中央研究院
三民主義研究所
專題選刊
(二十二)

廣義的吉尼係數

——所得有負時的吉尼係數及其修正——

曹添旺　賴東昇　陳昭南

中華民國
臺灣　臺北　南港
中華民國六十八年一月
(七十五年重新排印出版)

廣義的吉尼係數*
——所得有負時的吉尼係數及其修正——

曹 添 旺　　賴 東 昇　　陳 昭 南

當代民主國家的政府，每多運用所得重分配政策：如租稅征課、移轉性支付、公共財政支出等措施，藉以縮減貧富的差距，增進全社會的公平和福祉。我們若要評估所得重分配政策是否有效，就得測量並比較每期各類所得來源的分配狀況。我們知道，財政措施可使某些人增加所得，某些人減少所得（有的人繳稅，有的人接受補貼）。除了移轉性所得之外，像「農業淨收入」或「財產所得」之類的所得，也常常有負值（損失）出現。但衡量所得分配不平均度最常用的吉尼集中係數（Gini coefficient of concentration，簡稱吉尼係數），却不能適用於所得有正有負的情形，例如Pyatt-Chen-Fei (1978) 和曹添旺 (1978)，根據這類所得所計算的吉尼係數竟然大於一。大家都知道，吉尼係數的值必須介乎零與一之間才算是合理（其中零和一分別代表最平均和最不平均的極端情形）。本文的目的就在指出所得有負時傳統吉尼係數不合理的地方，並進而設計「廣義的」吉尼係數，使它既能適用於所得有負的情況，又能保留傳統吉尼係數的基本特性（介乎零與一之間）。

讓我們先檢討有負所得時的Lorenz曲線（Lorenz curve）跟平常的有什麼不同？為便於分析起見，假定n個家庭（或個人）中，雖有若干家的所得是負的，但

*本文在寫作過程中，曾與郭秋永、胡春田、張嘉璥、王春源、黃佳彬諸先生討論，獲益良多，特此致謝。

n家的總所得仍然爲正〔註一〕。把他們的所得從小到大排好，並以橫軸代表累積的戶（人）數百分比（其坐標從零到一），縱軸代表累積的所得百分比（其坐標從負數到一），則Lorenz曲線的形狀有如圖一所示。

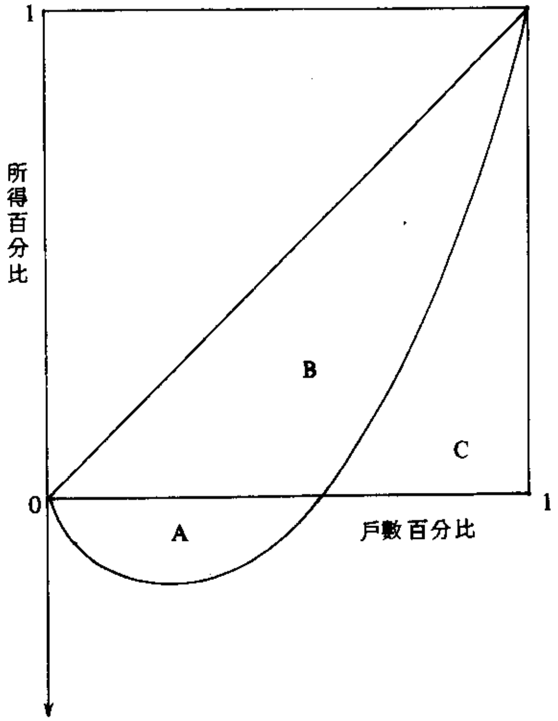


圖 一

因爲前幾家的所得是負的，所以Lorenz曲線有一部份落在橫軸底下。爲了以下討論的方便，我們令這段曲線與橫軸包圍的面積（絕對值）是A，在橫軸之上的Lorenz曲線與對角線（45° 線）包圍的面積爲B，而三角形扣去B後所剩的面積則稱爲C（即B＋C＝1／2）。

