UNIVERSITY OF CALGARY

A NONPARAMETRIC ANALYSIS OF THE PERSONAL INCOME DISTRIBUTION ACROSS THE PROVINCES AND STATES IN THE U.S. AND CANADA

By

Yibing Wang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FUFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

DEPARTMENT OF ECONOMICS

CALGARY, ALBERTA

DECEMBER, 2003

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled " A Nonparametric Analysis of the Personal Income Distribution across the Provinces and States in the U.S. and Canada" submitted by Yibing Wang in partial fulfillment of the requirements of the degree for the degree of Master of Arts.

____

Supervisor, Dr. John Boyce from Department of Economics

____

Dr. Patrick Coe from Department of Economics

____

Dr. Peter Ehlers from Department of Mathematics and Statistics

17 Dec 2003

Date

Abstract

Nonparametric methods are used in this paper to investigate the per capita personal income distribution and how it evolved over the period 1950 - 2000 across the 59 provinces and states in the U.S. and Canada. A nonparametric Silverman test is applied to examine whether multi-modality is present in the income distributions. Multi-modality is a significant feature of the per capita personal income distribution Across the provinces and states in the U.S. and Canada.

Both nonparametric stochastic kernel and Markov transition probability matrix are used to analyze the long run income distribution dynamics of the per capita personal income across the provinces and states in the U.S. and Canada. Signs of convergence and mobility are found in the 1950-1970 period, evidence in favor of convergence and mobility is weak in the 1971-1990 period, and a nearly perfect immobility is found in the 1990's. The income distribution for entire study period (1950 –2000) shows a form of club convergence.

Acknowledgements

I would like to express my greatest thanks to my supervisor, Dr. John Boyce. Without his idea and helpful suggestions and comments, this thesis could not have been completed.

Deep thanks are also conveyed to the Committee member: Dr. Patric Coe from Department of Economics and Dr. Peter Ehlers from Department of Mathematics and Statistics.

Finally I would like to thank my wife Wen and my parents for their supports.

Table of Contents

List of Tables

List of Figures

A Nonparametric Analysis of the Personal Income Distribution across the Provinces and States in the U.S. and Canada[1]

1. Introduction

In recent years there has been an increasing literature on income distribution analysis that makes use of nonparametric methods.

By estimating non-parametrically the cross-country income data, Bianchi (1997), Jones (1997) and Quah (1997) tested the growth *convergence*[2] and *convergence clubs*[3] hypothesis. They found evidence of "twin peaks" (*bimodality*) in the world relative income distribution via a thorough analysis of the changes in the shape of the income distribution.

Applying a tree-clustering algorithm on cross-country data, Durlauf and Johnson (1995) divide the countries into four different groups with different growth behavior. Multiple steady states (*multi-modality*) in cross-country growth behavior were found in their

[1] The author would like to thank John Boyce for his helpful suggestions and comments. The usual disclaimer may apply.

[2] Convergence is defined as the phenomenon of income levels in poor regions catching up in relative terms with those in the rich.

[3] Baumol (1986) and Quah (1993) have all been strong proponents of the idea of grouping similar countries or economies into different *convergence clubs*, such as the rich country club, the poor country club etc. While overall convergence may not exist, countries within a convergence club may show signs of convergence (*club convergence*).

regression results.

From the above one may realize that the test of multi-modality can be very useful in the research of income distributions. As indicated by Zhu (2002) important hypotheses such as income convergence, polarization, poverty traps and "vanishing middle class" can be formulated as a question of multi-modality.

In this paper, several nonparametric approaches, specifically, an adaptive kernel density estimator, one *mode* [4] testing procedures, stochastic kernel and Markov transition probability matrix were used to investigate convergence issues and analyze distribution dynamics of the personal income distribution across the 59 provinces and states in the US and Canada.

The non-parametric approach used in this paper is different from a conventional one. No particular functional form is imposed on the underlying density as in parametric approach. The empirical results of this work indicate that over the study period 1950-2000, multi-modality is a significant feature of the regional personal income distribution across the U.S. and Canada. Convergence (unimodality) is found over the 1950s and most of the 1960s. Evidence in favor of bimodality is found during the 1970's and the 1980's. In the 1990s, there is a strong indication that the number of the poor regions increased dramatically.

[4] A *mode* is defined as a point at which the gradient of the density changes from positive to negative.

The rest of the paper is outlined as follows. Section II provides a review of the conventional parametric estimation methods of the convergence of income distribution and a reasonable critique of these classic methods. Section III gives a detailed description of the kernel methodology used in this study. Section IV briefly discusses the source and quality of our income data. In section V a nonparametric bootstrap test (Silverman test) is performed with results reported. Section VI gives a brief introduction of DIP test. Section VII stochastic kernel and Markov transition probability matrix are used to analyze the income distribution dynamics. Section VIII concludes and gives future research suggestions.

2. Convergence: Classical Parametric Approach and Critics

Convergence, as stated by Abramovitz (1986), implies a long-run tendency towards the equalization of *per capita* income or product levels. In other words, convergence addresses the important question of whether poor countries, as measured by low per capita incomes, display faster growth rates in per capita income than rich countries with higher per capita incomes.

Numerous attempts have been made to provide a precise description of the convergence of income. There are three major classical convergence concepts: 1) Absolute β-convergence 2) Conditional β-convergence 3) Sigma (σ)-convergence.

In terms of methodology, there are three primary kinds of classical methods: (a) Cross-section approach (b) Time Series approach (c) Panel-data approach.

2.1 The Cross-section Approach

1) The absolute β-Convergence

The *absolute β-Convergence* approach, developed mainly by Baumol (1986) and Barro and Sala-i-Martín (1991, 1992), is to estimate cross-section growth rates on initial levels of income as follows:

$$\log(y_{i,t+T} / y_{i,t})/T = \alpha + \beta \log(y_{i,t}) + \varepsilon_{i,t} \tag{1}$$

where $\log(y_{i,t+T} / y_{i,t})/T$ is economy *i*'s growth rate of per capita income between *t* and *t+T*, $\log(y_{i,t})$ is the logarithm of economy *i*'s per capita income at time *t*, *T* is the length of time over which the growth of per capita income is measured, and $\varepsilon_{i,t}$ is a stochastic error term, α is a constant term representing the steady-state point of convergence which is the same for all economies and β is the convergence coefficient. According to the neo-classicists, a negative sign of the β coefficient indicates that the growth rates in per capita incomes over the *T* year period were negatively correlated with starting incomes, or in other words the initially poor regions grow faster than the initially richer ones. This is the idea of absolute β-convergence where all economies are assumed to converge to the same steady state represented by the coefficient α (Sala-i-Martin, 1994).

Absolute *β-convergence* has been widely tested for the U.S. regional convergence. Sala-i-Martin (1991, 1992) found evidence in favor of absolute *β-convergence* in both the per capita personal income and per capita gross state product with an estimated convergence speed at about 2% per annum for the period 1963-1986. [5]

The same tests for Canada have been done by Coulombe and Lee (1995), who found evidence of absolute *β-convergence* in six different measures of incomes across Canadian provinces for the period 1961-1991. Their findings also indicate that direct taxes and transfers may play crucial roles in increasing convergence speeds.

Sala-i-Martin (1996) pointed out that the convergence of all economies to the same steady state predicted by the neoclassical model relies heavily on the assumptions that the only difference across countries lies in their initial levels of capital. In reality, however, economies may differ in levels of technology, propensities to save, or population growth rates. With these differences in levels of technology and preferences, different economies most likely will have different steady states and the absolute β-convergence will be flawed by imposing the restriction that they are the same.

2) Conditional β-Convergence

To overcome the obvious flaw in the absolute β-estimation of the classical growth theory, authors such as Barro and Sala-i-Martín (1992), Mankiw, Romer and Weil (1992) and Durlauf (1996) among others, developed the idea of conditional *β-convergence.*

---

[5] Since this paper is concerned with regional convergence within the U.S. and Canada, only previous literature relevant to the U.S. and Canada is cited here.

Conditional β-convergence argues that economies converge to different steady-state points of growth since they have different economic structures. Convergence is conditional on the steady-state growth path which is a function of the differences in technology levels, human capital, investment and saving rates, among other structural variables.

To test the hypothesis of conditional β-convergence one has to extend equation (1) of the absolute β-convergence to incorporate some structural variables, such as the investment ratio, human capital, innovative activity, public expenditure, population growth, trade, and so on. In other words, instead of estimating (1) one estimates

$$\log\left(y_{i,t+T} / y_{i,t}\right)/T = \alpha + \beta \log\left(y_{i,t}\right) + \psi \mathrm{X}_{i,t} + \varepsilon_{i,t+T} \tag{2}$$

In equation (2), all the variables are the same as in equation (1), and $\mathrm{X}_{i,t}$ is a vector of structural variables (as proxies for the steady state) that affect the steady state of economy *i*. If the estimate of β is negative once $\mathrm{X}_{i,t}$ is held constant, then the data set is said to exhibit conditional β-convergence.

Conditional β-convergence has also been extensively tested for the U.S. regional convergence. Sala-i-Martin (1992) included sector shift parameters as well as regional dummies to hold constant possible shocks and found evidence of conditional β-convergence at a rate of about 2% per annum in per capita gross state product for the period 1963-1986. Sala-i-Martin claimed that there is no significant difference between absolute β-convergence and conditional β-convergence rate for the same study period.

A number of econometric problems have been identified with conditional β-convergence analysis. The initial level of technology, which should be included in a conditional β-convergence specification, is not observed. Since it is also correlated with other regressors (such as initial income), the conditional β-convergence studies suffer from an omitted variable bias.

3) Sigma (σ)-Convergence.

One of the drawbacks of conditional convergence is that it cannot tell if overall convergence among all economies exists or not. It can only tell if economies with similar characteristics are converging or not. This leads to another popular approach in measuring convergence across economies: *sigma-convergence*[6]. The concept of sigma-convergence is primarily developed by Easterlin (1960), Baumol (1986), Dowrick and Nguyen (1989), Barro and Sala-i-Martin (1991, 1992). The central idea is that if the standard deviations of per capita incomes across groups of economies exhibit a decreasing trend over time, then there is sigma-convergence. Formally, following Sala-i-Martin (1996), sigma-convergence occurs if

$$\sigma_{t+T} < \sigma_t, \tag{3}$$

[6] This terminology was first introduced by Sala-i-Martin (1990).

where σ_t is the time t standard deviation of $\log(y_{i,t})$ across i and T is the time period over which the standard deviation is observed.

The σ-convergence has been extensively tested for the U.S. case. Mitchener and McLean (1999) studied the US state personal income for 1880-1980 and found different sigma-convergence rates with different choices of series (nominal income, price adjusted income or labor productivity). They also showed that the West and the South play crucial roles in regional convergence at different times. Barro and Sala-i-Martin (1995) showed that the standard deviations of per capita personal income for 48 states declined from 1880 to 1920, rose in 1930, fell again from 1940 to mid-1970s, then increased from 1980 to 1988, and decreased from the end of 1980s to 1992. They attributed the fluctuations of the standard deviation to an agriculture price shock in 1920s and to oil shocks in 1970s.

As argued by Quah (1996), the main flaw of σ-convergence is that when σ_t is constant over a study period, thus signaling no convergence or divergence, the underlying economies may actually still be moving within an invariant distribution frame. For example, σ-convergence could not inform if clusters are forming within the cross section or if transitions occur within the distribution. Hence inter- and intra-distributional dynamics could not be uncovered by σ-convergence.

2.2 Time Series Approach

The idea to apply time-series approach to test convergence was mainly due to Bernard and Durlauf (1995, 1996), Durlauf (1989) and Quah (1992). Bernard and Durlauf (1996) defined time series convergence between two series X_t and Y_t as:

$$\lim_{k \to \infty} E(Y_{t+k} - X_{t+k} \mid I_t) = 0,$$

where I_t is information set at time t. In words, series Y and X converge if the long run forecasts for both series are equal at time t.

In practice, researchers use unit root tests and cointegration tests to investigate if time series convergence exists across all economies or for a specific pair of economies. Failure to reject unit root is equivalent to failure to find evidence of convergence.

A number of previous studies have used time series methods to test income convergence across countries. One interesting study was made by St Aubyn (1999). Using time series methods (augmented Dickey-Fuller (ADF) test and Kalman filter test), St Aubyn did not find evidence of convergence in per capita GDP between the U.S. and Canada through the post-war period 1947-1989 in either ADF test or Kalman filter test. St Aubyn attributed this result to the low power of time series method.

Brown, Coulson, and Engle (1990) found no time-series convergence across a number of U.S. states. Carlino and Mills (1993) incorporated trend breaks in their regression and could reject the unit root null in three out of eight US census regions for period 1929-

1990. Using the same data as Carlino and Mills, Loewy and Papell (1996) endogenized both the break date and the lag length and could reject the unit root in seven out of eight US regions for the same study period.

One criticism of time series approach to convergence was raised by Bernard and Durlauf (1996) who argued that the time series approach relies on the assumption that the per capita income differences between different economies are constant with zero mean, which contradicts the reality.

Moreover, in the time series approach convergence is based on different definitions from that of cross sections. As a result, time series often have different findings on the same dataset from its counterpart of cross-section approach.

2.3 The Panel Data Approach

Islam (1995) extends the cross-section regression to the panel case by formulating a dynamic panel data model as follows:

$$\ln y_{it} = \gamma \ln y_{i,t-1} + \beta \ln x_{it} + \mu_i + \eta_t + \upsilon_{it} \qquad (4)$$

where μ_i represents economy-specific effects (time invariant), η_t stands for time-specific effects (region invariant), υ_{it} is an error term that varies across regions and time periods and has mean zero and, x_{it} consists of variables such as saving rate, population growth rate, capital depreciation rate, technological growth rate and trade openness etc.

The major advantage of the panel data approach over the cross-section β-convergence approach is that the panel data specification allows one to control for differences in the initial level of technology, which could be captured by the region-specific fixed effects represented by the term μ_i. Thus the panel data approach makes it possible for one to correct the omitted-variable bias mentioned in the conditional β-convergence in section 2.1.

A panel unit root test is to test the null hypothesis that each series in the panel contains a unit root, i.e. H_0: $\gamma = 1$ against the alternative hypothesis that all individual series in the panel are stationary and display convergence, i.e. H_1: $\gamma < 1$.

The panel unit root technique has been applied to study regional convergence in Canada by Wakerly (2002) who rejected the null hypothesis of unit root.

However, the panel-data analysis could be problematic when it is applied to growth convergence analysis. Note that the major difference between equation (4) and equation (1) or (2) is that the constant α in equation (1) or (2) has been decomposed into economy-specific and time-specific effects μ_i and η_t:

$$\alpha = \mu_i + \eta_t \tag{5}$$

This decomposition is considered an advantage of the panel-data approach over the cross-section approach. However, for convergence analysis, this decomposition of α could result in problems.

Durlauf and Quah (1999) pointed out that there are at least two major problems with the panel data regression. First, equation (1) implies that the initial level of technology (and thus α through μ_i) forms part of the long run convergence path of the given economies. In case of absolute β-convergence, one could conclude that convergence occurs precisely when the poor catch up with the rich. Hence, a convergence finding could be *transparently* translated into a statement about catching up. On the contrary, when the initial level of technology is allowed to vary across economies, finding convergence to an underlying steady state could not be simply interpreted as catching up occurs between poor and rich. If instead the differences in initial level of technology are not modeled as functions of right-hand side variables, the question whether poor economies are catching up with rich ones would be left unanswered. Therefore panel-data approach makes it even more difficult to explain convergence results in terms of catching up from poor to rich.

Second, one of the traditional problems in the panel data regression like equation (4) is that the μ_i's may be correlated with some of the right hand side variables, thus causing inconsistency estimates (Nickell (1981), pp. 417-26). One solution to the inconsistency problem is to transform equation (4) to eliminate μ_i. Without μ_i, a researcher is left with only one choice to analyze a left-hand side variable without its long-run variations which he originally intends to investigate.

Chamberlain (1984) also argued that the solution provided by panel-data techniques in eliminating the correlation between individual effects and right-hand side variables ends up severely limiting its power in explaining patterns of cross-country growth and convergence.

The above classic approaches are frequently used to analyze the regional income distributions, which involve functions of the parameters of interest. Unfortunately, all the classic approaches explained so far are with defects in some ways.

3. Non-parametric Density[7] Estimation

Unlike the parametric approach, nonparametric density estimation allows one to draw a complete picture and hence provides full information on the entire income distribution.

Nonparametric estimation approach allows one to analyze data at hand without any *a priori* assumptions on the form of the underlying density of the data. The only requirement about the data, if any, is perhaps that the underlying density of the data should be smooth enough for meaningful analysis.

