

ISSN 0819-2642  
ISBN 0 7340 2582 3



**THE UNIVERSITY OF MELBOURNE**  
**DEPARTMENT OF ECONOMICS**

RESEARCH PAPER NUMBER 926

FEBRUARY 2005

**Estimating and Combining National Income  
Distributions using Limited Data**

by

Duangkamon Chotikapanich  
&  
William E. Griffiths  
&  
D. S. Prasada Rao

Department of Economics  
The University of Melbourne  
Melbourne Victoria 3010  
Australia.

# **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*

December 22, 2004

## ***Abstract***

A major problem encountered in studies of income inequality at regional and global levels is the estimation of income distributions from data that are in a summary form. In this paper we estimate national and regional income distributions within a general framework that relaxes the assumption of constant income within groups. A technique to estimate the parameters of a beta-2 distribution 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; beta-2 distribution

***JEL classification numbers:*** C13, C16, D31

## 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. Central to the ongoing debate about globalization are the problem of measuring the extent of economic inequality and the need to be able to meaningfully compare inequality across countries, regions and time periods. Unless we can measure well and compare well, we cannot easily evaluate whether various policy initiatives that move towards greater globalization, or have other related impacts, are increasing or reducing inequality. The current literature on the subject (Chotikapanich et al, 1997; Bourguignon and Morrisson, 1999; Dowrick and Akmal, 2001; Milanovic, 2002a, 2002b; Sala-i-Martin, 2002a, 2002b; Bhalla, 2002; Quah, 1999, 2002) shows varying results in terms of the levels and trends of inequality. In general, results appear to be sensitive to the measures of income, inequality data and the methodology used in studying regional and global inequality.

A crucial component of the estimation of global inequality is the intra-country inequality measurement for all the countries included in the study. Much of the earlier work in the area of global inequality, due to Theil (1989), was focused on inter-country inequality and ignored inequality within each country. In recent years it has become common to use global inequality measures that incorporate measures of inequality within each country. Typically, when using data for cross-country and regional comparisons, limited information is available on the level of inequality within each country. Most available data are either in the form of some measure of inequality such as the Gini coefficient or in the form of income shares of quintile or decile groups. Chotikapanich, Valenzuela and Rao (1997) use information on Gini

coefficients – their work was based on the assumption that the income distribution in each country follows a lognormal distribution. More recent work by Milanovic (2002a), Sala-i-Martin (2002a and 2002b) and Dowrick and Akmal (2001) makes use of income share data for deciles and quintiles of each country's population, with Deininger and Squire (1996) being the primary source of data. An important consideration when using decile or quintile share data is the treatment of inequality within each population (e.g., decile) group. In some studies there is an implied assumption that there is no within-group inequality and each individual in an income group receives the same income. In other studies kernel smoothing has been used. Sala-i-Martin (2002a and 2002b) uses kernel smoothing methods to derive the distribution of income within each country and both Milanovic (2002a) and Sala-i-Martin (2002a and 2002b) use kernel smoothing techniques to derive the global distribution of income. This approach is also employed in the more recent work of Dowrick et al (2004) where global poverty estimates are computed from national and global income distributions derived using kernel smoothing.

In a recent critique of Sala-i-Martin (2002b), Milanovic (2002b) demonstrates the sensitivity of the estimates of levels of and trends in global income distribution to the methodology used. He demonstrates, using a simulated example, that use of kernel smoothing to derive country-specific income distributions from quintile share data can produce strange outcomes. Milanovic also provides evidence that global income distributions and measures appear to be sensitive to the choice of methodology used to analyze intra-country inequality when using limited data. An alternative assumption of equal distribution of income within each income quintile or decile is equally untenable when income distributions are known to be highly skewed.

There is a general recognition that estimates of global income inequality and its underlying distribution would be vastly improved if country-specific income distributions could be modelled adequately. The main constraint to this endeavour is the limited nature of data available. We are typically faced with a problem of using decile group data, with ten pieces of information, to model more general and complex distributions than the lognormal distribution used in Chotikapanich, Valenzuela and Rao (1997). The main objective of this paper is to describe a method for estimating the parameters of a relatively flexible form of income distribution using a limited amount of data. The success and effectiveness of the proposed methodology is assessed by applying goodness-of-fit criteria to the fitted income distribution. We establish the feasibility and usefulness of estimating a beta-2 distribution (McDonald, 1984), a distribution that is known to be flexible in modelling a variety of income distributions and known to fit income distributions well.

Given a parametric description of the distribution of income in each country, we show how global and regional distributions can be studied in detail by considering them as a mixture of the country-specific distributions with population shares as weights. Methods for measuring regional inequality and its decomposition from the mixture distribution are also outlined. The empirical example in the paper should be regarded as illustrative in nature, designed to demonstrate the feasibility of the proposed method. The data used are those compiled and used by Milanovic (2002a); they comprise mean income for each of a number of population groups ordered from poorest to richest. The proposed methodology is illustrated using data from eight selected East Asian countries: Hong Kong, Japan, Malaysia, Philippines, Singapore, Korea, Taiwan and Thailand. We fit a beta-2 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 adequacy of the beta distribution is assessed through a comparison of predicted and actual income shares and Gini coefficients.

An outline of the paper is as follows. In Section 2 we describe the basic data and their sources. The methodology used to estimate the parameters of the income distribution of a given country, assuming that it follows a beta-2 distribution, is discussed in Section 3. An analytical framework to study regional inequality as a mixture of the country-specific distributions is also outlined. In Section 4 we present empirical results from application of the methodology to the eight selected countries. The results show that inequality within countries increased over the period 1988 to 1993, but inequality within the region comprised by those countries did not change. Concluding remarks and some possible areas for further research are provided in Section 5.

## **2. Description of Data and Sources**

Compilation of data on income distributions from a large set of countries spanning a long period is a major research problem. Fortunately, the World Bank has long been a major provider of income-distribution data for the purpose of cross-country research. Recent work by Milanovic (2002a) is based on a set of cross-country data that he compiled for the World Bank. We use the same set of data. For each country class mean incomes (or expenditures) are given in local currency for a number of income classes, ranging from 10 to 20. For each income class the population share is known. Data are available for more than 100 countries for the years 1988 and 1993. The data on our selected East Asian countries come from this source with the exception of Singapore in 1988. Singapore was not included in the

data set for this year and so we use ILO (1995) as an alternative source for this case. This data set is different from the ones obtained from Milanovic (2002a) and it consists of decile expenditure shares.

Ideally distribution data should refer either to income or expenditure of persons or households. In the current data set, there is a mix of per capita income and per capita expenditure. Most data are for the distribution of incomes; the exceptions where the distribution of expenditures is used are Singapore in 1988, Philippines in both 1998 and 1993 and Thailand in 1993. These differences could influence the estimates of the parameters of the respective “income” distributions.

