

鹿児島県の各経済圏を対象とする年齢層別の所得分布に関する定量分析

福井 昭吾*

Abstract

This paper attempts to perform empirical analysis of the income distribution by age groups of different economic areas in Kagoshima. The author estimates a joint density function of family income and the age of household head for each area with GMM. And then, we derive income distributions by the age groups from the joint distribution. The result shows that, in almost all areas, householders over 60 years old have less income and are subjected to larger income inequality than the younger ones. This study also shows that the further an area is from an urban region, the larger the income inequality of each age group within the area becomes.

1 はじめに

地方を対象とする年齢層別の所得分布の推定は、利用可能なデータが限定されているために、現状では極めて大きな困難を伴う。一般に、年齢層別に分かれた集団を対象に所得分布の推定や所得格差の測定などを行う場合は、個人または世帯ごとの所得と年齢を記録したマイクロデータ、あるいは所得と年齢の両方の階層ごとにマイクロデータを分類した分割表を用いる。日本全体を対象に分析を行うならば、世帯ごとのマイクロデータおよび分割表を比較的容易に入手することができる。しかし、国内諸地域を対象とするこれらのデータについて、その利用可能性は極めて低い。近年、公的機関等によるマイクロデータ利用の提供が進んでいる。しかし、所得に関するマイクロデータについて、個人情報保護の観点から、都道府県別あるいはそれより小さい地域のマイクロデータは現時点で提供の対象とはなっていない。一方、所得と年齢の分割表については、都道府県を対象とするものならばオーダーメイド集計を通じて利用可能な場合がある。しかし、マイクロデータと同様、それよりも小さい地域を対象とする分割表は現在のところ提供されていない。

本研究では、鹿児島県内の各経済圏を対象に、マイクロデータおよび分割表の代替として所得および年齢の階層別データを用いて、所得と年齢の同時分布について GMM 推定を試みる。その後、推定した同時分布に基づいて、各経済圏における若年層・高齢層の所得の現状を明らかにする。具体的には、総務省「平成 21 年度全国消費実態調査」における鹿児島県の各経済圏について、世帯年間収入の階層別データ（度数・階層別平均所得・階層別平均年齢）と、世帯主年齢の階層別データ（度数・階層別平均所得）

キーワード：鹿児島，所得分布，所得格差，コピュラ，GMM，階層別データ

* 本学経済学部准教授

を用いて、所得（世帯年間収入）と年齢（世帯主年齢）の同時分布を推定する。その後、推定結果から60歳以下世帯・60歳より上の世帯・40歳以下世帯の所得分布を経済圏ごとに導出し、それらの平均所得・標準偏差・不平等尺度などを計算する。

階層別データからそれらデータ間の関連を分析する手法は、これまでも数多く考案されてきた。特に、階層別データから分割表を推定する手法を提示した研究として、Haitovsky (1973), Farebrother (1979), Fukushima and Hatanaka (1991), Diaconis and Sturmfels (1998), 竹村・青木 (2006) などある。しかし、Haitovsky (1973) では二つの階層別データについて強い正の相関を仮定しており、Farebrother (1979), Diaconis and Sturmfels (1998), 竹村・青木 (2006) では二つの階層別データについてそれらの確率変数は互いに独立であると仮定している。したがって、二つの階層別データ間の関連がこのような仮定を満たさない場合、これらの方法を用いることは難しい。

本研究では、福井 (2015) で開発した手法を使い、同時分布を推定する。この方法では、同時分布をモデル化した後、階層別データに基づいてモーメント式を構築し、GMMにより同時分布のパラメータを推定する。そのため、階層別データのみ利用可能な状況でも、その度数・平均などについてのモーメント式を構築することにより同時分布の推定が可能となる。また、先行する手法の多くは、二つのデータが強い正の相関をもつことや互いに独立であることを仮定しているが、この手法では、同時分布のモデル化において、データ間の関連を分析者が柔軟に設定することができる。

2 GMM 推定の方法

本研究では福井 (2015) で提示した推定方法を一部修正し、所得と年齢の同時分布を推定する^{*1}。この方法は、同時分布をモデル化した後、その同時分布に基づいて度数・階層平均・条件付平均のモーメント式を設定し、GMM推定によって同時分布のパラメータを推定する。この節では、その推定方法を説明していく。

説明を進める前に、利用する階層別データの定義をしておこう。所得に関して、第*i*階層の範囲 $I_i^X = [L_i^X, U_i^X)$ ($i = 1, \dots, m$)、相対度数 R_i^X ($i = 1, \dots, m$)、階層平均 \bar{X}_i ($i = 1, \dots, m$)、および平均年齢 \bar{Y}_i ($i = 1, \dots, m$) が与えられている。また、年齢に関しては、第*j*階層の範囲 $I_j^Y = [L_j^Y, U_j^Y)$ ($j = 1, \dots, n$)、相対度数 R_j^Y ($j = 1, \dots, n$)、および平均所得 \bar{X}_j ($j = 1, \dots, n$) が与えられているとする。

所得と年齢をそれぞれ確率変数 (x, y) 、それらの同時分布を $P(x, y)$ として、所得の相対度数・所得の階層平均・各年齢階層の平均所得・各所得階層の平均年齢についてモーメント式を作成する。所得の相対度数のモーメント式は

$$E[1_i(x) - p_i] = \int_0^\infty \{1_i(x) - p_i\} dP(x), \quad (1)$$

所得の階層平均のモーメント式は

$$E[x - \mu_i] = \int_{I_i^X} (x - \mu_i) dP(x|x \in I_i^X), \quad (2)$$

各年齢階層の平均所得のモーメント式は

$$E[x - \phi_j] = \int_0^\infty (x - \phi_j) dP(x|y \in I_j^Y), \quad (3)$$

^{*1} 主な修正部分は、所得の階層平均についてのモーメント式を加えたこと、および、コンピュータの分割の仕方を変えたことである。

各所得階層の平均年齢のモーメント式は

$$E[y - \psi_i] = \int_0^{\infty} (y - \psi_i) dP(y|x \in I_i^X) \quad (4)$$

となる。ただし、 p_i は所得の相対度数の母数、 μ_i は各所得階層における所得の母平均、 ϕ_j は各年齢階層における所得の母平均、 ψ_i は各所得階層における年齢の母平均である。また、

$$1_i(x) = \begin{cases} 1 & x \in I_i^X \\ 0 & \text{otherwise} \end{cases}$$

である*2。

式 (1) から式 (4) で、期待値を計算する際の積分の領域、および確率測度（確率分布）が互いに異なっている。したがって、これらのモーメント式に基づいて GMM 推定を行うことは難しい。そこで、Radon-Nikodym の定理を用いて、これらすべての式で、積分の領域が $[0, \infty) \times [0, \infty)$ となり、また確率測度が $P(x, y)$ となるよう変形を行う。

式 (1), 式 (3), および式 (4) について、その変形の方法は福井 (2015) に示している。結果として、変形後のモーメント式はそれぞれ

$$E[1_i(x) - p_i] = \int_{[0, \infty) \times [0, \infty)} \{1_i(x) - p_i\} dP(x, y) \quad (5)$$

$$E[y - \psi_i] = \int_{[0, \infty) \times [0, \infty)} \left\{ (y - \psi_i) \frac{1_i(x)}{P(x \in I_i^X)} \right\} dP(x, y) \quad (6)$$

$$E[x - \phi_j] = \int_{[0, \infty) \times [0, \infty)} \left\{ (x - \phi_j) \frac{1_j(y)}{P(y \in I_j^Y)} \right\} dP(x, y) \quad (7)$$

と変形できる。ただし、

$$1_j(y) = \begin{cases} 1 & y \in I_j^Y \\ 0 & \text{otherwise} \end{cases}$$

である。

式 (2) については、以下の通り変形することができる。いま、確率測度 $P_1^i(x, y)$ を

$$P_1^i((x, y) \in (A \times B)) = \frac{P((x, y) \in ((A \cap I_i^X) \times B))}{P(x \in I_i^X)}$$

と置く。この式から、確率測度 $P_1^i(x, y)$ は確率測度 $P(x, y)$ に対して絶対連続である。したがって、Radon-Nikodym の定理により

$$\begin{aligned} P_1^i((x, y) \in (A \times B)) &= \int_{A \times B} dP_1^i(x, y) \\ &= \int_{A \times B} h_1^i(x, y) dP(x, y) \end{aligned} \quad (8)$$

*2 次節で説明するように、本研究で用いる同時分布では、年齢の周辺分布の形状はパラメータの値によらず常に固定される。したがって、年齢の度数・年齢の階層平均についてのモーメント式は推定に寄与しない。そのため、本稿で説明する推定方法では、年齢の相対度数・年齢の階層平均のモーメント式を含めていない。

