

集中曲線と弾力性計測法*

豊田 敬

1. はじめに

ローレンツ曲線は累積シェアをグラフ化した簡単な図法であって、不平等度や集中度を比較する際に広く利用されている。ローレンツ曲線の面積と関係のあるジニ係数もまた不平等尺度としてよく使われる。これらは、データに見られる不平等度といった側面の比較計測だけに用いられる特殊な方法とみなされがちであるが、標準偏差や変動係数と同様、データの広がりやを比較計測する方法でもあるという点に変わりはない。

本稿は後者の視点からの論考であり、その要点は次の二つからなる。第一点は、ローレンツ曲線とジニ係数の考え方を2変数データに適用すると、一連の弾力性計測法が得られるという事実を指摘することである。第二点は、それらの弾力性計測法はすべて、操作変数法で直線をあてはめて平均まわりの弾力性を計測する方法になるという推定法からの根拠を与えることである。

ローレンツ曲線が累積シェアという測定単位に関係しない量に基づく図法であることから、それが弾力性概念と関連するであろうという予想がたつ。実際、弾力性概念とローレンツ曲線の順序付け(ローレンツ順序)との一般的な関係は、Jakobsson(1976)、Kakwani(1977b)によって示されている。しかし、その関係は定性的であるがために、弾力性の計測法とは結びつけにくいという難点がある。ところが、1次式の場合の弾力性に限定すると、弾力性の計測法と操作変数法とが直接結びつく。これが本稿の主張である。直線のあてはめの方にウェイトをおいて、次のように言ってもよい。累積シェアの

曲線から操作変数法の一つのクラスが導出される。

本稿の構成は次のとおりである。2変数データには、ローレンツ曲線を含めて、累積シェアの曲線が3種類ある。この3つの集中曲線が本稿における考察のフレームワークになる。実証研究では、その区別があまり意識されず、ときに混同されていることもある。そこで、第2節では、集中曲線について少し詳しく説明する。第3節は、弾力性の計測法の図解であり、集中曲線を組み合わせた図を用いてそのことを指摘する。第4節では、ジニ係数等の集中指数を共分散形式の式で表現する。この結果を踏まえて、第3節で指摘した方法が操作変数法で直線をあてはめたときの平均まわりの弾力性になっているという結果を示す。それが第5節である。

2. 集中曲線(累積シェアの曲線)

本稿では、 x を正の説明変数、 y を正の被説明変数として、直線 $y = a + \beta x$ をあてはめる方法を念頭においている。具体的には租税関数(x :所得、 y :税負担額)やエンゲル関数(x :消費支出または可処分所得、 y :特定の支出費目)を考えればよい。

正の2変数 x, y のデータを $(x_i, y_i), i=1, \dots, n$ とし、その平均を $(\bar{x}, \bar{y})$ で表す。本稿では一貫して、 x のデータは昇順に並べられているとする。すなわち、 $0 < x_1 \leq x_2 \leq \dots \leq x_i \leq \dots \leq x_n$ である。もちろん、 y の方は昇順になるとは限らない。

変数 x の順序に従ったとき、このデータの累積シェアには次の3つがある：

$$U_k = \frac{k}{n}, V_k = \frac{\sum_{i=1}^k x_i}{n\bar{x}}, W_k = \frac{\sum_{i=1}^k y_i}{n\bar{y}},$$

$$k = 0, 1, \dots, n.$$

ただし、 $U_0 = V_0 = W_0 = 0$ とする。 U_k は度数の累積シェア(累積相対度数)、 V_k は昇順の変数 x の累積シェア、そして W_k は変数 x を昇順にしたときの変数 y の累積シェアであって、 $U_n = V_n = W_n = 1$ である。この3つの累積シェアに対して、3つのペア $\{U_k, V_k\}, \{U_k, W_k\}, \{V_k, W_k\}$ の組合せがある。この3つを図示すると、 (U_k, V_k) からローレンツ曲線、 (U_k, W_k) から集中曲線、 (V_k, W_k) から比率曲線といった広義の集中曲線が得られる¹⁾。以下、この3種類の曲線について説明する。

ローレンツ曲線

直交座標で

『 $n+1$ 個の点 $(U_k, V_k), k=0, 1, \dots, n$ を順次直線補完した曲線』

が変数 x のローレンツ曲線 L_x である(図1(a) 参照²⁾)。

$U_0 = V_0 = 0, U_n = V_n = 1$ により、この L_x は点 $(0, 0)$ と点 $(1, 1)$ を結ぶ曲線であることが直ちにわかる。また、 L_x の各線分の傾きが

$$\frac{V_k - V_{k-1}}{U_k - U_{k-1}} = \frac{x_k}{\bar{x}}, \quad k=1, \dots, n$$

であることから、データ x_k が昇順である点に注意すれば、 L_x は下に凸で単調増加の折れ線であることがわかる。さらに、その傾きは、平均より低い階層に対応する部分では1より小さく、平均より高い階層部分では1より大きいこともわかる。そして、 $x_1 = \dots = x_n$ のとき、 $U_k = V_k = k/n, k=0, \dots, n$ となるので、よく知られているとおり、点 $(0, 0)$ と点 $(1, 1)$ を結ぶ45度線は完全均等線を表す。ローレンツ曲線を上からの累積シェアとして捉えることもできる。この場合のローレンツ曲線は

『点 $(1-U_k, 1-V_k), k=n, n-1, \dots, 0$ を順次直線補完した曲線』

となる。このローレンツ曲線は、先のローレンツ曲線を点 $(0.5, 0.5)$ を中心にして180度回転したものにはかならず、単調増加で上に凸である。

ローレンツ曲線による不平等度や集中度の比較は、各 $k=1, \dots, n$ における累積シェア V_k が

小さいほど不平等度は大きい、すなわち、完全均等線の45度線からの隔たりが大きいほど不平等度は大きいという基準でなされる。図形的に言えば、図1(a)における凸レンズ状の部分が厚くふくらむほど不平等度は大きいという基準である³⁾。

