Estimating and Combining National Income Distributions using Limited Data

Duangkamon Chotikapanich
Monash University

William E. Griffiths
University of Melbourne

D.S. Prasada Rao
University of Queensland

June 15, 2004, 4.00pm

Abstract

Recently, there has been a resurgence of studies on the distribution of income and inequality at regional and global levels, largely driven by the concerns of economists, international development organizations and the general public about the overall effects of globalization on growth and inequality. A major data problem encountered in these studies is the nature of income distribution data that are available mainly in a summary form that includes mean (average) income and income shares of quintile or decile groups of the population. Past studies have either ignored distributional characteristics within each population sub-group, implying that all individuals in a quintile or decile group have the same income, or used simple distributions like the lognormal or Pareto to model income distribution within each country.

The aim of the paper is to estimate national and regional income distributions within a more general framework that relaxes the assumption of constant-income-within-groups and is based on a general and versatile class of income distributions. A technique to estimate parameters of a class of generalized beta distributions using grouped data is proposed. Regional income distribution is modelled using a mixture of country-specific distributions and its properties are examined. The techniques are used to analyse national and regional inequality trends for eight East Asian countries and two benchmark years 1988 and 1993.

Keywords**:** Gini coefficient; Generalized beta distribution

1. Introduction

In the current climate of increasing globalization and a push for free trade among nations, there is concern that increasing globalization may lead to increasing inequality, and that increasing global inequality may mean the unsustainability of the current international order. One major difficulty with the ongoing debate about globalization is the problem of measuring the extent of economic inequality, and being able to meaningfully compare inequality across countries, regions or time periods. Unless we can do this well, we cannot easily evaluate whether various policy initiatives such as moves towards greater globalization are increasing or reducing inequality.

The state of the art approach to measuring inequality on a global scale is found in a recent paper by Milanovic (2002). His work recognizes that inequality measures need to take account of the differences in income levels within countries, as well as differences in the averages across countries. However, Milanovic's approach still understates the magnitude of inequality, because it assumes that all people in a given group (for example the poorest 10% of people in a given country) receive that same income (the average income for that group). Thus, it ignores the distributional characteristics within each population sub-group in each of the countries. This is a potentially serious limitation, and could be critical to inequality measures and development policy evaluation.

In this paper we suggest and apply an alternative methodology to redress this deficiency. Using data on mean income for each of a number of population groups (usually decile groups from poorest to richest) for eight East Asian countries, we fit a generalized beta income distribution for each country and compute a regional income distribution as a weighted average of the country-specific income distributions. This procedure is applied to data for two years, 1988 and 1993.

The outline of this paper is as follows. Section 2 describes the data and sources. Section 3 discusses the methodology used to estimate income distribution for each country assuming that country-specific income distributions follow generalized beta distributions. It was found that inequality within countries increased over the period 1988 and 1993. Section 4 describes the methodology used to combine country-specific income distributions into regional income distribution. A few concluding remarks and possible areas for further research are provided in Section 5.

2. Description of data and sources

The empirical implementation of the estimation of the statistical distributions used in measuring inequality within each country and in the region as a whole is based on income distribution data from the countries covered. The main purpose of the empirical work reported here is to illustrate the econometric methodology developed in the paper, and it is restricted to a small selection of Asian countries. The authors intend to apply this methodology in the near future to a larger data set which could be considered truly global in its coverage.

The countries covered in the empirical investigation consist of a few countries in the East Asian region. These are: Hong Kong, Japan, Malaysia, the Philippines, Singapore, South Korea, Taiwan and Thailand.[1] The selection of these countries is also based on the fact that data for these countries would be more reliable.

The main objective of the current research is to provide an accurate assessment of the levels and trends in regional and global inequality. Such assessments could form the basis for informed debate on the effect of globalisation on inequality. Ideally, the study should cover a period of at least the last two decades. The current empirical application of the methodology has been restricted to two benchmark years 1988 and 1993, these are the two years of focus in the study of Milanovic (2002). At the time of preparation of this paper, it was not possible to obtain income distribution data for the most recent years.[2]

The methodology developed in the paper focuses specifically on the limited nature of the data available. The basic data used in this paper has three components:

- mean income/real per capita income of the population
- income shares of population in decile groups

In addition, data on population size is also utilised.

Real per capita income: In order to be able to derive income distribution for the region as a whole, nominal per capita income of each country needs to be adjusted for differences in prices across countries, and for purposes of temporal welfare comparisons these need to be adjusted for movements in prices over time. Such adjustments are made through the use of purchasing power parities. Data on real per capita income are drawn from the latest version of the Penn World Tables, PWT 6.1[3] which has data on real per capita incomes for over 150 countries spanning a 50-year period. PWT 6.1 also provides data on population size of each of the countries.

Income shares of population in decile groups: The empirical work of the paper makes use of income distribution data available in an aggregated form. The data used here are exactly the same as those used by Milanovic (2002)[4]. The data consists of the shares of total gross domestic product (income) received by people in groups of 10 per cent after the population is ranked from the poorest to the richest. These data are the same as those available from the World Bank inequality data set compiled by Deininger and Squire (1996). The Milanovic data set did not include data for Singapore for the year 1988, ILO (1995) is used as an alternative source of data for Singapore.

Income versus expenditure share distributions: Ideally distribution data should refer either to income or expenditure of persons or households. In the current data set, all the data used refer to the distribution of incomes with the exception of Singapore in 1988, Philippines in both 1998 and 1993 and Thailand in 1993 where data refer to expenditure distributions. These differences could influence the estimates of the parameters of respective income distributions.

The following are the main variables for which data are drawn from various sources described or derived from the data given. For each country in the study, let

---

[1] These countries may also be considered as the newly industrialized countries (NICs) of the region. See Chotikapanich and Rao (1998) for details on inequality in the NICs and non-Nics.

[2] It is anticipated that the authors would gain access to data on Asian countries for the year 2000 from the Asian Development Bank.

[3] The URL for PWT 6.1 is http://pwt.econ.upenn.edu/php_site/pwt_index.php.

[4] The authors are grateful to Mialnovic for making these data available for analysis.

S = Size of population

$\bar{y}$ = Average or real per capita income (or expenditure)

c_i = Share of population in i-th income group[5]

g_i = Income share of population in i-th income group

$\bar{y}_i$ = Mean income of population in i-th income group

(a_{i-1} , a_i) represents the income class for i-th income group

Of all these variables, only S, $\bar{y}$, c_i and g_i are available from various data sources. In this study, for each income group $\bar{y}_i$ is derived as:

$$\bar{y}_i \quad = \quad \frac{g_i \, \bar{y} \, S}{c_i \, S}$$

and the limits of the income class intervals, a_1, a_2, ...a_{i-1}, a_i, a_{i+1}, ... are treated as parameters to be estimated (see Section 3 for further details).