[7] This is the univariate case.

Although researchers such as Jones (1997), Bianchi (1997) and Quah (1997) have extensively used the nonparametric kernel estimator with *fixed* window width (explained in Section 3.1) to analyze the world income distribution and growth convergence, few studies were found to use the kernel estimator with *variable* window width to make relevant researches. Zhu (2002) pioneered in applying an adaptive kernel estimator to analyze the personal income distributions in the U.S. However, an in-depth analysis of the per capita personal income across provinces (states) in *both* the U.S. *and* Canada based on *adaptive kernel density estimation* (explained in Section 3.2) has not yet been found in the previous studies on income distribution dynamics.

Fixed kernel estimates are locally insensitive to the detail changes in the data. An adaptive kernel estimator has the advantage over the fixed kernel in that it allows the smoothing parameter to vary according to the local density of data.

This paper uses the adaptive kernel estimator, mainly due to two reasons. First, it is simple to calculate. Second, as Zhu (2002) argued when testing *mode*, due to the relative sparseness of data in the tails and abundance of data in the center, false modes in the tails may be reported well before true modes in the center could be detected. Local smoothing allows one to obtain a reasonably robust estimate of income density.

3.1 Fixed Kernel Estimator

There are a variety of ways of non-parametrically computing a density estimate[8]. The kernel estimator is one of the most popular estimators due to its simplicity to calculate and interpret.

Let $f = f(x)$ denote the continuous density function of a random variable X at a point x, and let $x_1, \ldots, x_n$ be the observations from f.

Rosenblatt (1956) defined a kernel function K as:

$$\int_{-\infty}^{\infty} K(\psi)d\psi = 1. \tag{3.1}$$

where $K(\psi) \geq 0$.

The general kernel estimator $\hat{f}(x)$ of $f(x)$ is defined by:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) = \frac{1}{nh}\sum_{i=1}^{n} K(\psi_i), \tag{3.2}$$

where $\psi_i = h^{-1}(x_i - x)$, n is the number of observations in the sample, h is the window-width (bandwidth) which is a function of the sample size and goes to zero as $n \rightarrow \infty$.

The above kernel is known as the *fixed window width kernel estimator* since the window width does not vary with the density.

---

[8] Density estimation is one of the most fundamental problems in statistics. Consider a univariate continuous random variable X distributed according to a probability density f. This means that for any interesting set B of real numbers we can find the probability that X belongs to this set by the formula $P(X \in B) = \int_B f(x)dx$. For instance, for B=[a, b] we get $P(a \leq X \leq b) = \int_a^b f(x)dx$. Then one observes n independent realizations $X_1, X_2, \ldots, X_n$ of X, and the aim is to find an estimate $\tilde{f}(x)$, based on these observations, that fits the underlying $f(x)$.

3.2 Adaptive Kernel Estimator

Breiman et al. (1977) and Abramson (1982) developed *adaptive kernel estimator* (*AKE*) which makes the window width vary inversely with the density. Formally the adaptive kernel estimator is defined as

$$\hat{f}_1(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_{ni}}K\left(\frac{x_i - x}{h_{ni}}\right), \tag{3.3}$$

The *AKE* is essentially (3.2) with $h_{ni} = h\delta_{ni}$, h is the overall window width and δ_{ni} is the local bandwidth parameter defined by

$$\delta_{ni} = \left[\frac{\tilde{f}(x_i)}{G}\right]^{-\lambda}, \tag{3.4}$$

where G is the geometric mean of some preliminary estimator of the density $\tilde{f}(x_i)$ over all x_i, i.e.

$$G = \left(\prod_{i=1}^{n}\tilde{f}(x_i)\right)^{1/n} \tag{3.5}$$

$0 < \lambda \leq 1$ is a sensitivity parameter, and $\tilde{f}(x_i)$ could be any convenient initial estimator. If $\lambda = 0$, the *AKE* is reduced to the fixed bandwidth case (3.2). Abramson (1982) and others suggest that λ should be set to 0.5, which reduces bias significantly compared to its fixed-width counterpart. In this study λ is set to 0.5 accordingly.

3.3 Kernel Estimator Properties and Choice of Bandwidth h

Pagan and Ullah (1999) indicated that the kernel K is a symmetric function around zero satisfying

(1) $\int K(\psi)d\psi = 1$

(2) $\int \psi^2 K(\psi)d\psi = m \neq 0$

(3) $\int K^2(\psi)d\psi < \infty$.

Devroye(1983) and Devroye and Gyorfi (1985) showed that the kernel estimator $\hat{f}$ is asymptotically *unbiased* and *consistent*.[9]

Hardle (1990), Silverman (1986) and others argued that the choice of kernel is not crucial to analysis since any kernel could be optimal for large enough samples. In contrast the selection of the window width h is critical. The window width determines the degree of smoothing of the estimated density. There is always a trade-off between bias and variance when choosing window width h. On the one hand, a very small h may result in an under-smoothed density estimate since there may not be enough points for smoothing. Under-smoothing may reduce the bias of the estimator but increase its variance. A large h, on the other hand, could result in an over-smoothed density because a large number of points are used in forming an estimate. Over-smoothing reduces variances but increases

[9] The proof of this can also be found in Pagan and Ullah (1999): pp 33-34.

bias, besides missing important details about the distribution. Hence great caution should be executed in selecting the window width h.

One has to have some criterion in order to make the optimal choice of h. So far the most popular method (originally introduced by Rosenblatt (1956)) has been to minimize $E\left\{\int\left[\hat{f}(x)-f(x)\right]^2 dx\right\}$, the mean integrated squared error (*MISE*).

$$
\begin{aligned}
MISE(\hat{f}) = E\left\{\int\left[\hat{f}(x)-f(x)\right]^2 dx\right\} &= \int\left[E\hat{f}(x)-f(x)\right]^2 dx + \int Var(\hat{f}(x))dx \\
&= \int\left[\left(Bias\ \hat{f}\right)^2 + Var\left(\hat{f}\right)\right]dx
\end{aligned}
$$

AMISE [10] (Asymptotic MISE) is used to approximate *MISE* which is difficult to calculate.

$$
\begin{aligned}
MISE(\hat{f}) \approx AMISE(\hat{f}) &= \frac{h^4}{4}\mu_2^2 \int\left(f^{(2)}(x)\right)^2 dx + (nh)^{-1}\int f(x)dx \int K^2(\psi)d\psi \\
&= \frac{h^4}{4}\mu_2^2 \int\left(f^{(2)}(x)\right)^2 dx + (nh)^{-1}\int K^2(\psi)d\psi
\end{aligned}
$$

Silverman (1986) showed that the value of h that minimizes *AMISE* is

$$
h = 0.9\left[\min(\hat{\sigma}, R/1.34)\right]n^{-1/5},
$$

[10] Proof of this can be found in Pagan and Ullah (1999): pp 22-24

where R stands for the inter-quartile range and $\hat{\sigma}$ is the sample standard deviation.

In the *AKE* case, the local bandwidth parameter δ_{ni} allows the window width to be adjusted according to the local density of the data. Therefore bandwidth chosen by the adaptive method gives a more accurate estimate. The Gaussian kernel[11] is chosen in this paper mainly due to the reason that the Gaussian kernel has 95.1% efficiency relative to the Epanechnikov kernel in minimizing *MISE* as shown by Fox (1990).

4. Data: Sources, Transformation and Sub-periods Division

Fifty states in the U.S. and nine provinces in Canada are included in this study[12]. The per-capita personal income over the period 1950 - 2000 for the 59 regions in the U.S. and Canada is used to analyze the regional income distribution in this paper.

The data of the regional per capita personal income in the U.S. is taken from the well-known and widely used Bureau of Economic Analysis Regional Accounts Data (1929 – 2001) and the data of the provincial per-capita personal income in Canada is extracted from CANSIM (Canadian Socio-Economic Information Management System)[13]. CANSIM is the Statistics Canada's computerized database of time series covering a wide variety of social and economic aspects of Canadian life.

---

[11] The Gaussian Kernel is defined as: $K(x) = \frac{1}{\sqrt{2\pi}} e^{-0.5x^2}$

[12] See appendix TableA.1 for detailed list of the provinces and states.

[13] CANSIM is Statistics Canada's computerized database and information retrieval service. Data before 1990 are from CANSIM II SERIES V501122, data after 1990 are from CANSIM II SERIES v691825.

The per-capita personal income in each year for each region is normalized by the average per-capita personal income for the total 59 economies in that year (with the average taking a value of 1.00).

Since one of the purposes of this study is to compare per capita personal income across years, incomes are reported in 1982-84 prices by correcting for inflation using the consumption price index (CPI)[14].

In order to make the provincial personal income in Canada comparable to that in the U.S., the personal incomes in Canada are converted to U.S. dollars as per average exchange rate in each year[15].

In this paper the period 1950 to 2000 is divided into three sub-periods: 1950-1970, 1971-1990 and 1991-2000.

5. A Preliminary Look at the Income Distribution

Figure 1 presents the entire distribution of per capita personal income (all relative to the North American average, excluding Mexico) across 9 Canadian provinces and 50 U.S.

---

[14] The Canada CPI and the U.S. CPI are from CANSIM Series P10000 and D19805 respectively.

[15] There are some disputes among economic researchers as to whether exchange rate or PPP should be used as a conversion factor. PPP is based on the assumption that if goods are priced differently in different countries, then consumers would simply switch to buying the cheaper good, thereby moving the exchange rate towards PPP. In reality, transactions costs plus customs regulations usually prevent such switching taking place. The response of trade to movements in real exchange rates is limited, contradicting the simple notion of arbitrage embodied in PPP. Further future research could be done as to compare the empirical results from these two methods of conversions.

states for the period 1950 – 2000.

The North American average is indicated by one on the vertical *(Z)* axis marked *Relative Income* in Figure 1. Time periods are sequentially marked along the *X*-axis of *Year*. Different provinces and regions are represented along the *Y*-axis marked *States and Provinces*.[16]

The Canadian provinces are indicated by vertical lines along the *Y*-axis at 1.0 (Prince Edward Island), 2.0 (New Brunswick), 3.0 (Nova Scotia), 4.0 (Saskatchewan), 5.0 (Quebec), 8.0 (Alberta), 9.0 (Manitoba), 14.0 (Ontario) and 15.0 (British Columbia) respectively. It can be noted that in 1950 the relative incomes in all the Canadian provinces are below the North American average.

From Figure 1, it is clear that the pure use of either cross-sectional distribution or time-series distribution would fail to provide complete intra-distribution information. Since Figure 1 does not give a clear view on how the relative personal incomes across the nine Canadian provinces evolve over the whole study period 1950 –2000, the relative incomes of the nine Canadian provinces were presented in Figure 2.

[16] Provinces and states are ordered from the lowest per capita income to the highest per capita income in 1950 as shown in the appendix Table A1.

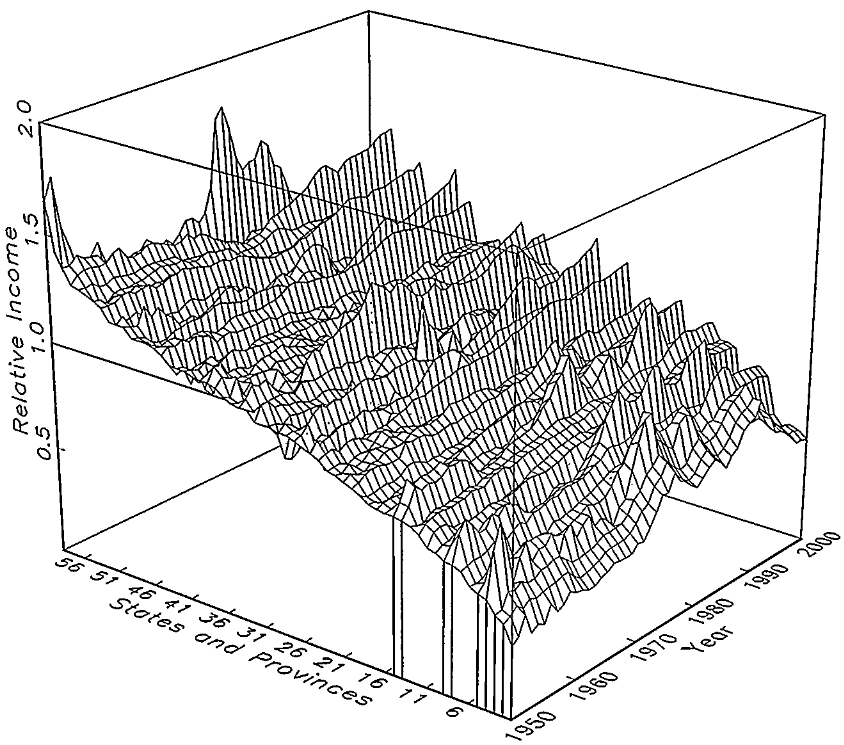


Figure 1. Relative incomes across 59 Provinces and States

in the U.S. and Canada for 1950-2000

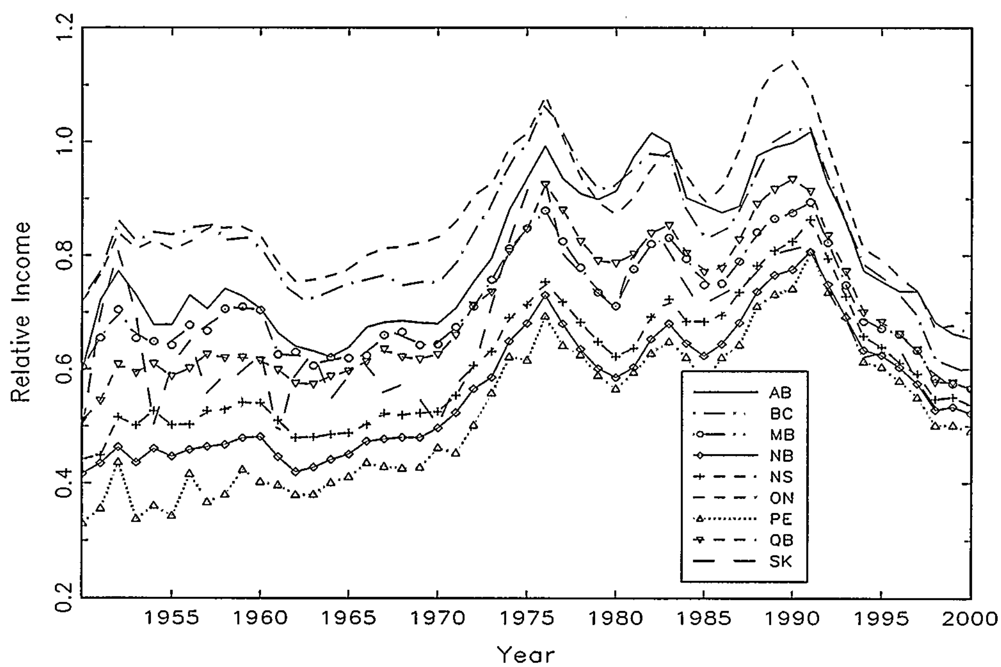


Figure 2. Relative incomes in Canadian provinces for 1950 -2000

From Figure 2, one may notice that the income levels in the Canadian provinces experienced a slow convergence to the North American average from 1950 to 1970, but became approximately stagnated between 1970 and 1990, then diverged from the North American average over the 1990's. It is also interesting to note that Ontario, British Columbia and Alberta have almost always been the three highest income provinces over this whole study period and all experienced income levels higher than the North American average for some years between 1970 and 1990.

5.1 The Shapes of the Distributions

As described in Section 3, the adaptive kernel estimator is preferred to the fixed kernel estimator in that it allows window width vary according to the local density of data. The merit of adaptive kernel estimator is more apparent when estimating long-tailed or multi-modality income distributions. Adaptive kernel estimator can reduce the variance of the estimates in regions with low density of data and decrease the bias of the estimates in regions with high density of data while a fixed bandwidth approach may result in under-smoothing in areas with few observations while over-smoothing in others.

