Are any growth theories linear? Why we should care about what the evidence tells us

Henderson, Daniel J.; Papageorgiou, Chris and Parmeter, Christopher F.
State University of New York at Binghamton

13. May 2008

Are any growth theories linear? Why we should care about what the evidence tells us*

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

Chris Papageorgiou‡
Research Department
International Monetary Fund

Christopher F. Parmeter§
Department of Agricultural and Applied Economics
Virginia Polytechnic Institute and State University

May 13, 2008

Abstract

Recent research on macroeconomic growth has been focused on resolving several key issues, two of which, specification uncertainty of the growth process and variable uncertainty, have received much attention in the recent literature. The standard procedure has been to assume a linear growth process and then to proceed with investigating the relevant variables that determine growth across countries. However, a more appropriate approach would be to recognize that a misspecified model may lead one to conclude that a variable is relevant when in fact it is not. This paper takes a step in this direction by considering conditional variable uncertainty with full blown specification uncertainty. We use recently developed nonparametric model selection techniques to deal with nonlinearities and competing growth theories. We show how one can interpret our results and use them to motivate more intriguing specifications within the traditional studies that use Bayesian Model Averaging or other model selection criteria. We find that the inclusion of nonlinearities is necessary for determining the empirically relevant variables that dictate growth and that nonlinearities are especially important in uncovering key mechanism of the growth process.

JEL Classification: C12, C14, C15, O10, O40.

Keywords: Growth Nonlinearities, Irrelevant Variables, Least Squares Cross Validation, Bayesian Model Averaging, Parameter Heterogeneity.

---

*The views expressed in this study are the sole responsibility of the authors and should not be attributed to the International Monetary Fund, its Executive Board, or its management.

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

‡Chris Papageorgiou, Research Department International Monetary Fund, Washington DC, 20431. Phone: 202-623-7503, Fax: 202-589-7503, E-mail: cpapageorgiou@imf.org.

§Christopher F. Parmeter, Department of Agricultural and Applied Economics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0401. Phone: 540-231-0770, Fax: 540-231-7417, E-mail: parms@vt.edu.

1 Introduction

Recently, much attention in the growth empirics literature has paid attention to four key tenets: (i) parameter heterogeneity, (ii) the consequences of competing economic growth theories, (iii) nonlinearities in the growth process and (iv) association versus causation.[1] Most research that has taken issue with any these points have done so in a singular fashion, focusing on any given tenet before dealing with another. Strategies that can tackle numerous empirical issues are thus warranted and conjectured to provide better insights into the growth process. The current study's aim is to address tenets (i) and (iii) in simultaneous fashion while keeping an eye towards (ii). This is empirically interesting as it can be seen as a step towards motivating more appropriate parametric specifications for use with Bayesian Model Averaging (BMA) and other model selection methods currently used to investigate competing growth theories,[2] while at the same time shedding new insights into growth processes for individual growth theories. Thus, empirical growth researchers of all ilks should find the discussion in this paper illuminating.

As it stands, point (ii) can be subdivided into further categories, two of which can be classified as variable uncertainty and specification uncertainty. Typically, model selection studies assume a linear (or at least functionally known) growth process so that specification uncertainty can be abrogated (see e.g., Fernadez, Ley and Steel; 2001, Brock and Durlauf; 2001, Sala-i-Martin, Doppelhoffer and Miller; 2004, Ciccone and Jarocinski; 2007, and Durlauf, Kourtellos and Tan; 2008). However, an emerging theme in the literature (see Massoumi, Racine, and Stengos; 2007 for the most current research) has been the appearance of significant nonlinearities in cross-country growth regressions.[3] From this vista, it is relevant to identify the nonlinearities in the growth process, *for a specific growth theory*, so that they can be used to extend the model space of BMA and other growth model selection investigations in uncovering the appropriate growth process, *assuming* a universal one exists.[4]

Our ability to deal with specification uncertainty and variable uncertainty stems from recent research in nonparametric model selection methods, see Hall, Li and Racine (2007). These methods

[1] See Temple (1999), Brock and Durlauf (2001), and Durlauf, Johnson and Temple (2005) for more on these issues in the growth empirics literature.

[2] See Durlauf, Kourtellos, and Tan (2008) for a look at dealing with (ii) and (iv) simultaneously.

[3] Kalaitzidakis, Mamuneas, and Stengos (2000) take a step in this direction by considering variable selection in the presence of possible nonlinearities, however, the main variables of interest enter into the model in a linear fashion.

[4] Masanjala and Papageorgiou (forthcoming) have documented that African countries may grow differently than the rest of the world.

are robust to functional form misspecification (specification uncertainty) and have the ability to remove irrelevant variables that have been added by the research (variable uncertainty). We make a caveat that our variable uncertainty can be thought of as conditional; if a researcher omits a relevant regressor from this exercise then the results provided may be erroneous and so we mention that removing the irrelevant variables is *conditional* on the variables included in the exercise. More specifically, since we are not engaging in model averaging, our results can be seen as thinking about alternative growth theories separate from one another to determine appropriate variable/functional specifications *for a given growth theory.* This information could then be used at a later date when concerns over theory robustness are investigated in a model averaging context. This type of approach is very important because Durlauf, Kourtellos, and Tan (2008, page 344) suggest that "more work needs to be done in systematically uncovering potential nonlinearities and heterogeneity in growth processes across countries."

While the insights of the empirical growth papers employing BMA and model selection are valuable in and of themselves, their foundation of *a priori* functional form misspecification limits the scope of these methods in truly uncovering the process dictating economic growth. It may turn out that a variable found to be statistically relevant in explaining growth is arrived at through an inappropriate specification of the growth process; or, alternatively, it may be that a theory was deemed weak given that the functional form used to dictate growth was inappropriate for the theory of interest. Here we argue that nonparametric model selection procedures are invaluable as a tool for uncovering the salient features of growth processes: those variables (conditionally) which are *relevant* for predicting growth and their *appropriate* influence on growth.

Our results highlight the importance of accounting for nonlinearities across the spectrum of growth variables, including the Solow model variables themselves. We note that few specific growth theories outperform the baseline Solow specification (Levine and Renelt; 1992 and Durlauf, Kourtellos, and Tan; 2008) with the exception of macroeconomic policy and institutions. Both of these theories have important policy implications for fostering long term economic growth.[5] We also find that nearly all *individual* growth theories appear to display some form of parameter heterogeneity as well as nonlinearities. This is important in three respects. First, it solidifies the growing consensus in the empirical growth literature that growth models exhibit functional forms that go beyond

[5]Exploiting the intricate relationship that exist between the proxy variables used here and the growth of nations is an important prospect for future research.

linear, parametric models. Second, the results here should prove useful to researchers looking for additional motivation for incorporating nonlinearities into the BMA paradigm. Lastly, for applied growth researchers, this paper outlines an approach for determining potential nonlinearities which may subsequently guide model selection.

The remainder of the paper is organized as follows. Section 2 reviews the findings of various model selection studies to get a sense for the variables and theories that are most relevant for studying growth across countries. These results will serve as a benchmark and guideline for proper perspective of our results to follow. Section 3 discuss the data used in the estimation and provides the econometric intuition and the mechanics behind nonparametric model selection. Section 4 presents the main results of the paper by contrasting two and three growth theories at the same time across the 8 main theories listed in Durlauf, Kourtellos, and Tan (2008) (DKT hereafter). Section 5 provides Monte Carlo evidence that the nonparametric model selection methods work well for the sample sizes used in the paper, and number of included covariates typical of a singular growth theory regression exercise. Section 6 discusses the results and concludes.

2 A Brief Review of the Literature

2.1 Growth Variable Robustness

Melding cross-country growth regressions with various conditioning sets dates back to the seminal work of Levine and Renelt (1992) who used Leamer's (1983) Extreme Bounds Analysis (EBA)[6] to check the robustness of the key economic, political and institutional variables that, at the time, were used extensively to detect empirical linkages with long-run growth rates. These authors looked at no more than seven growth variables at a time and focused on a cross section of anywhere from 64-106 countries depending on the variables used, investigating growth over the period 1960-1989. However, much of the study focuses on the shorter time horizon of 1974-1989 due to lack of specific conditioning variables for the period from 1960-1973.

Levine and Renelt (1992) also adopted the tradition of including a set of variables that appear in 'every' regression run. Typically these are chosen to be the Solow variables[7], but are not required

[6]Model uncertainty has long been recognized as a major econometric problem in regression analysis. The initial approach to model selection was to use stepwise methods developed by Efroymson (1960) and search over various classes of models choosing the one that best fits the data. Leamer (1978) developed a method we now call EBA that would be superior to stepwise regression in that it would account not only for the within model uncertainty but also the between model uncertainty associated with model selection.

[7]The traditional Solow variables are taken as initial income, population growth plus a constant designed to capture

to be. EBA can be seen as overtly restrictive in the face of non-robustness. To remedy this Sala-i-Martin (1997a,b) developed alternative methods that still penalized non-robust variables, albeit less harshly than EBA. Sala-i-Martin had 62 covariates but chose to follow the strategy of Levine and Renelt (1992) and only considered seven variables at a time and always included the Solow variables in every regression.[8]

Sala-i-Martin (1997a,b) reached an almost polar conclusion to that of Levine and Renelt (1992) documenting 22 variables that were robust according to his method as well as all three of the Solow variables included in all the regressions. While his methods were not based on any formal statistical theory, they did open up a debate on the relevant sources of growth and how one goes about parsing them out from a seemingly infinite pool of candidate variables. One point worth making is that the studies of Levine and Renelt (1992) and Sala-i-Martin (1997a,b) dealt with the robustness of key variables in cross-country growth regressions across varying specifications of said regressions.

2.2 Model Uncertainty and Model Averaging in Growth Empirics

It was not until the turn of the century that growth empiricists started attacking the issues raised by Levine and Renelt (1992) and Sala-i-Martin (1996, 1997) with model averaging methods, acknowledging that the model space for cross-country growth regressions was quite large. The basic idea behind model averaging is to estimate the distribution of unknown parameters of interest across different models. The principle of model averaging is to treat models and related parameters as unobservable and estimate their distributions based on observable data. In contrast to classical estimation, model averaging helps account for model uncertainty and consequently reduces related biases of the parameters.

To wit, Brock and Durlauf (2001) Fernandez, Ley and Steel (2001), DKT, and Masanjala and Papageorgiou (forthcoming) have all attacked the robustness of various growth theories (for various countries) using BMA, while Sala-i-Martin, Doppelhoffer and Miller (2004) have used Bayesian averaging of classical estimates (BACE) procedures, while Hendry and Krolzig (2004) and Hoover and Perez (2004) used general to specific modelling approaches. These methods are more parsimonious than EBA and are grounded in statistical theory (see Hoeting, Madigan, Raftery and Volinsky; 1999 for a nice overview of BMA).

---

depreciation rates and technological growth, the investment rate, and a measure of human capital.

[8]Sala-i-Martin (1997a,b) did not include population growth as one of his 'Solow' variables, thus he only has three variables that are in every regression.

Fernandez, Ley and Steel (2001) use the same data as Sala-i-Martin (1997a,b) but do not require that only seven variables appear at a time and also do not include the Solow variables in every regression. Using a posterior probability cutoff of 90% they find that of the 22 variables deemed 'robust' in Sala-i-Martin (1997a,b), only four (initial income, percent Confucian, life expectancy and equipment investment) are statistically relevant from their perspective. Their findings were appealing for a variety of reasons, one of the most important being that the regressions considered were not required to have at most seven variables. This lent further evidence that limiting the size of the model space of linear growth regressions had an impact on the findings.

Brock and Durluaf (2001) laid the terminology and foundation for the importance of model averaging when considering growth models and growth theories. Their discussion of model uncertainty brought to light several key facets of model uncertainty: theory uncertainty, functional form uncertainty and heterogeneity uncertainty. Their application of BMA focused on the study of Easterly and Levine (1997) on the impact of ethnic conflict on growth and its potential for explaining Africa's dismal growth performance compared to the rest of the world. Brock and Durlauf (2001) find that ethnic conflict is a robust predictor of growth in the face of theory uncertainty. Building on the fact that Africa may grow differently than the rest of the world found in Brock and Durlauf (2001), Masanjala and Papageorgiou (forthcoming) conducted a full scale study of model uncertainty focusing exclusively on Sub-Sahara African countries.

Ley and Steel (forthcoming) and Doppelhofer and Weeks (forthcoming) extend this literature by constructing alternative measures of jointness to explore dependence among growth regressors, in the context of Bayesian model selection. These papers show that some key growth determinants should occur jointly in growth regressions, while the majority of the regressors captures effects that can also be accounted for by other regressors. In other recent contributions, Ley and Steel (forthcoming) and Eicher, Papageorgiou and Raftery (2007) argue that the implementation of BMA is subject to the choice of priors: the priors for the parameters in each model, and the prior over the model space. Using predictive performance, a neutral criterion for comparing different priors, these papers show that model choice can be sensitive to the prior specification.

