グループデータからの所得分布の推定

西埜晴久

Chapter 1

はじめに

1.1　研究の背景と問題意識

所得分布の研究は古くからある研究であり，所得分布は，社会保障，労働，財政，マクロ，消費，経済統計など多くの領域にまたがる問題である．橘木俊詔（1998）『日本の経済格差』および，大竹文雄（2005）『日本の不平等』が，日本の経済格差の問題について大きな論争を呼びおこした．その結果，日本の不平等度が米国より高いということはないが，90 年代以降，少しずつ上昇していることにはコンセンサスができてきている．一方，*Journal of Economic Inequality* が 2003 年に発刊されるなど，世界的にも所得格差の問題は関心を集めている．不平等度と社会階層間のモビリティとの関係が重要とされている．日本では，大竹 (2005) によって格差拡大の原因として，高齢化仮説および世帯規模縮小仮説が唱えられている．

経済格差・不平等度は日本経済においても大きな論点である．そして，橘木・大竹論争では日本の経済格差が拡大したかが大きな論点だった．そこでは，ジニ係数を使って記述統計的に計算するのが一般的であった．個票データや階級別データ（グループデータ）を使ってローレンツ曲線を推定する研究は多かった．一方，本研究ではパラメトリックな分布を仮定し，扱いやすいグループデータを使って推定することを考えることにする．

1.2　研究にあたっての準備

1.2.1　日本のデータについて

本研究で利用する日本のデータについて説明する．まず，これまで所得分布を調べるために用いられている日本のデータは，おおまかに以下のように分類できる．

1. 家計調査（総務省統計局）＊本研究で用いているデータ（毎年ある）
2. 全国消費実態調査（総務省統計局）（5 年おき）
3. 所得再分配調査，国民生活基礎調査（厚生労働省）橘木 (1998) など
4. 税務統計 民間給与実態統計調査など（国税庁）Moriguchi and Saez (2008) など

本研究では使うデータは総務省統計局が出している家計調査のデータを用いる．家計調査は比較的古くからのデータが利用可能であり，また毎年公表されているために使い勝手も良いためである．一方，同じく総務省統計局が出している全国消費実態調査は発表されているのは 5 年ごとであるが，家計調査よりも標本サイズが大きいためにより精密に分布を調べるに

は有用な統計である．また，厚生労働省の発表している所得再分配調査および国民生活基礎調査も所得分布を調べるために良く利用されている統計である．ただ，所得再分配調査の当初所得の概念が特殊であるためにそのままジニ係数を算出すると高くなってしまうことが，大竹 (2005) によって指摘されている．また，さらに注意すべきことに，大竹 (2005) ではいずれの統計によっても 90 年代からジニ係数が全体的に上昇する傾向を示している．

そして，本研究では用いることはしないが，国税庁による税務統計も所得の不平等度をみるためにしばしば用いられている統計である．古くは高橋 (1955) が大蔵省主税局統計年報により明治 20 年 (1887 年) からの所得分布のデータを収集している．そして，Moriguchi and Saez (2008) は上位所得の納税者の所得が全体に占める割合という不平等度の指標を使うことで，19 世紀末からの日本，米国，フランスの不平等度の変遷を示している．それによれば，第 2 次世界大戦前は 3 国とも不平等度が大きいが，特に 1930 年代の日本の不平等度は米国やフランスよりも高いことが分かる．そして，第 2 次世界大戦後は 3 か国とも急速に不平等度は小さくなる．次に，1980 年代以降，米国の不平等度が非常に拡大してきていることが分かる．また，米国に比べると変化は小さいが日本とフランスも 1990 年代ごろから少しずつ不平等度が拡大する傾向にある．他に税務統計を使った研究として，Nirei and Souma (2007) は高額納税者公示制度（いわゆる長者番付）のデータを用いて所得分布の上の裾の分布について調べている．

1.3 さまざまな分布

次に，所得分布で用いられる分布について簡単に概観する．所得分布を表わす分布には多くの種類があるが，代表的なものに以下の分布がある．

- 2 パラメータ：対数正規分布，パレート分布
- 3 パラメータ：Singh-Maddala 分布　Singh and Maddala (1976)
- 4 パラメータ：一般化ベータ分布（第一種，第二種）McDonald(1984)

たとえば，こうした種々の分布については Kleiber and Kotz (2003) に詳しい説明がある．そして，本研究では 2 パラメータのパレート分布および対数正規分布を用いる．なお，Kleiber and Kotz (2003) によれば，第二種の一般化ベータ分布は 3 パラメータの Singh-Maddala 分布，第二種のベータ分布，Dagum 分布を含んだ分布であることが分かる．さらに，McDonald and Xu(1995) が提案しているパラメータ数の多い 5 つのパラメータをもつ一般化ベータ分布は 4 パラメータの第二種の一般化ベータ分布を包含している．

1.3.1 グループデータとその推定法

本研究で用いるグループデータについて説明しておこう．以下のデータは分位データの例である． 5 分位データの場合，20%, 40%, 60%, 80%の 4 つの分位点の値が観察されることになる．