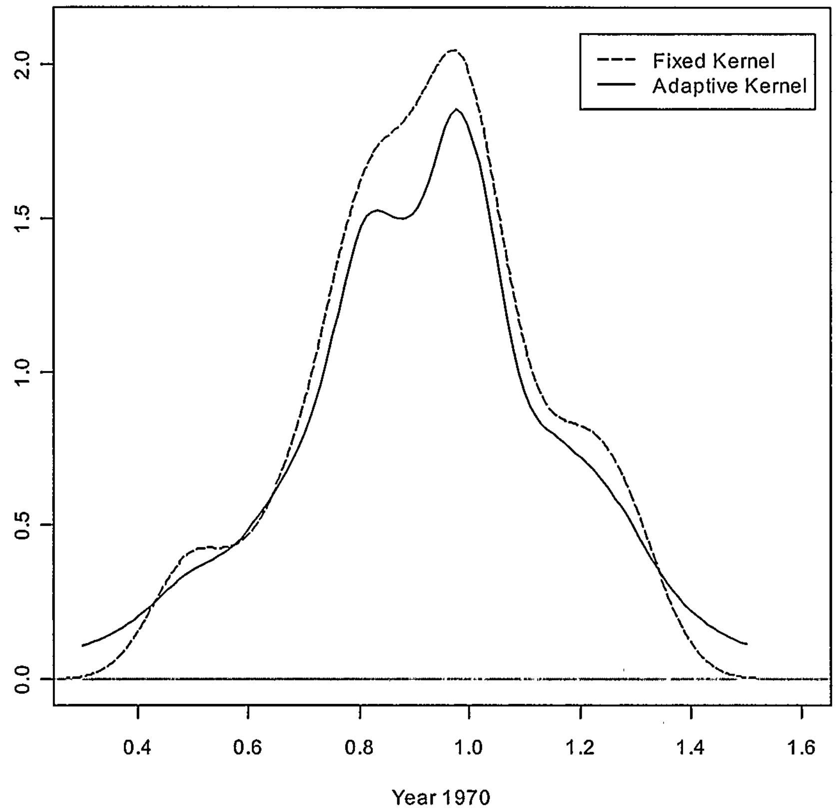


Figure 3. Fixed and Adaptive Kernel Density Estimates of the Personal Income in 1970

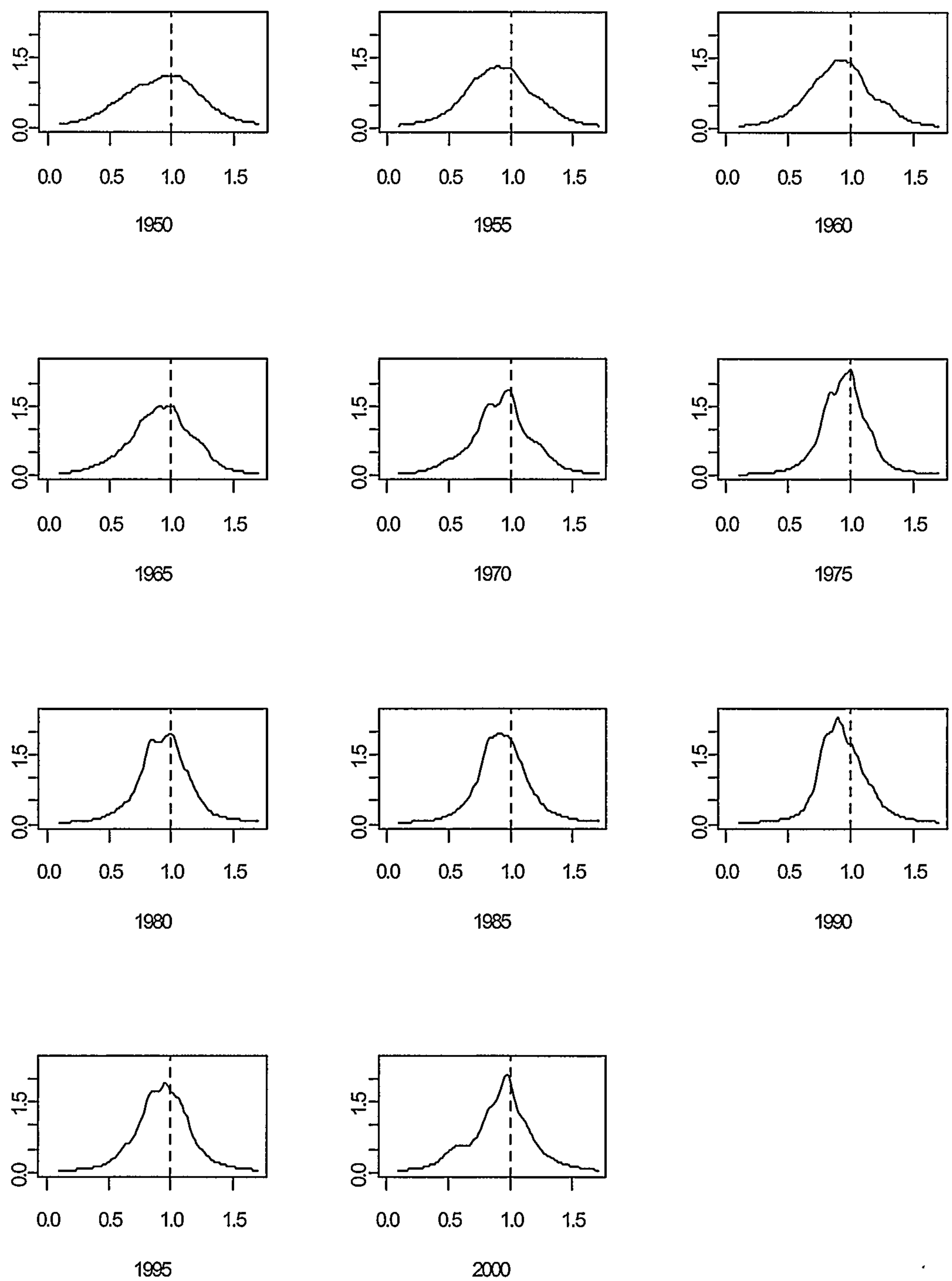
Figure 3 gives an example of the fixed and adaptive kernel density estimates of the per capita personal income distribution in 1970. It can be clearly seen that the adaptive kernel estimate successfully detects the bimodality in the distribution while the fixed kernel fails to do so. In this example, the fixed bandwidth over-smoothed the middle of the distribution and under-smoothed the tails of the distribution. On the contrary, the adaptive kernel estimate gave a proper estimate in the central distribution (with high density of data) and smoothed out the obvious left and right bumps in the left and right tail (with low density of data).

Due to the obvious advantages of adaptive kernel estimate, the study in this paper pertains to the use of adaptive kernel approach in density estimates. Figure 4 presents the non-parametric adaptive kernel density estimates of relative per capita personal incomes in 1950, 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995 and 2000.

The shapes of the income distributions estimated for the provinces and states have changed significantly for different years. In the density estimates of 1950, 1955, 1960 and 1965, there are no signs of multi-modality. In the density estimate of the 1970, there is strong evidence of bimodality with one peak at 1.0 and another peak at 0.8. The density estimates of 1975 and 1980 seem to be bi-modal too; however the two modes are very close. There is no sign of multi-modality in the density estimate of 1985. No multi-modality is detected in the density estimate of 2000, but there is a bump in the lower left tail though not very strong.

In order to effectively capture the changes of densities over 1950's, adaptive kernel density estimates of income distributions in 1950, 1955 and 1960 are presented in the upper left panel of Figure 5. Though the densities of all of these three years are unimodal, the shapes have changed a lot. By comparing the densities, a clear process of convergence can be observed, since the central mass of the density has increased substantially from 1950 to 1955 and from 1955 to 1960 as indicated by the height of the central peak which has increased from 1.1 in 1950 to about 1.4 in 1955 and to 1.5 in 1960. This figure shows that the strong convergence that occurred during this period is mainly caused by the noticeable decrease in the density of the left tail (poor regions

Figure 4. Adaptive Kernel Density (Gaussian) of the relative per capita Income for Provinces and States in Canada and the U.S. over 1950 – 2000



becoming better off) as well as the decrease in the density of the right tail (rich regions becoming relatively poorer). It is also worth mentioning that the regional incomes converge to 0.9 instead of the average 1.0 as indicated by the shift of the mode from 1.0 in 1950 to 0.9 in 1955 and in 1960.

Compared with the density of 1960, the density of 1965 does not change much though the mode has shifted back to 1.0 in 1965 from 0.8 in 1960. However, bimodality is clearly present in the density estimate of 1970. One may also notice that the process of convergence that occurred in 1950's still goes on because there is a clear right shift in the left tail and a decrease in the right tail in the density of 1970 compared with that of 1965. The densities of 1966, 1967, 1968 and 1969[17] are also bimodal by visual inspection. Therefore, a process of club convergence seems to emerge in the late 1960's.

In 1975 there were further decreases in both the density of the left tail and the density of the right tail, indicating the number of poor regions have decreased by becoming relatively better off while the number of rich regions further decrease by becoming relatively poorer. But the density of the right shoulder has increased, indicating there is an increase in the number of upper middle income regions which may be due to the fact some rich regions have decreased by becoming relatively poorer and moving to upper-middle income ranks. The bimodality in the density of 1975 is ambiguous. In 1980 a sign of divergence emerges since there are increases in both the density of left tail and the density of right tail, which result in a decrease in the central mass of density. Visual

[17] See Figure A2 in appendix.

inspection of the densities in 1970's[18] indicates that the density in 1971, 1974, 1977 and 1978 are bimodal too. In sum, there are both signs of convergence and divergence in 1970's.

The density of 1985 is clearly unimodal without any ambiguity of bimodality as shown in the density of 1980. A notable feature in the density of 1990 is that there is a clear left shift of the density mass and the central mode has shifted to the left to about 0.9 in 1985, indicating a large number of middle-income regions became relatively poorer than before. There is a slight right shift of the left tail of the 1980 density, signaling the poor regions become relatively richer than before. Visual inspection of the densities in the 1980's[19] indicates the density estimates of 1981, 1982 and 1989 are bimodal. In sum, same as in 1970's, both signs of convergence and divergence are present in the density estimates of 1980's.

In 1995 two modes are present: a major mode around 0.9 and another around 0.7. There is a significant increase in both the density of the left tail and the mass of the right shoulder of the 1995 density, indicating a large number of poor regions have actually become relatively poorer than before and some middle income regions have become relatively richer and moved to upper-middle income ranks. The lower-middle income peak around 0.7 exhibited in the central mass of the 1995 density has completely disappeared in 2000, implying that the lower-middle income regions are vanishing. The central peak moved back to 1.0 in 2000 from 0.9 in 1995, signaling that a larger number

[18] See Figure A3 in appendix.
[19] See Figure A4 in appendix.

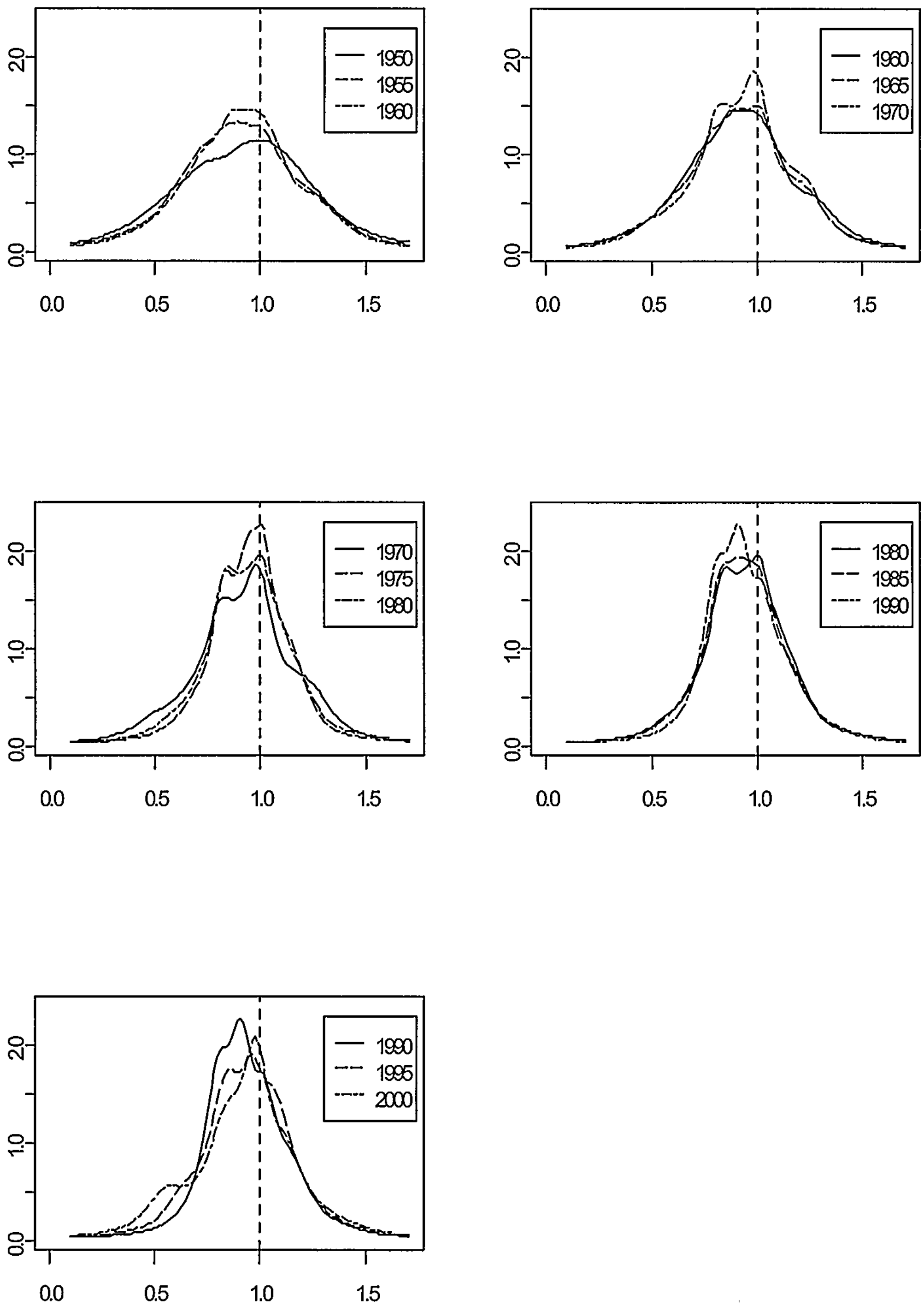
of middle-income regions have become relatively richer than before. The number of upper-middle income regions has decreased as indicated by the decrease in the density of the right shoulder of the 2000 density. The increase in the density of right tail is not significant, implying that the number of rich regions does not increase much. Furthermore, the increase in the central mass of 2000 may be partially due to the fact that quite a significant number of lower-middle income regions disappeared by becoming either poorer or richer as well as partially due to a large decrease in the upper-middle income regions. The number of the poor has increased as indicated by the appearing of a significant bump in the left tail of the density. An examination of the income data in 2000 indicates that the low-income group (less than 60% of the average income in 2000) consists of six Canadian provinces: Prince Edward Island, New Brunswick, Nova Scotia, Saskatchewan, Quebec and Manitoba. This indicates that at the end of 1990's Canadian provinces lag further behind their U.S. counterparts. The densities of 1992, 1994 and 1995[20] are bimodal by visual inspection. In sum, in the 1990's there is a strong indication that the number of the poor increased dramatically and a mode of the low-income group is likely to emerge.

The presence of multi-modality in the income distributions will be more thoroughly investigated in the next few sections.

---

[20] See Figure A5 in appendix.

Fig. 5. Change of the density shapes over different time periods



6. Multi-modality Tests

Multi-modality test for a density estimate usually involves two steps: estimating the density and rejecting the null hypothesis of uni-modality. In many previous empirical studies, however, it is a quite common practice that a density is taken as bimodal as long as it displays two peaks. But showing multi-modality in a density estimate is one thing, and proving it is multi-modal is quite another.

There exist numerous parametric methods of testing for multi-modality which, however, require parameters that need *a priori* assumptions. To avoid these *a priori* assumptions, nonparametric methods are applied to determine whether a density function is actually multi-modal.

6.1 Silverman Test for Multi-modality

The Silverman test for multi-modality was developed by Silverman (1981) on the basis of bootstrapping to test the null hypothesis H_0 that the density f has n modes, against the alternative H_1 that f has more than n modes ($n = 1, 2, \ldots$). Following Silverman (1981), the number of modes of density f can be defined as:

$$Mode(f) = \#\left\{x \in \Re_+ \mid f'(x) = 0 \ and \ f''(x) < 0\right\}$$

Silverman (1981) defined the n-th critical bandwidth $\hat{h}_{n,critical}$ as:

$$\hat{h}_{n,critical} = \inf\{h \mid Mode(\hat{f}_h) \leq n\}$$

where $\hat{f}$ is an adaptive kernel density estimate of f with overall bandwidth h.

The statistic to be used is $Mode(\hat{f}_h)$. Critical bandwidth is one of the key concepts in non-parametric bootstrap modality tests. Since the degree of smoothness depends on h, it follows that as h increases the number of modes will not increase. As an example, in Figure 6 the number of modes in the kernel density estimate of per capita personal income in 2000 is plotted against the window size h. It can be clearly seen that the number of modes either increases or remains unchanged when the window size h increases.