To derive the income distribution for the region as a whole, nominal per capita income for each country needs to be adjusted for differences in prices across countries, and for purposes of temporal welfare comparisons further adjustments are necessary for movements in prices over time. To describe how such adjustments are made, consider first the original data from each country in one particular year. Let  $\bar{x}_i$  = class mean income (or expenditure) in local currency, and  $c_i$  = population share for the  $i$ -th income class. Based on these data we calculate the income share for each income class as  $g_i = \bar{x}_i c_i / \sum \bar{x}_j c_j$ . To adjust for purchasing power parity (over countries and time) we obtain data on real per capita income from the latest version of the Penn World Tables, PWT 6.1,<sup>1</sup> which have data on real per capita incomes for over 150 countries spanning a 50-year period. PWT 6.1 also provides data on the population size of each of the countries. For each country and for a given year, let  $\bar{y}$  be the real per capita income adjusted for differences in prices across countries and over time and let  $S$  be the size of the population. For each income group in a country

---

<sup>1</sup> The URL for PWT 6.1 is [http://pwt.econ.upenn.edu/php\\_site/pwt\\_index.php](http://pwt.econ.upenn.edu/php_site/pwt_index.php).

the real class mean income for income class  $i$ ,  $\bar{y}_i$ , is derived as total income in the  $i$ -th group,  $g_i \bar{y} S$ , divided by total population in the  $i$ -th group,  $c_i S$ . That is,  $\bar{y}_i = g_i \bar{y} / c_i$ . Values for  $\bar{x}_i, c_i, \bar{y}$  and  $S$  that are used in this paper are given in Appendix Tables A1 and A2 for 1988 and 1993, respectively. The values  $\bar{y}_i$  and  $c_i$  are those used in the later analysis.

To provide an accurate assessment of the levels and trends in regional inequality, leading to a basis for informed debate on the effect of globalisation on inequality, our study should ideally 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 (2002a). At the time of preparation of this paper, it was not possible to obtain income distribution data for the most recent years.

### **3. Estimation of Country-Specific Income Distributions**

A large number of probability density functions have been suggested in the literature for modelling income distributions. See, for example, McDonald and Ransom (1979), McDonald (1984), McDonald and Xu (1995), Creedy and Martin (1997), Bandourian, McDonald and Turley (2002) and Kleiber and Kotz (2003). The one we have chosen for our analysis is a member of the McDonald family of distributions (see McDonald and Xu 1995) known as the beta-2 distribution. This distribution has analytical properties that make it well suited to the analysis, and, as we will see, it provides a very good fit to the observed data. The estimation problem is to estimate the beta-2 parameters for each of the countries in our study when only limited grouped data are available. A similar paper by Chotikapanich, Valenzuela and Rao (1997) used a lognormal distribution to model the income distributions for each country. In this case



the parameters of each lognormal distribution could be estimated relatively easily from information on the Gini coefficients and mean income for each country. The beta-2 distribution is a more flexible distribution, but estimating its parameters is more difficult, particularly when only grouped data are available and the class limits of the groups are not. In Section 3.1 we describe the beta-2 distribution and its characteristics. In Section 3.2 a method for estimating its parameters, and the class limits of the grouped data, is outlined. Methods for combining the country-specific income distributions and exploring the characteristics of the resulting regional income distribution are given in Section 3.3.

### 3.1 *The beta-2 income distribution*

The beta-2 distribution whose parameters  $b$ ,  $p$  and  $q$  we wish to estimate has probability density function (pdf)

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

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

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

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

$$F(y) = \frac{1}{B(p, q)} \int_0^{\lfloor y/(b+y) \rfloor} t^{p-1} (1-t)^{q-1} dt = B_{y/(b+y)}(p, q) \quad (2)$$

The function  $B_i(p, q)$  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

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

The mean, mode and variance of  $Y$  are given by

$$\begin{aligned} \mu &= \frac{bp}{q-1} & m &= \frac{(p-1)b}{q+1} \\ \sigma^2 &= \mu \left[ \frac{b(p+1)}{q-2} - \mu \right] = \frac{b^2 p(p+q-1)}{(q-1)^2(q-2)} \end{aligned} \quad (3)$$

The estimation procedure that we describe in Section 3.2 requires starting values for  $b$ ,  $p$  and  $q$ . It is often easier to suggest reasonable starting values for  $\mu$ ,  $m$  and  $\sigma^2$ . In this case corresponding values for  $b$ ,  $p$  and  $q$  can be found from the results:

$$\begin{aligned} b &= \frac{\mu^2(\mu - m) - (3m - \mu)\sigma^2}{\sigma^2 - \mu^2 + \mu m} \\ p &= \frac{\mu}{b} \left( \frac{2m + b}{\mu - m} \right) & q &= \frac{\mu + m + b}{\mu - m} \end{aligned} \quad (4)$$

For future reference we note that the Gini coefficient is given by

$$G = \frac{2B(2p, 2q-1)}{pB^2(p, q)} \quad (5)$$

### 3.2 Method of estimation of parameters of beta-2 distribution

Suppose we have  $N$  income classes  $(a_0, a_1), (a_1, a_2), \dots, (a_{N-1}, a_N)$ , with  $a_0 = 0$  and  $a_N = \infty$ . Let the mean class incomes for each of the  $N$  classes be given by  $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_N$ ; and let the population proportions for each class be given by  $c_1, c_2, \dots, c_N$ . Given available data on the  $\bar{y}_i$  and the  $c_i$ , but not the  $a_i$ , our problem is to estimate the parameters of a beta-2 distribution, along with the unknown class limits  $a_1, a_2, \dots, a_{N-1}$ . One approach is to fit a beta distribution to the data such that the sample moments  $\bar{y}_i$  and  $c_i$  are “close” to their population counterparts. This approach is equivalent to fitting a distribution such that  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{2N}$  are “close to zero” where

$$\int_{a_{i-1}}^{a_i} f(y)dy - c_i = \varepsilon_i \quad i = 1, 2, \dots, N \quad (6)$$

and

$$\frac{\int_{a_{i-1}}^{a_i} yf(y)dy}{\int_{a_{i-1}}^{a_i} f(y)dy} - \bar{y}_i = \varepsilon_{N+i} \quad i = 1, 2, \dots, N \quad (7)$$

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

$$B_{a_i/(b+a_i)}(p, q) - B_{a_{i-1}/(b+a_{i-1})}(p, q) - c_i = \varepsilon_i \quad (8)$$

and

$$\frac{bp}{q-1} \left( \frac{B_{a_i/(b+a_i)}(p+1, q-1) - B_{a_{i-1}/(b+a_{i-1})}(p+1, q-1)}{B_{a_i/(b+a_i)}(p, q) - B_{a_{i-1}/(b+a_{i-1})}(p, q)} \right) - \bar{y}_i = \varepsilon_{N+i} \quad (9)$$