ローレンツ曲線を提案した Lorenz(1905)で既に指摘されているとおり、ローレンツ曲線が交差するケースについては不平等比較ができないという意味で、ローレンツ順序は半順序である。

凸レンズ状の部分がふくらむほど不平等度が大きいという基準に従えば、その部分の面積を不平等度や集中度の指標とする方式が自然に導かれる。それがジニ係数 G_x である。ジニ係数を、完全均等線とローレンツ曲線とに囲まれた弓形の面積の2倍と定義してもさしつかえない。なお、ジニ係数等の集中指数の表現式については第4節で述べる。

完全均等線とローレンツ曲線 L_x との隔たりは、 U_k と V_k との差

$$U_k - V_k = \frac{k}{n} - \frac{\sum_{i=1}^k x_i}{n\bar{x}} = \frac{\sum_{i=1}^k (\bar{x} - x_i)}{n\bar{x}}, \quad k=1, \dots, n \quad (1)$$

である。したがって、 L_x においては、平均からの偏差 $x_i - \bar{x}, i=1, \dots, n$ の部分(もしくは累積和)

$\sum_{i=1}^k (\bar{x} - x_i) = \sum_{i=k+1}^n (x_i - \bar{x}), k=1, \dots, n$ がデータの広がり(散らばり)を測る基本量である。

集中曲線

次に、変数 y の集中曲線 $L_{y,x}$ は、直交座標で

『 $n+1$ 個の点 $(U_k, W_k), k=0, 1, \dots, n$ を順次直線補完した曲線』

として定義される(図1(b) 参照)。

y のデータを昇順、降順に並べたものを、それぞれ

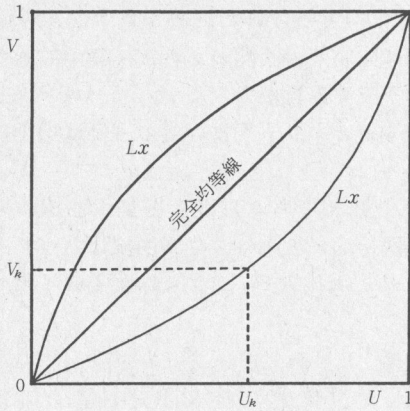
$$y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}, \quad y_{[1]} \geq y_{[2]} \geq \dots \geq y_{[n]}$$

で書き表せば、

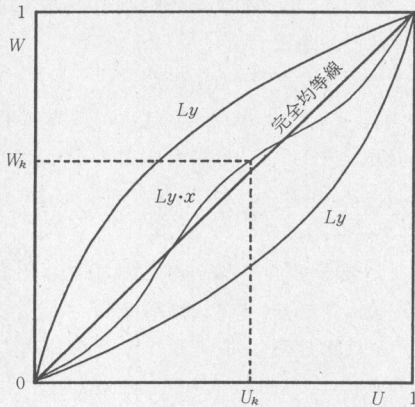
$$\begin{aligned} \frac{\min_p \sum_{i=1}^k y_{pi}}{n\bar{y}} &= \frac{\sum_{i=1}^k y_{(i)}}{n\bar{y}} \\ &\leq W_k \leq \frac{\max_p \sum_{i=1}^k y_{pi}}{n\bar{y}} = \frac{\sum_{i=1}^k y_{[i]}}{n\bar{y}}, \quad k=1, \dots, n \end{aligned}$$

図1 集中曲線

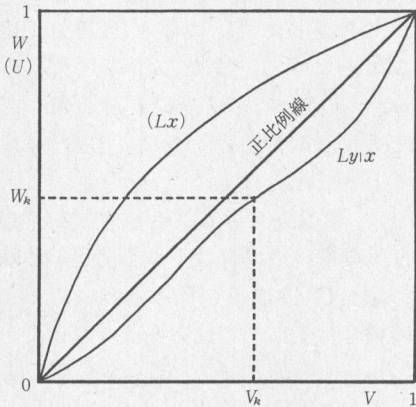
(a) ローレンツ曲線 L_x



(b) 集中曲線 $L_{y \cdot x}$



(c) 比率曲線 $L_{y/x}$



ただし、 $p = (p_1, \dots, p_i, \dots, p_n)$ は $(1, \dots, i, \dots, n)$ の順列、が成り立つ。この不等式から次の事実がわかる。

集中曲線 $L_{y \cdot x}$ は変数 y のローレンツ曲線 L_y と上からの累積シェアをとった y のローレンツ曲線との間にあり、 x と y の順位相関係数の値が ± 1 のとき、集中曲線 $L_{y \cdot x}$ はローレンツ曲線 L_y と一致する（正確には順位相関係数の値が -1 のときは上からの累積シェアをとった L_y ）。つまり、ローレンツ曲線 L_y は変数 y の累積シェアの列の最小部分とと最大部分とに対応するグラフであり、集中曲線 $L_{y \cdot x}$ はこの2つの曲線で囲まれた凸レンズ状の図形のなかを縫う曲線である。

集中曲線 $L_{y \cdot x}$ の場合も45度線は完全均等線を表す。それは、 $y_1 = \dots = y_n$ のとき、 $U_k = W_k = k/n$ となるからである。もとより、ここでの完全均等は、 x の変化に対して y が反応しないという意味においてである。そして、変数 x と y の相関が正の場合には完全均等を表す45度線より下になる可能性が大きく、相関が負の場合にはこの45度線より上になる可能性が大きいといった意味で、集中曲線 $L_{y \cdot x}$ は一種の相関関係を表現している。

この $L_{y \cdot x}$ の各線分の傾きは、形式的にはローレンツ曲線 L_x と同様で

$$\frac{W_k - W_{k-1}}{U_k - U_{k-1}} = \frac{y_k}{\bar{y}}, \quad k = 1, \dots, n$$