Suppose the true underlying density has two modes. To test the null hypothesis H_0 that the underlying density f has 1 mode against the alternative H_1 that f has more than 1 mode, h needs to be larger than $\hat{h}_{2,critical}$ because a considerable amount of smoothing is required in order to obtain a unimodal density estimate from a bimodal density. This suggests that $\hat{h}_{n,critical}$ can be used as a statistic to test the null hypothesis H_0 that $f(x)$ has n modes versus the alternative H_1 that $f(x)$ has more than n modes. A large value of $\hat{h}_{n,critical}$ indicates more than n modes, thus rejecting the null.

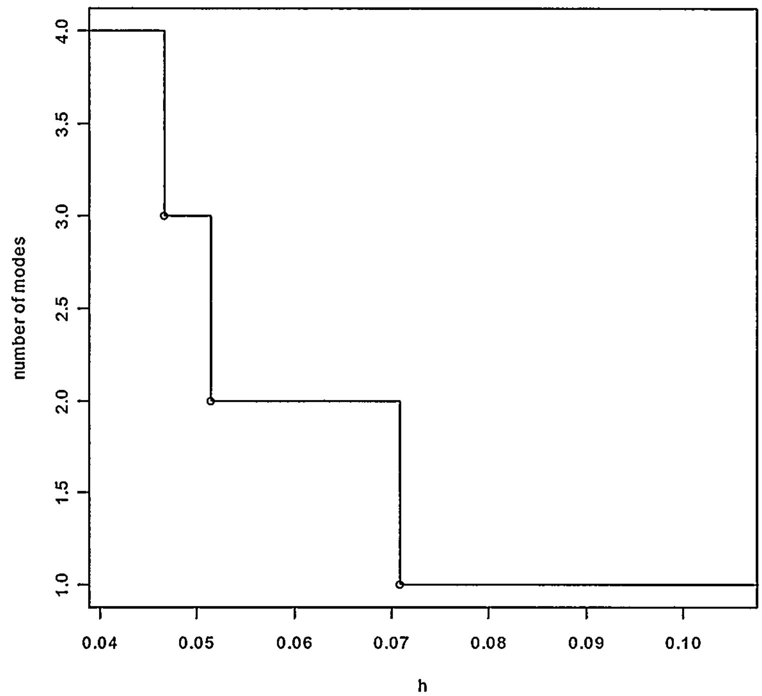


Figure 6. Number of modes in the kernel density estimate as a function of window size h (year 2000)

Silverman (1981) developed the multi-modality test on the basis of the following theorem:

Theorem[21]**:** Given any fixed X_1, ..., X_n, let $\hat{f}_h$ be a kernel density estimate of f with bandwidth h and the Gaussian kernel function $K(\cdot)$. Then, $Mode(\hat{f}_h)$ is a right continuous decreasing function of h.

[21] Proof of this theorem could be found in Silverman (1981), Journal of Royal Statistical Society, Series B., Vol. 43, No. 1, page 98.

It follows immediately from this theorem, $Mode(\hat{f}_h) > n$ if and only if $h < \hat{h}_{n,critical}$. As per this theorem either $\hat{h}_{n,critical}$ or $Mode(\hat{f}_h)$ could be used as a test statistic.

Following Silverman (1981), Efron and Tibshirani (1993), Davison and Hinkley (1997) and Zhu (2002), the significance of $\hat{h}_{n,critical}$ could be estimated from the data against its bootstrap distribution. The multi-modality bootstrap test procedure can be described as follows:

1. For a data set $\{Y_1, \cdots, Y_n\}$, randomly draw with replacement B bootstrap samples of size n $\{Y_1^b, \cdots, Y_n^b\}_{b=1}^B$;

2. Obtain B *smooth bootstrap*[22] samples $\{X_1^b, \cdots, X_n^b\}_{b=1}^B$ by computing

$$X_i^b = \overline{Y}^b + (1 + \hat{h}_{n,critical}^2 / \hat{\sigma}^2)^{-1/2}(Y_i^b - \overline{Y}^b + \hat{h}_{n,critical}\varepsilon_i) \tag{6.1.1}$$

   where $\overline{Y}^b$ is the mean of $Y_1^b, Y_2^b, \ldots, Y_n^b$, and ε_i is assumed to be distributed standard normal. The scaling factor $\left(1 + \hat{h}_{n,critical}^2 / \hat{\sigma}^2\right)^{-1/2}$ scales the variance of the bootstrap samples $\hat{\sigma}^2$ so that it has the same variance as that of the original data;

3. For $b = 1, \ldots, B$, obtain bootstrap estimates $\hat{f}_{hcrit}^b$ with the critical bandwidth $\hat{h}_{n,critical}$ computed as per adaptive kernel method. For each bootstrap sample,

[22] This is, rather than sampling with replacement from the data, a sampling from a smooth estimate of the population. For this reason it is called smooth bootstrap.

compute the n-th critical bandwidth $\hat{h}^{b}_{n,critical}$;

4. Compute the achieved significance level (ASL)[23] as:

$$A\hat{S}L_{boot} = \#\left\{\hat{h}^{b}_{n,critical} \geq \hat{h}_{n,critical}\right\}/B \qquad (6.1.2)$$

5. Reject the null hypothesis of n modes in the underlying density whenever $A\hat{S}L_{boot}$ is smaller than standard levels of significance.

In Table 1 the critical bandwidths and the p-values of the nonparametric bootstrap test are presented for each year from 1950 to 2000. In 1950 and 2000 the bootstrap test fails to reject the null of unimodality even at the 10 per cent significance level. In 1970 and 1992 the unimodality hypothesis can be rejected at the 10% and 5% significance level respectively. Specifically, in 1992 only as few as 23 times $\hat{h}^{b}_{1,critical}$ is greater than $\hat{h}_{1,critical}$ out of 500 bootstrap samples giving an achieved significance level of 0.046; in 1970 the ASL_{boot} is 0.09.

The null of unimodality can also be rejected at 10% significance level in 1951, 1967, 1978, 1989, and 1994 and at 5% significance level in 1966, 1974 and1969. In total, unimodality is rejected in ten years out of the sample of 51 years.

---

[23] The ASL or p-value of the test is: $ASL_{boot} = \Pr ob_{\hat{F}_n}\left\{\hat{h}^{b}_{n,critical} > \hat{h}_{n,critical}\right\}$

Table 1. Nonparametric Bootstrap test for *n*-modality: critical bandwidth and the *p* values

Year	*m*=1	*m*=2	*m*=3	Year	*m*=1	*m*=2	*m*=3
1950	0.0991 (0.738)	0.0926 (0.278)	0.0508 (0.550)	1975	0.0658 (0.592)	0.0512 (0.124)	0.0436 (0.056)
1951	0.1120 (0.078)*	0.0662 (0.316)	0.0540 (0.678)	1976	0.0548 (0.802)	0.0293 (0.59)	0.0274 (0.688)
1952	0.1463 (0.306)	0.0560 (0.416)	0.0445 (0.43)	1977	0.0832 (0.294)	0.048 (0.124)	0.0452 (0.196)
1953	0.0793 (0.336)	0.0709 (0.092)	0.0678 (0.144)	1978	0.0635 (0.082)*	0.0387 (0.748)	0.0244 (0.702)
1954	0.1028 (0.138)	0.0611 (0.328)	0.0552 (0.152)	1979	0.0568 (0.698)	0.0402 (0.698)	0.034 (0.63)
1955	0.0879 (0.71)	0.0655 (0.172)	0.0645 (0.204)	1980	0.0797 (0.398)	0.0351 (0.644)	0.031 (0.598)
1956	0.0697 (0.354)	0.0628 (0.564)	0.0540 (0.106)	1981	0.0756 (0.37)	0.0345 (0.582)	0.0336 (0.296)
1957	0.0667 (0.87)	0.062 (0.136)	0.0499 (0.496)	1982	0.0776 (0.358)	0.0411 (0.67)	0.0396 (0.368)
1958	0.0846 (0.472)	0.06 (0.514)	0.0591 (0.17)	1983	0.0665 (0.49)	0.0372 (0.72)	0.0216 (0.828)
1959	0.0783 (0.204)	0.0647 (0.428)	0.0465 (0.588)	1984	0.0547 (0.144)	0.0431 (0.13)	0.0319 (0.294)
1960	0.0808 (0.164)	0.0747 (0.23)	0.0653 (0.136)	1985	0.0593 (0.67)	0.0519 (0.028)	0.0379 (0.094)
1961	0.0968 (0.508)	0.0815 (0.212)	0.0724 (0.002)	1986	0.0636 (0.554)	0.0459 (0.548)	0.0351 (0.542)
1962	0.0796 (0.736)	0.0684 (0.47)	0.058 (0.28)	1987	0.0623 (0.126)	0.0349 (0.876)	0.0344 (0.2)
1963	0.0906 (0.118)	0.0646 (0.476)	0.0555 (0.358)	1988	0.0742 (0.458)	0.0337 (0.926)	0.0331 (0.21)
1964	0.098 (0.52)	0.0755 (0.31)	0.07 (0.008)	1989	0.071 (0.066)*	0.048 (0.436)	0.0361 (0.484)
1965	0.0992 (0.458)	0.0645 (0.106)	0.0612 (0.178)	1990	0.0576 (0.138)	0.0573 (0.176)	0.0347 (0.44)
1966	0.1018 (0.022)**	0.0514 (0.742)	0.0485 (0.178)	1991	0.0572 (0.586)	0.0534 (0.014)	0.0317 (0.148)
1967	0.0965 (0.058)*	0.0606 (0.472)	0.0336 (0.76)	1992	0.063 (0.046)**	0.033 (0.426)	0.028 (0.686)
1968	0.1127 (0.282)	0.0655 (0.392)	0.0357 (0.574)	1993	0.0815 (0.324)	0.0451 (0.358)	0.0368 (0.366)
1969	0.0983 (0.034)**	0.0584 (0.55)	0.0452 (0.562)	1994	0.0679 (0.066)*	0.0477 (0.456)	0.0334 (0.702)
1970	0.0791 (0.09)*	0.0508 (0.328)	0.0506 (0.088)	1995	0.0775 (0.448)	0.0521 (0.398)	0.0403 (0.118)
1971	0.0851 (0.386)	0.0546 (0.122)	0.0371 (0.774)	1996	0.0551 (0.344)	0.0495 (0.13)	0.0407 (0.386)
1972	0.0514 (0.418)	0.0467 (0.598)	0.0359 (0.326)	1997	0.0711 (0.114)	0.0485 (0.596)	0.047 (0.046)
1973	0.0534 (0.370)	0.0530 (0.074)	0.0449 (0.026)	1998	0.0657 (0.21)	0.06 (0.08)	0.0438 (0.486)
1974	0.0722 (0.036)**	0.0421 (0.556)	0.0309 (0.406)	1999	0.0687 (0.61)	0.0634 (0.072)	0.0432 (0.184)
				2000	0.0708 (0.190)	0.0515 (0.614)	0.0467 (0.422)

Note: Numbers without bracket represent critical bandwidths, numbers in bracket stands for p-values and m indicates the number of modes, * indicates unimodality is rejected at the 10% significance level and ** indicates the unimodality is rejected at 5% significance level.

The conservatism of Silverman test is reflected by the fact some density estimates showing multi-modality could not be detected by the Silverman test. Silverman (1983) acknowledged that his test might be conservative and suffers from low power in that it could fail to reject the null hypothesis when it is false, thus underestimating the number of modes. But a merit of this is that this test is unlikely to falsely reject a null hypothesis when it is true.

It can also be noted that there seems two groupings of bimodality distributions: one around 1966 – 1970 period and the other one around 1989 – 1994 period. This suggests that bimodality distributions tend to move continuously over a period of time.

This paper chooses to use a combination of Silverman test plus visual inspections of the density estimates to determine the number of modes in a distribution. Since the Silverman test indicates that income distributions in 1951, 1966, 1967, 1969, 1970, 1974, 1978, 1989, 1992 and 1994 are multi-modal, and a visual inspection of the density estimates in these years also confirms the results from the Silverman test, it can be safely concluded that bimodality exists in the income distributions of these years.

6.2 Dip Test of Unimodality

It is worth mentioning that another way to test multi-modality is the *DIP* test developed by J. A. Hartigan and P. M. Hartigan (1985) who defined the *DIP* as the maximum difference between the empirical distribution function and the unimodal distribution function that minimizes this maximum difference.

Let $\{x_1, \ldots, x_n\}$ be a data set from a density $f(x)$, and $F(x)$ be the empirical cumulative distribution function of the data. *DIP* test could be utilized to test if $f(x)$ is unimodal. Following Hartigan and Hartigan (1985), the DIP of F is defined as

$$D = D(F,U) \equiv \sup_x |F(x) - U(x)|$$

where U is the best fitted unimodal *cumulative* distribution function to F.

However, the *DIP* test is designed to test the null of unimodality against the alternative of multi-modality. Thus DIP could not be used to test the exact number of modes once the null is rejected. In addition, the DIP test is difficult to implement and rather conservative. The Silverman test, on the other hand, has the advantage of being able to test the null of exactly n modes against the alternative of more than n modes. Moreover, the Silverman test has much higher power than the DIP test. Therefore, the DIP test is not applied in this paper.

7. Income Distribution Dynamics

Figure 7 tracks the evolution of the distribution of the relative per capita personal income across the 59 US-Canada regions through the period 1950 – 2000. The middle line shows the 50th percentile (median) of the relative income distribution in each year. The top and bottom lines show, respectively the 75^{th} and 25^{th} percentiles of the relative income distribution. The distance between 75^{th} percentile line and the 25^{th} percentile line indicates the inter-quartile range (*IQR*). The IQR is defined as

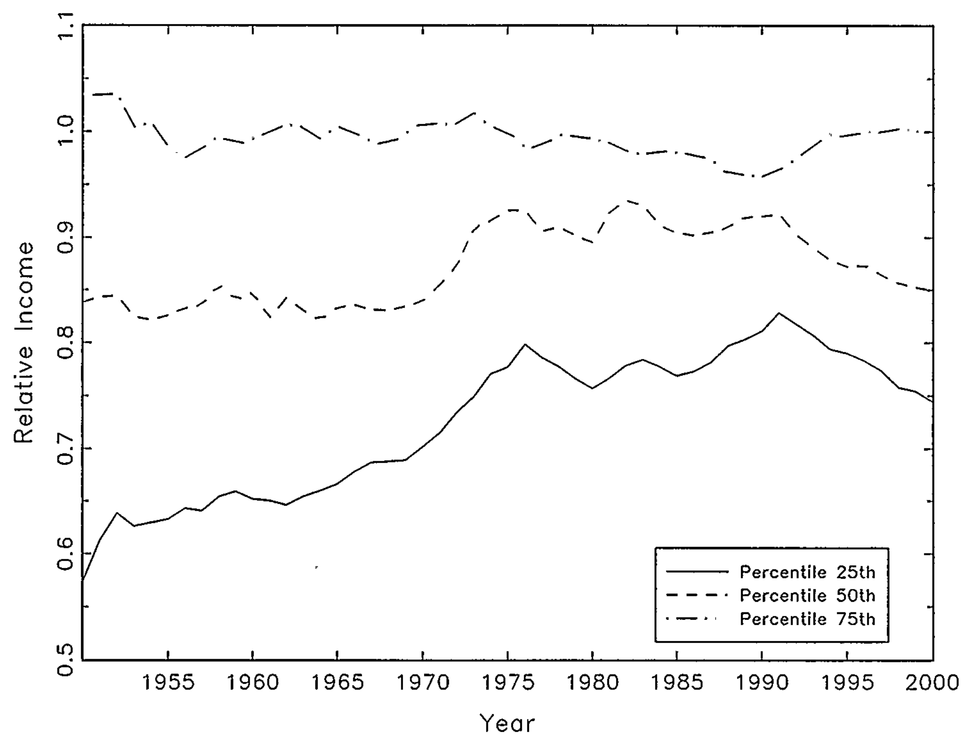
$$IQR = Q_3 - Q_1$$

where Q_3 indicates the 75^{th} percentile and Q_1 represents the 25^{th} percentile.

From 1950 to 1970, the 25^{th} percentile group experienced a significant increase in its relative income while the 75^{th} percentile and the 50^{th} percentile remained approximately unchanged. Over the following period 1971-1990, the relative incomes of the 25^{th} percentile group and the 50^{th} percentile group were fairly stable without significant change while the relative income of the 25^{th} percentile group became stable after a sharp increase in the beginning of this period. In the 1990's, the 25^{th} and the 50^{th} percentile group experienced some decreases and the 75^{th} percentile group was rising.