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

$$y f(y) = \frac{bp}{q-1} f^*(y) \quad (10)$$

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

$$w_i = B_{a_i/(b+a_i)}(p, q) - B_{a_{i-1}/(b+a_{i-1})}(p, q) \quad (11)$$

and

$$z_i = \frac{B_{a_i/(b+a_i)}(p+1, q-1) - B_{a_{i-1}/(b+a_{i-1})}(p+1, q-1)}{w_i} \quad (12)$$

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

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

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

Equations (8) and (9) can now be written as  $w_i - c_i = \varepsilon_i$  and  $(bp/(q-1))z_i - \bar{y}_i = \varepsilon_{N+i}$ , or, alternatively,

$$x = \sum_{i=1}^N d_i w_i + \frac{bp}{q-1} \sum_{i=1}^N d_{N+i} z_i + \varepsilon \quad (13)$$

Initially, we estimated the class limits and beta distribution parameters by finding those values of  $(b, p, q, a_1, a_2, \dots, a_{N-1})$  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,  $(bp/(q-1))z_i - \bar{y}_i = \varepsilon_{N+i}$ . It was possible to get estimates such that  $\hat{a}_{i-1} > \hat{a}_i$ . We overcame this problem, and ensured all  $2N$  equations played their part in estimation, by minimizing the sum of squares of percentage errors  $\varepsilon'V^{-1}\varepsilon$  where  $V = \text{diagonal}(x_1^2, x_2^2, \dots, x_{2N}^2)$ .

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 (4). 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  $(a_1, a_2, \dots, a_{N-1})$  were obtained as  $a_i = (\bar{y}_i + \bar{y}_{i+1})/2$ .

### 3.3 Modelling regional income distributions

After estimating the country-specific income distributions we are in a position to combine them to form a regional income distribution. Given  $M$  countries each with a beta income pdf  $f_j(y)$ ,  $j = 1, 2, \dots, M$ , and population proportions  $\lambda_1, \lambda_2, \dots, \lambda_M$ , the pdf for the regional income distribution is given by the mixture

$$f(y) = \sum_{j=1}^M \lambda_j f_j(y) \quad (14)$$

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

$$F(y) = \sum_{j=1}^M \lambda_j F_j(y) = \sum_{j=1}^M \lambda_j B_{y/(y+b_j)}(p_j, q_j) \quad (15)$$

Regional mean income is given by

$$\mu = \sum_{j=1}^M \lambda_j \mu_j = \sum_{j=1}^M \frac{\lambda_j b_j p_j}{q_j - 1} \quad (16)$$

The regional cumulative income shares are given by

$$\begin{aligned} \eta(y) &= \frac{1}{\mu} \int_0^y z f(z) dz \\ &= \frac{1}{\mu} \sum_{j=1}^M \lambda_j \int_0^y z f_j(z) dz \\ &= \frac{1}{\mu} \sum_{j=1}^M \lambda_j \mu_j B_{y/(y+b_j)}(p_j + 1, q_j - 1) \end{aligned} \quad (17)$$

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

A regional cumulative distribution function can be graphed by using (15) to compute  $F(y)$  for a grid of values of  $y$ . A regional Lorenz curve, relating income shares to population shares, can be graphed by using (15) and (17) to compute  $F(y)$  and  $\eta(y)$  for a grid of values of  $y$ .

The regional Gini coefficient is given by (see, for example, Lambert (1993, p.43))

$$\begin{aligned} G &= -1 + \frac{2}{\mu} \int_0^{\infty} y F(y) f(y) dy \\ &= -1 + \frac{2}{\mu} \int_0^{\infty} y \left( \sum_{j=1}^M \lambda_j F_j(y) \right) \left( \sum_{i=1}^M \lambda_i f_i(y) \right) dy \\ &= -1 + \frac{2}{\mu} \sum_{j=1}^M \sum_{i=1}^M \lambda_j \lambda_i \int_0^{\infty} y F_j(y) f_i(y) dy \end{aligned} \quad (18)$$

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

$$\int_0^{\infty} y F_i(y) f_i(y) dy = \mu_i \left[ \frac{B(2p_i, 2q_i - 1)}{p_i B^2(p_i, q_i)} + \frac{1}{2} \right] \quad (19)$$

$$= \frac{\mu_i}{2} (G_i + 1)$$

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

$$m_{ij} = \int_0^{\infty} y F_j(y) f_i(y) dy = E_{f_i} [y F_j(y)] \quad (20)$$

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

$$G = -1 + \frac{2}{\mu} \sum_{j=1}^M \sum_{i=1}^M \lambda_j \lambda_i m_{ij} \quad (21)$$

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

$$\hat{m}_{ij} = \frac{1}{H} \sum_{h=1}^H y_i^{(h)} F_j(y_i^{(h)}) \quad (22)$$

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

#### 4. Empirical Analysis

Our presentation and discussion of the results begins in Section 4.1 with consideration of the estimated income distributions for the eight East Asian countries. Goodness-of-fit of the distributions is assessed in Section 4.2. Levels and trends in inequality are

examined in Section 4.3. In Section 4.4 we discuss the regional income distribution and compare regional inequality over the two years.

#### 4.1 *Country-specific income distributions*

Table 1 shows the estimated parameters of the beta distributions obtained using the procedure described in Section 3.2. The estimated parameters provide meaningful income distributions, all of which are skewed and uni-modal. However, the very large values of  $p$  for Japan (and to a lesser extent Thailand in 1988) appear out of place. Further investigation of Japan revealed that the EViews program used to perform the calculations 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 quite different parameter estimates in 1988 and 1993 may be explained by the different sources of data. 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; they are consistent with general expectations. The locations of the distributions in terms of the mode and the mean appear to be ordered according to the real per capita incomes of these countries. Also informative are the distribution functions and Lorenz curves for each country in each of the two years. To find them we select a grid of income points  $(y_1, y_2, \dots, y_L)$  and compute  $F(y_i) = B_{y_i/(b+y_i)}(p, q)$  and  $\eta_i = B_{y_i/(b+y_i)}(p+1, q-1)$ . 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 that level, followed by Thailand and then the other countries. The ranking of these countries remained unaltered over the two periods. However, such a 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. If interest centers on inequality only, with no concern for mean income, these figures show clear Lorenz ordering with Japan having the least inequality, followed by Singapore and Hong Kong.

#### 4.2 *Goodness-of-fit of beta distributions*

It is useful to assess the goodness of fit of the beta distributions by comparing the observed income shares with the expected income shares derived using the estimated distributions. The empirical income shares are given by

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

To find those implied by each beta distribution we began with the population shares  $c_i$ , and corresponding cumulative proportions

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

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

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

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

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

The estimated income shares are given by

$$\hat{g}_i = \hat{\eta}_i - \hat{\eta}_{i-1}$$

A comparison of the estimated and observed income shares appears in Tables 2 and 3. The actual (observed) and estimated (expected) income shares are remarkably similar for all the countries in both years. In most cases the differences are in the third decimal place. This outcome is very encouraging given that the parameters of the distributions have been estimated from limited data, and given that the class limits  $a_i$  implied by the estimated parameters, not the  $a_i$  giving the “best fit”, were used to compute the expected income proportions.