其實，上圖的Lorenz曲線如改用數式來說明，可能更清楚。由於n戶家庭的所得已按「單調遞增」（Monotonic nondecreasing）的次序排列：

$$Y_1 \leq Y_2 \leq \cdots\cdots\cdots\cdots \leq Y_n \tag{1}$$

若令S_i（$i=1, 2, \cdots\cdots, n$）代表前i戶家庭的累積所得，即

$$S_i = Y_1 + \cdots\cdots\cdots\cdots\cdots + Y_i \tag{2}$$

並設 $S_n > 0$ ，即可將圖一的Lorenz 曲線，在 $i/n(i=1,2,\cdots\cdots,n)$點上的定義寫成：

$$L(0)=0\ ,\quad L(i/n)=S_i/S_n=s_i\ ,\ L(1)=s_n=1 \tag{3}$$

現在我們要問：這樣的所得分配的吉尼係數到底是多少？爲回答此問題，且讓我們探討傳統吉尼係數的定義。誠如Theil(1967)，Kendall and Stuart (1963), Gastwirth (1972) 指出的，吉尼係數可以由統計學上「均互差」(mean difference)的概念引伸得來。就式(1)的n 個序數(ordered numbers)來說，其均互差(M) 可以寫做〔註二〕

$$M=(2/n^2)\sum_{i=1}^{n}\sum_{j<i}(Y_i-Y_j) \tag{4}$$

換句話說，所謂均互差就是指這些序數相互間絕對差量的平均，而這均互差與傳統的吉尼係數 (G) 具有下列的關係：

$$G=M/(2\mu) \tag{5}$$

式中 $\mu=S_n/n$，代表n 戶家庭的平均所得。

把式 (4) 代入式 (5) 中，即得

$$G=(1/n)\sum_{i=1}^{n}\sum_{j<i}(y_i-y_j) \tag{6}$$

式中 $y_i=Y_i/S_n\ (i=1,2,\cdots\cdots,n)$，稱爲第 i 戶家庭的所得配份(income share)，而 $y_1+\cdots\cdots+y_i=s_i$ ，$y_1+\cdots\cdots+y_n=s_n=1$ 。

展開式 (6) ，並稍加整理，我們得到〔註三〕

$$G=\frac{n-1}{n}-\frac{2}{n}[(n-1)y_1+(n-2)y_2+\cdots\cdots+y_{n-1}] \tag{7}$$

因此可得

$$1-G=\frac{1}{n}+\frac{2}{n}[(n-1)y_1+(n-2)y_2+\cdots\cdots+y_{n-1}]$$

$$=\frac{2}{n}\left[\frac{1}{2}+(n-1)y_1+(n-2)y_2+\cdots\cdots+y_{n-1}\right]$$

$$=\frac{2}{n}\left[\frac{1}{2}(y_1+y_2+\cdots+y_n)+(n-1)y_1+(n-2)y_2+\cdots+y_{n-1}\right]$$

$$=\frac{2}{n}\left[\frac{1}{2}y_1+\left(\frac{1}{2}y_2+y_1\right)+\left(\frac{1}{2}y_3+y_2+y_1\right)+\cdots\right.$$

$$\left.+\left(\frac{1}{2}y_n+y_{n-1}+\cdots\cdots+y_1\right)\right] \tag{8}$$

仔細推敲上式的右邊，不難發現它也可以用圖二所示的面積來表示，即

$$\frac{1}{n}\left[\frac{1}{2}y_1+\left(\frac{1}{2}y_2+y_1\right)+\left(\frac{1}{2}y_3+y_2+y_1\right)+\cdots+\left(\frac{1}{2}y_n+y_{n-1}+\cdots+y_1\right)\right]$$

$$=-A+C \tag{9}$$

以式(9)代入式(8)，即得

$$1-G=2(-A+C) \tag{8a}$$

所以，

$$\begin{aligned} G &= 1+2A-2C \\ &= 1+2A-2(1/2-B) \\ &= 2(A+B) \end{aligned} \tag{10}$$

這個式子明白地告訴我們，由統計均互差推導得來的吉尼係數，也可以Lorenz曲線的概念表示。套Theil的話來說，「吉尼係數正是對角線與Lorenz 曲線所包圍面積的兩倍」(Theil (1967), 頁 121)。[註四]