Finally, DKT consider an unbalanced panel of countries and look not only at the importance of individual variables on the growth process, but the competing growth theories themselves. They also account for endogeneity in their model averaging exercises, thus emphasizing two of the four major tenets discussed in the introduction. Their findings suggest that many of the 'nouveau' growth theories are not as robust as previously believed and that very few of the commonly used variables have high posterior probabilities. These posterior probabilities signify the likelihood that a variable or theory is part of the 'true' growth process. The variables that appear to be robust are initial income and investment (two of the Solow variables), government consumption and inflation (variables classified as relating to macroeconomic policy), and the East Asian regional dummy (capturing regional heterogeneity). In unison with the results about variable robustness, DKT also determine that the theories with the highest posterior probabilities are the original Solow model original Solow model, macroeconomic policies and regional disparities across nations. Theories such as institutions, demography, geography, religion, and fractionalization do not appear to be robust (when put up against other theories) at empirically determining cross-country economic growth.

2.3 Nonlinearities and Heterogeneity in Growth Regressions

The robustness and model uncertainty exercises have shed new light on important and telling growth features. However, one area where these methods have been less used has been examining the impact of nonlinearities and parameter heterogeneity within the growth process. In fact, very few studies have paid much attention to the fact that growth may not be dictated by a global linear process. An exception to this is Durlauf and Johnson (1995), whose pioneering empirical work brought to the attention of growth empiricists heterogeneity in cross-country growth. Their work implies that different countries obeyed different *linear* growth processes using regression tree methods. They were able to account for parameter heterogeneity within the standard Solow framework, albeit using a linear model.

Lee, Pesaran and Smith (1997) estimated a stochastic Solow model that allowed for parameter heterogeneity by letting the convergence coefficient vary across countries in a panel data setting. Their findings showed significant heterogeneity in terms of the speed of convergence, typically taken as a transform on the coefficient of initial income. These findings have subsequently been reaffirmed by Durlauf, Kourtellos, and Minkin (2001) and Kourtellos (2003) using semiparametric smooth coefficient models. These two approaches are interesting because they model nonlinearities and parameter heterogeneity in a simultaneous fashion.

Another interesting extension of the insights from Duraluf and Johnson (1995) is to focus on the theoretical model that gives rise to the empirical specification. While the textbook Solow (1956) model yields a log linear econometric model, the constant elasticity of substitution model allows the growth process to behave in a nonlinear fashion in econometric settings (Masanjala and Papageorgiou 2004). Extending the regression tree approach of Durlauf and Johnson (1995) from linear models to nonlinear models thus allows the researcher to account for nonlinearities and parameter heterogeneity in the growth process simultaneously. This ability to do two things at once was not exploitable in the model of Durlauf and Johnson (1995) due to the linear in parameters nature of the Solow model's growth predictions.

Tan (2007) used GUIDE (general, unbiased interaction detection and estimation) to aid in identifying clustering of countries that obey a common growth model. This methodology is similar in spirit to that of Durlauf and Johnson (1995) but the methods employed by Tan (2007) look for interactions between covariates, thus introducing nonlinearities, and have the tendency to provide fewer regression splits. The evidence relayed in Tan (2007) show that institutional quality and ethnic fractionalization define convergence clubs. These results strengthen the implications of the Azariadis and Drazen (1992) model of threshold externalities for economic growth.

Much of the focus on nonlinearities using nonparametric kernel methods in empirical growth regressions has been due to the pioneering work of Thanasis Stengos. Liu and Stengos (1999) consider an additive partly linear growth specification. Specifically, they consider a semiparametric model where some of the variables enter linearly whereas others are allowed to enter in an unknown fashion. Their research influenced a large number of studies within the semiparametric domain (e.g., see Durlauf, Kourtellos and Minkin 2001, Ketteni, Mamuneas and Stengos 2007, Mamuneas, Savvides, and Stengos 2007, and Vaona and Schiavo 2007). All of these papers have shown significant nonlinearities for a variety of variables on cross-country economic growth. 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, Massoumi, Racine and Stengos (2007) consider a fully nonparametric growth structure. Specifically, they focus on what happens to predicted growth rates and residuals over time. Here we deviate from Massoumi, Racine and Stengos (2007)'s focus, but exploit their methodology to determine which growth theories display nonlinear tendencies.

To our knowledge the only paper that has combined robustness of economic variables in a growth regression context while allowing for nonlinearities has been Kalaitzidakis, Mamuneas and Stengos (2000). Their work used EBA, as in Levine and Renelt (1992), but allowed for nonlinearities by setting up the growth regression in a partly linear framework. They allowed the Solow variables to enter the growth regression in a linear fashion, consistent with the Solow model predictions, but the auxiliary variables used in Levine and Renelt (1992) were allowed to enter in a nonparametric fashion. Kalaitzidakis, Mamuneas and Stengos (2000) tested the linear specification of the auxiliary variables and then used these robust models to ascertain the significance of any variable in standard EBA fashion. Their findings confirmed that investment has a robust impact on growth, however, the omitted nonlinearities of Levine and Renelt (1992) showed that at least one variable from every major policy group was robust, contrary to their conclusions. In sum, Kalaitzidakis, Mamuneas and Stengos (2000, p. 616) note that "... the use of a simple linear regression framework is inappropriate for assessment of the specification of cross-country growth models and for addressing the robustness properties of variables that enter these models."

3 Data and Estimation Methods

3.1 Data

Our data come from DKT and represent an ample portion of the set of variables that have been used at one juncture or another to assess a growth theory.[9] We briefly look at several key features of the data before getting to our main results.

The DKT data set contains data for the traditional Solow model (initial income, investment rate, human capital, population growth) as well as variables that compose several of the contending growth theories being debated today: fractionalization, institutions, demographics, geography, religion, and macroeconomic policy. At least two variables for each theory are used. Given that region is an unordered discrete variable, which does not affect the asymptotic properties of the estimator, we include it when we compare all other theories.

3.2 Nonparametric Methods for Growth Empirics

Standard growth regressions take the following (linear) form:

$$g_i = \beta' w_i + \gamma' z_i + \varepsilon_i \tag{1}$$

where g_i is the growth rate of output over a predetermined time period, w_i is a vector composed of the 'Solow' variables, initial income, physical capital savings rate, human capital savings rate,

[9]The theories tested and the variables used are contained in an appendix available from the authors upon request.

and the joint depreciation term on both types of capital,[10] while z_i is a vector of unknown length that contains variables associated with several alternative growth theories. The exact variables within the z_i vector is what typically gives rise to model uncertainty; while there are many growth theories none refutes the others and so an exact specification of Equation (1) becomes increasingly difficult as more growth theories are constructed. Brock and Durlauf (2001) refer to this inability of growth theories to reject one another as 'openendedness'. Empiricists have used BMA to uncover just what variables matter in both the x_i and z_i vectors, but to date have yet to break free of the linear growth structure implicit in Equation (1).

Now, consider a general growth specification taking the unknown form:

$$g_i = m(x_i) + \varepsilon_i, \qquad i = 1, ..., N \tag{2}$$

where x_i is the union of w_i and z_i and g_i is the growth rate of country i. Further, m is the unknown smooth growth process. For the argument $x_i = [x_i^c, x_i^u, x_i^o]$ we make distinct reference to data type; x_i^c is a vector of continuous regressors (initial income, capital savings rate, percent Confucian), x_i^u is a vector of regressors that assume unordered discrete values (geographic regions, OECD membership), and x_i^o is a vector of regressors that assume ordered discrete values (time, number of conflicts, trade openness). An additive, mean zero error is captured through ε_i.

3.3 Nonparametric Regression

In this section we describe Li-Racine Generalized Kernel Estimation (see Li and Racine 2004 and Racine and Li 2004) of equation (2). Ignoring for the moment the fact that irrelevant regressors may have been included in Equation (2), we discuss its estimation using standard kernel techniques. To begin we model the unknown relationship through the conditional mean, i.e. $m(x_i) = E[g_i|x_i]$. This allows us to write the regression equation at a given point as

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} g_i K_h(x, x_i)}{\sum_{i=1}^{n} K_h(x, x_i)}. \tag{3}$$

where

$$K_h(x, x_i) = \prod_{s=1}^{q} h_s^{-1} l^c \left(\frac{x_s^c - x_{si}^c}{h_s} \right) \prod_{s=1}^{r} l^u \left(x_s^u, x_{si}^u, \widehat{\lambda^u}_s \right) \prod_{s=1}^{p} l^o \left(x_s^o, x_{si}^o, \widehat{\lambda^o}_s \right). \tag{4}$$

[10]The common $n_i + g + \delta$ term that includes population growth rate, technology growth rate, and factor depreciation rates, respectively.

$K_h(x, x_i)$ is the commonly used product kernel (see Pagan and Ullah 1999), where l^c is the standard normal kernel function with window width $h^c_s = h_s(N)$ associated with the s^{th} component of x^c. l^u is a variation of Aitchison and Aitken's (1976) kernel function and l^o is the Wang and Van Ryzin (1981) kernel function. See Li and Racine (2004) and Racine and Li (2004) for further details. Nonparametric regression of this type is known as local constant least squares (LCLS).

Equation (3) can be written in matrix notation to display it in a more compact form. Let $\mathbf{i}$ denote an $n \times 1$ vector of ones and let $\mathcal{K}(x)$ denote the diagonal n matrix with j^{th} element $K_h(x, x_j)$. Also, denote by g the $n \times 1$ vector of growth rates across countries. Then, we can express our LCLS estimator as

$$\hat{m}(x) = \left(\mathbf{i}'\mathcal{K}(x)\mathbf{i}\right)^{-1} \mathbf{i}'\mathcal{K}(x)g. \tag{5}$$

Another popular method of nonparametric regression, known as local linear least squares (LLLS), begins by taking a first-order Taylor expansion [11] of (2) around x, yielding,

$$g_i \approx m(x) + (x^c_i - x^c)\beta(x^c) + \varepsilon_i \tag{6}$$

where x^c refers to the continuous variables within x, $\beta(x^c)$ is defined as the partial derivative of $m(x)$ with respect to x^c. The estimator of $\delta(x) \equiv (m(x), \beta(x^c))'$ is given by

$$\widehat{\delta}(x) = \left[\sum_i K_h(x, x_i) \begin{pmatrix} 1 \\ x^c_i - x^c \end{pmatrix} \left(1, (x^c_i - x^c)'\right)\right]^{-1} \sum_i K_h(x, x_i) \begin{pmatrix} 1 \\ x^c_i - x^c \end{pmatrix} g_i. \tag{7}$$

The returns to the categorical variables are obtained separately. For example, the impact of OECD status, akin to the coefficient on an OECD dummy variable in a standard growth regression, is calculated as the counterfactual change in OECD status of a particular country (switches from zero to one), *ceteris paribus*. Consequently, the returns to the categorical variables also vary across observations. This type of analysis is not common in parametric and semiparametric procedures.[12]

Equation (7) can be written in vector-matrix form to reduce the notational burden. Let $\mathcal{X}$ be an $n \times (1 + q)$ matrix with j^{th} row being $\left(1, \left(x^c_j - x^c\right)'\right)$. Here q represents the number of continuous variables appearing in the unknown function. Our estimator takes the compact form

$$\hat{\delta}(x) = (\mathcal{X}'\mathcal{K}(x)\mathcal{X})^{-1}\mathcal{X}'\mathcal{K}(x)g. \tag{8}$$

[11] The Taylor expansion is only taken for the continuous variables.
[12] See Li and Racine (2007).

3.4 Cross-Validatory Bandwidth Selection

Estimation of the bandwidths $(h, \lambda^u, \lambda^o)$ is typically the most salient factor when performing nonparametric estimation. For example, choosing a very small h 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 h, 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 econometrics and thus we usually resort to automatic selection procedures to estimate the bandwidths. Although there exist many selection methods, Hall, Li, and Racine (2007, HLR hereafter) have shown that Least Squares Cross-Validation (LSCV) has the ability to smooth away irrelevant variables that may have been erroneously included into the unknown regression function. Specifically, the bandwidths are chosen to minimize

$$CV(h, \lambda) = \underset{\{h,\lambda\}}{argmin} \frac{1}{n} \sum_{i=1}^{n} (g_i - \hat{m}_{-i}(x_i))^2, \tag{9}$$

where $\hat{m}_{-i}(x_i)$ is the common leave-one-out estimator. Notice that even when one is selecting bandwidths to be used for LLLS estimation, the unknown function is all that enters into the CV criterion, *not* the partial derivatives.

For the discrete variables, the bandwidths indicate which variables are relevant, as well as the extent of smoothing in the estimation. From the definitions for the ordered and unordered kernels, it follows that if the bandwidth for a particular unordered or ordered discrete variable equals zero, then the kernel reduces to an indicator function and no weight is given to observations for which $x_i^o \neq x_j^o$ or $x_i^u \neq x_j^u$. On the other hand, if the bandwidth for a particular unordered or ordered discrete variable reaches its upper bound, then equal weight is given to observations with $x_i^o = x_j^o$ and $x_i^o \neq x_j^o$. In this case, the variable is completely smoothed out (and thus does not impact the estimation results). For unordered discrete variables, the upper bound is given by $(d_s - 1)/d_s$ where d_s represents the number of unique values taken on by the variable. For example, a categorical variable for geographic location which takes on 5 values would have an upper bound for its bandwidth of $4/5 = 0.8$. For ordered discrete variables, the upper bound is unity. See HLR for further details.

