INTERNATIONAL INCOME COMPARISONS AND LOCATION CHOICE: METHODOLOGY, ANALYSIS, AND IMPLICATIONS

Vivek H. Dehejia

Department of Economics, Carleton University, 1125 Colonel By Drive, Ottawa , Ontario K1S 5B6, Canada; and CESifo, Munich, Germany.

E-mail: `vdehejia@ccs.carleton.ca`

Marcel Voia

Department of Economics, Carleton University, 1125 Colonel By Drive, Ottawa , Ontario K1S 5B6, Canada.

E-mail: `mvoia@connect.carleton.ca`

February 5, 2008

Abstract. This paper contributes to ongoing debates on international income comparisons by deploying a novel methodology for constructing empirical distribution functions for the United States and Canada over the period 1993 - 2000. We also conduct tests for first, second, third order stochastic dominance and of intersection of distributions, to determine which,if either, country might be a preferred destination for migration. Our findings are for that all of the years for which there is comparable data, the Canadian income distribution second order stochastically dominates the US income distribution. We provide an interpretation in terms of expected utility theory, considering the case of log utility, and relate our findings to an argument by Joseph Stiglitz, that in the face of skewness of income distributions a potential migrant should look at the median rather than the mean. It turns out that Stiglitz's intuition is correct, at least in the context of our study.

JEL classification codes: C12, D31, D63, D81.

Key words and phrases: Non-parametrics, Finite Mixtures, Heterogeneous Income Distribution, Stochastic Dominance, Kolmogorov-Smirnov type statistic, Bootstrap.

1. Introduction and motivation

Comparison of income levels across countries and over time is one of the cornerstones of the economics of growth and development, and plays a role too in evolving debates around the effects of globalization, whether benign or malign, on income distributions in both rich and poor countries, on the incentives to migrate, and much else besides. In much of this literature, applied as well as scientific, the benchmark of comparison is mean or average income, i.e., income per capita, suitably adjusted to make comparison over time (e.g., deflating by a price index) and across countries (e.g., adjusting for purchasing power parity) legitimate. But, as we teach even our first year students, average income is a very imperfect measure of the level of economic development. Amongst many other problems, it collapses the entire income distribution into a single (albeit extremely important) statistic, and thus discards valuable information on inequality around the mean. (This is quite apart from concerns about the validity of income as an index of development, on the grounds that it misses out other relevant social indicators, such as the Sen-Nussbam capabilities criteria.)

To make matters more concrete, consider as an example the following statement by Joseph Stiglitz, extracted from a review of a recent book on globalization, and published in the influential policy journal, Foreign Affairs. Stiglitz writes: "Consider the following thought experiment: If you could choose which country to live in but would be assigned an income randomly from within that countrys income distribution, would you choose the country with the highest GDP per capita? No. More relevant to that decision is median income.... As the income distribution becomes increasingly skewed, with an increasing share of the wealth and income in the hands of those at the top, the median falls further and further below the mean. That is why, even as per capita GDP has been increasing in the United States, U.S. median household income has actually been falling." (Stiglitz, 2005)

Stiglitz's point, evidently, is that high mean income in the United States masks the stagnation of the median, which, if true, would suggest that a few high income individuals are pulling up the mean while not having much, if any, impact on the median. The implication that he draws for the behavior of a putative migrant is equally interesting, as it suggests that a risk averse individual would tend to look at the median, rather than the mean, income level when comparing countries she might migrate to, as the median is, presumably, a better indicator of roughly where she is likely to end up. This pair of observations raises a host of interesting and difficult questions, some of which this paper will explore. While looking at the median rather than the mean may be a good rule of thumb under conditions of uncertainty and with risk averse preferences (an argument taken up in Dehejia, 2008), a more precise analysis will surely require a knowledge of the entire income distribution in the United States and other comparator countries, over time. We select in our analysis a comparison of the US with Canada, for a variety of reasons. Reliable data is available for both countries, and, given both their geographical contiguity and similarity across a range of non-economic criteria, along with the existence of a long history of migration flows in both directions, and a shared history, the bilateral comparison appears to us legitimate.

In what follows, we are concerned both with the methodology of constructing empirical distribution functions for the US and Canada for a number of different years, and with considering formal tests for which income distribution would be "better" for, say, a putative migrant. Here, we need to move beyond the comparison of mean vs. median, intuitive and appealing though it may be. In the literature, the relevant concept is that of stochastic dominance of different orders, that we deploy below. In particular, the differences in income distributions can be compared by employing statistical tests for the equality of two distributions, for testing first-order stochastic dominance (FSD), second-order stochastic dominance (SSD), and, intersection of two distributions.

We find an extensive literature on testing for stochastic dominance, which essentially starts with the work by McFadden (1989) where he proposes and analyzes a Kolmogorov-Smirnov type test for stochastic dominance. Subsequently, Anderson (1996), Davidson and Duclos (2000), Barrett and Donald (2003), and Linton, Maasoumi and Whang (2003) develop powerful statistical inferential results for stochastic dominance of any order. Horvath, Kokoszka and Zitikis (2004) contribute to the literature by showing how to modify the statistics in order to test for stochastic dominance over non-compact intervals.

Our findings suggest that our stochastic dominance results are in line with our simple application of expected utility theory, and also coincide with the Stiglitz rule of thumb of looking at the higher median. Over the period of investigation Canada " higher than the US, and this is picked up by all of these different criteria considered.

The paper is organized as follows. Section 2 formulates the problem. Section 3 describes the data used. Section 4 analysis the data. Section 5 concludes.

2. Methodology

Our arguments about the non-parametric estimation follow Zhu (2005), while the subsequent heterogeneity and stochastic dominance tests follow Davidson and Duclos (2000), Linton, Maasoumi and Whang (2003) and Voia and Zitikis (2008).

2.1. **Nonparametric Income Density Estimation.** We consider a nonparameteric analysis of the income distributions of the two countries to help us understand the shape and potential heterogeneity of the two countries' income distributions. We use an Adaptive Kernel method. Varying bandwidths is very important when long-tailed or multi-modal density functions are estimated using kernel methods. Therefore, we use an adaptive kernel density estimation to avoid the potential problems of using kernels with fixed bandwidth, such as undersmoothing in areas with only sparse observations while oversmoothing in others. Adaptive kernels were introduced and discussed by the following authors: Silverman (1986), Bowman and Azzalini (1997), Pagan and Ullah (1999), Salgado-Ugarte et al. (1993), Salgado-Ugarte et al. (1995) and Salgado-Ugarte and Perez-Hernandez (2003). The sparseness of the data in the upper tail of the income distribution recommends the use of kernels with varying bandwidths. Varying the bandwidth along the support of the sample data is reducing the variance of the estimates in areas with few observations, and is reducing the bias of the estimates in areas with many observations.

