2	4	0.284	0.287	0.292	0.295	0.302	0.307	0.006
IEκG1	4	0.283	0.289	0.295	0.301	0.308	0.313	0.003
IEκG2	4	0.280	0.285	0.291	0.295	0.302	0.307	0.006
dPLN	4	0.282	0.286	0.292	0.295	0.302	0.307	0.006
GB	5	0.262	0.282	0.292	0.294	0.301	0.307	0.011
GdPLN	5	0.282	0.286	0.292	0.296	0.302	0.310	0.006
dP ² LN _{μr}	6	0.276	0.293	0.293	0.301	0.313	0.317	0.004
dP ² LN	7	0.286	0.472	0.294	0.301	0.319	0.320	0.073
公表値		0.280	0.293	0.297	0.301	0.308	0.311	

* 公表値は、十分位階級別集計値にシンプソン近似公式を適用して算出しており、0.001程度過小推定になっているとみられるため、公表値に0.001加算した値との差のRMSEをとっている。

● 推定密度分布, 2009年



● dP²LN, dP²LN_{μr}の最尤パラメータ

	dP ² LN							dP ² LN _{μr}					
	μ_L	μ_R	σ_L	σ_R	α	β	r	μ_L	μ_R	σ	α	β	r
1984	8.71	8.92	0.292	0.179	3.47	2.83	0.861	8.62	8.78	0.247	4.56	2.79	0.714
1989	9.03	8.46	0.355	0.442	3.74	2.24	0.455	9.03	8.32	0.381	2.58	2.53	0.680
1994	9.26	8.57	0.345	0.462	3.61	2.09	0.458	8.81	8.94	0.299	4.12	2.92	0.510
1999	9.18	8.10	0.419	0.138	1.76	2.63	0.898	9.33	8.31	0.349	2.62	2.01	0.616
2004	9.08	8.06	0.421	0.133	1.68	2.53	0.846	9.29	8.21	0.318	2.38	1.76	0.526
2009	9.00	8.03	0.444	0.105	1.72	2.57	0.861	9.25	8.16	0.320	2.41	1.64	0.511

p.d.fはいずれも単峰

比較結果 — アメリカSCFの総世帯の可処分所得と イタリアSHIWの総世帯の粗収入

統計分布モデルの適合度

	パラ メータ数	アメリカSCF, 1992～2010			イタリアSHIW, (2000～2012)		
		AIC	LRSSE	ジニ係数の誤差	AIC	LRSSE	ジニ係数の誤差
SM	3	6.9	1.695	0.0249	19.0	0.185	0.0022
Dagum	3	4.1	1.350	0.0198	33.6	0.515	0.0080
κ G	3	14.7	1.248	0.0198	68.5	0.377	0.0054
$l\kappa$ G	3	45.7	1.079	0.0170	121.0	1.376	0.0222
GB1	4	152.3	3.153	0.0395	273.1	0.510	0.0021
GB2	4	0.9	1.717	0.0256	15.1	0.174	0.0024
$E\kappa$ G1	4	33.3	1.519	0.0277	16.4	0.162	0.0019
$E\kappa$ G2	4	-5.6	1.767	0.0274	25.3	0.295	0.0044
$IE\kappa$ G1	4	-6.1	1.614	0.0247	25.5	0.188	0.0027
$IE\kappa$ G2	4	9.2	10.148	0.1929	13.4	2.332	0.0441
dPLN	4	0.0	1.490	0.0215	0.0	0.159	0.0022
GB	5	2.9	1.717	0.0256	16.4	0.166	0.0022
GdPLN	5	1.2	1.590	0.0245	-50.2	0.228	0.0033
dP^2LN_σ	5	-20.3	1.098	0.0179	0.4	0.158	0.0022
$dP^2LN_{\mu\sigma}$	6	-47.8	4.241	0.0582	-60.4	3.314	0.0581
dP^2LN	7	-49.1	1.691	0.0233	-88.7	0.107	0.0014
dP^2LN'	6	-42.6	0.500	0.0060	-86.2	0.127	0.0016
dP^2LN''	6	-43.1	0.534	0.0063	-84.6	0.114	0.0014