3.5 Nonparametric Model Selection

The abundance of asymptotic results that form the statistical backbone of nonparametric methods have always assumed that the bandwidth(s) converge to zero (at a certain rate) as the sample size gets larger. This means that as the sample size is increased the amount of data in a specific region is growing and so the kernel weighting function no longer needs to use points farther away to construct an accurate representation of the functional form. However, recent advances have shown that when the researcher includes irrelevant variables, this bandwidth condition is no longer true. Automatic bandwidth selection procedures actually increase the bandwidths associated with irrelevant regressors, essentially removing them from the sample. It is as if the researcher had failed to include them in the first place! It was commonly believed that the inappropriate inclusion of irrelevant variables harmed the performance of nonparametric methods, but this is not the case.

HLR have shown that the inclusion of *irrelevant* regressors does not add to the 'curse of dimensionality'.[13] Their paper shows that when one uses cross-validation procedures to select the appropriate amount of smoothness[14] of the unknown function, the covariates that are irrelevant are eliminated from the smoothing relationship. This property allows nonparametric estimators to not only allow for functional form misspecification, but relevant covariate selection at the same time. Thus, tenets (ii) and (iii) alluded to in the introduction can be handled simultaneously; a potentially elucidating advance for the growth empirics literature.

However, there is no free lunch for this method as it hinges on several facets that need to be considered on a case by case basis. First, the key assumption used by HLR asks that the irrelevant regressors are independent of the relevant regressors, something unlikely to hold in practice.[15] Second, it is not entirely clear how well this method works as the set of relevant regressors is increased. HLR's finite sample investigations looked at at most two relevant regressors while there empirical application considered six variables for 561 observations in which only two regressors were deemed relevant according to their procedure. Clearly more work needs to be done to assess the performance of this level for very small sample sizes and for large sets of potential regressors.

[13] Addition of other relevant variables still adds to the dimensionality issue however.

[14] See the Monte Carlo exercises in Section 5.

[15] This is not entirely damning as it was shown in finite samples that the HLR method worked even when dependence was allowed between relevant and irrelevant regressors. The assumption was made for ease of proof of the corresponding theorems in the paper. Indeed, in our small sample exercises we violate this condition and it appears to have no affect on the corresponding results.

4 Results

This section brings together all of our results from the methods discussed above.[16] We also deepen our results by using the consistent model misspecification test of Hsiao, Li and Racine (2007) and the consistent variable significance test of Lavergne and Vuong (2000).[17] These test buttress the appealing features for detecting departures from linearity and variable relevance from the local constant and linear estimation procedures, with LSCV bandwidth selection, discussed previously. The model misspecification test allows us to determine if nonparametric methods are needed and the test of significance provides insight into variables that may not have been smoothed away via the LSCV procedure,[18] thus resulting in needed dimension reduction to lessen the curse of dimensionality.

Our first goal is to examine the Solow growth variables and use these results as a baseline when additional theories are investigated. Then we consider variable robustness *within* each individual growth theory via the LSCV selected bandwidths and theory robustness, with respect *only* to the regional Solow model, using the Lavergne and Vuong test. This will allow us to determine which of the proxy variables for each model, and each individual theory, are relevant predictors of growth. From there we will examine which of the individual growth theories are nonlinear by testing the linear, parametric specification. Finally, we will examine in detail, potential nonlinearities in our Macro and Geographic growth theories. These results will showcase how nonparametric methods can be used to deepen one's understanding of some of the types of nonlinearities and parameter heterogeneity that may exist.

Our bandwidths for the exact DKT sample, across all theories, are presented in Tables 1 and 2. To increase the efficacy of our model selection exercises, as well as the power of the two tests employed, we also divided up the DKT data to maximize the number of observations for any given theory. Thus, in Tables 3 and 4 we present bandwidths found for local constant and linear regressions, respectively, with all available observations for a given theory. The last two rows of Tables 1 through 4 show $p-$values for 399 bootstrap replications of a consistent test of model misspecification (I_n^a) and a consistent test for variable insignificance (I_n^b). Again, the significance test is performed on all the variables in the model outside of the Solow variables and our regional

[16]All code used in this section is available from the authors upon request.
[17]These procedures are described in Appendices 1 and 2, respectively.
[18]This is due to the fact that the LSCV bandwidth selection method does not constitute a formal test.

indicator. In essence, we are testing whether the theory under consideration is a significant predictor of growth above and beyond the main empirical specification used in many growth studies.

Before discussing our results we make a note on interpreting the bandwidths. For the two discrete variables, time and region, their relevance (in either local constant or linear regression) is determined by how close the bandwidth is to its upper bound; the upper bound for *time* is 1 while that for *region* is 0.875. For the continuous regressors, the upper bound of the bandwidth in the local-linear regression suggest whether the variable enters in linearly, *not* if it is relevant as is the case when interpreting the local constant bandwidths. The upper bound for a continuous variable is infinity and thus is impossible to observe in practice. We follow the suggestion of HLR and use two standard deviations of the independent variable as the bound for relevance/linearity. Thus, if any bandwidth exceeds two standard deviations of its associated variable, we conclude that it enters in an irrelevant fashion (in a local constant setting) or linearly (in the local linear setting). However, this linearity does not mean that important interactions do not exist. One should also check for these in practice using the methods illustrated in the subsections focusing specifically on Macro and Geography as stand alone growth theories.

4.1 The Solow Variables

Our bandwidths for the Solow variables, when considering *only* the Solow model, provide a snap shot of the model's perceived fit when viewed as the main driver behind economic output. We note that population growth and human capital are smoothed out while there appear to be relevant nonlinearities occurring in both investment and initial income. It is noteworthy that population goes from being irrelevant to relevant once regional effects are accounted for with the relevance of human capital not depending on those same regional effects. The nonlinearities in initial income are in accord with the findings of Durlauf, Kourtellos, and Minkin (2001) and Kourtellos (2003). Aside from a handful of studies, most growth researchers ignore any type of nonlinear structure either between or across these variables, often resorting to standard fare linear models.

Also, we note in passing that just the inclusion of regional effects greatly improves the model's fit, bumping up the pseudo-R^2 from just under 0.5 to 0.73 (in both Tables 1 and 2). This is similar to the results of Temple (1998) who found that there were significant regional impacts on output. However, the test of significance for the regional variable suggests that it is irrelevant (In Table 2). One reason for this puzzling result is the small sample size of the original DKT dataset. The

Table 1: Bandwidths for DKT data using Local Constant Regression

Variable	Solow	Region	Demo	Geo	Macro	Frac	Rel	Inst
Population Growth	410210	0.0802	160790	2800122	0.1422	0.1725	0.3510	0.0024
Investment	0.1184	1.1490	0.6598	0.4721	0.2307	0.3958	0.4883	384712
Human Capital	2282240	8580899	495056	7146979	0.4076	2545186	2.0254	0.1235
Initial Income	0.6716	0.3729	417718	0.3120	0.1472	0.2740	0.2563	0.0099
Time	0.6755	0.7817	0.8355	0.4057	0.4194	0.9577	0.5581	0.1967
Region	.	0.1076	0.3359	0.4285	0.0131	0.1499	0.3227	0.0000
Fertility	.	.	783861	.	.	.	.	.
Life Expectancy	.	.	0.0015	.	.	.	.	.
Koeppen-Geiger	.	.	.	0.0688	.	.	.	.
% Ice Free Coast	.	.	.	0.2779	.	.	.	.
Openness	.	.	.	.	0.2268	.	.	.
Net Govt. Cons	.	.	.	.	0.0207	.	.	.
Inflation	.	.	.	.	0.0700	.	.	.
Language	.	.	.	.	.	0.1291	.	.
Ethnic Tension	.	.	.	.	.	1992069	.	.
Hindu	.	.	.	.	.	.	0.0586	.
Jewish	.	.	.	.	.	.	3455412	.
Muslim	.	.	.	.	.	.	1727793	.
Orthodox	.	.	.	.	.	.	40294432	.
Other Religion	.	.	.	.	.	.	0.0332	.
Protestant	.	.	.	.	.	.	533780	.
Eastern Religions	.	.	.	.	.	.	0.0060	.
Exec. Constraints	.	.	.	.	.	.	.	0.0250
Exprop. Risk	.	.	.	.	.	.	.	0.1026
KKZ96	.	.	.	.	.	.	.	1205089
Legal Formalism	.	.	.	.	.	.	.	0.0648
Model Fit	49.23	73.21	89.73	91.24	99.54	84.12	87.51	99.99
I_n^a (HLR)	0.122	0.075	0.000	0.020	0.003	0.053	0.123	0.003
I_n^b (LV)	N/A	0.068	0.736	0.045	0.575	0.131	0.000	0.333

same result is not found when testing the significance of regional effects in the other three tables ($p-$values of 0.068, 0.000, and 0.019, respectively). We also see that region is never smoothed away across all theories, again suggestive of the research of Temple (1998). We note however that since the bandwidth on region is not zero that there exist important interactions between region and the continuous variables entering the model that are not captured in the Temple (1998) setting.

Increasing the sample size shows that all the Solow variables are relevant in either the Solow only or the regional Solow models (see Table 3). It also appears that no matter the sample size the local linear bandwidths suggest that initial income *does not* enter in a linear fashion. Our model misspecification tests for the Solow and regional Solow models are a bit mixed. For the original

Table 2: Bandwidths for DKT data using Local Linear Regression

Variable	Solow	Region	Demo	Geo	Macro	Frac	Rel	Inst
Population Growth	363128	0.1361	1057716	314653	0.2047	1.4589	0.1162	183568
Investment	1.8282	2.3527	0.6598	1528366	0.6376	0.9153	0.5621	1247654
Human Capital	1.3544	2920565	642330	3596212	774069	10579156	0.6186	245344
Initial Income	1.0519	2.1663	1765849	0.6561	1.0178	954223	0.4876	672167
Time	1	1	0.8355	0.8177	0.7607	0.6708	0.4915	1
Region	.	0.1627	0.3359	0.3647	0.3062	0.4418	0.0493	0.3352
Fertility	.	.	887142	.	.	.	.	.
Life Expectancy	.	.	0.0015	.	.	.	.	.
Koeppen-Geiger	.	.	.	0.1589	.	.	.	.
% Ice Free Coast	.	.	.	0.3645	.	.	.	.
Openness	.	.	.	.	224686	.	.	.
Net Govt. Cons	.	.	.	.	75246	.	.	.
Inflation	.	.	.	.	1.0149	.	.	.
Language	.	.	.	.	.	0.2260	.	.
Ethnic Tension	.	.	.	.	.	0.0990	.	.
Hindu	.	.	.	.	.	.	0.0271	.
Jewish	.	.	.	.	.	.	0.0822	.
Muslim	.	.	.	.	.	.	0.1939	.
Orthodox	.	.	.	.	.	.	0.0082	.
Other Religion	.	.	.	.	.	.	0.0506	.
Protestant	.	.	.	.	.	.	0.0537	.
Eastern Religions	.	.	.	.	.	.	0.1001	.
Exec. Constraints	.	.	.	.	.	.	.	34051
Exprop. Risk	.	.	.	.	.	.	.	0.2364
KKZ96	.	.	.	.	.	.	.	0.5345
Legal Formalism	.	.	.	.	.	.	.	1510442
Model Fit	48.05	71.69	82.82	92.43	88.37	92.05	90.67	99.90
I_n^a (HLR)	0.087	0.038	0.010	0.010	0.010	0.048	0.040	0.000
I_n^b (LV)	N/A	0.206	0.000	0.000	0.000	0.200	0.587	0.000

DKT data we accept the null (at the 10% level of significance) of a linear Solow model but reject that the regional model is correctly specified in the local constant setting. The local linear results suggest that both models are misspecified as do our local constant results for the extended DKT dataset. When we use a local linear regression, however, we still find that the Solow model is not linear but once regional effects are considered we cannot reject linearity in the extended dataset.

Looking at the Solow variables across theories it is interesting to note that human capital is smoothed out in every setting except macroeconomic policy and institutions. We mention in passing that both investment and initial income are each relevant across all theories except that the institutions theory drives out the relevance of investment and the demography theory eliminates the relevance of initial income. Moving to the local linear results we see that while initial income and investment are relevant across the space of theories, assessing their perceived linearity is a bit harried. In the demography, fractionalization, and institution theories, initial income enters in linearly, albeit relevantly, except in the demography theory. The linearity of investment is relevant across the geography and institution theories, with investment being irrelevant within the institutions theory to begin with. Thus we again confirm that in general both investment and initial income are relevant predictors of economic growth and display a nonlinear effect.

It interesting to note that human capital enters linearly (for Table 2) in all models except the Solow and religion models. This is suggestive that our measure of human capital is picking up a potential omitted variable bias related to variables from other theories (this shows the danger of just focusing on individual theories, as noted by Brock and Durlauf, 2001). However, in Table 4 we see that human capital enters in a nonlinear fashion in the Solow, Region, Geography and Religion models, perhaps implying that the smaller sample may be clouding existing nonlinearities.

As noted in Section 5, the finite sample results for the nonparametric procedures improve drastically as more data is added. First, both investment and initial income are relevant across *all* theories individually (see Table 3). Second, population growth and human capital start to appear relevant across a wider array of theories than in the limited, homogenous sample. Lastly, the Solow model by itself fits the data much better than in the limited sample; a difference in fit of almost 0.17. Viewing Table 4 we see that initial income still appears to affect growth in a nonlinear manner, except in the geography, fractionalization, and religion theories. For investment, it always enters in the model in a linear fashion, a stark difference from our smaller sample results. In fact, it appears that the only Solow variable that is robustly nonlinear across the individual theories is initial income, something past research has touched upon.