であって、その傾きは正である。ただし、 y_k は変数 x の順序で並べられていて昇順であることが保障されないので、傾きの単調性まではいえない。

ジニ係数と同様にして、45度の完全均等線と集中曲線 $L_{y \cdot x}$ とに囲まれた図形の面積の2倍の指標を考えることができる。それを集中係数と呼び、 $G_{y \cdot x}$ で表す。ただし、面積の符号は、完全均等線より下になっている部分を正、完全均等線より上になっている部分を負とする。

完全均等線と、 $L_{y \cdot x}$ との隔たりは、

$$U_k - W_k = \frac{k}{n} - \frac{\sum_{i=1}^k y_i}{n\bar{y}} = \frac{\sum_{i=1}^k (\bar{y} - y_i)}{n\bar{y}}, \quad k = 1, \dots, n \quad (2)$$

であって、 y が必ずしも昇順になっていないという点を除けば、これも形式的にはローレンツ曲線のときと変わりはない。

特にデータに厳密な線形関係

$$y_i = \alpha + \beta x_i, \quad i = 1, \dots, n,$$

があるとすれば、 $\bar{y} = \alpha + \beta \bar{x}$ であって、

$$\begin{aligned} U_k - W_k &= \frac{\beta \sum_{i=1}^k (\bar{x} - x_i)}{n \bar{y}} \\ &= \beta (\bar{x} / \bar{y}) (U_k - V_k), \\ k &= 1, \dots, n \end{aligned} \quad (2')$$

が成り立つから、変数 y の集中曲線 $L_{y \cdot x}$ は、完全均等線を基準にしてローレンツ曲線 L_x を縦軸方向に $\beta (\bar{x} / \bar{y})$ 倍したものに一致する。このときの集中曲線 $L_{y \cdot x}$ は y のローレンツ曲線 L_y と一致することはいうまでもない(ただし、 $\beta < 0$ ときは上からの累積シェアをとった L_y)。 (2') 式は第3節の議論の要になる関係式である。

ところで、実証研究では度数分布の形に組み分けされた統計を使うことが多い。いま、

x の階層区分	度数	x の階層平均	y の階層平均
C_1	f_1	x_1	y_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_i	f_i	x_i	y_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_n	f_n	x_n	y_n

のような形にデータがまとめられているとすれば、各累積シェアは

$$U_k = \frac{\sum_{i=1}^k f_i}{\sum_{i=1}^n f_i}, \quad V_k = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^n x_i f_i}, \quad W_k = \frac{\sum_{i=1}^k y_i f_i}{\sum_{i=1}^n y_i f_i},$$

$$k = 0, \dots, n$$

となる。この場合の集中曲線 $L_{y \cdot x}$ は、変数 x で階層化された順序に従ったときの変数 y の累積シェアを表す。たとえば、「家計調査」における年間収入階級別の統計を使う際、 (U_k, V_k, W_k) はそれぞれ累積世帯数割合、年間収入の累積シェア、年間収入以外の収支項目の累積シェアとなるので、ローレンツ曲線は年間収入だけであって、他の収支項目に関するものはすべて集中曲線になる。このようなわけで、利用可能な統計データに制約のある実証研究では、ローレンツ曲線と呼んでいるもののうちには、事実上は本稿でいう集中曲線であるものが多い。

組み分けされたデータの集中曲線で不平等度や集中度の分析をする場合、次の点に注意する必要がある。

a) 階層化に用いられた変数と集中曲線の変数との間に強い正の相関があれば、集中曲線によっても不平等分析ができるが、ローレンツ曲線に基づくときよりも不平等度の評価は過小になる。

b) 階層化に用いられた変数と集中曲線の変数との相関が弱いときは、集中曲線に基づく不平等分析はミスリーディングな結果を導く可能性が強い。

比率曲線

変数 y の変数 x に対する比率曲線 $L_{y|x}$ は、直交座標で

『 $n+1$ 個の点 (V_k, W_k) , $k=0, 1, \dots, n$ を順次直線補完した曲線』

として定義される(図1(c) 参照)。

この $L_{y|x}$ も点 $(0, 0)$ と点 $(1, 1)$ を結ぶ単調増加の曲線であって、その各線分の傾きは

$$\frac{W_k - W_{k-1}}{V_k - V_{k-1}} = \frac{y_k / \bar{y}}{x_k / \bar{x}} = \frac{y_k / x_k}{\bar{y} / \bar{x}}, \quad k = 1, \dots, n$$

である。各線分のこの傾きが、対応する各個別の比率 y_k / x_k と全体の比率 $\bar{y} / \bar{x} = (\sum_{i=1}^n y_i) / (\sum_{i=1}^n x_i)$ との比を表していることから、次の事実がわかる。

2つの変数とが正比例しているとき、すべての線分の傾きは1となって、比率曲線 $L_{y|x}$ は45度線になる。そこで、この45度線を正比例線と呼ぶことにする。また、 x_k の増大に伴って比率 y_k / x_k が増加するとき、 $L_{y|x}$ は下に凸となり、そして x_k の増大に伴って比率 y_k / x_k が減少するとき、 $L_{y|x}$ は上に凸となる。

たとえば、変数 x を所得、変数 y を税負担額とすれば、累進税の場合は下に凸の比率曲線に、比例税の場合は45度線の比率曲線に、そして逆進税の場合は上に凸の比率曲線になる⁴⁾。

なお、 $y_1 = \dots = y_n$ のとき $W_k = k/n = U_k$ となるから、比率曲線 $L_{y|x}$ はローレンツ曲線 L_x に帰着する。ただし、座標点は (U_k, V_k) ではなく (V_k, U_k) である。また、 $y_1 \geq \dots \geq y_n$ の場合、比率曲線は図1(c)の L_x より上に位置する。

比率曲線 $L_{y|x}$ においても、ジニ係数や集中係数の考え方を適用して、45度の正比例線と比率曲線とに囲まれた図形の面積の2倍で定義される指標を比率係数と呼んで、 $G_{y|x}$ で書き表すことにする。正比例線より下の部分の面積の符号を正とし、正比例線より上の部分の面積の符号を負とするのも同じである。