を満たす関数 $h_1^i(x, y)$ がただ一つ存在する。実際には, $h_1^i(x, y) = 1_i(x)/P(x \in I_i^X)$ である。式 (8) から, $P(x, y)$ に関して積分可能な任意の関数 $g(x, y)$ について,

$$\int_{A \times B} g(x, y) dP_1^i(x, y) = \int_{A \times B} g(x, y) \frac{1_i(x)}{P(x \in I_i^X)} dP(x, y) \quad (9)$$

となる。ここで, $g(x, y) = x - \mu_i, A = I_i^X, B = [0, \infty)$ とすれば,

$$\begin{aligned} \int_{I_i^X \times [0, \infty)} (x - \mu_i) dP_1^i(x, y) &= \int_{I_i^X \times [0, \infty)} (x - \mu_i) \frac{1_i(x)}{P(x \in I_i^X)} dP(x, y) \\ &= \int_{[0, \infty) \times [0, \infty)} (x - \mu_i) \frac{1_i(x)}{P(x \in I_i^X)} dP(x, y) \end{aligned}$$

となる。一方, $P(x|x \in I_i^X) = P_1^i(x, y \in [0, \infty))$ であるから

$$\int_{I_i^X \times [0, \infty)} (x - \mu_i) dP_1^i(x, y) = \int_{I_i^X} (x - \mu_i) dP(x|x \in I_i^X)$$

でもある。この結果, 式 (2) は

$$\begin{aligned} E[x - \mu_i] &= \int_{I_i^X} (x - \mu_i) dP(x|x \in I_i^X) \\ &= \int_{[0, \infty) \times [0, \infty)} (x - \mu_i) \frac{1_i(x)}{P(x \in I_i^X)} dP(x, y) \\ &= E \left[(x - \mu_i) \frac{1_i(x)}{P(x \in I_i^X)} \right] \end{aligned} \quad (10)$$

と変形できる。

結果として, 上述のモーメント式は次のようにまとめることができる。

$$E[\mathbf{m}] = \begin{pmatrix} E[1_1(x) - p_1] \\ \vdots \\ E[1_m(x) - p_m] \\ E[(x - \mu_1) \{1_1(x)/P(x \in I_1^X)\}] \\ \vdots \\ E[(x - \mu_m) \{1_m(x)/P(x \in I_m^X)\}] \\ E[(x - \phi_1) \{1_1(y)/P(y \in I_1^Y)\}] \\ \vdots \\ E[(x - \phi_n) \{1_n(y)/P(y \in I_n^Y)\}] \\ E[(y - \psi_1) \{1_1(x)/P(x \in I_1^X)\}] \\ \vdots \\ E[(y - \psi_m) \{1_m(x)/P(x \in I_m^X)\}] \end{pmatrix} \cdot \quad (11)$$

一方、これらモーメント式の標本対応 $\bar{\mathbf{m}}$ は

$$\bar{\mathbf{m}} = \begin{pmatrix} R_1^X - \int_{X_1 \times [0, \infty)} dP(x, y) \\ \vdots \\ R_m^X - \int_{X_m \times [0, \infty)} dP(x, y) \\ \bar{X}_1 - \int_{X_1 \times [0, \infty)} x dP(x, y) / P(x \in I_1^X) \\ \vdots \\ \bar{X}_m - \int_{X_m \times [0, \infty)} x dP(x, y) / P(x \in I_m^X) \\ \tilde{X}_1 - \int_{[0, \infty) \times Y_1} x dP(x, y) / P(y \in I_1^Y) \\ \vdots \\ \tilde{X}_n - \int_{[0, \infty) \times Y_n} x dP(x, y) / P(y \in I_n^Y) \\ \tilde{Y}_1 - \int_{X_1 \times [0, \infty)} y dP(x, y) / P(x \in I_1^X) \\ \vdots \\ \tilde{Y}_m - \int_{X_m \times [0, \infty)} y dP(x, y) / P(x \in I_m^X) \end{pmatrix} \quad (12)$$

となる。

\mathbf{m} の共分散行列 Σ を求めよう。 Σ は、GMM 推定の最適ウエイトを設定する際に必要となる。 Σ の各要素のうち、その一部については福井 (2015) に基づいて以下のように計算できる。

$$\begin{aligned} \sigma_{k,l}^p &= E[\{1_k(x) - p_k\}\{1_l(x) - p_l\}] \quad (k, l = 1, \dots, m) \\ &= \begin{cases} p_k(1 - p_k) & k = l \\ -p_k p_l & \text{otherwise} \end{cases} \end{aligned} \quad (13)$$

$$\begin{aligned} \sigma_{k,l}^{p\phi} &= E\left[\{1_k(x) - p_k\} \frac{(x - \phi_l) 1_l(y)}{P(y \in I_l^Y)}\right] \quad (k = 1, \dots, m; l = 1, \dots, n) \\ &= \frac{1}{P(y \in I_l^Y)} \int_{I_k^X \times I_l^Y} (x - \phi_l) dP(x, y) \end{aligned} \quad (14)$$

$$\begin{aligned} \sigma_{k,l}^{p\psi} &= E\left[\{1_k(x) - p_k\} \frac{(y - \psi_l) 1_l(x)}{P(x \in I_l^X)}\right] \quad (k, l = 1, \dots, m) \\ &= 0 \end{aligned} \quad (15)$$

$$\begin{aligned} \sigma_{k,l}^{\phi} &= E\left[\frac{(x - \phi_k) 1_k(y)}{P(y \in I_k^Y)} \frac{(x - \phi_l) 1_l(y)}{P(y \in I_l^Y)}\right] \quad (k, l = 1, \dots, n) \\ &= \begin{cases} \frac{1}{\{P(y \in I_k^Y)\}^2} \int_{[0, \infty) \times I_k^Y} (x - \phi_k)^2 dP(x, y) & k = l \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (16)$$

$$\begin{aligned} \sigma_{k,l}^{\phi\psi} &= E\left[\frac{(x - \phi_k) 1_k(y)}{P(y \in I_k^Y)} \frac{(y - \psi_l) 1_l(x)}{P(x \in I_l^X)}\right] \quad (k = 1, \dots, n; l = 1, \dots, m) \\ &= \frac{1}{P(x \in I_l^X) P(y \in I_k^Y)} \int_{I_l^X \times I_k^Y} (x - \phi_k)(y - \psi_l) dP(x, y) \end{aligned} \quad (17)$$

$$\sigma_{k,l}^{\psi} = E\left[\frac{(y - \psi_k) 1_k(x)}{P(x \in I_k^X)} \frac{(y - \psi_l) 1_l(x)}{P(x \in I_l^X)}\right] \quad (k, l = 1, \dots, m)$$

$$= \begin{cases} \frac{1}{\{P(x \in I_k^X)\}^2} \int_{I_k^X \times [0, \infty)} (y - \psi_k)^2 dP(x, y) & k = l \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

$\{1_k(x) - p_k\}$ と $\{(x - \mu_l)1_l(x)/P(x \in I_l^X)\}$ の共分散 $\sigma_{k,l}^{p\mu}$ ($k, l = 1, \dots, m$) は

$$\sigma_{k,l}^{p\mu} = \int_{[0, \infty) \times [0, \infty)} \{1_k(x) - p_k\} \left\{ (x - \mu_l) \frac{1_l(x)}{P(x \in I_l^X)} \right\} dP(x, y)$$

より, $k = l$ のとき

$$\begin{aligned} & \int_{[0, \infty) \times [0, \infty)} \{1_k(x) - p_k\} \left\{ (x - \mu_l) \frac{1_l(x)}{P(x \in I_l^X)} \right\} dP(x, y) \\ &= \int_{I_k^X \times [0, \infty)} (1 - p_k) \left\{ (x - \mu_k) \frac{1}{P(x \in I_k^X)} \right\} dP(x, y) \\ &= (1 - p_k) \int_{I_k^X} (x - \mu_k) dP(x|x \in I_k^X) = 0, \end{aligned} \quad (19)$$

$k \neq l$ のとき

$$\begin{aligned} & \int_{[0, \infty) \times [0, \infty)} \{1_k(x) - p_k\} \left\{ (x - \mu_l) \frac{1_l(x)}{P(x \in I_l^X)} \right\} dP(x, y) \\ &= \int_{I_l^X \times [0, \infty)} (0 - p_k) \left\{ (x - \mu_l) \frac{1}{P(x \in I_l^X)} \right\} dP(x, y) \\ &= -p_k \int_{I_l^X} (x - \mu_l) dP(x|x \in I_l^X) = 0, \end{aligned} \quad (20)$$

であるから, $\sigma_{k,l}^{p\mu} = 0$ となる。 $\{(x - \mu_k)1_k(x)/P(x \in I_k^X)\}$ と $\{(x - \mu_l)1_l(x)/P(x \in I_l^X)\}$ の共分散 $\sigma_{k,l}^{\mu}$ ($k, l = 1, \dots, m$) は,

$$\begin{aligned} \sigma_{k,l}^{\mu} &= \int_{[0, \infty) \times [0, \infty)} \left\{ (x - \mu_k) \frac{1_k(x)}{P(x \in I_k^X)} \right\} \left\{ (x - \mu_l) \frac{1_l(x)}{P(x \in I_l^X)} \right\} dP(x, y) \\ &= \int_{[0, \infty) \times [0, \infty)} (x - \mu_k)(x - \mu_l) \frac{1_k(x)1_l(x)}{P(x \in I_k^X)P(x \in I_l^X)} dP(x, y) \\ &= \begin{cases} \frac{1}{\{P(x \in I_k^X)\}^2} \int_{I_k^X} (x - \mu_k)^2 dP(x) & k = l \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (21)$$