#### ***4.3 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. The Philippines, Korea and Taiwan are worthy of special mention. The income distribution in the Philippines remained virtually unchanged over the period whereas major structural shifts are evident in the case of Korea and Taiwan which have been labelled as the Asian tigers for their performance during the study period.

The levels and trends in inequality can be studied using Gini coefficients and Lorenz curves. Both observed and estimated Gini coefficients are computed and presented. The observed values of the Gini coefficient were obtained by applying the formula

$$G = \sum_{i=1}^N \eta_{i+1} \pi_i - \sum_{i=1}^N \eta_i \pi_{i+1}$$

to the grouped data. The estimated values were obtained by substituting estimates  $\hat{b}$ ,  $\hat{p}$ , and  $\hat{q}$  into the formula

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

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 a sufficient condition described in Wilfling (1996). 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 than (lies above) or equal to the Lorenz curve of  $H$ . Given that the 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)

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

The observed and estimated Gini coefficients for all countries are presented in Table 4. Overall, the estimated Gini's are higher than the observed ones. This outcome is expected because the Gini's estimated from the beta distribution take into account the distribution of income within classes. Trends in inequality shown in Table 4 are also interesting. With the exception of Korea, inequality within each country has

increased over the period 1988 to 1993. This result 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, the two coefficients may not be directly comparable because the data for the year 1993 were drawn from Milanovic (referring to income data) and the 1988 data were drawn from the ILO and refer to the expenditure distribution.

It is possible to draw conclusions on Lorenz dominance using the sufficient condition given above. Comparing estimated values of  $p$  and  $q$  for the years 1988 and 1993 shows that the 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.

The results reported for each of the countries demonstrate the feasibility of using the beta-2 distribution to model the distribution of income for the chosen Asian countries. The estimation procedure discussed in Section 3.2 provides a method for estimating the parameters of the distribution using grouped data in the form of population shares and class mean incomes. 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.

#### **4.4 Regional inequality**

In Figure 3 the 1988 and 1993 regional income distributions appear as weighted averages of the income distributions for each country. For both years, the regional income distributions exhibit some degree of bimodality. The apparent reason for the second mode towards the right tail is the relatively large population of Japan, leading to a relatively large weight being assigned to its distribution, coupled with the fact that the modes of Japan's distributions are to the right of those for other countries. For comparative purposes, in Figure 4 we present the 1988 and 1993 regional income distributions together in one graph. There is not an obvious shift in the regional distributions.

The regional Lorenz curves for 1988 and 1993 are shown in Figure 5. They are virtually identical. Visually, no difference can be detected from the Figure. As expected, the regional Gini coefficients calculated using equation (21) are also almost identical, being equal to 0.4818 and 0.4802 for 1988 and 1993, respectively.

### **5. Conclusions**

The main objective of the paper is to suggest improvements to current approaches used for estimating global and regional inequality. We employ an income distribution specification that is more general than the lognormal distribution that has been used in past research, and, at the same time, we relax the assumption of a uniform distribution of income within (quintile and decile) groups of population. Also, we describe a technique for estimating the parameters of the beta-2 distribution when only limited data in the form of population shares and class mean incomes for groups of the population are available. The empirical illustration comprises eight East Asian countries with income distribution data are for the years 1988 and 1993. The

empirical results demonstrate the feasibility of the technique and the goodness-of-fit results support its usefulness. 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 the income distributions of each country. Levels and trends in inequality in these countries and the region are examined. Properties based on Lorenz dominance are established. The empirical results show a clear increase in inequality in most of the East Asian countries over the period 1988 to 1993. There are several avenues for further research. Based on the technique developed here, the next step is to employ the methodology on a larger scale to derive improved estimates of inequality for the world, and for more recent years for which data may become available. Further research will also focus on the derivation of analytical properties of the mixture distribution used for purposes of studying regional inequality.

Table A1: Population Shares and Class Mean Income 1988

Population Shares ( $c_i$ )	Class Mean Income ( $\bar{x}_i$ )					Japan		Philippines		Singapore	
	Hong Kong	Malaysia	S. Korea	Taiwan	Thailand	Pop Shares ( $c_i$ )	Mean income ( $\bar{x}_i$ )	Pop Shares ( $c_i$ )	Mean income ( $\bar{x}_i$ )	Pop Shares ( $c_i$ )	Income Shares ( $g_i$ )
10.0	969.0	47.0	783 279.6	30 171.4	4 110.7	8.6	73 915.5	13.7	1 781.2	10.0	2.2
10.0	1 632.0	75.0	1 276 662.2	41 341.0	5 800.5	8.9	943 925.2	12.4	2 581.5	10.0	3.5
10.0	2 040.0	97.0	1 574 921.7	48 631.9	7 035.3	9.9	1 069 767.4	11.2	3 187.7	10.0	4.5
10.0	2 499.0	119.0	1 850 881.4	55 736.0	8 367.7	10.2	1 184 573.0	10.5	3 829.0	10.0	5.5
10.0	3 009.0	145.0	2 118 479.7	63 156.7	9 895.0	10.2	1 309 333.3	9.9	4 564.2	10.0	6.6
10.0	3 621.0	176.0	2 416 738.2	71 286.6	11 844.7	10.2	1 456 919.1	9.6	5 493.7	10.0	8.0
10.0	4 386.0	217.0	2 790 260.5	81 423.1	14 525.6	10.2	1 642 487.0	9.2	6 714.2	10.0	9.7
10.0	5 508.0	275.0	3 289 217.0	94 181.8	18 506.4	10.5	1 879 487.2	8.7	8 440.7	10.0	12.2
10.0	7 599.0	381.0	4 047 409.4	115 827.9	25 298.0	10.6	2 253 846.2	8.1	11 422.9	10.0	16.3
10.0	19 788.0	869.0	7 698 998.7	194 204.2	57 095.2	10.8	3 375 314.9	6.7	22 856.7	10.0	31.6
Total Pop ( $S$ )	5 626 600	17 144 390	42 031 000	19 357 000	53 687 208	122 610 000		59 369 000			
Mean Income ( $\bar{y}$ )	19 774.3	5 746.4	8 714.8	9 843.6	4 015.4	20 118.6		2 920.5			

Note: Source of data for  $S$  and  $\bar{y}$  are from PWI6.1. For  $c_i$  and  $\bar{x}_i$  for all countries except Singapore are from Milanovic (2002a).  $c_i$  and  $g_i$  for Singapore are from ILO (1995).