正比例線と比率曲線 $L_{y|x}$ との隔たりは

$$V_k - W_k = \frac{\sum_1^k x_i}{n\bar{x}} - \frac{\sum_1^k y_i}{n\bar{y}} = \frac{\sum_1^k (x_i\bar{y} - \bar{x}y_i)}{n\bar{x}\bar{y}}, \quad k=1, \dots, n \quad (3)$$

である。これが正比例関係からの乖離を測る量であるという点については、次のように考えるとう理解しやすい。

(1), (2) 式より、

$$\begin{aligned} V_k - W_k &= (U_k - W_k) - (U_k - V_k) \\ &= \frac{\sum_1^k (\bar{y} - y_i)}{n\bar{y}} - \frac{\sum_1^k (\bar{x} - x_i)}{n\bar{x}} \end{aligned}$$

であるから、データに厳密な線形関係 $y_i = \alpha + \beta x_i$ があると想定すると、

$$V_k - W_k = \frac{-\alpha \sum_1^k (\bar{x} - x_i)}{n\bar{x}\bar{y}} = \frac{-\alpha}{\bar{y}} (U_k - V_k), \quad k=1, \dots, n \quad (3')$$

が成り立つ。ここで $(U_k - V_k) \geq 0$ であるから、比率曲線は、正比例 ($\alpha=0$) のとき、正比例線の45度線に一致し ($V_k = W_k$)、 $\alpha > 0$ のとき、45度線より上に ($V_k < W_k$)、 $\alpha < 0$ のとき、45度線より下に ($V_k > W_k$) 位置する。そして、 α の絶対値が大きくなると、比率曲線は正比例線から遠ざかる。この (3') 式も第3節の議論で使われる。

(3) 式を行列式で表示して次のように解釈することもできる。基本量にあたる分子は

$$\begin{aligned} \sum_1^k (x_i\bar{y} - \bar{x}y_i) &= \sum_1^k \begin{vmatrix} x_i & y_i \\ \bar{x} & \bar{y} \end{vmatrix} \\ &= \sum_{k+1}^n (\bar{x}y_i - x_i\bar{y}) = -\sum_{k+1}^n \begin{vmatrix} x_i & y_i \\ \bar{x} & \bar{y} \end{vmatrix} \end{aligned}$$

と表される。行列式の部分は2つのベクトル (x_i, y_i) と $(\bar{x}, \bar{y})$ の張る平行四辺形の面積に等しく、 x と y とが正比例していれば、その交角は0となって平行四辺形の面積も0になる。また、 $y_i/x_i < \bar{y}/\bar{x}$ のとき、つまり個別比率が全体比率より小さいときは、その面積は正であり、

$y_i/x_i > \bar{y}/\bar{x}$ のとき、つまり個別比率が全体比率より大きいときは、その面積は負である⁵⁾。

3. 弾力性の計測法と直線のあてはめ(図解)

3種類の曲線、ローレンツ曲線 L_x 、集中曲線 $L_{y \cdot x}$ 、比率曲線 $L_{y|x}$ に対して、3組のペア $\{L_x, L_{y \cdot x}\}$, $\{L_x, L_{y|x}\}$, $\{L_{y \cdot x}, L_{y|x}\}$ がある。このペアから、それぞれ弾力性の計測法を導くことができる。それを図によって示すことにする。

1次式を $y = \alpha + \beta x$ とすると、 y の x に対する弾力性 η_y は

$$\eta_y = (dy/dx)(x/y) = \beta(x/y)$$

である。2変数データ (x_i, y_i) , $i=1, \dots, n$ がこの直線上に乗っていると仮定すれば、

$$y_i = \alpha + \beta x_i, \quad i=1, \dots, n; \quad \bar{y} = \alpha + \beta \bar{x}$$

であり、平均まわりの弾力性は $\eta = \beta(\bar{x}/\bar{y})$ となる。したがって、(2') 式より

$$\eta = \frac{U_k - W_k}{U_k - V_k} = \beta \frac{\bar{x}}{\bar{y}}, \quad k=1, \dots, n-1$$

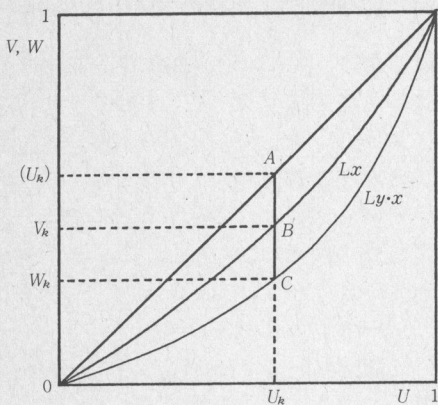
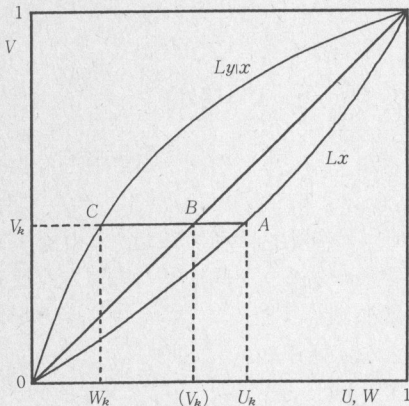
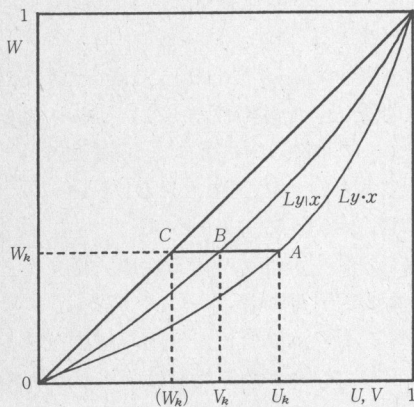
が成り立つ。つまり、データが直線上に並んでいるとき、比率 $(U_k - W_k)/(U_k - V_k)$ の値は平均まわりの弾力性の値を表し、 k の値に関係しない定数である。平均まわりの弾力性計測法はこの関係に着目することによって導出される。さらに、(3') 式より

