BANK FOR INTERNATIONAL SETTLEMENTS

BIS Working Papers

No 184

A nonparametric analysis of the shape dynamics of the US personal income distribution: 1962-2000

By Feng Zhu

Monetary and Economic Department

October 2005

Copies of publications are available from:

Bank for International Settlements
Press & Communications
CH-4002 Basel, Switzerland

E-mail: publications@bis.org

Fax: +41 61 280 9100 and +41 61 280 8100

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)

ISSN 1682-7678 (online)

Abstract

We provide stylized facts on the evolving shape dynamics of the US personal income distribution from 1962 to 2000. Based on adaptive kernel density estimation, we propose an adaptive bootstrap test for multimodality. Our results indicate that multimodality has been a predominant feature of the US income distribution. Both the number and location of modes change over time, revealing rich distributional dynamics and strong heterogeneity. Decomposing the sample by age, education, gender and race, all groups exhibit multiple modes and follow distinct distributional patterns. For all groups and for the population as a whole, income distribution improved over the 1960s, 1970s and 1980s, but deteriorated dramatically in the 1990s.

Keywords: Adaptive kernel method, adaptive multimodality test, bootstrap, income distribution, nonparametric density estimation, population heterogeneity.

JEL Classification: C12, C14, C15, E25

1 Introduction[1]

Using an impressive array of inequality, mobility and polarisation measures, recent studies indicate a clear change in the trends of income distributions in developed economies. For the last five decades, wages and earnings grew rapidly in the 1950s and 1960s and inequality was reduced slightly. Beginning in the 1970s, inequality rose rapidly whereas wages grew at a slow pace. Since wages and earnings constitute the lion's share of personal income for the vast majority of the population, these changes in labour market conditions were quickly reflected in changes in personal and family income distributions. There is substantial diversity in the behaviour of income inequality across developed countries. A "stylised fact" for the US economy seems to emerge, suggesting a period of fast economic growth and decreasing inequality in the 1960s and 1970s, but slow growth in income and enlarging income inequality in the 1980s and 1990s.[2]

This paper builds on the original work of Bianchi (1997), Jones (1997) and Quah (1996a, 1996b, 1997). They used nonparametric methods to test the hypotheses of growth convergence and formation of convergence clubs, examining changes in the global, external shape of the cross-country income distribution and in intradistributional interaction patterns. Their major finding was the emergence of "twin peaks" (or bimodality) in the world income distribution, implying divergence and polarisation.[3] Using a regression tree clustering algorithm, Durlauf and Johnson (1995) run cross-section regressions that were allowed to adapt to subsamples. They were able to uncover "local basins of attraction" and identify countries which share a common model. This result was interpreted as the existence of multiple steady states or regimes in cross-country growth behaviour.[4]

Most previous studies of income distributions focussed on a number of summary statistics which, at best, only partially capture the characteristics of the probability distribution of income. Some authors tried to fit a parametric model to the data, only to find their models overwhelmingly rejected by data. In this paper, we take a nonparametric approach to examine the historical evolution of the shape and intradistributional dynamics of the US personal income distribution. In this framework, "income distribution" is perceived as the physical distribution of income across the population as well as the probability distribution of income.

We make use of adaptive kernel density estimates and propose an *adaptive* mode testing procedure for the empirical measurement of the *entire* income distribution. It is well-known that for kernel

[1] I thank Giuseppe Moscarini and Peter Phillips for helpful discussions and detailed comments. Comments from Steve Arthur, Richard Blundell, Claudio Borio, Steve Durlauf, Andy Filardo, Gabriele Galati, John Hartigan, Helmut Lütkepohl, Herman van Dijk, Chris Sims and Zhijie Xiao also brought improvement. In addition, I thank participants of sessions at the 2002 European Economic Association Meeting, the 2002 European Winter Meeting and the 2003 North American Summer Meeting of the Econometric Society, and of seminars at the Bank for International Settlements, Princeton University, Universidade Nova de Lisboa and Yale University for helpful comments. Financial support from the Portuguese Foundation for Science and Technology and the Cowles Foundation for Research in Economics is gratefully acknowledged. The views expressed here are those of the author and do not necessarily reflect those of the BIS. E-mail: feng.zhu@bis.org.

[2] Much of this literature has been summarised in the review articles of Levy and Murnane (1992), Gottschalk (1997) and Gottschalk and Smeeding (1997). Not all economists agree on the "stylised fact", see, for example, Wolfson (1997).

[3] The emerging "twin peaks" property corresponds to the gradual clustering of the very rich economies and the poor into two separate groups, with the middle-income class slowly disappearing.

[4] This literature is summarised in Durlauf and Quah (1999).

density estimates, the choice of the bandwidth h determines the amount of smoothing. When applied to long-tailed distributions, conventional fixed-bandwidth kernel method suffers from the drawback that h is constant across the entire sample. In our adaptive version, we compute a local bandwidth parameter γ_{ni}, which allows us to adapt the smoothing parameter to the local density of the data. Adaptive kernel density estimation makes it possible to purge spurious noises in the tails while preserving important characteristics in the central part of distribution.

Based on adaptive density estimates, we propose an *adaptive* version of Silverman's (1981) bootstrap test for multimodality. Compared with other mode testing procedures, the Silverman test is less conservative and easy to implement. However, conventional fixed-bandwidth versions of the bootstrap test focus on one single bandwidth h that is constant over the whole density range. In particular, when h is small, there is a tendency for spurious modes to appear in the tails of the estimates. On the other hand, if one increases h to eliminate these modes, over-smoothing becomes inevitable and true modes in the main part of distribution may be sacrificed. Our adaptive bootstrap multimodality test is based on an overall smoothing parameter h, while the local bandwidth parameter γ_{ni} adopts the density estimates to local details of data. Adaptive bandwidth tests are more accurate.

Mode testing is useful for assessing the degree of population heterogeneity, which enables us to uncover distinct income distribution dynamics for different population subgroups. Important hypotheses such as "income equalisation and convergence", polarisation, poverty traps and "vanishing middle class" can be examined visually by scrutinising changes in income density estimates over time. Our comparative statics approach focusses on the evolution of the shape of kernel density estimates over time, and it can be easily extended to study the world income distribution, and in fact any other size distribution.

From our work, some clear empirical regularities emerge. Over the last four decades, multimodality has been a predominant feature of the US income distribution. Both the number and location of modes change over time, revealing important distributional dynamics and strong population heterogeneity. Over the 1960s, 1970s and 1980s, we observe a general trend of diminishing low-end spike to the benefit of the middle class, suggesting that the hypotheses of poverty trap and polarisation cannot be validated. However, the trend was reversed in the 1990s, the low-end spike re-emerged and a large amount of population mass moved from the middle and upper-middle relative income ranges to the lower-middle range. The observed low-quality convergence might not be desirable, with the middle class potentially disappearing into the lower range.

In order to examine the apparent heterogeneity in the income distribution dynamics of population subgroups with distinct characteristics, we further decompose the sample by age, education, gender and race and carry out the adaptive kernel estimation and mode testing procedures for each subgroup. Interestingly, all subgroups exhibit multiple modes and follow distinct distributional patterns. This extra layer of distributional complexity presents a challenge to theorists and empirical researchers alike.

In Section II, we provide a detailed account of our methodology. Section III describes our income data and the data transformation process. The empirical results are reported in Section IV, and Section V concludes and discusses possible extensions.

2 Methodology

2.1 Empirical measurement of income distribution

Measures of income inequality can be largely classified into two categories: fractile methods which are based either on the fraction of income belonging to a particular fractile or population range or on the fraction of population pertaining to a certain income class; and scalar inequality indicators which summarise information in a single statistic. Summary inequality measures could provide, at best, only incomplete and sometimes misleading information on the shape and dynamics of income distribution.

To better understand income distributions in the whole income range, parametric models are often fitted to income data. Unfortunately, estimation methods and test statistics are quite sensitive to outliers and arbitrary truncation. McDonald and Ransom (1979) compared the goodness-of-fit for the lognormal, gamma, beta and Singh-Maddala functions using different estimation techniques, namely, least squares, method of scoring and minimum chi-square. Their results suggested that estimates of population characteristics (eg, mean income, Gini coefficient) depend on both the functional specification and the adopted method of estimation. McDonald and Mantrala (1995) showed that different forms of data sets and alternative estimation techniques are responsible for apparently contradictory results and inconsistency of the empirical findings.

More importantly, almost all specifications possess a unimodal density function, ignoring the possibility that the income distributions may exhibit multimodality and may be better modelled as a mixture of two or more component densities. Indeed this may be the reason behind the observation that most parametric specifications usually fit well for only part of the income distribution but not the full range. There are neither intrinsic reasons nor sound theoretical justifications for choosing any particular functional form. Even when a model is claimed to provide an excellent fit to data, it may not be straightforward to find reasonable economic interpretations for the estimated parameters.

Like the parametric approach, nonparametric density estimation techniques allow us to provide full information on the *entire* income distribution. The nonparametric approach has the advantage of letting the data speak for itself, freeing us from the arbitrariness of parametric specifications. Minimal assumptions are imposed on the data, namely that the income density exists and satisfies some smoothness properties. Instead of taking a stand on unimodality from the start, the nonparametric approach enables us to explicitly test for the number of modes present in the underlying income distribution. Multimodality, if confirmed, points to population heterogeneity, implying the existence of population groups that follow distinct distributional dynamics over time and are sorted into different steady states. By conditioning on certain population characteristics (age, education, gender and race), we may explain the apparent population heterogeneity and sorting.

The nonparametric approach avoids many theoretical difficulties and empirical fragility encountered in traditional measures of income distribution. We directly estimate the empirical frequency distribution from data and address the problem of population heterogeneity by testing the number of component densities using an adaptive version of Silverman's (1981, 1983) multimodality test. Estimates of density and number of modes are provided for different age, gender, racial and educational groups, factors potentially underlying observed income heterogeneity.

2.2 Adaptive kernel density estimation

The original fixed-bandwidth kernel density estimator was introduced by Rosenblatt (1956) and Parzen (1962). It is flexible and robust.[5] In the spirit of "letting the data speak for itself", minimal assumptions are imposed upon the data and the underlying true density function. Compared to other nonparametric techniques, such as the nearest neighbour, variable bandwidth, orthogonal series and penalised likelihood methods, kernel estimators enjoy good theoretical properties which are well understood. Under mild regularity conditions, kernel density estimates are asymptotically unbiased, consistent and asymptotically normal, although they converge at a slower than root-n rate.[6]

Fixed-bandwidth kernel density estimates were obtained by Bianchi (1997), Jones (1997) and Quah (1996a, 1997) to analyse world income distribution and growth dynamics. Analysis based on kernel density estimation is rare in the vast literature of personal or household income distribution dynamics. Champernowne (1974) intended to estimate "a large number of empirical frequency distributions of income as the basis of this study", but the lack of appropriate statistical apparatus and computational power forced him to adopt the less ambitious parametric approach, "using families of theoretical frequency distributions to represent the empirical distributions. Hildenbrand and Hildenbrand (1986) used maximum penalised likelihood method to obtain a nonparametric density estimate for UK income data. However, they took it as an intermediate estimate for the estimation of mean income effect.

Our use of adaptive kernel estimation and mode testing as a way to study distributional heterogeneity is novel. We adopt the adaptive kernel method for two reasons. First, ordinary kernel estimates with fixed bandwidth are known to be locally insensitive to data details. Empirical income distributions typically have long and thick tails. Income data are relatively sparse in the upper range and may be top-coded and aggregated in order to enhance confidentiality. Secondly, for mode testing, relative sparseness of data in the tails and profusion of data in the centre makes it easy for spurious modes to pop up in the tails well before true modes in the centre are detected. A local adaptive smoothing procedure is necessary, with the smoothing parameter allowed to vary with local data density.[7] Local smoothing not only provides an improved estimate of the underlying probability density, particularly when the tail structure is complicated, but it also reduces the risk of obtaining spurious modes with a multimodality test.

Suppose at each period t, we have a set of univariate observations $Y_1, \ldots, Y_n$, the adaptive kernel density estimate can be computed in the following two-step procedure, adapted from Breiman et al (1977).

[5] By *robustness* we mean the insensitivity of a density estimate to the removal or replacement of a small number of data points.

[6] For details, see Devroye (1987), Härdle and Linton (1994), Pagan and Ullah (1999), Robinson (1998) and Silverman (1986).

[7] Hildenbrand and Hildenbrand (1986) noticed the problem with fixed-bandwidth kernel methods: "The kernel estimator typically fluctuates and is unstable for small and large values in the range of net income since in that range there are not enough data points."