となる。 $\{(x - \mu_k)1_k(x)/P(x \in I_k^X)\}$ と $\{(x - \phi_l)1_l(y)/P(y \in I_l^Y)\}$ の共分散 $\sigma_{k,l}^{\mu\phi}$ ($k = 1, \dots, m; l = 1, \dots, n$) は

$$\begin{aligned} \sigma_{k,l}^{\mu\phi} &= \int_{[0, \infty) \times [0, \infty)} \left\{ (x - \mu_k) \frac{1_k(x)}{P(x \in I_k^X)} \right\} \left\{ (x - \phi_l) \frac{1_l(y)}{P(y \in I_l^Y)} \right\} dP(x, y) \\ &= \frac{1}{P(x \in I_k^X)P(y \in I_l^Y)} \int_{I_k^X \times I_l^Y} (x - \mu_k)(x - \phi_l) dP(x, y) \end{aligned} \quad (22)$$

と展開される。また, $\{(x - \mu_k)1_k(x)/P(x \in I_k^X)\}$ と $\{(y - \psi_l)1_l(x)/P(x \in I_l^X)\}$ の共分散 $\sigma_{k,l}^{\mu\psi}$ ($k, l = 1, \dots, m$) は

$$\begin{aligned} \sigma_{k,l}^{\mu\psi} &= \int_{[0,\infty) \times [0,\infty)} \left\{ (x - \mu_k) \frac{1_k(x)}{P(x \in I_k^X)} \right\} \left\{ (y - \psi_l) \frac{1_l(x)}{P(x \in I_l^X)} \right\} dP(x, y) \\ &= \begin{cases} \frac{1}{\{P(x \in I_k^X)\}^2} \int_{I_k^X \times [0,\infty)} (x - \mu_k)(y - \psi_l) dP(x, y) & k = l \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (23)$$

となる。

以上より, m の共分散行列 Σ は,

$$\Sigma = \begin{pmatrix} \Sigma_p & \mathbf{0} & \Sigma_{p\phi} & \mathbf{0} \\ \mathbf{0} & \Sigma_\mu & \Sigma_{\mu\phi} & \Sigma_{\mu\psi} \\ \Sigma'_{p\phi} & \Sigma'_{\mu\phi} & \Sigma_\phi & \Sigma_{\phi\psi} \\ \mathbf{0} & \Sigma'_{\mu\psi} & \Sigma'_{\phi\psi} & \Sigma_\psi \end{pmatrix} \quad (24)$$

と表すことができる。ただし,

$$\begin{aligned} \Sigma_p &= \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_m \\ -p_1p_2 & p_2(1-p_2) & \cdots & -p_2p_m \\ \vdots & \vdots & \ddots & \vdots \\ -p_1p_m & -p_2p_m & \cdots & p_m(1-p_m) \end{pmatrix}, \quad \Sigma_{p\phi} = \begin{pmatrix} \sigma_{1,1}^{p\phi} & \cdots & \sigma_{1,n}^{p\phi} \\ \vdots & \ddots & \vdots \\ \sigma_{m,1}^{p\phi} & \cdots & \sigma_{m,n}^{p\phi} \end{pmatrix}, \\ \Sigma_\mu &= \begin{pmatrix} \sigma_{1,1}^\mu & & 0 \\ & \ddots & \\ 0 & & \sigma_{m,m}^\mu \end{pmatrix}, \quad \Sigma_{\mu\phi} = \begin{pmatrix} \sigma_{1,1}^{\mu\phi} & \cdots & \sigma_{1,n}^{\mu\phi} \\ \vdots & \ddots & \vdots \\ \sigma_{m,1}^{\mu\phi} & \cdots & \sigma_{m,n}^{\mu\phi} \end{pmatrix}, \quad \Sigma_{\mu\psi} = \begin{pmatrix} \sigma_{1,1}^{\mu\psi} & & 0 \\ & \ddots & \\ 0 & & \sigma_{m,m}^{\mu\psi} \end{pmatrix}, \\ \Sigma_\phi &= \begin{pmatrix} \sigma_{1,1}^\phi & & 0 \\ & \ddots & \\ 0 & & \sigma_{n,n}^{\mu\phi} \end{pmatrix}, \quad \Sigma_{\phi\psi} = \begin{pmatrix} \sigma_{1,1}^{\phi\psi} & \cdots & \sigma_{1,m}^{\phi\psi} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1}^{\phi\psi} & \cdots & \sigma_{n,m}^{\phi\psi} \end{pmatrix}, \quad \Sigma_\psi = \begin{pmatrix} \sigma_{1,1}^\psi & & 0 \\ & \ddots & \\ 0 & & \sigma_{m,m}^\psi \end{pmatrix} \end{aligned}$$

である。

いま, 同時分布の形状がパラメータ θ によって決まると仮定し, その同時分布を $P(x, y; \theta)$ と表すことにしよう。このとき, 標本対応と共分散行列について, それらの値もまたパラメータ θ によって決まることになる。そこで, 標本対応と共分散行列が同時分布のパラメータ θ の関数であることを明示するため, 標本対応を $\bar{m}(\theta)$, 共分散行列を $\Sigma(\theta)$ と表す。

GMM 推定は, 標本対応について共分散行列の逆行列をウェイトとする二次形式を最小にするパラメータをその推定値とする。つまり, 今回の推定では

$$\min_{\theta} \bar{m}(\theta)' \Sigma(\theta)^{-1} \bar{m}(\theta) \quad (25)$$

を満たすパラメータ θ の値が GMM による推定値となる。

福井 (2015) と同様の理由により, 上記の GMM 推定量 $\hat{\theta}$ は, 一致性および漸近的正規性を持つ。このとき, 漸近共分散行列 S は

$$\begin{aligned} S &= \frac{1}{N} \left(G(\hat{\theta})' \Sigma(\hat{\theta})^{-1} G(\hat{\theta}) \right)^{-1} \\ G(\hat{\theta}) &= \left. \frac{\partial m(\theta)}{\partial \theta'} \right|_{\theta=\hat{\theta}} \end{aligned} \quad (26)$$

表 1: 全国を対象とした所得と年齢の分割表 (同時確率, 総務省「平成 21 年全国消費実態調査」より筆者作成)

	30歳未満	30歳以上 40歳未満	40歳以上 50歳未満	50歳以上 60歳未満	60歳以上 70歳未満	70歳以上
200万円未満	0.0012	0.004455	0.004892	0.006134	0.013028	0.014299
200万円以上 300万円未満	0.003225	0.007994	0.007314	0.008461	0.028842	0.028051
300万円以上 400万円未満	0.006312	0.019255	0.014151	0.014553	0.044249	0.052185
400万円以上 500万円未満	0.005246	0.028586	0.018953	0.018603	0.039994	0.033397
500万円以上 600万円未満	0.003053	0.029298	0.022594	0.018965	0.031436	0.018697
600万円以上 800万円未満	0.003244	0.036975	0.051721	0.043206	0.039185	0.018135
800万円以上 1000万円未満	0.001128	0.014405	0.032696	0.039953	0.020846	0.00853
1000万円以上 1250万円未満	0.00019	0.005938	0.019065	0.031111	0.012899	0.004692
1250万円以上 1500万円未満	3.17E-05	0.002239	0.008702	0.015017	0.006487	0.00204
1500万円以上		0.001076	0.006988	0.015193	0.007462	0.003423

である (ただし, N は総度数である)。

3 同時分布のモデル化

今回の推定では, 所得と年齢に対して単変量の確率分布を仮定し, それらをコンピュータにより結合することで同時分布を構築する。すなわち, 所得と年齢をそれぞれ確率変数 (x, y) とし, 所得の周辺確率分布を $F_x(x)$, 年齢の周辺確率分布を $F_y(y)$ とすると, それらの同時確率分布 $F_{x,y}(x, y)$ は

$$F_{x,y}(x, y) = C(F_x(x), F_y(y)) \tag{27}$$

となる。ここで, $C(F_x(x), F_y(y))$ がコンピュータと呼ばれる関数である*3。

所得と年齢の同時分布については, 60歳近辺を境界とする構造変化が見られる。これは, 日本における多くの労働者が60歳周辺で退職し, 所得の源泉および金額がその前後で大きく異なるためである。実際に, 全国消費実態調査における全国を対象とした年齢と所得の分割表から, この傾向が見て取れる(表1)。