The basic data used in the paper is presented in Appendix Table.

3. Modelling Country-specific Income Distributions

This paper assumes that the income in different countries follow a generalized beta distribution with different parameters. A similar paper by Chotikapanich *et. al.* (1997) used a lognormal distribution to model income distribution for each country. Generalized beta distribution is a more flexible distribution and it has been shown to provide a good fit to variety of empirical income distributions (see McDonald (1984) and McDonald and Ransom (1979)). The main problem in terms of estimating parameters of the distribution is the lack of detailed income data. When only a limited number of observations are available (10 in the case of decile shares) the standard approach of estimation cannot be used. The proposed method of estimation used in this paper is described below.

3.1 The Beta Income Distribution

A number of generalizations of beta distribution have been used to fit income distributions. See, for example, McDonald (1984) and Kleiber and Kotz (2003, Ch6.). A flexible one that we have chosen has probability density function (pdf)

$$f(y) = \frac{y^{p-1}}{b^p B(p,q)\left(1+\dfrac{y}{b}\right)^{p+q}} \qquad\qquad y > 0$$

where $b > 0, p > 0$ and $q > 0$ are parameters and

$$B(p,q) \quad = \quad \frac{\Gamma(p)\ \Gamma(q)}{\Gamma(p+q)} \quad = \quad \int_0^1 t^{p-1}(1-t)^{q-1}\,dt$$

---

[5] If the income group refers to decile groups then each $c_i = 0.1$.

For the mode of $f(y)$ to be nonzero we require $p > 1$; for the mean to exist $q > 1$ is required.

Its corresponding cumulative distribution function (cdf) is given by

$$F(y) = \frac{1}{B(p,q)} \int_0^{[y/(b+y)]} t^{p-1}(1-t)^{q-1}\,dt$$
$$= B_{y/(b+y)}(p,\ q)$$

The function $B_t(p,q)$ is the cdf for the normalized beta distribution defined on the (0,1) interval. It is a convenient representation because it is commonly included as a readily-computed function in statistical software.

If T is a standard beta random variable defined on the interval (0, 1), then the relationship between T and Y is

$$T = \frac{Y}{b+Y} \qquad\qquad Y = \frac{bT}{1-T}$$

The mean, mode and variance of Y are given by

$$\mu = \frac{bp}{q-1} \qquad\qquad m = \frac{(p-1)b}{q+1}$$

$$\sigma^2 = \mu\left[\frac{b(p+1)}{q-2} - \mu\right] = \frac{b^2 p(p+q-1)}{(q-1)^2(q-2)}$$

Solving these expressions for b, p and q yields

$$b = \frac{\mu^2(\mu - m) - (3m - \mu)\sigma^2}{\sigma^2 - \mu^2 + \mu m}$$

$$p = \frac{\mu}{b}\left(\frac{2m+b}{\mu - m}\right) \qquad\qquad q = \frac{\mu + m + b}{\mu - m}$$

The Gini coefficient is given by

$$G = \frac{2B(2p, 2q-1)}{pB^2(p,q)}$$

3.2 Estimation method

The data available for each country are in the form of mean incomes for each of N classes $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_N$ each of which related to population proportions $c_1, c_2, \ldots, c_N$. Let the unknown class limits for these classes be given by $a_0, a_1, \ldots, a_N$ with $a_0 = 0$ and $a_N = \infty$. Then, we would like to fit a beta distribution to these data such that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{2N}$ are "close to zero" where

$$\int_{a_{i-1}}^{a_i} f(y)\,dy - c_i = \varepsilon_i \qquad\qquad i = 1, 2, \ldots, N$$

and

$$\frac{\int_{a_{i-1}}^{a_i} y f(y) dy}{\int_{a_{i-1}}^{a_i} f(y) dy} - \bar{y}_i = \varepsilon_{N+i} \qquad i = 1, 2, \ldots, N$$

In terms of the beta distribution function, these equations can be written as

$$B_{a_i/(b+a_i)}(p,q) - B_{a_{i-1}/(b+a_{i-1})}(p,q) - c_i = \varepsilon_i$$

and

$$\frac{bp}{q-1}\left(\frac{B_{a_i/(b+a_i)}(p+1,q-1) - B_{a_{i-1}/(b+a_{i-1})}(p+1,q-1)}{B_{a_i/(b+a_i)}(p,q) - B_{a_{i-1}/(b+a_{i-1})}(p,q)}\right) - \bar{y}_i = \varepsilon_{N+i}$$

where $B_{a_0/(b+a_0)}(p,q) = 0$ and $B_{a_N/(b+a_N)}(p,q) = 1$.

The second result is obtained by recognizing that

$$y\, f(y) = \frac{bp}{q-1} f^*(y)$$

where $f(y)$ is a beta pdf with parameters (b, p, q) and $f^*(y)$ is a beta pdf with parameters $(b, p+1, q-1)$.

To set up a framework for estimation we write

$$w_i = B_{a_i/(b+a_i)}(p,q) - B_{a_{i-1}/(b+a_{i-1})}(p,q)$$

and
$$z_i = \frac{B_{a_i/(b+a_i)}(p+1,q-1) - B_{a_{i-1}/(b+a_{i-1})}(p+1,q-1)}{w_i}$$

Also, define a $(2N \times 1)$ vector *x* as the dependent variable

$$x = \left(c_1, c_2, \ldots, c_N, \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_N\right)'$$

and let $d_1, d_2, \ldots, d_{2N}$ be dummy variables with d_i having unit in the *i*-th position and zeros elsewhere. Note that w_i and z_i are scalars and $x, d_1, d_2, \ldots, d_{2N}$ are $(2N \times 1)$ vectors.

The equations $w_i - c_i = \varepsilon_i$ and $z_i - \bar{y}_i = \varepsilon_{N+i}$ can now be written as

$$x = \sum_{i=1}^{N} d_i w_i + \frac{bp}{q-1} \sum_{i=1}^{N} d_{N+i} z_i + \varepsilon$$

Initially, we estimated the class limits and beta distribution parameters by finding those values of $(b, p, q, a_1, a_2, \ldots, a_{N-1})$ such that $\varepsilon'\varepsilon$ was minimized. However, because the first *N* elements of *x* are relatively small (proportions), and the last *N* elements are relatively large (income class means), these estimates were largely determined by the last *N* equations, $z_i - \bar{y}_i = \varepsilon_{N+i}$. It was possible to get estimates such that $\hat{a}_{i-1} > \hat{a}_i$. We overcame this problem, and ensured all 2*N* equations played their part in estimation, by minimizing the sum of squares of percentage errors $\varepsilon' V^{-1}\varepsilon$ where $V = \text{diagonal}(x_1^2, x_2^2, \ldots, x_{2N}^2)$.