Another fact derived from Figure 7 is that the cross-sectional distribution of income is quite tight, centered on approximately 0.9. The distance of the lines from each other also provides important information about the income gap. The closer are the lines, the lower is the income gap. Figure 7 suggests personal income gap in the U.S. and Canada experienced a period of rapid decrease from the 1950's to the 1970's, but slow increase in the 1990's. This finding is consistent with the previous findings in Section 5.

Figure 7. Evolution of Relative Income



7.1. Three-Dimensional Representations: The Stochastic Kernel[24]

Quah (1996) argued that a stochastic kernel (as well as its contour) is a graphical representation of the transition probabilities with the advantage that it gives estimates for continuous states of transition probabilities.

Let $\{x_1, \dots, x_n\}$ be a set of income data at time *t* from density $f_t(x)$, and after period *k* the income changes to $\{y_1, \dots, y_n\}$ and its corresponding density evolves to $f_{t+k}(y)$. Then

[24] The stochastic kernel is a conditional density function. Estimation of the kernel is carried out by first estimating the joint density function of the process at time *t* and *t*+ *k* and then normalizing it by the marginal in *t*. See Quah (1996a).

the relation between these two densities can be described by[25]:

$$f_{t+k}(y) = \int_0^{\infty} T_k(y \mid x) f_t(x) dx \tag{7.1}$$

where $T_k(y|x)$ is the stochastic kernel (transition probability) that could be used to describe the income distribution from time t to time $t+k$.

Following Quah (1997), a stochastic kernel conveys important information on income distributions. In this paper, the relative incomes in period t and $t+k$ are respectively represented along the t axis and the $t+k$ axis. The 45-degree diagonal line represents constant income. Thus, points lying along the 45-degree line indicate that relative incomes remain unchanged, while points to the left (right) of the diagonal signal a rise (decrease) in relative incomes between any two periods studied. *Unimodality* in a stochastic kernel indicates convergence while *multi-modality* may indicate club convergence or divergence.

Furthermore, the tendency to convergence is also indicated if most of the stochastic kernel mass becomes more parallel to the t axis by making a clock-wise movement around the center of the 45-degree line while the tendency to divergence is implied if most of the stochastic kernel mass becomes more vertical to the $t + k$ axis by making a counter-clockwise movement around the center of the 45-degree line.

7.2. Empirical Evidence from the Stochastic Kernel

[25] First order Markov process is assumed here.

Figure 8 ~ Figure 15 present the stochastic kernel and contour plots for the 59 regions in each of the three sub-periods: 1950 – 1970, 1971 – 1990, 1991 – 2000 and the entire study period (1950 – 2000)[26].

The stochastic kernel for the 1950-1970 period is characterized by multi-modality. Two central peaks are present: one clear middle-income mode at about 1.0 and another less clear lower-middle income mode at about 0.8. Most mass of the stochastic kernel does not run parallel to the 45-degree line, thus intra-distributional mobility in regional income rankings is rather active throughout this period. Though most mass of the two central peaks sits on the 45-degree line, neither the high-income mode nor the low-income mode is centered on the 45-degree line, instead both exhibit clockwise movement, thus signaling a strong tendency of convergence: the rich becomes relatively poorer and the poor becomes relatively richer over the six year horizon. In sum, mobility and convergence are strongly indicated by the stochastic kernel of this period. The evidence of convergence exhibited by the stochastic kernel of this period is consistent with the previous results in section 5.2 and 5.3.

In the 1971-1990 period, the middle-income mode once clearly presented in the 1950-1970 stochastic kernel has disappeared while the lower-middle income mode becomes significant and clear. Though most of the central mass of the stochastic kernel is approximately on the 45-degree diagonal (actually some portion of it is below the diagonal), most of the low-income peak is above the 45-degree line and actually located

[26] The stochastic kernels are of transitions of 6 years, 6 years, 3 years and 15 years respectively for each study period. Transition of years can be set at other numbers and will give similar results.

to the left of the diagonal. Most poor regions were becoming richer and caught up with the lower-middle income regions while most middle-income regions remained unchanged and a small number of them even fell back in their income ranks. As a result of this middle-income regions and low-income regions tend to converge to the lower-middle income peak. The high-income group is splitting (since while most portion of it is below the 45-degree diagonal a small portion of it is above the diagonal), most of them become poorer than before while only a small portion may become richer than before. The underlying converging force is not powerful as that in 1950-1970 but still evident. Since most of the middle-income peak is on the 45-degree line, mobility is not very significant. In sum, the 1971-1990 period shows less mobility and convergence than the 1950-1970 period.

The most striking feature exhibited by the stochastic kernel in the 1991-2000 period is immobility as the whole mass of the stochastic kernel is nearly perfectly symmetric on the 45-degree line. However, the obvious immobility in the New Age is not a complete picture. A little noise to the nearly perfect symmetric picture is that the poor income cluster is not on the 45-degree line, instead it is below the diagonal and exhibits a clear counter-clockwise movement. This indicates that while most regions remain where they are in their income ranks the poor regions are actually becoming even poorer. Thus the income gap between the average and the poor regions is actually increasing - an evidence of divergence. Another noticeable feature is that the low-middle-income peak exhibited in the 1971-1990 period has disappeared in this period, and a middle-income peak emerges again. This may be due to a significant number of the lower-middle income regions vanishing by becoming either poorer or richer.

The fact that the poor regions become even poorer in the 1990's is somewhat surprising and contradictory to the conventional wisdom that suggests income level for every region should increase due to the remarkable economic expansion caused in part by the extensive application of IT technology everywhere through this period. This may be partially due to the fact that the poor regions are unable to make sufficient IT investments or attract highly skilled IT professionals due to their financial resource limitations.

