Are any Growth Theories Linear?*

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

February 1, 2010

Abstract

Recent research on growth empirics has been focused on resolving model and variable uncertainty. The conventional approach has been to assume a linear growth process and then to proceed with investigating the relevant variables that determine cross-country growth. This paper questions the linearity assumption underlying the vast majority of such research and uses recently-developed nonparametric techniques to consider nonlinearities and variable selection jointly. We show that inclusion of nonlinearities is necessary for determining the empirically relevant variables and uncovering key mechanisms of the growth process. We demonstrate this by considering a specific theory of growth, geography, and show important differences in the impact of geographic conditions on economic growth overlooked by conventional parametric models.

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

Keywords: Growth nonlinearities, parameter heterogeneity, nonparametric regression methods, variable selection.

*We thank Antonio Ciccone and two anonymous referees for insightful comments as well as participants of talks given at the University of Nevada, Las Vegas, the 2009 Meeting of the Midwest Econometric Group at Purdue University and the 2009 North American Summer Meeting of the Econometric Society at Boston University. We also acknowledge superb research assistance from William Larson funded by a research grant from the Research Department of the IMF. The data and code used in this paper are available from the authors upon request. A previous version of this paper was circulated under the title "Are Any Growth Theories Linear? Why We Should Care What the Evidence Tells Us". 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

The intrigue of uncovering a singular process that dictates economic growth has led to numerous theories explicating a link between some variable(s) of interest and economic output. This explosion of theories on economic growth has simultaneously put forth an empirical conundrum known as "theory openendedness" (Brock and Durlauf, 2001). Theory openendedness suggests that while numerous theories may indeed explain economic output, no one theory rules out another theory as a definitive predictor of cross-country growth. This has sharpened the need for viable model selection and averaging techniques to parse through the vast amounts of data being used to empirically test growth models. Such techniques allow growth empiricists to tenably focus on the variables which produce robust relationships with economic growth.

Most model selection and averaging studies assume a linear growth process so that functional form uncertainty can be abrogated - see e.g., Fernandez, Ley and Steel (2001), Sala-i-Martin, Doppelhoffer and Miller (2004), and Durlauf, Kourtellos and Tan (2008), DKT hereafter. However, an emerging theme in the literature has been the appearance of significant nonlinearities in cross-country growth regressions - see Maasoumi, Racine, and Stengos (2007) for the most current research. From this vantage, it is important to identify the nonlinearities in the growth process, for a specific growth theory, so that they can be used to extend the model space in model averaging/selection exercises.

While the insights of the empirical growth papers employing model averaging techniques are valuable in and of themselves, their foundation of a priori functional form specification limits the scope of these methods in 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.

In this paper, we use recently-developed methods for nonparametric regression to investigate potential non-linearities in the growth process at the same time as selecting relevant variables. 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 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 are robust to functional form misspecification (specification uncertainty) and have the ability to remove irrelevant variables that have been added by the researcher (variable uncertainty).

Our results highlight the importance of accounting for nonlinearities across the spectrum of growth variables, including the Solow model variables themselves. We find that nearly all individual growth theories possess nonlinearities and have heterogenous partial effects. 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 linear, parametric models. Second, the results here should prove useful to researchers looking for additional motivation for incorporating nonlinearities into the BMA paradigm. For example, we find that introducing simple nonlinearities into the parametric model for the macroeconomic theory produces reliable results. Third, this paper outlines an approach for determining potential nonlinearities which may subsequently guide model selection. Subsequently, we showcase how our approach can unravel important insights that parametric growth models would overlook by focusing on geography as a potential theory of growth.

The remainder of the paper is organized as follows. Section 2 provides a brief literature review while Section 3 takes a look at the data used in estimation. Section 4 presents an intuitive overview of recently-developed nonparametric methods which our results stem from. Section 5 presents the main results. Section 6 concludes.

2 A Brief Literature Review

2.1 Model uncertainty in growth empirics

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, 1983) developed a method we now call Extreme Bounds Analysis (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.