**Table A2: Population Shares and Class Mean Income 1993**

Population Shares ( $c_i$ )	Class Mean Income ( $\bar{x}_i$ )							Japan	
	Hong Kong	Malaysia	S. Korea	Taiwan	Thailand	Singapore	Philippines	Pop Shares ( $c_i$ )	Class Mean Income ( $\bar{x}_i$ )
10.0	1 600.0	879.0	1 493 547.1	56 662.5	4 718.4	182.0	3 116.5	8.6	917 603.0
10.0	2 800.0	1 403.1	2 563 469.9	77 749.4	6 694.8	287.0	4 487.6	8.9	1 248 275.9
10.0	3 600.0	1 846.3	3 231 492.8	92 186.6	7 035.3	369.0	5 548.8	9.7	1 364 741.6
10.0	4 500.0	2 286.0	3 801 756.2	106 145.3	8 318.4	453.0	6 691.0	10.2	1 508 620.7
10.0	5 500.0	2 793.9	4 388 312.9	120 442.6	10 042.8	541.0	8 029.7	9.9	1 676 880.2
10.0	6 600.0	3 422.9	5 034 611.4	136 289.6	12 034.8	659.0	9 717.1	10.1	1 869 209.8
10.0	8 000.0	4 246.9	5 740 651.9	156 443.0	14 659.2	799.0	11 904.7	10.2	2 088 235.3
10.0	10 100.0	5 435.0	6 718 246.3	182 645.8	18 132.0	1 000.0	15 034.3	10.4	2 380 952.4
10.0	14 100.0	7 547.6	8 168 344.8	225 716.0	23 492.4	1 354.0	20 619.4	10.7	2 855 643.0
10.0	43 200.0	17 996.0	13 170 369.7	368 717.1	33 236.4	2 767.0	42 820.5	11.3	4 352 644.8
Total Pop ( $S$ )	5 901 000	19 609 110	44 195 000	20 848 250	58 064 000	3 315 000	67 092 660		124 670 000
Mean Income ( $\bar{y}$ )	24 292.8	7 606.4	11 717.3	13 211.1	5 832.1	20 761.3	2 884.6		22 906.2

Note: Source of data for  $S$  and  $\bar{y}$  are from PWI6.1. For  $c_i$  and  $\bar{x}_i$  for all countries are from Milanovic (2002a).



## References

- Bandourian, R., J.B. McDonald and R.S. Turley (2002), "A Comparison of Parametric Models of Income Distribution Across Countries and Over Time", Mimeo. Brigham Young University.
- Bhalla, S.S. (2002), *Imagine there is no country*, Washington: Institute for International Economics.
- Bourguignon, F., and C. Morrisson (1999). "The size distribution of income among world citizens: 1820-1990." Mimeo. DELTA, Paris.
- Chotikapanich, D., M.R. Valenzuela and D.S.P. Rao (1997), "Global and Regional Inequality in the Distribution of Income: Estimation with Limited/Incomplete Data", *Empirical Economics*, 20, 533-546.
- 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.
- Creedy, J. and V.L. Martin, editors (1997), *Nonlinear Economic Models: Cross-sectional, Time Series and Neural Network Applications*, Cheltenham: Edward Elgar.
- Deininger, K. and L. Squire (1996), *Measuring Inequality: A new data base*, The World Bank, Washington, DC, Mimeographed.
- Dowrick, S., R. Eckland and B. Freyens (2004), "Global Poverty Measurement: Why PPP Methods Matter?", Paper presented at the 28th General conference of the International Association for Research in Income and Wealth, 22-28 August, 2004, Cork, Ireland.

- Dowrick, S. and Muhammed Akmal (2001), "Contradictory Trends in Global Inequality: A Tale of Two Biases", Draft 29 March, 2001, available from <http://ecocomm.anu.edu.au/economics/staff/dowrick/dowrick.html>.
- 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.
- Lambert, P. J. (1993), *The Distribution and Redistribution of Income*, Manchester: Manchester University Press.
- 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", *Econometrica*, 6, 1513 – 1525.
- McDonald, J.B. and Y.J. Xu (1995), "A Generalization of the Beta Distribution with Applications", *Journal of Econometrics* 66, 133-152. Errata 69 (1995), 427-428.
- Milanovic, B. (2002a), "True World Income Distribution, 1988 and 1993: First Calculations based on Household Surveys Alone", *The Economic Journal*, 112, 51-92.
- Milanovic, B. (2002b), "The Ricardian Vice: Why Sala-i-Martin's calculations of world income inequality are wrong?", mimeographed, Carneige Mellon University.
- Quah, D. (1999), "Calculations of world income distribution", unpublished notes.

- Quah, D. (2002), "One third of the world's growth and inequality", mimeographed.
- Sala-i-Martin, X. (2002a) "The Disturbing "Rise" Of Global Income Inequality, NBER Working Paper No. 8904: 1-75.
- Sala-i-Martin, X. (2002b) "The World Distribution of Income (Estimated from Individual Country Distributions), NBER Working Paper No. 8905: 1-68.
- Theil, H. (1989) "The Development of International Inequality, 1960 – 1985", *Journal of Econometrics*, 42, 145-155.
- 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		
<i>b</i>	2746.5230	2958.5740	<i>b</i>	1456.9800	7077.6610
<i>p</i>	9.5631	8.6944	<i>p</i>	42.2231	6.0023
<i>q</i>	2.3293	2.0609	<i>q</i>	4.8383	3.0465
Japan			Korea		
<i>b</i>	6.0287	11.4284	<i>b</i>	4083.4100	27515.8700
<i>p</i>	16794.4200	9834.7310	<i>p</i>	7.7662	4.2235
<i>q</i>	6.0346	5.9103	<i>q</i>	4.6501	10.9322
Malaysia			Taiwan		
<i>b</i>	1337.3480	1800.073	<i>b</i>	997.8207	2336.1680
<i>p</i>	6.7416	6.1139	<i>p</i>	36.5787	22.1082
<i>q</i>	2.5691	2.4468	<i>q</i>	4.7087	4.9097
Philippines			Thailand		
<i>b</i>	308.3341	361.1595	<i>b</i>	11.6411	177.2501
<i>p</i>	17.6414	13.3538	<i>p</i>	480.8511	39.8189
<i>q</i>	2.8638	2.6737	<i>q</i>	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