- 家計調査 (2006)　勤労者世帯 5 分位データ
- 標本サイズ：$n = 10000$，各セル：$10000/5 = 2000$

x_1	x_2	x_3	x_4
429	570	724	938

(単位：万円)
この場合，全体の標本サイズが 10000 であることから，各グループ（セル）の標本サイズは 2000 である．また，分位データだけがグループデータではない．たとえば，各階級の幅が固定されている階級データもグループデータである．

次に，グループデータからのこれまで用いられてきた推定法を 2 種類説明する．一つは多項最尤法であり，もう一つはピアソンの最小カイ 2 乗法である．これらの二つの推定法はグループデータの推定法として伝統的に用いられてきたものである．これらは Cox and Hinkley (1979) などの多項分布のセルの確率を推定する方法を適用したものである．他には，McDonald and Ransom (1979) がある．

1. **多項最尤法**
   セルの数は $g = k + 1$．各セルの標本サイズは，$m_i = n_1$, $m_i = n_i - n_{i-1}$ and $m_g = n - n_k$ で，総計で $\sum_{i=1}^{g} m_i = n$) とする．多項分布の尤度は

$$L = n! \prod_{i=1}^{g} \frac{\pi_i^{m_i}}{m_i!}, \tag{1.1}$$

   となる．ただし，各セルの確率は

$$\begin{aligned} \pi_1 &= F(x_1), \\ \pi_i &= F(x_i) - F(x_{i-1}), \quad (i = 2, \ldots, g-1), \\ \pi_g &= 1 - F(x_{g-1}) \end{aligned}$$

   である．そこで，$\log L$ を最大化するようにパラメータを求める．

2. **ピアソンの最小カイ 2 乗法**
   これは適合度のカイ 2 乗検定を推定へと応用したものである．

$$n \prod_{i=1}^{g} \frac{\{(m_i/n) - \pi_i\}^2}{\pi_i}, \tag{1.2}$$

   これを最小化するようにパラメータを求める

なお，van Dijk and Kloek (1980) では多項最尤法と最小カイ 2 乗法は大標本では同値であることに触れられている．

1.3.2 日本のデータからの推定

最後に，本研究と同じアプローチであるパラメトリックな分布をあてはめて，日本のデータから所得分布を推定した研究を見ておこう．まずは，代表的な研究として所得再分配調査のグループデータに対して 2 つから 3 つのパラメータのパラメトリックな分布の推定を行った Atoda et. al. (1988) がある．この研究では日本のデータに対して Singh-Maddala 分布のあてはまりが良いことを示している．次に，Tachibanaki et. al. (1997) では，所得再分配調査の個票データを入手して最尤法で推定を行い，グループデータを用いた推定との比較を行っている．

1.4 本稿の構成

本論文の構成を以下に述べる．2 章のパレート分布の推定の話は西埜 et. al. (2009) によるものであるが，Koutrouvelis (1981) のパレート分布の議論をそのまま用いたものであり，部分順序統計量にもとづく推定量 (SOE) は GLS 推定量として陽表的に得ることができる．この SOE 推定量を用いて，経済格差が変化したかを検定する話である．パレート分布を確率変数を対数変換すると指数分布が得られるが，指数分布の部分順序統計量の漸近正規近似を使った推定の研究は，Sarhan et. al. (1963), Saleh and Ali (1966) などを代表として，1960 年代に盛んに行われていた．なお，Saleh and Ali (1966) では，この推定量を ABLUE(Asymptotically BLUE) であるとしている．

同じ議論を，パレート分布から対数正規分布に移して行うというアイデアが 3 章の議論である．1960 年代には分位点が陽表的に得られる指数分布を対象に研究が行われていたが，現在では，正規分布の分位関数はエクセルなどのスプレッドシートソフトでも求めることができるため，対数正規分布を考えることは有用であると考えている．さらに，現実のデータでは対数正規分布の当てはまりが良いという利点がある．なお，3 章は Nishino and Kakamu (2011) を日本語訳したものにあたる．

そして，3 章で導かれた部分順序統計量を漸近正規近似することで線形モデルとして扱い，その線形モデルに確率的ボラティリティ(SV) モデルをいれることで時系列構造を入れる話が 4 章の話になる．なお，4 章は Nishino et. al. (2012) を日本語訳したものである．最後に，5 章では今後に残された問題について論じる．

Chapter 2

パレート分布を用いた経済格差の検定

2章は西埜 et al. (2009) にもとづくものである．ここでのパレート分布の推定の話は，Johnson et. al. (1994, 20章) および Koutrouvelis (1981) に示されているものを用いたものである．この部分順序統計量を使った SOE(selected order statistics based estimator) は陽表的に得ることができる．そして，この SOE を用いて経済格差が変化したかを検定する方法を提案する．なお，Koutrouvelis (1981) はパレート分布の確率変数を対数変換すると指数分布になることを利用して，指数分布の部分順序統計量の漸近正規近似を使った推定をパレート分布へと応用していた．指数分布の部分順序統計量にもとづく推定の研究は Saleh and Ali (1966) など，1960 年代に盛んに行われていたが，Saleh and Ali (1966) では，この推定量を ABLUE(Asymptotically BLUE) としている．