4.2 Estimating Alternative Theories

While examining the impact of the Solow variables on economic growth is interesting and insightful, much of the focus on economic growth has focused on alternative explanations aside from factor accumulation and initial conditions. Theories such as geography and institutions have permeated the literature in recent years and created quite a stir among academics.[19] To determine how each

[19] See the papers by Sachs (2003) and Rodrik, Subramanian, and Trebbi (2004) for one glimpse of the ongoing debates over the causes of growth.

Table 3: Bandwidths for Full Theory data sets using Local Constant Regression

Variable	Solow	Region	Demo	Geo	Macro	Frac	Rel	Inst
Population Growth	0.1427	0.1817	357571	0.4266	0.2139	0.3708	795171	1506792
Investment	0.1888	0.4698	0.5312	0.4576	0.4755	0.5192	0.3834	0.4570
Human Capital	1.1703	2.3535	2139032	0.7527	0.5129	0.9569	2974283	7445294
Initial Income	0.3945	0.4879	0.3877	0.5342	0.2711	0.5170	0.5367	0.2607
Time	0.5379	0.2110	0.5595	0.4586	0.1671	0.7268	0.5692	0.8634
Region	.	0.6389	0.1074	0.2216	0.5375	0.1188	0.5624	0.2247
Fertility	.	.	0.1769	.	.	.	.	.
Life Expectancy	.	.	9841	.	.	.	.	.
Koeppen-Geiger	.	.	.	0.1018	.	.	.	.
% Ice Free Coast	.	.	.	1461365	.	.	.	.
Openness	.	.	.	.	0.1625	.	.	.
Net Govt. Cons	.	.	.	.	0.0382	.	.	.
Inflation	.	.	.	.	0.1692	.	.	.
Language	.	.	.	.	.	0.1796	.	.
Ethnic Tension	.	.	.	.	.	0.8328	.	.
Hindu	.	.	.	.	.	.	0.1886	.
Jewish	.	.	.	.	.	.	272966	.
Muslim	.	.	.	.	.	.	8868439	.
Orthodox	.	.	.	.	.	.	0.0407	.
Other Religion	.	.	.	.	.	.	1778613	.
Protestant	.	.	.	.	.	.	1621897	.
Eastern Religions	.	.	.	.	.	.	0.0416	.
Exec. Constraints	.	.	.	.	.	.	.	0.2481
Exprop. Risk	.	.	.	.	.	.	.	8795918
KKZ96	.	.	.	.	.	.	.	1.2687
Legal Formalism	.	.	.	.	.	.	.	0.1206
Sample Size	271	271	267	256	265	247	269	173
Model Fit	66.03	61.48	69.60	81.14	97.62	72.96	63.42	92.07
I_n^a (HLR)	0.018	0.001	0.008	0.003	0.025	0.003	0.003	0.000
I_n^b (LV)	N/A	0.000	0.024	0.200	0.492	0.514	0.512	0.773

theory *on its own* affects growth aside from factor accumulation, as well as the variables that may be seen as suitably characterizing the theory under consideration, we keep the same Solow variables, as well as region and time effects, in the models.

In terms of improvement in model fit we see from Table 1 that geography, macroeconomic policy and institutions are the highest among the individual theories for the homogeneous data set. While fit is only one way to judge the adequacy of a model we mention in passing that these three theories are the most intensely studied of those considered in this paper and in all three theories, more than one of the proxy variables are relevant and at least one of them enters in a nonlinear fashion. The fit of the macro model seems to slightly degrade when using local linear least squares while that of fractionalization improves by almost 10% (see Table 2). When looking at model fits for the heterogeneous samples, Tables 3 and 4 show that macro policy, geography and institutions dominate the other models when looking at the local constant regression while the fit of the geography model degrades somewhat in the local linear regression.

For the geography theory both the Koeppen-Geiger measure and % ice free coast measures are relevant and enter in nonlinearly. We see the same story emerge in the macro policy setting with all three of our proxies, openness, government consumption, and inflation all being relevant, however, only inflation appears to have a nonlinear impact on growth. Our setting for studying institutions uses four proxy variables, of which three are relevant and only expropriation risk entering in a relevant and nonlinear fashion.

This is suggestive that future research focusing exclusively on any of these individual theories should consider nonlinear impacts of the proxy variables. In fact, given that no variable completely captures the underlying theory being investigated, it is useful to have a means to discern both relevance and impact simultaneously, which is exactly what these nonparametric model selection techniques give us.

The demography theory set up provides a considerable improvement in fit over the basic Solow model, however, it is the only theory that suggests initial income is irrelevant. What's more, three of the four Solow variables are deemed irrelevant in the demography theory setup; the only theory of the six that displays this type of behavior. Of the demography variables, fertility and the reciprocal of life expectancy at the age of one, fertility is seen to be irrelevant while our life expectancy measure is relevant and enters in a nonlinear fashion. The results here are suggestive that after region and time effects have been controlled for increasing investment in both capital and health should lead to higher growth.

Our last two individual theories under consideration, fractionalization and religion are the worst fitting of the six theories, however, each predicts that three of the four Solow variables are relevant for explaining growth and the religion theory shows that these same three relevant predictors enter in a nonlinear manner. The fractionalization setup shows that of the three relevant Solow variables, only investment enters in nonlinearly.

If we compare these results for the homogeneous data set to those of the larger heterogeneous data sets we reach some striking similarities. First, every theory has at least one proxy variable that is relevant. Second, at least one proxy variable from each theory enters the model in a

Table 4: Bandwidths for Full Theory data sets using Local Linear Regression

Variable	Solow	Region	Demo	Geo	Macro	Frac	Rel	Inst
Population Growth	325616	734717	0.1768	1.1537	0.2865	388851	0.4981	227913
Investment	2.1206	8054010	3237715	4029275	1.5667	2.5226	2.6697	1060182
Human Capital	1.5165	2.1397	7581277	1.7506	3271349	1750581	1.7987	1016826
Initial Income	0.9804	0.9934	0.6631	3977940	0.4237	401174	5.7774	0.4815
Time	0.8860	0.4847	0.5322	0.6052	0.5590	0.5434	0.6435	1
Region	.	0.5518	0.4653	0.5130	0.6864	0.5810	0.6686	0.3727
Fertility	.	.	0.8755	.	.	.	.	.
Life Expectancy	.	.	0.0038	.	.	.	.	.
Koeppen-Geiger	.	.	.	0.2086	.	.	.	.
% Ice Free Coast	.	.	.	337765	.	.	.	.
Openness	.	.	.	.	0.7604	.	.	.
Net Govt. Cons	.	.	.	.	0.1786	.	.	.
Inflation	.	.	.	.	0.8172	.	.	.
Language	.	.	.	.	.	0.9043	.	.
Ethnic Tension	.	.	.	.	.	0.2206	.	.
Hindu	.	.	.	.	.	.	4.7708	.
Jewish	.	.	.	.	.	.	0.0994	.
Muslim	.	.	.	.	.	.	2.5299	.
Orthodox	.	.	.	.	.	.	0.2593	.
Other Religion	.	.	.	.	.	.	0.0831	.
Protestant	.	.	.	.	.	.	0.7395	.
Eastern Religions	.	.	.	.	.	.	0.1906	.
Exec. Constraints	.	.	.	.	.	.	.	0.5429
Exprop. Risk	.	.	.	.	.	.	.	1049243
KKZ96	.	.	.	.	.	.	.	2104962
Legal Formalism	.	.	.	.	.	.	.	0.2720
Sample Size	271	271	267	256	265	247	269	173
Model Fit	52.80	57.95	73.64	70.62	81.53	68.06	77.74	91.61
I_n^a (HLR)	0.075	0.155	0.028	0.198	0.013	0.005	0.033	0.003
I_n^b (LV)	N/A	0.019	0.017	0.030	0.000	0.061	0.002	0.000

nonlinear fashion. This is suggestive that there are numerous sources of economic growth and that nonlinearities play an important role in determining growth.

From the larger data set exercises we also note a few additional aspects afforded from the larger sample. In the demography theory, the relevance of the proxy variables has switched. Previously, life expectancy was a relevant predictor for growth, but it now appears that it is irrelevant and in fact fertility is driving predictions of growth. The Koeppen-Geiger measure remains relevant moving to a larger sample, while the percentage of land within 100km of ice free coast has turned irrelevant. Our macro theory proxy variables appear to be robust to the addition of almost 100 more observations with all three variables under study again showing relevance and as we see from Table

4, nonlinearly. Both fractionalization variables appear to impact growth in a relevant and nonlinear fashion while many of the religion variables are smoothed away. The additional six observations to the institutions theory have not shed new light on the variables being used as again two variables are relevant and impact growth nonlinearly.

Turning to our test results we see that for the original DKT data *only* the religion theory in the local constant setting cannot be rejected as a linear theory. When we move to the extended dataset we again see that only geography cannot be rejected as a linear theory (when using the local linear results). These results affirm the previous discussion, nearly all of the alternative growth theories discussed in the previous literature display some type of parameter heterogeneity or nonlinearities. We state again, these results suggest that when considering an individual growth theory it is important to account for nonlinearities/parameter heterogeneity.

Moving to theory significance we get a mix of results. In the smaller DKT dataset we note from our local constant results that demography, macro, fractionalization, and institutional theories, the significance test does not support the additional variables, suggesting that these theories are not robust predictors of economic growth. The local linear results however, suggest that both demography and macro are robust predictors of growth while fractionalization remains a weak predictor of growth but now religion is deemed to be an insignificant predictor of growth. The larger sample provides conflicting results as well. Our tests in the local constant setting suggest that macro, fractionalization, religion, and institutions are weak predictors, while our local linear results suggest that all of these theories are robust predictors of growth. In fact, the local linear results for the extended DKT data suggest that no theory can be rejected as a viable predictor of growth above and beyond the original Solow variables and regional effects. We note in passing that there is no consensus amongst nonparametric econometricians on which method (local constant vs. local linear) to perform testing within.[20] Previous experience leads us towards local linear results and these are what we use when we investigate deeper into the macro and geography theories.

4.3 Macro as a Growth Theory

In this section we take a detailed look at the implication of the macroeconomic variables used as proxies for determining economic growth along with the traditional Solow variables. We choose to focus on the implications from the Macro model for a variety of reasons. First, as DKT found, the

[20]This may prove a fruitful area of research in the future.

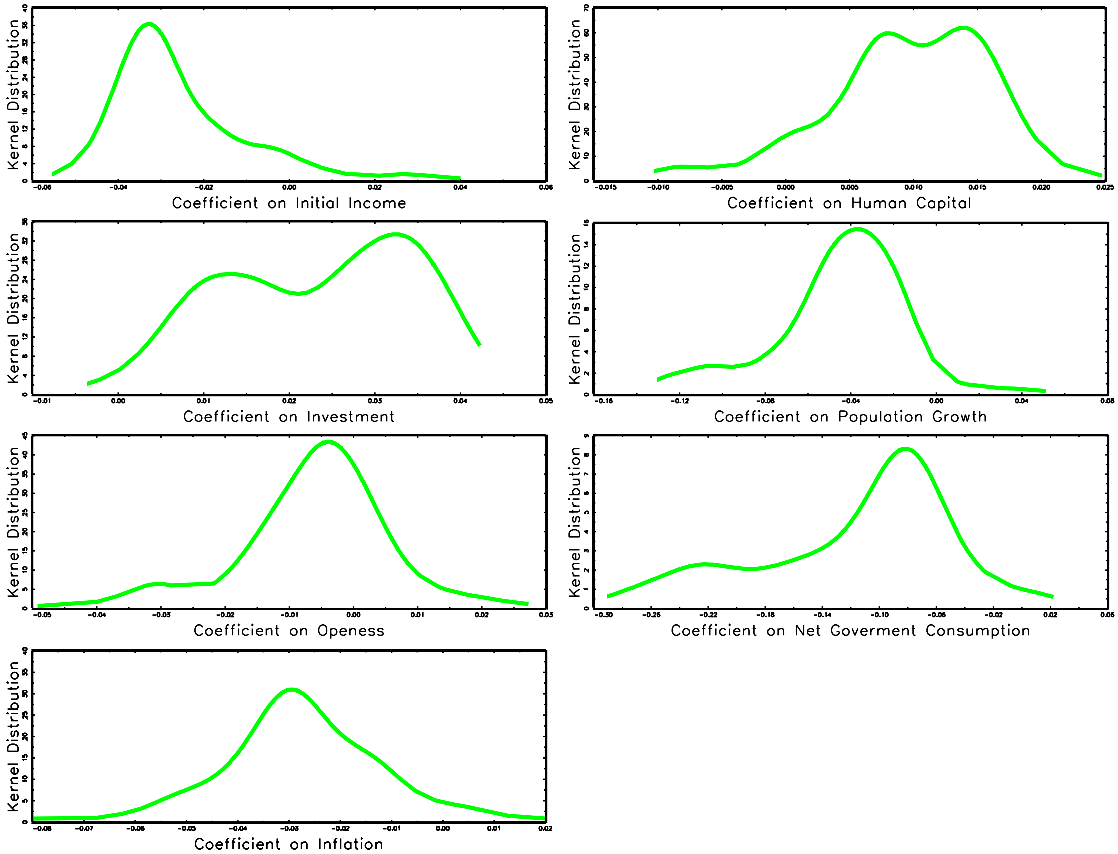
Table 5: Partial effects for all continuous regressors for the DKT macro model

Variable	Mean	$Q1$	$Q2$	$Q3$
Population Growth	-0.0453	-0.0581	-0.0409	-0.0254
	0.0140	0.0170	0.0126	0.0243
Investment	0.0232	0.0135	0.0250	0.0329
	0.0057	0.0061	0.0064	0.0074
Human Capital	0.0096	0.0064	0.0100	0.0143
	0.0043	0.0072	0.0041	0.0042
Initial Income	-0.0255	-0.0365	-0.0301	-0.0187
	0.0062	0.0242	0.0166	0.0080
Openness	-0.0062	-0.0116	-0.0049	0.0005
	0.0073	0.0079	0.0071	0.0064
Net Govt. Cons	-0.1171	-0.1604	-0.0914	-0.0747
	0.0477	0.0583	0.0628	0.0748
Inflation	-0.0281	-0.0352	-0.0287	-0.0187
	0.0105	0.0063	0.0209	0.0056

macro model (as a theory of growth) had a high posterior probability and two of the three variables used as proxies (government consumption and inflation) also had posterior probabilities very close to 1. Second, our findings suggest that the macro variables enter the model in a nonlinear and relevant fashion (using either the actual DKT or the extended dataset) for the local linear regression estimates. These two pieces of information suggest that a deeper look at how macroeconomic variables impact growth and convergence is warranted.