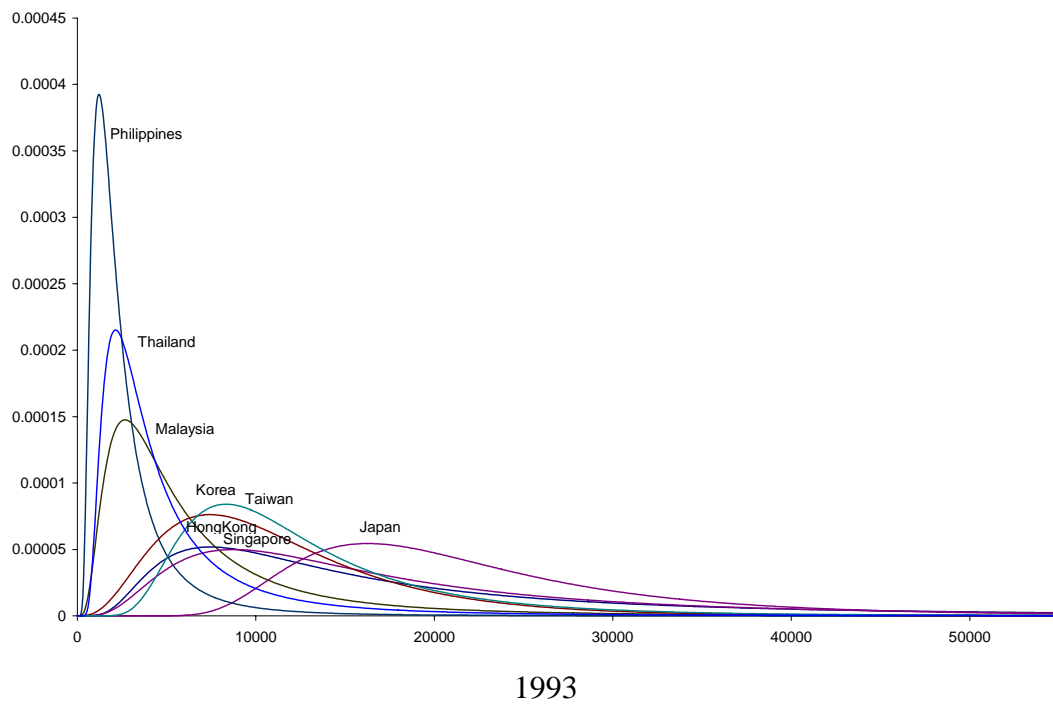
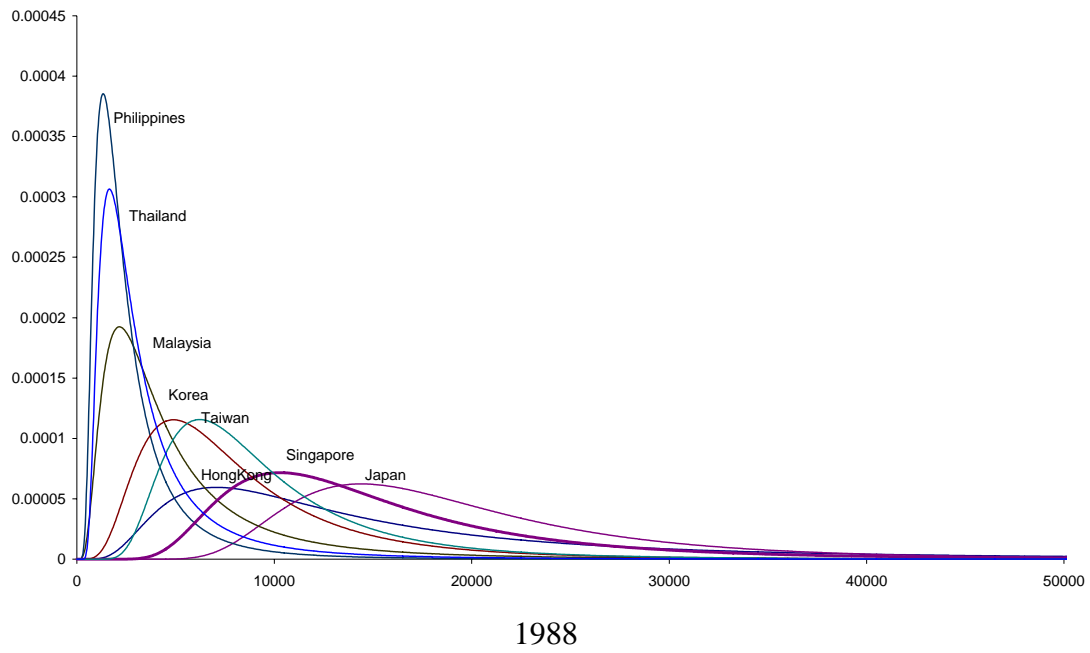
Table 3: Income shares 1993

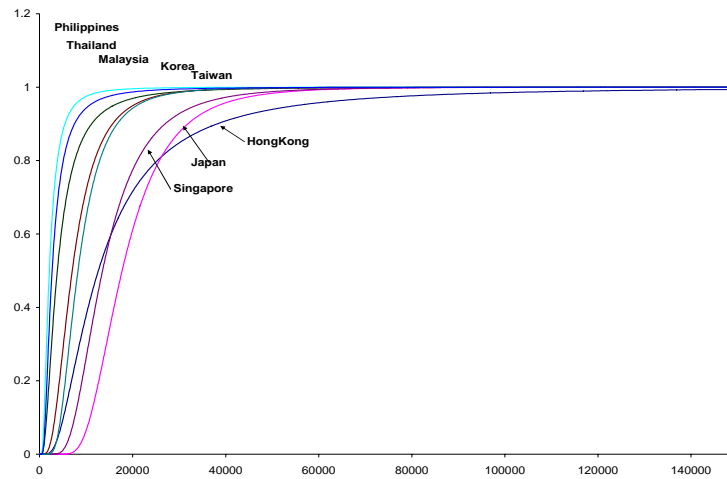
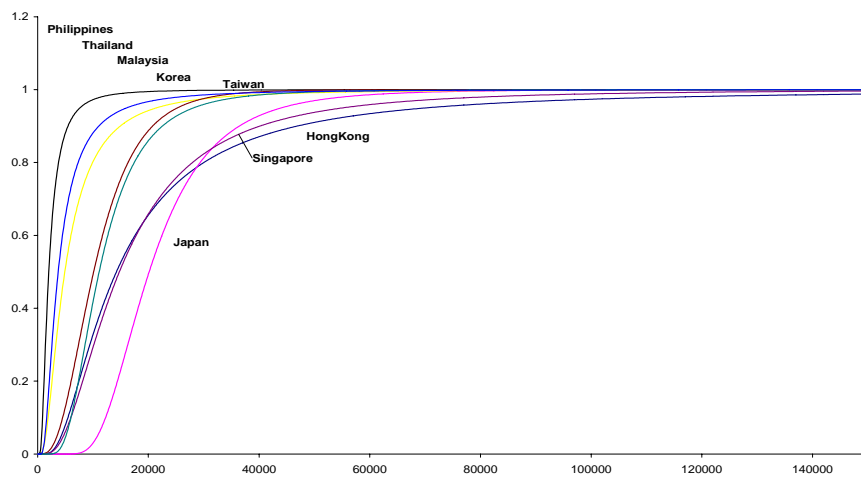
Hong Kong		Japan		Malaysia		Philippines	
actual	estimated	actual	estimated	actual	estimated	actual	estimated
0.016	0.017	0.038	0.039	0.018	0.018	0.024	0.024
0.028	0.027	0.053	0.050	0.029	0.029	0.035	0.035
0.036	0.035	0.064	0.063	0.039	0.039	0.043	0.044
0.045	0.043	0.075	0.074	0.048	0.048	0.052	0.054
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