The estimation procedure follows a few steps: the first step computes an initial (fixed bandwidth) density estimate to get an idea of the density at each of the data points, and in the second step, this pilot estimate is used to adapt the size of the bandwidth over the data points when computing a new kernel density estimate.

Kernel density estimates are not unbiased; they are asymtotically biased, with a bias varying with the bandwidth and the shape of the true density function. For a given bandwidth, the bias does not tend to 0 as the sample size increases; therefore, we should be careful about the inference that we make using this approach.

To estimate the densities using an adaptive kernel we use Abramson (1982). Let $X_1, X_2, ...X_n$ be *iid* random variables with continuous distribution function $F(x) = \Pr(X_i \leq x)$.

The estimator constructs the local bandwidth h_i as a product of an estimated local bandwidth factor λ_i and a fixed bandwidth h at each sample point ($h_i = \lambda_i * h$). The local bandwidth factor stretches or shrinks the sample points' bandwidths to adapt to the density of the data, while the fixed bandwidth controls for the overall degree of smoothing.

$$\widehat{f_{h_i}}(x) = \frac{1}{\sum_{i=1}^{n}} \sum_{i=1}^{n} \frac{w_i}{h_i} K\left(\frac{x - X_i}{h_i}\right)$$

where X_i are data points associated to the weights w_i and $K(u)$ is the kernel (window) function.

The kernel function is a weight function that puts different weights on different points. Typically, it puts more weight on points near x and the weights decline as X_i gets farther away from x. "Near" and "far" from x is determined by the bandwidth parameter h_i. The local bandwidth factors are proportional to the square root of the underlying density functions at the sample points:

$$\lambda_i = \lambda(X_i) = (\frac{G}{\widetilde{f}(X_i)})^{0.5}$$

where G is the geometric mean over all i of the pilot density $\widetilde{f}(X)$. The pilot density estimate is the Kernel density estimate with fixed bandwidth h.

One can construct bands around the estimated density functions using the fact that the variance of the adaptive Kernel density estimator can be expressed as

$$V(\widehat{f_{h_i}}(x)) = (\sum_{i=1}^{n} \frac{w_i^2}{n^2})\frac{f(x)}{h_i}\int (K(s))^2 ds.$$

2.2. **Testing for internal heterogeneity.** Using the information from the nonparametric income density estimation we test for internal heterogeneity by fitting the data using a finite mixture model. To fit our data we use the following steps:

- Check if the income density plot shows that our data is a mixture of distributions and,
- assume that the true density is the weighted sum of log-normal densities with different expected values (E(y)) and variances (Var(y)) (the density plot shows that our data mimic a mixture of log-normal distributions), therefore,
- we estimate the parameters of such mixture by maximum likelihood.

The following likelihood function is used:

$$f(y,\theta) = \sum_{k=1}^{K} p_k \frac{1}{y\sigma_k\sqrt{2\pi}} e^{\frac{-(\ln y - \mu_k)^2}{2\sigma_k^2}},$$

where the parameters of interest are $\theta = \{K, p_k, \mu_k, \sigma_k\}$ with $k = 1, ..., K$ and $\sum_{k=1}^{K} p_k = 1$. Estimates of the parameter of interest are obtained by maximizing the log-likelihood:

$$\widehat{\theta} = argmax_{\theta} \sum_{i=1}^{n} \ln f(y_i, \theta).$$

2.3. **Decision Making Under Uncertainty Using the Expected Utility Criterion.** As a matter of econmic theory, the "correct" way to set up the problem that we have discussed in the introduction is to set up a single-person choice-theoretic-problem, in which an individual has to choose between any two (or more) given lotteries. The most commonly used criterion for comparing lotteries is the expected utility criterion : each decision maker has a certain utility function defined on money, assumed to be increasing in y. This utility function is called the Von Neumann-Morgenstern utility (VNM utility) of this decision maker.

For each lottery L an individual's expected utility from L is given by:

$$V(L) = \sum_{k=1}^{n} p_k u(y_k).$$

If an individual is risk-neutral, then maximizing expected utility boils down to maximizing the expected value of the lottery : this reflects the linearity of the sub-utility function in the case of risk-neutrality. However, in the more usual case of risk-aversion, the entire shape of the distribution of each lottery is, in principle, relevant, and there is no general rule of thumb on choosing between two uncertain lotteries. Stiglitz's intuition, mentioned in the introduction, is that in the face of skewed distributions such as income a risk-averse individual will want to look at the median rather than the mean. This may or may not coincide with the technically "correct" concepts of stochastic dominance, that come from the economics of uncertainty literature, that we shall look at subsequently.

We can illustrate the Stiglitz intuition, following Dehejia (2008), with the following simple exercise: suppose for simplicity that someone has a VNM expected utility function, with the underlying state utility function being log utility. This corresponds to an Arrow-Pratt coefficient of relative risk aversion of unity. The expected utility is then just the integral of log of income times the density of income at that point. We can then compute this individual's expected utility for each year for the US and Canadian income distribution, and find out which maximizes her income. We can also check whether this corresponds to the highest mean or median, as a test of Stiglitz's intuition.

Tables 3 and 4 present the summary statistics of the income data for US and Canada as well as the results for the expected utility. The results show that expected utility is higher for Canada than the US in each of the years, even though the US has a higher mean income in most of those years. Also, in most of the cases, the highest median accurately "picks" the highest expected utility. This is consistent with the Stiglitz intuition: for further details consult Dehejia (2008).

To move beyond the comparison of mean vs. median, we consider tests of stochastic dominance of different orders, to reinforce the results of the simple approach discusssed in this section. Indeed, the case of log utility is a special case of second-order stochastic dominance, that we will test for below, since, as a matter of economic theory, any risk-averse individual with VNM utility will prefer a distribution that second-order stochastically dominates another one, in a binary contest over the two distributions. We will return to a discussion of this later.