到這裏我們已經有了粗略的印象：當負所得家庭的戶數越多，其所得配份(絕對值)越高，則圖中的A也就越大，果爾，傳統的吉尼係數就越可能超過一。不過，爲要進一步確知傳統吉尼係數大於一的條件，我們將式(8a)改寫如下：

$$G=1+2(A-C) \tag{8b}$$

這就是說，

$$G \gtreqless 1 \quad 端視 \quad A \gtreqless C 而定 \tag{11}$$

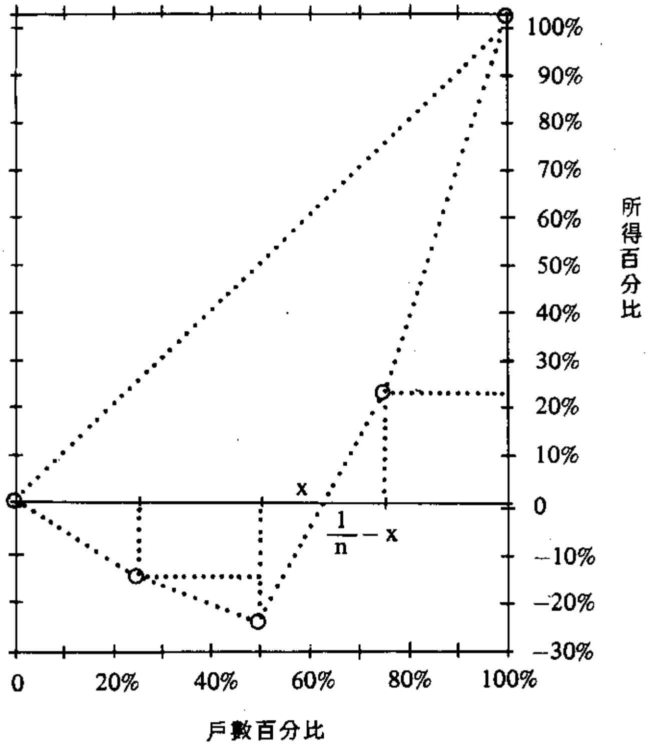


圖 二

換言之，一旦面積A比面積C大，則傳統的吉尼係數大於一；反之，則吉尼係數小於一。

我們在此擬借用 Schutz（1951）的兩個例子——負所得小（如表一）和負所得大（如表三）——分別討論傳統的吉尼係數大於一及小於一的情形。並假設一組吉尼係數恰恰爲一的家庭所得資料（如表二），再將各情況所對應的Lorenz 曲線分別繪於圖三至圖五，以便比較三者並指出傳統吉尼係數不合理的地方。

下列圖表已經顯示了傳統吉尼係數應用於所得有負時的缺點。第一、即使G＝1也不能代表通常所謂的絕對不平均的情況。前面說過，若A愈大，G就愈高。我們可以很容易地想像：可能有相當大的A使G趨於無窮大。換句話說，這時傳統吉

表一　G小於一的情況

家庭		所得				
戶數 1	累積戶數 2	每戶所得 (Y) 3	累積所得 (S) 4	每戶所得配份 (y) 5	累積所得配份 (s) 6	累積所得配份之累加(Σs_i) 7
1	1	-$ 50	− 50	-.05	− .05	− .05
1	2	20	− 30	.02	− .03	− .08
1	3	50	20	.05	.02	− .06
1	4	80	100	.08	.10	.04
1	5	100	200	.10	.20	.24
1	6	100	300	.10	.30	.54
1	7	120	420	.12	.42	.96
1	8	150	570	.15	.57	1.53
1	9	180	750	.18	.75	2.28
1	10	250	1,000	.25	1.00	3.28