そこで, 60歳を境界として所得分布とコンピュータのパラメータが異なるよう同時分布 $F_{x,y}(x, y)$ を構築する。所得分布は, 60歳以下の世代と60歳より上の世代とでその形状が異なると仮定し, それらの周辺分布関数をそれぞれ $F_x(x; \lambda_1), F_x(x; \lambda_2)$ と置く。一方, 年齢分布に関しては, 単一の周辺分布関数 $F_y(y)$ を仮定する。コンピュータについては, 所得分布と同様, 60歳を境界としてその上下の世代でパ

*3 コピュラが満たすべき条件などについては, Joe (1997), Nelsen (2006), または Jondeau et al. (2007) を参照せよ。

ラメータが異なるとして、60歳以下世帯のコピュラを $C_1(F_x(x; \lambda_1), F_y(y); \rho_1)$ 、60歳より上の世帯のコピュラを $C_2(F_x(x; \lambda_2), F_y(y); \rho_2)$ とする。

以上より、60歳以下世帯の同時分布は

$$F_1(x, y | y \leq 60) = \begin{cases} C_1(F_x(x; \lambda_1), F_y(y); \rho_1) / C_1(1, F_y(60); \rho_1) & y \leq 60 \\ 0 & y > 60 \end{cases} \quad (28)$$

と、60歳より上の世帯の同時分布は

$$F_2(x, y | y > 60) = \begin{cases} 0 & y \leq 60 \\ C_2(F_x(x; \lambda_2), F_y(y); \rho_2) / (1 - C_2(1, F_y(60); \rho_2)) & y > 60 \end{cases} \quad (29)$$

となるため、全体の同時分布は

$$F_{x,y}(x, y) = P(y \leq 60)F_1(x, y | y \leq 60) + P(y > 60)F_2(x, y | y > 60) \quad (30)$$

と表すことができる。

パラメータ $\theta = (\lambda_1, \lambda_2, \rho_1, \rho_2)'$ を推定するには、所得分布、年齢分布、およびコピュラをモデル化しなくてはならない。今回のモデル化では、所得分布として第2種の一般化ベータ分布を用いる*4。すなわち、所得分布の密度関数 $f_x(x; \lambda)$ を

$$\begin{aligned} f_x(x; \lambda) &= f_x(x; (a, b, p, q)') \\ &= \frac{ax^{ap-1}}{b^{ap}B(p, q) \{1 + (x/b)^a\}^{p+q}} \end{aligned} \quad (31)$$

とする。ただし、 $a > 0, b > 0, p > 0, q > 0$ である。

年齢分布は、カーネル関数を正規カーネルとするノンパラメトリック推定を行う*5。第 j 年齢階層の階級値 M_j を各階層の上限と下限の平均値として、年齢分布の密度関数 $f_y(y)$ を

$$f_y(y) = \frac{1}{Nh} \sum_j N_j K\left(\frac{y - M_j}{h}\right) \quad (32)$$

と置く。ここで、 N は総度数、 N_j は第 j 年齢階層の度数、 $K(\cdot)$ はカーネル関数、 h はバンド幅である。また、カーネル関数として、正規カーネル

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

を用いる。なお、カーネル推定でのバンド幅について、今回の分析では目算により設定する。

コピュラについては、正規コピュラを仮定する*6。上記の GMM 推定では標本対応や共分散行列の計算の際、同時分布について積分を求めなくてはならない。その数値計算において、同時密度関数の形状

*4 第2種の一般化ベータ分布については、McDonald (1984) および McDonald and Xu (1995) を参照。

*5 ノンパラメトリック法による密度関数の推定については、竹澤 (2007) が詳しい。

*6 正規コピュラについては、例えば Trivedi and Zimmer (2005) を参照。

が明らかであることが望ましい。正規コピュラの場合、その同時密度関数 $f(x, y; \rho)$ は

$$\begin{aligned}
 f(x, y; \rho) &= c(x, y; \rho) f_x(x) f_y(y) \\
 c(x, y; \rho) &= \frac{\exp B}{\sqrt{1 - \rho^2}} \\
 B &= \frac{2\rho\Phi^{-1}(u)\Phi^{-1}(v) - \rho^2\{(\Phi^{-1}(u))^2 + (\Phi^{-1}(v))^2\}}{2(1 - \rho^2)} \\
 u &= \int_{-\infty}^x f_x(z) dz \\
 v &= \int_{-\infty}^y f_y(z) dz
 \end{aligned} \tag{33}$$

である。ただし、 $\Phi^{-1}(\cdot)$ は標準正規分布の累積密度関数の逆関数である。

以上より、同時分布 $F_{x,y}(x, y; \theta)$ について、その密度関数 $f_{x,y}(x, y; \theta)$ を

$$\begin{aligned}
 f_{x,y}(x, y; \theta) &= P(y \leq 60) f_1(x, y | y \leq 60; \theta) + P(y > 60) f_2(x, y | y > 60; \theta) \\
 f_1(x, y | y \leq 60; \theta) &= \frac{c(x, y; \rho_1) f_x(x | \lambda_1) f_y(y)}{\int_0^{60} \int_0^{\infty} \{c(x, y; \rho_1) f_x(x | \lambda_1) f_y(y)\} dx dy} \\
 f_2(x, y | y > 60; \theta) &= \frac{c(x, y; \rho_2) f_x(x | \lambda_2) f_y(y)}{\int_{60}^{\infty} \int_0^{\infty} \{c(x, y; \rho_2) f_x(x | \lambda_2) f_y(y)\} dx dy}
 \end{aligned}$$

と置くことができる。ただし、

$$\begin{aligned}
 \theta &= (\lambda_1, \lambda_2, \rho_1, \rho_2)' \\
 &= ((a_1, b_1, p_1, q_1), (a_2, b_2, p_2, q_2), (\rho_1), (\rho_2))'
 \end{aligned}$$

である。

今回の分析で、年齢分布 $F_y(y)$ はノンパラメトリック推定により求めているため、パラメータ θ の推定において $F_y(y)$ は所与とみなせる。したがって、推定の過程で $P(y \leq 60)$ と $P(y > 60)$ は常に一定であり、また、コピュラの性質より

$$\begin{aligned}
 P(y \leq 60) &= \int_0^{60} \int_0^{\infty} f_1(x, y) dx dy \\
 P(y > 60) &= \int_{60}^{\infty} \int_0^{\infty} f_2(x, y) dx dy
 \end{aligned}$$

となることから、今回の分析における同時密度関数 $f_{x,y}(x, y; \theta)$ を

$$\begin{aligned}
 f_{x,y}(x, y; \theta) &= f_1(x, y; \theta) + f_2(x, y; \theta) \\
 f_1(x, y; \theta) &= \begin{cases} c(x, y; \rho_1) f_x(x; \lambda_1) f_y(y) & y \leq 60 \\ 0 & y > 60 \end{cases} \\
 f_2(x, y; \theta) &= \begin{cases} 0 & y \leq 60 \\ c(x, y; \rho_2) f_x(x; \lambda_2) f_y(y) & y > 60 \end{cases}
 \end{aligned} \tag{34}$$

と簡略化できる。

以上より、同時分布、すなわち同時密度の形状が確定したため、GMMにおける目的関数 $\bar{m}(\theta)' \Sigma(\theta)^{-1} \bar{m}(\theta)$ が計算可能となる。

表 2: 平成 21 年全国消費実態調査における鹿児島県調査対象市町村（カッコ内の数字は集計世帯数）

経済圏	市町村名
経済圏 A (353)	鹿児島市, 枕崎市, 指宿市, 日置市, いちき串木野市, 南さつま市, 南九州市
経済圏 B (107)	阿久根市, 出水市, 薩摩川内市, さつま町
経済圏 C (94)	霧島市, 伊佐市, 始良市
経済圏 D (123)	鹿屋市, 垂水市, 曾於市, 志布志市, 錦江町
経済圏 E (93)	西之表市, 奄美市, 屋久島町, 和泊町

(出典：総務省「平成 21 年全国消費実態調査」)

4 実証分析

本研究では、第一に、総務省「平成 21 年全国消費実態調査」の鹿児島県内各経済圏の階層別データを用いて、それら経済圏における所得と年齢の同時分布を推定する。第二に、推定した所得と年齢の同時分布に基づいて、各経済圏における年齢層別の所得分布を求め、所得格差についての定量分析を試みる。なお、以下では、所得を全国消費実態調査における世帯別年間収入、年齢を同調査における世帯主の年齢とする。

平成 21 年全国消費実態調査において、鹿児島県内の調査対象市町村およびその経済圏ごとの分類を表 2 に示している。また、表 3 は各経済圏の所得の階層別データ、表 4 は各経済圏の年齢の階層別データである。表 3 と表 4 の両方で、最上位階層の上限が設定されていない。しかし、GMM 推定における数値計算の際、最上位所得階層の上限および最上位年齢階層の上限が必要である。そこで、以下の数値計算では、表 3 の最上位所得階層の上限を 3000、表 4 の最上位年齢階層の上限を 100 とする。