In Table 5 we present the quartile and mean values of the estimated coefficients for each of the continuous variables in the macro growth regression. The associated standard errors are listed underneath each estimate. We also plot out the entire distribution of estimated coefficients in Figure 1. The table and figure both suggest that there is significant dispersion in the estimated impact that any given variable has on growth across country/time. We see that the majority of the mass of the density for initial income is skewed to the right of zero suggesting 'beta' convergence, while the corresponding distributions for human capital and investment appear to posses multiple modes. Both distributions of the coefficients on openness and inflation seem to be symmetric, albeit not around zero suggesting that 1) the majority of the influences for these variables results in a negative impact on growth and 2) these impacts appear equally distributed about the mean impact. The density for net government consumption has a large mass around -0.09 and is heavily skewed to the right, even though the vast majority of the estimated effects are negative.

Figure 1: Estimated marginal effects for the Macro Theory (DKT data)



Assessing the significance of any specific estimated marginal effect is a tough task to tackle. We briefly note that while no variable has across the board significance in Table 5, every variable has at least one quartile or its mean show up as significant at the 10% level. The 10% level is arguably appropriate for such a small dataset. Given the bandwidths we found in the local linear setting as well as our results from the model misspecification and variable significance tests used previously, we feel that the macro model estimated by DKT is rife with nonlinearities for both the original Solow variables as well as the three proxy macro variables.

We can also paint a similar picture of parameter heterogeneity by examining Table 6 and Figure 2. The table shows the quartiles and mean estimates for initial income when we partition the data by the median of each of the other continuous variables values. The figure shows the entire distributions overlaid. The rightmost column of Table 6 provides the p-value of the Li (1996) test for equality of distribution for these two densities.[21] Only when considering the variable measuring

[21]See Appendix 3 for an explanation of this test.

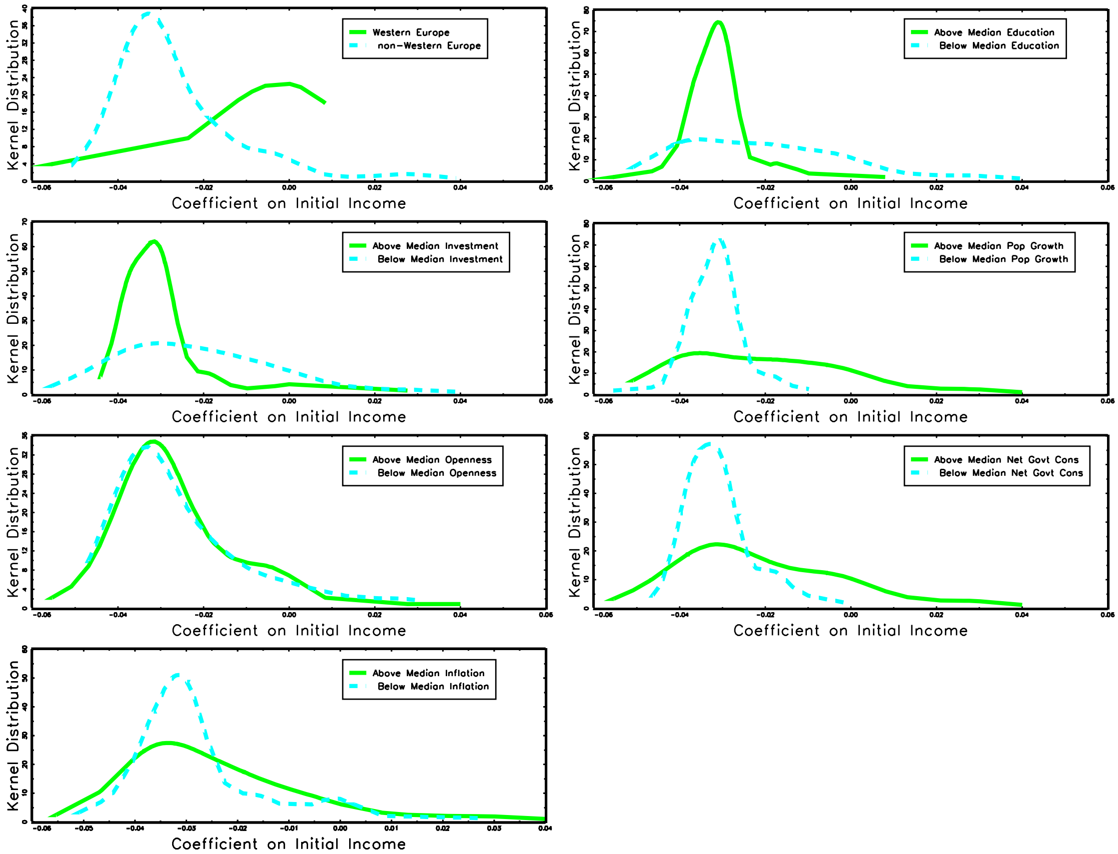
Table 6: Partial effect of initial income across various groups in Macro Model (DKT data).

Variable	Mean	$Q1$	$Q2$	$Q3$	I_n^c (L)
Western Europe	-0.0321	-0.0364	-0.0316	-0.0298	0.0000
	0.0065	0.0082	0.0084	0.0118	
non-Western Europe	-0.0228	-0.0366	-0.0277	-0.0134	
	0.0077	0.0126	0.0257	0.0109	
Above Median Human Capital	-0.0311	-0.0357	-0.0315	-0.0290	0.0000
	0.0092	0.0104	0.0116	0.0090	
Below Median Human Capital	-0.0198	-0.0376	-0.0222	-0.0060	
	0.0073	0.0286	0.0092	0.0138	
Above Median Investment	-0.0300	-0.0365	-0.0320	-0.0290	0.0000
	0.0081	0.0065	0.0078	0.0107	
Below Median Investment	-0.0208	-00354	-0.0231	-0.0090	
	0.0065	0.0067	0.0115	0.0088	
Above Median Population Growth	-0.0194	-0.0366	-0.0213	-0.0057	0.0000
	0.0064	0.0106	0.0204	0.0092	
Below Median Population Growth	-0.0316	-0.0365	-0.0315	-0.0292	
	0.0082	0.0091	0.0082	0.0149	
Above Median Openness	-0.0257	-0.0364	-0.0298	-0.0189	0.6923
	0.0116	0.0082	0.0070	0.0138	
Below Median Openness	-0.0252	-0.0366	-0.0308	-0.0182	
	0.0127	0.0111	0.0100	0.0083	
Above Median Net Govt. Cons.	-0.0202	-0.0337	-0.0226	-0.0061	0.0000
	0.0092	0.0286	0.0134	0.0115	
Below Median Net Govt. Cons.	-0.0308	-0.0366	-0.0323	-0.0285	
	0.0082	0.0111	0.0063	0.0115	
Above Median Inflation	-0.0235	-0.0369	-0.0286	-0.0176	0.0046
	0.0077	0.0118	0.0112	0.0082	
Below Median Inflation	-0.0274	-0.0364	-0.0303	-0.0264	
	0.0138	0.0070	0.0083	0.0077	

openness, does the test reject the null of equality. This suggests that the impact of initial income varies with the level of the other associated variables in the model which is indicative of interactions and parameter heterogeneity. The plots in the figure relay exactly the same information.

One interesting point is how condensed the density of the initial income coefficients are for above median education and investment compared to those below the median. The exact opposite effect is seen for population growth and net government consumption. However, given that developed countries are more likely to have above median education and investment and below median population growth, this suggests that the impact of initial income is quite condensed for highly developed countries. This is seen in the first two rows of the table where the quartiles differ dramatically for western European countries (and the U.S. and Canada) versus all other countries in the data. It is also indicative of parameter heterogeneity based on the notion of growth in stages put forth in Galor and Weil (2000).[22] Our feeling from these results are that explaining economic growth using a macro oriented theory requires accounting for nonlinearities in all variables as well as allowing for parameter heterogeneity in at least the initial income variable.

Figure 2: Density of estimated marginal effects across initial income in Macro model (DKT data).

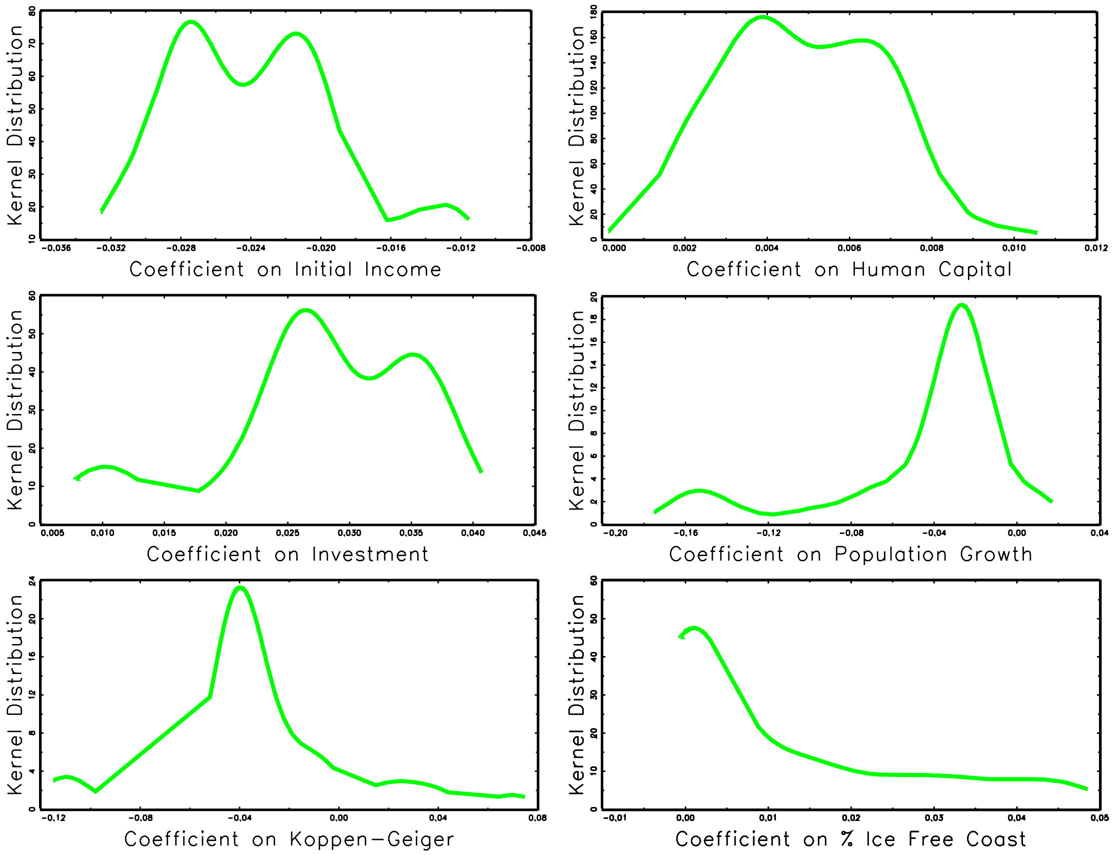


4.4 Geography as a Growth Theory

Investigating the geographic growth regression is interesting because on the surface there are many differences between this model and the macro model previously discussed. First, there has been a lively discussion over the placement of geography in growth regressions given that it is nearly immune to policy decisions (aside from starting a war to acquire more land). Second, Tables 1 and 2 suggests that the original DKT dataset implies that both geographic proxies enter in a significant

[22]In an appendix available upon request, we enlarged our macro model to 265 observations (almost 100 more than in the original DKT dataset), and came to almost identical conclusions.

Figure 3: Estimated marginal effects for the Geography theory (DKT data).



and nonlinear fashion and the linear model is misspecified. However, Table 4 suggests that while both geographic proxies are still jointly significant, % ice free coast now has a linear impact and the Hsiao, Li and Racine (2007) specification test fails to reject the null of a correctly specified linear regression. Given that the sample size has increased by more than 50% from the previous one we put faith in the notion that the previous results could have appeared due to a small sample problem. Note that the results from the macro model were consistent across local linear estimations with differing samples sizes.