1. Obtain a pilot kernel estimate $\tilde{f}(y)$ of the income density $f(y)$ in such a way that $\tilde{f}(Y_i) > 0$, for all i:

$$\tilde{f}(y) = (nh_0)^{-1} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{h_0}\right) \tag{1}$$

where the initial bandwidth h_0 is chosen by minimising the approximate mean integrated square error (MISE)[8], with reference to the normal distribution with mean zero and variance equal to the sample variance $\hat{\sigma}^2$. Specifically, set $h_0 = 0.9An^{-1/5}$, where $A = \min\{\hat{\sigma}^2, \hat{R}/1.34\}$ is an adaptive estimate of spread and $\hat{R}$ is the empirical interquartile range of the underlying normal distribution.

The kernel $K(\cdot)$ is any symmetric function satisfying standard regularity conditions. We use the Gaussian kernel at all stages of estimation and testing,[9]

$$K(y) = (2\pi)^{-1/2} \exp\left(-y^2/2\right) \tag{2}$$

The kernel function is itself a legitimate density function. Our density estimate inherits all smoothness (continuity and differentiability) properties of the standard normal density.

2. Compute the adaptive kernel estimate $\hat{f}(y)$ as follows:

$$\hat{f}(y) = n^{-1} \sum_{i=1}^{n} h_i^{-1} K\left(\frac{y - Y_i}{h_i}\right) \tag{3}$$

where $h_i = h\gamma_{ni}$. h is the overall bandwidth or smoothing parameter and γ_{ni} is the local bandwidth parameter defined by

$$\gamma_{ni} = \left(\frac{\tilde{f}(Y_i)}{\bar{f}}\right)^{-\alpha} \tag{4}$$

Here $\bar{f}$ is the geometric mean of the pilot estimates $\tilde{f}(Y_i)$ at all data points, ie,

$$\bar{f} = \left(\prod_{i=1}^{n} \tilde{f}(Y_i)\right)^{1/n} \tag{5}$$

and α is a sensitivity parameter satisfying $0 \le \alpha \le 1$. In the fixed bandwidth case, $\alpha = 0$. Abramson (1982) argues that setting α to $1/2$ leads to bias reduction compared to the fixed-width kernel estimate.

For kernel density estimates, the choice of bandwidth h determines the amount of smoothing. When applied to long-tailed distributions, the fixed bandwidth method suffers from the drawback that h is constant across the entire sample. If h is too small, there is a tendency for spurious noise to appear in the tails of the estimates. On the other hand, if one increases h to eliminate the noise, oversmoothing is often inevitable and essential detail in the main part of distribution will be sacrificed.

---

[8]Breiman et al (1977) showed that the adaptive kernel density estimate was "surprisingly insensitive" to the choice of bandwidth in the pilot estimate $\tilde{f}$. In fact, the adaptive method is in general insensitive to the details of first-stage pilot estimation.

[9]The choice of kernel in the first stage makes little difference. The Gaussian kernel has unbounded support. To speed up computations, we use a modified Gaussian kernel truncating $K(y)$ at three standard deviations.

In the variable bandwidth case, the local bandwidth parameter γ_{ni} allows us to adapt the smoothing parameter to the local density of the data, relative to its geometric mean and properly rescaled. Adaptive kernel density estimation therefore makes it possible for us to purge spurious noise in the tails while preserving important characteristics in the fat part of the distribution. Bandwidth tests based on the adaptive method are more accurate.

2.3 Population heterogeneity and adaptive mode testing

Broadly speaking, there are two basic approaches to testing for homogeneity of a population. The first approach is embedded in finite mixture modelling, which has a long tradition in the statistics literature. It approximates the sampling distribution by a parametric mixture of unimodal distributions and then evaluates the goodness of fit. A finite mixture model for the income data can be written as:

$$p(y) = \sum_{i=1}^{m} \pi_i f_i(y) \qquad y \in \Re_+ \tag{6}$$

where $\pi_i > 0$, $\sum_{i=1}^{m} \pi_i = 1$, $f_i(y) \geq 0$, and $\int_{\Re_+} f_i(y)dy = 1$, for $i = 1, \ldots, m$. In this case, we say that the data Y is generated from a finite mixture distribution and $p(\cdot)$ is a finite mixture density function. The parameters $\pi_1, \ldots, \pi_m$ are known as the mixing proportions or mixing weights, and $f_1, \ldots, f_m$ are the component densities.

In some cases, it is possible to specify functional forms for the component densities. It is known that the log-normal distribution fits well the lower part, while the gamma distribution fits well the middle part and the Pareto distribution fits well the upper part of income distributions. One can fit a finite mixture of these three distributions to income data. We do not adopt the parametric approach, because we have little or no *a priori* information about the functional forms assumed by each component density f_i: on the one hand, economic theory provides little guidance about the specific composition of income distribution; on the other hand, inference for parametric models depends, to a large extent, on the validity of the assumed functional forms. In our case, the appropriate value for the fundamental parameter m must be estimated first, so one needs to detect and assess the presence of mixtures in the first place.

The nonparametric approach includes distribution-free methods of bump-hunting and mode testing. The most striking feature of a mixture density is often the presence of multiple bumps and modes, which can be assured if the component densities in a population are unimodal and are well-separated.[10] Heuristically, a *mode* in a density f is a local maximum, and a *bump* with no flat part is the portion of the density f lying between two points of inflection a and b, such that f is concave in the interval $[a, b]$ but not outside (see Figure 1). A density f is *multimodal* if it has more than one mode. Multimodality or the existence of multiple bumps is indicative of the existence of distinct

[10]We say two modes are *well-separated* if the distance in the means of the corresponding component distributions is large relative to their variances. Take the example of a mixture of two normal components ϕ_1 and ϕ_2, the first component being distributed as $N(\mu_1, \sigma^2)$ and the second as $N(\mu_2, \sigma^2)$. If the mixing weights are equal, the mixture of ϕ_1 and ϕ_2 will not be bimodal if $|\mu_1 - \mu_2| < 2\sigma$. In general, a mixture of distinct distributions does not necessarily give rise to the same number of modes as the number of mixture components. Furthermore, the location of modes in a mixture may not correspond to that of the component densities.

component densities, hence population heterogeneity.[11]

There are several known multimodality tests in the statistical literature: Good and Gaskins' (1980) iterative surgery procedure, Hartigan and Hartigan's (1985) dip test, Müller and Sawitzki's (1991) excess mass approach, and Silverman's (1981, 1983) bootstrap test. The dip and excess mass tests share difficulties of implementation and may lead to considerable conservatism. In fact, as Cheng and Hall (1998) pointed out, for each non-zero value of nominal level, the asymptotic levels of these tests are zero. Silverman's bootstrap test has a non-zero asymptotic level for each non-zero nominal level.

The original fixed-bandwidth Silverman test can be sensitive to spurious noise in the tails, as it inherits the difficulty of fixed-bandwidth kernel methods in obtaining fine distributional details in the centre before spurious modes appear in the tails. Without properly controlling the sensitivity to local data density, multimodality tests error-prone. In our adaptive version, the bandwidth h_i has two components, an overall bandwidth h, which controls the overall amount of data smoothing in the density estimate, and a local bandwidth factor γ_{ni} which varies according to the local data density. It is known that fixed-bandwidth kernel methods tend to produce spurious modes in the tails while masking essential details in the central part of size distributions. The single fixed critical bandwidth required to eliminate the spurious modes are determined more by tail behaviour than by the presence of true modes in the centre. Moreover, identification of a mode in any particular region of a density depends on the distance from the nearest mode, which compels us to use an adaptive kernel. In the adaptive version, we base our multimodality test on the overall bandwidth h, and use the local bandwidth factor γ_{ni} to prevent spurious modes from popping up too easily due to the possibly complicated tail structure.

Define the number of modes of density f as:

$$M(f) = \#\left\{y \in \Re_+ : f'(y) = 0 \text{ and } f''(y) < 0\right\} \tag{7}$$

For $m = 1, \dots, M$, we test the null hypothesis that the underlying density f has m modes (H_0 : $M(f) \leq m$), against the alternative that f has more than m modes ($H_1 : M(f) > m$). In our adaptive version, the statistic we use is $M(\hat{f}_h)$, where $\hat{f}$ is the kernel density estimate and h is the overall bandwidth.

The bootstrap multimodality test is based on the notions of critical smoothing and critical bandwidth. Define the m-th critical bandwidth $\hat{h}_{m,crit}$ by:

$$\hat{h}_{m,crit} = \inf\left\{h : M(\hat{f}_h) \leq m\right\} \tag{8}$$

where, for any density estimate $\hat{f}_h$, $\hat{h}_{m,crit}$ is the smallest overall bandwidth such that $\hat{f}_h$ has at most m modes. Consequently, $\hat{f}_{\hat{h}_{m,crit}}$ is known as the m-th critical density. The idea of critical smoothing is crucial. Intuitively, a large value of h leads to oversmoothing and therefore to a smaller number of modes. If the underlying density f has m modes, in order to eliminate one or two modes in $\hat{f}$, h

[11] Multimodality and mixtures are two different concepts: unimodality can mask the existence of distinct subpopulations, and bimodality may sometimes be generated by a single homogeneous population. Multimodality is a stronger concept in that it implies mixture.

needs to be larger than $\hat{h}_{m,crit}$. When $h = \hat{h}_{m,crit}$, $\hat{f}_h$ will display m modes with a shoulder.[12] When h is further decreased, an additional $(m+1)$-th mode will show up in place of the shoulder. The null H_0 will be rejected for large values of $\hat{h}_{m,crit}$. The multimodality test is based on the following result.

Theorem: *Given any fixed* $\{X_1, \ldots X_n\}$, *let* $\hat{f}_h$ *be a kernel density estimate of* f *with bandwidth* h *and the Gaussian kernel function* $K(\cdot)$ *as defined in (2). Then,* $M\left(\hat{f}_h\right)$ *is a right continuous non-increasing function of* h. *Moreover,* $M\left(\hat{f}_h\right) > m$ *if and only if* $h < \hat{h}_{m,crit}$.

Proof: *See Silverman (1981).*

Remark 1: Silverman (1981) noted that this result does not apply for general kernels. Moreover, the convolution of unimodal densities needs not be unimodal.

Remark 2: This result allows us to use either $\hat{h}_{m,crit}$ or $M\left(\hat{f}_h\right)$ as the test statistic. The null hypothesis can be written as $H_0 : h > \hat{h}_{m,crit}$ and the alternative as $H_0 : h \leq \hat{h}_{m,crit}$. If too large a value of h is required to eliminate the m-th mode, then we can accept the null that there are at least m modes in the true density f.

Following Silverman (1986) and Efron and Tibshirani (1993), we assess the significance of $\hat{h}_{m,crit}$ estimated from data against its bootstrap distribution. To simplify notation, we write h_{crit} and $\hat{h}_{m,crit}$ interchangeably. The bootstrap resampling procedure can be described as follows:

1. Given the original data $\{Y_1, \ldots, Y_n\}$, randomly draw (with replacement) B independent bootstrap samples $\left\{Y_1^b, \ldots, Y_n^b\right\}_{b=1}^B$ from the empirical distribution $\hat{F}_n$, each of size n;
2. Obtain B *smoothed* bootstrap samples $\left\{X_1^b, \ldots, X_n^b\right\}_{b=1}^B$ by computing

$$X_i^b = \bar{Y}^b + \left(Y_i^b - \bar{Y}^b + \hat{h}_{m,crit} e_i\right)\left(1 + \hat{h}_{m,crit}/\hat{\sigma}^2\right)^{-1/2} \tag{9}$$

   where e_i is an $i.i.d.$ $N(0,1)$ variable. The scaling factor $\left(1 + \hat{h}_{m,crit}/\hat{\sigma}^2\right)^{-1/2}$ in equation (9) ensures that the variance of the bootstrap samples X_i^b is the same as the sample variance $\hat{\sigma}^2$ of the original data;
3. For $b = 1, \ldots, B$, use the adaptive kernel method to obtain bootstrap estimates $\hat{f}^b_{h_{crit}}$ based on the critical bandwidth $\hat{h}_{m,crit}$ computed from the original data. For each bootstrap sample, calculate the corresponding number of modes $M\left(\hat{f}^b_{h_{crit}}\right)$ and the m-th critical bandwidth h^b_{crit};
4. Compute the estimate of the achieved significance level (ASL) or p-value:[13]

$$\widehat{ASL}_m = p_m = \frac{1}{B}\sum_{b=1}^{B}\{M\left(\hat{f}^b_{h_{crit}}\right) > m\} \tag{10}$$

[12] A *shoulder point* y can be defined by $f'(y) = 0$, $f''(y) = 0$, $f'''(y) \neq 0$.

[13] The achieved significance level or p-value is $ASL_m = \Pr\left\{M \quad \hat{f}_h \quad > m | Y_1, \ldots, Y_n \text{ is a random draw from } f\right\}$.