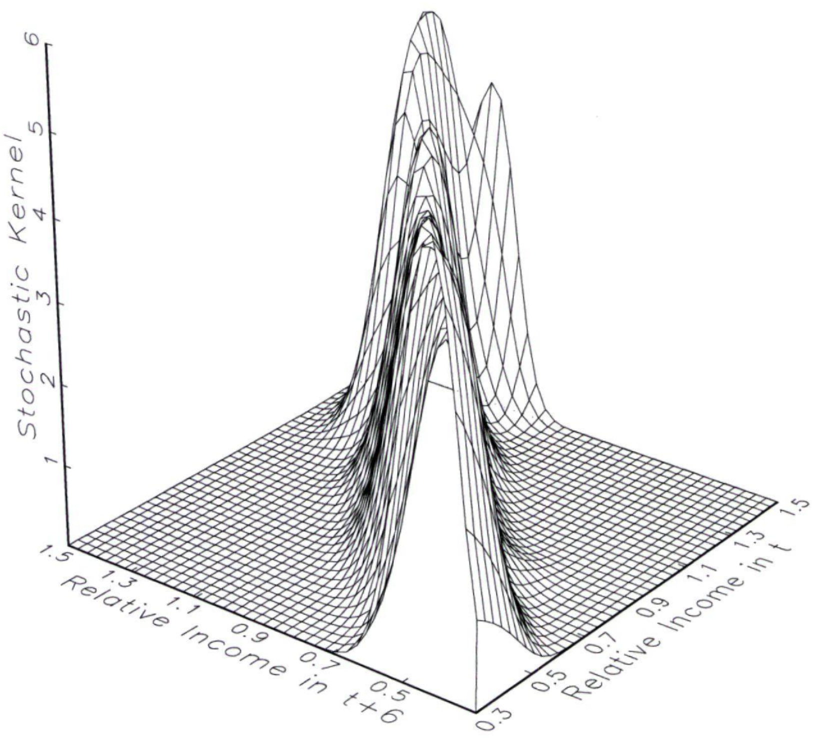


Figure 8. Stochastic Kernel 1950-1970

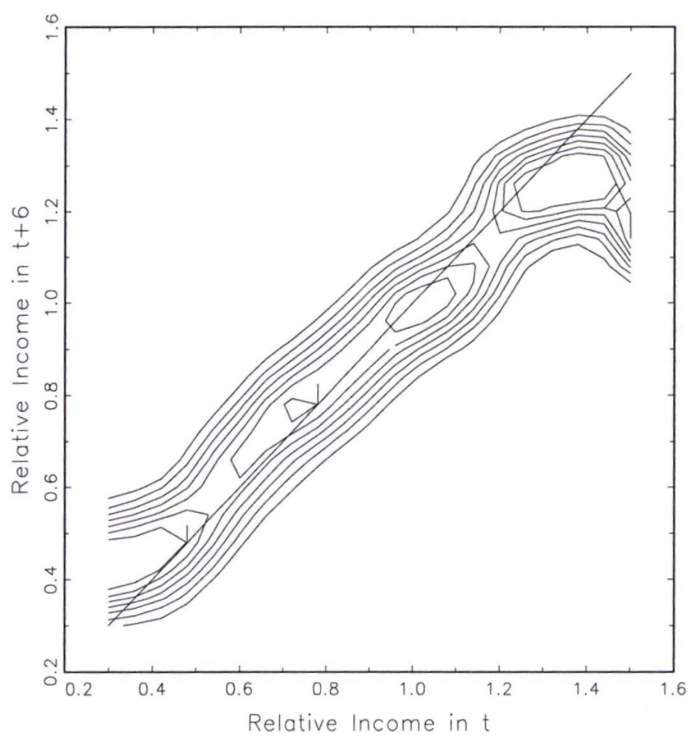


Figure 9. Contour for Figure 8

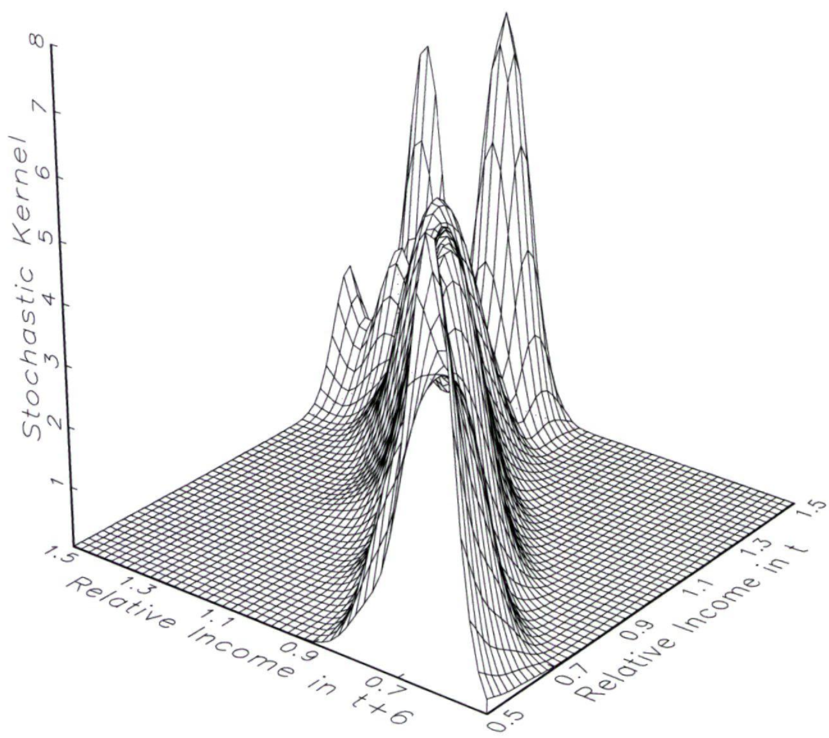


Figure 10. Stochastic Kernel 1971-1990

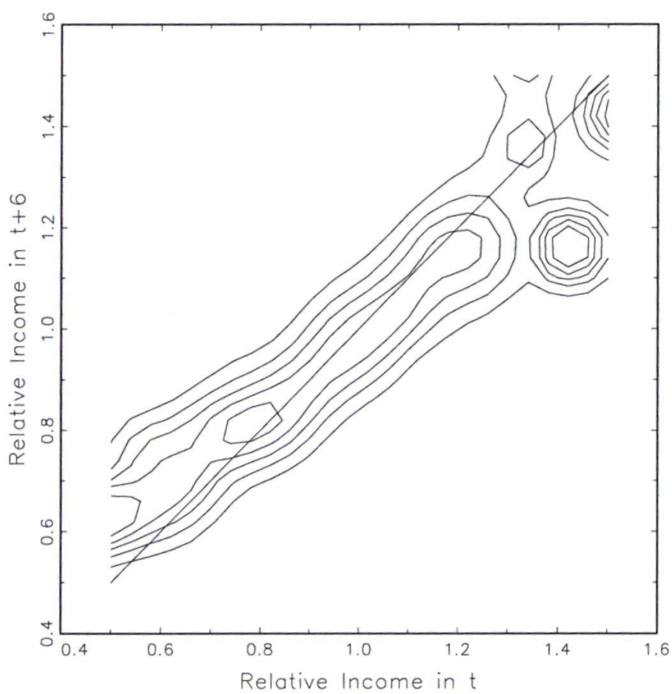


Figure 11. Contour for Figure 10

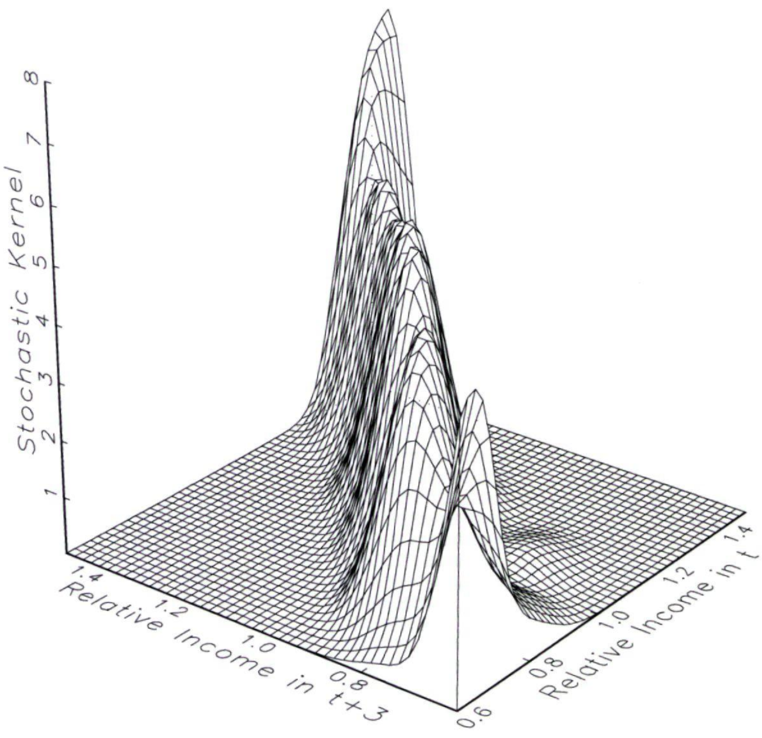


Figure 12. Stochastic Kernel 1991-2000

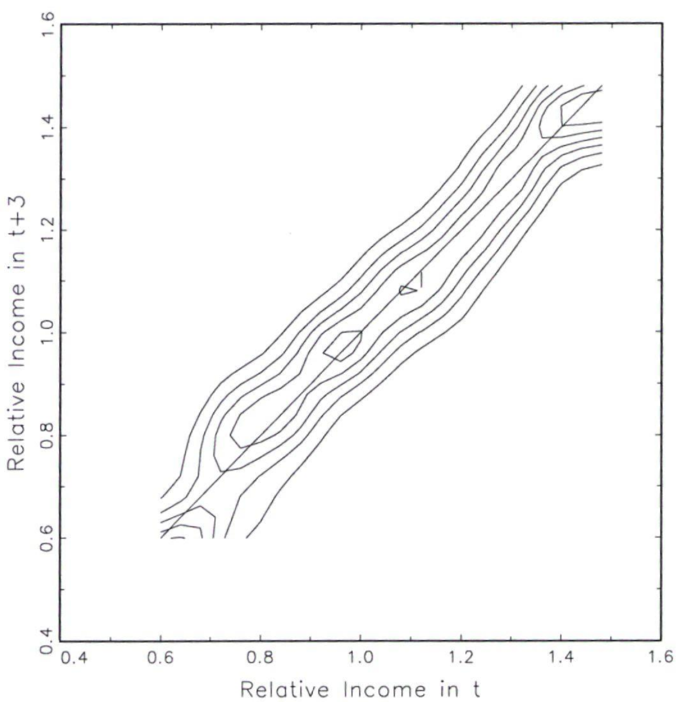


Figure 13. Contour for Figure 12

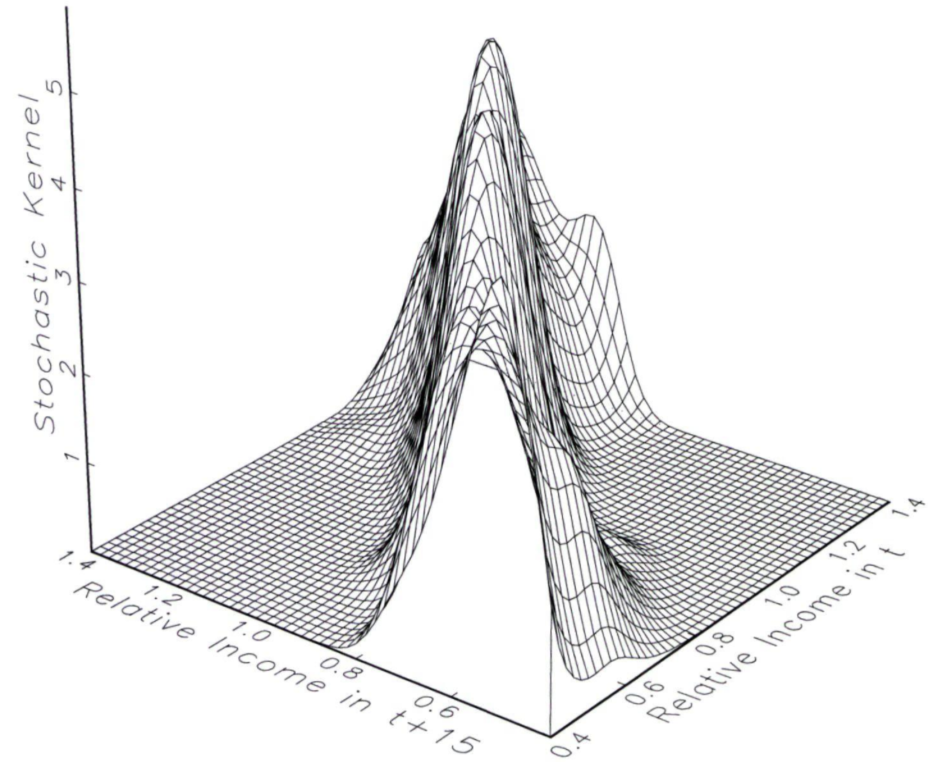


Figure 14. Stochastic Kernel 1950 - 2000

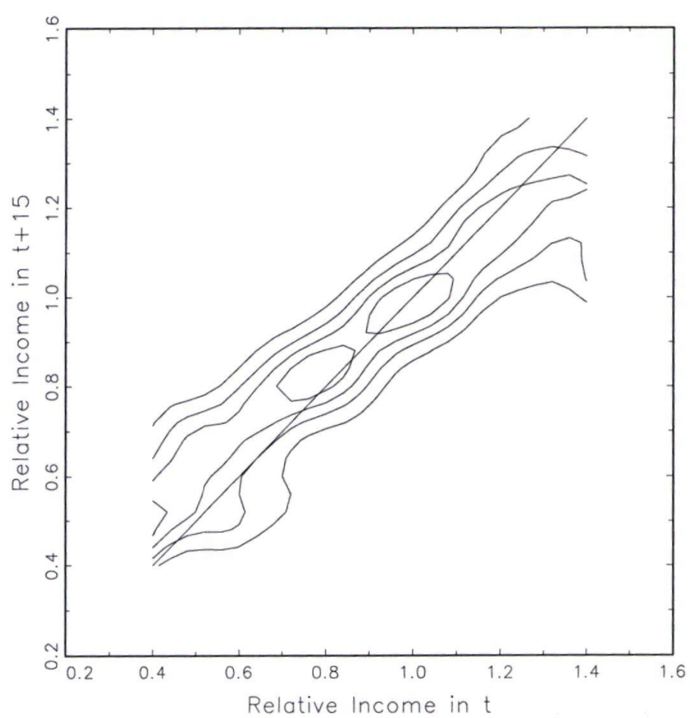


Figure 15. Contour for Figure 14

At this point, one would like to know how the stochastic kernel of the whole study period 1950–2000 looks like. Figure 14 and figure 15 shows the stochastic kernel of the relative income distribution between 1950 and 2000. There are two distinct features in the stochastic kernel of the entire period: twin peaks (bimodality) - the average income peak at 1.0 and the low-middle income peak at about 0.8. Since most mass of the stochastic kernel is on the 45-degree line, one can conclude mobility of income distributions is not very strong throughout the whole study period. It seems regions tend to converge to different steady states and exhibit a form of club-convergence in the long run.

7.3. Markov Transition Probability Matrix (MTPM)

Another traditional nonparametric way to describe income distribution dynamics is transition probability matrix, which is actually the discrete version of the stochastic kernel.

A simple first order Markov chain can be described as

$$X_{t+1} = M_{t+1} X_t$$

where X_t is a vector giving the number of economies in each income state at time t and X_{t+1} is a vector a period later. If the probabilities of the economies moving between different income states from period t to $t+1$ can be described by M_{t+1}, then M_{t+1} is a transition probability matrix. In practice, a transition probability matrix is estimated by using discrete states of income distributions.

Following Quah (1993), the first column in a transition probability matrix table gives the total numbers of observations in a specific state in a study period, the first row provides the upper points of the corresponding cells. Each row *i* gives the estimated probabilities of staying in that state *i* and of moving from state *i* to other states. The last row presents the ergodic distribution. Numbers along the diagonal indicate the level of immobility while off-diagonal numbers imply the degree of mobility.

7.4 Results from MTPM

Table 2 presents one-year horizon transition probabilities between different income states for the period 1950 - 2000.

The values in the main diagonal of Table 2 for the one-year transition are around 90%, indicating a high degree of immobility during this period of study. This finding is consistent with the previous results: the stochastic kernel for the period 1950 – 2000 indeed exhibits a low degree of mobility.

Mobility is considerably higher in the second state than in other states of the income distribution. It shows that a region in state 2 has 10% probability to move to the lower state and 11% probability to move to the higher state 3. This indicates that there is a strong tendency for the upper-low-middle income regions to move out of their initial income states.

Table 2. Markov Transition Probability Matrix, 1-Year Horizon (1950 – 2000)

Number	Upper endpoint			
	0.85	*0.95*	*1.1*	*1.6*
(931)	*0.94*	*0.06*		
(536)	*0.10*	*0.79*	*0.11*	
(855)		*0.07*	*0.90*	*0.03*
(563)			*0.06*	*0.94*
Ergodic	*0.292*	*0.191*	*0.325*	*0.193*

Table 3. Markov Transition Probability Matrix, 5-Year Horizon (1950 – 2000)

Number	Upper endpoint			
	0.85	*0.95*	*1.1*	*1.6*
(873)	*0.83*	*0.15*	*0.02*	
(536)	*0.17*	*0.57*	*0.25*	
(855)	*0.02*	*0.12*	*0.77*	*0.09*
(563)	*0.01*		*0.14*	*0.85*
Ergodic	*0.233*	*0.185*	*0.356*	*0.226*

The ergodic distribution was reported in the last row of the table. The ergodic distribution indicates the long run tendency of an economy staying in a given state regardless of its initial state. The results indicate that over the long run, the probability of an economy staying in the 3rd state is the highest, a little over 32% and the probability of landing in the 1st state is the second highest, nearly 30%. This is encouraging as it indicates that overall the middle-income regions are stable in the long term in the U.S and Canada despite a seeming form of club convergence.

Similar results are presented in Table 3 which gives a five-year horizon transition probability between different income states for the same period. It shows that the degree of mobility is higher over a five-year transition period. The results for the second state group are still noticeable in that they show a probability of a region moving out of its own initial state is much higher than that of any other regions.

It is noteworthy that Table 3 shows the probability that an economy in the fourth state has a one percent probability to move to the first sate. An examination of the data shows that it is most likely that Montana, which started as a high-income region in 1950, has suffered from one of the worst economic contractions and experienced a constant decrease in per capita personal income from 1960's and fell into the low-income group by the end of 1990's.

In sum, the results from Markov transition probability confirm the previous findings in section 7.2.

8. Conclusions and Suggestions for Future Research

Several nonparametric methods, namely kernel density, Silverman test, stochastic kernel and Markov chain, have been used to examine the per capita personal state (provincial) income distribution in the U.S and Canada in this paper.

The Silverman test shows that the state (provincial) personal income distribution across the province and states in the U.S. and Canada has been bimodal in 1970 and 1992 and unimodal in 1950 and 2000.

Both stochastic kernel and Markov transition probability matrix are applied to examine the long run state income distribution dynamics, such as convergence and mobility across the provinces and states in the U.S. and Canada over the period 1950-2000.

There are evidence of convergence and mobility for the period 1950 – 1970. Multi-modality has characterized the 1971-1990 period, the 1991-2000 period and the entire study period (1950 – 2000). The 1971-1990 period shows weak sign of convergence and mobility. New Ages exhibits almost perfect immobility. The entire study period shows a form of club convergence.

Income gap experienced a rapid decrease in the 1950-1970 period and kept almost unchanged in the 1971-1990 period. An obvious reversal to this general trend is in 1991-2000, a decade marked with remarkable economic expansion and the advent of a "New Economy"; most poor regions become even poorer than before.

This paper shows the merits of several nonparametric methods in analyzing income distributions when multi-modality is present. Most states or provincial governments make polices to eliminate regional income disparities on the basis of results from classical parametric methods, but this may be inappropriate because, for example, a mean value (first moment), such as of income, from a bimodality density distribution does not provide the same inferences as that from a unimodality density distribution.

One limitation of this paper is that it does not provide economic rationales as why bimodality may or may not occur in different periods. The goal of further research will be to apply the nonparametric methods employed in this paper to analyze the per capita personal income distribution conditional on variables (such as saving rates, population growth rates, trade openness, physical neighbors and taxation levels etc.) that can provide additional information on income distributions.

Appendix

Table A1. Provinces and States in Canada and the U.S.[27]

Region	Code	Region	Code
Prince Edward Island	1	North Dakota	30
New Brunswick	2	Texas	31
Nova Scotia	3	Arizona	32
Saskatchewan	4	Missouri	33
Quebec	5	Hawaii	34
Mississippi	6	Minnesota	35
Arkansas	7	Kansas	36
Alberta	8	Wisconsin	37
Manitoba	9	Colorado	38
Alabama	10	Indiana	39
South Carolina	11	Iowa	40
Kentucky	12	Pennsylvania	41
Tennessee	13	Rhode Island	42
Ontario	14	Nebraska	43
British Columbia	15	Ohio	44
West Virginia	16	Maryland	45
Georgia	17	Montana	46
North Carolina	18	Massachusetts	47
Louisiana	19	Oregon	48
Oklahoma	20	Michigan	49
Vermont	21	Wyoming	50
Maine	22	Washington	51
New Mexico	23	New Jersey	52
Virginia	24	Illinois	53
South Dakota	25	New York	54
Florida	26	California	55
Idaho	27	Connecticut	56
New Hampshire	28	Nevada	57
Utah	29	Delaware	58
		Alaska	59

27 Canadian provinces are in bold letters. Canadian province Newfoundland is not included in this paper as Newfoundland joined Canada in 1950, it has long been isolated from the main Canada economy and its data may suffer from major structure changes and are unreliable.