第一種のパレート分布 $P(I)(\theta,\alpha)$ はパラメータが 2 つある．一つは分布の下限を表すパラメータ θ であり，もう一つは分布の形状を表すパラメータ α である．パレート分布のジニ係数 G は $G = 1/(2\alpha - 1)$ となるので，ジニ係数は形状パラメータ α だけによって決まる．

そして全体の標本がサイズ n で得られ，さらにその部分標本がサイズ k で得られて，サイズ k の部分標本が小さい順に並び替えられているとする．つまり，$1 \le n_1 < n_2 < \ldots < n_k \le n$ に対し，$\{X(n_1) \le X(n_2) \le, \cdots, \le X(n_k)\}$ の順序統計量が得られるとする．ただし，i 番目の順序統計量 $X(n_i)$ は，ある一定の比率 $p_i = n_i/n$ に対応して得られているものとする．また，計算の便宜上 $u_i = -\log(1-p_i)$ の表記を導入する．確率変数 X がパレート分布 $P(I)(\theta,\alpha)$ に従うとき，対数変換した確率変数 $Y = \log X$ は，確率密度関数

$$f(y) = \frac{1}{\sigma}\exp\left\{-\frac{(y-\mu)}{\sigma}\right\}, \qquad y > \mu,\ \sigma > 0, \tag{2.1}$$

の指数分布に従う．なお，$(\mu,\sigma) = (\log(\theta), 1/\alpha)$ とパラメータを変換した．そこで，不平等度についての情報は形状パラメータ σ に集約される．

(μ,σ) の推定量は Koutrouvelis (1981) より，

$$\hat{\mu} = \log X(n_1) - \hat{\sigma}u_1, \tag{2.2}$$

$$\hat{\sigma} = \sum_{i=1}^{k} b_i \log X(n_i) \tag{2.3}$$

となる．なお，

$$
\begin{aligned}
b_1 &= -\frac{1}{L}\frac{u_2-u_1}{e^{u_2}-e^{u_1}},\\
b_i &= \frac{1}{L}\left[\frac{u_i-u_{i-1}}{e^{u_i}-e^{u_{i-1}}}-\frac{u_{i+1}-u_i}{e^{u_{i+1}}-e^{u_i}}\right], \quad (i=2,\ldots,k-1)\\
b_k &= \frac{1}{L}\frac{u_k-u_{k-1}}{e^{u_k}-e^{u_{k-1}}},\\
L &= \sum_{i=2}^{k}\frac{(u_i-u_{i-1})^2}{e^{u_i}-e^{u_{i-1}}},
\end{aligned}
$$

である．したがって，(2.2) と (2.3) から，$(\hat{\mu},\hat{\sigma})$ は陽表的に計算できることが分かる．また，2 章ではモンテカルロ実験によってこの推定量が最尤推定量と比較して効率性がほぼ同等であることを確認した．

さらに，順序統計量に基づく推定量の漸近正規性を用いて，経済格差が異なるかを検定する検定統計量を提案した．$\hat{\sigma}$ は標本の大きさ n が十分に大きいと漸近正規性が成り立つ．そこで，σ_t, σ_s を推定した際の標本サイズを n_t, n_s とし，推定に用いた L をそれぞれ L_t, L_s とすれば，

$$
\hat{\sigma}_t \sim \mathcal{N}\left(\sigma_t, \frac{\sigma_t^2}{L_t n_t}\right), \quad \hat{\sigma}_s \sim \mathcal{N}\left(\sigma_s, \frac{\sigma_s^2}{L_s n_s}\right), \tag{2.4}
$$

と近似できる．正規分布の平均の差の検定を考えると，$\hat{\sigma}_t, \hat{\sigma}_s$ が独立に分布するとの仮定できて, 帰無仮説：$\sigma_t = \sigma_s$ の下では検定統計量 Z は, 以下のように,

$$
Z = \frac{\hat{\sigma}_t - \hat{\sigma}_s}{\sqrt{\frac{\hat{\sigma}_t^2}{L_t n_t}+\frac{\hat{\sigma}_s^2}{L_s n_s}}} \sim \mathcal{N}(0,1), \tag{2.5}
$$

となり，大標本で標準正規分布で近似できることが分かる．そこで，標準正規分布により棄却域を設定して検定を構成できる．すなわち，検定統計量 Z が棄却域に入れば帰無仮説：$\sigma_t = \sigma_s$ は棄却されることになり，有意に格差が異なっていると判断できる．さらには，最尤推定から計算できる尤度比検定により格差の変化を検定する方法をも示した．

最後に，これらの手法を用いて総務省統計局の家計調査のデータで実証分析を行い，その結果，統計的に有意な所得分布の不平等度の拡大や縮小を確認することができた．

Chapter 3

対数正規分布を用いたジニ係数のグループデータ推定と検定

3 章は Nishino and Kakamu (2011) を日本語訳したものにあたる．そして，2 章のパレート分布を対数正規分布に変えるというアイデアが 3 章の議論である．指数分布は分位点が陽表的に得られるが，現在では，正規分布の分位関数も容易に求めることができる．また現実のデータでは対数正規分布の当てはまりが良いことを考えれば，対数正規分布を考えることは有用であると考えている．