Consistent with the implication of linearity is that the quartile and mean values from Table 7 suggest that the spread of the estimated marginal effects is limited. For human capital the interquartile range is 0.0031, while that for initial income is 0.0072. The largest interquartile range is 0.0339 for population growth. These estimated coefficient densities are indicative of variables that have roughly constant effects, possibly implying linearity in that variable.

Table 7: Partial effects for all continuous regressors for the DKT geography model

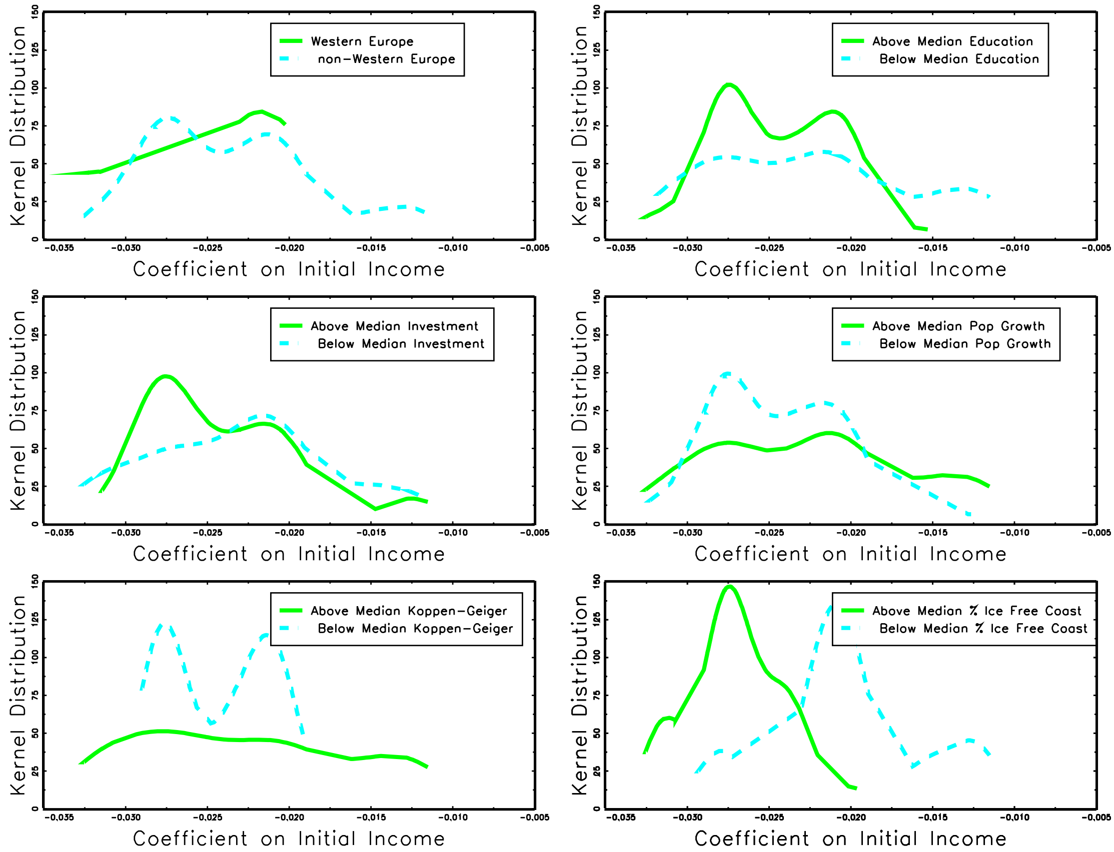
Variable	Mean	$Q1$	$Q2$	$Q3$
Population Growth	-0.0471	-0.0562	-0.0299	-0.0223
	0.0165	0.0344	0.0132	0.0160
Investment	0.0273	0.0244	0.0271	0.0345
	0.0050	0.0057	0.0050	0.0057
Human Capital	0.0049	0.0034	0.0048	0.0065
	0.0040	0.0035	0.0031	0.0036
Initial Income	-0.0235	-0.0277	-0.0238	-0.0205
	0.0039	0.0051	0.0051	0.0042
Koeppen-Geiger	-0.0302	-0.0430	-0.0378	-0.0150
	0.0094	0.0224	0.0219	0.0111
% Ice Free Coast	0.0116	0.0005	0.0018	0.0195
	0.0067	0.0046	0.0051	0.0057

Even with this evidence of linearity in the geographic growth model, looking at differences in estimated coefficient densities for initial income across splits along the median for the other five variables as well as the western Europe and U.S./Canada split. We see that these results are again indicative of parameter heterogeneity. The Li test rejects equality of estimated densities for every split in the table suggesting potential interactions between initial income and the other variables used in the model. This is true whether looking at the original DKT data (Table 8) or the extended DKT dataset (Table E2). Keep in mind that while we failed to reject a linear geographic growth regression, this evidence of parameter heterogeneity is suggestive of the small sample interfering with the power properties of the test. Thus a more parsimonious model that included, say, an interaction between the regional covariate and initial income may also fail to be rejected.

5 Monte Carlo Experiments

Our main findings rest on the parameter estimates that we report in the previous tables and figures. A natural question concerns the reliability of the estimates we have obtained using nonparametric estimation techniques for the growth specification given our "small" samples and potential problems. Since these estimates are the primary concern of our study, we felt it pertinent to undertake a set of Monte Carlo experiments to assess the (very) small sample properties of nonparametric model selection in the face of more than one relevant covariate as well as many irrelevant covariates. This should lend credibility and insight into our assessment of growth theories found above. We

Figure 4: Comparison of estimated marginal effects for Geography theory (DKT data).



notice that due to lack of information on certain variables that for any given theory we have samples as small as 167 observations and as large as 271. Therefore we conduct our small sample analysis using both $n_1 = 167$ and 271 observations. Our setup follows the Monte Carlo exercise in Hall, Li and Racine (2007), except that we include more relevant and irrelevant regressors. We judge the performance of the nonparametric model selection exercise based on out-of-sample predictive performance and the behavior of the cross-validated bandwidths.

To be firm, for $i = 1, \ldots, n_1$, with $n_1 = 167$ or 271 we generate the following random variables: $(z_{1i}, z_{2i}, z_{3i}) \in \{0, 1\}$, $Pr[z_{1i} = 1] = .62$, $Pr[z_{2i} = 1] = .71$, $Pr[z_{3i} = 1] = .82$, $(w_{1i}, w_{2i}) = \{0, 1, \ldots, 3\}$, with $Pr[w_{1i} = \ell] = .25$, $\forall \ell$, $Pr[w_{2i} = 0] = .4$ and $Pr[w_{2i} = \ell] = .2$, $\ell \in \{1, 2, 3\}$, while $(x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i})$ are all distributed normally with mean zero and variance one. The variables are drawn so that they exhibit a 0.50 degree of correlation.

Table 8: Partial effect of initial income across various groups for the DKT geography model

Variable	Mean	$Q1$	$Q2$	$Q3$	I_n^c (L)
Western Europe	-0.0246	-0.0276	-0.0246	-0.0216	0.0178
	0.0037	0.0037	0.0037	0.0041	
non-Western Europe	-0.0231	-0.0277	-0.0232	-0.0197	
	0.0037	0.0039	0.0037	0.0039	
Above Median Human Capital	-0.0246	-0.0276	-0.0248	-0.0212	0.0000
	0.0042	0.0037	0.0042	0.0048	
Below Median Human Capital	-0.0224	-0.0277	-0.0224	-0.0191	
	0.0051	0.0037	0.0040	0.0055	
Above Median Investment	-0.0240	-0.0277	-0.0249	-0.0209	0.0000
	0.0039	0.0038	0.0048	0.0041	
Below Median Investment	-0.0230	-0.0276	-0.0225	-0.0202	
	0.0041	0.0037	0.0054	0.0039	
Above Median Population Growth	-0.0224	-0.0277	-0.0222	-0.0195	0.0164
	0.0055	0.0039	0.0054	0.0037	
Below Median Population Growth	-0.0246	-0.0277	-0.0248	-0.0216	
	0.0052	0.0054	0.0041	0.0050	
Above Median Koppen-Geiger	-0.0227	-0.0278	-0.0230	-0.0189	0.0000
	0.0038	0.0055	0.0039	0.0037	
Below Median Koppen-Geiger	-0.0244	-0.0275	-0.0242	-0.0215	
	0.0042	0.0038	0.0050	0.0048	
Above Median % Ice Free Coast	-0.0269	-0.0285	-0.0271	-0.0248	0.0000
	0.0042	0.0050	0.0054	0.0039	
Below Median % Ice Free Coast	-0.0201	-0.0220	-0.0206	-0.0183	
	0.0037	0.0044	0.0047	0.0037	

We generate y_i according to

$$y_i = z_{1i} + x_{1i} + x_{2i} + x_{1i} \cdot x_{2i} + \varepsilon_i,$$

or

$$y_i = z_{1i} + \sqrt{w_{1i}} \cdot x_{1i} + x_{2i} + x_{1i} \cdot x_{2i} + x_{3i}^2 + \varepsilon_i.$$

For both models, ε_i is drawn from a $\mathcal{N}(0, 1)$ distribution. In each model there is more than one relevant continuous variable and there are both categorical and continuous variables that are irrelevant. Both setups also contain nonlinearities to fully highlight the nonparametric approach. We feel that while limited, these two models should provide good insight into how this method performs with a small sample and more than one relevant continuous covariate. Indeed, Fernandez, Ley, and Steel (2001) and Sala-i-Martin, Doppelhofer, and Miller (2004) have both shown using BMA (BACE) that four continuous variables are a part of the true growth model with very high

Table 9: Summary of cross-validated bandwidths for the discrete covariates NP LSCV estimator.

	Median, [10th Percentile, 90th Percentile] of $\hat{\lambda}$				
	$\hat{\lambda}_{z_1}$	$\hat{\lambda}_{z_2}$	$\hat{\lambda}_{z_3}$	$\hat{\lambda}_{w_1}$	$\hat{\lambda}_{w_2}$
$n_1 = 167$					
Model 1	0.26	0.50	0.50	1.00	0.96
	[0.02,0.50]	[0.28,0.50]	[0.29,0.50]	[0.48,1.00]	[0.46,1.00]
Model 2	0.50	0.50	0.50	0.49	0.91
	[0.08,0.50]	[0.32,0.50]	[0.26,0.50]	[0.21,0.90]	[0.44,1.00]
$n_1 = 271$					
Model 1	0.17	0.50	0.50	0.97	1.00
	[0.04,0.48]	[0.33,0.50]	[0.35,0.50]	[0.60,1.00]	[0.62,1.00]
Model 2	0.48	0.50	0.50	0.49	0.91
	[0.13,0.50]	[0.32,0.50]	[0.37,0.50]	[0.27,0.70]	[0.62,1.00]

probability.[23]

Our first assessment is the ability of the cross-validation procedure to smooth away the variables that are indeed not present in the data generating process. We use LCLS to assess if both continuous and discrete variables have been correctly smoothed away. For the categorical variables we use the rule of thumb that if the bandwidth is within 5% of its upper bound that the variable has been smoothed out and for the continuous variables we look at the bandwidth compared to the standard deviation of the data drawn. If the bandwidth is larger than two standard deviations of the regressor we conclude that the continuous variable has been smoothed out of the exercise. For our 1000 replications we note the median, 10th and 90th percentiles of the cross-validated bandwidths.

We see from Tables 9 and 10 that the median results suggest that the method is correctly smoothing away irrelevant discrete and continuous variables. For instance, in model 1, only z_1, x_1 and x_2 are relevant. Table 9 shows that h_{z_1} is the only categorical bandwidth whose median value is significantly different from its upper bound. At the same time, the median bandwidths for x_1 and x_2 in Table 10 correctly suggest that they are relevant while each of the other median bandwidths correctly suggest irrelevance. Although the results are good for the smaller sample,

[23]The four that each found are different, with the exception of initial income, but both winnow the large set of potential covariates down to a relatively small set that is manageable for empirical studies employing nonparametric estimation methods.

Table 10: Summary of cross-validated bandwidths for the continuous covariates NP LSCV estimator.

	Median, [10th Percentile, 90th Percentile] of $\hat{h}$				
	$\hat{h}_{x_1}$	$\hat{h}_{x_2}$	$\hat{h}_{x_3}$	$\hat{h}_{x_4}$	$\hat{h}_{x_5}$
$n_1 = 167$					
Model 1	0.43	0.46	35.70	27.97	49.44
	[0.28,0.55]	[0.32,0.69]	[1.22,4898.59]	[1.19,4898.59]	[1.35,4959.27]
Model 2	0.47	0.57	0.41	28.61	28.43
	[0.29,0.59]	[0.42,0.87]	[0.23,0.64]	[1.34,4971.78]	[1.17,5127.24]
$n_1 = 271$					
Model 1	0.43	0.42	480.37	58.49	48.56
	[0.29,0.53]	[0.32,0.51]	[1.61,3064.97]	[1.28,3010.38]	[1.26,4480.14]
Model 2	0.48	0.55	0.41	34.94	44.95
	[0.34,0.59]	[0.40,0.69]	[0.28,0.54]	[1.08,4395.69]	[1.31,4954.60]

it is obvious that the ability to smooth away irrelevant regressors is enhanced by additional data. We note again that this is also for data that are drawn to have a 0.5 degree of correlation, lending further evidence that the method works well when variables are correlated.