$$\frac{V_k - W_k}{U_k - V_k} = -\frac{\alpha}{\bar{y}}, \quad k=1, \dots, n-1$$

が成り立つが、この比率も k の値に関係しない定数である。

図で考察することにしよう。図2(a)はローレンツ曲線 L_x と集中曲線 $L_{y \cdot x}$ を組み合わせた図である。横軸に共通の U 、縦軸に V, W をとってある。この図で、 $AB = U_k - V_k$, $AC = U_k - W_k$, $BC = V_k - W_k$ であるから、比率 AC/AB が平均まわりの弾力性に対応する。また、比率 BC/AB からは $-\alpha/\bar{y}$ に対応する値を読み取ることができる。厳密な線形関係が成り立つとき、この2つの比率は k に関係しない定数になるという事実から、次の基準が示唆される。すなわち、データに直線をあてはめる場合、それぞれの推定値として面積比を対応させればよい、という基準である。この面積比基準

図2 集中曲線のペア

(a) $\{Lx, Ly \cdot x\}$ (b) $\{Lx, Ly/x\}$ (c) $\{Ly \cdot x, Ly/x\}$ 

を適用すれば、図2(a)により、平均まわりの弾力性 η の計測値を

$$\hat{\eta}_1 = G_{y \cdot x} / G_x, \quad (4)$$

そして、 β, α の推定値を

$$\hat{\beta}_1 = (G_{y \cdot x} / G_x) (\bar{y} / \bar{x}), \quad (4')$$

$$\hat{\alpha}_1 = ((G_x - G_{y \cdot x}) / G_x) \cdot \bar{y}$$

とする方式が自然なかたちで導かれる。

次に、図2(b)はローレンツ曲線 Lx と比率曲線 Ly/x を組み合わせた図であって、縦軸に共通の V 、横軸に U, W がとってある。この図でも、 $AB = U_k - V_k$ 、

$$\begin{aligned} AC &= AB + BC = (U_k - V_k) + (V_k - W_k) \\ &= U_k - W_k \end{aligned}$$

であるから、比率 AC/AB が平均まわりの弾力性を指し、比率 BC/AB が $-\alpha/\bar{y}$ を指す。これにも面積比基準を適用すれば、弾力性 η の計測値、 β, α の推定値はそれぞれ、

$$\hat{\eta}_2 = (G_x + G_{y/x}) / G_x, \quad (5)$$

$$\hat{\beta}_2 = ((G_x + G_{y/x}) / G_x) (\bar{y} / \bar{x}), \quad (5')$$

$$\hat{\alpha}_2 = -(G_{y/x} / G_x) \cdot \bar{y}$$

となる。

最後の図2(c)は集中曲線 $Ly \cdot x$ と比率曲線 Ly/x を組み合わせた図である。この図においては、

$$\begin{aligned} AB &= AC - BC = (U_k - W_k) - (V_k - W_k) \\ &= U_k - V_k, \end{aligned}$$

であるから、同様にして比率 AC/AB が平均まわりの弾力性に、そして、比率 BC/AB が $-\alpha/\bar{y}$ に対応する。したがって、弾力性 η の計測値、 β, α の推定値は、面積比基準を適用して、それぞれ、

$$\hat{\eta}_3 = G_{y \cdot x} / (G_{y \cdot x} - G_{y/x}), \quad (6)$$

$$\hat{\beta}_3 = (G_{y \cdot x} / (G_{y \cdot x} - G_{y/x})) (\bar{y} / \bar{x}), \quad (6')$$

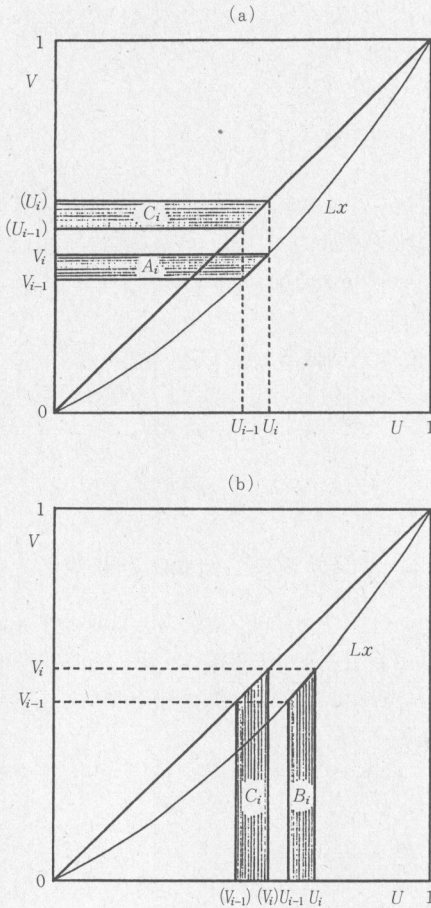
$$\hat{\alpha}_3 = (G_{y/x} / (G_{y/x} - G_{y \cdot x})) \cdot \bar{y}$$

である。

以上に挙げた弾力性の計測法 (4), (5), (6) の相違は、結局のところ1次式の推定法の相違に帰する。この3つの弾力性の計測法は、それぞれ (4'), (5'), (6') によって1次式を推定したときの弾力性になっているからである。

ところで、図解からも予想されるように、推定された1次式はいずれも平均の点を通る。つ

図3 ジニ係数



まり,

$$\hat{\alpha}_j = \bar{y} - \hat{\beta}_j \bar{x}, \quad j = 1, 2, 3$$

が成り立つ。したがって、 $\hat{\beta}_j, j = 1, 2, 3$ の推定法の違いを検討すれば十分である。そのためには、準備としてジニ係数 G_x 、集中係数 $G_{y \cdot x}$ 、比率係数 $G_{y|x}$ がいずれも共分散形式で表現できることを示す必要がある。それが次節の課題である。

4. 集中指数

完全均等線とローレンツ曲線で囲まれた弓形の面積で格差・不平等度を測る指標とするというジニ係数の考え方を、集中曲線と比率曲線に適用したのが集中係数 $G_{y \cdot x}$ と比率係数 $G_{y|x}$ で