対数正規分布 $LN(\mu,\sigma)$ のジニ係数 G は $\Phi(\cdot)$ を標準正規分布の分布関数として，$G = 2\Phi\left(\sigma/\sqrt{2}\right) - 1$, と計算される．ジニ係数 G はパラメータ σ だけに依存し，$G(\sigma)$ は σ の単調増加関数である．そこで不平等度の変化に興味があれば σ だけを推定すればよいことになる．

以下でグループデータからの部分順序統計量に基づいて得られる推定量を selected order statistics にもとづく推定量 (SOE) と呼ぶ．順序統計量の漸近理論 (たとえば, Mosteller, 1946, Theorem 2, David and Nagaraja, 2003, Theorem 10.3) によって $\lim_{n\to\infty} n_i/n = p_i$ を満たすとき，部分順序統計量 $(\log X(n_1), \log X(n_2), \ldots, \log X(n_k))'$ の同時分布が $n \to \infty$ において平均 $(\mu+\sigma u_1, \mu+\sigma u_2, \ldots, \mu+\sigma u_k)'$，分散共分散行列 $(\sigma^2/n)\mathbf{W}$ の多変量正規分布になる．分散共分散行列 $\mathbf{W}$ の (i,j) 成分は

$$
\begin{aligned}
W_{ii} &= \frac{p_i(1-p_i)}{\varphi(\Phi^{-1}(p_i))^2} \\
W_{ij} &= W_{ji} = \frac{p_i(1-p_j)}{\varphi(\Phi^{-1}(p_i))\varphi(\Phi^{-1}(p_j))} \quad (i<j), \quad i,j = 1,2,\ldots,k, \qquad (3.1)
\end{aligned}
$$

であり，$\varphi(x)$ は標準正規分布の確率密度関数を表わす.

$\mathbf{y} = (\log X(n_1), \log X(n_2), \ldots, \log X(n_k))'$ の同時分布が漸近的に多変量正規分布で近似できることより，

$$
\log X(n_i) = \mu + \sigma u_i + \varepsilon_i, \quad (i = 1, \ldots, k), \qquad (3.2)
$$

と書ける．ただし，$(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)'$ は近似的に平均 $(0,0,\ldots,0)'$，分散共分散行列 $(\sigma^2/n)\mathbf{W}$ の多変量正規確率ベクトルである．したがって，(3.2) は $(\sigma^2/n)\mathbf{W}$ の分散の誤差項を持つ線形回帰モデルとみなすことができる．$\mathbf{W}$ が既知であるので，一般化最小 2 乗法 (GLS) によって，パラメータ (μ,σ) を推定することができる．被説明変数ベクトルを $\mathbf{y}$ として，

$= (\log X(n_1), \log X(n_2), \ldots, \log X(n_k))'$ と書き，説明変数 $\mathbf{Z}$ を

$$\mathbf{Z} = \begin{pmatrix} 1 & u_1 \\ 1 & u_2 \\ \vdots & \vdots \\ 1 & u_k \end{pmatrix}.$$

とし，$\gamma = (\mu, \sigma)'$ とすると，線形回帰モデルは

$$\mathbf{y} = \mathbf{Z}\gamma + \boldsymbol{\epsilon}, \tag{3.3}$$

と書ける．ただし，$\boldsymbol{\epsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)'$ とする．すると $\hat{\gamma}$ の GLS 推定量は,

$$\hat{\gamma} = \left(\mathbf{Z}'\mathbf{W}^{-1}\mathbf{Z}\right)^{-1} \mathbf{Z}'\mathbf{W}^{-1}\mathbf{y}, \tag{3.4}$$

となる．$\hat{\gamma}$ の分散共分散行列は

$$Var(\hat{\gamma}) = \frac{\sigma^2}{n} \left(\mathbf{Z}'\mathbf{W}^{-1}\mathbf{Z}\right)^{-1}. \tag{3.5}$$

となる．$\hat{\gamma} = (\hat{\mu}, \hat{\sigma})'$ は明示的に得られ，これを SOE とする．すなわち行列計算だけで推定量を求められる．

さらに，経済格差の変化を検定するために，(3.4) の GLS 推定量である SOE の漸近正規性を用いた検定を提案する．ジニ係数の推定量 $\hat{G}$ は $\hat{\sigma} > 0$ の単調非減少関数であるので，ジニ係数の変化の検定は，σ の変化の検定と同等である．不平等度の変化を検定するために，σ_t と σ_s の間の平均値の差の検定 z-test を考える．

２つのパラメータ σ_t と σ_s について, 帰無仮説, 対立仮説を

$$H_0 : \sigma_t = \sigma_s$$
$$H_1 : \sigma_t \neq \sigma_s$$

と設定する．つまり，帰無仮説では格差の程度は変わらないであり，対立仮説では格差の程度は異なっているである．

モンテカルロ実験により標本サイズ n が十分大きい時に $\hat{\sigma}$ は近似的に正規分布にしたがいことが確かめられている．その漸近分散は (3.5) で得られる．σ_t, σ_s を推定した際の標本サイズを n_t, n_s とする．$\hat{\sigma}_t$ and $\hat{\sigma_s}$ の漸近分散は (3.5) より，$Var(\hat{\sigma_t})$ and $Var(\hat{\sigma_s})$ である．つまり，

$$\hat{\sigma}_t \ \sim \ \mathcal{N}\left(\sigma_t, Var(\hat{\sigma_t})\right), \quad \hat{\sigma}_s \ \sim \ \mathcal{N}\left(\sigma_s, Var(\hat{\sigma_s})\right). \tag{3.6}$$