There is one issue in the above tables. For the second model, z_1, a relevant categorical regressor, has a median bandwidth equal to its upper bound when $n_1 = 167$. This means that in at least half of the Monte Carlo replications that the variable is incorrectly smoothed out. We do note here that the tenth percentile (0.08) is far smaller than the tenth percentile of the 'truly irrelevant' regressors. At the same time, when the sample is increased to 276, the median bandwidth is 0.48. This is an improvement, but it is still very close to the upper bound. This shows that it is not sufficient to simply look at the bandwidths and we must also use formal tests to determine relevance, especially in small samples with many regressors.

Our second assessment involves the model's predictive performance where we generate data, *independent* from the original draw, from the same DGP of size $n_2 = 1,000$. Predictive performance is assessed via $PMSE = 1/n_2 \sum_{j=1}^{n_2} (\hat{y}_j - y_j)^2$. We consider three parametric models, an incorrect linear model (PI-ALL) that includes all the variables, an incorrect linear model that only includes the relevant variables (PI-ONLY) and the correct nonlinear, interactions model (PC) as well as the

Table 11: Out-of-sample predictive PMSE performance for parametric and nonparametric models containing irrelevant regressors for $n_1 = 100$ ($\rho = 0.5$).

	Median, [10th Percentile, 90th Percentile] of PMSE			
	NP-LSCV	PI-ALL	PI-ONLY	PC
$n_1 = 167$				
Model 1	1.44	2.41	2.40	1.02
	[1.26,2.19]	[2.19,2.65]	[2.17,2.59]	[0.97,1.09]
Model 2	2.49	7.92	8.00	1.04
	[2.00,3.15]	[7.09,8.96]	[7.09,8.91]	[0.98,1.11]
$n_1 = 271$				
Model 1	1.33	2.34	2.39	1.02
	[1.20,1.51]	[2.14,2.57]	[2.18,2.59]	[0.96,1.07]
Model 2	2.13	7.79	8.04	1.02
	[1.85,2.52]	[6.92,8.88]	[7.13,9.23]	[0.97,1.10]

LCLS cross-validated results. The estimates for the first two models should lead to inconsistent estimates while the second two are consistent estimators. Table 11 suggests that while the *correctly* specified parametric model dominates all the competitors, as expected, the performance of the nonparametric model relative to the two incorrect models is notable. For model 1, in the smaller sample, the relative performance is over 40% better than the incorrectly linear model with every variable included and nearly 40% better than the incorrectly specified linear model with only the relevant variables. We also note that this relative performance improves with the sample size as more data helps the nonparametric estimates, but does not ameliorate the inconsistent parametric estimators.

In summary, we see that even with the threat of the curse of dimensionality, the nonparametric estimators perform well in small samples with relatively large numbers of relevant and irrelevant variables. We note here that this level of performance testing with such small samples and so many regressors has not been attempted in the literature. We also ran a sample with $n_1 = 100$ data points as in HLR and found the results to be acceptable.[24] That being said, the performance of the estimators improves as we increase the sample size. The ability to smooth out irrelevant

[24]These tables can be obtained from the authors upon request.

regressors and the prediction power with $n_1 = 271$ observations relative to $n_1 = 167$ observations leads us to put more faith in the results from the extended DKT data set. This leads us to suggest that future research in nonparametric growth regressions should not only focus on the appropriate control variables, but it should also attempt to maximize the number of observations.

6 Discussion and Conclusion

This paper has offered a unique perspective into the debate over 'relevant' growth theories while allowing for specification uncertainty. The use of nonparametric modelling techniques allows the inclusion of irrelevant variables at almost no harm to the predictions of the model given the ability to automatically remove them. This is an appealing feature of nonparametric methods in general and is critical for studying growth given recent findings that the growth process may be highly nonlinear coupled with the fact that many variables may be weak predictors of growth. While these methods are still plagued by the curse of dimensionality, excluding them from being used in a kitchen sink type manner, they do allow a litmus test of potential nonlinearity for a handful of variables, the results of which can then be used to guide BMA parametric specifications as well as more in depth studies of any given growth theory.

Our results for the singular theories follow along the lines of DKT, with several of the Solow variables, most notably initial income, robust to moving across theories. In contrast however, we see that while initial income is a relevant regressor for explaining growth, its appearance in the growth model seems to suggest a nonlinear impact on growth rates. We also agree with DKT that the macro variables seem to generate the greatest improvement in fit over other theories and once again that these macro variables display a nonlinear effect on overall country growth. This is suggestive that both the BMA results of DKT and the nonparametric model selection techniques employed here are coming to the same conclusions about which variables impact growth, but are differing in the explicit nature of that impact.

Our deeper investigation into the macro and geography theories reveals that thinking of each individual growth theory as impacting growth in a linear fashion is incorrect in two facets: linearity and individuality. When we examined the macro theory we found credible evidence of both nonlinearities in many of the variables as well as parameter heterogeneity arising due to potential interactions between initial income and the other variables in the model. However, for the geography model we found that linearity was a common theme for many of the variables in the model but parameter heterogeneity in the initial income variable was still prevalent. What is interesting is that many of the attempts to model parameter heterogeneity have focused on initial income, which is also present in our results. It appears that for any given theory initial income plays a role in determining how the other variables impact growth. This is also an important finding in terms of poverty or development traps (see Azariadis and Stachurski, 2005). Thus, researchers using BMA to determine the posterior probability of inclusion for either a growth theory or a specific variable may wish to think of alternative model specifications to enlarge the model space of potential growth models. This should provide more illuminating insights into the correct representation of a country wide growth model and the variables it is composed of.

While the results here do not account for endogeneity or model uncertainty, they have shed light on the prevalence of both nonlinearities and parameter heterogeneity in growth models while allowing for inclusion of irrelevant proxy variables in a given growth theory under examination. Future attempts to perform Bayesian Model Averaging or other model averaging methods on growth regressions can use these findings as motivation to widen the model space and to develop methods to handle nonparametric estimators. Also, for those researchers focused on an individual theory, the results here imply that nonlinearities and parameter heterogeneity not be taken lightly in the modeling approach. Failure to account for these features of the growth model may seriously mislead the researcher, as would failure to account for endogeneity or model uncertainty.

The main point is that regardless if one is interested in *which* growth theories are robust or the implications of a specific growth theory, failure to account for nonlinearities, variable interactions, and parameter heterogeneity could lead to gross misconceptions about what is really going on. Furthermore, one needs to draw into question any growth study that does not think seriously about nonlinearities as their impact can sharply dictate parameter estimates. In fact, without properly controlling the model space, one cannot learn about theory robustness. And, if one ignores nonlinearities, policy recommendations based off a specific growth theory may not offer the correct prescription. In either setting, nonlinearities provide evidence about the underlying growth dynamics and ignoring their testimony may result in a mistrial when judging any specific growth theory.

As more data becomes available, both for variables within a theory as well as for countries in general, these nonparametric model selection methods will prove invaluable, given the results from our small sample exercises. Indeed, even looking at the individual theories with an expanded, albeit heterogeneous, data set, we find that all of the Solow variables began to display themselves as relevant across theories. We also reaffirm many of the same insights drawn from the smaller dataset, providing further credence to the small sample performance of the nonparametric model selection techniques.

References

[1] Azariadis, C. and J. Stachurski, 2005. Poverty Traps, in P. Aghion and S.N. Duraluf, eds. Handbook of Economic Growth, Vol. 1 (Elsevier, USA).

[2] Brock, W. and S. Durlauf, 2001. "Growth Empirics and Reality," *World Bank Economic Review*, 15(2): 229-272.

[3] Ciccone, A. and M. Jarocinski, 2007. "Determinants of Economic Growth: Will Data Tell?," working paper, Universitat Pompeu Fabra.

[4] Doppelhofer, G. and M. Weeks, (forthcoming). "Jointness of Growth Determinants," *Journal of Applied Econometrics*.

[5] Durlauf, S., and P. Johnson, 1995. "Multiple Regimes and Cross Country Growth Behavior," *Journal of Applied Econometrics*, 10(4): 365-384.

[6] Durlauf, S.N., P. Johnson, and J. Temple, 2005. "Growth Econometrics," in *Handbook of Economic Growth*, P. Aghion and S.N. Durlauf (Eds.), Amsterdam, North Holland, Volume 1A, 555-677.

[7] Durlauf, S., A. Kourtellos, and A. Minkin, 2001. "The Local Solow Growth Model," *European Economic Review*, 45(4-6): 928-940.

[8] Durlauf, S., A. Kourtellos, and C. Tan, 2008. "Are any Growth Theories Robust?" *Economic Journal* 118, 329-346.

[9] Easterly, W., and R. Levine, 1997. "Africa's Growth Tragedy: Politics and Ethnic Divisions," *Quarterly Journal of Economics*, 112(4): 1203-1250.

[10] Efroymson, M.A.,1960, "Multiple Regression Analysis," in *Mathematical Methods for Digital Computers*, edited by A. Ralston and H. S. Wilf. Wiley, New York.

[11] Fernandez, C., E. Ley and M. Steel, 2001. "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics* , 16(5): 563-576.

[12] Galor, O. and D. Weil 2000. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond," *American Economic Review* 90, 806-829.

[13] Hall, P., Q. Li, and J. Racine 2007. "Nonparametric Estimation of Regression Functions in the Presence of Irrelevant Regressors," *Review of Economics and Statistics* 89, 784-789.

[14] Hendry, D. and H.-M. Krolzig, 2004. "We Ran One Regression," *Oxford Bulletin of Economics and Statistics*, 66(5): 799-810.

[15] Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky, 1999. "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14: 382-417.

[16] Hoover, K., and S. Perez, 2004. "Truth and Robustness in Cross-country Growth Regressions," *Oxford Bulletin of Economics and Statistics*, 66(5): 765-798.

[17] Hsiao, C., Q. Li and J. Racine, 2007. "A Consistent Model Specification Test with Mixed Categorical and Continuous Data," *Journal of Econometrics* 140(2): 802-826.

[18] Kalaitzidakis, P., T. Mamuneas, and T. Stengos, 2000. "A Non-Linear Sensitivy Analysis of Cross Country Growth Regressions," *Canadian Journal of Economics*, 33(3): 604-617.

[19] 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.

[20] Kourtellos, A., 2003. "Modeling Parameter Heterogeneity in Cross-Country Growth Regression Models," *mimeo*, University of Cyprus.

[21] Lavergne, P., and Q. Vuong (2000). "Nonparametric Significance Testing," *Econometric Theory* 16, 576-601.

[22] Leamer, E.E.,1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, Wiley, New York.

[23] Leamer, E.E., 1983. "Let's Take the Con Out of Econometrics," *American Economic Review*, 73(1): 31-43.

[24] Lee, K., M. Pesaran, and R. Smith, 1997. "Growth and Convergence in a Multi Country Empirical Stochastic Solow Model," *Journal of Applied Econometrics*, 12(4): 357-392.

[25] Levine, R. and D. Renelt, 1992. "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, 82(4): 942-963.

[26] Ley, E. and M.F.J. Steel, 2007. "Jointness in Bayesian Variable Selection with Applications to Growth Regression," *Journal of Macroeconomics* 29(3): 476-493.

[27] Ley, E. and M.F.J. Steel, (forthcoming). "On the Effect of Prior Assumptions in BMA with Applications to Growth Regression," *Journal of Applied Econometrics.*

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

[29] Liu, Z., and T. Stengos, 1999. "Non-Linearities in Cross Country Growth Regressions: A Semiparametric Approach," *Journal of Applied Econometrics*, 14(5): 527-538.

[30] Mamuneas, T., 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.

[31] Masanjala, W., and C. Papageorgiou, 2004. "The Solow Model with CES Technology: Nonlinearities with Parameter Heterogeneity," *Journal of Applied Econometrics*, 19(2): 171-201.

[32] Masanjala, W., and C. Papageorgiou, forthcoming. "Rough and Lonely Road to Prosperity: A Reexamination of the Sources of Growth in Africa Using Bayesian Model Averaging," *Journal of Applied Econometrics* .

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

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

[35] Sala-i-Martin, X., 1997a. "I Just Ran Two Million Regressions," *American Economic Review*, 87(2): 178-183.

[36] Sala-i-Martin, X., 1997b. "I Just Ran Four Million Regressions," *National Bureau of Economic Research Working Paper no. 6252.*

[37] Sala-i-Martin, X., G. Doppelhoffer, and R. Miller, 2004. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, 94(4): 813-835.

[38] Tan, C., 2007. "No One True Path to Development: Uncovering the Interplay Between Geography, Institutions, and Ethnic Fractionalization in Economic Development," *mimeo*, Tufts University.

[39] Temple, J., 1998. "Robustness Tests of the Augmented Solow Model," *Journal of Applied Econometrics*, 13(4): 361-375.

[40] Temple, J., 1999. "The New Growth Evidence," *Journal of Economic Literature*, Vol. 37(1): 112-156.