It is important to get reasonable starting values. Those for b, p and q were obtained by finding estimates of the mean, mode and variance and then substituting into the equations for b, p and q given in Section 3.1. Note that for a sensible income distribution, we require $b > 0$, $p > 1$ and $q > 1$. It was often necessary to change the estimate of the mode to satisfy these inequalities. Starting values for $(a_1, a_2, \cdots, a_{N-1})$ were obtained as

$$a_i = \frac{\bar{y}_i - \bar{y}_{i+1}}{2}$$

3.3 Empirical Analysis

This section provides a brief analysis of the country-specific results obtained for the two years 1988 and 1993.

3.3.1 Country-specific income distributions

Table 1 shows the estimated parameters of the beta functions derived using the procedure described in Section 3.2. The estimated parameters provide meaningful income distributions, all of which show skewed uni-modal distributions. The very large values of p for Japan (and to a lesser extent Thailand in 1988) appear out of place. For the case of Japan the EViews program took a large number of iterations to converge. Estimation of q was stable, but for p and b it was not. This instability did not appear to be a problem, however. The parameters b and p were highly correlated and alternative pairs of (b, p) close to the convergence point led to virtually identical income distributions. For Singapore, the data for 1993 obtained from Milanovic are income data. This source did not have the data for 1988. Instead, we used 1988 expenditure data from ILO; this difference may explain the quite different parameter estimates for the two years. For Thailand, the data for both years are from Milanovic but the 1988 data are for income while the 1993 data are for expenditure.

Figure 1 shows the plots of the density functions and these are consistent with the general expectations. The locations of these distributions in terms of the mode and the mean appear to be ordered according to the real per capita incomes of these countries.

In fact it is more informative to examine the distribution functions and Lorenz curves for each country in each of the two years. To do so we can select a grid of income points $(y_1, y_2, \ldots, y_L)$ and compute

$$\pi_i = B_{y_i/(b+y_i)}\left(p, q\right)$$

and

$$\eta_i = B_{y_i/(b+y_i)}\left(p+1, q-1\right)$$

Figures 2a and 2b show the distribution functions for all the countries in the study. The Philippines, Thailand, Malaysia, Korea and Taiwan appear to be consistently ranked from the poorest to the richest. For any given income level, the Philippines has the highest proportion of people whose incomes are below the level, followed by Thailand and other countries. The ranking of these countries remained unaltered over the two periods. However, such clear dominance pattern is not evident in the case of

Japan, Singapore and Hong Kong – for these three countries the distribution functions cross-over at some income levels.

Figures 3a and 3b depict the Lorenz curves for Japan, Singapore and Hong Kong. These figures show clear Lorenz ordering of these three countries with Japan with least inequality followed by Singapore and Hong Kong.

3.3.2 Goodness-of-fit of Beta distributions

It is useful to assess the goodness of fit of the Beta distribution by comparing the observed and expected income shares derived using estimates of parameters of the distributions involved. The empirical income shares are given by

$$g_i = \frac{c_i\,\bar{y}_i}{\sum_{j=1}^{N} c_j\,\bar{y}_j}$$

To find those implied by a beta distribution we began with the population shares c_i, computed cumulative proportions

$$\pi_i = \sum_{j=1}^{i} c_i$$

and then found class limits a_i (not necessarily the same as the previously-estimated class limits) such that

$$B_{a_i/(b+a_i)}\left(p,q\right) = \pi_i$$

Corresponding cumulative income shares were found from the first moment distribution function

$$\begin{aligned}\eta_i &= \frac{1}{\mu}\int_0^{a_i} y\,f(y)\,dy \\ &= \frac{1}{\mu}\frac{pb}{q-1}\,B_{a_i/(b+a_i)}\left(p+1,q-1\right) \\ &= B_{a_i/(b+a_i)}\left(p+1,q-1\right)\end{aligned}$$

The estimated income shares are given by

$$g_i = \eta_i - \eta_{i-1}$$

A comparison of the estimated and observed income shares appears in Tables 2 and 3. The actual (observed) and estimated (expected) income shares of people in different decile groups are remarkably similar for all the countries in both years. In most cases the differences are in third decimal places. This is quite encouraging given that parameters of the distribution have been estimated using very limited data.

3.3.3 Temporal analysis of shifts in income distribution and levels and trends in inequality

Figure 4 shows the density functions for the years 1988 and 1993 for each of the countries included in the current study. These distributions appear to be plausible. Income distribution in the Philippines remained virtually unchanged over the period where as major structural shifts are evident in the case of Korea and Taiwan which are labelled as the Asian tigers for their performance during the study period.

The levels and trends in inequality can be studied using the Gini coefficients and the Lorenz curves. The observed and estimated Gini coefficients are computed and presented here.

The observed values of the Gini coefficient were obtained by applying the formula

$$G = \frac{2}{\mu} \operatorname{cov}\left[y, F(y)\right]$$

to the grouped data. The estimated values were obtained by substituting estimates *b*, *p* and *q* into the formula

$$G = \frac{2B\left(2p, 2q-1\right)}{pB^2\left(p,q\right)}$$

In addition to a comparison of the Gini coefficients, Lorenz dominance properties of the estimated income distributions for the years 1988 and 1993 are examined using sufficient conditions described in Wilfling (1996). The following sufficient condition is used here.

A distribution function $F(y)$ is said to exhibit less inequality in the Lorenz sense than a distribution $H(y)$, $F \leq_L H$, if the Lorenz curve of *F* is greater (lies above) than or equal to the Lorenz curve of *H*.

Given that income distributions of country *i* and *j* follow a Beta distribution, then a sufficient condition for the income distribution of country *i* to Lorenz dominate (have less inequality) than that for country *j* is (Wilfling, 1996, p.383) if

$$p_j \leq p_i \quad \text{and} \quad q_j \leq q_i$$

Table 4 presents the observed and estimated Gini coefficients for all the countries. Overall, the estimated Gini's are higher than the observed ones. This is expected because the estimated Gini is estimated from the beta distribution and therefore it takes into account the distribution within the classes. The exception is the result for Philippines for 1988 where estimated Gini is less than actual one. For this case the data for population shares are not even (not 10%) for different classes.

Trends in inequality shown in Table 4 are also quite interesting. With the exception of Korea, inequality within each country has increased over the period 1988 to 1993. This is consistent with the general notion that inequality may increase in countries experiencing rapid growth. The only surprising result is for Singapore where the Gini coefficient increased significantly. However, this may largely be due to the fact that the data for the year 1993 was drawn from Milanovic (referring to income data) and the 1988 data was drawn from the ILO and it referred to the expenditure distribution.

