A Test for Multimodality of Regression Derivatives with an Application to Nonparametric Growth Regressions

Henderson, Daniel J.
State University of New York at Binghamton

18. April 2008

A Test for Multimodality of Regression Derivatives with an Application to Nonparametric Growth Regressions

Daniel J. Henderson*
Department of Economics
State University of New York at Binghamton

April 18, 2008

Abstract

This paper presents a method to test for multimodality of an estimated kernel density of parameter estimates from a local-linear least-squares regression derivative. The procedure is laid out in seven simple steps and a suggestion for implementation is proposed. A Monte Carlo exercise is used to examine the finite sample properties of the test along with those from a calibrated version of it which corrects for the conservative nature of Silverman-type tests. The test is included in a study on nonparametric growth regressions. The results show that in the estimation of unconditional β-convergence, the distribution of the parameter estimates is multimodal with one mode in the negative region (primarily OECD economies) and possibly two modes in the positive region (primarily non-OECD economies) of the parameter estimates. The results for conditional β-convergence show that the density is predominantly negative and unimodal. Finally, the application attempts to determine why particular observations posess positive marginal effects on initial income in both the unconditional and conditional frameworks.

Keywords: Nonparametric Kernel, Convergence, Modality Tests
JEL Classification: C14, C15, O10, O40

*Daniel J. Henderson, Department of Economics, State University of New York, Binghamton, NY 13902-6000, (607) 777-4480, Fax: (607) 777-2681, e-mail: djhender@binghamton.edu.

1 Introduction

Nonparametric and semiparametric kernel methods are becoming increasingly popular tools for econometricians. Researchers have begun to gravitate toward nonparametric and semiparametric methods when there is little prior knowledge on specific functional forms or some known parametric specifications are deemed inadequate for the problem at hand. This often occurs when formal rejection of a parametric model yields no clues as to the direction in which to search for an improved parametric model. This growing popularity of nonparametric methods stems from their ability to relax functional form assumptions of an unknown model and let the data determine a function tailored to the data.

Another benefit of nonparametric kernel methods is that they give a plethora of results. Observation specific estimates are given for each regressor in a local-linear regression. The problem with $n \times q$ parameter estimates is that one has $n \times q$ parameter estimates. It is often difficult and/or impractical to present this many values (along with their corresponding standard errors) in a paper. Therefore researchers often devise ways to present the results. Some authors simply look at the mean of the estimates for a particular varying coefficient. However, this ignores possible heterogeneity in the estimates. Others attempt to look at the quartiles and/or percentiles, but these also may hide some interesting findings. Another approach creates counterfactual multivariate regression surfaces via two-dimensional plots. A problem with this approach (when there are multiple regressors) is that the results are counterfactual and do not necessarily represent any particular observation. One increasingly popular method to present the results is to plot kernel densities of the estimates. This allows one to examine the entire set of estimates for a particular regressor in one simple to view figure. The question then arises, how do we view these types of figures?

What does it mean if the density appears to be multimodal? How would one test for this? This is the motivation for the test proposed in this paper. This paper provides a procedure to test for multimodality in a kernel density of estimates for a nonparametric regression derivative. The procedure is based on the test by Silverman (1981) for testing multimodality in a population density and is similar to the test for monotonicity of a nonparametric regression by Bowman, Jones and Gijbels (1998).

Nonparametric tests of modality are a distribution-free way of assessing evidence about heterogeneity in a population, provided that the potential subpopulations are sufficiently well separated. Past research on these type of procedures will be extended to assess evidence about heterogeneity of returns within a population. Evidence of multimodality will be taken as evidence that the returns to a specific variable vary significantly across different groups, time periods, and/or values of the exogenous regressor. However, as is the case with the Silverman (1981) test, a rejection of the null does not necessarily lead to identification of the cause of the multimodality. For example, we may not know if multimodality arises from the function or where the covariates appear. All that we know is that multimodality is present. This is still quite useful as reporting means and/or quartiles of the parameter estimates can mask important information. Informal inspection of the density is also inadequate because modes that are not very prominent could be anomalies attributable to measurement error or other stochastic phenomena.

The proposed test is used to study the extensively researched growth regression literature (Barro 1991; Baumol 1986; Islam 1995) to examine unconditional and conditional β-convergence. If one finds multimodality in the distribution of parameter estimates on initial income, then it shows not only that convergence rates differ, but it shows that particular groups of countries are converging at similar rates, but different to those of other groups. The analysis finds that the density of the coefficients on initial income is multimodal in the unconditional β-convergence case. Specifically, the group of OECD countries possess predominantly negative coefficients while all other groups of countries possess predominantly positive coefficients, consistent with the beliefs of past research (Baumol 1986; Durlauf and Johnson 1995) that suggest that unconditional β-convergence does not hold for all countries. Alternatively, the density of the coefficients on initial income is unimodal in the conditional β-convergence setup with negative coefficients for the majority, but not all, of the sample. The negative partial effects are consistent with past theories on conditional convergence (Barro and Sala-i-Martin 1992; Mankiw, Romer and Weil 1992). Finally, the study attempts to determine why particular observations possess positive marginal effects on initial income in both the unconditional and conditional frameworks.

The remainder of the paper is organized as follows: Section 2 outlines the Silverman (1981) test for multimodality of a population density as well as the calibration method for correct size of the Silverman test proposed by Hall and York (2001). The third section describes the proposed test while section 4 gives suggestions on how to implement the procedure, including methods for bandwidth selection and bootstrapping. The fifth section gives the results of the simulations for both the calibration method as well as the finite sample performance of the proposed test. Section 6 gives the application of the proposed method to the growth regression literature and the final section concludes.

2 Silverman Test for Multimodality

To determine the shape of an underlying population density, one can explicitly estimate the density and infer the number of modes. Silverman (1981) used this insight to develop a test that allows direct comparison between a k-modal density and a density with more than k modes. Given a sample realization $\{x_1, x_2, ..., x_n\}$ from an unknown population with density f, one can construct an estimate of this density by applying the kernel density estimator

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

where h is a smoothing parameter and $K(\cdot)$ is a kernel function. Silverman (1981) showed that if $K(\cdot)$ is the Gaussian kernel, then there is a strict nonincreasing relationship between the bandwidth and the number of modes of $\widehat{f}_h(x)$. In other words, as the bandwidth decreases, the number of modes does not decrease. Hall and York (2001) show that another benefit of using the standard normal kernel to determine modality is that unless the first three derivatives of the density are zero simultaneously (at a given mode), the bumps and troughs of the density remain separated as the bandwidth is decreased. Further, Silverman (1981) showed that the quantity

$$\widehat{h}_{crit} = \inf\left\{h : \widehat{f}_h(x) \text{ has precisely } k \text{ modes}\right\} \tag{2}$$

is well defined given the monotonicity of the number of modes. It becomes evident that if the true density was characterized by k (> 1) modes then an unduly amount of smoothing would be necessary to make $\widehat{f}(\cdot)$ appear unimodal. This provides the key to testing between the number of modes within a kernel density estimate. In fact, $\widehat{h}_{crit}$ can be used to test the null hypothesis that $f_h(x)$ has k modes against the alternative of more than k modes. By applying bootstrap methods, the size of $\widehat{h}_{crit}$ can be assessed and large values of the smoothing parameter are taken as evidence against the null hypothesis.

For a given bootstrap sample, $\{x_1^*, x_2^*, ..., x_n^*\}$, one can construct the conditional kernel density

$$\widehat{f}_h^*(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x_i^* - x}{h}\right), \tag{3}$$

and determine the number of modes of $\widehat{f}_h^*(x)$. If the number of modes is greater than k, then this provides evidence in favor of the null hypothesis, as more smoothing would be required to produce a k-modal conditional density.

A formal test of the size of $\widehat{h}_{crit}$ would be to generate B bootstrap samples from the data and determine the number of times that $\widehat{f}_h^*(x)$ possesses more than k modes. Silverman (1981) suggested that failure to reject the null hypothesis should be based on

$$\widehat{P} = P\left(\widehat{h}_{crit}^* \geq \widehat{h}_{crit}\right), \tag{4}$$

which is equivalent to finding

$$\widehat{P} = \left(\#\text{ of occurrences that } \widehat{f}_h^*(x) \text{ has more than } k \text{ modes}\right)/B \tag{5}$$

If $\widehat{P}$ is higher than some preset level α, this would imply a failure to reject the null hypothesis of a k-modal population density. The intuition here is that if there are many occurrences where one observes more than k modes then the smoothing parameter is not smoothing the density enough and should be increased. However, remember that $\widehat{h}_{crit}$ is the smallest bandwidth such that one witnesses a k-modal density, therefore increasing the bandwidth should still leave one with a density with k modes. If one sees a small value for $\widehat{P}$, then they have evidence of over smoothing and the bandwidth should be decreased. However, doing so will cause the density estimate to have more than k modes which contradicts the hypothesis of a k-modal population density.

Mammen, Marron and Fisher (1992) show that the Silverman test is conservative in its asymptotic limit. Specifically, they show that the expected number of modes from the bootstrap density is higher than the expected number of modes from the true density. The conservatism of the Silverman test leads to reduced power which can lead to misleading results when the goal is to determine unimodality versus multimodality.