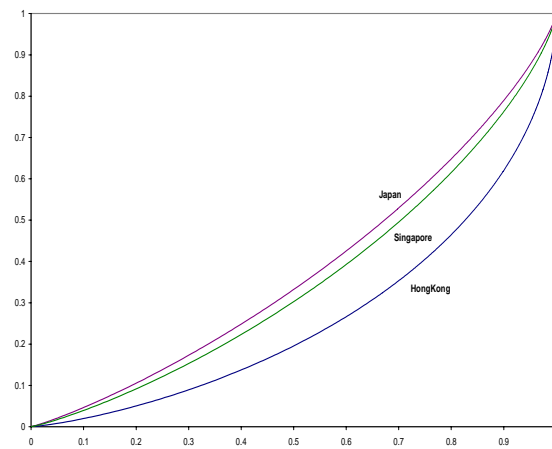
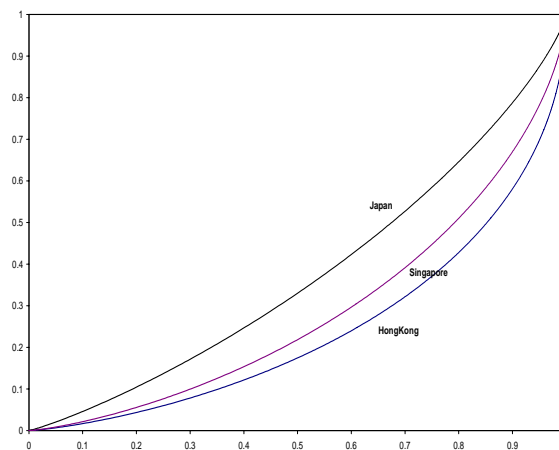
  

Singapore		Korea		Taiwan		Thailand	
actual	estimated	actual	estimated	actual	estimated	actual	estimated
0.022	0.021	0.028	0.029	0.037	0.037	0.022	0.022
0.034	0.034	0.047	0.045	0.051	0.051	0.032	0.032
0.044	0.044	0.060	0.057	0.061	0.060	0.039	0.040
0.054	0.054	0.070	0.069	0.070	0.069	0.048	0.049
0.064	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

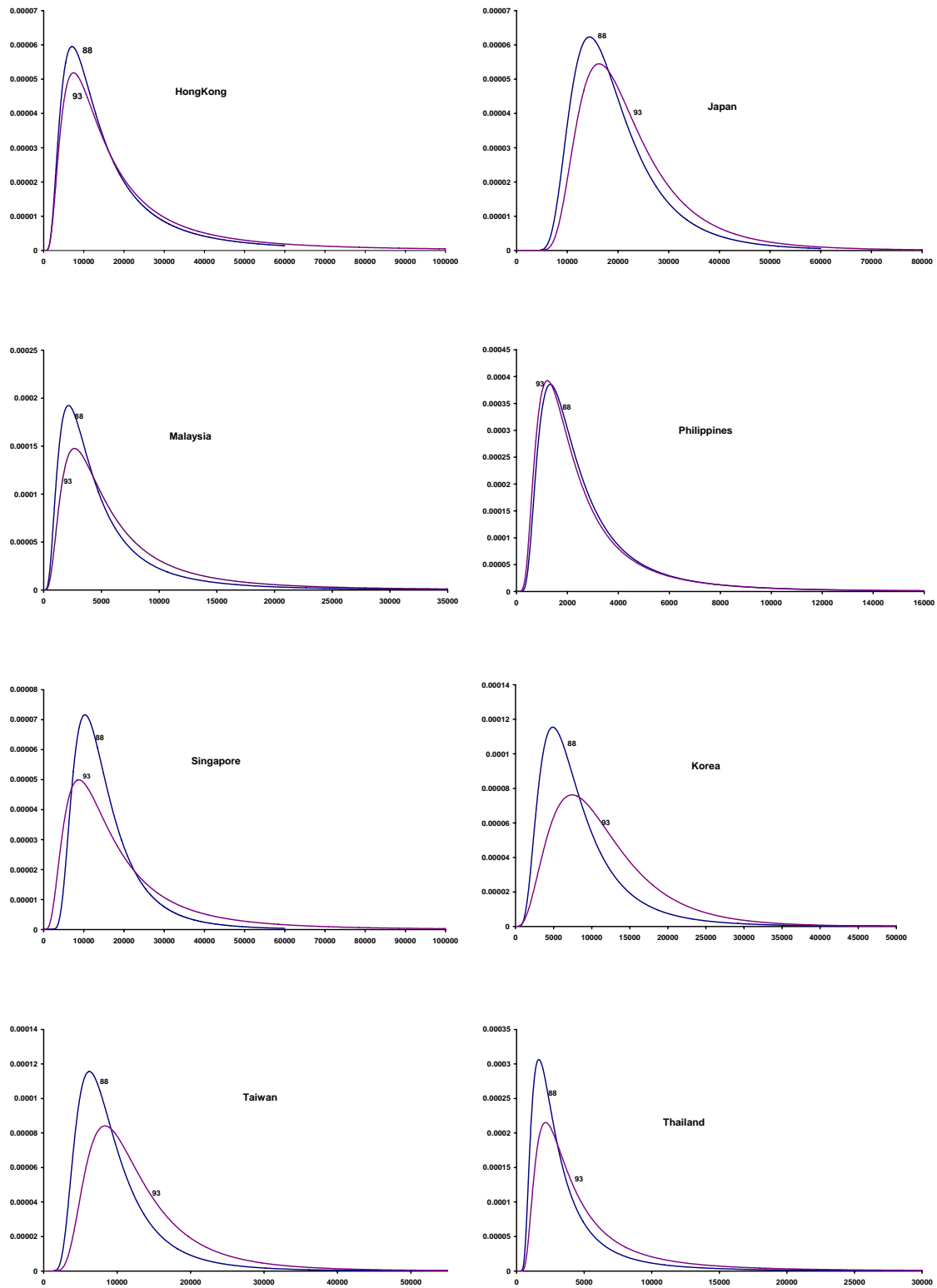
**Note:** All shares are decile shares with the exception of Japan for 1988 and 1993 and Philippines for 1988 where the population proportions were not equal for each class.