5. Do not reject the null of m modes in the underlying density if $\widehat{ASL}_m$ or p_m is large enough.

The test is carried out by simultaneously computing $\widehat{ASL}_m$ for all $m = 1, \ldots, M$, where M is a preset number. For the underlying density f, we select an estimate $\hat{m}$ as the critical number of modes for which a robust and sufficiently large ASL is obtained. Simultaneous testing is necessary because we suspect that a possibly long and complicated tail structure is present in our data, which may prevent the sequence of p-values from being monotonely increasing in m. In fact, large fluctuations in p-values are possible, depending on the location of modes. This may give rise to inconsistency but it also leaves us space for judgement. Bianchi's (1997) fixed-bandwidth test results also exhibit such inconsistency, but he stopped testing the number of modes at $m = 2$, which may not be appropriate for cross-country income data.

Silverman's original multimodality test is intuitive and easy to implement. Our adaptive and simultaneous version of the test is designed to cope with the relative sparseness of data in the tails and possible fluctuations in the estimated values of the ASL. Silverman (1983) acknowledged that his test might be conservative, in the sense that the bootstrap procedure may produce p-values larger than the true significance level, therefore underestimate the number of modes $M(f)$. This conservatism leads to reduced power for the bandwidth test. Indeed this is a general feature of many nonparametric mode testing and bump-hunting procedures, and Silverman's test seems to be the least conservative of all.

2.4 Heterogeneous Income Distribution Dynamics

Based on income inequality measures obtained from data, regressions are often run from these measures on potential explanatory variables such as growth rates, trade and openness, human capital, savings and socio-political factors. However, such regressions contain as much information as the conditional average of a distribution allows them to carry. Given that these variables are interdependent and homogeneous, OLS regressions are as problematic and uninformative as are summary statistics.

As an alternative to regression analysis, and to give a clear picture of the distinct income distribution dynamics present in different population subgroups, we directly condition our kernel income density estimates on both endogenous (education) and exogenous (age, gender and race) population characteristics, and examine possible changes in the shape of income distributions of population subgroups using adaptive density estimation and mode testing. Conditioning on such population characteristics provides a detailed analysis of distributional heterogeneity.

3 Empirical analysis

3.1 Data

3.1.1 Source

The personal income data are extracted from the March Income Supplement of the Current Population Surveys (CPS) 1962 - 2000, compiled by the US Census Bureau. The CPS are administered monthly by the Census Bureau to over 50,000 households. The surveys gather a wealth of information on the education, labour force status, demographics and other aspects of the US population. The March surveys contain the Annual Demographic File and the Income Supplement, which report the income-related aspects of US households in detail.

The main advantage of using CPS data is the fact that the samples are large, nationally representative and designed to reflect demographic changes, a blessing for our nonparametric approach, in which the samples are treated themselves as the population for density estimation. The CPS use complex multi-stage cluster design, and the income data are constructed in a hierarchical structure, going from households to families and then to persons. Data collection is not a random sampling process, and sampling variability may be large. This may result in inaccuracy and larger variances of inequality and mobility measures. However, Bound and Krueger's (1991) validation study of the "matched" CPS data indicates that the errors are limited to a tolerable range for statistical analysis. Compared to income data compiled in longitudinal surveys such as the Panel Study of Income Dynamics (PSID), less age-on-income effects are present because the CPS are a wide cross-section with only half of its sample rotated in a two-year interval.

As a potential drawback, income data are classified and grouped into small intervals, either because the questions allow the surveyed to classify income into different ranges, or because the surveyed rounds income into convenient integer numbers.[14] Rounding by the surveyed is apparent in the data: in histograms drawn using the raw data, large and distinctive spikes appear around almost all values which are multiples of 100 or 1000 US dollars. We deal with this problem by uniformly "redistributing" these possibly rounded values into a reasonably small neighbourhood of each centre of mass, which much alleviates the severity of the problem. The CPS sample sizes are quite large and the grids quite refined, which makes it possible to obtain an accurate estimate of the underlying true density without resorting to bounding or interpolation methods.

We focus on the total personal income for individuals aged between 25 and 54. This range includes the prime age for employment, at which both employment and the income level are becoming more stable. We use *total personal income*, because it not only reflects an individual's earning capacity but also measures his potential overall welfare. For the vast majority of the population, personal income is the main, if not the only source of current spending and living. It is sometimes argued that studies on mobility should be based on wages and earnings, and analyses of income inequality should be based on household or family data, since both the size and composition of a household or

[14] With grouped data, detailed information on the distribution within each interval $[x_i, x_{i+1})$, for $i = 1, \ldots, I$, is usually not available. Cowell (2000) pointed out that "this information structure will result in the loss of some important information and will complicate some standard statistical problems such as inference".

family may have a large impact on their members' general welfare. The emphasis of our paper on *both* the individual earning capacity and potential welfare makes the total personal income a more pertinent choice.

3.1.2 Data transformation

Let y_{it} denote the income level of individual i in a particular year t and let n_t be the total number of usable observations for that year. We normalise the data dividing them by the maximum income of each year:

$$Y_{it} = \frac{y_{it}}{\max_i\{y_{i_t}\}} \qquad \text{for } i = 1, \ldots, n_t \text{ and } t = 1, \ldots, T \tag{11}$$

Thus Y_{it} is the relative income of individual i with respect to the highest income level at t. Normalisation by maximum has been shown to have better properties than other forms of transformation, including those by dividing the whole set of data by the sum, mean, median or range. Being a purely scale transformation, the relative income transformation has the advantage of making the income data unit-free while not affecting the shape of the original empirical distribution. The transformation dispenses the need to use a price deflator to obtain real income data. It emphasises the fact that in income studies, an individual's relative position in the income distribution matters, besides the absolute income level. Zhu (2002) provides an asymptotic analysis of commonly used forms of normalisation. His results indicate that under possible multimodality, as the sample size goes to infinity, data normalised by maximum will follow the same distribution as the original data. In particular, the transformed data will be multimodal if the original data are.

Normalisation by maximum also facilitates temporal comparisons of income distributions. It is well-known that steadily increasing income levels along with an ever-expanding income range make such comparisons difficult.[15] To avoid the possibility that movements of a small number of high-income individuals influence the shape of overall distribution,[16] we trim out the top 1% of each year's sample data, even though the the CPS data are already top-coded.[17] The likelihood of distortion in the relative income distributions caused by high-end mobility is thus reduced.

Other forms of transformation are possible. It is common to take logarithms of the raw data, which effectively "trims" outliers and cuts short the long tails present in data. Sen (1973) stressed the advantage of taking logarithms to "stagger" income levels and eliminate the arbitrariness of units of measurement. Bianchi (1997) pointed out that log transformation may remove skewness and outliers present in the original data, therefore alter some of the data's fundamental properties. Logged data

[15] Normalisation by median or maximum may give rise to different shape dynamics of income distributions, since the distance between data points may vary over time according to different normalisation methods. The length of relative income ranges obtained through a division by median or mean changes over time, while transformation by maximum guarantees a fixed range of $[0, 1]$. If density estimates based on median-transformed data are scaled and centred to $[0, 1]$, density estimates will have the same shape as those obtained from maximum-transformed data.

[16] For details, see Appendix C.

[17] The observed maximum income levels for the 1962, 1972, 1982, 1992 and 2000 samples are, respectively, 121,900, 50,000, 197,500, 396,772 and 424,770 US dollars. Top coding seems to be carried out from 1972 onwards, with increasing cutoffs. From 1992 to 2000, the maximum income changed very little. After dealing with the rounding errors by randomisation and trimming out both negative values and the top 1% outliers, the new maximum values are: 17,200, 29,900, 60,250, 100,299 and 198,252 US dollars, much smaller than the values without trimming. Proportional 10-year increments are now of comparable size: 73.8%, 101.5%, 66.5% and 97.7%.

often closely approximate a normal distribution, even though the original data may be nowhere close to a normal representation. Indeed, logarithmic transformation is often applied in such a way that the transformed data better conform to the assumption of normality. In our work, we are interested in the whole range of income distribution. The tail properties of the income data matter, and logarithmic transformation penalises the high-income range and throws away valuable information contained in the upper tail.[18]

3.2 Main results

We first apply the adaptive kernel density estimation and adaptive bootstrap multimodality test for the CPS personal income data for 1962, 1972, 1982, 1992 and 2000. Subsequently the same non-parametric analysis is conducted for population subgroup, differentiated by exogenous (age, gender and race) and endogenous population characteristics (education). There are three age groups: young (25-34), middle-aged (35-44) and older (45-54); two gender groups: male and female; two major racial groups: African-American and Caucasian; and three educational groups: elementary level, secondary level (high school graduates) and college and graduate level. Conditioning on exogenous population characteristics, we search for possible explanations for distributional heterogeneity in the overall US income distribution.

In Figures 5-9 we present histograms for the original CPS income data, in current US dollars, and the transformed data.[19] For both raw and normalised data, negative values and those in the top 1% range are trimmed out.[20] For all five years under analysis, we observe, in the original data, large spikes around zero-income level and at the hundreds and thousands. This may be due to reporting errors as the surveyed often find it convenient, for various reasons, to report rounded values. We uniformly "redistribute" these values in small bands around the rounding values and then proceed with the above-mentioned transformation. Apparently, the transformation seems to preserve the fine details of the original data well while it greatly reduces the problem of false spikes that may contaminate our kernel density estimates.

While the spikes at rounding values are large in the original data and much smaller in the transformed data in 1962 and 1972, the problem seems to become less serious in 1992 and 2000. This may be attributed to a gradual improvement in the quality of CPS surveyed data. The broad similarity between the original and transformed data can also be seen in Tukey's boxplots, as shown in Figure 10. The boxes have lines at the lower quartile, median and upper quartile values. Dashed lines extend from each end of the box to the most extreme data value within 1.5 times the inter-quartile range, beyond which lie outliers (indicated by "+"). Focussing attention on transformed data, boxplots show

---

[18] The fact that logarithmic transformation penalises the high-income range and highlights differences in the lower-income levels underlies findings that lognormal distributions fit the lower part of income density well but not the upper part.

[19] For 1962, 1972, 1982 and 1992, histograms for the original data have large spikes which are cut off so that the original and transformed data can be compared on the same scale. Data in the trimmed spikes are used throughout the analysis.

[20] Values belonging to the top 1% income class are outliers that do not affect the general shape of the income distribution, but they may greatly distort our kernel density estimates by exerting a disproportionate impact on overall smoothing.

that the CPS income data are very much skewed towards the lower end of the (relative) income range. As the boxes first increase and then decrease in length, individuals became more dispersed in the 1960s and 1970s, but more condensed in the 1980s and 1990s. Moreover, the median relative income rose steadily over the 1960s and 1970s, became stagnated in the 1980s, then dropped dramatically over the 1990s. We observe a great number of outliers in the upper range but essentially no outliers in the lower range (by construction) throughout all years under study.

As a result of our bootstrap[21] tests for the number of modes, we observe that for almost all years, for the surveyed as a whole and across various groupings (in terms of age, education, gender and race), the null of unimodality has been frequently rejected at the 1% significance level. Bimodality and trimodality are the rule, with four as the maximum number of modes that we could accept with our tests. In Table 1, we list the critical bandwidths for the overall income distribution, computed as a function of the numbers of modes, increasing from 1 to 6. The corresponding p-values are included in parentheses. Clearly, critical bandwidth is a monotonically decreasing function of the number of modes (see also Figure 11). For the group as a whole, the US income distribution has been bimodal throughout the period, except for 1972, when it became briefly trimodal. The "stylised fact" of unimodal income distributions so often implicitly assumed in the literature has been strongly rejected by our mode testing evidence. The multimodality test results for population subgroups are reported in a similar fashion in Table 2.

Initial pilot and final critically smoothed kernel density estimates as described in Section 2.3 are plotted in Figures 12 and 13. Our initial density estimates (estimated using a Gaussian kernel and an "optimal" bandwidth, selected with reference to normal distributions and an adaptive estimate of spread) do a reasonable job in describing the general shape of the US income distributions for the last four decades. But compared to the final income densities estimated using adaptive local bandwidths and critical smoothing, initial pilot estimates are much more wiggled in the long tails and are also distorting in the centre. For instance, the estimated lower modal spike for the 1972 income distribution has much less mass than that indicted by the critical adaptive estimate. More importantly, many small and possibly inconsequential modes that appear in the long tails are smoothed out in the adaptive estimates, while important features in central parts remain in place.

Changes in the shape of the US income distribution over the entire range is better illustrated in three-dimensional diagrams (Figure 13). The pictures suggest important stylised facts in terms of relative income: the general shape of the US income distribution changed little in the 1960s and 1970s, with a large spike in the very low end of the income range and a sizeable concentration of mass in the $[0.1, 0.4]$ range. While the spike grew higher over the 1960s, it was greatly reduced in the 1970s and 1980s, much to the benefit of the middle class. Improvement in the US income distribution during the long economic recovery over the 1980s is particularly noticeable. The messages from the 1990s are mixed: while the lower middle class in the range $[0.04, 0.25]$ became greatly inflated, this seemed

[21]We have been able to obtain test results for pooled 1962-2000 income data, each with 500 bootstrap replications. The running time for one single program ranges from around 307 hours (13 days) to about 1,287 hours (54 days) of CPU time on an Intel Pentium IV 2.0 GHz computer.