Figure A1

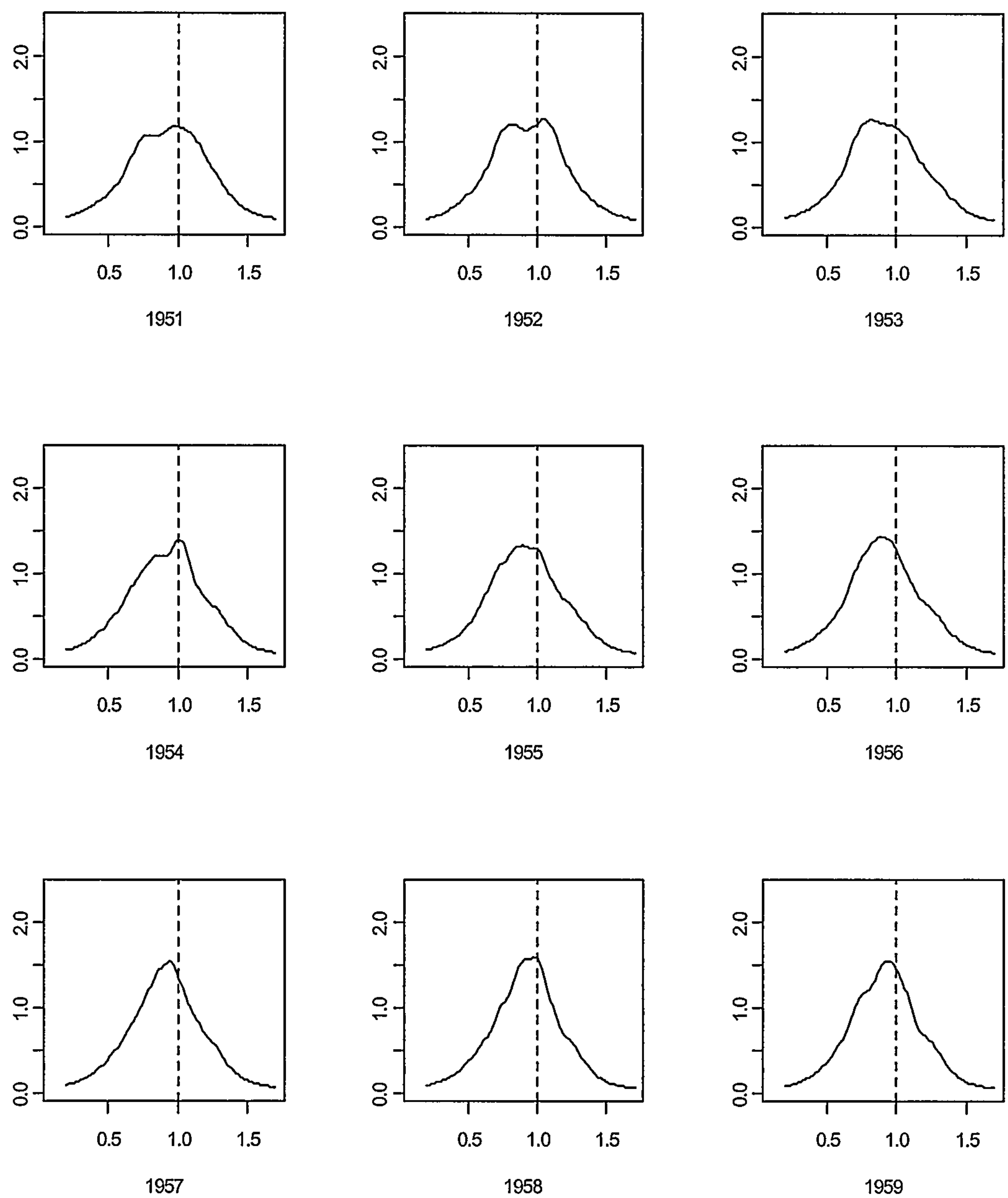


Figure A2

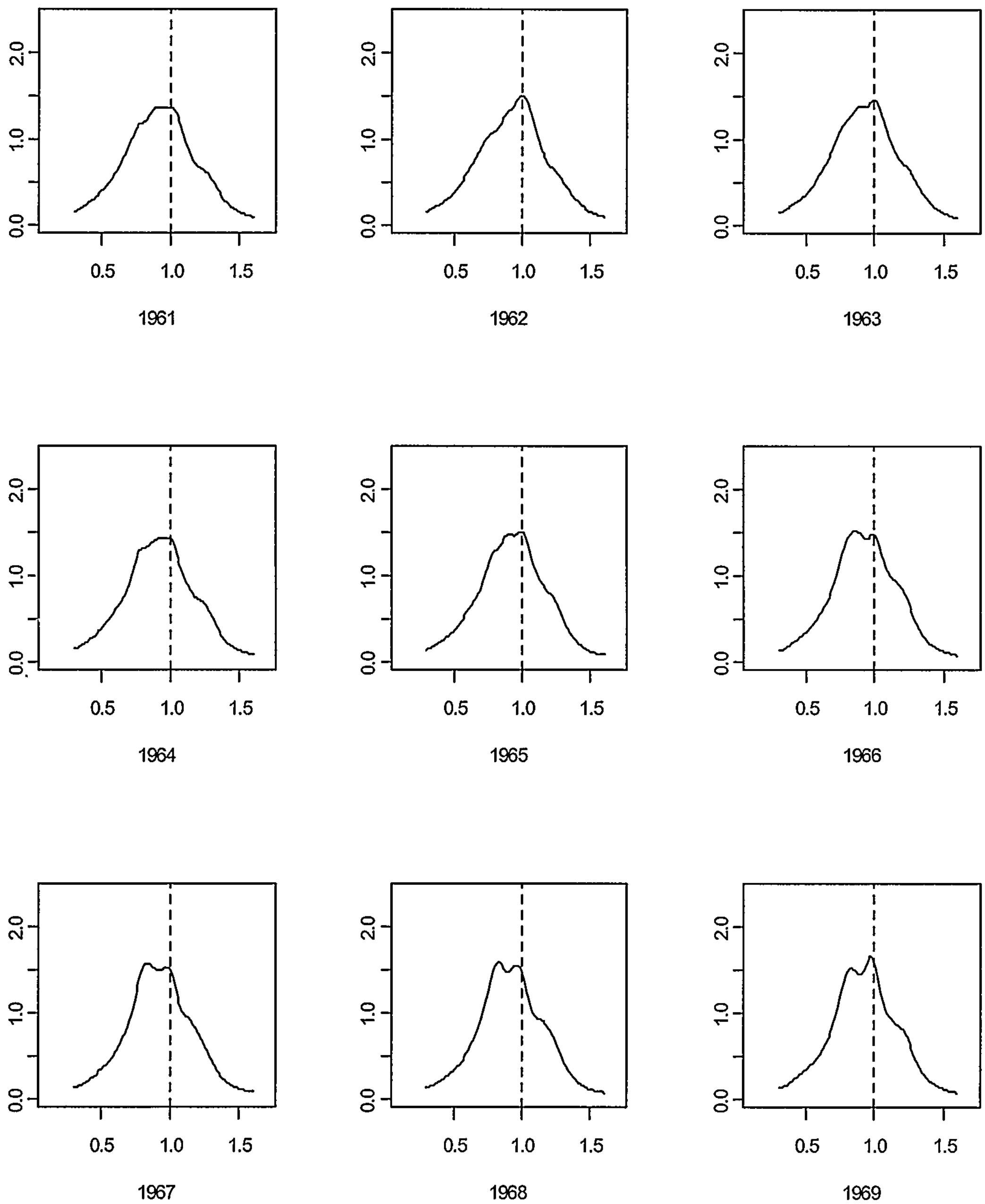


Figure A3

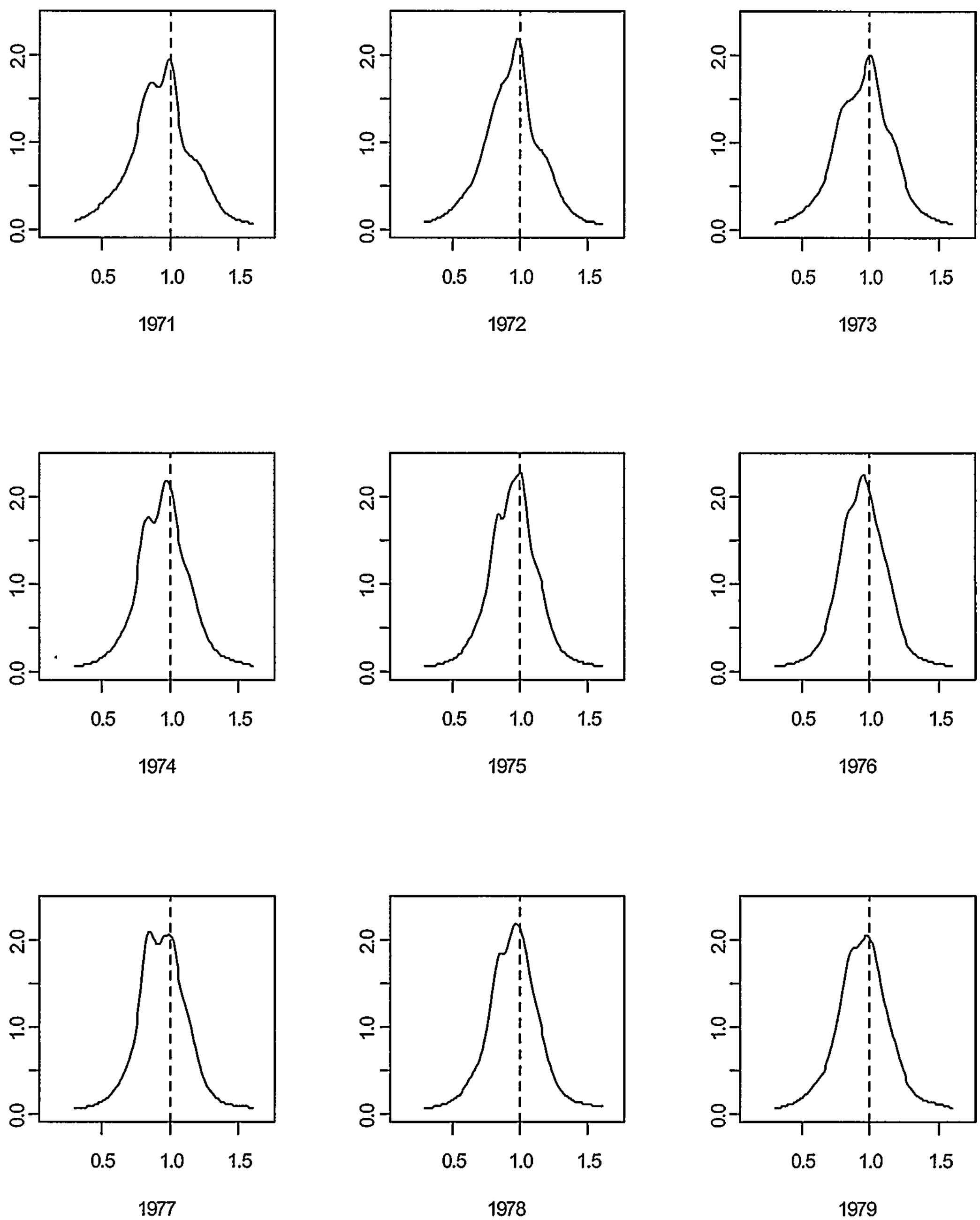


Figure A4

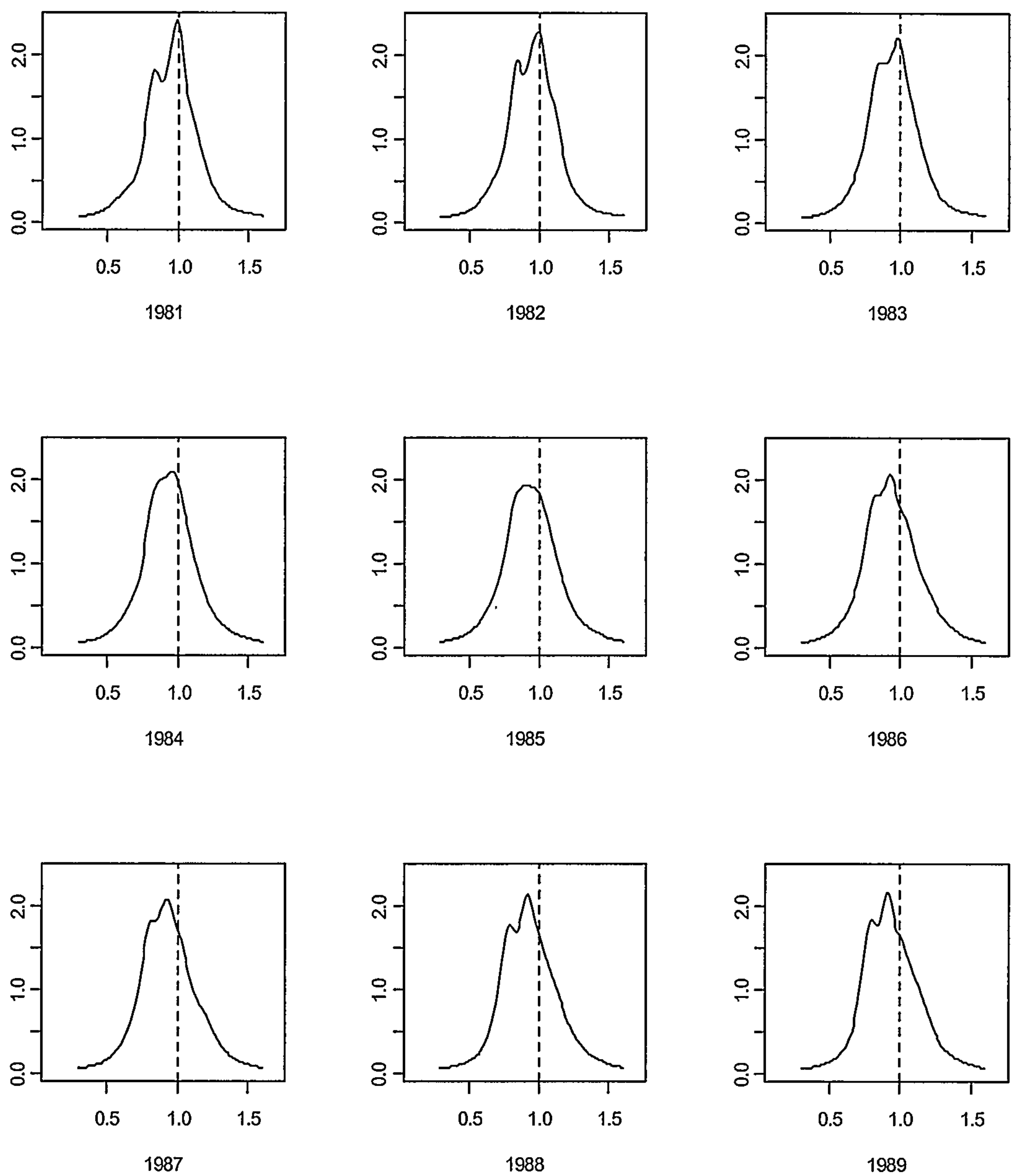
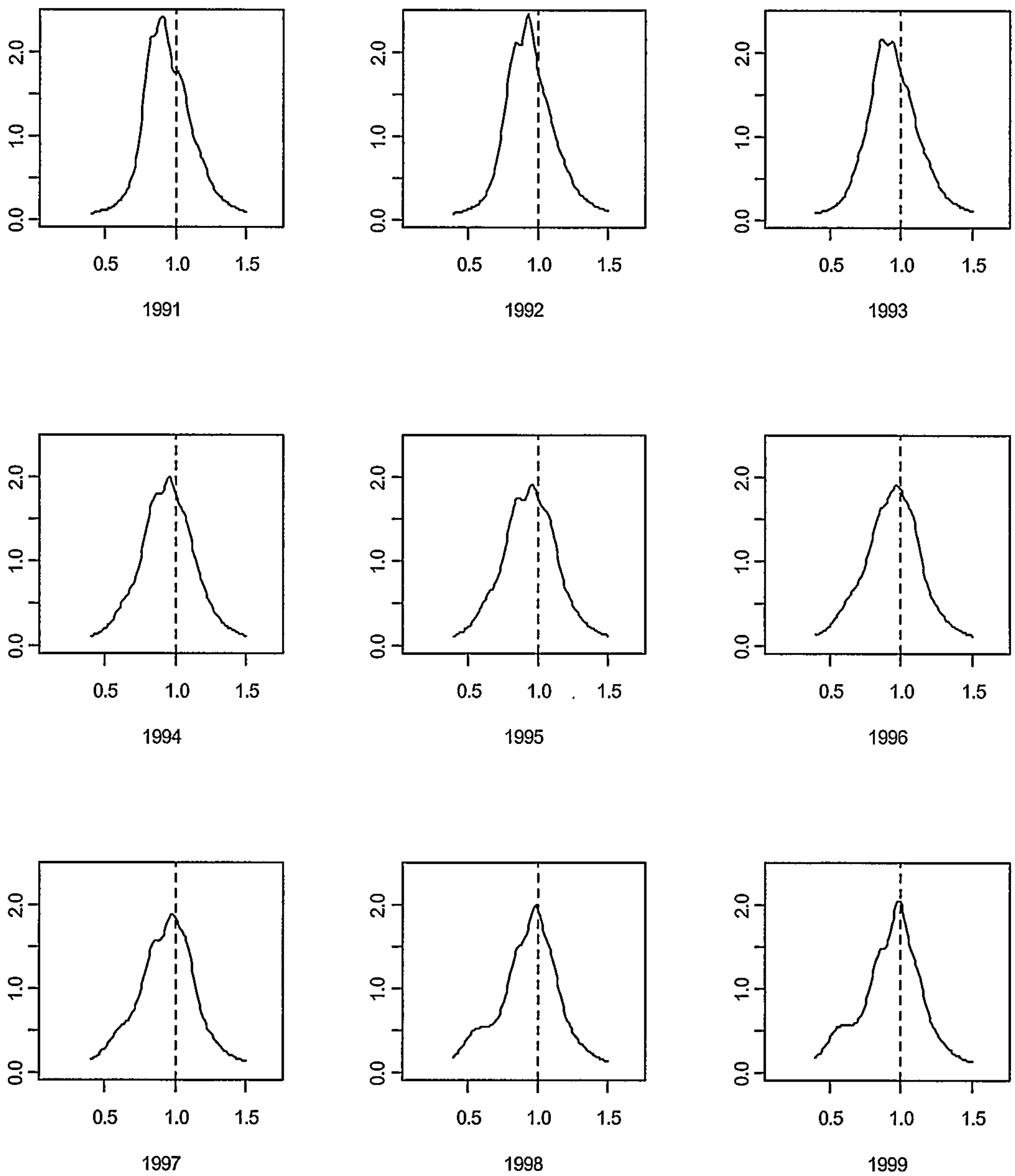


Figure A5



Computer Code

>> <u>**Silverman Test**</u>

```
---code starts here----

# This set of S-Plus functions is for testing significance
# for multimodality
# using the algorithm described by Efron & Tibshirani 1993.
# It is still under development; please report bugs and
# suggestions for improvement to Aaron Ellison
#(aellison@mtholyoke.edu).
# Written for S-Plus version 3.x, 4.x, 2000 for Windows
# Copyright (c) 1997, 2000 by Aaron M. Ellison

bootmode <- function(y, nm, nboot)
{
# Computes Achieved Significance Level of bootstrap test
# statistic for multimodality
# Algorithm 16.3 of Efron & Tibshirani 1993 (An
# introduction to the bootstrap)
# Requires as input data vector y, number of modes to test
# for nm, and number of bootstraps nboot
# Calls function critmode to calculate critical value of
# window size hk
# Calls function smoothboot to implement equation 16.22 of
# Efron & Tibshirani
# Calls function nummodes, a modification of Tibshirani's
# nmodes, that returns only the number of modes

ystar <- matrix(sample(y, size = length(y) * nboot, replace
= T),
nrow = nboot)
        hk <- critmode(y, nm)
        xstar <- smoothboot(y, ystar, hk)
        modelist <- apply(xstar, 1, nummodes, hk)
        omode <- nummodes(y, hk)
        asl <- sum(modelist - omode)/nboot
        return(asl)
}

critmode <- function(y, modenum)
{
#calculates critical window size hk for a given data vector
# y and requested number of modes modenum.  Based on
# windows of .5 to .01 of data width
```

```
#
        start <- NULL
        imodes <- NULL
        for(i in 1:99) {
        start <- c(start, ((max(y) - min(y))/(i + 1)))
        imodes <- c(imodes, nummodes(y, start[i]))
        }
#start is set of window values for nummodes
#imodes is set of modes returned from nummodes for each
#value of start
#
   hk <- min(start[imodes == modenum])
   return(hk)
}

smoothboot <-function(y, ystar, hk)
{
# does a smoothed bootstrap, a la Efron & Tibshirani 1993,
# eqn. 16.22 requires as input data vector y, bootstrapped
# matrix ystar, and window size hk
#
        vardata <- var(y)
        lendata <- length(y)
        xstar <- mean(ystar) + sqrt(1 + hk^2/vardata) *
(ystar - mean(ystar) + hk * rnorm(lendata))
        return(xstar)
}

nummodes <- function(y, h, w = rep(1, length(y)))
{
# finds number of modes- Bradley Efron's method
# modified from code (function nmodes) supplied by Rob
# Tibshirini to return only number of modes and to
# accommodate generic histogram range
# calls function kdens takes as arguments data vector y and
# window size h
#
        a <- min(y)
        b <- (max(y) - min(y))/40
        xx <- a + (0:40) * b
        temp <- kdens(y, h, xx = xx, w = w)
        junk <- temp$y
        xx <- temp$x
        n <- length(junk)
        mcount <- 0
        pos <- NULL
        for(i in 1:39) {
```

```
        if((junk[i + 1] > junk[i]) & (junk[i + 1] > junk[i
+
2])) {
         mcount <- mcount + 1
         pos <- c(pos, xx[i])
                 }
        }
        return(mcount)
}

kdens <- function(x, h, w = rep(1, length(x)), xx = NULL)
{
# returns gaussian kernel at np points
# code written by Rob Tibshirani
#
       if(!is.null(xx)) {
        np <- length(xx)
        }
        if(is.null(xx)) {
        np <- 100
        xx <- seq(min(x) - 3 * h, max(x) + 3 * h, length =
np)
        }
        y <- rep(0, length(xx))
        n <- length(x)
        for(i in 1:np) {
        y[i] <- (1/(n * h)) * sum(w * dnorm((xx[i] - x)/h))
        }
        return(x = xx, y)
}

---code ends here---
```