である．$\hat{\sigma}_t$ と $\hat{\sigma}_s$ が独立に分布していることを仮定することで，正規標本の平均値の差の検定を考えることができる．つまり，$\hat{\sigma}_t - \hat{\sigma}_s$ の分布は

$$\hat{\sigma}_t - \hat{\sigma}_s \ \sim \ \mathcal{N}\left(\sigma_t - \sigma_s,\ Var(\hat{\sigma_t}) + Var(\hat{\sigma_s})\right). \tag{3.7}$$

と近似できる．

帰無仮説：$\sigma_t = \sigma_s$ の下では検定統計量 Z は大標本で標準正規分布で近似できる．

$$Z = \frac{\hat{\sigma}_t - \hat{\sigma}_s}{\sqrt{Var(\hat{\sigma_t}) + Var(\hat{\sigma_s})}} \ \sim \ \mathcal{N}(0, 1). \tag{3.8}$$

両側検定の棄却域は，5%有意水準で $|Z| > 1.96$，1%有意水準で $|Z| > 2.576$ と定められる．検定統計量 Z が棄却域に入れば，帰無仮説：$\sigma_t = \sigma_s$ は棄却される．つまり，不平等度の変化が統計的に有意であると判断される．片側検定も同様に考えられるだろう．

最後にこれらの推定と検定を日本経済のデータに対して応用する．

Chapter 4

対数正規SVモデルによる持続的な所得不平等度のベイズ推定

4 章は Nishino et. al. (2012) を日本語訳したものにあたる．そして，3 章で導かれた部分順序統計量を漸近正規近似することで得られた線形モデルに確率的ボラティリティ(SV) モデルをいれることで時系列構造を入れたモデルが 4 章の話になる．この対数正規分布にもとづく SV モデルはジニ係数を含む不平等度を含むモデルであり，不平等度の持続性を考慮したモデルになる．

まず，先に述べた (3.2) は線形回帰モデルとみなすことができ，誤差分散行列 $(\sigma^2/n)W$ は既知である．そこで，ウェイト行列 W が既知なので一般化最小 2 乗法を用いて推定できる．所得不平等度の持続性を把握するために，部分順序統計量にもとづく線形回帰モデル (3.2) を動学モデルへと拡張する．これは以下の SV モデルで表される．

$$
\begin{aligned}
y_{i,t} &= \mu_t + \sigma_t u_i + \sigma_t \eta_{i,t}, && (4.1)\\
\sigma_t &= \exp(h_t/2) && (4.2)\\
h_t &= \omega + \phi(h_{t-1} - \omega) + \xi_t, && (4.3)
\end{aligned}
$$

ただし，

$$
\begin{aligned}
&y_{i,t} = \ln(X(n_i)),\\
&\boldsymbol{\eta}_t = (\eta_{1,t}, \eta_{2,t}, \ldots, \eta_{k,t})' \sim \mathcal{MVN}(\mathbf{0}, W/n),\\
&\xi_t \sim \mathcal{N}(\mathbf{0}, \sigma_\xi^2).
\end{aligned}
$$

である．$X(n_i)$ が対数正規分布に従う場合に，$u_i = \Phi^{-1}(p_i)$ とし，W_{ij} は (3.1) で定義され，$\Phi(\cdot)$ は標準正規分布の累積分布関数とする．(4.1) の μ_t は SV モデルでは各 t で別々に推定され，不平等度に影響する σ_t は (4.2) により時変構造をもつ．σ_t からジニ係数を得ることができる．(4.1), (4.2) および (4.3) による SV モデルをマルコフ連鎖モンテカルロ (MCMC) 法によって推定する．

SV モデルの推定では総務省統計局によって公表されている家計調査のデータを用いる．1971 年から 2007 年までの暦年の年次データである．２つの種類の家計を用いている．勤労者世帯家計と二人以上世帯家計である．家計調査における世帯の種類について触れておこう．全世帯は全ての種類の世帯を含んでいる．これは，二人以上世帯と単身者世帯から成っている．二人以上世帯は，勤労者世帯と農業，林業，漁業などの非勤労者世帯から成ってい

Table 4.1: Marginal Likelihood (Model Comparison)

LogNormal	LogNormal-SV
勤労者世帯	
-7.7845	247.46
二人以上世帯	
-84.261	206.91

Table 4.2: Marginal Likelihood (with exogenous variables)

Constant	GDP	CPI	Nikkei	elderly ratio
勤労者世帯				
247.46	236.14	238.07	237.02	236.24
二人以上世帯				
206.91	190.07	191.17	191.40	190.82

る．1995 年以降に単身者世帯のデータが利用可能になったことから，1995 年以前を含む長期のデータをとるために二人以上世帯および勤労者世帯のデータを用いることとする．