前節までの議論から、GMM によるパラメータ θ の推定が可能となる。そこで、数値最適化の方法を用いて、式 (25) の目的関数の値を最小にするパラメータ θ を求めよう。ただし、推定に際して、年齢分布の形状を事前に決定する必要がある。先述の通り、ノンパラメトリック推定のバンド幅を目算により決めるとし、以下の推定においては、すべての経済圏でバンド幅の値を 6 とする。図 1 は、バンド幅を 6 とした場合の、各経済圏における年齢の密度関数を示している。

実際の数値最適化では、Nelder-Mead 法を複数回行った後に BFGS 法を適用する。目的関数を $\Lambda(\theta)$ とすると、本研究での BFGS 法による最適化では

$$\left\| \frac{\partial \Lambda(\hat{\theta})}{\partial \hat{\theta}} \right\|_2 < 0.1$$

を収束判定条件とする。また、最適化を行う際の初期値を設定する必要がある。所得分布として第 2 種の一般化ベータ分布を仮定し、表 3 のデータについて、多項分布に基づく最尤法で推定したパラメータを λ_1 および λ_2 の初期値とする^{*7}。また、 (ρ_1, ρ_2) の初期値を $(0.3, -0.3)$ とする。数値最適化の過程で、同時密度関数の二重積分を計算しなければならない。今回の分析では、二重積分の数値計算方法としてシンプソン法による数値積分を用いる。

*7 多項分布に基づく最尤法を用いた所得分布の推定については、例えば Kleiber and Kots (2003) を参照。

表 3: 平成 21 年全国消費実態調査における鹿児島県各経済圏の世帯年間収入階層別データ (抽出率調整済み)

		経済圏 A			経済圏 B			経済圏 C		
下限 (万円)	上限 (万円)	度数 (1 万分比)	平均所得 (万円)	平均年齢 (歳)	度数 (1 万分比)	平均所得 (万円)	平均年齢 (歳)	度数 (1 万分比)	平均所得 (万円)	平均年齢 (歳)
0	200	590	137.8	62.8	664	170.5	67.3	614	110.7	54.4
200	300	1595	256.4	57.2	1283	248.4	58.6	1789	257.2	59.1
300	400	1840	349.8	57.3	1983	359.2	54.5	1995	342	62.9
400	500	1840	450.7	55.8	2249	446.9	54.4	1347	430.5	57.3
500	600	1162	550.9	52.7	1584	547.1	59.7	1266	534.1	55.3
600	800	1659	682.1	51.9	1473	684.5	49.2	1866	653.9	52.8
800	1000	715	879	53	469	913.9	56.6	603	898	56.4
1000	1250	368	1165.8	52.1				260	1107.3	55.3
1250	1500	110	1372.9	59						
1500	3000	121	2141.6	60.6						
		経済圏 D			経済圏 E					
下限 (万円)	上限 (万円)	度数 (1 万分比)	平均所得 (万円)	平均年齢 (歳)	度数 (1 万分比)	平均所得 (万円)	平均年齢 (歳)			
0	200	893	129	63.9	1425	145.1	61.3			
200	300	1211	240.2	66.8	1589	245.2	60.5			
300	400	1737	354.9	61.5	1630	342.3	52.9			
400	500	1366	436	56.1	1919	453.7	58			
500	600	1316	549.1	47	488	560.7	54.3			
600	800	1837	690	52.2	1013	699.7	50.9			
800	1000	728	826.3	52.2	1105	883.8	51.9			
1000	1250	519	1122.9	51.9	425	1080.5	54			
1250	1500	257	1388.6	56.4	407	1298	53.5			
1500		135	2330.6	50.5						

(出典：総務省「平成 21 年全国消費実態調査」)

表 4: 平成 21 年全国消費実態調査における鹿児島県各経済圏の世帯主年齢階層別データ (抽出率調整済み)

下限 (歳)	上限 (歳)	経済圏 A		経済圏 B		経済圏 C		経済圏 D		経済圏 E	
		度数 (1 万分比)	平均所得 (万円)	度数 (1 万分比)	平均所得 (万円)	度数 (1 万分比)	平均所得 (万円)	度数 (1 万分比)	平均所得 (万円)	度数 (1 万分比)	平均所得 (万円)
25	30	355	413.8	286	522.9			272	415.4		
30	35	604	487.6	838	456	448	437	512	528.6	504	378.8
35	40	1146	509.8	554	465.3	888	390.2	467	672.9	514	646.8
40	45	630	667.3	748	433.9	1265	477.3	951	616.1	699	605.2
45	50	803	548.3	470	515.1	254	455.4	956	581.5	740	589.1
50	55	808	560.9	801	601.1	1127	605.8	1168	746.5	1696	599.3
55	60	1253	679.7	1430	688	1779	688.8	1540	654.9	1484	627.1
60	65	1113	562.8	1954	396.9	787	441.6	1089	542.4	1107	438.2
65	70	999	435.8	1000	538	1248	434.3	1154	497.2	1534	445
70	75	1136	445.3	1041	322.2	576	304.4	786	252.4	371	243.9
75		1091	396.4	876	402.1	1538	394.9	1105	345.9	1059	287.5

(出典：総務省「平成 21 年全国消費実態調査」)

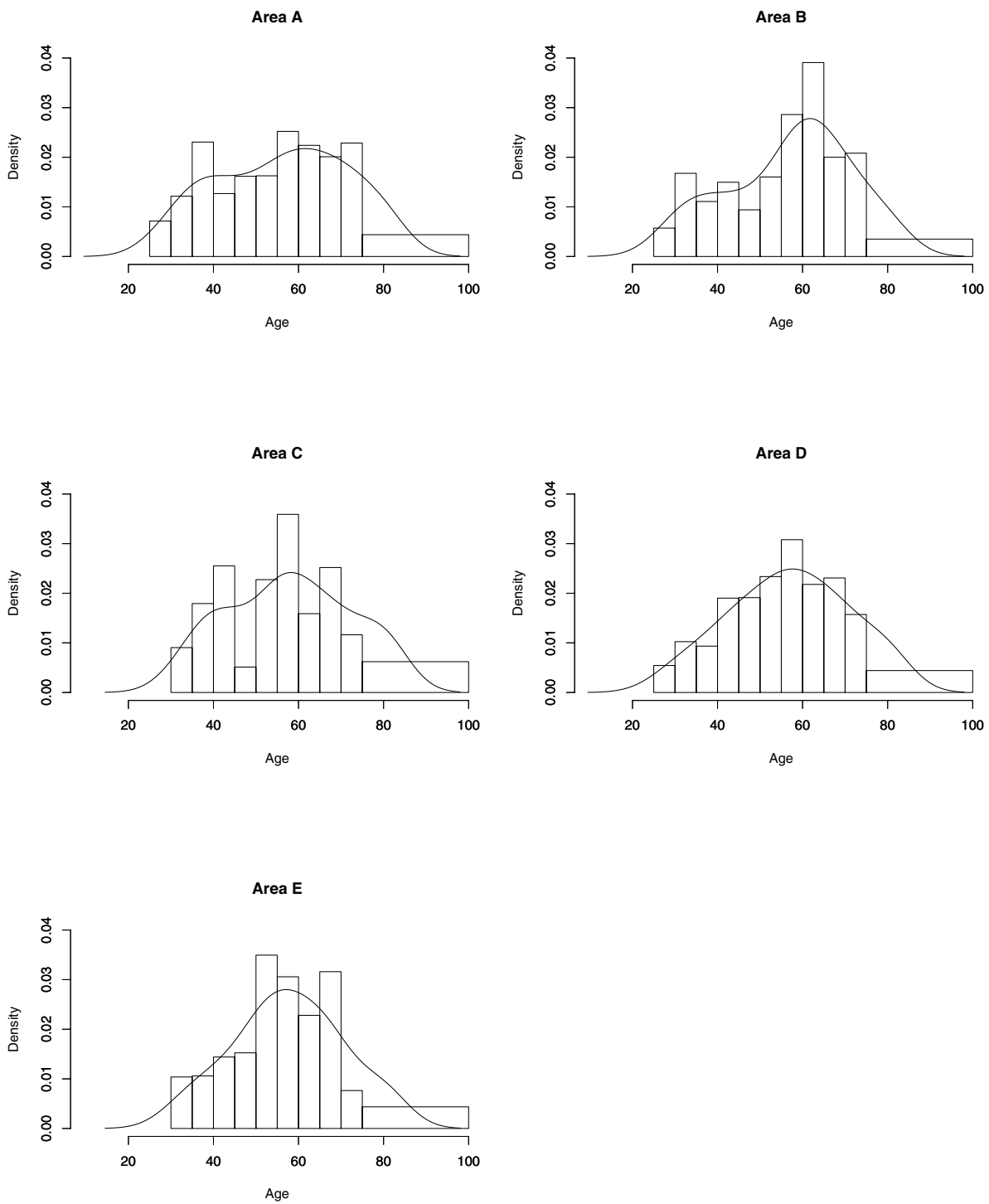


図 1: 年齢の密度関数 (ヒストグラムは元のデータ, 実線は推定結果)