あった。これら3つの係数を総称して集中指数と呼ぶことにする。

ジニ係数

ジニ係数は、多様な形の数式で書き表わすことができる。ここでは図を利用して、共分散の形の表現式を2つ導く⁶⁾。

まず、図3(a)から始める。この図で、台形の面積 A_i の総和の2倍 ($2\sum A_i$) から台形の面積 C_i の総和の2倍 ($2\sum C_i = 1$) を引算すれば、弓形の面積の2倍、すなわちジニ係数 G_x が求められる：

$$2\sum^n A_i = \sum^n (U_i + U_{i-1})(V_i - V_{i-1})$$

$$2\sum^n C_i = \sum^n (U_i + U_{i-1})(U_i - U_{i-1})$$

であるから、

$$\begin{aligned} G_x &= \sum^n (U_i + U_{i-1}) \{ (V_i - V_{i-1}) \\ &\quad - (U_i - U_{i-1}) \} \\ &= \sum^n (U_i + U_{i-1}) \{ (U_{i-1} - V_{i-1}) \\ &\quad - (U_i - V_i) \} \end{aligned} \quad (7)$$

が得られる。これが第一の表現形式である。

次に、図3(b)に移る。この図では、台形の面積 C_i の総和の2倍 ($2\sum C_i = 1$) から台形の面積 B_i の総和の2倍 ($2\sum B_i$) を引算すれば、ジニ係数 G_x が求められる：

$$2\sum^n B_i = \sum^n (V_i + V_{i-1})(U_i - U_{i-1})$$

$$2\sum^n C_i = \sum^n (V_i + V_{i-1})(V_i - V_{i-1})$$

であるから、第二の表現形式

$$\begin{aligned} G_x &= \sum^n (V_i + V_{i-1}) \{ (V_i - V_{i-1}) \\ &\quad - (U_i - U_{i-1}) \} \\ &= \sum^n (V_i + V_{i-1}) \{ (U_{i-1} - V_{i-1}) \\ &\quad - (U_i - V_i) \} \end{aligned} \quad (8)$$

が得られる。(1)式より

$$\begin{aligned} (V_i - V_{i-1}) - (U_i - U_{i-1}) &= (x_i / (n\bar{x})) \\ &\quad - (1/n) = (x_i - \bar{x}) / (n\bar{x}) \end{aligned}$$

が成り立つので、(7)、(8)式は、それぞれ

$$G_x = (1/n\bar{x}) \sum^n (U_i + U_{i-1})(x_i - \bar{x}) \quad (7')$$

$$G_x = (1/n\bar{x}) \sum^n (V_i + V_{i-1})(x_i - \bar{x}) \quad (8')$$

となる。

ところで、一般に、2変数 z と x の共分散は

$$\begin{aligned} cov(z, x) &= (1/n) \sum^n (z_i - \bar{z})(x_i - \bar{x}) \\ &= (1/n) \sum^n z_i (x_i - \bar{x}) \end{aligned}$$

で定義されるから、(7')式は

$$z_i = (U_i + U_{i-1}) / \bar{x} = (2i - 1) / (n\bar{x}),$$

$$i = 1, \dots, n$$

と置いたときの z と昇順に並べた x の共分散であることを意味し、(8') 式は

$$z_i = (V_i + V_{i-1})/\bar{x} = (2\sum_{i=1}^n x_i - x_i)/(n\bar{x}),$$

$$i = 1, \dots, n$$

と置いたときの z と昇順に並べた x の共分散であることを意味する。

$$\sum_{i=1}^n (2i-1)(x_i - \bar{x}) = 2\sum_{i=1}^n i(x_i - \bar{x})$$

が成り立つことに留意すると、(7') 式は、変数 x とその順位の共分散を総和 $\sum_{i=1}^n x_i$ で除して 2 倍した量と解釈することもできる。

集中係数と比率係数

集中係数と比率係数は、ジニ係数と同様に考えればよいので簡単である。図 3 で、 V を W に置き換えれば、集中曲線 $L_{y,x}$ になる。したがって集中係数 $G_{y,x}$ は、(7), (8) 式から直ちに

$$G_{y,x} = \sum_{i=1}^n (U_i + U_{i-1}) \{ (W_i - W_{i-1}) - (U_i - U_{i-1}) \}$$

$$= \sum_{i=1}^n (U_i + U_{i-1}) \{ (U_{i-1} - W_{i-1}) - (U_i - W_i) \} \quad (9)$$

$$G_{y,x} = \sum_{i=1}^n (W_i + W_{i-1}) \{ (W_i - W_{i-1}) - (U_i - U_{i-1}) \}$$

$$= \sum_{i=1}^n (W_i + W_{i-1}) \{ (U_{i-1} - W_{i-1}) - (U_i - W_i) \} \quad (10)$$

である。さらに (2) 式により

$$(W_i - W_{i-1}) - (U_i - U_{i-1}) = (y_i / (n\bar{y}))$$

$$- (1/n) = (y_i - \bar{y}) / (n\bar{y})$$

が成り立つから、(7'), (8') 式は

$$G_{y,x} = (1/n\bar{y}) \sum_{i=1}^n (U_i + U_{i-1}) (y_i - \bar{y}) \quad (9')$$

$$G_{y,x} = (1/n\bar{y}) \sum_{i=1}^n (W_i + W_{i-1}) (y_i - \bar{y}) \quad (10')$$

となって両式ともに共分散の形の式になる。

次に、図 3 で U, V をそれぞれ V, W に置き換えると比率曲線 $L_{y|x}$ になるので、(7), (8) 式により、比率係数 $G_{y|x}$ は

$$G_{y|x} = \sum_{i=1}^n (V_i + V_{i-1}) \{ (W_i - W_{i-1}) - (V_i - V_{i-1}) \}$$

$$= \sum_{i=1}^n (V_i + V_{i-1}) \{ (V_{i-1} - W_{i-1}) - (V_i - W_i) \} \quad (11)$$

$$G_{y|x} = \sum_{i=1}^n (W_i + W_{i-1}) \{ (W_i - W_{i-1}) - (V_i - V_{i-1}) \}$$

$$= \sum_{i=1}^n (W_i + W_{i-1}) \{ (V_{i-1} - W_{i-1}) - (V_i - W_i) \} \quad (12)$$