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

Appendices

1 Testing for Correct Parametric Specification

To assess the correct estimation strategy, we utilize the Hsiao, Li and Racine (2007) specification test for mixed categorical and continuous data. The null hypothesis is that the parametric model $(f(x_i, \beta))$ is correctly specified $(H_0 : \Pr[E(y_i|x_i) = f(x_i, \beta)] = 1)$ against the alternative that it is not $(H_1 : \Pr[E(y_i|x_i) = f(x_i, \beta)] < 1)$. The test statistic is based on $I \equiv E\left(E(u|x)^2 f(x)\right)$, where $u = y - f(x, \beta)$. I is non-negative and equals zero if and only if the null is true. The resulting test statistic is

$$T_n^a = \frac{n\sqrt{h_1 h_2 \cdots h_q}\widehat{I}_n}{\widehat{\sigma}_n^a} \sim N(0,1), \tag{A1}$$

where

$$\begin{aligned}\widehat{I}_n^a &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\widehat{u}_i\widehat{u}_j K_{\hat{h},\hat{\lambda}^u,\hat{\lambda}^o}, \\ \widehat{\sigma}_n^{a2} &= \frac{2h_1h_2\cdots h_q}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\hat{u}_i^2\hat{u}_j^2 K^2_{\hat{h},\hat{\lambda}^u,\hat{\lambda}^o},\end{aligned}$$

with $\widehat{u}_i = y_i - f\left(x_i, \widehat{\beta}\right)$ the residual from the *parametric* model, $K_{\hat{h},\hat{\lambda}^u,\hat{\lambda}^o}$ is the product kernel discussed previously, q is the number of continuous regressors, and $\hat{h}, \hat{\lambda}^u, \hat{\lambda}^o$ are the bandwidths obtained via LSCV. If the null is false, T^a diverges to positive infinity. Unfortunately, the asymptotic normal approximation performs poorly in finite samples and a bootstrap method is generally suggested for approximating the finite sample null distribution of the test statistic. Formally, the steps involved in computing the wild bootstrap statistic are as follows:

1. For $i = 1, 2, \ldots, n$, generate the two-point wild bootstrap error $u_i^* = \left[\left(1-\sqrt{5}\right)/2\right]\widehat{u}_i$, where $\widehat{u}_i = y_i - f\left(x_i, \widehat{\beta}\right)$ with probability $r = \left(1-\sqrt{5}\right)/2\sqrt{5}$ and $u_i^* = \left[\left(1+\sqrt{5}\right)/2\right]\widehat{u}_i$ with probability $1-r$.

2. Create $y_i^* = f\left(x, \widehat{\beta}\right) + u_i^*$ $(i = 1, 2, \ldots, n)$. The resulting sample $\{x_i, y_i^*\}_{i=1}^n$ is called the bootstrap sample.

3. Obtain bootstrap residuals $\widehat{u}_i^* = y_i^* - f\left(x_i, \widehat{\beta}^*\right)$ $(i = 1, 2, \ldots, n)$, where $\widehat{\beta}^*$ is the parametric estimator of β estimated from the bootstrap sample.

4. Use the bootstrap residuals to compute the test statistic $T_n^{a*} = n\left(h_1h_2\cdots h_q\right)^{1/2}\widehat{I}_n^{a*}/\widehat{\sigma}_n^{a*}$, where $\widehat{I}_n^{a*}$ and $\widehat{\sigma}_n^{a*}$ are the same as $\widehat{I}_n^a$ and $\widehat{\sigma}_n^a$ except that $\widehat{u}_i$ is replaced by $\widehat{u}_i^*$.

5. Repeat steps (1-4) a large number (B) of times and then construct the empirical distribution of the B bootstrap test statistics, $\{T_n^{a*}\}_{b=1}^B$. This bootstrap empirical distribution is used to approximate the null distribution of the test statistic T_n^a. We reject H_0 if $T_n^a > T_{n(\alpha B)}^{a*}$, where $T_{n(\alpha B)}^{a*}$ is the upper α -percentile of $\{T_n^{a*}\}_{b=1}^B$.

Steps 2 through 4 heuristically ensure that conditional on the random sample, the bootstrap sample is generated by the null model. Conditional on $\{x_i, y_i\}_{i=1}^n$, u_i^* has zero mean and the bootstrap statistic obtained in step 3 approximates the null distribution of the test statistic whether the null hypothesis is true or not.

2 Testing for Variable Significance

While the properties of LSCV discovered by HLR suggest that irrelevant variables are removed, statistically there is no way to determine joint (in)significance by simply appealing to the bandwidths returned. A formal test for joint significance of variables is thus warranted to make more precise statements about the relevance of variables entering into the model.

To determine whether or not a set of variables are jointly significant, we utilize the Lavergne and Vuong (2000) test modified to allow for mixed categorical and continuous data. Consider a nonparametric regression model of the form

$$y_i = m(w_i, z_i) + u_i. \tag{B1}$$

Here we discuss the case where all the variables in z are continuous, but w may contain mixed data. Let w have dimension r and z have dimension $q - r$. The null hypothesis is that the conditional mean of y does not depend on z.

$$H_0 : E(y|w, z) = E(y|w) \tag{B2}$$

Define $u = y - E(y|w)$. Then $E(u|x) = 0$ under the null and we can construct a test statistic based on

$$E\{uf_w(w)\,E[uf_w(w)\,|x]\,f(x)\} \tag{B3}$$

where $f_w(w)$ and $f(x)$ are the pdf's of w and $x = (w, z)$, respectively. A feasible test statistic is given by

$$\widehat{I}_n^b = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} (y_i - \widehat{y}_i)\,\widehat{f}_w(w_i)\,(y_j - \widehat{y}_j)\,\widehat{f}_w(w_j)\,W(x_i, x, h, \lambda^o, \lambda^u) \tag{B4}$$

where $W(x_i, x, h, \lambda^o, \lambda^u) = \prod_{s=1}^{q_c} K\left(\frac{x_{si}^c - x_s^c}{h_s}\right) \prod_{s=1}^{q_u} l^u(x_{si}^u, x_s^u, \lambda_s^u) \prod_{s=1}^{q_o} l^o(x_{si}^o, x_s^o, \lambda_s^o)$ is the product kernel mentioned previously and

$$\widehat{f}_w(w_i) = \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} W(w_i, w, h_w, \lambda_w^o, \lambda_w^u) \tag{B5}$$

is the leave-one-out estimator of $f_w(w_i)$. The leave one out estimator of $E(y_i|w_i)$ is

$$\widehat{y}_i = \frac{1}{(n-1)\widehat{f}_w(w_i)} \sum_{j=1, j\neq i}^{n} y_j W(w_i, w, h_w, \lambda_w^o, \lambda_w^u). \tag{B6}$$

One shortcoming of this test is that it requires the researcher to estimate two sets of bandwidths, one for the model under the null and another for the model under the alternative. For large samples this may be computationally expensive. However, given the typical size of a growth related dataset this expense is not too severe to steer away from.

Under the null we have that

$$T_n^b = (nh_1h_2\cdots h_q)^{1/2}\,\widehat{I}_n^b/\widehat{\sigma}_n^b \to N(0,1) \tag{B7}$$

where

$$\widehat{\sigma}_n^{b2} = \frac{2h_1h_2\cdots h_q}{n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(y_i-\widehat{y}_i)^2\widehat{f}_w(w_i)(y_j-\widehat{y}_j)^2\widehat{f}_w(w_j)W(x_i,x,h,\lambda^o,\lambda^u) \tag{B8}$$

Again, the asymptotic distribution does not work well for finite samples. A bootstrap procedure is suggested instead. The bootstrap test statistic is obtained via the following steps:

1. For $i = 1, 2, \ldots, n$, generate the two-point wild bootstrap error $u_i^* = \left[\left(1-\sqrt{5}\right)/2\right]\widehat{u}_i$, where $\widehat{u}_i = y_i - \widehat{y}_i$ with probability $r = \left(1-\sqrt{5}\right)/2\sqrt{5}$ and $u_i^* = \left[\left(1+\sqrt{5}\right)/2\right]\widehat{u}_i$ with probability $1-r$.

2. Use the wild bootstrap error u_i^* to construct $y_i^* = \widehat{y}_i + u_i^*$, then obtain the kernel estimator of $E^*\left(y_i^*|w_i\right)f_w\left(w_i\right)$ via

$$\begin{aligned}\widehat{y}_i^*\widehat{f}_w(w_i) &= \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} y_j^*W\left(w_i,w,h_w,\lambda_w^o,\lambda_w^u\right)\\ \widehat{y}_i^* &= \frac{1}{(n-1)\widehat{f}_w(w_i)}\sum_{=1,j\neq i}^{n} y_j^*W\left(w_i,w,h_w,\lambda_w^o,\lambda_w^u\right)\end{aligned}$$

   The estimated density-weighted bootstrap residual is

$$\begin{aligned}\widehat{u}_i^*\widehat{f}_w\left(w_i\right) &= \left(y_i^*-\widehat{y}_i^*\right)\widehat{f}_w\left(w_i\right)\\ &= y_i^*\widehat{f}_w\left(w_i\right)-\widehat{y}_i^*\widehat{f}_w\left(w_i\right)\end{aligned}$$

3. Compute the standardized bootstrap test statistic T_n^{b*} where y^* and $\widehat{y}^*$ replace y and $\widehat{y}$ wherever they occur.

4. Repeat steps 1-3 a large number (B) of times and obtain the empirical distribution of the B bootstrap test statistics. Let $T_{n(\alpha B)}^{b*}$ denote the the α-percentile of the bootstrap distribution. We will reject the null hypothesis at significance level α if $T_n^b > T_{n(\alpha B)}^{b*}$.

3 Testing Equality of Two PDFs

To test whether two vectors of data $\{x_i\}_{i=1}^{n_1}$ and $\{z_i\}_{i=1}^{n_2}$ are drawn from the same distribution we employ the Li (1996) test. The Li (1996) test, which tests the null hypothesis $H_0: f(x) = g(x)$ for all

x, against the alternative $H_1 : f(x) \neq g(x)$ for some x, works with either independent or dependent data. The test statistic used to test for the difference between the two unknown distributions (which Fan and Ullah 1999 show goes asymptotically to the standard normal), predicated on the integrated square error metric on a space of density functions, $I(f,g) = \int_x (f(x) - g(x))^2 \, dx$, is

$$T_n^c = \frac{(n_1 n_2 h_1 h_2 \cdots h_q)^{1/2} \, \widehat{I}_n^c}{\widehat{\sigma}_n^c} \sim N(0,1), \tag{C1}$$

where

$$\begin{aligned} \widehat{I}_n^c &= \frac{1}{n_1^2} \sum_{i=1}^{n_1} \sum_{j=1, j \neq i}^{n_1} K_{h,ij}^x + \frac{1}{n_2^2} \sum_{i=1}^{n_2} \sum_{j=1, j \neq i}^{n_2} K_{h,ij}^z \\ &\quad - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1, j \neq i}^{n_2} K_{h,ij}^{xz}, \end{aligned}$$

and

$$\widehat{\sigma}_n^{c2} = \frac{h_1 h_2 \cdots h_q}{n_1 n_2} \left\{ \sum_{i=1}^{n_1} \sum_{j=1, j \neq i}^{n_1} \frac{\left[K_{h,ij}^x\right]^2}{n_1/n_2} + \sum_{i=1}^{n_2} \sum_{j=1, j \neq i}^{n_2} \frac{\left[K_{h,ij}^z\right]^2}{n_2/n_1} + 2 \sum_{i=1}^{n_1} \sum_{j=1, j \neq i}^{n_2} \left[K_{h,ij}^{xz}\right]^2 \right\}, \tag{C2}$$

where $K_{h,ij}^x = \prod_{s=1}^q h_s^{-1} K((x_{is} - x_{js})/h_s)$, $K_{h,ij}^z = \prod_{s=1}^q h_s^{-1} K((z_{is} - z_{js})/h_s)$, and $K_{h,ij}^{xz} = \prod_{s=1}^q h_s^{-1} K((x_{is} - z_{js})/h_s)$.

Again, if the null is false, T^c diverges to positive infinity. Unfortunately, the asymptotic normal approximation performs poorly in finite samples and a bootstrap method is generally suggested for approximating the finite sample null distribution of the test statistic. Formally, this is accomplished by randomly sampling with replacement from the pooled data. The steps are as follows:

1. Randomly draw $n_1 + n_2$ observations with replacement from the pooled data set. Call the first n_1 observations $\{x_i^*\}_{i=1}^{n_1}$ and the remaining n_2 observations $\{z_i^*\}_{i=1}^{n_2}$.

2. Use the bootstrap data to compute the test statistic $T_n^{c*} = (n_1 n_2 h_1 h_2 \cdots h_q)^{1/2} \, \widehat{I}_n^{c*} / \widehat{\sigma}_n^{c*}$, where $\widehat{I}_n^{c*}$ and $\widehat{\sigma}_n^{c*}$ are the same as $\widehat{I}_n^c$ and $\widehat{\sigma}_n^c$ except that $\{x_i\}_{i=1}^{n_1}$ and $\{z_i\}_{i=1}^{n_2}$ are replaced by $\{x_i^*\}_{i=1}^{n_1}$ and $\{z_i^*\}_{i=1}^{n_2}$, respectively.

3. Repeat steps (1-2) a large number (B) of times and then construct the empirical distribution of the B bootstrap test statistics, $\{T_n^{c*}\}_{b=1}^B$. This bootstrap empirical distribution is used to approximate the null distribution of the test statistic T_n^c. We reject H_0 if $T_n^c > T_{n(\alpha B)}^{c*}$, where $T_{n(\alpha B)}^{c*}$ is the upper α -percentile of $\{T_n^{c*}\}_{b=1}^B$.