吉尼係數　$G=\frac{1}{10}[(10+1)-2(3.28)]=\frac{1}{10}[4.44]=.444<1$

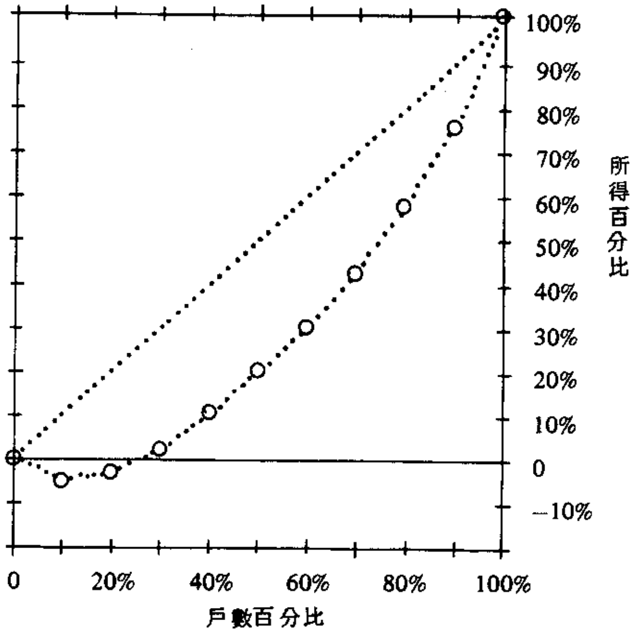


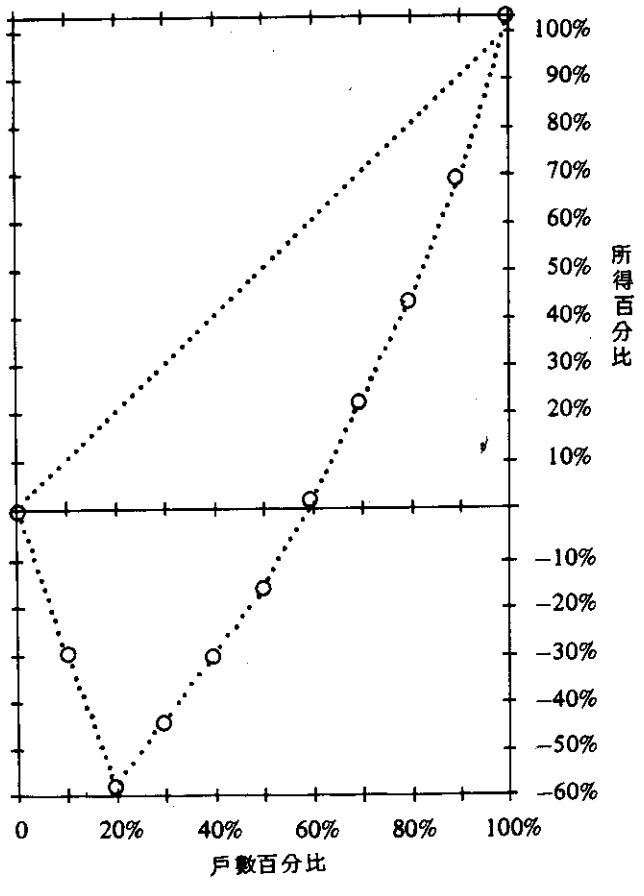
圖　三

表二　G等於一的情況

家庭		所得				
戶數	累積戶數	每戶所得 (Y)	累積所得 (S)	每戶所得配份 (y)	累積所得配份 (s)	累積所得配份之累加 (Σs_i)
1	2	3	4	5	6	7
1	1	−$300	−$ 300	−.30	−.30	−.30
1	2	− 280	− 580	−.28	−.58	−.88
1	3	130	− 450	.13	−.45	−1.33
1	4	140	− 310	.14	−.31	−1.64
1	5	140	− 170	.14	−.17	−1.81
1	6	180	10	.18	.01	−1.80
1	7	200	210	.20	.21	−1.59
1	8	210	420	.21	.42	−1.17
1	9	250	670	.25	.67	−.50
1	10	330	1,000	.33	1.00	.50