である。

ここで、比率係数と他の 2 つの係数との関係を表す式を示す。

$$(V_{i-1} - W_{i-1}) - (V_i - W_i)$$

$$= \{ (U_{i-1} - W_{i-1}) - (U_i - W_i) \}$$

$$- \{ (U_{i-1} - V_{i-1}) - (U_i - V_i) \}$$

$$= (1/n\bar{y})(y_i - \bar{y}) - (1/n\bar{x})(x_i - \bar{x})$$

と (11), (8') 式より

$$G_{y|x} = (1/n\bar{y}) \sum_{i=1}^n (V_i + V_{i-1}) (y_i - \bar{y}) - G_x \quad (13)$$

が成り立つ。同様に (12), (10') 式より

$$G_{y|x} = G_{y,x}$$

$$- (1/n\bar{x}) \sum_{i=1}^n (W_i + W_{i-1}) (x_i - \bar{x}) \quad (14)$$

が成り立つ。

5. 直線のあてはめ(操作変数法)

変数 x に関して昇順に並べた 2 変数データ $(x_i, y_i), i = 1, \dots, n$ に直線 $y = a + \beta x$ をあてはめる方式として操作変数法を採用すれば、 β の推定値 b は

$$b = \frac{\text{cov}(z, y)}{\text{cov}(z, x)} = \frac{\sum (z_i - \bar{z})(y_i - \bar{y})}{\sum (z_i - \bar{z})(x_i - \bar{x})}$$

$$= \frac{\sum z_i (y_i - \bar{y})}{\sum z_i (x_i - \bar{x})}$$

と書き表される。ただし、 z は操作変数であり、特に $z_i = x_i$ と置くと普通の最小 2 乗法に帰着する。

第 3 節で得た 3 つの推定値 $\hat{\beta}_j, j = 1, 2, 3$ はどれもこの b の型になる。それを示すことにしよう。

まず、(4'), (7'), (9') より

$$\hat{\beta}_1 = \left(\frac{G_{y,x}}{G_x} \right) (\bar{y}/\bar{x})$$

$$= \frac{\sum_{i=1}^n (U_i + U_{i-1}) (y_i - \bar{y})}{\sum_{i=1}^n (U_i + U_{i-1}) (x_i - \bar{x})}$$

$$= \frac{\sum_{i=1}^n i (y_i - \bar{y})}{\sum_{i=1}^n i (x_i - \bar{x})}$$

となる。これは、変数 x の順位を操作変数としたケースで、Durbin の操作変数法である⁸⁾。よって、(4) の弾力性計測法は、Durbin の操作変数法で 1 次式をあてはめたときの平均まわりの弾力性を計測する方法ということになる。

$\hat{\beta}_2, \hat{\beta}_3$ については次のとおりである。

(5'), (8'), (13) より

$$\begin{aligned}\hat{\beta}_2 &= \left(\frac{G_x + G_{y|x}}{G_x} \right) (\bar{y}/\bar{x}) \\ &= \frac{\sum_1^n (V_i + V_{i-1}) (y_i - \bar{y})}{\sum_1^n (V_i + V_{i-1}) (x_i - \bar{x})}\end{aligned}$$

(6') (10') (14) より

$$\begin{aligned}\hat{\beta}_3 &= \left(\frac{G_{y|x}}{G_{y \cdot x} - G_{y|x}} \right) (\bar{y}/\bar{x}) \\ &= \frac{\sum_1^n (W_i + W_{i-1}) (y_i - \bar{y})}{\sum_1^n (W_i + W_{i-1}) (x_i - \bar{x})}\end{aligned}$$

よって、(5), (6) の弾力性計測法は、操作変数法によって1次式をあてはめたときの平均まわりの弾力性を計測する方法である。そしてその違いは、操作変数として x の累積シェアを使った

$$z_i = V_i + V_{i-1} = (2 \sum_{j=1}^i x_j - x_i) / (n\bar{x}), \quad i = 1, \dots, n$$

を選択するか、 y の累積シェアを使った

$$z_i = W_i + W_{i-1} = (2 \sum_{j=1}^i y_j - y_i) / (n\bar{y}), \quad i = 1, \dots, n$$

を選択するかにある。

ここで、(2') の関係を直接適用して得られる操作変数法についても簡単に触れておく⁹⁾。

(2') より

$$\begin{aligned}\hat{\beta}_4(k) &= \frac{U_k - W_k}{U_k - V_k} (\bar{y}/\bar{x}) = \frac{\sum_1^k (\bar{y} - y_i)}{\sum_1^k (\bar{x} - x_i)} \\ &= \frac{\sum_{k+1}^n (y_i - \bar{y})}{\sum_{k+1}^n (x_i - \bar{x})}\end{aligned}$$

$k=1, \dots, n-1$, なる型の推定値が考えられる。

これは $z_1 = \dots = z_k = -1, z_{k+1} = \dots = z_n = 0$, あるいは $z_1 = \dots = z_k = 0, z_{k+1} = \dots = z_n = 1$ と置いた操作変数法に相当する。 k に関する分母の最大値は平均に近い順位のときであるから、

$$\begin{aligned}\max_k \sum_1^k (\bar{x} - x_i) &= \sum_1^m (\bar{x} - x_i) \\ &= (\sum_1^n |x_i - \bar{x}|) / 2,\end{aligned}$$

ただし、 m は $x_m \leq \bar{x} < x_{m+1}$ なる x の順位であって、このタイプの操作変数と x の共分散を最大にする推定値は $\hat{\beta}_4(m)$ である。