For some population subgroups, we first use importance sampling to obtain smaller data samples, with sample sizes varying between 2,000 and 6,000, as inputs for bootstrap modality tests. In these cases, the number of bootstrap replications varies between 300 and 500. For the selected years, the analysis is carried out for the population as a whole and the 10 population subgroups, resulting in a total of 55 programs.

to come at the expense of the middle and upper-middle classes. A large density mass originally located in the range $[0.25, 0.6]$ has moved downwards to the $[0.04, 0.25]$ range, and the trend of diminishing spike in the very low end of income range has also been reversed. Of course, in terms of absolute income, the whole US income distribution has shifted upwards over the years, but in terms of relative income, improvement in the 1960s, 1970s and 1980s was brought to a stop and the quality of the US income distribution deteriorated quickly in the 1990s.[22] Although the number of modes has been maintained roughly constant at two, the shape dynamics of the US income distribution provide information over the entire range that may not be available otherwise.

To examine the income distribution dynamics for population subgroups differentiated by important characteristics such as age, education, gender and race, we collect adaptive kernel density estimates for all our population subgroups in both two and three-dimensional plots (Figures 14-17). The estimates are presented after we run the Silverman test for each subsample and select the corresponding critical number of modes. Two important observations immediately stand out: first, almost all subsample density estimates are multimodal during the sample periods under study (refer to Table 2); almost all population subgroups have their own distinct distributional dynamics over the last four decades.

As we observe in Figures 14 and 15, historically, male and female groups follow very different distributional patterns. The male population group has income distributions that are bimodal throughout the period of study, with the mode in the very low end of income range swiftly disappearing and a second mode beginning to appear in the central part during the 1960s. For most part of the 1960s, 1970s and 1980s, density mass gradually spread out from the central part to both lower and higher neighbouring areas. Then the great shake-up set in over the 1990s, when a sizeable density mass shifted from the $[0.22, 0.65]$ range to the low $[0, 0.22]$ range. This movement certainly had a large impact on the overall income distribution dynamics.

The female population group followed a different trend, being bimodal in 1962, 1972 and 2000, but unimodal in 1982 and 1992. While the great mass of the female population has been concentrated in the lower relative income range, the mass has been spreading out to both higher and lower-income ranges in the 1960s and 1970s, but mainly to the low end. Over the 1980s and 1990s, the spike in the lower end became greatly reduced and the lower middle class range $[0.05, 0.3]$ became inflated. So when a sizeable proportion of the male population moved downwards in the 1990s, the upward movement of much of the female population was noticeable.

The African-American and Caucasian population groups also differ greatly in their distributional dynamics. Convergence across the African-American population is evident, with the distribution moving from trimodality in 1962 to bimodality in 1972 and 1982 and further to unimodality in 1992 and 2000. Given that the maximum income for the African-American population is rather low compared to that of the Caucasian population, and that a sizeable portion of its converging population mass moved from the upper income range to the low central part, the quality of convergence does not seem to be laudable. On the contrary, the Caucasian population is much larger in absolute size

[22] This observation should be interpreted with some caution, as we discussed in Section 3.1.2. However, given that we have trimmed off the top 1% of income data and the fact that the maximum level of maximum income level after trimming has increased at comparable growth rates over the last four decades, there does not seem to be any extraordinary upward movement by the richest income group. The conclusions should be considered robust.

and dominates the overall US income distributional dynamics. As a matter of fact, the distributional pattern of the Caucasian population over the sample period is quite similar to that of the US population as a whole.

Income distributions of population subgroups differentiated by age all have general shapes and trends that closely mimic those of the overall US income distribution. Again bimodality and trimodality are the rule, with the large spike in the very low end of income range gradually thinning out and its mass moving to the middle-income range. The big shake-up again came in the 1990s, with much of the mass in the middle or upper-middle ranges reallocated to the lower-middle range. This type of income convergence may not be desirable.

Decomposition of the overall U.S. income distribution by levels of education also reveal complex and distinct distributional dynamics. Although the group with only elementary education has numbers of modes that vary between two to four, while the secondary group has bimodal income distributions throughout the whole period, the general shape dynamics are quite similar for these two subgroups. The spikes at the low end of income range became gradually sharper and thinner, with the mass in low and high ranges shifting towards the central part of income distribution. The group of people with college or higher educational background has a distinct pattern, with more accentuated movements in density masses. The general shape was clearly bimodal, but the deep dip at 0.08 became shallower and the low-end spike diminished gradually over the years. Sizeable density masses moved out from low to the middle and upper income ranges in the 1960s, 1970s and 1980s, fattening the tails. The dramatic reversal occurred in the 1990s, with a large, thick spike emerging around the value 0.135. Again there was convergence but unfortunately it occurred because of the downward movement of a large population mass.

Summing up, these results confirm our conjecture that both the overall and the disaggregated income distributions in the US have been multimodal over the last four decades. While some subgroups have income dynamics similar to those of the entire population, most groups have their own distinct distributional patterns and shape changes. The presence of multiple modes implies that strong population heterogeneity and complex distributional dynamics are essential features of the US personal income data that cannot be ignored by economists. We are able to identify distinct distributional characteristics for various mutually exclusive socioeconomic groups. Population subgroups with distinct income dynamics interact to create complex and evolving overall distributional dynamics. Socioeconomic factors and institutional details matter for income distribution dynamics.

4 Conclusion

In this paper, we propose to nonparametrically estimate the US income distribution over a span of four decades by an *adaptive* kernel density estimation method. We propose an *adaptive* version of Silverman's (1981, 1983) bootstrap test to examine the number of modes in income distributions. Changes in the shape of the income distribution over the entire income range provide rich information and shed light on issues of income inequality, poverty traps and convergence.

Our results indicate that the US personal income distribution has been bimodal in 1962, 1982, 1992 and 2000 and trimodal in 1972. Evolution of the shape of income density reveals important information. Though the shape of the US income distributions did not change much during the 1960s and 1970s, but there was a noticeable shift of the probability mass from the very low end to the middle of the relative income range. The shift continued until the early 1990s, implying that hypotheses of poverty traps and vanishing middle class cannot be validated in that period. The big change in this general trend was set in motion in the "roaring" 1990s, a decade of remarkable economic expansion: not only did the low-end spike become inflated, but a great population mass has also shifted downwards from the upper-middle ($[0.25, 0.6]$) to the lower-middle ($[0.04, 0.25]$) income range. What caused this reversal is an important question.

Multimodality implies strong population heterogeneity and distinct ongoing distributional dynamics across population subgroups with distinct characteristics. To further investigate the depth of multimodality existing in data, we decompose the survey population into age, gender, racial and educational subgroups and apply the same nonparametric analysis. Almost all population subgroups exhibit multimodality during the sample period, and they follow distinct distributional patterns, revealing complex income distribution dynamics. The abrupt trend reversal in the 1990s is noticeable in income distributions of all population subgroups, albeit to different degrees of severity.

Evidence of multimodality suggests that previous studies analyzing income data within the framework of unimodal parametric distributions are not an accurate description of reality. Instead, a nonparametric approach or parametric modelling with multimodal finite mixtures may be more fruitful. Moreover, theoretical models should be capable of explaining and generating multimodal income distributions. The adaptive method employed in this paper can be easily extended to a wide range of topics that involve estimation and testing of size distributions of interesting economic variables, such as industrial firms, farms, consumption expenditure, commodity prices, stock prices, dividends, wages and earnings.

Appendices

A Traditional empirical measurement of income distributions

A.1 Conventional fractile and scalar measures

The family of fractile measures include income shares for different quintiles, deciles or percentiles of the population, ratios such as $P90/P10$, the income share of the richest 10% of population divided by the income share of the poorest 10%, and log ratios thereby derived, eg, $log(P90/P10)$. These measures are easy to calculate and provide useful information about the changes at relevant "focal points" of income distributions. However, they are also very crude in that they only provide a segmented snapshot of the shape of income distributions, ignoring distributional dynamics within and beyond each fractile. Fractile methods give an incomplete picture of the income distributions and may be grossly misleading.[23]

The second family of inequality measures are designed to characterise income distributions and compress information on income inequality into one single statistic, facilitating the comparison and ranking of different income distributions.[24] Simple statistics like mean, median, variance, skewness and kurtosis only reveal partial information about the income distribution. More sophisticated measures are often based on the Lorenz criterion. But once two Lorenz curves derived from different data cross one or more times, the associated measures are deemed non-comparable.[25] Different measures based on the same set of income data may disagree. In these cases unambiguous inequality judgement is not possible. One remedy often used in empirical studies is to compute a large number of scalar inequality measures in such a way that a conclusion is drawn while most statistics concur (see, for example, Biewen (2000)). This is certainly a problematic way to overcome potential inconsistencies between conventional inequality measures.

The summary nature of these measures alludes to the important fact that they ignore much of the information in the shape of the *entire* range of income distribution and obscure key facts of intradistributional dynamics (persistence, churning and crisscrossing). Density estimation, either parametric or nonparametric, remedy the shortcomings of conventional inequality measures by providing more complete information not available otherwise. In fact, as described in Cowell (2000), many inequality measures can be expressed as functionals of the income density, which can be consistently estimated by kernel methods.

[23] Wolfson (1997) constructed an illustrative example showing how sensitive fractile measures could be to a seemingly harmless change in the definition of income classes.

[24] Champernowne (1953, 1974), Cowell (2000), Foster (1985) and Sen (1973) provide detailed descriptions of these measures.

[25] Foster (1985) describes two general approaches to get around the difficulty. The partial ranking or quasi-ordering approach is concerned with "finding and characterising unanimous and unambiguous inequality rankings", which is valid only for certain comparisons. In the axiomatic approach, complete inequality rankings are possible by assigning numerical values to each measure.

A.2 Parametric estimation of income distributions

Numerous attempts have been made to provide a precise mathematical description for the size distribution of income. Using Dagum's classification, these attempts can be grouped into three broad categories: 1) functional forms proposed to describe an income distribution that has been generated by a stochastic process, as in Champernowne (1953) and Fisk (1961); 2) functional forms obtained by adopting a pragmatic approach, ie, solely by their ability to provide a good fit to empirical data, as in McDonald (1984); 3) specification of a differential equation to capture the characteristics of regularity and permanence observed in an empirical distribution of income, as in Singh and Maddala (1976). In most cases, a good fit to the data is the primary objective, with sound economic interpretations and computational simplicity a bonus.

Many parametric specifications have been experimented. These include, among others, the Pareto, lognormal, exponential and Weibull distributions, the beta distribution of Thurow (1970), the gamma density of Salem and Mount (1974), the Singh-Maddala (1976) distribution based on the hazard rate, the generalised beta (GB) distributions of McDonald (1984), the generalised gamma distribution (GG) of Estaben (1986) based on an "income share elasticity" approach. Majumder and Chakravarty (1990) provided a four-parameter model of personal income distribution based on the same approach. Broadly speaking, the Pareto distribution gives excellent fit to the upper part of income distribution but fails elsewhere, while the lognormal fits well the lower part but badly in the upper tail. The gamma distribution fits well the middle part, but seems to be inferior to the Singh-Maddala distribution, which some authors claim to deliver an overall good fit for the full range of income distribution.

B Some classical multimodality tests

Good and Gaskins (1980) proposed an iterative surgery or repeated smoothing procedure for their maximum "penalised" likelihood (MPL) estimates $\hat{f}_1$. The significance of any bump is assessed by a Bayesian interpretation of the MPL method. More precisely, a new smoothed density estimate $\hat{f}_2$ is obtained by removing the particular bump under question (see Figure 2). Then, the difference between the penalised log likelihoods, $\hat{f}_1$ and $\hat{f}_2$, ie, $|\ell(\hat{f}_1) - \ell(\hat{f}_2)|$, is taken as the posterior log-odds in favour of the bump. Unlike kernel-based testing procedures, the theoretical properties of the Good and Gaskins' (1980) estimator are not yet well understood.

The dip and excess mass approaches to mode testing were introduced by Hartigan and Hartigan (1985) and Müller and Sawitzki (1991), respectively. For the one-dimensional case, the dip test and excess mass test are equivalent. For $Y_1, \ldots, Y_n$, a random sample from the distribution function F, the empirical distribution function F_n is defined as:

$$F_n(y) = n^{-1} \sum_{i=1}^{n} 1\left\{Y_i \leq y\right\}$$

where $1\{\cdot\}$ is the indicator function. The *dip* statistic DIP can then be defined as the maximum difference, over all sample points, between the empirical distribution function F_n and the unimodal distribution that minimises this maximum difference:

$$DIP(F_n) = \inf_{F \in U_1} \sup_{y \in Y} |F_n(y) - F(y)| \tag{12}$$

where U_1 is the class of all distribution functions with unimodal densities and Y is the support of y (see Figure 3). The null hypothesis of unimodality can be rejected against the alternative of m modes when DIP is too large. Hartigan (1985) developed a computational algorithm for implementing the dip test. Hartigan (2000) extended the original unimodality dip test to the case of multimodalily, where U_m, the class of uniform mixtures with at most m modes, is the benchmark in a test of m modes.