Hall and York (2001) develop a testing procedure that has better power than the original test and the correct asymptotic level. They theoretically show how the Silverman test may be calibrated for any hypothesis, although the numerical calculation of the calibration factors is only derived for the null hypothesis of unimodality vs. multimodality. The reason that further calibrations are not calculated is that once the null hypothesis shifts from unimodality to bimodality, the calibration factors cannot be found with simple Monte Carlo techniques. Their procedure follows directly from that of Silverman (1981) except when calculating the bootstrapped distribution, $h = \lambda_\alpha \widehat{h}_{crit}$, where λ_α is chosen so that the test has asymptotic level α. They determine λ_α for testing the null of one mode versus the alternative of more than one mode from the bootstrap distribution of $\widehat{h}^*_{crit}/\widehat{h}_{crit}$, where $\widehat{h}^*_{crit}$ is the infimum of all bandwidths h such that the bootstrapped distribution has exactly one mode. They set up an α-level test that rejects the null hypothesis if

$$\widehat{P} = P\left(\widehat{h}^*_{crit}/\widehat{h}_{crit} \leq \lambda_\alpha\right) \geq 1 - \alpha, \tag{6}$$

which can be recast as

$$\widehat{P} = P\left(\widehat{h}^*_{crit} \leq \lambda_\alpha \, \widehat{h}_{crit}\right) \geq 1 - \alpha. \tag{7}$$

This is almost equivalent to (4) except that the inequality sign has been reversed and $\lambda_\alpha = 1$.

The calibration factor λ_α is used because Hall and York (2001) show that the distribution of $\widehat{U} = P\left(\widehat{h}^*_{crit} \leq \widehat{h}_{crit}\right)$ is not uniform on the interval $(0,1)$. By reformulating the test as (6), Hall and York (2001) are able to account for the non-uniformity of the distribution of $\widehat{U}$ and achieve level accuracy. Specifically, they show that the bootstrap distribution function $\widehat{G}_n(\lambda) = P\left(\widehat{h}^*_{crit}/\widehat{h}_{crit} \leq \lambda\right)$ converges in probability to a stochastic process, the distribution of which is independent of unknowns. This property allows them to determine λ_α uniquely for every α by using a rational polynomial approximation based off of Monte Carlo simulations. Hall and York (2001) calculated the size of the Silverman test for $\alpha = 0.001, .0002, \ldots, 0.999$ and fit λ_α

through the plot of the desired size versus the actual size. Given that the stochastic process does not depend on unknowns, this is an applicable approach to calibration of the Silverman test. To adapt this calibration procedure to that of Silverman (1981), the same algorithm is followed by using (3) but evaluating the bootstrapped estimates using $\lambda_\alpha \widehat{h}_{crit}$. In practice, Hall and York (2001) suggest (their Method 2) that the values for λ_α be calculated via Monte Carlo simulations for a sample equal to the size of the dataset $\lambda_\alpha(n)$. For the test being proposed in this paper, one would simulate from a data generating process which would produce a unimodal density of coefficient estimates in order to determine the appropriate value of $\lambda_\alpha(n)$.

3 The Method

In order to implement the proposed test one needs both a good method for nonparametric regression (note that the method can also be applied to nonlinear parametric models) and a good bandwidth selection criteria. Thus it is suggested that one use the well-known local-linear regression estimator (Fan and Gijbels 1992) with a bandwidth selection criteria powered by a cross-validation procedure (Härdle 1993). Local-linear estimation estimates both the conditional mean ($m(x,h)$) and derivative ($\beta(x,h)$) simultaneously. The estimators of the conditional mean and derivative will be denoted by $\widehat{m}(x,h)$ and $\widehat{\beta}(x,h)$, respectively, where h is the bandwidth. Using $\widehat{\beta}(x,h)$, the method is:

1. Estimate $\widehat{\beta}(x,h_{cv})$ by local-linear regression where h_{cv} is the bandwidth selected by using a cross-validation procedure.

2. Find the critical bandwidth h_{crit}, say the smallest h such that the density of $\widehat{\beta}(x,h_{cv})$, i.e. $\widehat{f}_{h_{crit}}\left(\widehat{\beta}(x,h_{cv})\right)$, has k modes.

3. For $i = 1, 2, \ldots, n$, calculate $\widehat{\varepsilon}_i = y_i - \widehat{m}(x_i, h_{cv})$. Generate a bootstrap sample $\widehat{\varepsilon}_1^*, \widehat{\varepsilon}_2^*, \ldots, \widehat{\varepsilon}_n^*$ from $\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \ldots, \widehat{\varepsilon}_n$ and hence a bootstrap data set $y_i^* = \widehat{m}(x_i, h_{cv}) + \widehat{\varepsilon}_i^*$, $i = 1, 2, \ldots, n$.

4. Using the bootstrap data set, estimate $\widehat{\beta}^*(x, h_{cv})$ by local-linear regression with bandwidth h_{cv}.

5. Using h_{crit}, see if $\widehat{f}^*_{h_{crit}}\left(\widehat{\beta}^*(x, h_{cv})\right)$ has k modes.

6. Repeat steps 3-5 a large number (B) of times.

7. Construct the p-value by determining the proportion of estimates which have more than k modes.

The careful reader will notice that the main departure from Silverman (1981) is that the data in question ($\beta(x, h)$) is unknown and thus must be estimated (see Guerre, Perrigne and Vuong 2000 for an example of kernel estimation with estimated quantities). Therefore, instead of bootstrapping from $\widehat{\beta}(x, h_{cv})$, it is important to bootstrap from the residuals of the regression ($\widehat{\varepsilon}_i$) and re-estimate $\beta(x, h_{cv})$ in each replication. This allows one to to approximate the sampling distribution of the derivative.

The calibrated version of the test will modify the method in two ways. First, as in Hall and York (2001), the null of unimodality versus the alternative of multimodality is the only test feasible given the procedure used to estimate the calibration parameter. Higher level tests are not possible given the Monte Carlo approach used to calculate $\lambda_\alpha(n)$. Thus k is restricted to be equal to unity. Second, given that the expected number of modes from the bootstrap density is higher than the expected number of modes from the true density, in step five, h_{crit} will be replaced with $\lambda_\alpha(n)h_{crit}$ wherever it occurs.

4 Implementation

The method discussed above is relatively simple, but in order to implement the procedure, more discussion is necessary. This section will outline a suggested implementation of the above method. First, local-linear estimation of the regression function can be completed by using generalized kernel estimation. The benefit of this approach is that it allows for smoothing of both continuous and categorical (ordered and unordered) variables. Second, the choice of bandwidth for the local-linear regression by a popular cross-validatory procedure is discussed. Although many procedures are similar as the sample size tends towards infinity, some perform better in finite samples than others. Therefore the AIC_c bandwidth selection criteria is suggested. Third, estimating the kernel density of a vector of derivative estimates is briefly discussed. Finally, the smoothed bootstrap used in this paper is outlined. This is important because the true parameters are unknown and we must bootstrap from the residuals and re-estimate the model in each replication in order to simulate the sampling distribution of the parameter estimates.

4.1 Generalized Kernel Estimation

In this subsection Li-Racine Generalized Kernel Estimation (Li and Racine 2004; Racine and Li 2004) is described. It will be used in order to estimate the conditional mean and gradient. First, consider the nonparametric regression model

$$y_i = m(x_i) + \varepsilon_i, \qquad i = 1, ..., n \tag{8}$$

where y_i is the dependent variable measured for observation i. m is the unknown smooth function with argument $x_i = [x_i^c, x_i^u, x_i^o]$, x_i^c is a vector of continuous regressors, x_i^u is a vector of regressors that assume unordered discrete values, x_i^o is a vector of regressors that assume ordered discrete values, ε is an additive error, and n is the number of observations. Taking a first-order Taylor expansion of (8) with respect to x_j yields

$$y_i \approx m(x_j) + (x_i^c - x_j^c)\beta(x_j) + \varepsilon_i \tag{9}$$

where $\beta(x_j)$ is defined as the partial derivative of $m(x_j)$ with respect to x^c. The parameter $\beta(x_j)$ is interpreted as a varying coefficient.