ローレンツ曲線 L_x における凸レンズ部分の厚さを縦軸方向に測ることにすれば、(2') より次の推定値が得られる。

$$\hat{\beta}_5(k) = \frac{(U_k - W_k) + (U_{n-k} - W_{n-k})}{(U_k - V_k) + (U_{n-k} - V_{n-k})} (\bar{y}/\bar{x})$$

$$= \frac{(y_{n-k+1} + \dots + y_n) - (y_1 + \dots + y_k)}{(x_{n-k+1} + \dots + x_n) - (x_1 + \dots + x_k)}, \quad k = 1, \dots, n-1.$$

この分母は、上位 $100 \times (k/n) \%$ の和と下位 $100 \times (k/n) \%$ の和との格差であって、操作変数法を、 $z_1 = \dots = z_k = -1, z_{n-k+1} = \dots = z_n = 1$, そして、 $k < [n/2]$ のとき、 $z_{k+1} = \dots = z_{n-k} = 0$ と置いたケースにあたる(ただし、 $[]$ はガウス記号)。 k に関する分母の最大値は $k = [n/2]$ のときで、このとき

$$\begin{aligned}& (x_{[(n+3)/2]} + \dots + x_n) - (x_1 + \dots + x_{[n/2]}) \\ &= (\sum_1^n |x_i - med|) / 2,\end{aligned}$$

ただし、 med はメディアンであり、Waldの方法になる。また、 $k = [n/3]$ のときは、Bartlettの方法になる。

6. むすび

本稿では、集中曲線を利用すると一連の弾力性計測法が得られるが、それらはすべて、操作変数法で1次式をあてはめたときの平均まわりの弾力性の計測法に帰する、という事実を指摘した。これは、集中曲線から得られる操作変数法の一つのクラスを提示したということでもある。

最後に、租税関数への応用例として、Kakwani(1977a)とSuits(1977)の累進度尺度について述べておこう。 x を所得、 y を税負担額とし、Kakwani 尺度を K , Suits 尺度を S で表す。本稿の記号を使えば、

$$K = G_{y \cdot x} - G_x, \quad S = G_{y|x}$$

である。この2つの尺度を修正して

$$K^* = (K/G_x) + 1, \quad S^* = (S/G_x) + 1$$

とすれば、(4), (5) により両者とも弾力性(税負担累進度)を測っていることになる¹⁰⁾。修正した新尺度の K^*, S^* は元の K, S に比べて次の利点をもっている。

a) 計測値が弾力性であるから、その数値自体に意味がある。

b) 1次式の租税関数を想定することが妥当でない場合、 K^*, S^* で累進度を計測するのには無理があるという点で尺度の適用限界がはっきりしている。

(法政大学経営学部)

注

* 本稿は田口時夫教授(統計数理研究所)に負うところが多い。記してお礼申し上げる。本稿には、平成4年度法政大学特別研究助成を得た部分がある。

1) 本稿では、これら3種の曲線の総称としても集中曲線という用語を使う。この節で述べる枠組みは田口(1984)によってとられたものである。比率曲線の名称も同書による。

2) 本稿の図はいずれも概略図にすぎないので、折れ線ではなく簡単な曲線で描いてある。

3) majorization の概念によれば、ローレンツ順序は, submajorize かつ supermajorize として捉えられるので、下からと上からの累積シェアを組み合わせた凸レンズをローレンツ曲線とするほうがよい。majorization については Marshall and Olkin(1979)を参照。

4) Suits(1977)は、この特徴に着目して比率曲線を租税累進度の計測に応用している。

5) このような考え方は田口(1984)でなされている。ローレンツ曲線や集中曲線 $L_{y,x}$ についても同様な扱いが可能である。

6) ジニ係数のその他の表現については、David(1981), Stuart and Ord(1987)を参照。

7) 本稿の範囲の操作変数法に関しては Jonston(1984)で十分である。以下で述べる Durbin, Wald, Bartlett の方法も同教科書に載っている。

8) これは 豊田(1986, 1987)で指摘した。Olkin and Yitzhaki(1992)でも指摘されている。

9) このことは 豊田(1988)で指摘した。

10) この事実は理論・計量経済学会 1988 年度大会で報告した。なお、比例税のときに 0 となるように基準化するとすれば、 $K^*=(K/G_x)$, $S^*=(S/G_x)$ である。

引用文献

David, H. A.(1981) *Order Statistics*, 2nd ed., Wiley, New York.

Jakobson, U.(1976) "On the Measurement of the Degree of Progression," *Journal of Public Economics*, Vol. 5, pp. 161-168.

Jonston, J.(1984) *Econometric Methods*, 3rd ed., McGraw-Hill, New York.

Kakwani, N. C.(1977a) "Measurement of Tax Progressivity: An International Comparison," *Economic Journal* Vol. 87, pp. 71-80.

Kakwani, N. C.(1977b) "Applications of Lorenz Curves in Economic Analysis," *Econometrica*, Vol. 45, pp. 719-727.

Lorenz, M. O.(1905) "Methods of Measuring Concentration of Wealth," *Journal of the American Statistical Association*, Vol. 9, pp. 209-219.

Marshall, A. W. and I. Olkin(1979) *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.

Olkin, I and S. Yitzhaki(1992), "Gini Regression Analysis," *International Statistical Review*, Vol. 60, pp. 185-192.

Stuart, A and J. K. Ord(1987) *Kendall's Advanced Theory of Statistics*, 5th ed., Charles Griffin, London.

Suits, D. B.(1977) "Measurement of Tax Progressivity," *American Economic Review*, Vol. 67, pp. 747-752.

田口時夫(1984)『経済分析と多次元解析』東洋経済新報社。

豊田 敬(1986)「ローレンツ曲線による弾力性の計測方法」林周二・中村隆英編『日本経済と経済統計』東京大学出版会, pp. 205-223.

豊田 敬(1987)「税の累進度と所得再分配係数」『経済研究』第38巻, pp. 166-170.

豊田 敬(1988)「累積とグラフによる直線のあてはめ」『第56回日本統計学会講演報告集』pp. 124-125.