Define *excess mass* as $E(\lambda) = \int (f(y) - \lambda)^+ dy$, the probability mass exceeding the density level λ. It measures the strength of a mode. Müller and Sawitzki's (1991) excess difference test statistic $\Delta_{n,m}$ for a sample of size n can be defined in terms of the *empirical excess mass* for m modes:

$$E_{n,m}(\lambda) = \sup_{C_1,\ldots,C_m} \sum_{j=1}^{m} \{F_n(C_j) - \lambda \|C_j\|\}$$

where λ is a positive constant, $\{C_j\}_{j=1}^m$ are pairwise disjoint connected sets of the form $C \subset \{y \in Y : f(y) \geq \lambda\}$, $F_n(C)$ is the empirical measure of C, and $\|C\|$ is the length of C (see Figure 4). Putting $D_{n,m}(\lambda) = E_{n,m}(\lambda) - E_{n,m-1}(\lambda)$, the *excess difference* test statistic for multimodality is $\Delta_{n,m} = \sup_\lambda D_{n,m}(\lambda)$. The null hypothesis of $m-1$ modes can be rejected against the alternative of m modes when $\Delta_{n,m}$ is too large. It is not difficult to verify that for $m=2$, $\Delta_{n,2} = 2DIP(F_n)$.

Both the dip and excess difference test statistics are closely related to the Kolmogorov-Smirnov (KS) statistic, often used as a measure of goodness of fit:

$$KS = \|F_n - F\|_\infty = \sup_t |F_n(t) - F(t)|$$

From the Glivenko-Cantelli theorem, if $Y_1, \ldots, Y_n$ are i.i.d. random variables with distribution function F, then as $n \longrightarrow \infty$, we have:

$$KS \longrightarrow_{a.s.} 0$$

This imples $DIP(F_n) \longrightarrow_{a.s.} DIP(F)$ as $n \longrightarrow \infty$. The dip test is consistent and asymptotivally distinguishes any unimodal distribution from any multimodal distribution.

Recently, Cheng and Hall (1998) exploit the fact that the limiting distributions of the excess difference statistic under the null depends on unknowns only through a constant and suggest a way to calibrate the dip and excess mass tests that may substantially improve their level accuracy and power. To overcome conservatism inherent in these tests, Cheng and Hall (1999) propose the use of a boundary null hypothesis of one mode and one shoulder and calibrating the bandwidth, dip and excess mass tests by resampling from a "template" density with exactly one mode and one shoulder. The mode-and-shoulder test, when properly calibrated, has asymptotically correct level. Hartigan (2000) generalises the dip test to handle the possibility of many modes. His test statistic is based on a "shoulder interval", a maximal interval of constant density that is neither a mode nor an antimode, in a best fitting m-modal uniform mixture. The test is calibrated by reference to an appropriate null distribution.

Despite these exciting new developments in the theory of mode testing, we use the bandwidth test of Silverman (1981) in our analysis, for its appealing and well-understood theoretical properties and interesting practical implications. Furthermore, our version of the test is a natural extension of the adaptive kernel estimation of income density.

C Simulations on data transformation

Steadily increasing income levels expands the income range over time, and normalisation by maximum may not distinguish the following two cases. Suppose that in the normalised data, we observe a shift of large probability mass from $[0.30, 0.55]$ to $[0.15, 0.25]$. One may conclude that the welfare of a sizeable portion of the population has deteriorated in relative terms. One possibility is that the middle-class has lost ground in absolute terms. But it may also occur that a very small number of high-income earners, of approximately zero probability mass, are earning much more. These few individuals would greatly expand the absolute income range, without much change in the rest of income distribution.

We run simulations to examine how sensitive our normalisation is to these two competing cases under bimodal and trimodal densities, where the data are generated by a mixture of two normal densities located in the low and middle ranges, and a uniform density of little mass located at the high end of the income range:

$$\begin{aligned} Y_1 &\sim 0.29\aleph(2,500,\ 2,000) + 0.69\aleph(25,000,\ 20,000) \\ &\quad +0.02\Im[90,000,\ 120,000] \qquad (13) \\ Y_2 &\sim 0.29\aleph(2,500,\ 2,000) + 0.69\aleph(14,000,\ 14,000) \\ &\quad +0.02\Im[90,000,\ 120,000] \qquad (14) \\ Y_3 &\sim 0.29\aleph(2,500,\ 2,000) + 0.69\aleph(25,000,\ 20,000) \\ &\quad +0.02\Im[150,000,\ 200,000] \qquad (15) \end{aligned}$$

where $\aleph(\mu, \sigma)$ stands for a Gaussian distribution with mean μ and variance σ^2, $\Im[a, b]$ stands for the continuous uniform distribution of length $[a, b]$, and "$\sim$" means equivalence in distribution. Case 1 in (14) corresponds to relocating the middle-range normal density of Y_1 to a lower range, while Case 2 in (15) corresponds to moving the upper-range uniform density upwards in absolute terms.

Normalising the data, we apply the adaptive kernel estimation technique and the adaptive Silverman test. We find that data originated from two different scenarios but normalised by maximum typically lead to the same picture, normalisation by maximum cannot distinguish these two cases. However, other forms of normalisation, including divisions by sum, mean, median and range, cannot distinguish the two scenarios.

D Tables

Table 1: **Critical bandwidths versus numbers of modes: all population.**

Modes	1962	1972	1982	1992	2000
1	0.090 (0.00)	0.093 (0.00)	0.080 (0.00)	0.052 (0.00)	0.031 (0.00)
2	0.017 (0.64)	0.034 (0.02)	0.018 (0.32)	0.024 (0.02)	0.015 (0.03)
3	0.013 (0.61)	0.015 (0.15)	0.016 (0.02)	0.018 (0.01)	0.011 (0.00)
4	0.012 (0.25)	0.014 (0.00)	0.015 (0.00)	0.017 (0.00)	0.010 (0.00)
5	0.011 (0.11)	0.011 (0.02)	0.014 (0.00)	0.016 (0.00)	0.009 (0.00)
6	—	0.010 (0.01)	—	0.015 (0.00)	0.008 (0.00)

Note: The numbers in parentheses are the corresponding P-values of the Silverman test.

Table 2: **Critical bandwidths and numbers of modes: population subgroups.**

Groups	**1962**	**1972**	**1982**	**1992**	**2000**
All	0.017 (0.64, 2)	0.015 (0.73, 3)	0.018 (0.32, 2)	0.024 (0.02, 2)	0.015 (0.03, 2)
Young	0.028 (0.78, 2)	0.026 (0.73, 2)	0.023 (0.85, 2)	0.032 (0.33, 2)	0.020 (0.67, 2)
Middle-Aged	0.015 (0.97, 2)	0.027 (0.29, 2)	0.022 (0.60, 2)	0.019 (0.64, 3)	0.014 (0.55, 2)
Older	0.021 (0.34, 2)	0.022 (0.60, 2)	0.017 (0.50, 2)	0.021 (0.58, 3)	0.015 (0.41, 3)
Elementary	0.022 (0.41, 3)	0.029 (0.41, 2)	0.018 (0.50, 3)	0.018 (0.32, 3)	0.017 (0.40, 4)
Secondary	0.027 (0.76, 2)	0.026 (0.83, 2)	0.029 (0.44, 2)	0.026 (0.69, 2)	0.022 (0.71, 2)
College	0.016 (0.62, 2)	0.014 (0.84, 2)	0.020 (0.75, 2)	0.021 (0.85, 3)	0.012 (0.78, 3)
Male	0.015 (0.13, 2)	0.012 (0.36, 2)	0.021 (0.09, 2)	0.023 (0.51, 2)	0.017 (0.38, 2)
Female	0.022 (0.59, 2)	0.019 (0.43, 2)	0.026 (0.89, 1)	0.026 (0.51, 1)	0.023 (0.14, 2)
African-American	0.024 (0.40, 3)	0.028 (0.47, 2)	0.025 (0.73, 2)	0.028 (0.64, 1)	0.036 (0.41, 1)
Caucasian	0.012 (0.93, 2)	0.028 (0.28, 2)	0.018 (0.68, 2)	0.017 (0.39, 3)	0.014 (0.51, 2)

Note: The first number in each cell is the critical bandwidth, the corresponding P-value and critical number of mode(s) are included in parentheses.

E Figures

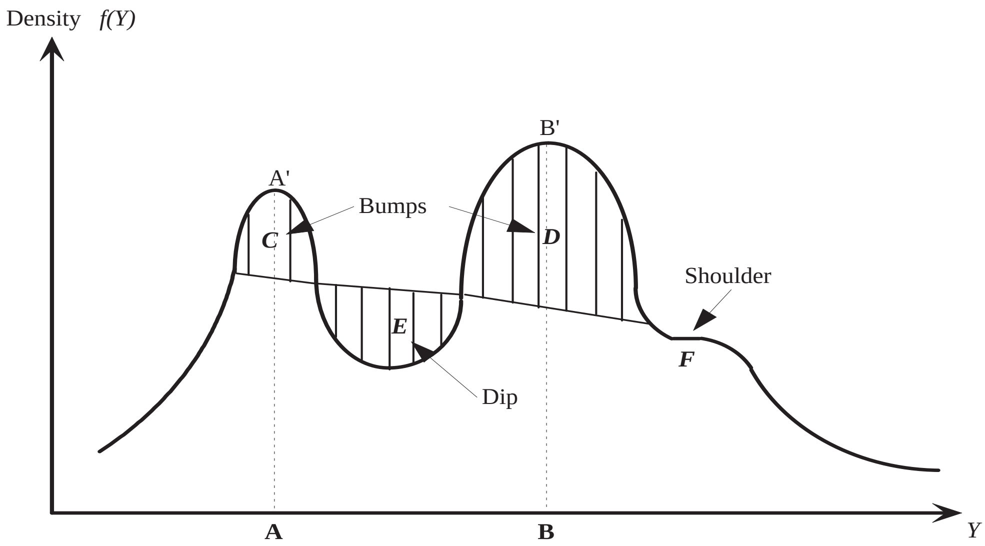

Figure 1: **Modes, bumps, dip and shoulder.** *Points A and B are modes, shaded areas C and D are bumps, area E is a dip and F is a shoulder point.*

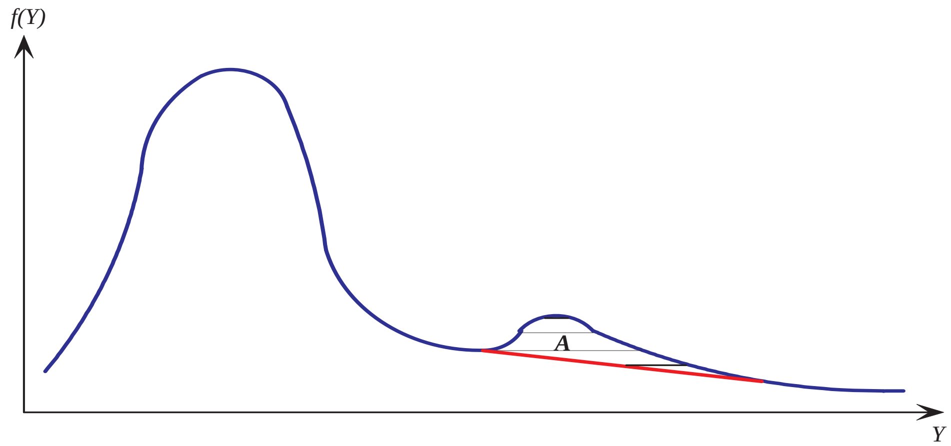

Figure 2: **Iterative surgery procedure.** *The bump A is smoothed out to obtain a unimodal density.*

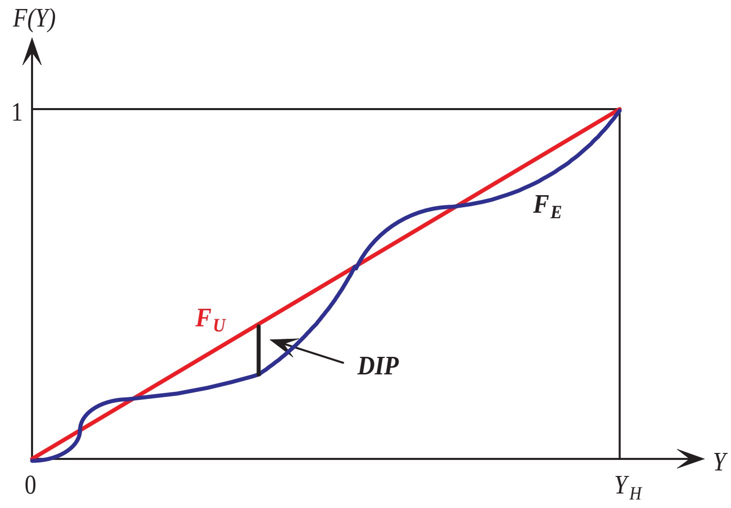

Figure 3: **DIP test statistic.** *It is the largest vertical difference between the empirical cumulative distribution F_E and the uniform distribution F_U.*