The estimator of $\delta(x_j) \equiv \binom{m(x_j)}{\beta(x_j)}$ is given by

$$\begin{aligned} \widehat{\delta}(x_j) &= \begin{pmatrix} \widehat{m}(x_j) \\ \widehat{\beta}(x_j) \end{pmatrix} = \left[\sum_{i=1}^{n} K_{\widehat{h}} \begin{pmatrix} 1 & (x_i^c - x_j^c) \\ (x_i^c - x_j^c) & (x_i^c - x_j^c)(x_i^c - x_j^c)' \end{pmatrix} \right]^{-1} \\ & \sum_{i=1}^{n} K_{\widehat{h}} \begin{pmatrix} 1 \\ (x_i^c - x_j^c) \end{pmatrix} y_i, \end{aligned} \tag{10}$$

where $K_{\widehat{h}} = \prod_{s=1}^{q_u} l^u \left(x_{sj}^u, x_{si}^u, \lambda_s^u\right) \prod_{s=q_u+1}^{q_u+q_o} l^o \left(x_{sj}^o, x_{si}^o, \lambda_s^o\right) \prod_{s=q_u+q_o+1}^{q_u+q_o+q_c} \frac{1}{\lambda_s^c} l^c \left(\frac{x_{sj}^c - x_{si}^c}{\lambda_s^c}\right)$. K_h can be constructed by generalizing the standard product kernel (see Pagan and Ullah 1999, and Racine and Li 2007). l^c is the standard normal kernel function with bandwidth λ_s^c associated with the s^{th} component of x^c, l^u is a variation of Aitchison and

Aitken's (1976) kernel function with bandwidth λ_s^u associated with the s^{th} component of x^u, and l^o is the Wang and Van Ryzin (1981) kernel function with window width λ_s^o associated with the s^{th} component of x^o. Formally, the kernel function for continuous variables is given by

$$l^c\left(\frac{x_{sj}^c - x_{si}^c}{\lambda_s^c}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_{sj}^c - x_{si}^c}{\lambda_s^c}\right)^2\right); \tag{11}$$

the kernel function for unordered categorical variables is given by

$$l^u\left(x_{si}^u, x_{sj}^u, \lambda_s^u\right) = \begin{cases} 1-\lambda_s^u & \text{if } x_{si}^u = x_{sj}^u \\ \frac{\lambda_s^u}{d_s-1} & \text{otherwise} \end{cases}; \tag{12}$$

where d_s is the number of unique values x_s^u can take (e.g., if x_s^u is binary, $d_s = 2$) and the kernel function for ordered categorical variables is given by

$$l^o\left(x_{si}^o, x_{sj}^o, \lambda_s^o\right) = \begin{cases} 1-\lambda_s^o & \text{if } x_{si}^o = x_{sj}^o \\ \frac{1-\lambda_s^o}{2}(\lambda_s^o)^{|x_{si}^o - x_{sj}^o|} & \text{otherwise} \end{cases}. \tag{13}$$

4.2 Bandwidth Selection

Estimation of the bandwidths $h = (\lambda^c, \lambda^u, \lambda^o)$ is typically the most salient factor when performing nonparametric estimation. For example, choosing a very small bandwidth means that there may not be enough points for smoothing and thus we may get an undersmoothed estimate (low bias, high variance). On the other hand, choosing a very large bandwidth, we may include too many points and thus get an oversmoothed estimate (high bias, low variance). This trade-off is a well-known dilemma in applied nonparametric estimation and thus one usually resorts to automatic determination procedures to estimate the bandwidths. Although there exist many selection methods, one increasingly popular method is Hurvich, Simonoff and Tsai's (1998) Expected Kullback Leibler (AIC_c) criteria. This method chooses smoothing parameters using an improved version of a criterion based on the Akaike Information Criteria. AIC_c has been shown to perform well in small samples and avoids the tendency to undersmooth as often happens with other approaches such as Least-Squares Cross-Validation. Specifically, the bandwidths are chosen to minimize

$$AIC_c(\lambda^c, \lambda^u, \lambda^o) = \log\left(\widehat{\sigma}^2\right) + \frac{1 + tr(H)/n}{1 - \left[tr(H) + 2\right]/n}, \tag{14}$$

where

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{1}{n}\sum_{j=1}^{n}\left(y_j - \widehat{m}(x_j)\right)^2 \\ &= \left(\frac{1}{n}\right) y'(I-H)'(I-H)y, \end{aligned} \tag{15}$$

where I is an identity matrix of dimension n and $\widehat{m}(x_j) = Hy_j$.

A nice feature of this cross-validation procedure is that it does not require the use of a leave-one-out estimator. While this offers no additional gain in computation time – there are still n^2 calculations to be performed – the intuitive appeal of this method is attractive. The criterion is composed of two distinct parts, one that rewards fit and another that penalizes fit. The objective is not simply to interpolate the data by connecting all of the points together. Rather, it is concerned with how the estimate of the function predicts the counterfactuals. A model that fits the data well may not be the best model for constructing counterfactuals and so the AIC_c criteria punishes bandwidths that are interpolating the data rather than determining the underlying population data generating process. The set of bandwidths that minimize the AIC_c function are those that are utilized in the final estimation. As the sample size grows and the number of regressors increases, computation time increases dramatically. However, it is highly recommended that one use a bandwidth selection procedure as opposed to a rule of thumb selection, especially in the presence of discrete data as no rule of thumb exists.

One potential complication exists for these type of bandwidth selectors. The local-linear method with AIC_c bandwidth selection is capable of detecting linearity (Hall, Li and Racine 2007). In this case, the gradient will have point mass at the (global) OLS estimator. Kernel methods for density estimation will fail in this case. Given that such cases occur with positive measure, it is important to address this point. Fortunately, if in fact the estimates are all equal, the test for multimodality is unnecessary as the conclusion is obvious.

4.3 Estimation of the Density of the Derivative

Here the modality of $\beta(x, h)$, and not x, is in question. This subsection discusses how to construct an estimate of $f\left(\widehat{\beta}(x, h_{cv})\right)$. Plotting kernel densities of predicted values and/or derivatives is common practice in applied nonparametric estimation and the procedure is analogous to that for a simple vector of data. Let $\widehat{\beta}_i = \widehat{\beta}(x_i, h_{cv})$, then the kernel density estimate for the estimated derivative is defined as

$$\widehat{f}_h\left(\widehat{\beta}\right) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{\widehat{\beta}_i - \widehat{\beta}}{h}\right), \tag{16}$$

where again h is the bandwidth and $K(\cdot)$ is the kernel function. The proof of the monotonicity of the bandwidth in the number of modes in (16) follows from Silverman (1981).

4.4 Smoothed Bootstrap

Here we discuss the bootstrap procedure necessary in order to simulate the sampling distribution of the parameter estimates. Let $\widehat{\varepsilon}_i = (\widehat{\varepsilon}_1, \widehat{\varepsilon}_2, \ldots, \widehat{\varepsilon}_n)'$, where $\widehat{\varepsilon}_i = y_i - \widehat{m}(x_i, h_{cv})$, and $\widehat{m}(x, h_{cv})$ is the local-linear estimator of $m(x, h)$ with bandwidth h_{cv}. Compute the smoothed bootstrap errors by randomly sampling with replacement from $\widehat{\varepsilon}_i$, thereby obtaining $\widehat{\varepsilon}_i^* = (\widehat{\varepsilon}_1^*, \widehat{\varepsilon}_2^*, \ldots, \widehat{\varepsilon}_n^*)'$. Then generate y_i^* via $y_i^* = \widehat{m}(x_i, h_{cv}) + \widehat{\varepsilon}_i^*$. Call $\{x_i, y_i^*\}_{i=1}^n$ the bootstrap sample. Using the bootstrap sample to estimate $\beta(x, h_{cv})$ via local-linear regression, denote the estimate by $\widehat{\beta}^*(x, h_{cv})$, and then estimate the bootstrap density of the estimated derivative, $\widehat{f}_{h_{crit}}^*\left(\widehat{\beta}^*(x, h_{cv})\right)$. The bootstrap estimate of the density is obtained as before except that $\widehat{\beta}(x, h_{cv})$ is replaced by $\widehat{\beta}^*(x, h_{cv})$ wherever it occurs. This process is repeated a large number (B) of times. The empirical distribution of the B bootstrap estimates is then used to approximate the sampling distribution of the derivative. The finite sample performance of this bootstrap procedure is examined via simulations in Section 5.

5 Simulations

5.1 Uncalibrated Test

The performance of the test in a finite setting is examined in this section. The size is computed by generating a data generating process which gives a unimodal density of the derivative. The actual size of the test is calculated by the proportion of times the null hypothesis of unimodality is rejected. The power is computed by creating a data generating process which gives a bimodal density of the derivative. The actual power is similarly calculated by the proportion of times the null hypothesis of unimodality is rejected.

The data is generated using a quadratic functional form. The difference between the two data generating processes will be the distribution of the independent variable. The uniform distribution of the independent variable leads to densities of the estimates of the partial effects which are (heterogenous but) unimodal and the mixed normal distribution of the independent variable will lead to densities of the estimates of the partial effects which are bimodal. It should be noted that alternative data generating processes do not significantly change the conclusions of the experiment assuming that the process leads to densities of the gradients which possess the appropriate number of modes. The specific form of the technology is

$$y = \beta_0 + \beta_1 x^2 + u, \tag{17}$$

where the intercept term is zero and β_1 takes the value of unity. For the size of the test, x is generated uniformly from negative one to one and u is generated as a standard normal. For the power of the test, x is distributed $N(-1, 1)$ with probability one-half and $N(1, 1)$ with probability one-half. All the remaining variables and parameters are the same as in the unimodal case. Note that, as expected, the power of the test increases as the difference between the means increases as this results in increasing the distance between the modes of the partial effects.