2.4. **Distributional Analysis.** This section follows Davidson and Duclos (2000), Linton, Maasoumi and Whang (2003) and Voia and Zitikis (2008). To compare the income distributions of two countries, we transformed the data to be in the same dollar amount. In this way we can compare the income distribution of Canada with the income distribution of US. Next, consider that we observe multiple time periods. Define the associated cumulative distribution functions for the two countries as $F^{(C,t)}$, where $F^{(C,t)}$ is the cumulative income distribution of country C at time t. Hence, we have pairs of income distributions at different times. The variable of interest is $Y^{(C,t)}$, where $Y^{(C,t)}$

is the income distribution of country C at time t. We shall be interested in various properties of the conditional distribution functions

$$F^{(C,t)}(y) := \mathbf{P}\left[Y^{(C,t)} \leq y \middle| \, C = country\text{“}j\text{”}\right].$$

Let

$$D_1^{(C,t)}(y) := F^{(C,t)}(y),$$

and define the higher orders of D_1 by

$$D_s^{(C,t)}(y) := \int_0^y D_{s-1}^{(C,t)}(x)dx.$$

We can also express D_s as:

$$D_s^{(C,t)}(y) := \frac{1}{(s-1)!}\int_0^y (y-x)^{s-1}dF(x).$$

We consider four possibilities for $F^{(C,t)}$, but, if necessary we can add more:

(1) The income distributions of the tested countries are equal. In this case we write the null hypothesis as

$$H_0^{(1)} : F^{(USA,t=i)} \equiv F^{(Canada,t=i)},$$

where i is the year under investigation.

(2) One of the distributions first order stochastically dominates another one. We shall consider the case when $F^{(USA,t=i)}(y) \leq F^{(Canada,t=i)}(y)$ for all y. We formulate the corresponding null hypothesis as

$$H_0^{(2)} : F^{(USA,t=i)} \leq F^{(Canada,t=i)}.$$

(Data might suggest testing the null hypothesis $F^{(USA,t=i)}(y) \geq F^{(Canada,t=i)}(y)$, which can be done analogously by interchanging the roles of $F^{(USA,t=i)}(y)$ and $F^{(Canada,t=i)}(y)$.)

(3) The two distributions intersect, and there are two points y_0 and y_1 such that $F^{(Canada,t=i)}(y_0) < F^{(USA,t=i)}(y_0)$ and $F^{(Canada,t=i)}(y_1) > F^{(USA,t=i)}(y_1)$. This case is important to emhasize the fact that there are sections of

the income distributions that are dominant for each country. In this case we write the null hypothesis as

$$H_0^{(3)} : F^{(USA,t=i)} \bowtie F^{(Canada,t=i)}.$$

(4) One distribution second order stochastically dominates (SSD) the other one. This is the most relevant case for our analysis. In this case we write the null hypothesis as

$$H_0^{(4)} : \int_0^y (y-x)dF^{((Canada,t=i))}(x) \geq \int_0^y (y-x)dF^{((USA,t=i))}(x).$$

A SSD test is sufficient to validate our results obtained via an VNM expected utility function, as noted previously. Next, the stochastic dominance tests used in the paper are presented.

2.4.1. *Testing $H_0^{(1)}$ vs $H_1^{(1)}$.* Considerations in this subsection are based on the classical Komogorov-Smirnov test. Namely, with the help of the parameter

$$\kappa := \sup_y \left| F^{(CANADA,t=i)}(y) - F^{(USA,t=i)}(y) \right|,$$

we rewrite the null and the alternative hypotheses under considerations as follows:

$$H_0^{(1)} : \ \kappa = 0 \quad \text{vs} \quad H_1^{(1)} : \quad \kappa > 0. \tag{2.1}$$

An estimator of κ can be defined by

$$\widehat{\kappa} := \sup_y \left| F^{(\widehat{CANADA},t=i)}(y) - F^{(\widehat{USA,t=}i)}(y) \right|.$$

The estimator $\widehat{\kappa}$ is consistent. Based on its asymptotic distribution we obtain that

$$\widehat{K} := \sqrt{\frac{nm}{n+m}}\, \widehat{\kappa}$$

is an appropriate statistic for testing the null hypothesis $H_0^{(1)}$ against the alternative $H_1^{(1)}$. Here n and m and sample sizes for the two distributions. The corresponding rejection (i.e., critical) region is $R : \ \widehat{K} > k_\alpha$ and the acceptance region is $A : \ \widehat{K} \leq k_\alpha$, where k_α is the α-critical value of the (classical) Kolmogorov-Smirnov test.

Testing $H_0^{(2)}$ vs $H_1^{(2)}$. Considerations in this subsection follow those in Linton, Maasoumi and Whang (2003). Namely, with the help of the parameter

$$\delta := \sup_y \left(F^{(CANADA,t=i)}(y) - F^{(USA,t=i)}(y)\right),$$

we rewrite the hypotheses $H_0^{(2)}$ and $H_1^{(2)}$ as follows:

$$H_0^{(2)}: \ \delta = 0 \quad \text{vs} \quad H_1^{(2)}: \quad \delta > 0. \tag{2.2}$$

The empirical estimator of δ is given by

$$\widehat{\delta} := \sup_y \left(F^{(\widehat{CANADA},t=i)}(y) - F^{(\widehat{USA,t=}i)}(y)\right).$$

The estimator $\widehat{\delta}$ is consistent. Therefore,

$$\widehat{D} := \sqrt{\frac{nm}{n+m}}\,\widehat{\delta}$$

is an appropriate statistic for testing the null hypothesis $H_0^{(1)}$ against the alternative $H_1^{(1)}$. The corresponding rejection (i.e., critical) region is $R: \ \widehat{D} > d_\alpha$ and the acceptance region is $A: \ \widehat{D} \le d_\alpha$, where d_α is the α-critical value of the maximum of a Gaussian stochastic process Γ which depends on both distributions $F^{(USA,t=i)}$ and $F^{(CANADA,t=i)}$. Since the distributions are not, in general, identical, the critical value d_α is not distribution free and has to be therefore estimated. For this we can use bootstrap as follows: from $Y_1^{(USA,t=i)}, \dots, Y_n^{(USA,t=i)}$ we sample with replacement and obtain n values $Y_1^{(USA,t=i)*}, \dots, Y_n^{(USA,t=i)*}$. Let $F^{(\widehat{USA,t=}i)}{}^*(y)$ be the corresponding empirical distribution function. Next, from $Y_1^{(CANADA,t=i)}, \dots, Y_m^{(CANADA,t=i)}$ we sample with replacement and obtain m values $Y_1^{(CANADA,t=i)*}, \dots, Y_m^{(CANADA,t=i)*}$. Let $F^{(\widehat{CANADA},t=i)}{}^*(y)$ be the corresponding empirical distribution function. With the notation above, we define the process

$$\begin{aligned}\Delta^*(y) &:= \sqrt{\frac{nm}{n+m}}\ \left(F^{(\widehat{USA,t=}i)}{}^*(y) - F^{(\widehat{USA,t=}i)}(y)\right) - \\ &-\sqrt{\frac{nm}{n+m}}\ \left(F^{(\widehat{CANADA},t=i)}{}^*(y) - F^{(\widehat{CANADA},t=i)}(y)\right),\end{aligned}$$

and then, in turn,

$$\widehat{D}^* := \sup_y \Delta^*(y)$$

We repeat the above sampling procedure B times and in this way obtain B values of $\widehat{D}^*$. Now we are in the position to define the estimator d^*_α as the smallest value of y such that at least $100(1-\alpha)\%$ of the obtained B values of $\widehat{D}^*$ are at or below y. With the just defined d^*_α, the rejection and the acceptance regions for testing the null hypothesis $H_0^{(2)}$ against the alternative $H_1^{(2)}$. are, respectively, $R: \ \widehat{D} > d^*_\alpha$ and $A: \ \widehat{D} \le d^*_\alpha$.