全体の標本サイズが $n = 10000$ とし，グループデータとして $n_1 = 2000$, $n_2 = 4000$, $n_3 = 6000$ および $n_4 = 8000$ が記録されているとする．つまり，$p_1 = 0.2$, $p_2 = 0.4$, $p_3 = 0.6$, および $p_4 = 0.8$ の分位データが記録されている．標本サイズ $n = 10000$ は漸近的な近似を適用するには十分大きい．SV モデルのパラメータ ω, ϕ, σ^2_ξ および $\{\mu_t; 1971 \leq t \leq 2007\}$ を MCMC 法を用いて推定する．なお，burn-in に $300,000$ 回を使った後に，繰り返し標本に $300,000$ 回を使う．その繰り返し標本の中から 10 個おきにパラメータの事後標本をサンプルする．これらの計算は Ox 5.1 (see Doornik (2008)) を用いた．

はじめにモデル比較を行う．このために SV モデルと SV を含まないモデルを MCMC 法で推定する．SV を含まないモデルは (4.1) のモデルの推定を含み，σ_t は σ_{t-1} と独立に推定する．

Table 4.1 に，勤労者世帯と二人以上世帯のデータに対し，対数正規分布を仮定したときの SV モデルと SV を含まないモデルを推定した周辺尤度が計算されている．Table 4.1 からは両方の種類の世帯において SV モデルのほうが SV を含まないモデルよりも良いということがわかる．そこで，このことは SV モデルを採用する根拠となる．そして，周辺尤度の差は時折大きな値をとる．これは，$\mathbf{y}$ の漸近分散が (**??**) より $\sigma^2 W/n$ であるので，標本サイズ $n = 10000$ のときは，小さくなるためである．また，勤労者世帯の周辺尤度のほうが二人以上世帯の周辺尤度よりも大きい．このことは勤労者世帯のデータほうがモデルの当てはまりが良いことを示しているだろう．

そして，不平等度に対する外生変数の効果を評価する．外生変数を SV モデルに組み込む．つまり，SV モデルの (4.3) 式を

$$h_t = \boldsymbol{\gamma}'\mathbf{z}_t + \phi(h_{t-1} - \boldsymbol{\gamma}'\mathbf{z}_{t-1}) + \xi_t, \tag{4.4}$$

に置き換えるのである．この時，$\mathbf{z}_t = (1, z_t)'$ は外生変数である．外生変数としては，GDP 成長率，CPI 変化率，日経平均のリターンおよび高齢化率をとった．なお，CPI 変化率は日本の消費者物価指数の変化率を表し，高齢化率は日本における 65 歳以上の全人口に占める

Table 4.3: 推定結果

	Mean	Stdev	95%L	Median	95%U
	勤労者世帯				
ω	-1.691	0.261	-1.966	-1.690	-1.424
ϕ	0.914	0.0649	0.758	0.928	0.996
σ_ξ^2	0.00149	0.000448	0.000832	0.00142	0.00256
μ_{71}	4.876	0.00432	4.868	4.876	4.885
μ_{85}	6.227	0.00460	6.218	6.227	6.236
μ_{00}	6.535	0.00486	6.525	6.535	6.544
	二人以上世帯				
ω	-1.360	0.273	-1.824	-1.361	-0.902
ϕ	0.960	0.0337	0.873	0.969	0.999
σ_ξ^2	0.00141	0.000430	0.000779	0.00134	0.00245
μ_{71}	4.844	0.00500	4.835	4.844	4.854
μ_{85}	6.159	0.00533	6.148	6.159	6.169
μ_{00}	6.415	0.00582	6.404	6.415	6.426

割合である．データは 1971 年から 2007 年までであり，外生変数の事前分布には正規分布を用いた．

Table 4.2 にはこれらの SV モデルの周辺尤度を載せている．定数項だけのモデルが周辺尤度が一番大きい．つまり，ここでは外生変数は不平等度をうまく説明していないことになる．Table 4.2 の数値を細かく見てみると，CPI 変化率と日経平均リターンは他の外生変数（GDP 成長率および高齢化率）よりも比較的良く不平等度を説明しているように見える．

したがって，Table 4.1 と Table 4.2 の結果から判断すると，外生変数を含まない SV モデルが日本経済の不平等度の変化をモデル化するには適切だということになる．すなわち，SV のような時系列構造をジニ係数で表現される不平等度を含む経済モデルに取り入れることにはもっともらしい根拠があるが，外生変数を経済の不平等度に含むことには、根拠を与えられないということである．不平等度と他の外生変数を共に含むより良い動学的なモデルの開発は今後の研究課題であると期待している．

Table 4.3 は勤労者世帯と二人以上世帯のデータに対し対数正規分布を使って SV モデルを推定した時の ω, ϕ, σ_ξ^2 および $\{\mu_t; 1971 \le t \le 2007\}$ の推定結果を与えている．これらの推定されたモデルは外生変数を含んでいない．Table 4.3 では，$\{\mu_t\}$ の中で特に 1971 年, 1985 年, 2000 年の μ_t の値である μ_{71}, μ_{85}, μ_{00} を掲げる．また，Table 4.3 はパラメータの事後平均，標準誤差，95%信用区間とメディアンを載せている．