表 5: 同時密度関数のパラメータの推定値 (括弧内は漸近標準偏差の推定値)

経済圏 A	a	b	p	q	ρ	平均	標準偏差	Theil 指数
60 歳以下	0.89666 (1.13273)	739.397 (932.594)	9.03293 (20.4591)	11.2799 (31.9432)	0.399959 (4.8161E-2)	557.19	282.88	0.115541
60 歳より上	0.797125 (2.30565)	715.829 (2311.59)	7.92093 (39.1455)	11.1251 (71.1643)	-0.337592 (5.14089E-2)	449.229	276.844	0.162999
経済圏 B	a	b	p	q	ρ	平均	標準偏差	Theil 指数
60 歳以下	0.245199 (0.926909)	4827.12 (51482.2)	99.6829 (859.294)	160.383 (982.391)	0.613689 (6.23419E-2)	604.631	282.99	0.100658
60 歳より上	0.229543 (1.87405)	4652.09 (172006)	114.886 (1782.04)	199.993 (3580.46)	-0.169263 (0.128726)	435.309	233.525	0.12708
経済圏 C	a	b	p	q	ρ	平均	標準偏差	Theil 指数
60 歳以下	2.31463 (1.29976)	47605.1 (1.244E+6)	0.799073 (0.697171)	10688.3 (645126)	0.578494 (3.69881E-2)	530.3	284.804	0.148529
60 歳より上	1.45561 (0.530248)	42912.6 (2.96725E+6)	2.26146 (1.27004)	1733.08 (172471)	-0.197118 (0.120205)	394.237	186.388	0.110368
経済圏 D	a	b	p	q	ρ	平均	標準偏差	Theil 指数
60 歳以下	2.14276 (1.16413)	1093.88 (612.368)	1.37768 (1.23107)	4.05087 (6.04748)	0.279008 (9.45482E-2)	601.386	303.954	0.121515
60 歳より上	0.823788 (2.07209)	1261.27 (7924.78)	4.62467 (18.0264)	10.2813 (70.6285)	-0.445161 (7.8455E-2)	413.663	290.84	0.209379
経済圏 E	a	b	p	q	ρ	平均	標準偏差	Theil 指数
60 歳以下	0.793921 (0.402899)	52270.3 (421137)	4.50875 (4.51282)	143.405 (838.275)	0.331186 (0.101001)	605.844	367.461	0.1691
60 歳より上	0.0962131 (0.191345)	50474.2 (825640)	327.651 (1501.74)	515.79 (1968.24)	-0.51757 (7.59124E-2)	396.55	289.561	0.21695

表 5 は、表 3 と表 4 のデータに基づいて、同時密度関数の式 (34) のパラメータ θ を GMM により推定した結果である。また、表 6 および 7 は、推定された同時分布に基づいて計算した階層別データであり、図 2 と図 3 は、推定結果に基づく、各経済圏における所得分布・コピュラ密度を示している。

表 5 および図 2 から分かるように、いずれの経済圏も、60 歳以下世帯の所得分布と 60 歳より上の世帯を比較したとき、後者の方が平均所得・標準偏差がともに小さい。60 歳より上の世帯は、収入の源泉が年金などの社会保障給付に限られることが多く、60 歳以下世帯と比較して年間収入が世帯間で均一化していることを示している。Theil 指数については、経済圏 C を除いて 60 歳以下世帯よりも 60 歳より上の世帯でその値が高い。したがって、所得の不平等度は高齢層の方が高いと言える。また、表 5 から、60 歳以下世帯および 60 歳より上の世帯のそれぞれにおける地域内・地域間所得格差が明らかとなる。60 歳以下世帯の場合、経済圏 C・E は他の経済圏と比較して Theil 指数が大きく、経済圏 A・C の平均所得はそれ以外の経済圏と比べて低い。60 歳より上の世帯の場合、経済圏 D・E は経済圏 B・C と比較して Theil 指数が大きく、経済圏 C・D・E の平均所得はそれ以外の経済圏と比較して低い。

以上の結果から、すべての経済圏で 60 歳以下世帯は 60 歳より上の世帯よりも平均所得が高く、60 歳より上の世帯では、都市部から遠い地域ほど地域内所得格差が大きく平均所得が低くなる傾向がある。

表 6: 推定した同時分布に基づく鹿児島県各経済圏の世帯年間収入階層別データ

下限 (万円)	上限 (万円)	経済圏 A			経済圏 B			経済圏 C		
		相対度数	平均所得 (万円)	平均年齢 (歳)	相対度数	平均所得 (万円)	平均年齢 (歳)	相対度数	平均所得 (万円)	平均年齢 (歳)
0	200	0.0748904	160.353	67.5644	0.0528997	171.179	68.9191	0.124686	140.743	57.8234
200	300	0.148378	262.505	60.9311	0.137101	265.573	63.1763	0.153955	255.752	61.3525
300	400	0.177364	361.59	57.7952	0.174776	364.444	59.8103	0.166404	361.183	62.3826
400	500	0.161447	459.946	56.072	0.164575	464.291	57.8332	0.150746	460.392	60.9386
500	600	0.127209	557.249	55.1119	0.132701	565.221	56.8819	0.121846	554.76	58.5894
600	800	0.156115	695.899	54.6787	0.16497	715.598	56.9916	0.150367	710.976	57.4703
800	1000	0.0721489	891.152	54.6083	0.0759109	929.335	57.5809	0.0687549	917.453	55.4397
1000	1250	0.0366644	1109.17	55.1177	0.0390666	1112.09	55.7512	0.0289954	1161.88	55.9942
1250	1500	0.0135971	1350.89	55.3873	0.013435	1362.24	56.157	0.00668604	1340.91	53.295
1500	3000	0.00902825	1783.61	56.5429	0.00717699	1801.65	58.2109	0.00104087	1642.51	55.2779
下限 (万円)	上限 (万円)	経済圏 D			経済圏 E					
		相対度数	平均所得 (万円)	平均年齢 (歳)	相対度数	平均所得 (万円)	平均年齢 (歳)			
0	200	0.119444	143.697	68.8058	0.115476	185.184	84.7933			
200	300	0.136497	257.125	62.7476	0.147136	274.107	66.5698			
300	400	0.144451	358.831	59.1188	0.131894	397.958	66.2151			
400	500	0.134462	463.416	56.8956	0.114783	493.098	61.7882			
500	600	0.116655	560.781	54.8488	0.0900887	614.061	61.5747			
600	800	0.164524	704.049	53.7329	0.132679	720.905	56.3021			
800	1000	0.0869405	902.1	53.3668	0.0730831	959.38	57.5148			
1000	1250	0.0467985	1115.21	53.5706	0.0476064	1174.87	56.4248			
1250	1500	0.0169842	1359.96	54.2439	0.0220809	1422.68	55.975			
1500	3000	0.0102839	1776.06	55.9848	0.0188144	1827.18	54.8784			

表 7: 推定した同時分布に基づく鹿児島県各経済圏の世帯主年齢階層別データ

下限 (歳)	上限 (歳)	経済圏 A		経済圏 B		経済圏 C		経済圏 D		経済圏 E	
		相対度数	平均所得 (万円)	相対度数	平均所得 (万円)	相対度数	平均所得 (万円)	相対度数	平均所得 (万円)	相対度数	平均所得 (万円)
25	30	0.0366937	445.844	0.0318239	462.699	0.0194894	309.188	0.0263162	506.151	0.0169455	472.589
30	35	0.0615576	484.332	0.0505169	512.987	0.0451563	381.958	0.04355	534.26	0.0351687	512.313
35	40	0.0776974	519.471	0.0599925	557.406	0.0709702	450.271	0.0617573	559.764	0.0536829	547.815
40	45	0.0813239	550.325	0.0632715	596.2	0.0826205	509.444	0.0814256	584.136	0.0736481	580.641
45	50	0.0839247	578.093	0.0718021	633.112	0.0882898	560.142	0.100403	608.489	0.100962	613.986
50	55	0.0936733	606.029	0.0955479	675.398	0.104907	609.18	0.116441	633.591	0.128241	650
55	60	0.104868	637.104	0.126622	730.468	0.118993	661.683	0.123754	659.884	0.138731	688.467
60	65	0.108034	503.411	0.136839	458.622	0.11258	419.126	0.117031	488.707	0.114515	544.497
65	70	0.102034	475.68	0.118936	446.009	0.0956303	410.504	0.100292	447.087	0.0882166	515.246
70	75	0.0890051	447.154	0.0892858	433.5	0.0808209	401.882	0.0805002	408.024	0.0607098	484.546
75	100	0.138031	392.618	0.117977	410.905	0.154023	381.651	0.12557	337.917	0.0828227	421.631