Testing $H_0^{(3)}$ vs $H_1^{(3)}$. Considerations in this subsection follow those in Voia and Zitikis (2008). First we note that if there is an y_0 such that the strict inequality $F^{(USA,t=i)}(y_0) > F^{(CANADA,t=i)}(y_0)$ holds, then the earlier introduced parameter δ is strictly positive. Likewise, the existence of y_1 such that $F^{(USA,t=i)}(y_1) < F^{(CANADA,t=i)}(y_1)$ results in a strictly positive value of the parameter

$$\theta := \sup_y (F^{(USA,t=i)}(y) - F^{(CANADA,t=i)}(y)).$$

Hence, if the two distributions $F^{(USA,t=i)}$ and $F^{(CANADA,t=i)}$ intersect, then the parameter

$$\tau := \min(\delta, \theta)$$

is strictly positive. In view of the discussion above, we reformulate the null hypothesis as $H_0^{(3)} \ : \ \tau > 0$. Under the alternative, the two distribution functions dominate each other. Hence, the parameter τ will never be positive. In fact, we have $\tau = 0$ since $F^{(USA,t=i)}(y)$ and $F^{(CANADA,t=i)}(y)$ always coincide at $y = \pm\infty$. Hence, we reformulate the alternative as $H_1^{(3)} : \ \tau = 0$.

The way the null and alternative hypotheses appear above poses a serious problem in developing a statistical test of desired size or level. To circumvent the problem, we shall formulate our problem somewhat differently. That is, we shall test the null hypothesis

$$H_0^{(\text{not } 3)} : F^{(USA,t=i)} \text{ dom } F^{(CANADA,t=i)},$$

where " $F^{(USA,t=i)}$ dom $F^{(CANADA,t=i)}$ " means that one of the distributions dominates another one, without specifying whether $F^{(USA,t=i)} \leq F^{(CANADA,t=i)}$ or $F^{(USA,t=i)} \geq F^{(CANADA,t=i)}$. The alternative $H_1^{(\mathrm{not}\,3)}$, which is the complement of $H_0^{(\mathrm{not}\,3)}$ by definition, coincides with the earlier specified $H_0^{(3)}$: $F^{(USA,t=i)} \bowtie F^{(CANADA,t=i)}$. Hence, if we reject the null hypothesis $H_0^{(\mathrm{not}\,3)}$: $\tau = 0$, then we shall have significant evidence to claim that the two distributions $F^{(USA,t=i)}$ and $F^{(CANADA,t=i)}$ intersect. In summary, we shall test

$$H_0^{(\mathrm{not}\,3)}: \ \tau = 0 \quad \mathrm{vs} \quad H_1^{(\mathrm{not}\,3)}: \quad \tau > 0. \tag{2.3}$$

We define an estimator of τ by

$$\widehat{\tau} := \min(\widehat{\delta}, \widehat{\theta}),$$

where $\widehat{\delta}$ is same as above, and

$$\widehat{\theta} := \sup_x (F^{(\widehat{CANADA},t=i)}(y) - F^{(\widehat{CANADA},t=i)}(y)).$$

We have that

$$\widehat{T} := \sqrt{\frac{nm}{n+m}}\, \widehat{\tau}$$

is an appropriate statistic for testing the null hypothesis $H_0^{(\mathrm{not}\,3)}$ against the alternative $H_1^{(\mathrm{not}\,3)}$ (recall that it coincides with $H_0^{(3)}$). The corresponding rejection region is $R: \ \widehat{T} > t_\alpha$, where t_α is the α-critical value of a distribution that depends on $F^{(USA,t=i)}$ and $F^{(CANADA,t=i)}$. Hence, we need to estimate t_α, for which we use a bootstrap as follows.

With the same process $\Delta^*(x)$ as defined earlier, let

$$\widehat{T}^* := \max(\sup_x \Delta^*(x), \ \sup_x(-\Delta^*(x)))$$

(the maximum is not a typographical error). We repeat the above sampling procedure M times and in this way obtain M values of $\widehat{T}^*$. Now we define the estimator t^*_α as the smallest x such that at least $100(1-\alpha)\%$ of the obtained M values of $\widehat{T}^*$ are at or below x. With the t^*_α, the rejection region for testing $H_0^{(3)}$ against $H_1^{(3)}$. is $R: \ \widehat{T} > t^*_\alpha$.

Testing $H_0^{(4)}$ *vs* $H_1^{(4)}$. Considerations in this subsection follow those in McFadden (1989) and Davidson and Duclos (2000). Hence, if $F^{(CANADA,t=i)}$ SSD $F^{(USA,t=i)}$, then the parameter

$$\gamma := \sup_y (D_2^{(CANADA,t)}(y) - D_2^{(USA,t)}(y))$$

is strictly positive. Therefore, we shall test the null hypothesis using

$$H_0^{(4)}: \ \gamma = 0 \quad \text{vs} \quad H_1^{(3)}: \quad \gamma > 0. \tag{2.4}$$

that one of the distributions SOSD another one.

Define an estimator of γ by

$$\widehat{\gamma} := \sup(D_2^{\widehat{(CANADA,t)}}(y) - D_2^{\widehat{(USA,t)}}(y)),$$

The estimator $\widehat{\gamma}$ is consistent and we have that

$$\widehat{G} := \sqrt{\frac{nm}{n+m}}\,\widehat{\gamma}$$

The corresponding rejection (i.e., critical) region is $R: \ \widehat{G} > g_\alpha$ and the acceptance region is $A: \ \widehat{G} \le g_\alpha$, where $g\alpha$ is the α-critical value of a distribution that depends on $F^{(USA,t=i)}$ and $F^{(CANADA,t=i)}$. Hence, g_α is not distribution free and has to be estimated. For this, we use a bootstrap approximation.