The following **S-Plus\R** procedure is used to do the Silverman test in thesis.

```
## S-plus is a commercial statistical software while R
## statistic software is available at
## http://cran.stat.ucla.edu/
## Copyright © Yibing Wang (2003)
## Contact wang_yibing@yahoo.com for permission to use this
## program
## Program starts here

nbmodes <- function (x, k)
## Function to calculate the number of modes
##
 {
     z1=seq(min(x),max(x),length=2*length(x))
     modes <- akj(x,z1,h=k)
     ## Calculates the adaptive kernel
     ## Use package quantreg available at
     ## http://cran.stat.ucla.edu/
     modes <- diff (diff (modes$dens) / abs (diff
(modes$dens)))
     modes <- rep(1, length(modes))[modes==-2]
     modes <- sum (modes)
     return (modes)
 }

hcrit <- function (x, n, e=.0001)
## Function to calculate the critical bandwidth
 {
   minb <- min (abs (diff (x)))
   maxb <- (max (x) - min (x))/2
   hb <- maxb
   lb <- minb
   while (abs(hb-lb)>e)
     {
       modes <- nbmodes (x, lb)
       lb <- hb
       if (modes > n)
         {
           min <- hb
           hb <- (hb + maxb)/2
         }
       else
         {
           maxb <- hb
           hb <- (hb - minb)/2
         }
```

```
      }
return (hb)
 }
smoothboot <-function(y, ystar, hk)
## Function to make smoothboot
## Use Aaron Ellison's methods
 {
         vardata <- var(y)
         lendata <- length(y)
         xstar <- mean(ystar) + sqrt(1 + hk^2/vardata) *
(ystar - mean(ystar) + hk * rnorm(lendata))
         return(xstar)
 }

bootmode <- function(y, n, nboot)
## Function to calculate ASL value
## Use Aaron Ellison's methods
 {
         ystar <- matrix(sample(y, size = length(y) * nboot,
replace = T), nrow = nboot)
         hk <- hcrit(y, n, e=.0001)
         xstar <- smoothboot(y, ystar, hk)
         modelist <- apply(xstar, 1, nbmodes, hk)
         omode <- nbmodes(y, hk)
         asl <- sum(modelist > omode)/nboot
         ## Corrections are made here
         return(asl)
 }

## Program ends here
```

>> **<u>Stochastic Kernel 3-D Procedure</u>**

```
@The following GAUSS procedure, NPSK © Yibing Wang (2003),
estimates stochastic kernels@
@ Contact wang_yibing@yahoo.com for permission to use this
program@
@ This program is used to draw 3-D Figure 8 ~ 15@
@Program starts here@

library pgraph;
load data[119,2] = GoldenAge10.txt;
@Load data@
y = data[2:119,.];
yi = reshape(y, rows(y), 2);
y1=seqa(0.3,0.03,45);
@ Create grids@
```

```
y2=seqa(0.3,0.03,45);
h=0.085|0.067;
@ Bandwidth@
p1=pdfn((y1-yi[.,1]')./h[1]);
p2=pdfn((y2-yi[.,2]')./h[2]);
f21=(p1*p2')./sumc(p1');
@ Calculates conditional probability@
@ Gratitude is given to Mark Trede for his advice on how to
calculate f21@
f22=f21/0.067;
ndpclex;
graphset;
_pdate="";
_paxht=0.2;
_pnumht=0.15;
_pframe={0,0};
xlabel("Relative Income in t");
ylabel("Relative Income in t+6");
zlabel("Stochastic Kernel")
surface(y1',y2,f22');
@ Draw a 3-D figure@
end;

@ Program ends here@
```

>> **<u>Stochastic Kernel Contour Procedure</u>**

```
@The following GAUSS procedure, NPCT © Yibing Wang (2003),
estimates contours of stochastic kernels@
@Contact wang_yibing@yahoo.com for permission to use this
program@
@This program is used to draw the corresponding contours in
Figure 8 ~ 15@
@Program starts here@

library pgraph;
load data[119,2] = GoldenAge10.txt;
@Load data@
y = data[2:119,.];
yi = reshape(y, rows(y), 2);
y1=seqa(0.3,0.03,45);
y2=seqa(0.3,0.03,45);
h=0.085|0.067;
p1=pdfn((y1-yi[.,1]')./h[1]);
p2=pdfn((y2-yi[.,2]')./h[2]);
f21=(p1*p2')./sumc(p1');
f22=f21/0.067;
```

```
ndpclex;
graphset;
begwind;
makewind(6.5,6.5,1.3,0.1,1);
@Create windows@
makewind(6.5,6.5,1.3,0.1,1);
setwind(1);
_pdate="";
_paxht=0.2;
_pnumht=0.15;
ztics(1,4,0.5,0);
xlabel("Relative Income in t");
ylabel("Relative Income in t+6");
contour(y1',y2,f22');
@Draw contours@
setwind(2);
_pltype={6};
xy(y1,y1);
@Draw diagonal@
endwind;

@Program ends here@
```

>> **Markov Transition Probability Matrix Procedure**

The following procedure estimates Markov transition probability matrix by using **tsrf** (available at http://econ.lse.ac.uk/~dquah/tsrf.html) from D. Quah.

```
**Program starts here
tsrf[1]->> dataFile -n suc4rr.dtl ;
** Read file name
tsrf[2]->> calendar -f 1 ;
** Calendar the data
tsrf[3]->> crsuSmpl 1 to 59 ;
** Cross-section dimension of the data
tsrf[4]->> timeSmpl 1950:1 to 2000:1 ;
** Time dimension of the data
tsrf[5]->> readData -rf -k 1 -n th -names -verbose ;
** Load data
tsrf[6]->> timeSmpl 1951:1 to 1999:1 ;
** Time sample
tsrf[7]->> transProb -rf -stationary -n th -nStates 4 -lag 1 -gridSpecify (0, 0.5, 0.7, 0.9,
1.1, 1.5)
```

>> **Adaptive Kernel Density Procedure**

```
## The following R procedure is used to draw four kernel
diagrams within a panel.
## Program starts here.

z = read.csv("C:\\Temp\\USCAmobilityComplete.csv", header =
TRUE, sep = ",", quote="\"", dec=".", fill = TRUE)
## Load data
attach(z)
x1 = sort(y1950)
x2 = sort(y1970)
x3 = sort(y1992)
x4 = sort(y2000)
b1 = seq(0.4,1.6,length=2*length(x1))
## Creates grids
b2 = seq(0.4,1.6,length=2*length(x2))
b3 = seq(0.4,1.6,length=2*length(x3))
b4 = seq(0.4,1.6,length=2*length(x4))
library(quantreg)
## Package quantreg is available at
## http://cran.stat.ucla.edu/
xd1 <- akj(x1,b1)
xd2 <- akj(x2,b2)
xd3 <- akj(x3,b3)
xd4 <- akj(x4,b4)
par(mfrow = c(2,2))
plot(b1,xd1$dens,type="l", xlim=range(min=0.3, max=1.7),
ylim=range(min=0, max=2.4), xlab=" ", ylab=" ", main=" ")
lines(b2, xd2$dens, lty=5)
legend(1.2, 2.3, c("1950", "1970"), lty=c(1,5))
plot(b2,xd2$dens,type="l", xlim=range(min=0.3, max=1.7),
ylim=range(min=0, max=2.4), xlab=" ", ylab=" ", main=" ")
lines(b3, xd3$dens, lty=5)
legend(1.2, 2.3, c("1970", "1992"), lty=c(1,5))
plot(b3,xd3$dens,type="l", xlim=range(min=0.3, max=1.7),
ylim=range(min=0, max=2.4), xlab=" ", ylab=" ", main=" ")
lines(b4, xd4$dens, lty=5)
legend(1.2, 2.3, c("1992", "2000"), lty=c(1,5))
plot(b1,xd1$dens,type="l", xlim=range(min=0.3, max=1.7),
ylim=range(min=0, max=2.4), xlab=" ", ylab=" ", main=" ")
lines(b4, xd4$dens, lty=5)
legend(1.2, 2.3, c("1950", "2000"), lty=c(1,5))

## Program ends here.
```

References

Abramson, I.S. (1982), "*On bandwidth Variation in Kernel Estimates – A Square Root Law*", Annals of Statistics, 10, 1217-1223

Abramovitz, Moses. (1986), "Catching Up, Forging Ahead, and Falling Behind." *Journal of Economic History*, 46:2, 385-406.

Barro, R. (1991), "Economic Growth in a Cross Section of Countries", *The Quarterly Journal of Economics*, 106 (2): 407-443

Barro, R. and Sala-i-Martín, X. (1991), "Convergence Across States And Regions", *Brookings Papers on Economics Activity*, 1991 (1): 107-182.

Barro, R. and Sala-i-Martín, X. (1992), "Convergence", *Journal of Political Economy*, 100 (2): 223-251

Barro, R. and Sala-i-Martín, X. (1995), Economic Growth. New York: McGraw-Hill, 1995

Baumol, W. (1986), "Productivity Growth, Convergence and Welfare. What Long-Run Data Show", *American Economic Review*, 76(5): 1072-85, December.

Bernard, A. B., and S.N. Durlauf (1995), "Convergence in international output", *Journal of Applied Econometrics*, 10 (2): 97-108

Bernard, A. B., and S.N. Durlauf (1996), "Interpreting tests of the convergence hypothesis", *Journal of Econometrics*, 71 (1-2): 161-174

Bianchi, M. (1997), "Testing for Convergence: Evidence from Nonparametric Multimodality Tests," *Journal of Applied Econometrics*, Vol. 12, pp. 393-409.

Breiman, L., W. Meisel, and E. Purcell (1977), "Variable Kernel Estimates of Multivariate Densities," *Technometrics*, 19, 135-144

Brown, S.J., N.E. Coulson, and R.F. Engle (1990), "Non-cointegration and econometric evaluation of models of regional shift and share", Working paper 3291 (NBER, Cambridgae, MA).

Carlino, G.A. and L.O. Mills (1993), "Are U.S. regional incomes converging? A time series analysis", *Journal of Monetary Economics* 32: 335-346

Chamberlain, G (1984), "Comments to 'Adaptive Estimation of Non-Linear Regression Models'," *Econometric Reviews*, 3, 199-202

Columbe, S. and Lee, F. C. (1995), "Convergence across Canadian provinces, 1961-1991", *Canadian Journal of Economics*, 4a, November, 1995

Davison, A.C. and Hinkley, D.V. (1997), Bootstrap Methods and Their Application, Cambridge University Press

Devroye, L.P. (1983), "The Equivalence of Weak, Strong and Complete Convergence in L1 for Kernel Density Estimates," *Annals of Statistics*, 11, 896-904

Devroye, L.P. and L. Gyorfi (1985), Nonparametric Density Estimation, New York, Wiley

Dowrick, Steve and Nguyen, Duc Tho (1989), "OECD comparative economic growth 1950-85: catch-up and convergence", *American Economic Review*, 79 (5): 1010-30

Durlauf, S.N. (1989), "Output persistence, economic structure, and the choice of stabilization policy", *Brookings Papers on Economic Activity*, 1989 (2): 69-116

Durlauf, S.N. (1996), "A Theory of persistent income inequality", *Journal of Economic Growth*, 1 (1): 75-93

Durlauf, S. N. and Johnson, P. A. (1995), "Multiple Regimes and Cross-Country Growth Behaviour," *Journal of Applied Econometrics*, Vol. 10, pp. 365-384.

Durlauf, S. N. and Quah, D. T. (1999), "The New Empirics of Economic Growth," *Handbook of Macroeconomics*, Vol. 1A, Amsterdam: Elsevier Science.

Easterlin, Richard A., (1960) Regional Growth in Income: Long Term Tendencies." Population Redistribution and Economic Growth: United States, 1870-1950, Vol. 2, 140-203, The American Philosophical Society

Efron, Bradley and Tibshirani, Robert J, (1993), An introduction to the Bootstrap, 227-233, Chapman & Hall

Fox, J. (1990), "Describing Univariate Distributions." Modern Methods of Data Analysis, pp. 58-125, Sage Publications

Hardle, W. and Linton, O. (1994), "Applied Nonparametric Methods," in R. F. Engle and D. L. McFadden, eds., Handbook of Econometrics, Vol. IV, pp. 2295-2339, Amsterdam: Elsevier Science.

Hardle, W. (1990), Applied Nonparametric Regression, New York, Cambridge University Press

Hartigan, J. A. and Hartigan, P. M. (1985), "The Dip Test of Unimodality," *Annals of Statistics*, Vol. 13:1, pp. 70-84.

Hartigan, P. M. (1985), "Computation of the Dip Statistic to Test for Unimodality," *Applied Statistics*, pp. 320-325.

Islam, N. (1995), "Growth empirics: a panel data approach", *Quarterly Journal of Economics*, 110 (443): 1127-1170

Jones, C. I. (1997), "On the Evolution of the World Income Distribution," *Journal of Economic Perspectives*, Vol. 11:3, pp. 19-36.

Loewy, M. B. and Papell, D.H. (1996), " Are U.S. regional incomes converging? Some further evidence", *Journal of Monetary Economics*, 38: 587-598

Mankiw, N.G., D. Romer and D.N. Weil (1992), "A contribution to the empirics of economic growth", *Quarterly Journal of Economics*, 107(2): 407-437

Mitchener, K. J. and W.Mclean. I. (1999), "U.S. Regional Growth And convergence, "1880-1980, *The Journal of Economic History*

Majumder, A. and Chakravarty, S. R. (1990), "Distribution of Personal Income: Development of a New Model and Its Application to U.S. Income Data," *Journal of Applied Econometrics*, Vol. 5, pp. 189-196.

Nickell (1981) “Biases in dynamic models with fixed effects”, *Econometrica*, 49 (6)

Pagan, A. and Ullah, A. (1999), Nonparametric Econometrics, Cambridge: Cambridge University Press.

Parzen, E (1962), “On Estimation of a Probability Density and Mode”, *Annals of Mathematical Statistics*, 33, 1065-1076

Quah, D. T. (1992), “International patterns of growth: I. Persistence in cross-country disparities”, Working Paper, LSE, London, October

Quah, D.T. (1993), “Empirical cross-section dynamics in economic growth”, *European Economic Review*, 37(2/3): 426-434

Quah, D. T. (1996a), “Convergence Empirics Across Economies with (Some) Capital Mobility”, *Journal of Economic Growth*, 1 (1):95-124

Quah, D. T. (1996), “Empirics for economic growth and convergence”, *European Economic Review*, 40 (6): 1353-1375

Quah, D. T. (1997), “Empirics for Growth and Distribution: Strati-fication, Polarization, and Convergence Clubs,” *Journal of Economic Growth*, Vol. 2, pp. 27-59.

Rosenblatt, M (1956), “Remarks on Some Nonparametric Estimates of a Density Function”, *Annals of Mathematical Statistics*, 27, 832-837

Sala-i-Martin, Xavier (1990), “On Growth and States” Ph.D Dissertation, Harvad University

Sala-i-Martin, Xavier (1994). "Regional Cohesion: Evidence and Theories of Regional Growth and Convergence," *Economic Growth Center Discussion Paper*, 715:46

Sala-i-Martin, Xavier (1996), “The Classical Approach to Convergence Analysis,” *The Economic Journal*, 106, 1019-1036

Silverman, B. W. (1981), “Using Kernel Density Estimates to Investigate Multimodality,” *Journal of the Royal Statistical Society*, Series B., Vol. 43:1, pp. 97-99.

Silverman, B.W. (1986), Density Estimation for Statistics and Data Analysis, New York, Chapman and Hall

Silverman, B.W. (1983), "Some Properties of a Test for Multimodality Based on Kernel Density Estimates," Probability, Statistics and Analysis, Cambridge University Press

St. Aubyn, M. (1999), "Convergence across industrialized countries (1890-1989): New results using time series methods", *Empirical Economics*, 24, 23-44

Tukey, J.W. (1977), Exploratory Data Analysis, Reading, Massachusetts, Addision Wesley

Wakerly, Elizabeth C. (2002), "Disaggregate dynamics and economic growth in Canada," *Economic Modelling*, 19, 197-219

Zhu, Feng (2002), "A Nonparametric Analysis of the Shape Dynamics of the U.S. Personal Income Distribution: 1962 – 2000", *Econometrics Society Winter European Meeting*, http://www.ceu.hu\eswm