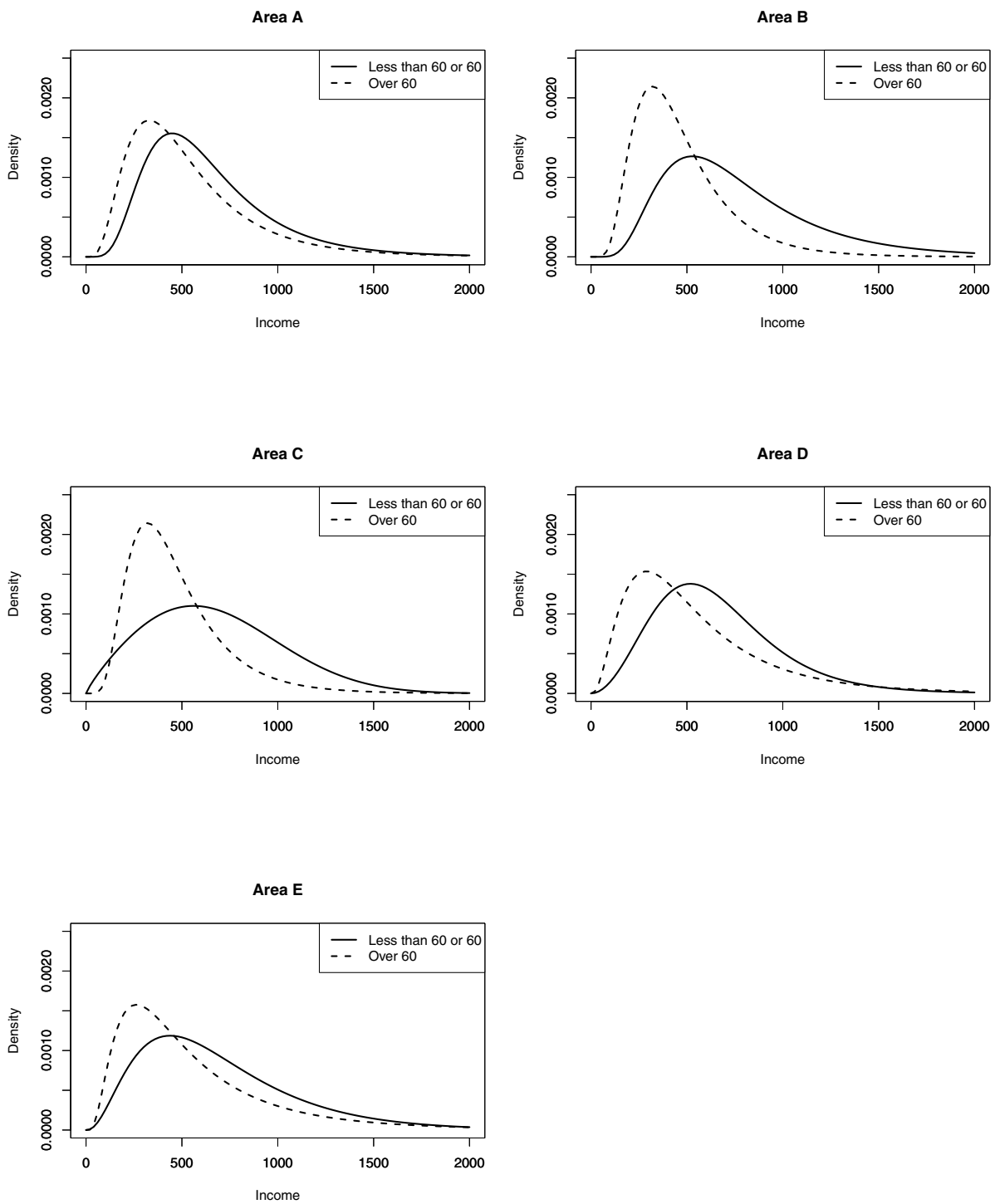
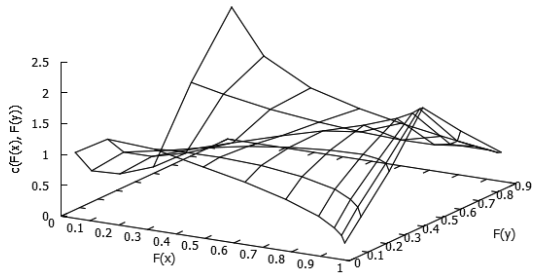
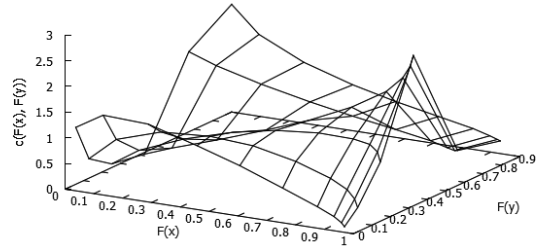


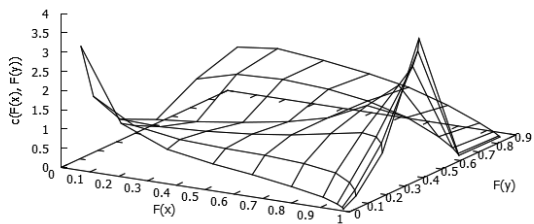
図 2: 各経済圏における所得分布の推定結果 (実線は 60 歳以下世帯, 破線は 60 歳より上の世帯)



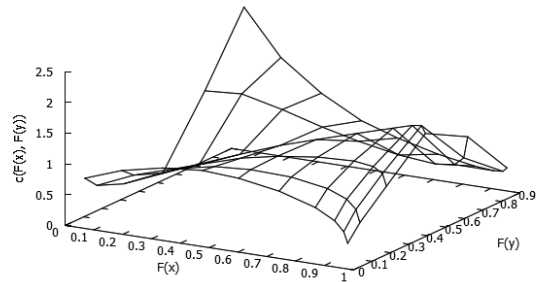
(a) 経済圏 A



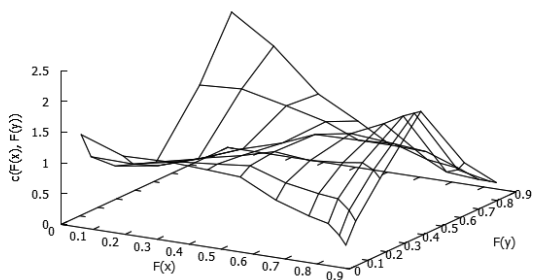
(b) 経済圏 B



(c) 経済圏 C



(d) 経済圏 D



(e) 経済圏 E

図 3: 各経済圏におけるコピュラ密度の推定結果

表 8: 世帯主年齢が 40 歳以下の世帯における各種代表値の推定結果

	平均値 (万円)	標準偏差 (万円)	Theil 指数	年収 200 万円以下 世帯割合 (%)	年収 300 万円以下 世帯割合 (%)
経済圏 A	482.747	241.780	0.112809	7.49471	22.3131
経済圏 B	475.270	237.904	0.112723	5.34468	18.9108
経済圏 C	384.228	237.116	0.189833	12.4747	27.8621
経済圏 D	531.753	276.162	0.128858	12.0471	25.7096
経済圏 E	495.502	312.545	0.180831	11.7281	26.9184

一方、60 歳以下世帯では、都市部から遠い地域ほど地域内所得格差が大きく平均所得が高くなる傾向がみられる。ただし、60 歳より上の世帯と比較して、60 歳以下世帯の地域差はそれほど明確ではない。

表 5 から、所得分布のパラメータについて漸近標準偏差が著しく大きいことが分かる。これは、第 2 種の一般化ベータ分布における識別性の問題を反映している。西埜 (2012) は、第 2 種の一般化ベータ分布にはある種の識別性の問題が存在することを指摘する。ここでの識別性の問題は、互いに値の大きく異なるパラメータの組であっても、その分布が極めて類似した形になることを意味する。そのため、識別性の問題が発生している状況では、推定結果と異なるパラメータの組が推定結果と同様の分布形を示しうる。したがって、パラメータの漸近標準偏差の大きさは、推定結果と大きく異なるパラメータの組が多数存在することを反映していると考えられる。

コピュラ密度から、所得と年齢間の関連を見ることができる。図 3 では、 $F(x)$ は所得の累積度数を、 $F(y)$ は年齢の累積度数を、縦軸はコピュラ密度を表している。図 3 より、いずれの経済圏においても、60 歳以下では年齢が高くなるにつれて所得が高くなり、60 歳より上では年齢が高くなるほど所得が低くなるという関連が確認できる。これは、福井 (2015) で全国のデータを用いて推定した場合のコピュラ密度と同様である。

一方、コピュラ密度を見ると、経済圏 B・C では 60 歳近辺における高所得層での所得と年齢の関連が著しく大きい。その原因として、高所得層におけるデータの不足が考えられる。表 3 にあるように、経済圏 B・C について、高所得層のデータが存在しない。したがって、これらの経済圏で、高所得層の度数・平均所得・平均年齢といった情報は GMM 推定に含まれない。この情報の欠落により、高所得層における所得と年齢の関連が経済圏 B・C では正しく推定されていない可能性がある。