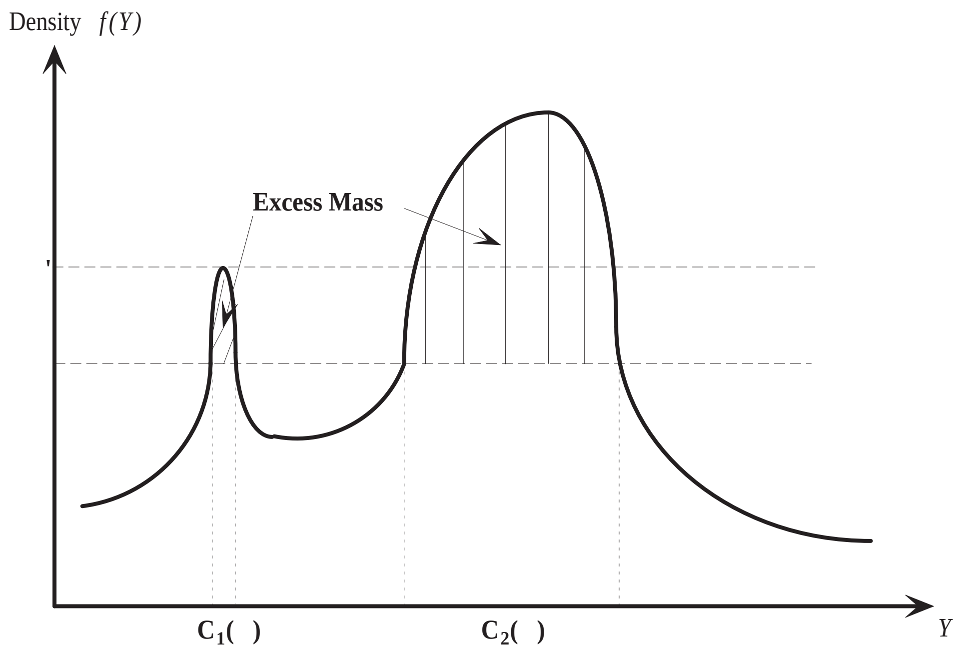

Figure 4: **Excess mass at level λ.** *Sum of shaded areas corresponds to the connected sets $C_1(\lambda)$ and $C_2(\lambda)$.*