Table 4.3 は勤労者世帯と二人以上世帯のデータに対し対数正規分布を使って SV モデルを推定した時の ω, ϕ, σ_ξ^2 および $\{\mu_t; 1971 \le t \le 2007\}$ の推定結果を与えている．これらの推定されたモデルは外生変数を含んでいない．Table 4.3 では，$\{\mu_t\}$ の中で特に 1971 年, 1985 年, 2000 年の μ_t の値である μ_{71}, μ_{85}, μ_{00} を掲げる．また，Table 4.3 はパラメータの事後平均，標準誤差，95%信用区間とメディアンを載せている．

Figure 4.1 は推定された σ_t の値で計算された勤労者世帯のデータのジニ係数の変化のグラフをプロットしている．Figure 4.2 は同じように二人以上世帯のデータの時のグラフプロットしている．これらのグラフにはジニ係数の 95%信用区間も描かれている．

これらのグラフからジニ係数の時系列にはおおまかに 4 つの山があることがわかる．第

一は，石油ショックの 1974 年ごろ，第二には，バブル経済の 1980 年代後半，第三にはドットコムバブルの 2000 年周辺，そして第四は 2005 年ごろである．しかしながら，Figure 4.1 と Figure 4.2 には違いがある．たとえば，90 年代初頭の格差の縮小は Figure 4.1 では明瞭であるが，Figure 4.2 ではそうではない．これらの差異は勤労者世帯と二人以上世帯の性質の違いに由来しているものと言えるだろう．

本研究では日本の不平等度を調べるために簡単なパラメトリックモデルである対数正規分布を仮定して，グループデータからジニ係数を推定した．部分順序統計量の漸近定理を用いることで，パラメータを推定し，近似された正規線形モデルを動学モデルへと拡張した．その動学モデルは SV モデルで表現され，MCMC 法で推定した．このようにしてジニ係数を含む不平等度の動学的なモデルを構築することができた．

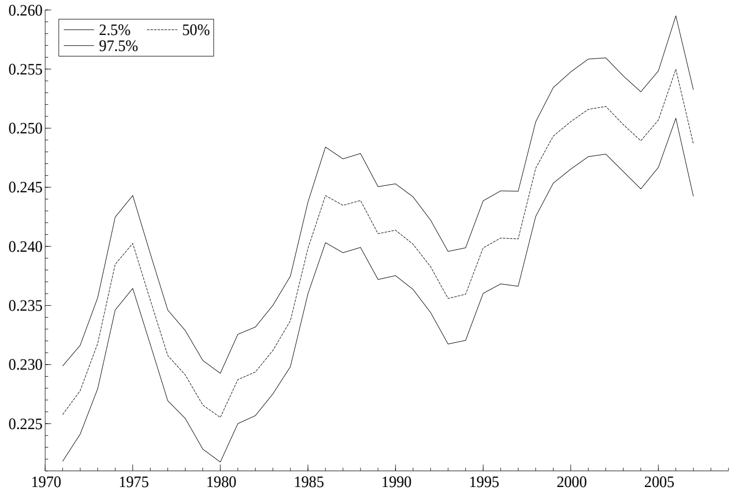

Figure 4.1: ジニ係数（勤労者世帯）

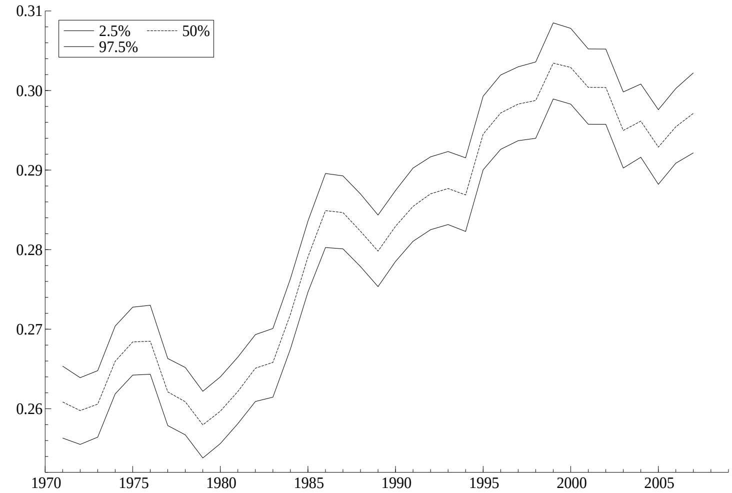

Figure 4.2: ジニ係数（二人以上世帯）

Chapter 5

今後の課題

本研究の今後の課題として，以下のようなことを考えている．

第一に，グループデータを部分順序統計量とみなし，その同時密度を尤度として MCMC によるベイズ推定を行う．Chotikapanich and Griffiths (2000) においてはグループデータに多項分布をあてはめた尤度に対し，ベイズ推定を行っているが，この尤度は部分順序統計量の同時密度を尤度としたものとは異なっている．このことは現在研究中のテーマである．

第二に，パラメータの数が 3 つ以上の分布の場合が興味深いと考えている．本研究では，パレート分布と対数正規分布といったパラメータ数が 2 つの分布を中心に扱っていたが，3 つ以上のパラメータを持つより複雑な分布の推定を考えるのは興味深いと考えている．現在，部分順序統計量の同時密度を尤度として推定する方法で研究中である．