- AICは、dPLNのAICとの差を表章
- 「ジニ係数の誤差」は、標本値との差
- AICは7回の調査に対する適用結果の単純平均、LRSSE及び「ジニ係数の誤差」はRMSE
- κ Gは、 κ -一般化分布 (Clementi, 2007), $l\kappa$ Gは、 κ Gに従う確率変数の逆数が従う分布
- アメリカSCFについては、期待値が無限大になるのを防ぐために、フルモデル dP^2LN の当てはめに $\sigma_L < \sigma_R$ の制約を課している

dP^2LN_σ : dP^2LN で $\mu_L = \mu_R$ と $r = \alpha/(\alpha + \beta)$ の制約を課した場合

$dP^2LN_{\mu\sigma}$: dP^2LN で $r = \alpha/(\alpha + \beta)$ の制約を課した場合

dP^2LN' : dP^2LN で $\mu_L > \mu_R$ と $\beta = 1.01$ の制約を課した場合

dP^2LN'' : dP^2LN で $\sigma_L < \sigma_R$ と $\beta = 1.01$ の制約を課した場合

各回への適用結果及び推定密度分布の比較についてはOkamoto (2014)参照

◎ dP²LN, dP²LN'の最尤パラメータ, アメリカSCF

	dP ² LN (with constraint $\sigma_L < \sigma_R$)							dP ² LN'					
	μ_L	μ_R	σ_L	σ_R	α	β	r	μ_L	μ_R	σ_L	σ_R	α	r
1992	11.5 (0.03)	10.0 (0.08)	0.33 (0.052)	0.87 (0.035)	2.1 (0.18)	1.1 (0.05)	0.36 (0.043)	11.5 (0.03)	10.1 (0.05)	0.31 (0.042)	0.87 (0.024)	2.2 (0.16)	0.34 (0.032)
1995	10.4 (0.53)	10.6 (0.25)	0.12 (0.140)	0.32 (0.283)	1.9 (0.09)	1.3 (0.11)	0.38 (0.070)	11.4 (0.03)	10.2 (0.05)	0.35 (0.026)	0.88 (0.033)	2.1 (0.13)	0.46 (0.031)
1998	11.6 (0.04)	10.0 (0.06)	0.31 (0.051)	0.80 (0.026)	1.6 (0.09)	1.1 (0.05)	0.41 (0.044)	11.6 (0.03)	10.1 (0.05)	0.29 (0.034)	0.80 (0.022)	1.7 (0.08)	0.38 (0.026)
2001	11.7 (0.04)	10.0 (0.09)	0.36 (0.046)	0.79 (0.040)	1.4 (0.10)	1.2 (0.06)	0.40 (0.056)	11.7 (0.03)	10.1 (0.04)	0.29 (0.030)	0.80 (0.018)	1.5 (0.06)	0.33 (0.031)
2004	11.5 (0.05)	10.0 (0.08)	0.52 (0.031)	0.84 (0.039)	1.5 (0.09)	1.4 (0.08)	0.51 (0.043)	11.7 (0.04)	10.2 (0.04)	0.36 (0.030)	0.83 (0.018)	1.8 (0.07)	0.30 (0.028)
2007	11.6 (0.05)	9.9 (0.08)	0.44 (0.053)	0.76 (0.037)	1.3 (0.07)	1.5 (0.37)	0.40 (0.053)	11.7 (0.01)	10.2 (0.03)	0.00 (0.094)	0.81 (0.015)	1.7 (0.07)	0.16 (0.032)
2010	11.6 (0.04)	10.0 (0.05)	0.41 (0.036)	0.73 (0.023)	1.4 (0.06)	1.4 (0.07)	0.29 (0.038)	11.7 (0.05)	10.1 (0.02)	0.27 (0.025)	0.75 (0.012)	1.7 (0.05)	0.16 (0.017)

多重補定法のために提供されている5組の所得データのうち4組目に適用した結果

括弧内の数値は標準偏差で, ブートストラップ・タイプのReplicate weightsと多重補定法により算出

1995年のdP²LNのp.d.f.は双峰(次々スライド参照)