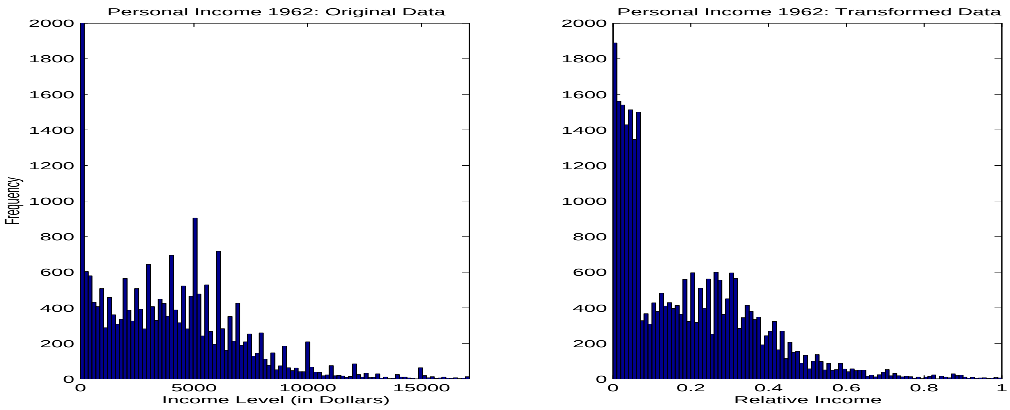

Figure 5: **Histograms for original and transformed US income data: 1962**

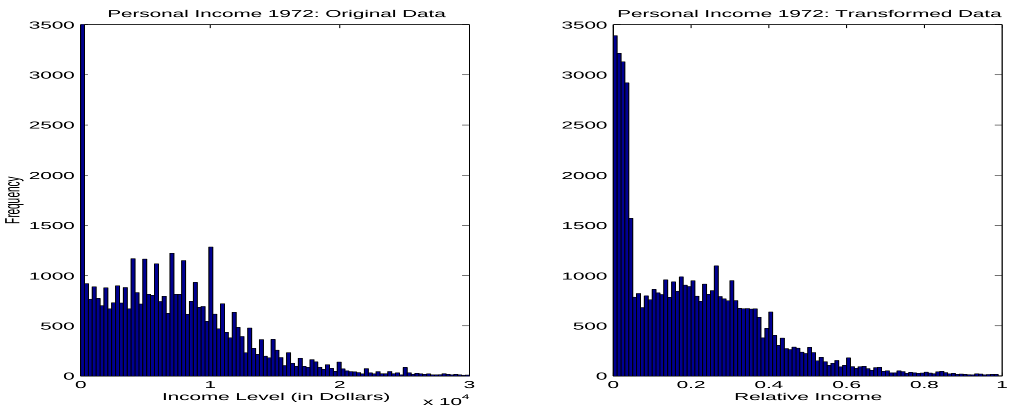

Figure 6: **Histograms for original and transformed US income data: 1972**

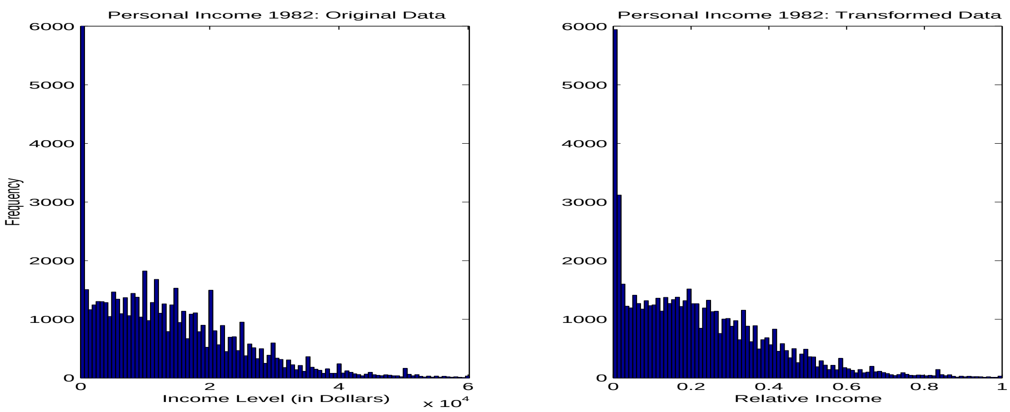

Figure 7: **Histograms for original and transformed US income data: 1982**

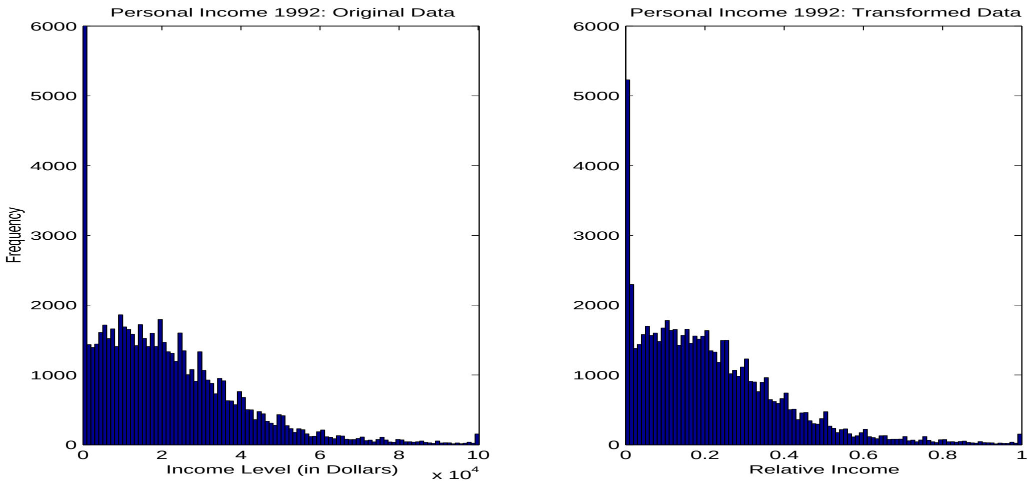

Figure 8: **Histograms for original and transformed US income data: 1992**

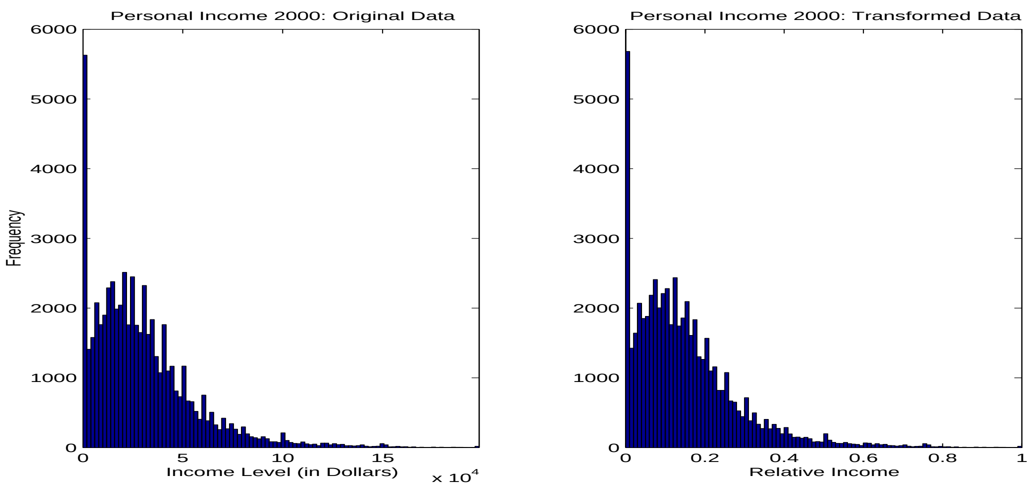

Figure 9: **Histograms for original and transformed US income data: 2000**

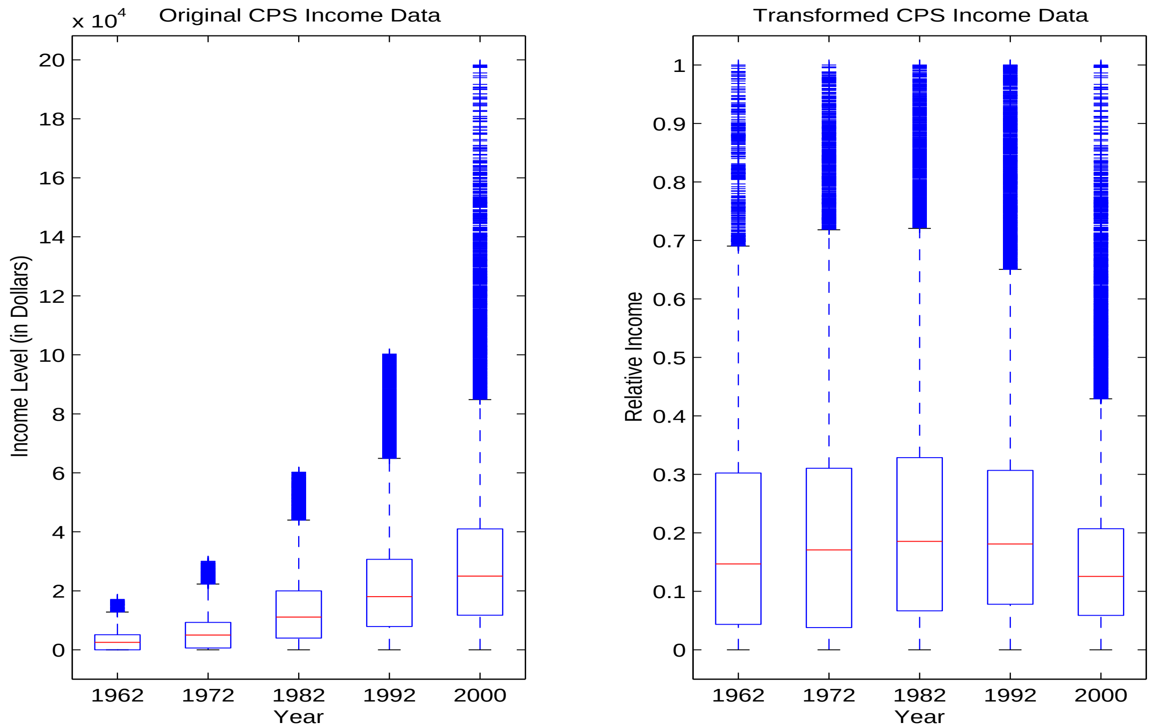

Figure 10: **Boxplots for original and transformed US income data**

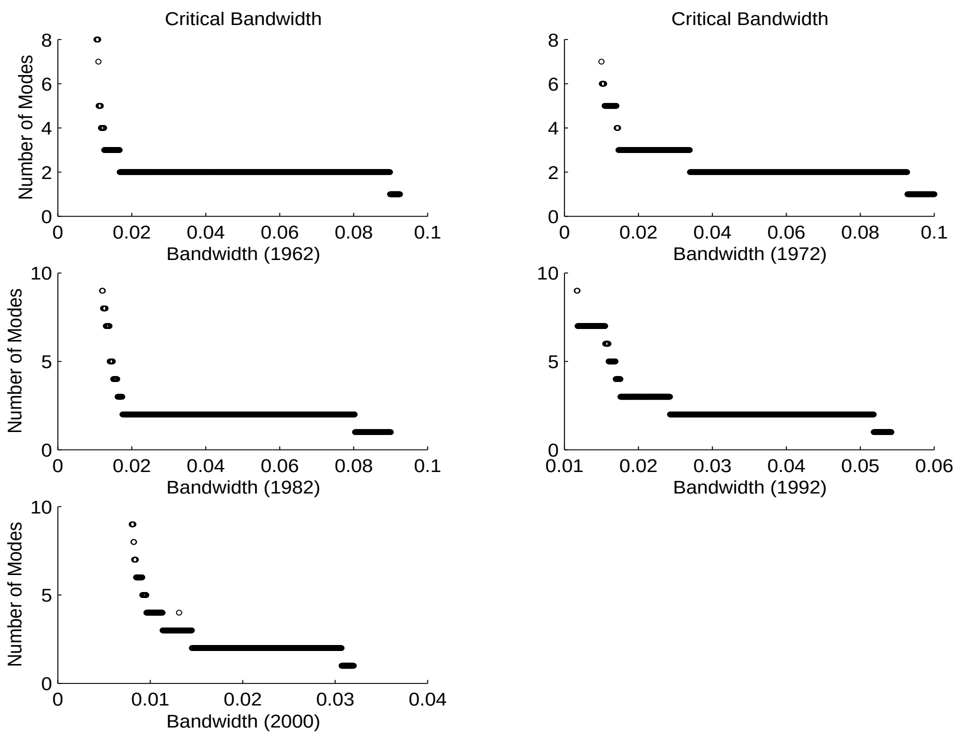

Figure 11: **Critical bandwidth versus number of modes: 1962 - 2000**

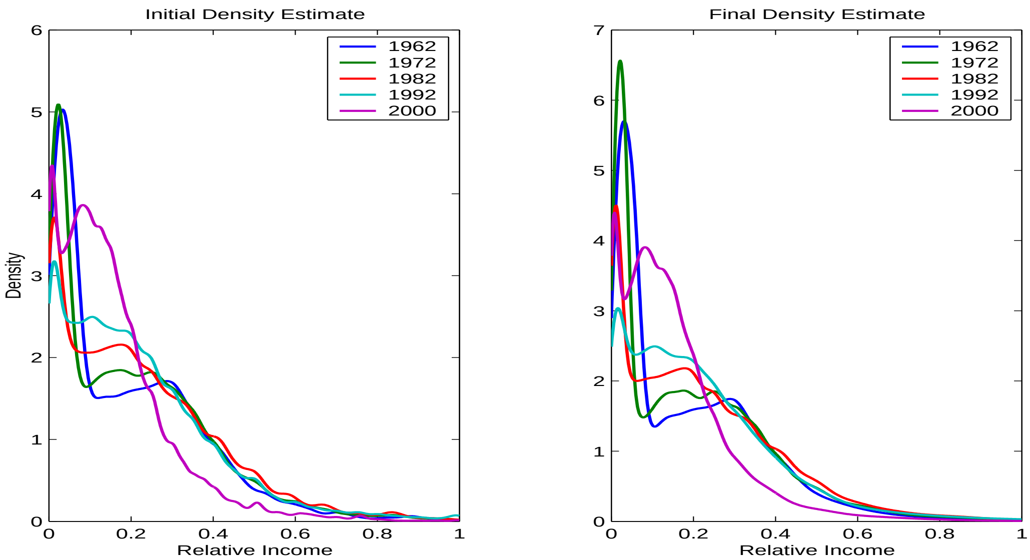

Figure 12: **US income distribution (two-dimensional), 1962-2000**

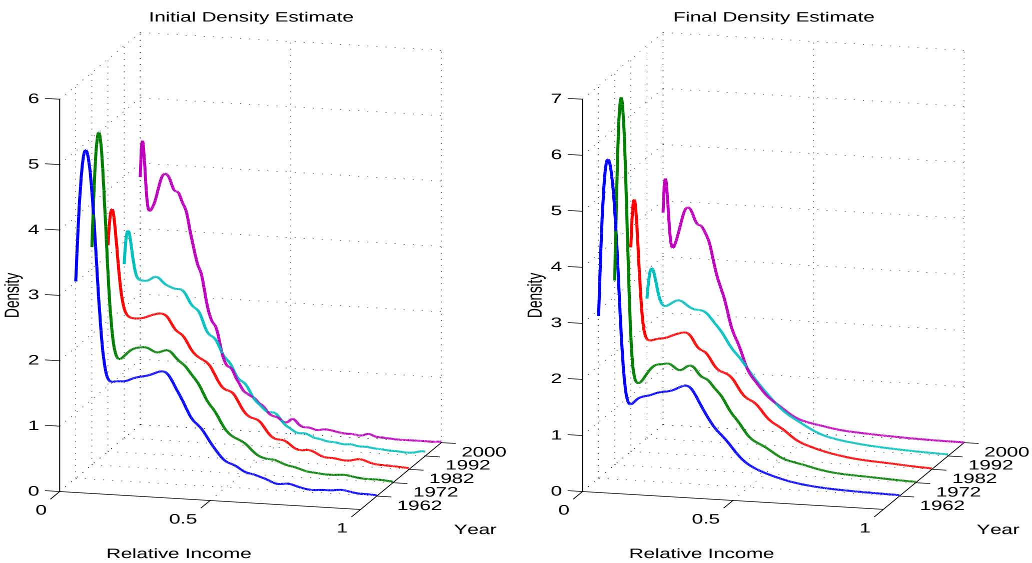

Figure 13: **US income distribution (three-dimensional), 1962-2000**

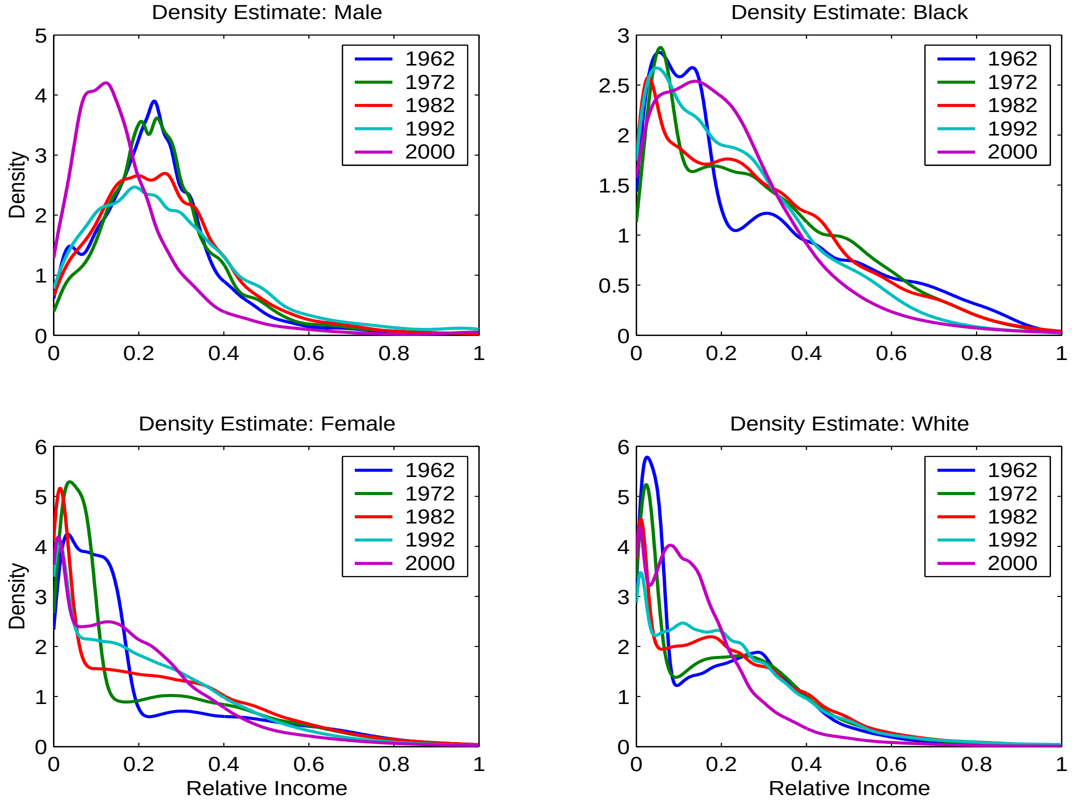

Figure 14: **US income distribution (two-dimensional).** *Gender and racial groups, 1962-2000*

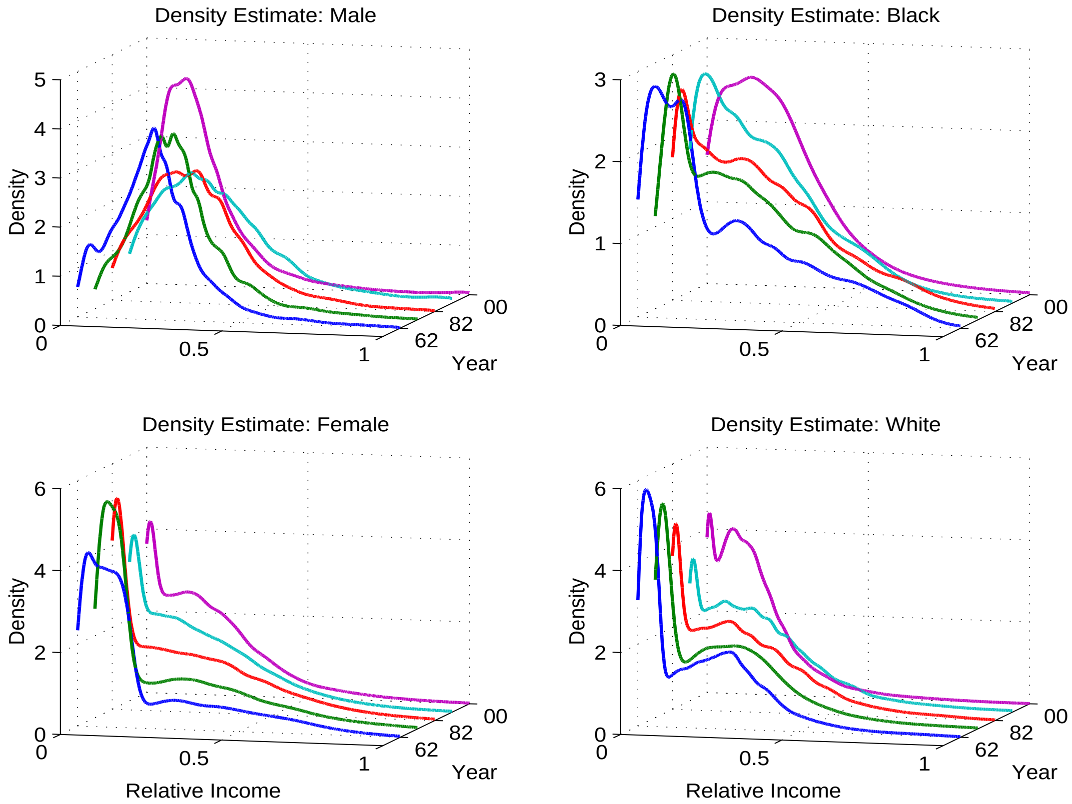

Figure 15: **US income distribution (three-dimensional).** *Gender and racial groups, 1962-2000*