第三に，本研究ではグループデータの境界の値だけを使って推定を行っていた．しかし，実際には各グループの平均値といった局所モーメントの情報が利用できることがある．たとえば，Hitomi et. al. (2008) では局所モーメントの情報も使った推定量を提案している．こうした局所モーメントの情報を用いた推定も今後の研究課題として考えている．

第四に，さらには，経済の不平等および所得格差はどのように生じるのかといった問題にも踏み込んでいくことを考えている．4 章で用いた時系列モデルでは外生変数を入れて不平等度の時系列モデルを説明することはできなかったが，異なったモデリングにより不平等度の原因をも含めたモデルを考えることができるかもしれない．また，不平等度を所得階層，地域あるいは職業などで分解することで，経済格差の変動のメカニズムに迫ることができるかもしれない．いずれにせよ，今後に追究すべき問題であると考えている．

Bibliography

[1] Atoda, N., T. Suruga and T. Tachibanaki (1988). "Statistical Inference of Functional Form for Income Distribution," *Economic Studies Quarterly*, **39**, 14–40.

[2] Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics*, Chapman and Hall, London.

[3] Chotikapanich D. and W. E. Griffiths (2000). "Posterior Distributions for the Gini Coefficient using Grouped Data," *Australian and New Zealand Journal of Statistics*, **4**, 901–910.

[4] Hitomi, K., Q.-F. Liu, Y. Nishiyama, and N. Sueishi (2008). "Efficient Estimation and Model Selection for Grouped Data with Local Moments," *Journal of Japan Statistical Society*, **38**, 131–143.

[5] David, H.A. and H.N. Nagaraja (2003). *Order Statistics* (3rd. ed.), Wiley, New York.

[6] Doornik, J.A. (2008). *Ox: Object Oriented Matrix Programming Language*, Timberlake Consultants Press, London.

[7] Johnson, N. L., S. Kotz and N. Balakrishnan (1994). *Continuous Univariate Distributions* (2nd ed.), **vol. 1**, Wiley, New York.

[8] Kleiber, C and S. Kotz (2003). *Statistical Size Distributions in Economics and ActuarialSciences*, Wiley, New York.

[9] Koutrouvelis, I. A. (1981). "Large-Sample Quantile Estimation in Pareto Laws," *Communications in Statistics-Theory and Methods, A*, **10**, 189–201.

[10] McDonald, J. B. (1984). "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, **52**, 647–663.

[11] McDonald, J. B. and M. R. Ransom (1979). "Functional Forms, Estimation Techniques and the Distribution of Income," *Econometrica*, **47**, 1513–1525.

[12] McDonald, J. B. and Y. J. Xu (1995). "A generalization of the beta distribution with applications," *Journal of Econometrics*, **66**, 133–152.

[13] Moriguchi, C. and E. Saez (2008). "The Evolution of Income Concentration in Japan, 1886-2005: Evidence from Income Tax Statistics," *The Review of Economics and Statistics*, **90**, 713–734.

[14] Mosteller, F. (1946). "On Some Useful "Inefficient" Statistics," *Annals of Mathematical Statisitcs*, **17**, 377–408.

[15] Nirei, M. and W. Souma (2007). "A Two Factor Model of Income Distribution Dynamics," *Review of Income and Wealth*, **53**, 440–459.

[16] Nishino, H. and K. Kakamu (2011). "Grouped data estimation and testing of Gini coefficients using lognormal distributions," *Sankhyā B*, **73**, 193–210.

[17] Nishino, H., K. Kakamu and T. Oga (2012). "Bayesian estimation of Persistent Income Inequality by Lognormal Stochastic Volatility model", *Journal of Income Distribution*, **21**, 88–101.

[18] Saleh, A. K. Md. E. and M. M. Ali (1966). "Asymptotic Optimum Quantiles for the Estimation of the Parameters of the Negative Exponential Distribution," *Annals of Mathematical Statisitcs*, **37**, 143–151.

[19] Sarhan, A. E., B. G. Greenberg and J. Ogawa (1963). "Simplified Estimates for the Exponential Distribution," *Annals of Mathematical Statistics*, **34**, 102–116.

[20] Singh, S. K. and G. S. Maddala (1976). "A Function for Size Distribution of Incomes," *Econometrica*, **44**, 963–970.

[21] Tachibanaki, T., T. Suruga and N. Atoda (1997). "Estimation of Income Distribution Parameters for Individual Observations by Maximum Likelihood Method," *Jornal of the Japan Statistical Society*, **27**, 191–203.

[22] van Dijk, H. K. and T. Kloek (1980). "Inferential Procedures in Stable Distributions for Class Frequency Data on Incomes," *Econometrica*, **48**, 1139–1148.

[23] 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, **38**, 151–164.

[24] 大竹文雄 (2005). 『日本の不平等』 日本経済新聞社.

[25] 橘木俊昭 (1998). 『日本の経済格差』 岩波書店.

[26] 高橋長太郎 (1955). 『所得分布の変動様式』 岩波書店.