吉尼係數　$G=\frac{1}{10}[(10+1)-2(.5)]=\frac{1}{10}[10]=1$

圖四



尼係數的上限不再是一，而是無窮大。準此，我們很難直接從係數的值去感受所得分配不平均的程度。第二、儘管是相當大的吉尼係數，但其分配狀況未必比吉尼係數較小但所得全爲正的分配更不平均。擧例來說，如果社會上有一百個人，其中一人獨囊全社會的所得，其他人的收入則皆爲零。這樣的所得分配當然比表三所列的資料更不平均。但後者的吉尼係數高達1.94，而前者——絕對不平均的特例——的吉尼係數却只是0.99。這樣看來，傳統的吉尼係數的大小已不能反映「眞正的」不平均度。

有鑑於此，Schutz (1951)即摒棄了傳統吉尼係數「面積」的概念，而從Lorenz曲線的斜率着眼，建立另一種不平均度的指標——斜率指標。他認爲，對角線是代表絕對平均的所得分配，其斜率恰好爲一。因此，只要計算 Lorenz 曲線各點的斜率與一的差量（絕對值）之和，即可用以表示所得分配不均的程度〔註五〕，縱使是

表三 G大於一的情況

家庭		所得				
戶數 1	累積戶數 2	每戶所得 (Y) 3	累積所得 (S) 4	每戶所得配份 (y) 5	累積所得配份 (s) 6	累積所得配份之累加 (Σs_i) 7
1	1	-$500	-$ 500	-.50	- .50	- .50
1	2	- 300	- 800	-.30	- .80	-1.30
1	3	- 300	- 1,100	-.30	-1.10	-2.40
1	4	- 100	- 1,200	-.10	-1.20	-3.60
1	5	200	- 1,000	.20	-1.00	-4.60
1	6	300	- 700	.30	- .70	-5.30
1	7	300	- 400	.30	- .40	-5.70
1	8	400	0	.40	0	-5.70
1	9	500	500	.50	.50	-5.20
1	10	500	1,000	.50	1.00	-4.20
吉尼係數		$G=\frac{1}{10}[(10+1)-2(-4.2)]=\frac{1}{10}[19.4]=1.94>1$				