It is possible to drawn conclusions on Lorenz dominance using the sufficient conditions above. For each of the countries, comparing estimated values of *p* and *q* for the years 1988 and 1993 show that distribution in 1988 Lorenz dominates 1993 for Hong Kong, Japan, Malaysia, the Philippines, Singapore and Thailand. The sufficient condition is not satisfied for Korea and Taiwan. It is also possible to use this condition to assess Lorenz dominance across countries. For example, Taiwan Lorenz dominates

Malaysia in both 1988 and 1993. Japan, Singapore and Korea provide a Lorenz ordering as demonstrated in Figure 3.

3.4 Concluding Remarks

The results reported in this Section clearly demonstrate the feasibility of using generalised Beta distribution to model the distribution of income in Asian countries. The estimation procedure discussed in Section 3.2 provides a method of estimating parameters of the distribution using grouped data in the form of income shares of decile groups of population. Results on the levels and trends of inequality are meaningful and support the general notion that inequality within countries increased over the period 1988 and 1993. The next Section focuses on inequality in the region as a whole.

4. Modelling regional income distributions

The results of the country-specific income distributions estimated in Section 3 can now be combined together to obtain the regional income distribution. The procedure for combining the country-specific income distributions is given in following Section.

4.1 Regional distribution as a mixture of distributions

Given *M* countries each with a beta income pdf $f_j(y), j = 1,2,\ldots,M$, and population proportions $\lambda_1, \lambda_2, \ldots \lambda_M$, the pdf for the regional income distribution is given by the mixture

$$f(y) = \sum_{j=1}^{M} \lambda_j f_j(y)$$

The regional cumulative distribution function is given by the same weighted average of the country cdf's

$$F(y) = \sum_{j=1}^{M} \lambda_j F_j(y) = \sum_{j=1}^{M} \lambda_j B_{y/(y+b_j)}\left(p_j, q_j\right)$$

Regional mean income is given by

$$\mu = \sum_{j=1}^{M} \lambda_j \mu_j = \sum_{j=1}^{M} \frac{\lambda_j b_j p_j}{q_j - 1}$$

The regional cumulative income shares are given by

$$\begin{aligned}\eta(y) &= \frac{1}{\mu} \int_0^y z\, f(z)\, dz \\ &= \frac{1}{\mu} \sum_{j=1}^{M} \lambda_j \int_0^y z\, f_j(z)\, dz \\ &= \frac{1}{\mu} \sum_{j=1}^{M} \lambda_j\, \mu_j\, B_{y/(y+b_j)}\left(p_j + 1, q_j - 1\right)\end{aligned}$$

where $\mu_j = b_j p_j \big/ (q_j - 1)$.

A regional cumulative distribution function can be graphed by computing $F(y)$ for a grid of values of y. A regional Lorenz curve can be graphed by computing $F(y)$ and $\eta(y)$ for a grid of values of y.

The regional Gini coefficient is given by

$$\begin{aligned} G &= -1 + \frac{2}{\mu}\int_0^\infty y\,F(y)\,f(y)\,dy \\ &= -1 + \frac{2}{\mu}\int_0^\infty y\left(\sum_{j=1}^{M}\lambda_j\,F_j(y)\right)\left(\sum_{j=1}^{M}\lambda_j\,f_j(y)\right)dy \\ &= -1 + \frac{2}{\mu}\sum_{j=1}^{M}\sum_{i=1}^{M}\lambda_j\,\lambda_i\int_0^\infty y\,F_j(y)\,f_i(y)\,dy \end{aligned}$$

For the case where $i = j$ the integral in the above equation can be written as (after some extensive algebra)

$$\begin{aligned} \int_0^\infty y\,F_i(y)\,f_i(y)\,dy &= \mu_i\left[\frac{B(2p_i, 2q_i - 1)}{p_i\,B^2(p_i, q_i)} + \frac{1}{2}\right] \\ &= \frac{\mu_i}{2}(G_i + 1) \end{aligned}$$

where G_i is the Gini coefficient for the *i*-th country. However, a convenient analytical expression for the corresponding integral where $i \neq j$ does not appear to be available. As an alternative, we suggest estimating the relevant integrals using a large number of draws from the various beta distributions. To describe this process first note that

$$m_{ij} = \int_0^\infty y\,F_j(y)\,f_i(y)\,dy = E_{f_i}\left[\,y\,F_j(y)\right]$$

is the expectation of $y\,F_j(y)$ with respect to the pdf $f_i(y)$. We can then write

$$G = -1 + \frac{2}{\mu}\sum_{j=1}^{M}\sum_{i=1}^{M}\lambda_j\,\lambda_i\,m_{ij} \qquad (4.1.1)$$

To estimate the m_{ij} we can draw observations $y_i^{(h)}$, $h = 1, 2, \ldots, H$ from the pdf for each country $f_i(y)$, compute values $y_i^{(h)}\,F_j(y_i^{(h)})$, $j = 1, 2, \ldots, M$ for each draw, and take the averages. That is

$$\hat{m}_{ij} = \frac{1}{H}\sum_{h=1}^{H} y_i^{(h)}\,F_j(y_i^{(h)})$$

For large H (we chose $H = 10{,}000$), the $\hat{m}_{ij}$ will be accurate estimates of the m_{ij}.

4.2 Empirical analysis

Figure 3 presents the 1988 and 1993 regional income distributions as the weighted average of country's income distributions. For both years, the regional income distributions exhibit some degrees of bimodal distributions. The intuitive reason for the second mode towards the right tail seems to be due to the fact that Japan has relatively big population sizes for both years and for this reason the weight assigned for Japan is relatively large. And because the peaks of the density functions for Japan for both years are toward the right of other countries, together with the large weights it pulls the regional distribution up and exhibits a second mode. Figure 4 presents the 1988 and 1993 regional income distributions together for comparison in one graph. There is not an obvious shift in the regional distributions.

Figure 5 shows the regional Lorenz curves for 1988 and 1933. The regional Lorenz curves for both years are virtually identical. From Table 4, the regional Gini coefficients calculated using equation 4.1.1 are 0.4818 and 0.4802 for 1988 and 1993, respectively. These two regional Gini coefficients are almost identical, supporting the fact that the Lorenz curves look the same.