For the Monte Carlo exercise, samples of size $n = 50, 100$, 200, and 616 (the sample size of the application) are generated. AIC_c is used to select the bandwidth in each of the 999 Monte Carlo replications for each sample size. After computing the bandwidth, generalized kernel estimation is used to estimate both the unknown function and its derivative. The kernel density estimates use the Gaussian kernel and the the critical bandwidth h_{crit} is determined by computing kernel density estimates on an equally spaced grid of points, and searching for the smallest bandwidth that yields a unimodal density estimate. For unimodal distributions, the size of the test is estimated by the proportion of times (999 bootstrap replications) the null hypothesis is rejected. A similar approach is used to estimate the power of the test for bimodal distributions.

As common in Silverman-type tests, Table 1 shows that the uncalibrated test is conservative. The test is undersized in each case, and the power of the test is relatively low, even at the ten-percent level. These results are similar, but the size and power are significantly smaller than that found in York (1998) for the uncalibrated Silverman test. This likely reflects the fact that one has to estimate the data whose distribution is in question.

5.2 Calibrated Test

To implement the calibrated forms of the test, the constant λ_α must be specified. A Monte Carlo is used to compute the value of λ_α for each sample size in question ($\lambda_\alpha(n)$). The constant is chosen to produce a test with correct size accuracy for each value of α. This is done by finding the value of lambda for each value of alpha ($\alpha = 0.001, 0.002, \ldots, 0.999$) for a given sample size. Following Hall and York (2001), the values of lambda are then regressed on a seventh order polynomial function of alpha. These are used to obtain fitted values of lambda and hence the specific values of $\lambda_\alpha(n)$ for the corresponding test. This process is computer intensive, but is necessary in order to obtain a test with the appropriate size.

Table 2 gives the results for the calibrated test. The data generating processes used are the same as in the uncalibrated tests and the random seeds are the same (for a given sample size) to ensure a fair comparison. The table shows that for a relatively small sample ($n = 100$), the test gives the appropriate size. Here the size is close to the nominal level. Further, there is a drastic increase in the finite sample power of the test. These results give an indication of the importance of using the calibrated form of the test in practice. It should be noted that using an 'asymptotic' set of λ_α's would not be appropriate here. For example, using the values of $\lambda_\alpha(616)$ in place of $\lambda_\alpha(200)$ lead to results which are oversized when $n = 200$. Thus it is suggested that when employing this test that one calibrate the value of $\lambda_\alpha(n)$ for the nominal level and sample size in question.

6 Nonparametric Growth Regressions

Abramovitz (1986) introduced the notion that under certain conditions, 'backward' countries would tend to grow faster than rich countries, in order to close the gap between the two groups of countries, or to 'catch-up'. Initially envisioning this phenomenon, Abramovitz's argument is based on the discovery of a considerable reduction in the coefficient of variation of growth rates within a group of 16 industrialized countries. Since the publication of his paper, there has been a plethora of empirical research on economic convergence.

One of the most common and exhaustively studied hypotheses is that of β-convergence. This is typically tested by regressing growth rates of output on initial levels of output ($g_{it} = \alpha + \beta y_{bit} + \varepsilon_{it}$ – unconditional convergence), sometimes while controlling for other exogenous variables ($g_{it} = \alpha + \beta y_{bit} + \gamma x_{it} + \varepsilon_{it}$ – conditional convergence). In this setup, a negative regression coefficient is interpreted as an indication of β-convergence (Barro 1991; Baumol 1986). Historically the (conditional) convergence rate is assumed constant across countries, although there is evidence that countries with lower levels of education have slower convergence rates and hence different conditional distributions of growth rates (Jones 1997; Quah 1997).

Another method to study convergence that has become popular in the recent literature is to uncover multimodality in the distribution of output per worker using kernel estimation (Bianchi 1997; Fiaschi and Lavezzi 2003; Henderson and Russell 2005; Johnson 2005) or mixture models (Paap and van Dijk 1998; Pittau 2005; Pittau and Zelli 2006). Loosely speaking, convergence is said to occur if a multimodal distribution of output per worker tends towards unimodality over time, or if the distribution is unimodal, the dispersion of the distribution tends to decline.

This 'bump hunting' approach was partly conceived in order to circumvent the perceived problems with β-convergence. Specifically, β-convergence has been widely criticized because it essentially focuses only on average behavior. Further, most typical growth regressions rely on (linear) parametric models which often assume that the aggregate production function is identical across countries or that it only differs with respect to a country-specific effect. There is plenty of reason to believe that technologies across vastly different economies differ by more than a linear shift.

However, this does not suggest that one should throw the Barro out with the bath water. Semiparametric and nonparametric kernel regression methods have the potential to alleviate some of these shortcomings. A large number of studies have used semiparametric methods to study growth (e.g., see Durlauf, Kourtellos and Minkin 2001, Ketteni, Mamuneas and Stengos 2007, Liu and Stengos 1998, Vaona and Schiavo 2007). Although these studies are able to relax functional form assumptions and lessen the curse of dimensionality, their consistency still depends on restrictive assumptions. As an alternative, Maasoumi, Racine and Stengos (2007) use fully nonparametric generalized kernel estimation to estimate some of the most widely used growth regressions. Specifically, they focus on what happens to predicted growth rates and the residuals over time. This paper will deviate from their focus, but exploit their methodology which allows for interactions among all variables as well as for nonlinearities in and among all variables.

In addition to the added flexibility, nonparametric regression gives observation specific estimates. Hence, countries who differ widely in social, political and institutional characteristics are not required to have the same parameter estimate. This is especially important because parameter heterogeneity is considered to be a severe problem in growth economics (Brock and Durlauf 2001; Temple 1999) and several authors provide strong evidence for widespread heterogeneity (Durlauf and Johnson 1995; Masanjala and Papageorgiou 2004). Further, Durlauf and Quah (1999) argue that linear regressions with homogenous parameters will often fail to distinguish growth models which posses multiple equilibria.

6.1 Data

The data for this study come from Maasoumi, Racine and Stengos (2007) and thus will only be explained briefly. The sample is a panel data set of 88 developed, newly developed and developing countries over seven nonoverlaping five-year periods from 1965 to 1995 ($88 \times 7 = 616$ observations). The dependent variable (g_{it}) is the growth rate of income per capita during each five-year period, y_{bit} is the per capita income at the beginning of each period, h_{it} is human capital, n_{it} is the average annual population growth for each country in each period and k_{it} is the share of output devoted to physical capital accumulation. Initial income estimates are from the Penn World Tables (Summers and Heston 1998) as are the share of output devoted to physical capital for each five year period. The average growth rate of per capita income as well as the average annual population growth each period are from the World Bank. The average years of schooling in the population, ages fifteen and older, are obtained from Barro and Lee (2000). Finally, an unordered categorical variable (dummy variable) is included for OECD status. An ordered categorical variable to control for time effects is also included.

6.2 Results

6.2.1 Unconditional Convergence

Following the methodology in Section 4.1, the estimates for the simplest model, the regression of the growth rate of per capita output solely on the initial period output per capita, are given in Figure 1 and Table 3. These represent the unconditional convergence regressions attributed to Baumol (1986) with a twist. Instead of estimating a single parameter estimate for all countries over all time periods, by using local-linear estimation, a separate coefficient is given for each observation (additionally, the linear model is rejected when employing the test of Hsiao, Li and Racine 2007 – p-value = 0.0000). In other words, it is possible to see who is unconditionally converging and who is unconditionally diverging. The figure plots the kernel density of the parameter estimates while the table reports the mean along with the 25th, 50th and 75th percentile (labeled as $q1$, $q2$, and $q3$) estimates along with their corresponding (999 bootstrapped) standard errors. It is obvious from the graph that the density covers both negative and positive values. In other words, a certain percentage of the sample shows evidence of unconditional convergence, while more than half the sample exhibits evidence of divergence. In fact, there appears to be a single mode for observations with a negative coefficient and possibly two modes for observations with positive coefficients. The null of a unimodal density at the five-percent level cannot be rejected when using the standard (p-value = 0.2915) version of the test, but is firmly rejected with the calibrated (p-value = 0.0012) version of the test. This shows the importance of using the calibration method. The conservative form of the test fails to reject the null of a visually obvious multimodal density.

The question then becomes, who converges and who diverges? These results are shown in Table 3 and Figure 2. Table 3 gives the mean and quartile estimates for different groups of countries while the second figure plots the kernel density estimates of the parameter estimates for selected groups of countries. Consistent with past research, a majority of OECD countries appear to (significantly) converge unconditionally while all other groups appear to diverge. It appears that most OECD countries show behavior consistent with the unconditional β-convergence hypothesis whereas most non-OECD countries do not.

Although this evidence is convincing, one can ask whether the convergence rates for OECD countries are uniformly smaller than those for non-OECD economies. Following Eren and Henderson (2006), stochastic dominance techniques are used to determine if the coefficients for one group are uniformly larger than those for another group. Using this approach, the null that the two distributions of parameter estimates are equal is firmly rejected (p-value = 0.0000). Further, the test is unable to reject the null that the parameter estimates for non-OECD countries first order dominate those for OECD economies (p-value = 0.9879). In other words, the unconditional rate of convergence for OECD economies is uniformly smaller than that for non-OECD countries. Although this is not surprising, it gives a stronger conclusion than that of Table 3.