3. Data

3.1. **Panel study of income dynamics (SLID).** We used the available public used data. The description of the data follows Giles and Philip (1999). The sample is drawn from the Canadian Labour Force Survey (LFS). The LFS covers the population of the 10 provinces, with the exception of Indian reserves, the military and residents of institutions. The coverage of the SLID sample is identical with LFS, with one small difference: SLID includes Armed Forces personnel living out of barracks. The size of each six-year panel is 15,000 households. This includes about 40,000 persons, of which 31,000 are aged 16 years and over. SLID is intended to continue indefinitely. Starting with Panel 2, two panels will always be overlapping. The approach of rotating overlapping panels ensures that the sample remains representative. Panel 1 was selected in January 1993. The second panel started with reference year 1996. The third panel began with reference year 1999, when the first panel was "retired". The choice of a six-year panel duration depended on the initial design and considerations about the burden costs for the respondents. When only a few consecutive years of longitudinal data are required, the sample size can be doubled by combining data from the last three years of one panel and the first three years of the next, overlapping panel. All longitudinal respondents are identified at the beginning of the panel, and they are followed for six years, whether they move or not. If a household splits up, all "branches" are followed. Likewise, new people who start to live with a longitudinal respondent during the six years are also included in the survey, although they do not contribute to the longitudinal sample. They are called "cohabitants" in SLID. The reasons for interviewing cohabitants are: a) to maintain complete family and household data on longitudinal respondents; b) to obtain data on a representative cross-section of the population each year. Cohabitants follow a very similar interview format to longitudinal respondents. However, they do not begin the process until they enter the household of a longitudinal respondent, and they cease to be interviewed or followed as soon as they cease to live with a longitudinal respondent.

3.2. **Survey of Income and Program Participation (SIPP).** The description of the data follows the description from the official SIPP website. SIPP as SLID is a multi-panel, nationally representative dataset created by the U.S. Census. The first SIPP panel was begun in the mid-1980s and the latest one was begun in 2004. The SIPP tracks individuals for two to four years, depending on the panel. SIPP respondents are asked questions every fourth month about their experiences over the prior four months. The information supplied by this survey provides a better understanding of the level, and changes in the level of well-being of the population. SIPP can provide information on how economic situations are related to the demographic and social characteristics of individuals. The data collected in SIPP will be especially useful in studying Federal transfer programs, estimating program cost and effectiveness, and assessing the effect of proposed changes in program regulations and benefit levels. The first interviews in the SIPP took place in October 1983, nearly 8 years after the research and developmental phase, the Income Survey Development Program (ISDP), was initiated by the Department of Health, Education, and Welfare, in 1975. Between 1975 and 1980 extensive research was undertaken to design and test new procedures for collecting income and related socioeconomic data on a subannual basis in a longitudinal framework. The design of the questionnaire for the first interview was similar in structure to that used in the 1979 ISDP panel study with two important exceptions. First, the reference period for the questions was extended from 3 months to 4 months in order to reduce the number of interviews and, therefore, lower costs. Second, the questions covering labor force activity were expanded in order to provide estimates that were closer, on a conceptual basis, to those derived from the Current Population Survey (CPS). The design also incorporated a number of other modifications resulting from experience with the 1979 pilot study.

There are three basic elements contained in the overall design of the survey content. The first is a control card which is used to record basic social and demographic characteristics for each person in the household at the time of the initial interview. Households are interviewed a total of eight or nine times, therefore the card is also used to record changes in characteristics such as age, educational attainment, and marital status, and to record the dates when persons enter or leave the household. Finally, during each interview, information on each source of income received and the name of each job or business is transcribed to the card so that this information can be used in the updating process in subsequent interviews. The second major element of the survey content is the core portion of the questionnaire. The core questions are repeated at each interview and cover labor force activity, the types and amounts of income received during the four- month reference period, and participation status in various programs. Income amounts are recorded on a monthly basis with the exception of amounts of property income (interest, dividends, rent, etc.). Data for these types of income are recorded as totals for the four- month period. The core also contains questions covering attendance in postsecondary schools, private health insurance coverage, public or subsidized rental housing, low-income energy assistance, and school breakfast and lunch participation. The third major element is the various supplements or topical modules that will be included during selected household visits. The topical modules cover areas that need not be examined every four months. Certain of these topical modules are considered to be so important that they are viewed as an integral part of the overall survey. Other topical modules have more specific and more limited purposes. A list of topical modules includes work history, health characteristics (including disability), assets and liabilities, pension plan coverage, housing characteristics, child care, child support agreements, support for non-household members, program participation history, reasons for not working, calender year income and benefits, taxes, and education and training.

The sample design for the first SIPP panel in 1984 consisted of about 20,000 households selected to represent the noninstitutional population of the United States. The most recent 1993 panel has also a sample size of approximately 20,000 households. Households in this SIPP panel are scheduled to be interviewed at four-month intervals over a period of 3 years. The reference period for the questions is the four-month period preceding the interview. For example, households interviewed in February 1993 were asked questions for the months October, November, December 1992, and January 1993. This household was interviewed again in June 1993 for the February through May period. The sample households within a given panel are divided into four samples of nearly equal size. These subsamples are called rotation groups and one rotation is interviewed each month. In general, one cycle of four interviews covering the entire sample using the same questionnaire is called a wave. SIPP panels have been introduced in February of each year succeeding the 1984 panel. This overlapping design provides a larger sample size from which cross- sectional estimates can be made. The overlap also insures smaller standard errors on differences between estimates for two points in time.

4. Results

To test our hypotheses we do a comprehensive distribution analysis using non-parametric representations of the income distributions, fitting the distributions using finite mixtures, which are necessary to identify the degree of heterogeneity of the distributions, using the expected utility criterion to understand decision making under uncertainty, and employing stochastic dominance tests to reinforce our empirical results.

Figures 6.1 and 6.2 use a non-parametric representation of the income densities of the two countries to show the degree of heterogeneity in these income distributions. The graphs show that both the US and Canada have distributions that are rightward-skewed, with a lot of mass at lower incomes toward medium incomes. Also, the graphs show that these distributions are multimodal, which suggest that both distributions are heterogeneous.

To test the heterogeneity assumption, we fit our distributions using finite mixtures. Our findings show that the US data is fitted by a mixture of four log-normal distributions (see Table 1) for all years excepting 1999 (which is fitted by a mixture of three log-normal distributions).

For Canadian data we find that the data is fitted by a mixture of three log-normal distributions with parameters as in Table 2.

The results show the degree of heterogeneity of the income data for the US and Canada. In both countries the income distributions are heterogeneous;

however, US data shows more heterogeneity, as there are more mixtures identified and a larger dispersion of the mean of these mixtures.

These results clearly suggest that simple comparisons of mean or median between the US and Canada at different points in time, while suggestive, cannot capture the rich heterogeneity in the income distributions of the two countries. It could be potentially misleading given that the distributions are multimodal and that each can be characterized as a mixture of three (or four) log normal distributions. Perforce, an expected utility maximizer will need to consider the entire shape of each distribution when deciding which "lottery" (in this case, income distribution) to pick. Tests of stochastic dominance do precisely this, that we discuss next.