● dP²LN, dP²LN''の最尤パラメータ, イタリアSHIW

	dP ² LN							dP ² LN''					
	μ_L	μ_R	σ_L	σ_R	α	β	r	μ_L	μ_R	σ_L	σ_R	α	r
2000	11.4	10.4	0.12	0.60	3.7	1.1	0.18	11.4	10.4	0.09	0.60	3.9	0.15
2002	10.7	9.8	0.15	0.60	4.1	1.1	0.15	10.8	9.8	0.12	0.60	4.3	0.13
2004	10.9	9.7	0.24	0.50	2.7	1.4	0.23	11.0	9.8	0.08	0.54	3.3	0.11
2006	10.9	10.2	1.30	0.59	292.4	0.9	0.03	10.9	9.9	0.14	0.54	3.5	0.11
2008	10.9 (0.08)	9.9 (0.04)	0.21 (0.060)	0.57 (0.017)	3.6 (0.38)	1.3 (0.13)	0.12 (0.039)	10.9 (0.01)	9.9 (0.03)	0.00 (0.000)	0.59 (0.015)	4.2 (0.62)	0.06 (0.014)
2010	10.9 (0.06)	9.9 (0.03)	0.19 (0.066)	0.57 (0.017)	3.7 (0.34)	1.1 (0.09)	0.18 (0.035)	10.9 (0.05)	10.0 (0.03)	0.15 (0.043)	0.57 (0.015)	3.8 (0.35)	0.15 (0.016)
2012	11.0 (0.07)	9.9 (0.04)	0.14 (0.042)	0.58 (0.019)	3.9 (0.56)	0.8 (0.12)	0.11 (0.038)	11.0 (0.06)	9.8 (0.03)	0.20 (0.061)	0.56 (0.019)	3.4 (0.33)	0.16 (0.027)

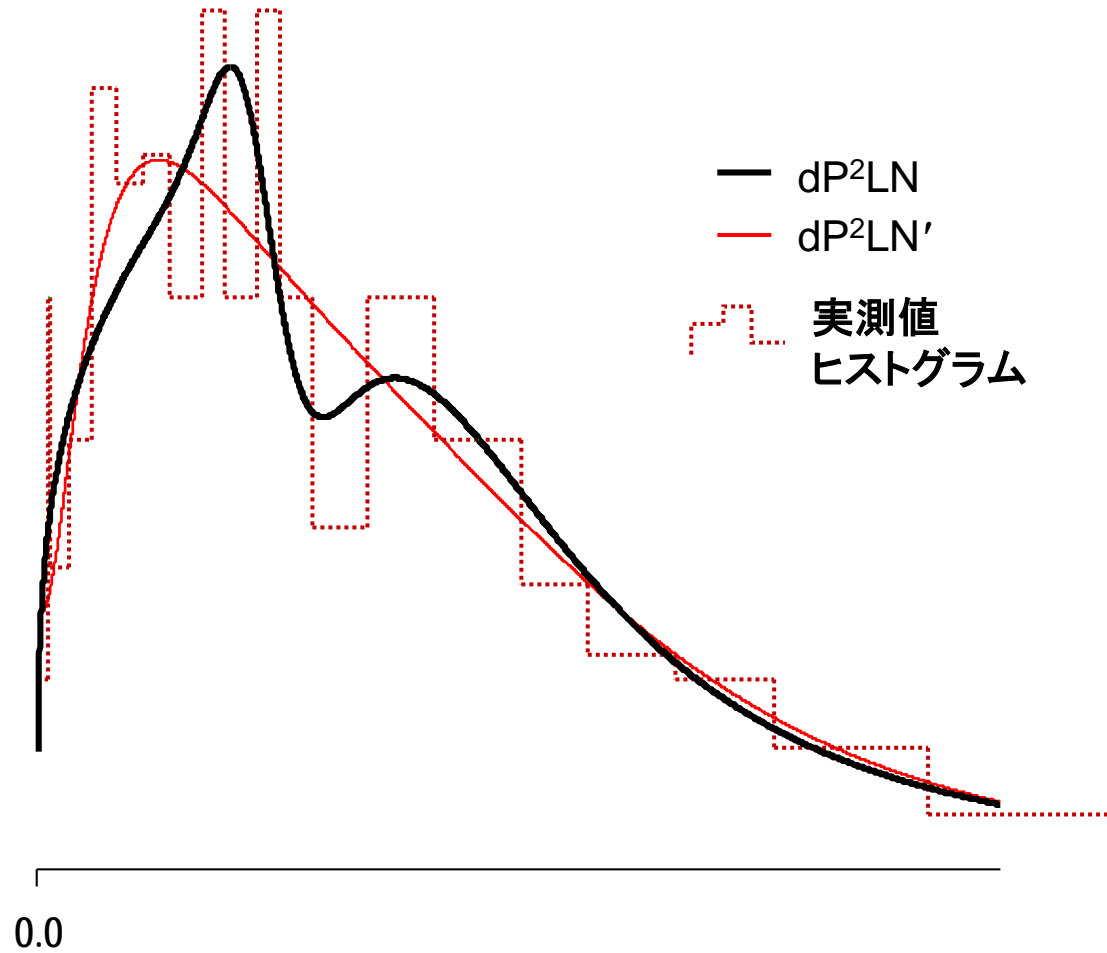
括弧内の数値は標準偏差で, (2008年調査から提供されている)ジャックナイフ・タイプのReplicate weightsにより算出