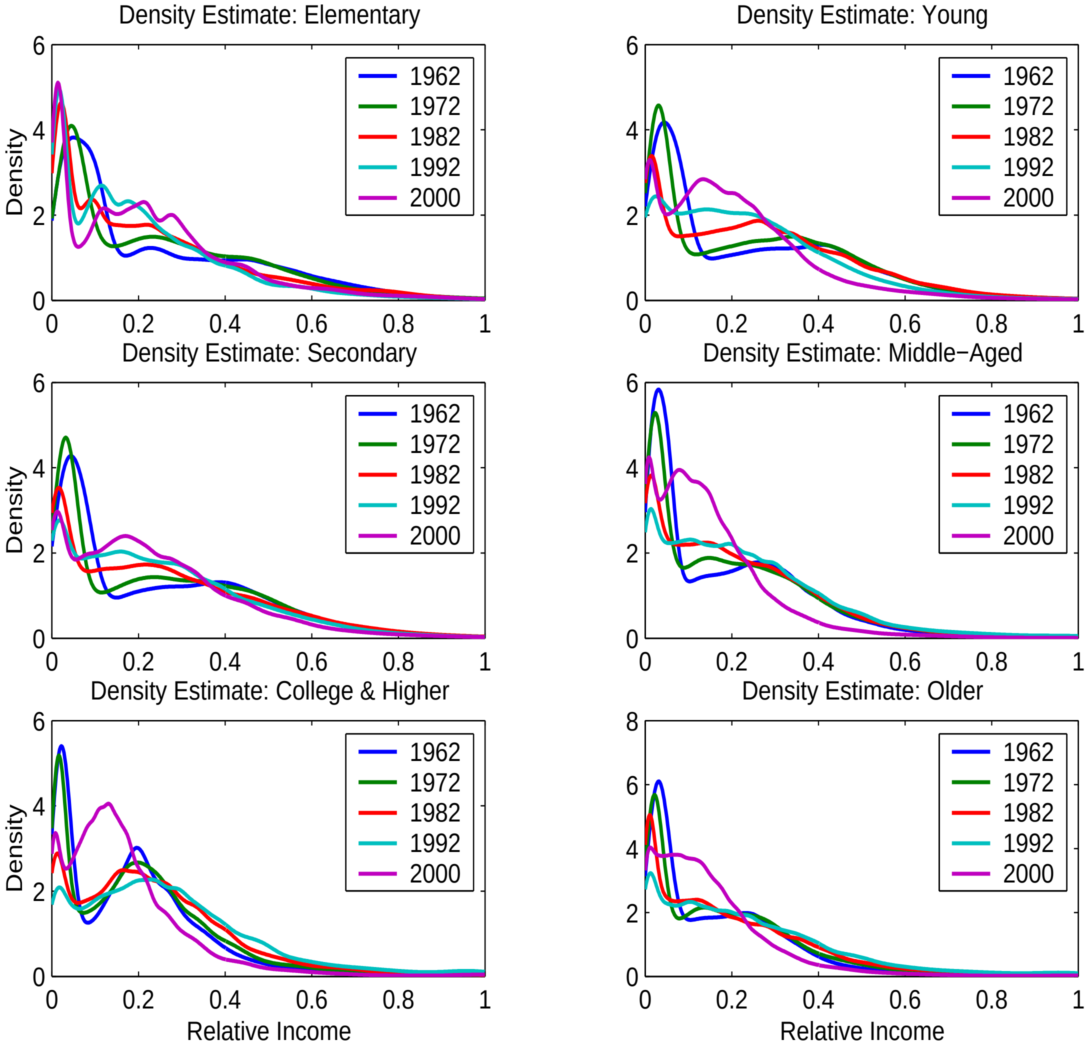

Figure 16: **US income distribution (two-dimensional).** *Educational and age groups, 1962-2000*

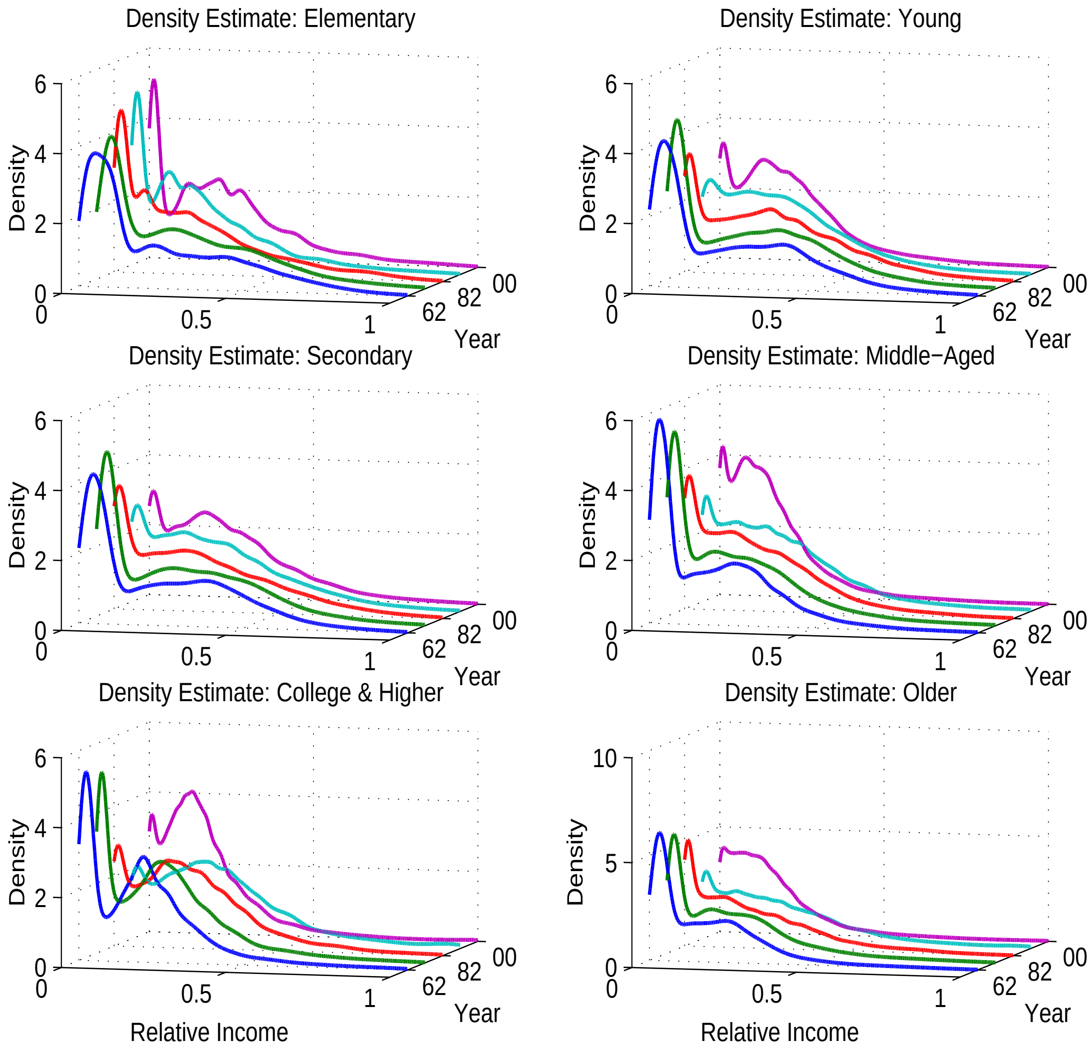

Figure 17: **US income distribution (three-dimensional).** *Educational and age groups, 1962-2000*

References

[1] Abramson, I. S. (1982), "On Bandwidth Variation in Kernel Estimates - A Square Root Law," *Annals of Statistics*, Vol. 10, pp. 1217-1223.

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

[3] Biewen, M. (2000), "Income Inequality in Germany During the 1980s and 1990s," *Review of Income and Wealth*, Series 46:1, pp. 1-19.

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

[5] Champernowne, D. G. (1953), "A Model for Income Distribution," *Economic Journal*, Vol. 53, pp. 318-351.

[6] Champernowne, D. G. (1974), "A Comparison of Measures of Inequality of Income Distribution," *Economic Journal*, Vol. 84, pp. 787-816.

[7] Cheng, M.-Y. and Hall, P. (1998), "Calibrating the Excess Mass and Dip Tests of Modality," *Journal of Royal Statistical Society, Series B*, Vol. 60:3, pp. 579-589.

[8] Cheng, M.-Y. and Hall, P. (1999), "Mode Testing in Difficult Cases," *Annals of Statistics*, Vol. 27:4, pp. 1294-1315.

[9] Cowell, F. A. (2000), "Measurement of Inequality," in Anthony B. Atkinson and Francois Bourguignon, eds., *Handbook of Income Distribution*, Vol. 1, Amsterdam: Elsevier Science, pp. 87-166.

[10] Devroye, L. (1987), *A Course in Density Estimation*, Boston: Birkhäuser.

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

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

[13] Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman Hall.

[14] Estaben, J. (1986) "Income Share Elasticity and the Size Distribution of Income," *International Economic Review*, Vol. 27, pp. 439-444.

[15] Fisk, P. R. (1961), "The Graduation of Income Distribution," *Econometrica*, Vol. 29, pp. 171-185.

[16] Foster, J. E. (1985), "Inequality Measurement," *Proceedings of Symposia in Applied Mathematics*, Vol. 33, pp. 31-68.

[17] Good, I. J. and Gaskins, R. A. (1980), "Density Estimation and Bump-hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data," with discussions, *Journal of the American Statistical Association*, Vol. 75, pp. 42-73.

[18] Gottschalk, P. (1997), "Inequality, Income Growth, and Mobility: The Basic Facts," *Journal of Economic Perspectives*, Vol. 11:2, pp. 21-40.

[19] Gottschalk, P. and Smeeding, T. M. (1997), "Cross-National Comparisons of Earnings and Income Inequality," *Journal of Economic Literature*, Vol. 35, pp. 633-687

[20] Härdle, 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.

[21] Hartigan, J. A. (2000), "Testing for Antimodes," manuscript, Department of Statistics, Yale University.

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

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

[24] Hildenbrand, K. and Hildenbrand, W. (1986), "On the Mean Income Effect: A Data Analysis of the U.K. Family Expenditure Survey," in W. Hildenbrand and A. Mas-Colell, eds., *Contributions to Mathematical Economics*, Chapter 14, Amsterdam: Elsevier Science, pp. 247-268.

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

[26] Levy, F. and Murnane, R. J. (1992), "US Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature*, Vol. 30, pp. 1333-1381.

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

[28] McDonald, J. B. (1984), "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, Vol. 52:3, pp. 647-664.

[29] McDonald, J. B. and Mantrala, A. (1995), "The Distribution of Personal Income: Revisited," *Journal of Applied Econometrics*, Vol. 10, pp. 201-204.

[30] McDonald, J. B. and Ransom, M. R. (1979), "Functional Forms, Estimation Techniques and the Distribution of Income," *Econometrica*, Vol. 47:6, pp. 1513-1525.

[31] Müller, D. W. and Sawitzki, G. (1991), "Excess Mass Estimates and Tests for Multimodality," *Journal of the American Statistical Association*, Vol. 86, pp. 738-746.

[32] Pagan, A. and Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge: Cambridge University Press.

[33] Parzen, E. (1962), "On Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics*, Vol. 33, pp. 1065-1076.

[34] Quah, D. T. (1996a), "Twin Peaks: Growth and Convergence in Models of Distribution Dynamics," *Economic Journal*, Vol. 106, pp. 1045-1055.

[35] Quah, D. T. (1996b), "Empirics for Economic Growth and Convergence," *European Economic Review*, Vol. 40, pp. 1353-1375.

[36] Quah, D. T. (1997), "Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs," *Journal of Economic Growth*, Vol. 2, pp. 27-59.

[37] Robinson, P. M. (1998), *Notes on Nonparametric and Semiparametric Estimation*, manuscript, London School of Economics.

[38] Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics*, Vol. 27, pp. 832-837.

[39] Salem, A. B. Z. and Mount, T. D. (1974), "A Convenient Descriptive Model of Income Distribution: the Gamma Density," *Econometrica*, Vol. 42:6, pp. 1115-1128.

[40] 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.

[41] Silverman, B. W. (1983), "Some Properties of a Test for Multimodality Based on Kernel Density Estimates," in J. F. C. Kingman and G. E. H. Reuter, eds., *Probability, Statistics and Analysis*, Cambridge: Cambridge University Press, pp. 248-259.

[42] Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

[43] Singh, S. K. and Maddala, G. S. (1976), "A Function for Size Distribution of Incomes," *Econometrica*, Vol. 44:5, pp. 963-970.

[44] Thurow, L. C. (1970), "Analyzing the American Income Distribution," *American Economic Review*, Papers and Proceedings, Vol. 60, pp. 261-269.

[45] Wolfson, M. C. (1997), "Divergent Inequalities: Theory and Empirical Results," *Review of Income and Wealth*, Series 43:4, pp. 401-421.

[46] Zhu, F. (2002), "Asymptotics and Exact Distributions for Normalization of Finite Mixture Data," *manuscript*, Yale University.