This result, however, should not be misinterpreted. This does not imply that all OECD economies tend towards convergence or that none of the non-OECD countries are converging. Figure 2 shows that the density of coefficient estimates for OECD economies covers positive regions, and at the same time, the density for non-OECD countries covers negative regions. In fact, of the 416 positive coefficients, 38 belong to OECD economies. Likewise, of the remaining 200 negative coefficients, 77 belong to non-OECD economies. This leads one to question what is it about these countries which makes the coefficients positive or negative. Status in the OECD does not seem to be the sole determinant. Figure 3 gives kernel density plots of the 'typical' Solow variables separated into those with a positive coefficient on the initial income variable and those with a negative coefficient on the initial income variable. It is obvious from the figure that counties with higher initial income, lower population growth, higher investment and higher levels of human capital are more likely to have a negative coefficient on the initial income variable.

Several authors noting these same phenomenon have tried to determine if there are thresholds which exist such that once countries cross this threshold, convergence can occur. Liu and Stengos (1999) suggest that convergence only holds for countries with an initial per capital GDP above $1800. Without controlling for other attributes, the results here find evidence of negative coefficients on initial income for per capita incomes as low as $312. This finding is related to Sala-i-Martin (2006) who finds evidence of convergence even at very low levels of income per capita. In fact, panel (a) of Figure 3 shows a large overlap at lower levels of initial income per capita. However, for higher levels of initial income per capita, there appears to be only evidence of convergence. Thus, what this figure suggests is that there is an initial level of income per capita which is sufficient (but not necessary) for convergence. This amount is somewhere around $5000 per capita. In contrast to previous studies, here a bound is found for where only convegence exists as opposed to a minimum bound to where convergence is feasible.

It should be noted that most of the remaining panels have a large amount of overlap. The set of population growth rates corresponding to the negative coefficients are a proper subset of the growth rates corresponding to the positive coefficients. Although it is obvious from the figure that the average population growth rate is lower for countries with negative coefficients, there can be no discussion of thresholds here. Similarly, for investment, the values for countries with negative coefficients on initial income are bounded by the values for countries with positive coefficients. Again, the average is larger for countries with negative coefficients, but there do not appear to be any necessary or sufficient thresholds for convergence with respect to investment.

Finally, for human capital, the range associated with negative coefficients is not contained by the values associated with positive coefficients. This occurs for large values of human capital. In fact, for countries where the average number of years of education is in excess of 9.70, there is only evidence of convergence. Again, this implies the possibility of a sufficient (but not necessary) threshold, but this only applies to a very small percentage of the sample.

6.2.2 Conditional Convergence

The above results leave out several important inputs in the regression model. Theory only states that the convergence hypothesis will hold among countries that have the same steady state. Once we control for the important covariates, we should see countries converge. This notion of conditional convergence (Barro and Sala-i-Martin 1992; Mankiw, Romer and Weil 1992) is shown in Table 4, panel (a) of Figure 4 and Figure 5. Table 4 gives the mean and quartile values for the estimates on the base period income for all countries as well as for several different groups of countries. In contrast to Table 3, now controlling for other factors (e.g., physical and human capital), all groups of countries show a majority of evidence of convergence. In addition, the empirical density of the parameter estimates for all observations appears to be unimodal. This is confirmed by the proposed test which is unable to reject the null hypothesis of unimodality at any conventional level using either the uncalibrated test (p-value = 0.9899) or the calibrated test (p-value = 0.9698).

However, this says nothing about the relative speed of convergence for different groups. That being said, by using stochastic dominance techniques to examine the convergence rates between OECD and non-OECD economies, the null that the two distributions of parameter estimates are equal cannot be rejected (p-value = 0.8150).

One should be careful with the above results. Although we find the first and third quartiles to be negative for all groups of countries considered, this still represents only half the observations for each classification. Figure 4 shows that a non-trivial percentage of countries still have positive coefficients on initial capital. These results, however, should not be surprising. Kalaitzidakis, Mamuneas, Savvides and Stengos (2001) and Liu and Stengos (1999) each find some positive partial effects on initial income when controlling for other variables in their respective semiparametric models. One may think that these solely belong to non-OECD economies, but Figure 5 reveals that all groups of countries have kernel densities which have observations in the positive region. In fact, 26% of the 110 observations with positive coefficients belong to OECD countries.

Given the analysis in the previous section, it seems obvious to analyze the differences between observations with positive coefficients and those with negative coefficients. Figure 6 plots the kernel density estimates of each of the continuous regressors split between those which have positive partial effects on initial income and those which have negative coefficients on initial income. In contrast to the third figure, here there is almost a complete overlap between the two distributions in each case. There appears to be no discernible difference between the two groups of observations based solely on their corresponding conditioning variables. The Li (1996) test for differences between two unknown distributions is unable to reject the null that each of the distributions are equal. All p-values are in excess of 0.15. In comparison, each null is rejected in the unconditional convergence case (all p-values being less than zero to four decimal places). Deeper examination shows that these positive partial effects are not more prominent in any particular region or time period. This result suggests that even with nonparametric estimation, the 'typical' Solow variables are not sufficient to explain convergence for all groups of countries. In fact, even 18% of the OECD countries have positive partial effects on initial income when controlling for the standard variables. Thus, controlling for possible model misspecification by using nonparametric methods is not sufficient to explain convergence and it is likely that additional variables are necessary in order to explain convergence for the remaining observations.

That being said, separate interest lies in the results for the remaining continuous regressors. The mean parameter value for human and physical capital each (shown in Table 5) take their expected signs. However, while the coefficients for the share of physical capital are significant, those of human capital and population growth are primarily insignificant. Although this result is common in empirical studies (Temple 1999), it is not necessarily expected here. One of the main criticisms of typical (linear) parametric growth regressions is the failure to address specification bias (e.g., failure to account for nonlinearities and important interactions between regressors). While nonparametric regression allows for interactions and nonlinearities among all variables, it is unable to identify a statistically significant relationship for population growth or human capital. These finding are not necessarily reasons to reject the nonparametric model (again the Hsiao, Li and Racine 2007 test rejects the null of linearity – p-value = 0.0000), but may simply be due to the use of poor proxies (e.g., years of schooling), due to bias related to a missing variable (e.g., institutions), or simply due to the relatively small sample employed.

Although two of the regressors are predominantly insignificant, the results for each are interesting. The results here are in line with Mamuneas, Savvides and Stengos (2006) in that there is evidence of large variations in the return to human capital across countries and time. Also, as in their semiparametric model, the majority of the partial effects on human capital are insignificant. Different from the other variables, the density of the returns to human capital appears to be multimodal, but this cannot be confirmed with either test. This result is consistent with that found in Henderson, Parmeter and Russell (2006), who find evidence of unimodality in the density of the returns to human capital before 1990, but evidence of multimodality in the density in 1990 and 2000.

The results for population growth are equally interesting. Although the popular view is that population growth is a detriment to economic growth (Kelley 1988), it is not well supported by the data. In fact, the result in panel (b) of Figure 4 shows that the result is negative only for a little over half the sample (both tests fail to reject unimodality). Closer examination shows that for non-OECD economies, increases in population growth lead to increases in the GDP growth rate when the population growth rate is small, while increases in the population growth rate lead to decreases in the GDP growth rate when the population growth rate is large. However, in the case of OECD countries, increases in the rate of population growth tend to hamper economic growth. These results are the same as those found in Maasoumi, Racine and Stengos (2007).

It should be noted that for the parameter estimates of any particular regressor, the remaining covariates are not constrained to be equal across or within groups. That being said, there may be some interest to examine this phenomenon. It is possible that there can be an effect of other covariates on a nonparametric gradient. You could get multimodality due to movements in other regressors. However, if you control for the presence of the other regressors (e.g., holding them constant at, say, their median) you are now in the uncomfortable position of having a multimodal gradient for one value of the covariates held constant but not for others. This possibility suggests a nontrivial future research agenda.

In summary, this nonparametric regression study of β-convergence is able to uncover several findings: (1) In the estimation of unconditional β-convergence, there is significant parameter heterogeneity with a large percentage of the estimates being negative, but a majority are positive. (2) The proposed test provides evidence that the density of the parameter estimates is multimodal. (3) An overwhelming majority of the negative parameter estimates belong to OECD economies while all other groups of countries possess primarily positive estimates. This result is consistent with the conclusions of past research that suggest that unconditional β-convergence does not hold for all countries. (4) In the estimation of conditional β-convergence, parameter heterogeneity still exists and the values lie on both sides of zero, but a majority of the observations have negative parameter estimates. The negative partial effects on initial income are consistent with past theories on conditional convergence. (5) Here the hypothesis that the density of parameter estimates is unimodal cannot be rejected at any conventional level. Finally, (6) the study looked at the characteristics of observations with positive coefficients on initial income. The results for unconditional convergence pointed to the usual suspects, but little information was available in terms of determining why some of the countries possessed positive coefficients when controlling for the 'typical' Solow variables.

7 Conclusion

In this paper, a method to test for multimodality of an estimated kernel density of parameter estimates from a local-linear least-squares regression derivative is presented. The procedure is laid out in seven simple steps and a suggestion for implementation is proposed. The finite sample performance of the test is analyzed with a Monte Carlo study along with a calibrated version of the test which corrects for the conservative nature of Silverman-type tests.