2006年のdP²LNのp.d.f.は双峰(次スライド参照)

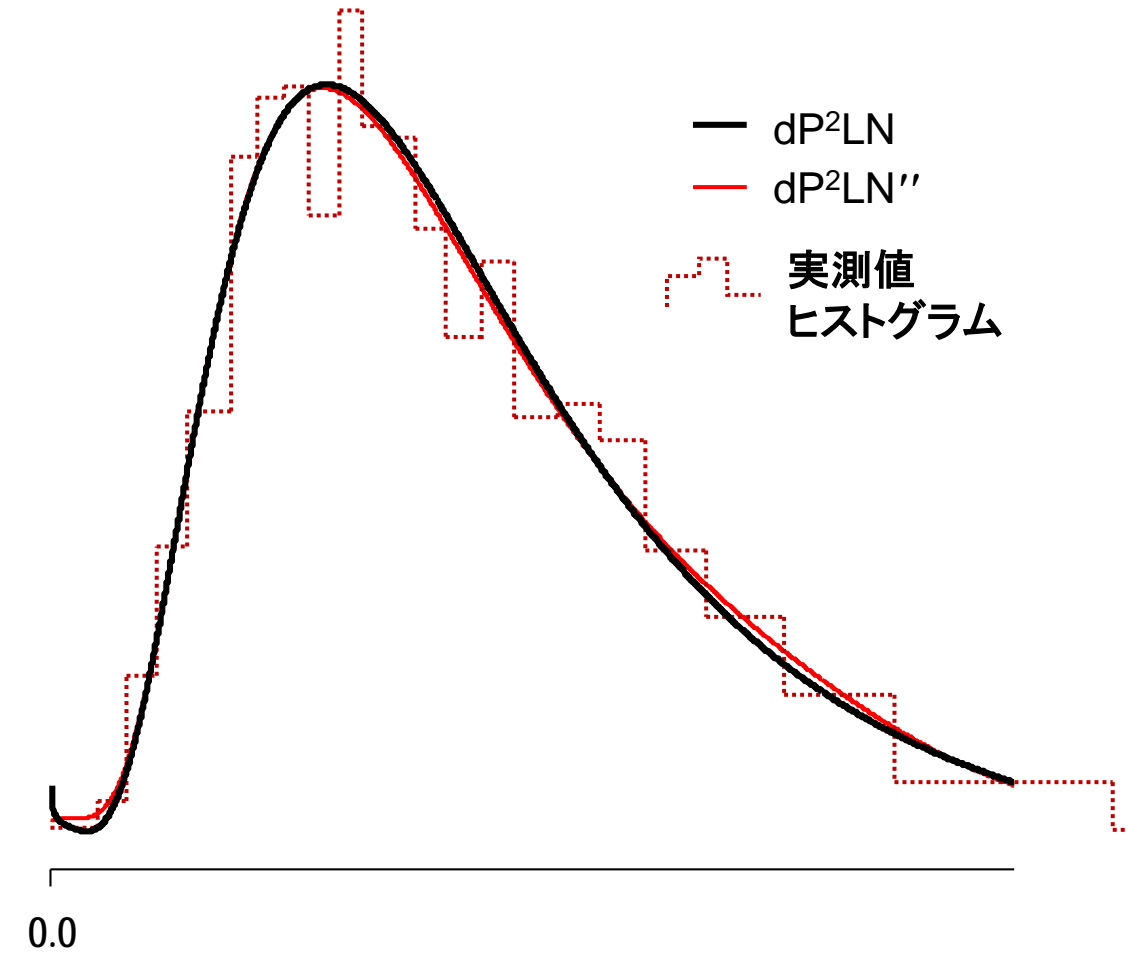


dP²LNのp.d.f.が双峰になった事例

アメリカSCF, 1995, 5組の所得データのうち4組目



イタリアSHIW, 2006





消費分布への当てはめ結果

	パラ メー タ数	AIC (dPLNのAICとの差を表章)						
		1984	1989	1994	1999	2004	2009	平均
GG	3	345.8	280.4	180.4	228.2	345.6	444.8	304.2
SM	3	1.0	-7.2	0.0	23.2	1.6	1.6	3.4
Dagum	3	-2.2	-7.2	-3.6	5.8	-5.8	0.4	-2.1
GB2	4	-3.0	-5.4	-2.2	-1.2	-3.8	1.4	-2.4
EkG1	4	-2.8	-6.8	-2.0	4.6	-0.2	3.4	-0.6
EkG2	4	-2.0	-11.6	-1.0	-2.4	-8.4	-6.6	-5.3
IEkG1	4	-3.0	-5.4	-2.2	-2.4	-4.2	2.4	-2.5
IEkG2	4	-2.0	1.4	-1.8	1.8	3.2	13.4	2.7
dPLN	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GB	5	-1.2	-3.4	-0.6	0.8	-1.8	3.4	-0.5
GdPLN	5	-3.4	-8.2	-0.8	0.6	-4.6	1.8	-2.4
dP ² LN [#]	6	-1.2	-11.6	1.4	0.2	-12.2	-11.2	-5.8
dP ² LN ^{##}	6	-1.4	-12.8	1.0	0.6	-11.2	-11.0	-5.8
dP ² LN	7	-1.8	-11.0	2.2	0.4	-10.6	-9.4	-5.0

使用データ: 全国消費実態調査2人以上の世帯の9~11月の世帯当たり消費支出

dP²LNにパラメータの制約を適切に課すと, AICがEkG2などの既存モデルを若干下回るが, 年収分布とは異なりに適合度の向上は僅か.

dP²LN[#]: $\beta = 3$ に固定; dP²LN^{##}: $r = 0.4238494$ に固定

1984年のdP²LNのp.d.f.は双峰

第2種拡張κ-一般化分布 (EκG2) (Okamoto, 2013; 岡本, 2013)

$$X \sim E\kappa G2(a, b, p, q) \quad \text{if} \quad Z = \frac{\left(\frac{X}{b}\right)^a}{\sqrt{1 + \left(\frac{X}{2b}\right)^{2a}} + \left(\frac{X}{2b}\right)^a}, \quad Z \sim \text{Beta}(p, q)$$

$$\Leftrightarrow X = b \left(\frac{Z}{1-Z} \right)^{\frac{1}{2a}}$$

p.d.f. :

$$f_{E\kappa G2}(x; a, b, p, q) = \frac{a}{bB(p, q)} \frac{z^{p-\frac{1}{a}}(1-z)^{q+\frac{1}{2a}}}{1 - \frac{1}{2}z}, \quad z = \frac{\left(\frac{x}{b}\right)^a}{\sqrt{1 + \left(\frac{x}{2b}\right)^{2a}} + \left(\frac{x}{2b}\right)^a}$$