5. Conclusions

The main objective of the paper is to make improvements in the current approaches used in estimating global and regional inequality. The paper employs a more general specification of income distribution than the lognormal distribution that is used in past research, and at the same time relaxes the assumption of uniform distribution of income within quintile and decile groups of population. The paper outlines an econometric technique to estimate parameters of the generalised-Beta distribution when only limited data in the form of income shares of decile groups of the population are available. The empirical illustration method includes eight East Asian countries and the income distribution data are for the years 1988 and 1993. The empirical results demonstrate the feasibility of the econometric technique and the goodness-of-fit results are very encouraging. The paper also focuses on the derivation of regional income distributions using country-specific distributions. Properties of the regional distribution are examined by expressing the distribution as a mixture of income distributions of countries. Levels and trends in inequality in these countries and the region are examined. Properties based on Lorenz dominance are established. Standard decomposition analysis of inequality is also conducted. The empirical results show a clear increase in inequality in most of the East Asian countries over the period 1988 to 1993. The paper also identifies several avenues for further research. Based on the econometric technique developed here, the next step is to employ the methodology on a larger scale and derive improved estimates of inequality for the world and for more recent years for which data may become available. Further research will focus on the derivation of analytical properties of the mixture distribution used for purposes of studying regional inequality.

References

Chotikapanich, D. and D.S. Prasada Rao (1998), "Inequality in Asia 1975-1990: A Decomposition Analysis", *The Asia Pacific Journal of Economics and Business*, Vol. 2, No.1, 63-78.

Deininger, K. and L. Squire (1996), *Measuring Inequality: A new data base*, The World Bank, Washington, DC, Mimeographed.

ILO (1995), *Household Income and Expenditure Statistics, 1997-1991, No. 4*, International Labour Office, Geneva.

Kleiber, C. and S. Kotz (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, New York: John Wiley and Sons.

McDonald, J.B. (1984), "Some Generalized Functions of the Size Distribution of Income," *Econometrica*, 52, 647-663.

McDonald, J.B. and M.R. Ransom (1979), "Functional Forms, Estimation Techniques and the Distribution of Income", *Econometrics*, 6, 1513 – 1525.

Milanovic, B. (2002), "True World Income Distribution, 1988 and 1993: First Calculations based on Household Surveys Alone", *The Economic Journal*, 112, 51-92.

Wilfling, B. (1996), "Lorenz Ordering of Generalized Beta-II Income Distributions", *Journal of Econometrics*, 71, 381-388.

Table 1: Estimated coefficients from Beta Distributions

	1988	1993		1988	1993
HongKong			Singapore		
b	2746.5230	2958.5740	*b*	1456.9800	7077.6610
p	9.5631	8.6944	*p*	42.2231	6.0023
q	2.3293	2.0609	*q*	4.8383	3.0465
Japan			Korea		
b	6.0287	11.4284	*b*	4083.4100	27515.8700
p	16794.4200	9834.7310	*p*	7.7662	4.2235
q	6.0346	5.9103	*q*	4.6501	10.9322
Malaysia			Taiwan		
b	1337.3480	1800.073	*b*	997.8207	2336.1680
p	6.7416	6.1139	*p*	36.5787	22.1082
q	2.5691	2.4468	*q*	4.7087	4.9097
Philippines			Thailand		
b	308.3341	361.1595	*b*	11.6411	177.2501
p	17.6414	13.3538	*p*	480.8511	39.8189
q	2.8638	2.6737	*q*	2.3950	2.2117

Table 2: Income shares 1988

Hong Kong		Japan		Malaysia		Philippines	
actual	estimated	actual	estimated	actual	estimated	actual	estimated
0.019	0.020	0.039	0.039	0.020	0.020	0.040	0.039
0.032	0.031	0.052	0.051	0.031	0.031	0.052	0.053
0.040	0.039	0.065	0.065	0.040	0.040	0.058	0.059
0.049	0.048	0.075	0.075	0.050	0.050	0.066	0.067
0.059	0.058	0.082	0.083	0.060	0.061	0.074	0.075
0.071	0.071	0.092	0.093	0.073	0.073	0.086	0.086
0.086	0.087	0.103	0.104	0.090	0.090	0.100	0.099
0.108	0.111	0.121	0.122	0.115	0.115	0.120	0.117
0.149	0.156	0.147	0.147	0.159	0.158	0.152	0.149
0.387	0.380	0.224	0.221	0.362	0.362	0.252	0.256

Singapore		Korea		Taiwan		Thailand	
actual	estimated	actual	estimated	actual	estimated	actual	estimated
0.040	0.039	0.028	0.030	0.038	0.038	0.025	0.025
0.052	0.052	0.046	0.044	0.052	0.051	0.036	0.035
0.060	0.061	0.057	0.054	0.061	0.061	0.043	0.044
0.069	0.070	0.066	0.064	0.070	0.069	0.052	0.052
0.079	0.079	0.076	0.075	0.079	0.079	0.061	0.062
0.090	0.090	0.087	0.087	0.090	0.090	0.073	0.074
0.104	0.103	0.100	0.102	0.102	0.103	0.090	0.089
0.122	0.120	0.118	0.123	0.118	0.120	0.114	0.112
0.150	0.147	0.145	0.155	0.146	0.148	0.156	0.152
0.234	0.237	0.276	0.265	0.244	0.241	0.351	0.355

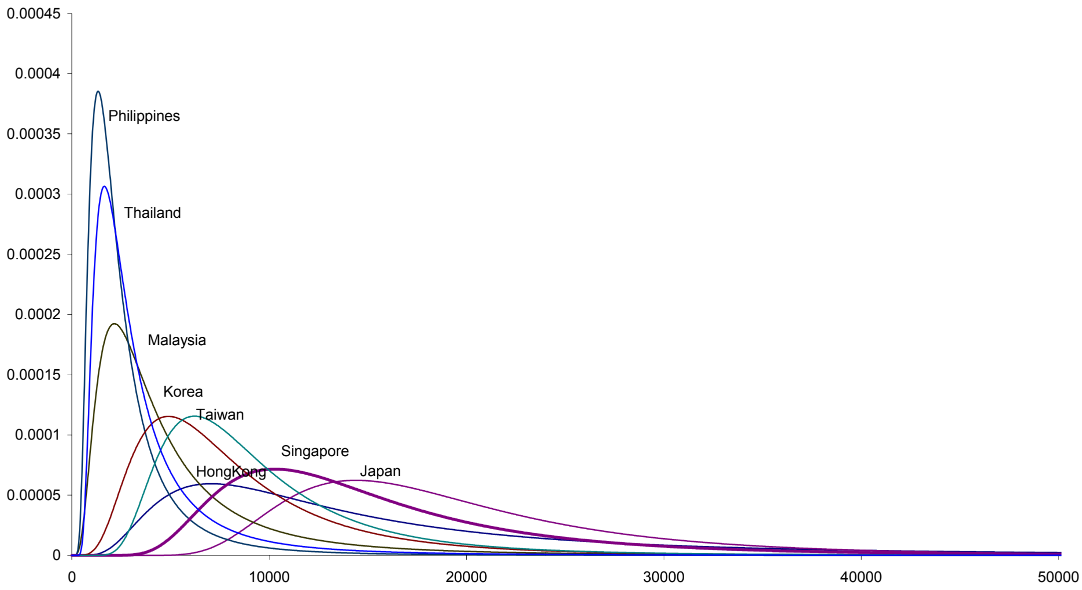
Note: All shares are decile shares with the exception of Japan for 1988 and 1993 and Philippines for 1988 where the population proportions were not equal for each class.