The proposed test is included in a study on nonparametric growth regressions. The results show that in the estimation of unconditional β-convergence, the density of the parameter estimates is multimodal with one mode in the negative region and possibly two modes in the positive region of the parameter estimates. The negative results primarily correspond to OECD countries and confirm past evidence that not all economies converge unconditionally. The results for conditional β-convergence show that the density is predominantly negative and unimodal.

Although most of the results found were in line with theory, many in the conditional convergence case need further examination. A non-trivial percentage of the partial effects on initial income were still positive. In fact, over fifteen percent of the OECD economies had positive coefficients, even after controlling for the 'typical' Solow variables. This shows that the added flexibility of the nonparametric methods is not sufficient to explain convergence itself with the standard variables. It appears that further research in nonparametric growth needs to incorporate additional explanatory variables which are known to affect growth.

References

[1] Abramovitz, M. (1986), "Catching Up, Forging Ahead and Falling Behind," *Journal of Economic History*, 46, 386-406.

[2] Aitchison, J. and C. G. G. Aitken (1976), "Multivariate Binary Discrimination by Kernel Method," *Biometrika*, 63, 413-420.

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

[4] Barro R. J. and J.-W. Lee (2001), "International Data on Educational Attainment: Updates and Implications," *Oxford Economic Papers,* 3, 541-563.

[5] Barro, R., and X. Sala-i-Martin (1992), "Convergence," *Journal of Political Economy*, 100, 223-251.

[6] Baumol, W. J. (1986), "Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show," *American Economic Review,* 76, 1072-1085.

[7] Brock, W. A., and S. N. Durlauf (2001), "What Have we Learned from a Decade of Empirical Research on Growth? Growth Empirics and Reality ," *World Bank Economic Review*, 15, 229-272.

[8] Bianchi, M. (1997), "Testing for Convergence: Evidence from Non-Paramteric Multimodality Tests," *Journal of Applied Economics,* 12, 393-409.

[9] Bowman, A. W., M. C. Jones, and I. Gijbels (1998), "Testing Monotonicity of Regression," *Journal of Computational and Graphical Statistics*, 7, 489-500.

[10] Cheng, M.-Y. and P. Hall (1998), "Calibrating the Excess Mass and Dip Tests of Modality," *Journal of the Royal Statistical Society, Series B*, 60, 579-590.

[11] De Long, J. B. (1988), "Productivity Growth, Convergence, and Welfare: Comment," *American Economic Review,* 78, 1138-1154.

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

[13] Durlauf, S. N., A. Kourtellos, and A. Minkin (2001), "The Local Solow Growth Model," *European Economic Review*, 45, 928-940.

[14] Durlauf, S. N. and D. T. Quah (1999), "The New Empirics of Economic Growth," in the *Handbook of Macroeconomics*, J. Taylor and M. Woodford (Eds.), North-Holland.

[15] Eren, O. and D. J. Henderson (2006), "The Impact of Homework on Student Achievement," *Econometrics Journal*, forthcoming.

[16] Fan, J. and I. Gijbels (1992), "Variable Bandwidth and Local Linear Regression Smoothers," *Annals of Statistics,* 20, 2008-2036.

[17] Fiaschi, D. and A. M. Lavezzi (2003), "Distribution Dynamics and Nonlinear Growth," *Journal of Economic Growth*, 8, 379-401.

[18] Galor, O. (1996), "Convergence? Inferences from Theoretical Models," *Economic Journal,* 106, 1056-1069.

[19] Guerre, E., I. Perrigne, and Q. Vuong (2000), "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica,* 68, 525-574.

[20] Hall, P., Q. Li and J. Racine (2007), "Nonparametric Estimation of Regression Functions in the Presence of Irrelevant Regressors," *Review of Economics and Statistics*, 784-789.

[21] Hall, P. and M. York (2001), "On the Calibration of Silverman's Test for Multimodality," *Statistica Sinica*, 11, 515-536.

[22] Härdle W. (1993), *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

[23] Henderson, D. J., C. F. Parmeter, and R. R. Russell (2006), "Modes, Weighted Modes and Calibrated Modes: Evidence of Clustering Using Modality Tests," *Journal of Applied Econometrics*, forthcoming.

[24] Henderson, D. J. and R. R. Russell (2005), "Human Capital and Convergence: A Production-Frontier Approach," *International Economic Review,* 46, 1167-1205.

[25] Hsiao, C., Q. Li, and J. Racine (2007), "A Consistent Model Specification Test with Mixed Discrete and Continuous Data," *Journal of Econometrics*, 140, 802-826.

[26] Hurvich, C. M., J. S. Simonoff, and C.-L. Tsai (1998), "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion," *Journal of the Royal Statistical Society, Series B*, 60*,* 271-293.

[27] Islam, N. (1995), "Growth Empirics: A Panel Data Approach," *Quarterly Journal of Economics*, 110, 1127-1170.

[28] Johnson, P. (2005), "A Continuous State Space Appraoch to "Convergence by Parts"," *Economics Letters,* 86, 317-321.

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

[30] Kalaitzidakis, P., T. P. Mamuneas, A. Savvides, and T. Stengos (2001), "Measures of Human Capital and Nonlinearities in Economic Growth," *Journal of Economic Growth*, 6, 229-254.

[31] Kelly, A. C. (1988), "Economic Consequences of Population Change in the Third World," *Journal of Economic Literature,* 26, 1685-1728.

[32] Ketteni, E., T. P. Mamuneas, and T. Stengos (2007), "Nonlinearities in Economic Growth: A Semiparametric Approach Applied to Information Technology Data," *Journal of Macroeconomics*, 29, 555-568.

[33] Li, Q. (1996), "Nonparametric Testing of Closeness between Two Unknown Distribution Functions," *Econometric Reviews*, 15, 261-274.

[34] Li, Q. and J. Racine (2004), "Cross-Validated Local Linear Nonparametric Regression," *Statistica Sinica*, 14, 485-512.

[35] Li, Q. and J. Racine (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton.

[36] Liu, Z. and T. Stengos (1999), "Non-linearities in Cross-country Growth Regressions: A Semiparametric Approach," *Journal of Applied Econometrics*, 14, 527-538.

[37] Maasoumi, E., J. Racine, and T. Stengos (2007), "Growth and Convergence: A Profile of Distribution Dynamics and Mobility," *Journal of Econometrics*, 136, 483-508.

[38] Mamuneas, T. P., A. Savvides, and T. Stengos (2006), "Economic Development and the Return to Human Capital: A Smooth Coefficient Semiparametric Approach," *Journal of Applied Econometrics*, 21, 111-132.

[39] Masanjala, W. and C. Papageorgiou (2004), "The Solow Model with CES Technology: Nonlinearities and Parameter Heterogeneity," *Journal of Applied Econometrics,* 19, 171-201.

[40] Mammen, E., J. S. Marron, and N. I. Fisher, (1992), "Some Asymptotics for Multimodality Tests Based on Kernel Density Estimates," *Probability Theory and Related Fields*, 91, 115-132.

[41] Mankiw, N. G., D. Romer, and D. Weil (1992), "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics,* 107*,* 407-37*.*

[42] Paap, R. and H. K. van Dijk (1998), "Distribution and Mobility of Wealth of Nations," *European Economic Review,* 42, 1269-1293.

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

[44] Pittau, M. G. (2005), "Fitting Regional Income Distributions in the European Union," *Oxford Bulletin of Economics and Statistics,* 67, 135-161.

[45] Pittau, M. G. and R. Zelli (2006), "Empirical Evidence of Income Dynamics Across EU Regions," *Journal of Applied Econometrics,* 21, 605-628.

[46] Quah, D. T. (1993a), "Empirical Cross-Section Dynamics in Economic Growth," *European Economic Review*, 37, 426-434.

[47] Quah, D. T. (1993b), "Galton's Fallacy and Tests of the Convergence Hypothesis," *Scandinavian Journal of Economics*, 95, 427-443.

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

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

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

[51] Racine, J. and Q. Li (2004), "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data," *Journal of Econometrics,* 119, 99-130.

[52] Sala-i-Martin, X. (1996), "The Classical Approach to Convergence Analysis," *Economic Journal,* 106, 1019-1036.

[53] Sala-i-Martin, X. (2006), "The World Distribution of Income: Falling Poverty and ... Convergence, Period," *Quarterly Journal of Economics*, 121, 351-397.

[54] Silverman, B. W. (1981), "Using Kernel Density Estimates to Investigate Multimodality," *Journal of the Royal Statistical Society, Series B*, 43, 97-99.

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

[56] Summers, R. and A. Heston (1988), "A New Set of International Comparisons of Real Product and Prices: Estimates for 130 Countries," *Review of Income and Wealth,* 34, 1-26.

[57] Temple, J. (1999), "The New Growth Evidence," *Journal of Economic Literature,* 37, 112-156.

[58] Vavona, A. and S. Schiavo (2007), "Nonparametric and Semiparametric Evidence on the Long-Run Effects of Inflation on Growth," *Economics Letters*, 94, 452-458.