**Figure 1: Cross-country distributions**

**Figure 2a: Cumulative Distribution Functions, 1988****Figure 2b: Cumulative Distribution Functions, 1988**

**Figure 3a: Selected Lorenz Curves, 1988****Figure 3b: Selected Lorenz Curves, 1993**

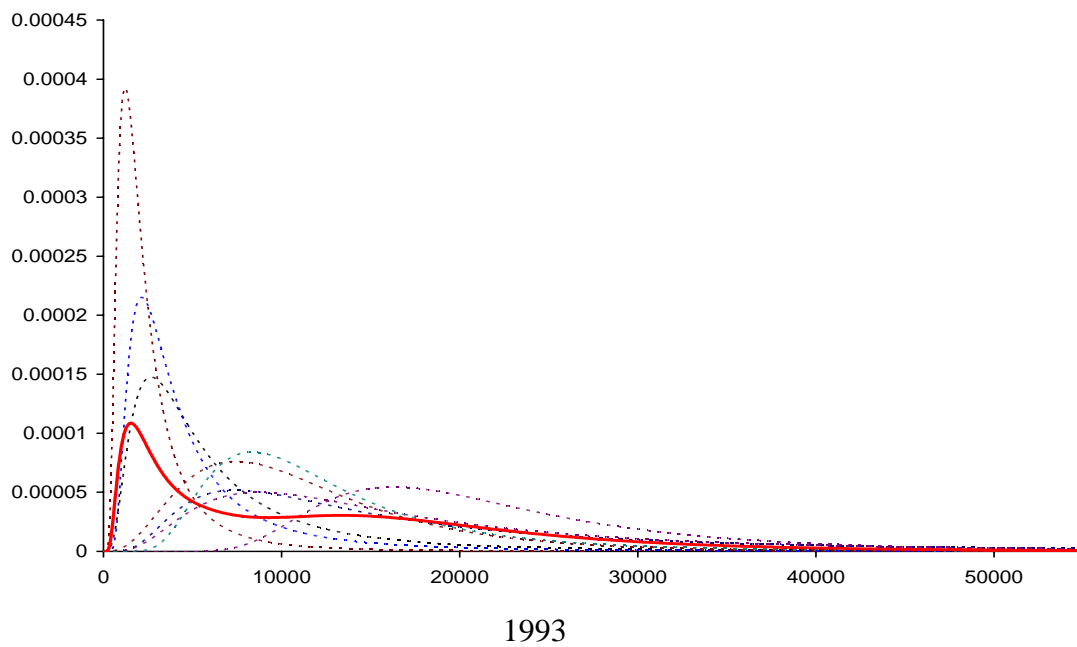
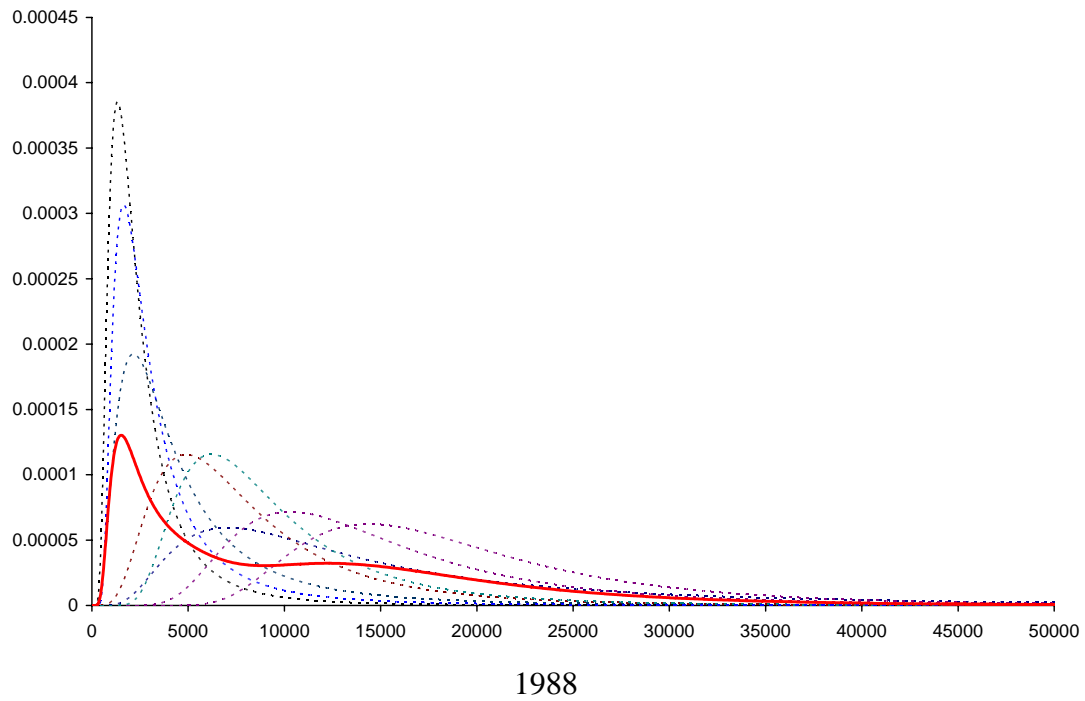


**Figure 4: Shifts in the distributions over time**

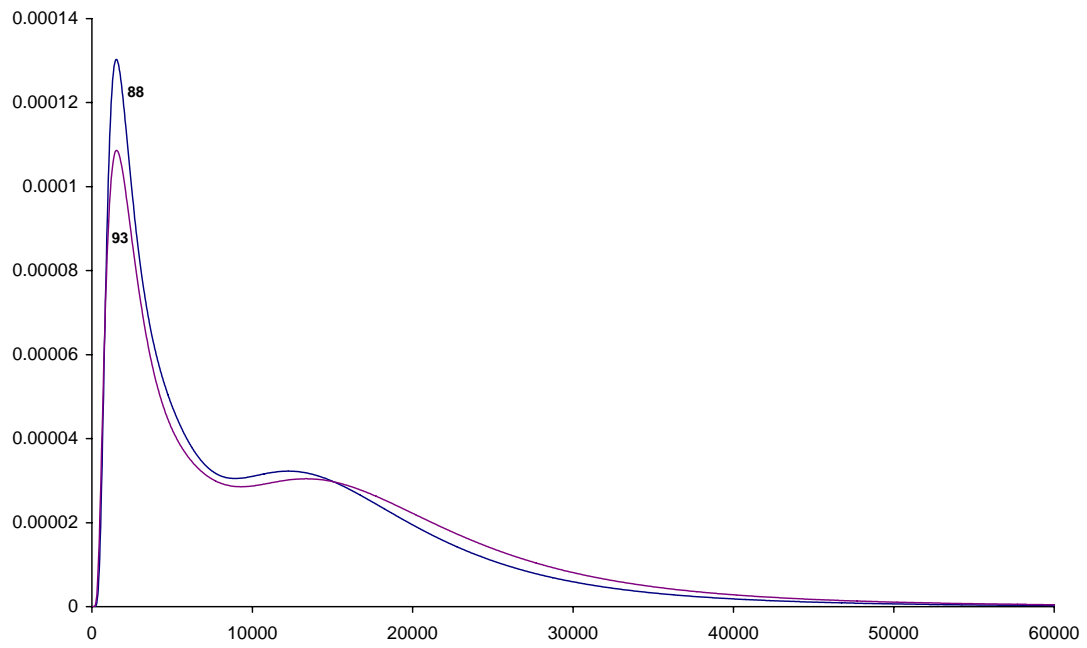
**Table 4: Observed and Estimated Gini Coefficients**

	1988		1993	
	Observed	Estimated	Observed	Estimated
HongKong	0.4598	0.4755	0.4974	0.5168
Japan	0.2409	0.2453	0.2428	0.2483
Malaysia	0.4474	0.4607	0.4629	0.4773
Philippines	0.4001	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**



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**