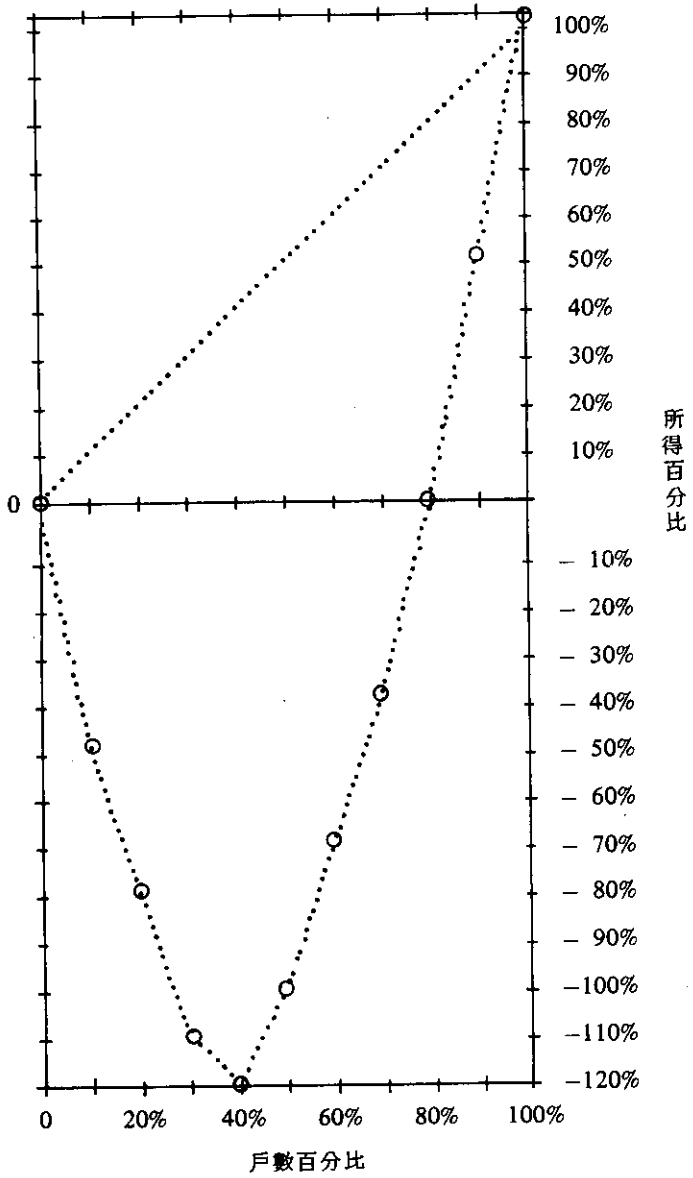

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0
– 10%
– 20%
– 30%
– 40%
– 50%
– 60%
– 70%
– 80%
– 90%
–100%
–110%
–120%
所得百分比
0
20%
40%
60%
80%
100%
戶數百分比

圖　五

所得有正有負的情況，也同樣可以比較。然而，Schutz 的指標並沒有克服所有的困難，因爲它的上限仍是無窮大。也就是說，它照樣犯有上述傳統吉尼係數的第一個缺點。

其實，我們如果把式(10)改變一個形式寫做

$$G=2(A+B)=(A+B)/(1/2)=(A+B)/(B+C) \qquad (10a)$$

就可以看出，傳統的吉尼係數應用在所得有負的情況會得到不合理的結果（高估不平均度），正是因爲它只在分子裏考慮A，而分母裡却忽略了。倘若我們也把A帶進分母裡來，那麼一切問題便可以迎双而解。因此，我們把吉尼係數設定爲下列的形式：

$$G^*=(A+B)/(A+B+C) \qquad (12)$$

使它既能保留吉尼係數原來的特性，又能應用在所得有負的情形：

㈠假使沒有負所得（即$A=0$），則$G^*=B/(B+C)=2B$，這正是所得全爲正的傳統的吉尼係數（即$G^*=G$），它的值當然介於零與一之間。

㈡當所得有負時（即$A>0$），即$G^*=(A+B)/(A+B+C)<1$，換句話說，G^* 的上限仍然是一，即使有負所得時也是一樣。

㈢在面積A既定的情況下，若C越小，則G^*越大。例如當C等於零時，G^*的值就等於一（達到上限）。所以，廣義的吉尼係數也涵蓋了傳統的絕對不平均的情況〔註六〕。

由此可見，不論負所得是否存在，G^*都沒有違反傳統吉尼係數的基本精神。因之，我們不妨稱G^*爲「廣義的」吉尼係數。

我們現在來探討「廣義的」吉尼係數與「傳統的」吉尼係數之間的關係。從式(12) 可得

$$G^*=(A+B)/(\frac{1}{2}+A)=2(A+B)/(1+2A)=aG \qquad (13)$$

式中$a=1/(1+2A)$，稱做調整係數。換言之，我們的廣義吉尼係數不外是把傳

統吉尼係數乘上一個調整係數而已，這個調整係數就是A——橫軸底下的面積——的函數。我們知道，A的出現正是傳統吉尼係數發生問題的癥結所在。因此，必須把A納入調整係數，才能解決問題。倘若沒有負所得（即A＝0），則調整係數爲一，廣義的吉尼係數跟傳統的吉尼係數相同（即$G^*=G$）；所得如果有負的時候（即A＞0），調整係數就小於一，因而廣義的吉尼係數必然小於傳統的吉尼係數（即$G^*<G$）。由此可知，G只是G^*的一個特殊情形而已。這就是爲什麼我們可以把G^*稱做「廣義的」吉尼係數的理由。

現在，我們回過頭來就前面的各個實例分別計算廣義的吉尼係數。在計算的過程中，我們先利用下式求得面積A的值〔註七〕：

$$A=\frac{-1}{n}\sum_{i=1}^{k-1}s_i+\frac{s_{k-1}\cdot s_k}{2n(s_k-s_{k-1})} \tag{14}$$

式中k爲累積所得恰好由負轉正的家庭戶別（$0\leq k\leq n$）。換句話說，s_k是「非負的」累積所得百分比中的極小值，亦即k滿足：$s_k\geq 0$，但$s_{k-1}<0$的條件。