表 2 に示しているように、すべての経済圏で集計世帯数すなわち標本が十分に大きいとは言えない。従来の計量モデルと同様、標本が小さいほど、推定した同時分布と真の同時分布との乖離は大きくなり、推定した同時分布は各経済圏の現状を良く説明できなくなる。それが、60 歳以下世帯における地域ごとの特徴付けの不確かさや、経済圏 B・C におけるコピュラ密度の極端な形状といった、推定上の問題の原因となっている可能性がある。

所得と年齢の同時分布を推定したことで、各経済圏を対象とする所得格差の現状についてより詳細な分析が可能となる。そこで、40 歳以下の世帯について、経済圏ごとにその所得分布を推定してみよう。図 4 は、推定した同時分布より導出した、各経済圏の 40 歳以下世帯の所得分布である。また、表 8 は、その各種代表値である。

表 8 より、40 歳以下世帯を対象とする場合、地域内所得格差および低所得層の割合は地方ほど高い傾

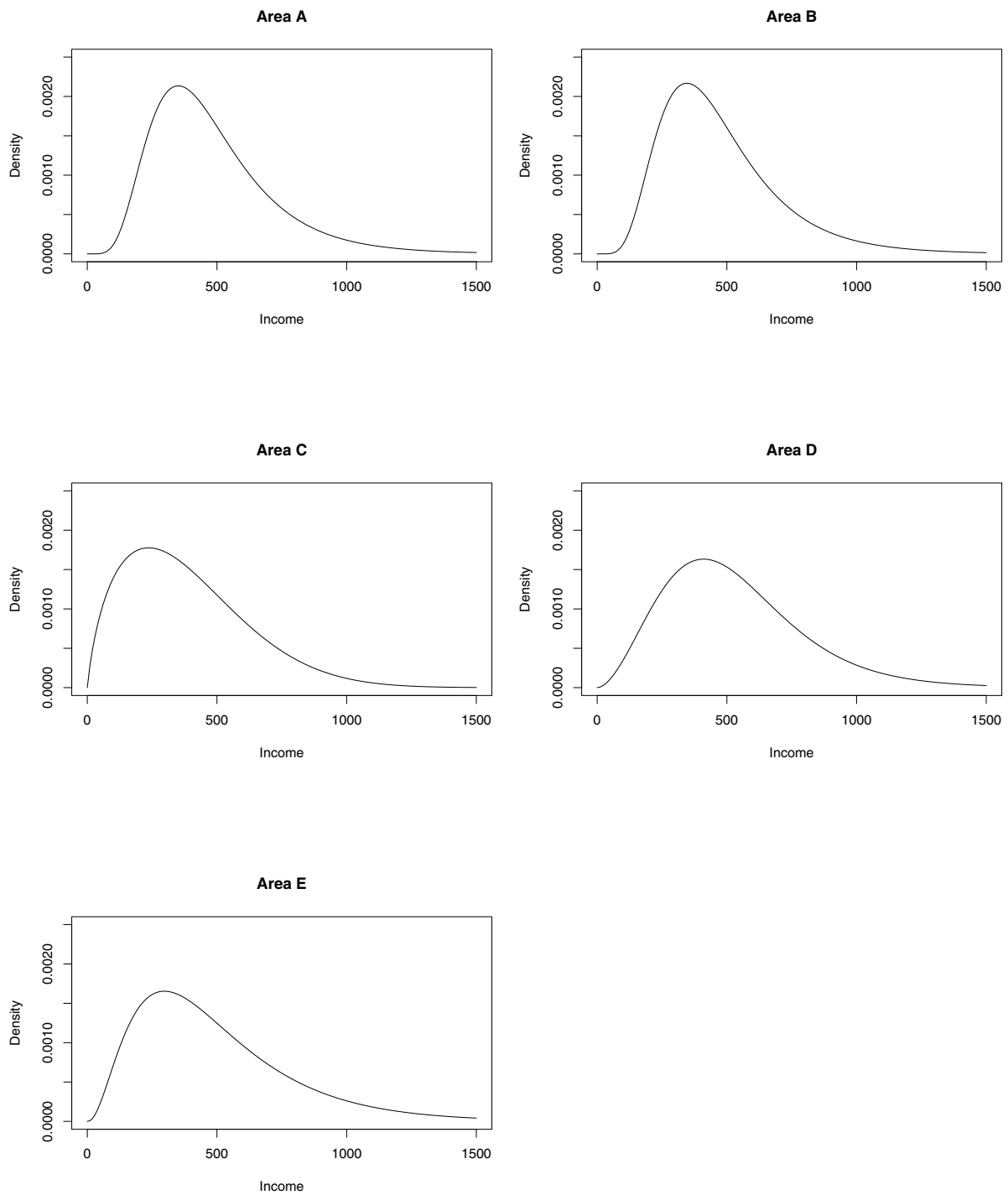


図 4: 各経済圏における 40 歳以下世帯の所得分布の推定結果

向があるが、平均所得の地域差は明確ではないと結論づけられる。地域内所得格差を Theil 指数で測った場合、経済圏 A・B・D よりも経済圏 C・E でその格差が大きい。低所得層の割合は、経済圏 A・B と比較して、経済圏 C・D・E でその割合が高い。また、他の地域と比較して経済圏 C の平均所得が低く、経済圏 D の平均所得は高い。

本研究では、福井 (2015) の方法を用いて、鹿児島県の各経済圏について、階層別データから所得と年齢の同時分布を推定し、各経済圏を対象に年齢層別の所得分布について詳細な分析を行った。実際に、60歳以下世帯・60歳より上の世帯・40歳以下世帯を対象に、平均所得と所得格差を定量化し分析した結果、それらの指標について一定程度の地域差が存在することが明らかとなった。

一方、各経済圏で標本が小さい・特定の地域において利用可能なデータが少ないなどの、データについての制限が存在し、それが推定結果に影響を及ぼしている可能性がある。この問題への対処については、今後の研究において引き続き考察したい。

また、以上の分析で明らかとなった平均所得・所得格差・低所得層の割合などの地域差については、それが地域の特性に真に由来するかを判断するための継続的な分析が必要である。上述の通り、各経済圏の標本が小さいことで、推定した同時分布と真の同時分布との間に乖離が生じ、平均所得・所得格差などの地域差に対してその乖離が強い影響を与えている可能性がある。そこで、より新しいデータに対しても本研究の分析を行い、その結果として同様の地域差が確認できれば、その地域差が地域の特性に由来する可能性が高いと判断できるだろう。

謝辞

本研究は JSPS 科研費 JP16K21558 の助成を受けたものです。

参考文献

- Diaconis, P. and B. Sturmfels (1998) "Algebraic Algorithms for Sampling from Conditional Distributions," *Annals of Statistics*, Vol. 26, No. 1, pp. 363-397.
- Farebrother, R. W. (1979) "Estimation with Aggregated Data," *Journal of Econometrics*, Vol. 10, pp. 43-55.
- Fukushige, M. and M. Hatanaka (1991) "Estimation of a Regression Model on Two or More Sets of Differently Grouped Data," *Journal of Econometrics*, Vol. 47, pp. 207-226.
- Haitovsky, Y. (1973) *Regression Estimation from Grouped Observations*: Griffin.
- Joe, H. (1997) *Multivariate Models and Dependence Concepts*: Chapman and Hall/CRC.
- Jondeau, E., S.-H. Poon, and M. Rockinger (2007) *Financial Modeling under Non-Gaussian Distributions*: Springer.
- Kleiber, C. and S. Kots (2003) *Statistical Size Distributions in Economics and Actuarial Sciences*: Wiley-Interscience.
- McDonald, J. B. (1984) "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, Vol. 52, No. 3, pp. 647-663.
- McDonald, J. B. and Y. J. Xu (1995) "A Generalization of the Beta Distribution with Applications,"

Journal of Econometrics, Vol. 69, No. 2, pp. 427-428.

Nelsen, R. B. (2006) *An Introduction to Copulas*: Springer.

Trivedi, P. K. and D. M. Zimmer (2005) "Copula Modeling: An Introduction for Practitioners,"
Foundations and Trends in Econometrics, Vol. 1, No. 1, pp. 1-111.

竹澤邦夫 (2007) 『みんなのためのノンパラメトリック回帰 (下)』, 吉岡書店, 第 3 版.

竹村彰通・青木敏 (2006) 「統計学におけるグレブナー基底」, 日比孝之 (編) 『グレブナー基底の現在』,
数学書房, 第 3 章, 60-84 頁.

西埜晴久 (2012) 「所得分布を推定するための 3 つ以上のパラメータを持つ分布」, 『千葉大学経済研究』,
第 27 卷, 321-338 頁.

福井昭吾 (2015) 「分割表が利用できない状況における世帯主の所得と年齢の同時分布の GMM 推定」,
『応用経済学研究』, 第 8 卷, 42-68 頁.