As noted, we employ tests of stochastic dominance to reinforce the conclusion of our choice-theoretic results discussed earlier. The results are presented in Table 5 and are represented in Figure 6.3. The results show that for all years the distribution of income in Canada second order stochastically dominates (SSD) the US distribution. These results do, indeed, correspond to the results obtained using the expected utility criterion method, as theory would predict. This is because, as a matter of theory, anyone with risk-averse preferences (that would include log utility) would prefer a distribution that SSD another one. So the results obtained in our heuristic test with log utility do indeed match the formal results based on tests of SSD, as we would expect. We presented both sets of results since, although the SSD tests are more rigorous and are theoretically the "right" ones to use, one can often get intuitively appealing results by testing for a specific state utility function, such as log utility, and computing an actual numerical value for expected utility.

5. Conclusion

In this paper we have considered a perennial question of the economics of growth and development, which is the international comparison of incomes.

Our motivation was a passage from Joseph Stiglitz, that considered the choice problem confronting a putative migrant, and suggesting, implicitly, that skewness in income distributions would be an important factor, and one would need to look beyond a comparison of means as is generally done. We have taken this admonition seriously, and have developed a new methodology in this paper to construct empirical income distribution (and corresponding density) functions for two comparator countries, the United States and Canada.

To summarize our main results, we find that that our stochastic dominance results are in line with our simple application of expected utility theory, and also coincide with the Stiglitz rule of thumb of looking at the higher median: in all cases, Canada "scores" higher than the US, and this is picked up by all of these different criteria. This methodology could easily be extended to a comparison of others pairs of countries, or to a comparision of income levels within a large federal state, such as comparing state-level incomes in the US, provincial-level incomes in Canada, etc. The econometric implications of our novel methodology deserve further investigation and exploration, which we plan to take up in a future paper. The implications in terms of the choice-theoretical problem and the Stiglitz intuition are taken up further in Dehejia (2008).

Bibliography

Abramson, I. S. (1982), On bandwidth variation in kernel estimates: a square root law, Annals of Statistics 10(4), 1217-1223.

Anderson, G. (1996), Nonparametric tests of stochastic dominance in income distributions, Econometrica 64, 1183-1193.

Barrett, G.F. and Donald, S.G. (2003), Consistent tests for stochastic dominance, Econometrica 71, 71-104.

Bowman, A. W. and Azzalini, A. (1997), Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations, Vol. 18 of Oxford Statistical Science Series, Oxford University Press, Oxford, UK.

Burkhauser, R. V., Crews, A. D., Daly, M. C. and Jenkins, S. P. (1999), Testing the significance of income distribution changes over the 1980s business cycle: A cross-national comparison, Journal of Applied Econometrics 14(3), 253272.

Csörgő, S. and Mason, D.M. (1989), Bootstrapping empirical functions, Annals of Statistics 17, 1447-1471.

Davidson, R. and Duclos, J.-Y. (2000), Statistical inference for stochastic dominance and for the measurement of poverty and inequality, Econometrica 68, 1435-1464.

Dehejia, Vivek H. (2008), Risk Aversion, Stochastic Dominance, and Rules of Thumb: Concept and Application, Carleton Economics Papers 2008-1. Available at: http://www.carleton.ca/economics .

Fraker, Thomas and Rebecca Maynard (1987), The Adequacy of Comparison Group Designs for Evaluations of Employment Related Programs, Journal of Human Resources 22, 194-227.

Giles and Philip (1999), Overview of the Survey of Labour and Income Dynamics (SLID), Statistics Canada paper.

Kaur, A., Prakasa Rao, B. L. S. and Singh, H. (1994), Testing for second-order stochastic dominance of two distributions, Econometric Theory 10, 849-866.

Horvatha, L., Kokoszkab, P., Zitikis R. (2006), Testing for stochastic dominance using the weighted McFadden-type statistic, Journal of Econometrics 133 (2006) 191205.

Linton, O., Maasoumi, E. and Whang, Y. (2003), Consistent Testing for Stochastic Dominance under General Sampling Schemes. Manuscript.

McFadden, D. (1989), Testing for stochastic dominance. In: Studies in the Economics of Uncertainty (eds. T.B. Fomby and T.K. Seo), Springer-Verlag, New York.

Pagan, A. and Ullah, A. (1999), Nonparametric Econometrics, Themes in Modern Econometrics, Cambridge University Press, New York, USA.

Salgado-Ugarte, I. H., Perez-Hernandez, M. A. (2003), Exploring the use of variable bandwidth kernel density estimators, Stata Journal, StataCorp LP, vol. 3(2), pages 133-147.

Salgado-Ugarte, I. H., Shimuzu, M. and Taniuchi,T. (1995), snp6.2: practical rules for bandwidth selection in univariate density estimation, Stata Technical Bulletin 27, 519.

Schmid, F. and Trede, M. (1996a), Testing for first order stochastic dominance in either direction, Comput. Statist. 11, 165-173.

Schmid, F. and Trede, M. (1998), A Kolmogorov-type test for second-order stochastic dominance, Statist. Probab. Lett. 37, 183-193.

Shaked, M. and Shanthikumar, J.G. (1994), Stochastic Orders and their Applications, Academic Press, Boston, MA, 1994.

Silverman, B. W. (1986), Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability, Chapman and Hall, London, UK.

Van Der Vaart, A. (1998), Asymptotic Statistics, Cambridge University Press, Cambridge, UK.

Voia, M.C. and R. Zitikis (2008), Analyzing treatment effects on subpopulations in a socioeconomic experiment, manuscript, Carleton University.

Zhu, F. (2005), A nonparametric analysis of the shape dynamics of the US personal income distribution: 1962-2000, BIS Working Papers, No. 184.

6. Tables and Figures

Table 1. Tabulated mixtures for the US data.

Year	Type	Proportion	E(y)	Std(y)
1993	I	0.0332582	101.8712	97.608
1993	II	0.3313026	6418.953	4019.39
1993	III	0.4756854	23705.8	11359.87
1993	IV	0.1597538	58731.36	29105.55
1994	I	0.0460916	224.8212	210.323
1994	II	0.3506448	7146.149	4260.474
1994	III	0.467055	25060.59	11915.1
1994	IV	0.1362086	61843.41	30430.85
1996	I	0.0250294	52.197	48.469
1996	II	0.4589643	9035.792	5880.17
1996	III	0.4713546	33376.25	17367.35
1996	IV	0.0446517	114978.3	83492.1
1997	I	0.023783	41.681	37.724
1997	II	0.4573882	9234.246	6022.806
1997	III	0.4770831	34206.55	17928.17
1997	IV	0.0417457	121278.1	88987.21
1998	I	0.0185062	24.795	18.786
1998	II	0.4655374	9872.275	6503.125
1998	III	0.473136	35676.66	18541.3
1998	IV	0.0428204	122269.0	87100.75
1999	I	0.4262812	9109.642	6302.796
1999	II	0.5239947	35156.2	18646.86
1999	III	0.0497241	122205.7	86510.69

Table 6.1

Table 1. Tabulated mixtures for the Canadian data.