[59] Wang, M. C. and J. Van Ryzin (1981), "A Class of Smooth Estimators for Discrete Estimation," *Biometrika*, 68, 301-309.

[60] York, M. G. (1998), "Some Problems in Testing for Modality," PhD Thesis, Australian National University.

	α	n = 50	100	200	616
	0.01	0.0000	0.0000	0.0000	0.0000
Size	0.05	0.0201	0.0050	0.0000	0.0000
	0.10	0.0603	0.0251	0.0000	0.0000
	0.01	0.0162	0.0048	0.0059	0.0434
Power	0.05	0.2942	0.1926	0.1934	0.3089
	0.10	0.5410	0.4885	0.4996	0.5821

Table 1 – Simulated size and power for the uncalibrated test

	α	n			
		50	100	200	616
	0.01	0.0060	0.0120	0.0096	0.0103
Size	0.05	0.0361	0.0461	0.0481	0.0503
	0.10	0.1182	0.1162	0.0961	0.1020
	0.01	0.1947	0.8366	0.9762	0.9944
Power	0.05	0.6630	0.8972	0.9887	1.0000
	0.10	0.8298	0.9160	0.9891	1.0000

Table 2 – Simulated size and power for the calibrated test: The values for $\lambda_\alpha(n)$ are calibrated via Monte Carlo simulations for each alpha, sample size combination.

	mean	$q1$	$q2$	$q3$
All	0.0039	-0.0074	0.0054	0.0157
	0.0037	0.0035	0.0039	0.0041
OECD	-0.0106	-0.0197	-0.0150	-0.0029
	0.0034	0.0042	0.0035	0.0036
non-OECD	0.0091	0.0048	0.0076	0.0201
	0.0041	0.0041	0.0041	0.0043
Africa	0.0121	0.0051	0.0157	0.0222
	0.0059	0.0040	0.0041	0.0056
Latin America	0.0065	0.0048	0.0057	0.0103
	0.0041	0.0041	0.0040	0.0041
Asia	0.0087	0.0049	0.0090	0.0191
	0.0041	0.0040	0.0041	0.0042
Middle East	0.0070	0.0040	0.0055	0.0122
	0.0080	0.0088	0.0039	0.0069

Table 3 – Nonparametric estimates of growth on initial income per capita (without controls) by group: $q1$, $q2$, and $q3$ refer to the first, second and third quartile, respectively, of the distribution of the estimated coefficients. AIC_c used for bandwidth selection. Standard errors are listed below each estimate and are obtained via bootstrapping.

	mean	$q1$	$q2$	$q3$
All	-0.0070	-0.0126	-0.0087	-0.0036
	0.0044	0.0050	0.0047	0.0050
OECD	-0.0064	-0.0131	-0.0087	-0.0031
	0.0038	0.0044	0.0047	0.0050
non-OECD	-0.0072	-0.0125	-0.0088	-0.0040
	0.0037	0.0043	0.0033	0.0044
Africa	-0.0084	-0.0129	-0.0093	-0.0060
	0.0050	0.0035	0.0045	0.0044
Latin America	-0.0071	-0.0137	-0.0087	-0.0018
	0.0044	0.0035	0.0047	0.0064
Asia	-0.0057	-0.0119	-0.0085	-0.0036
	0.0043	0.0032	0.0044	0.0044
Middle East	-0.0088	-0.0138	-0.0087	-0.0052
	0.0032	0.0035	0.0047	0.0056

Table 4 – Nonparametric estimates of growth on initial income per capita (with controls) by group: $q1$, $q2$, and $q3$ refer to the first, second and third quartile, respectively, of the distribution of the estimated coefficients. AIC_c used for bandwidth selection. Standard errors are listed below each estimate and are obtained via bootstrapping.

	mean	$q1$	$q2$	$q3$
y_b	-0.0070	-0.0126	-0.0087	-0.0036
	0.0044	0.0050	0.0047	0.0050
n	-0.0151	-0.0389	-0.0145	0.0045
	0.0210	0.0254	0.0196	0.0230
k	0.0269	0.0199	0.0268	0.0357
	0.0086	0.0053	0.0066	0.0066
h	0.0038	0.0009	0.0045	0.0084
	0.0032	0.0041	0.0040	0.0068

Table 5 – Nonparametric estimates for all continuous variables: $q1$, $q2$, and $q3$ refer to the first, second and third quartile, respectively, of the distribution of the estimated coefficients for each of the variables. AIC_c used for bandwidth selection. Standard errors are listed below each estimate and are obtained via bootstrapping.

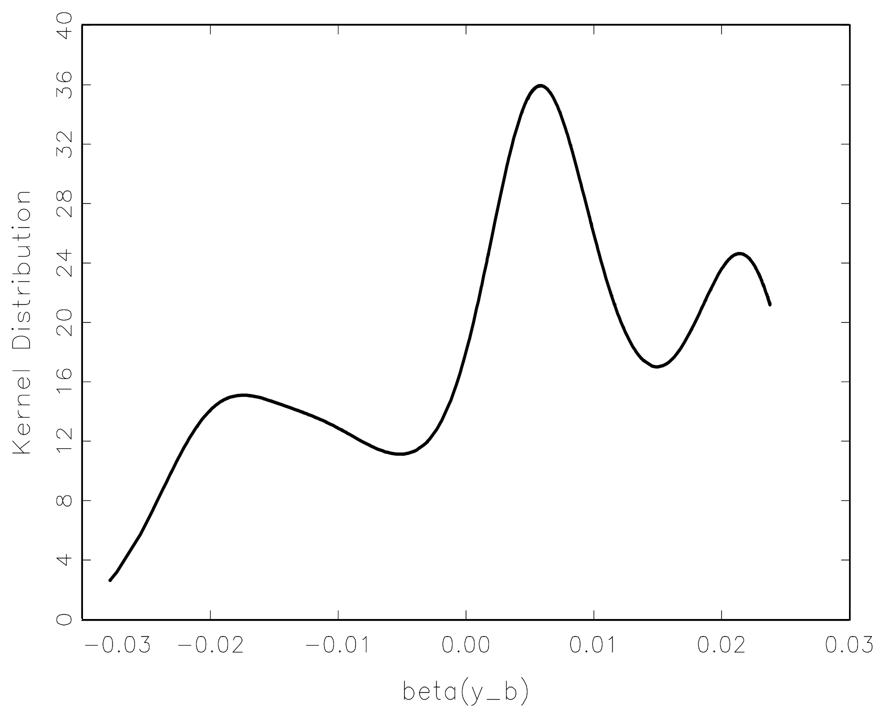

Figure 1: Kernel density estimate of the slope coefficients from the nonparametric regression of growth on initial income per capita (without controls)

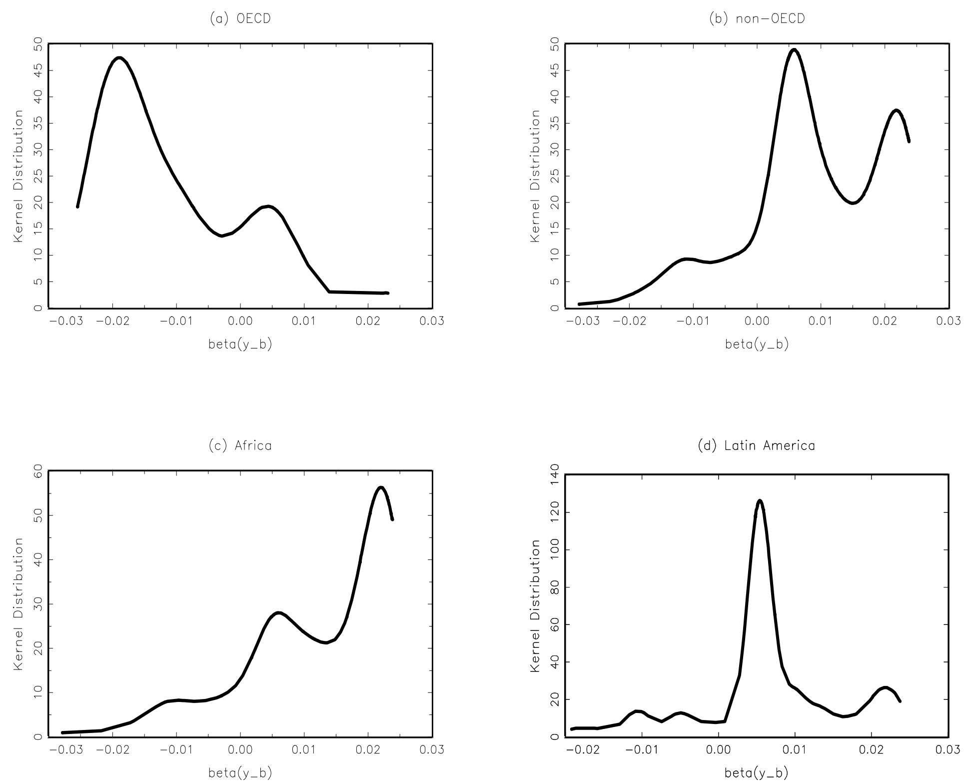

Figure 2: Kernel density estimates of the slope coefficients for select groups of countries from the nonparametric regression of growth on initial income per capita (without controls)