Table 3: Income shares 1993

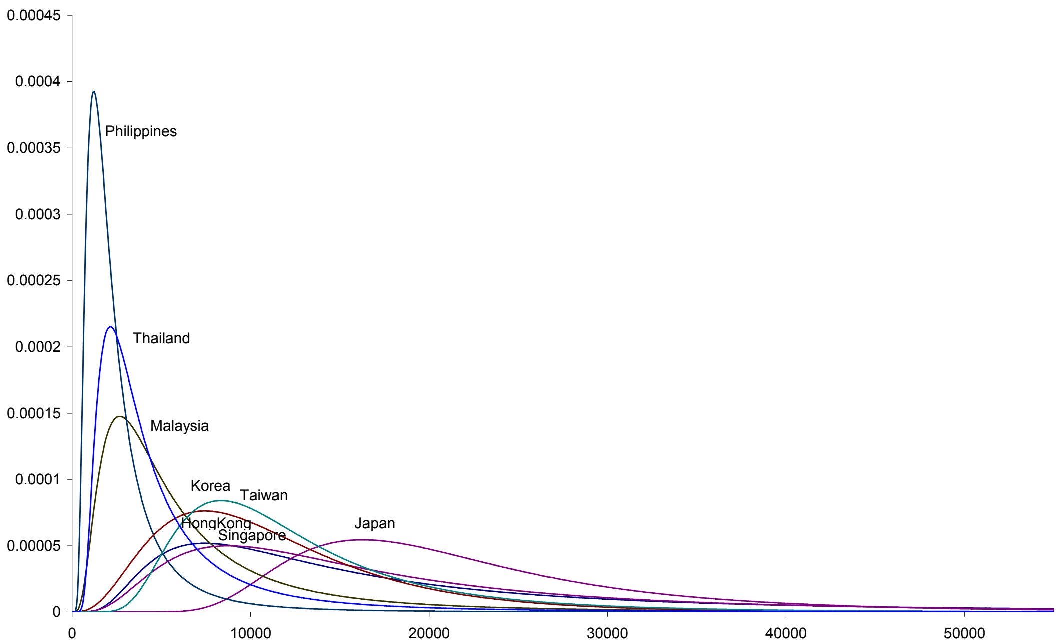
Hong Kong		Japan		Malaysia		Philippines	
actual	estimated	actual	estimated	actual	estimated	actual	estimated
0.016	0.017	0.038	0.039	0.018	0.018	0.024	0.024
0.028	0.027	0.053	0.050	0.029	0.029	0.035	0.035
0.036	0.035	0.064	0.063	0.039	0.039	0.043	0.044
0.045	0.043	0.075	0.074	0.048	0.048	0.052	0.054
0.055	0.053	0.080	0.080	0.058	0.058	0.063	0.064
0.066	0.065	0.090	0.091	0.072	0.071	0.076	0.076
0.080	0.082	0.102	0.103	0.089	0.088	0.093	0.092
0.101	0.106	0.119	0.120	0.114	0.113	0.117	0.115
0.140	0.153	0.147	0.147	0.158	0.158	0.161	0.156
0.432	0.419	0.235	0.233	0.376	0.376	0.335	0.340

Singapore		Korea		Taiwan		Thailand	
actual	estimated	actual	estimated	actual	estimated	actual	estimated
0.022	0.021	0.028	0.029	0.037	0.037	0.022	0.022
0.034	0.034	0.047	0.045	0.051	0.051	0.032	0.032
0.044	0.044	0.060	0.057	0.061	0.060	0.039	0.040
0.054	0.054	0.070	0.069	0.070	0.069	0.048	0.049
0.064	0.065	0.081	0.080	0.079	0.079	0.057	0.058
0.078	0.078	0.093	0.093	0.089	0.090	0.070	0.070
0.095	0.095	0.106	0.108	0.103	0.103	0.086	0.086
0.119	0.119	0.124	0.127	0.120	0.121	0.111	0.109
0.161	0.160	0.150	0.156	0.148	0.149	0.158	0.152
0.329	0.330	0.243	0.235	0.242	0.241	0.377	0.382