再把式(14)代入式(13)，即可求計各個情形的廣義的吉尼係數。它們與傳統的吉尼係數都列在表四：

表四　「廣義的」吉尼係數與「傳統的」吉尼係數的比較

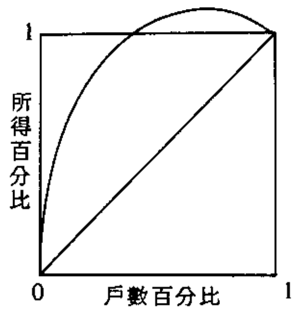
資料		傳統的吉尼係數 G	橫軸底下的面積 A	調整係數 a	廣義的吉尼係數 $G^*=aG$
有負所得時	如表一	.44	.0074	.9854	.4375
	如表二	1.00	.1805	.7347	.7347
	如表三	1.94	.5700	.4672	.9065
無負所得時	百分之九十九的人所得為零，一人獨得全部所得	.99	0	1.0000	.9900

該表顯示，在所得有正有負的情況下，廣義吉尼係數的值仍然落在零與一之間，所以它可以直接和所得都是正的所得分配比較不平均的程度。換句話說，本文提出的廣義吉尼係數，已經解決了傳統吉尼係數應用在所得有負時的不合理現象。

儘管如此，我們還是沒有完全克服傳統吉尼係數的困難。大家都知道，從分配狀況不同而其 Lorenz 曲線却相交的所得資料，可能求得等值的吉尼係數，我們的廣義吉尼係數也犯了這個「傳統的」毛病。此外，如果正的所得與負的所得恰恰相抵而使總所得等於零時，我們就不能照前述的方法定義吉尼係數。這時若要表示所得分配的狀況，就必須建立另一種不平均度的指標。

註　釋

〔註　一〕如果總所得是負的，則 Lorenz 曲線可能形如下圖：



不過，基於吉尼係數「尺度無關」(Scale irrelevant) 的特性，我們將各戶所得統乘負一，就可以使 Lorenz 曲線回復到圖一的形狀。因此，我們不必討論總所得爲負的情形。

〔註　二〕統計學 (如 Kendall and Stuart (1963)) 上均互差的定義有兩種：

$$M = \frac{1}{n(n-1)} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |x_j - x_k| f(x_j) f(x_k), \quad j \neq k \qquad (\mathrm{I})$$

$$M = \frac{1}{n^2} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |x_j - x_k| f(x_j) f(x_k) \qquad (\mathrm{II})$$

我們採用式 (II) 的定義。又，由於我們的所得資料已按大小次序排列，故其均互差可以直接寫如正文式 (4) 的形式。

〔註 三〕式(6)展開後，得

$$G = \frac{1}{n}\,[\,(y_2 - y_1) + (y_3 - y_1) + (y_4 - y_1) + \ldots\ldots + (y_n - y_1)$$

$$+ (y_3 - y_2) + (y_4 - y_2) + \ldots\ldots + (y_n - y_2)$$

$$+ (y_4 - y_3) + \ldots\ldots\ldots\ldots\ldots$$

$$\vdots$$

$$+ (y_n - y_{n-1})\,]$$

$$= \frac{1}{n}\,[\,y_2 + 2y_3 + \ldots\ldots + (n-1)\,y_n - (n-1)\,y_1 - (n-2)\,y_2 \ldots\ldots - y_{n-1}\,]$$

$$= \frac{1}{n}\,[\,y_2 + 2y_3 + \ldots\ldots + (n-2)y_{n-1} + (n-1)\,(1 - y_1 - y_2 \ldots\ldots - y_{n-1})$$

$$- (n-1)\,y_1 - (n-2)\,y_2 \ldots\ldots - y_{n-1}\,]$$

$$= \frac{1}{n}\,[\,(n-1) - 2\,(n-1)\,y_1 - 2\,(n-2)\,y_2 - 2\,(n-3)\,y_3 \ldots\ldots - 2y_{n-1}\,]$$

$$= \frac{2}{n}\,[\,\frac{n-1}{2} - (n-1)y_1 - (n-2)y_2 - (n-3)y_3 \ldots\ldots - y_{n-1}\,]$$

$$= \frac{n-1}{n} - \frac{2}{n}\,[\,(n-1)y_1 + (n-2)y_2 + \ldots\ldots + y_{n-1}\,]$$