Year	Type	Proportion	E(y)	Std(y)
1993	I	0.490245	9233.37	6420.73
1993	II	0.4800183	31013.32	15095.73
1993	III	0.0297368	75033.6	43233.08
1994	I	0.4882041	8920.146	6397.98
1994	II	0.4757695	30795.44	15057.59
1994	III	0.0360265	72708.55	41548.27
1996	I	0.4951843	9166.381	6497.109
1996	II	0.4812286	31654.1	16141.36
1996	III	0.0235871	94898.56	72505.2
1997	I	0.489361	9141.655	6435.47
1997	II	0.4853356	31554.95	16084.55
1997	III	0.0253034	96069.82	73172.89
1998	I	0.4839589	9476.342	6564.319
1998	II	0.4889037	31993.99	16228.13
1998	III	0.0271374	97308.83	78686.86
1999	I	0.4668998	9392.132	6397.98
1999	II	0.5064845	32222.89	15057.59
1999	III	0.0266156	98598.42	41548.27

TABLE 6.2

Table 3. Tabulated statistics and Expected Utility for SIPP (USA) data.

year	obs	Mean	Median	Mode	Expected Utility
1993	62721	21576	16690	1498.38	0.00101
1994	62721	21744	15684	8689.29	0.00095
1995	62721	21599	15731	6858.57	0.00124
1996	65439	22332	15228	5372.53	0.00063
1997	65438	22038	15223	7940.92	0.00049
1998	65435	22023	15344	7978.89	0.00036
1999	65435	22127	15644	7375.43	0.00039
2000	x	x	x	x	x
2001	65445	20135	14094	11297.37	0.00129

TABLE 6.3

Note: The Income data is conditional on a maximum income of 250,000 Canadian dollars. 1993 is used as a base year.
x =missing data for the given year.

Table 4. Tabulated statistics and Expected Utility for SLID (Canadian) data.

year	obs	Mean	Median	Mode	Expected Utility
1993	29536	21583.84	16690	3117.14	0.0013
1994	29362	21577.82	16267.4	9912.28	0.0011
1995	x	x	x	x	x
1996	61064	21749.14	16303	8760.95	0.0016
1997	61455	21957.46	16485	13464.00	0.00069
1998	62140	22491.71	17125.5	9987.22	0.00053
1999	58051	23049.02	17643.72	11497.90	0.00039
2000	57380	23376.8	17770.38	14405.81	0.00043

Table 6.4

Note: x =missing data for the given year.

Table 5. Tabulated results for the Stochastic dominance tests between US and Canadian Income distributions.

Year	US	Canada	p-value
1993	-2	2	0.000
1994	-2	2	0.000
1996	-2	2	0.000
1997	-2	2	0.000
1998	-2	2	0.000
1999	-2	2	0.000

TABLE 6.5

Note: -2 stands for second order stochastically dominated distribution and 2 stands for second order stochastic dominance.

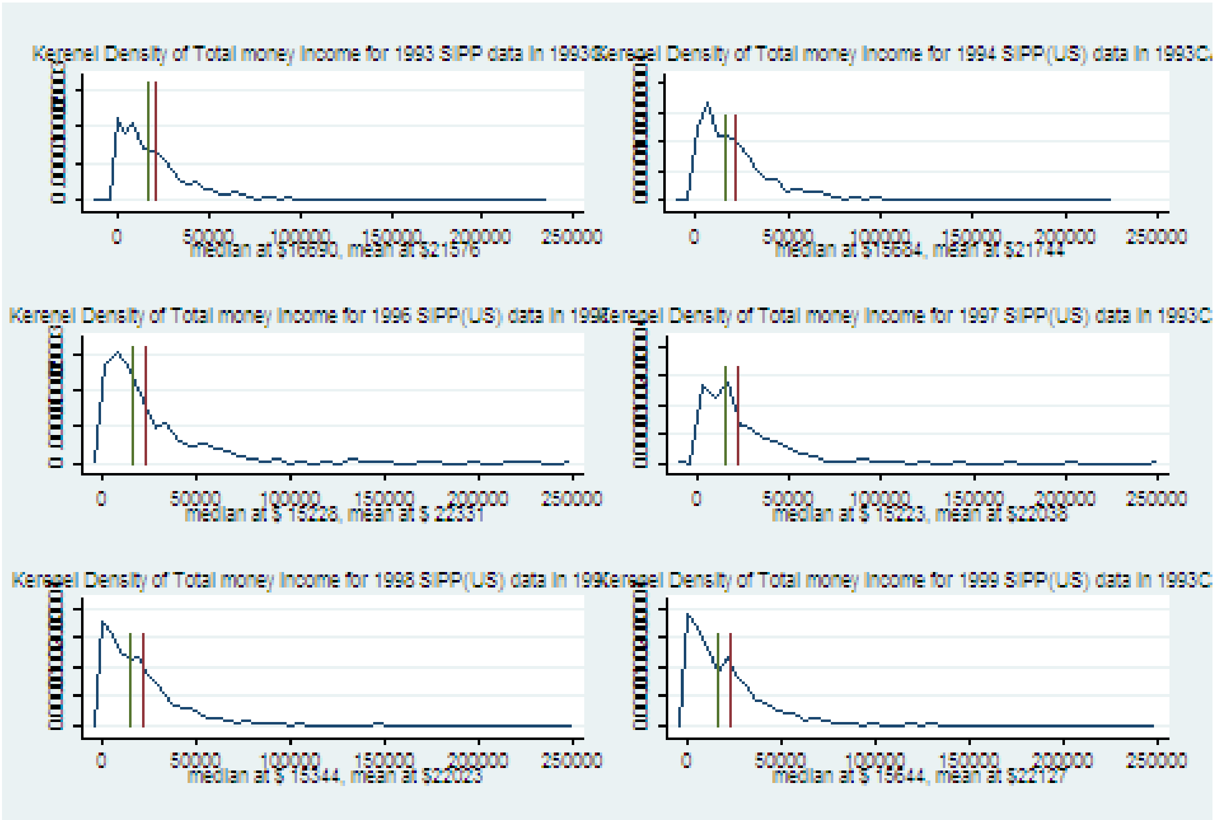

FIGURE 6.1. Kernel Densities of Total Money Income, US data in 1993 Canadian Dollars