Figure 1: Cross-country distributions



1988



1993

Figure 2a: Cumulative Distribution Functions, 1988

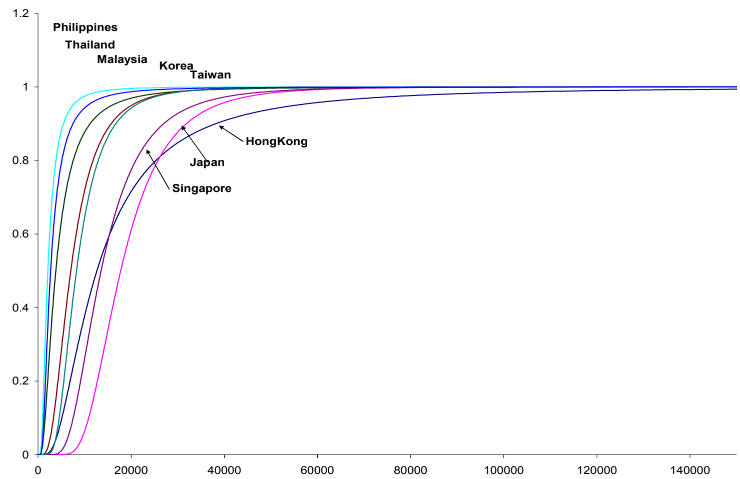


Figure 2b: Cumulative Distribution Functions, 1988

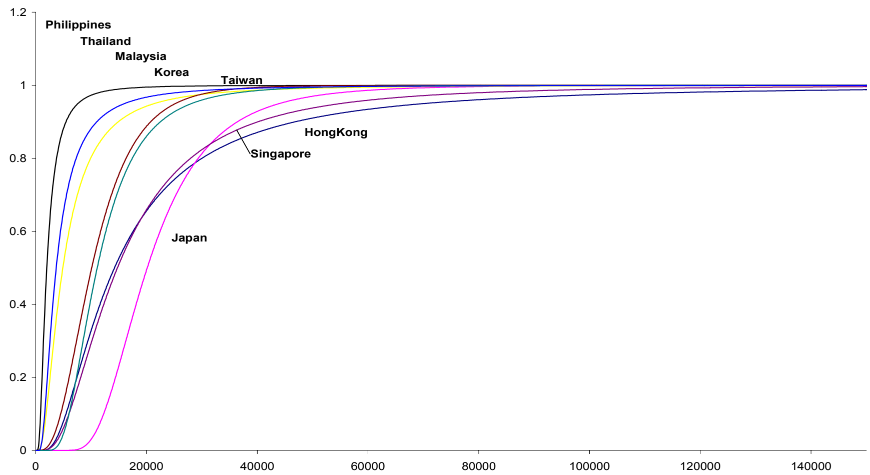


Figure 3a: Selected Lorenz Curves, 1988

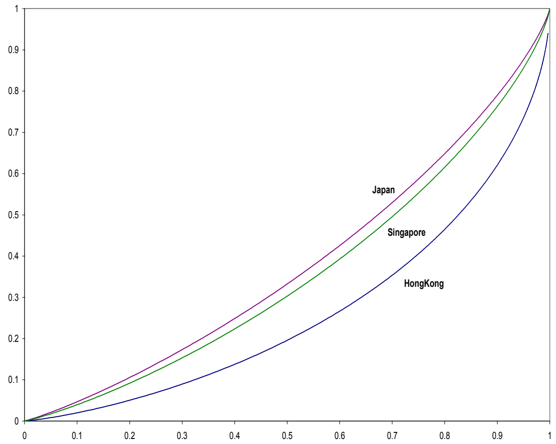


Figure 3b: Selected Lorenz Curves, 1993

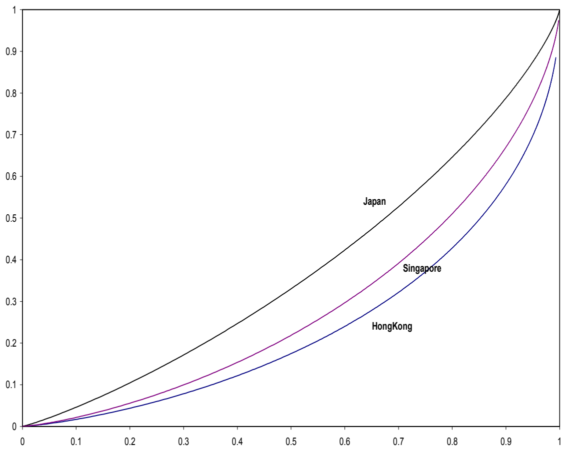


Figure 4: Shifts in the distributions over time

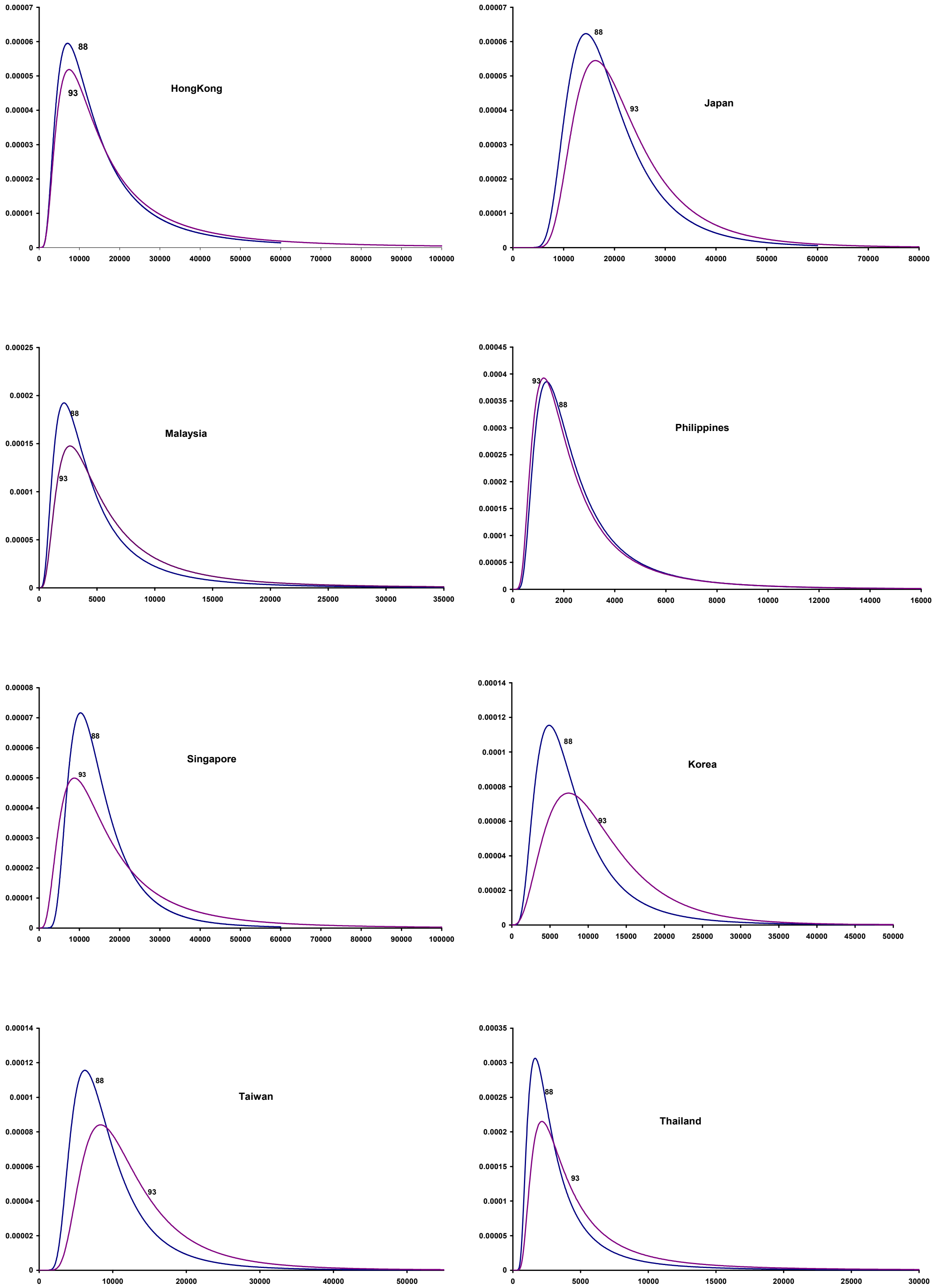
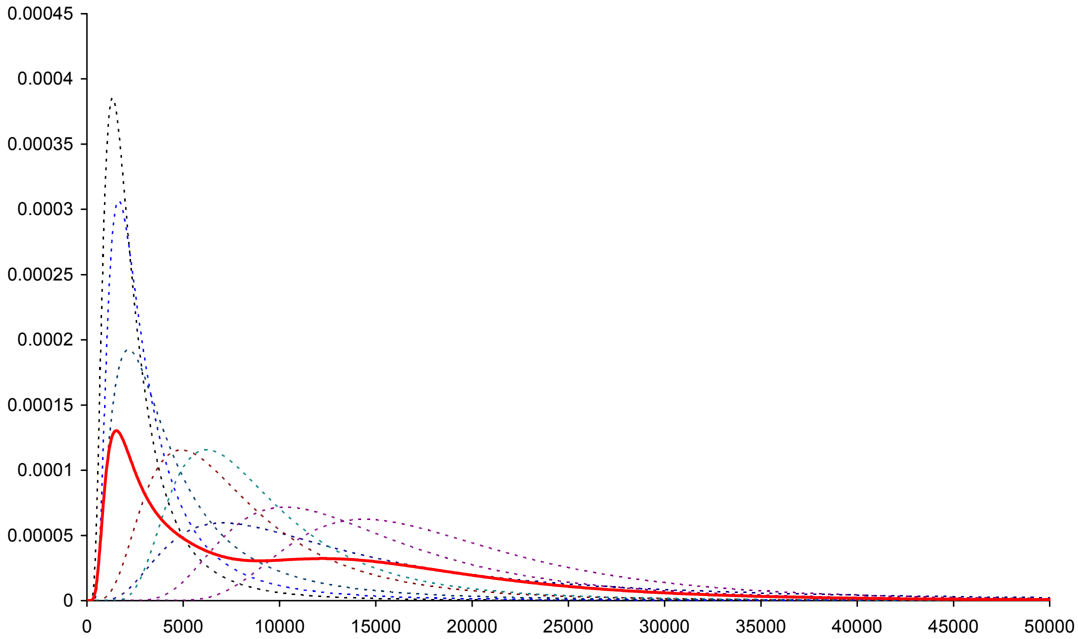