Melding cross-country growth regressions with various conditioning sets dates back to the work of Levine and Renelt (1992) who used Leamer's EBA to examine 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.[1] It was not until the turn of the century that growth empiricists, notably Brock and Durlauf (2001), and Fernandez, Ley and Steel (2001), incorporated model averaging methods, and specifically Bayesian Model Averaging (BMA), in growth regression. 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.[2]

2.2 Incorporating nonlinearities in model uncertainty

The model averaging exercises have shed new light on important and telling growth features. However, one area where these methods have been less used has been in examining the simultaneous impact of model uncertainty and nonlinearities within the growth process. Nonparametric kernel methods have been used to uncover nonlinearities in empirical growth regressions. Much of this research is due to the work of Thanasis Stengos which was in turn heavily influenced by the pioneering contribution of Durlauf and Johnson (1995). Liu and Stengos (1999) consider an additive partly linear growth specification. Their research influenced a large number of studies within the semiparametric domain (e.g., see Durlauf, Kourtellos and Minkin 2001, Mamuneas, Savvides, and Stengos 2006, Ketteni, Mamuneas and Stengos 2007, Vaona and Schiavo 2007, and Minier 2007a,b). To our knowledge the only paper that has considered model uncertainty 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.

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.

[1] Sala-i-Martin (1997) developed alternative methods that still penalized non-robust variables, albeit less harshly than EBA. While these 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 very large pool of candidate variables.

[2] Model averaging in growth empirics has shown an amazing expansion with such contributions as Sala-i-Martin, Doppelhoffer and Miller (2004), Ley and Steel (2007), Kourtellos, Tan and Zhang (2007), DKT and Ciccone and Jarocinski (forthcoming), just to name a few.

As an alternative, Maasoumi, Racine and Stengos (2007) consider a fully nonparametric growth structure. Specifically, they focus on what happens to predicted growth rates and residuals over time. In this paper, we deviate from Maasoumi, Racine and Stengos (2007)'s focus, but exploit their methodology to determine which growth theories display nonlinear tendencies. It is also worth noting that Hoeting, Raftery and Madigan (2002) and more recently Gottardo and Raftery (2007) make the same point about the need to consider variable selection and functional form jointly. However, their potential nonlinearities are typically far more restrictive than those we use in this paper.

3 The Data

Our primary data source is DKT, which to our knowledge, is the richest panel dataset for cross-country growth regressions. While we have considered other datasets in the literature, we have concluded that DKT is well-suited for our analysis as it allows us to maximize the number of country-time observations, a primary objective in the use of our nonparameteric method to help ensure that our results are reliable.[3] In its original form the DKT dataset features an unbalanced panel over three periods; 1965-74 (53 countries), 1975-84 (54 countries), and 1985-94 (57 countries). We extend DKT in two important dimensions to maximize our sample size; first we enlarge the time horizon to 1960-2000 when possible, and second we consider 5-year rather than 10-year intervals.[4]

Next we briefly review several key features of the DKT and the main variables considered. The DKT dataset contains data for the traditional Solow model (initial income, investment rate, human capital and population growth) as well as variables that compose several of the contending growth theories being debated today: fractionalization, institutions, demographics, geography and macroeconomic policy. At least two variables for each theory are used on our statistical analysis. The dependent variable is the average growth rate of real per worker GDP. Data for income are from PWT 6.1 while data for capital per worker are from Caselli (2005).

Following DKT we organize the determinants of growth into 6 theories and follow the existing

[3]Subsequent analysis with alternative datasets or with a richer set of theories is also an interesting possibility, but we leave that for future research.