$$X \sim E\kappa G2(a, b, p, q) \quad \Leftrightarrow \quad \left(\frac{X}{b}\right)^a = \frac{Y}{\sqrt{1+Y}} = \sqrt{Y \cdot \frac{Y}{1+Y}}, \quad Y \sim GB2(1, 1, p, q)$$

$$\frac{Y}{1+Y} \sim GB1(1, 1, p, q) = \text{Beta}(p, q)$$



結論と課題

■ 新モデル dP^2LN は、パラメータの制約条件を適切に選択することで、尤度などの頻度ベースの評価指標だけでなく、ローレンツ曲線や格差指標の精度に関しても所得分布への適合度が既存モデルよりも向上することが期待できる。

- 他国の所得分布への適用
- 効率的な当てはめ手順
- パラメータの制約条件の選択肢の確立



参 考 文 献

- Clementi, F., M. Gallegati and G. Kaniadakis (2007) “ κ -generalized statistics in personal income distribution,” *European Physical Journal B* **52**, 187–193.
- Gabaix, X. (1999) “Zipf’s law for cities: an explanation,” *Quarterly Journal of Economics* **114**, 739–767.
- McDonald, J. B. (1984) “Some generalized functions for the size distribution of income,” *Econometrica* **52**, 647–663.
- McDonald, J. B. and Y. J. Xu (1995) “A generalization of the beta distribution with applications,” *Journal of Econometrics* **66**, 133–152.
- Okamoto, M. (2013) “Extension of the κ -generalized distribution: new four-parameter models for the size distribution of income and consumption,” LIS working paper 600.
- Okamoto, M. (2014) “A flexible descriptive model for the size distribution of incomes,” *Economics Bulletin* **34**, 1600–1610.
- Reed, W.J. (2003) “The Pareto law of incomes – an explanation and an extension,” *Physica A* **319**, 579–597.
- Reed, W.J. and F. Wu (2008) “New four- and five-parameter models for income distributions,” in *Modeling Income Distributions and Lorenz Curves* by D. Chotikapanich Ed., New York: Springer, 211–223.
- 岡本政人 (2013) “所得格差と消費格差の比較,” 2013年度統計関連学会連合大会, 大阪大学.