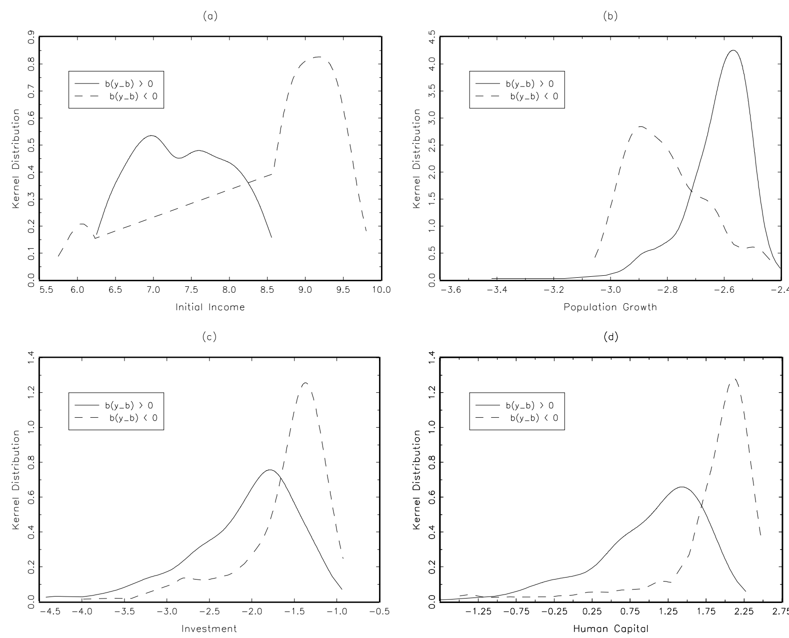

Figure 3: Kernel density estimates of initial income per capita, population growth, the capital-output ratio and human capital for both positive and negative slope coefficients on the initial income variable (without controls)

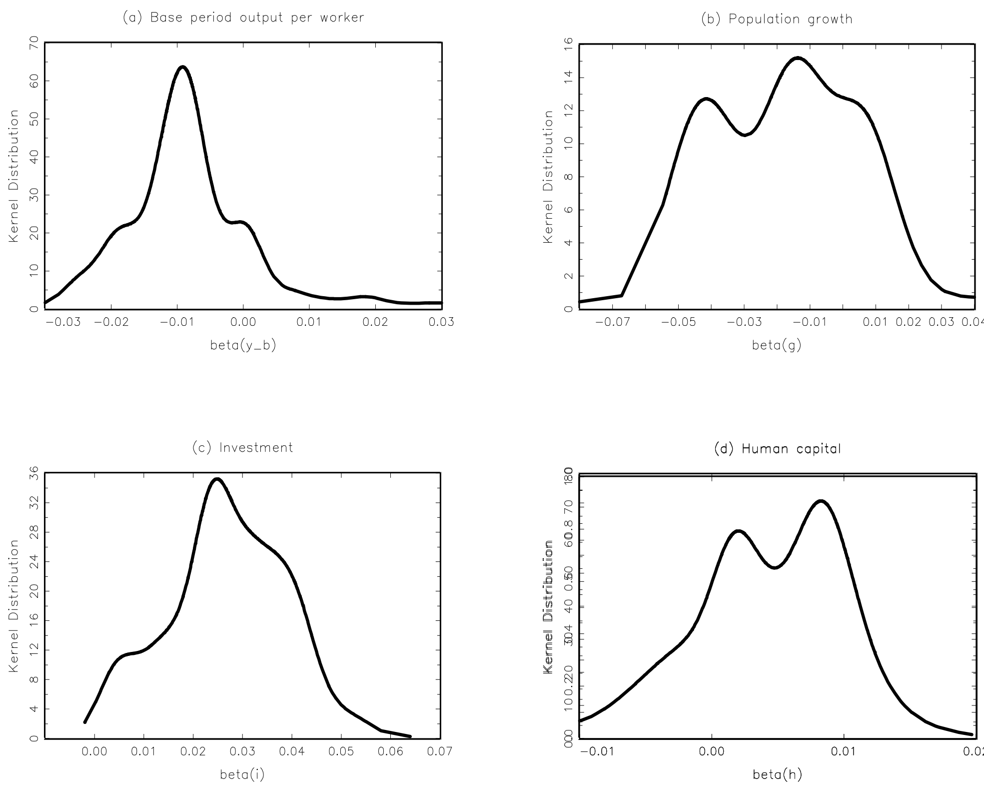

Figure 4: Kernel density estimates of the slope coefficients for the regression of output growth on initial income per capita, population growth, the capital-output ratio and human capital

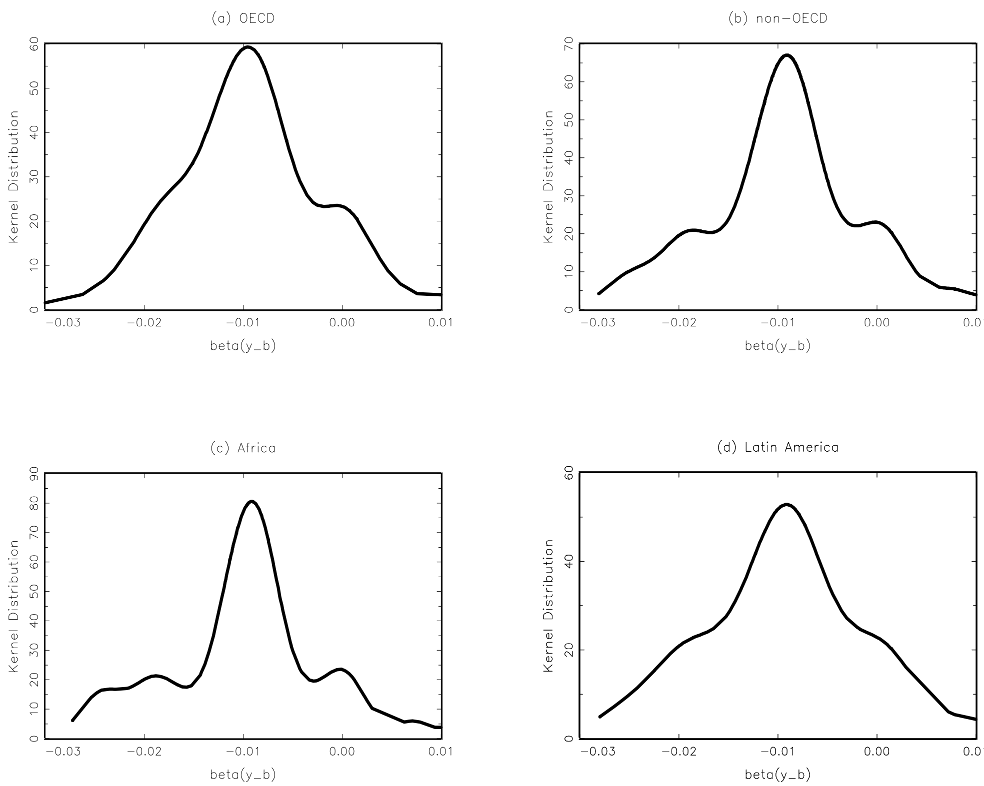

Figure 5: Kernel density estimates of the slope coefficients for select groups of countries from the nonparametric regression of growth on initial income per capita (with controls)

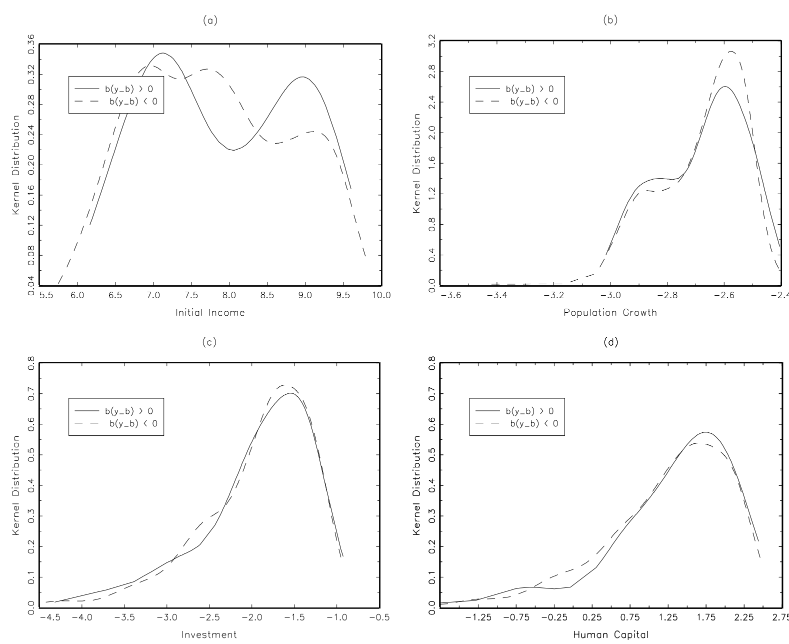

Figure 6: Kernel density estimates of initial income per capita, population growth, the capital-output ratio and human capital for both positive and negative slope coefficients on the initial income variable (with controls)