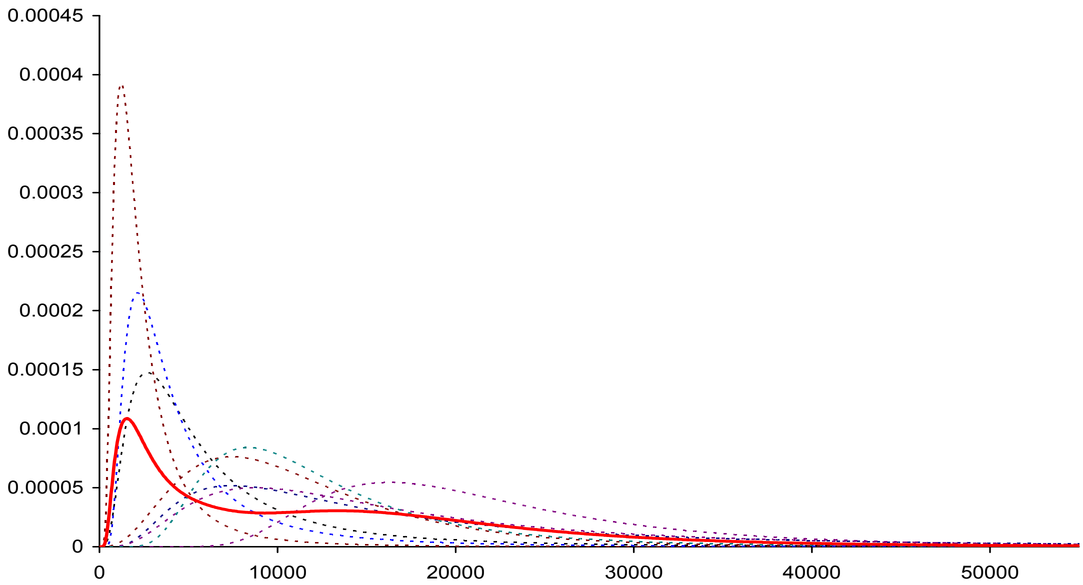
Table 4: Observed and Estimated Gini Coefficients

	1988		1993	
	Observed	Estimated	Observed	Estimated
HongKong	0.4598	0.4755	0.4974	0.5168
Japan	0.2433	0.2453	0.2415	0.2483
Malaysia	0.4474	0.4607	0.4629	0.4773
Philippines	0.4326	0.4064	0.4181	0.4293
Singapore	0.2858	0.2911	0.4167	0.4276
Korea	0.3351	0.3442	0.3097	0.3170
Taiwan	0.2903	0.2972	0.2931	0.2996
Thailand	0.4254	0.4381	0.4559	0.4704
Region		0.4818		0.4802

Figure 3: Regional income distributions as the weighted average



1988



1993

Note: The dotted lines are country income distributions. The solid lines are the weighted average regional income distribution.

Figure 4: Shift in regional distributions

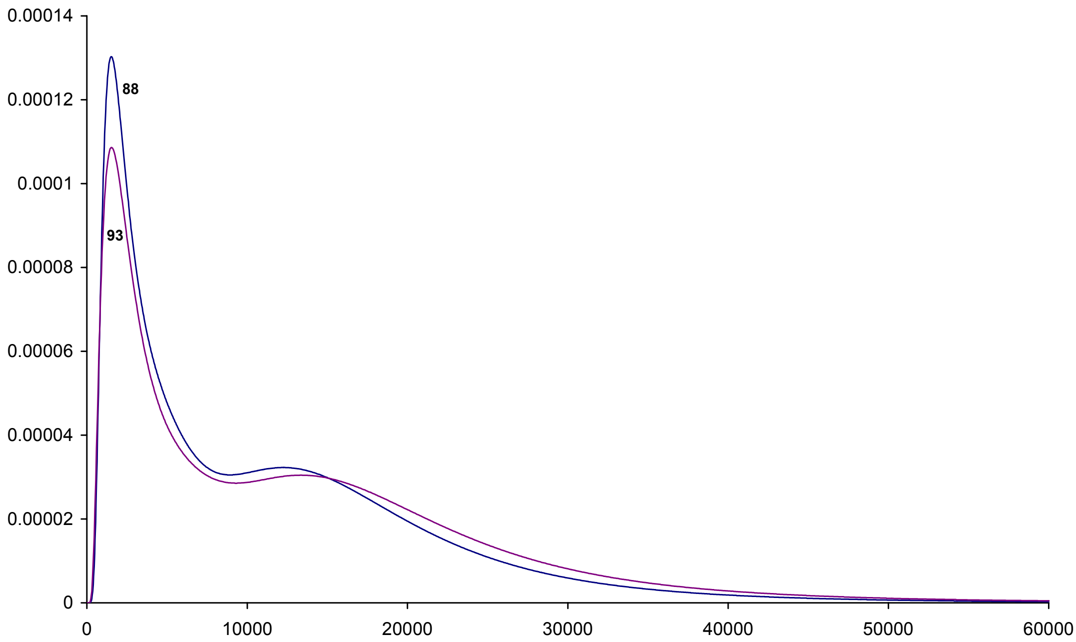


Figure 5: Regional Lorenz curves: 1988 and 1993