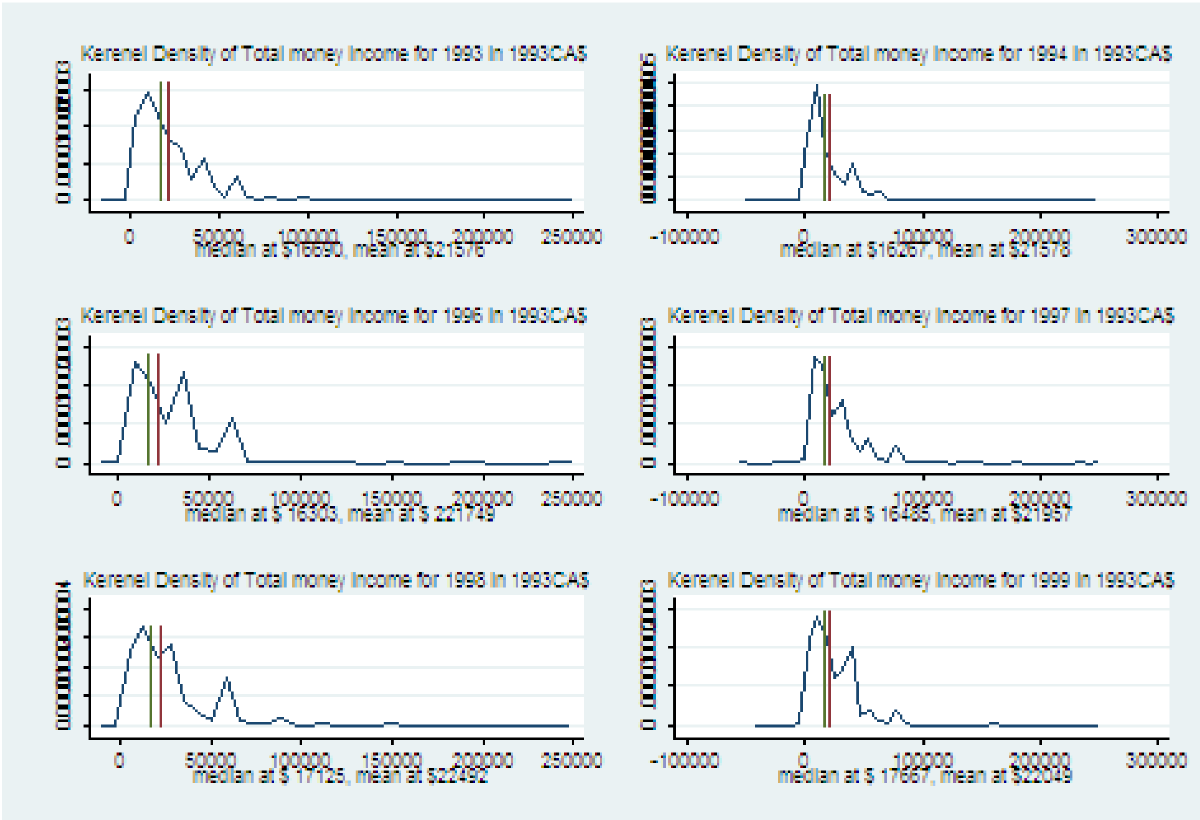

Figure 6.2. Kernel Densities of Total Money Income, Canadian data in 1993 Canadian Dollars

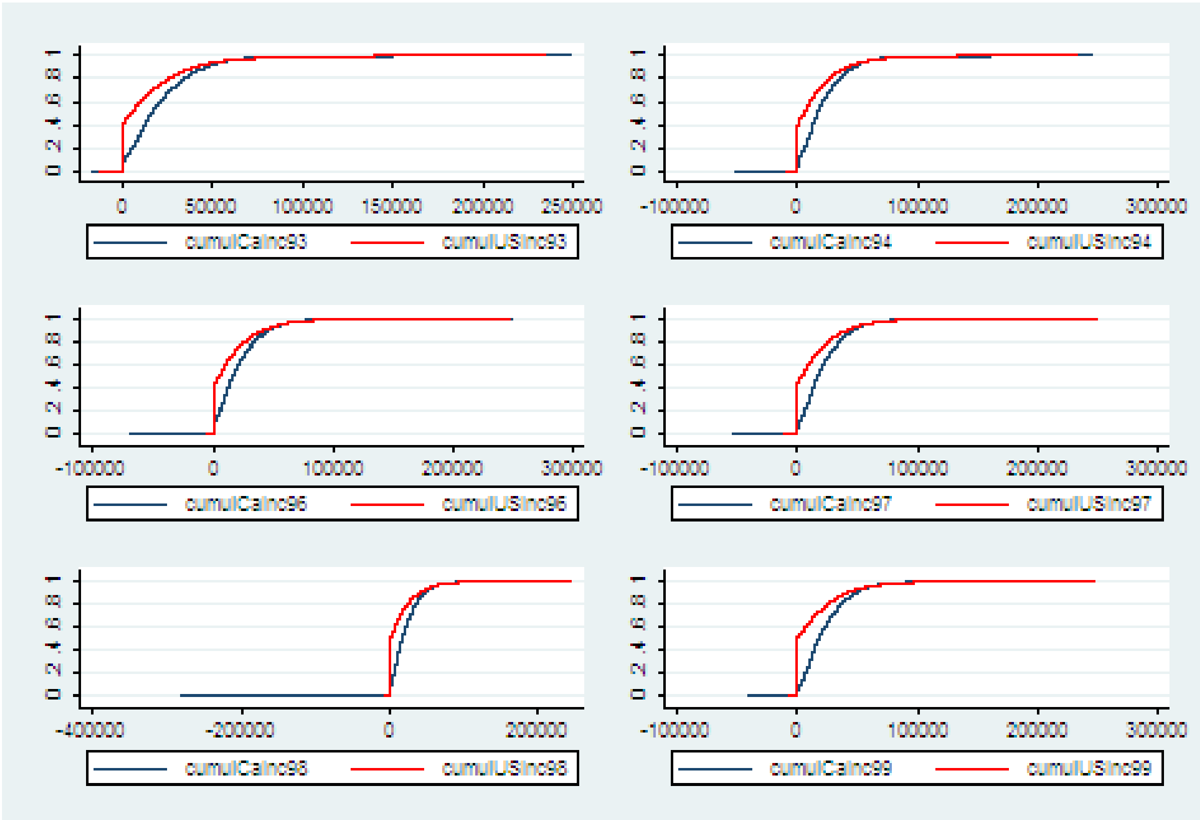

FIGURE 6.3. CDF's of SIPP (US) and SLID (Ca) distributions