不過，爲了實際計算的方便，我們在表一至表三中利用下列的公式：

$$G = \frac{1}{n}\,[\,(1 - s_1) + (1 - s_2) + \ldots\ldots + (1 - s_{n-1}) - (s_1 + s_2 + \ldots\ldots + s_{n-1})\,]$$

$$= \frac{1}{n}\,[\,(n-1) - 2\,(s_1 + s_2 + \ldots\ldots + s_{n-1})\,]$$

$$= \frac{1}{n}\,[\,(n+1) - 2\,(s_1 + s_2 + \ldots\ldots + s_n)\,]$$

〔註 四〕Fei-Ranis-Kuo（1978，頁 45）也有類似的證明。

〔註 五〕Schutz（1951）是以 Lorenz 曲線各點斜率與一的差量（絕對值）之和的一半代表

不均係數（ Coefficient of inequality ）。不過，誠如 Gastwirth（1972）指出，Schutz 的不均係數就是 Lorenz 曲線與對角線（即45°線）之間的最大距離（頁307）。事實上，這正相當於把所有家庭區分成「低於『平均』所得」與「不低於『平均』所得」兩組後，再根據分組資料（各組的平均所得）所計算的吉尼係數。

〔註　六〕在沒有負所得的通常情況下（即A＝0），若C＝0，則G＝G*＝1。

〔註　七〕從文中的圖二可以看出，面積A的值是：

$$A = \left[\frac{-1}{n}s_1 + \frac{1}{2n}y_1\right] + \left[\frac{-1}{n}s_2 + \frac{1}{2n}y_2\right] + \ldots\ldots + \left[\frac{-1}{n}s_{k-1} + \frac{1}{2n}y_{k-1}\right]$$

$$-\frac{1}{2}s_{k-1}x \qquad \text{(III)}$$

式中x代表 Lorenz 曲線與橫軸交點的橫坐標減去（k－1）／n 後的值。從相似三角形的性質知：

$$\frac{s_k}{\frac{1}{n} - x} = \frac{-s_{k-1}}{x}$$

故得　$$x = \frac{-1}{n}\,\frac{s_{k-1}}{s_k - s_{k-1}} \qquad \text{(IV)}$$

以式(IV)代入(III)，可得

$$A = \frac{-1}{n}(s_1 + s_2 + \ldots\ldots + s_{k-1}) + \frac{1}{2n}(y_1 + y_2 + \ldots\ldots + y_{k-1})$$

$$+\frac{1}{2n}s_{k-1}\frac{s_{k-1}}{s_k - s_{k-1}} \qquad \text{(V)}$$

利用$y_1 + y_2 + \cdots\cdots + y_{k-1} = s_{k-1}$的關係式，代入式(V)並加整理，我們求得文中的式(14)。

參考文獻

1. Fei J. C. H., Ranis G., and Kuo S. W. Y., "Growth and the Family Distribution of Income," *Quarterly Journal of Economics*, 92 (February 1978), 17-53.

2. Gastwirth J. L., "The Estimation of the Lorenz Curve and Gini Index," *Review of Economics and Statistics*, 54 (August 1972), 306-316.

3. Kendall M. G., and A. Stuart, *The Advanced Theory of Statistics* 1. 2nd ed. London: Charles Griffen and Company, (1963).

4. Pyatt G., Chen C. N., and Fei J. C. H., "The Distribution of Income by Factor Components," unpublished, (1978).

5. Schutz R. R., "On the Measurement of Income Inequality," *American Economic Review*, 41 (March 1951), 107-122.

6. Theil H., *Economics and Information Theory*, Chicago, Rand McNally and Company, (1967).

7. 曹添旺：「分組資料與家庭所得不平均度的關係」，中央研究院三民主義研究所專題選刊（十八），民國六十八年一月。