[4]Our extended dataset along with a Data Appendix reporting in detail the steps followed to construct these data are available upon request. We have also used 10-year panel data and have confirmed robustness of our results. This was done in response to Johnson, Larson, Papageorgiou and Subramanian (2009) who pointed out serious measurement problems in using higher frequency GDP growth data from Penn World Tables. These results are also available upon request.

literature choosing empirical proxies as follows: (1) Neoclassical growth variables: consist of the logarithm of real GDP per worker in the initial year (*Initial Income*), the logarithm of the average percentage of a country's working age population in secondary school (*Human Capital*), the logarithm of the average investment to GDP ratio (*Investment*), and the logarithm of population growth plus 0.05 over the corresponding periods (*Population Growth*). (2) Demography: is measured using the reciprocal of life expectancy at age 1 (*Life Expectancy*) and the fertility rate (Fertility). (3) Macroeconomic policy: is measured using three proxies; within-period ratio of exports plus imports to GDP, filtered for the relation of this ratio to the logs of population and area (*Openness*), the inflation rate for each period (*Inflation*), and within-period ratio of government consumption, net of outlays on defense and education, to GDP (*Net. Govt. Cons*). (4) Geography: is measured using a climate variable, the percentage of a country's land area classified as tropical and subtropical based on the Köppen-Geiger classification system for climate zones (*Köppen-Geiger*), and a geographic accessibility/isolation variable, the percentage of a country's land area within 100km of an ice-free coast based on Rodrik, Subramanian, and Trebbi (2002) and Sachs (2003) (*% Ice Free Coast*). (5) Fractionalization: is measured by linguistic fractionalization as constructed by Easterly and Levine (1997), and Alesina et al. (2003); a measure of "the degree of tension within a country attributable to racial, nationality (*Ethnic Tension*), or language divisions" from the International Country Risk Guide (*Language*). (6) Institutions: is measured using eight variables; the within-period average constraints on executive power (*Exec. Constraints*), the risk of expropriation of private investments, as in Acemoglu, Johnson, and Robinson (2001) (*Exprop. Risk*), an index of legal formalism based on the number of procedures for collecting on a bounced check developed by Djankov, La Porta, Lopez-de-Silanes, and Shleifer (2002) (*Eviction*), an index for the quality of governance in 1996 using a composite of governance index developed by Kaufmann, Kraay and Mastruzzi (2005) (*KKZ96*), in addition to *Bureaucratic Quality*, *Civil Liberties*, *Political Rights* and *Rule of Law*.[5]

Finally, regional heterogeneity and time variation are captured using categorical variables. In a nonparametric analysis these variables are the equivalent of standard dummy variables in linear parametric modelling. However, we do not need to introduce separate variables for Sub-Saharan African countries and East Asian countries, say. Instead, we introduce a variable that contains the region to which a country is deemed to belong to or the time period under investigation. In our data we follow the regional country classification of the World Bank while our categorical variable

[5]We thank Andros Kourtellos for providing detailed data on these variables.

for time is constructed based on the time interval under estimation. A further benefit of the nonparametric method is that our categorical variables are allowed to interact with the continuous regressors in an unknown fashion. Hence, these categorical variables allow for more than simple intercept shifts.

4 Nonparametric Methods for Growth Empirics

In regression we are typically concerned with predicting the left-hand-side variable given specific values of one or more right-hand-side variables. For a particular observation, this is the conditional expectation $E\left(y_i|x_i = x\right)$. The general regression model, with an additive mean zero random error, is written as

$$y_i = E\left(y_i|x_i\right) + u_i, \quad i = 1, 2, \ldots, n.$$

We often omit this step and assume that $E(y_i|x_i)$ is linear in x, i.e. $E(y_i|x_i) = \alpha + \beta x_i$. If this functional form is true and the other Gauss-Markov assumptions hold, then the estimators of α and β are the best linear unbiased estimators and we can proceed with inference and policy suggestions. However, if the true model is nonlinear and we ignore this, estimation may not only lead to inconsistent estimates, but it can also mask important heterogeneity in the partial effects. For example, suppose the true model is quadratic in x, but we fit a linear model. In a linear model, the estimated partial effect $\partial y/\partial x = \beta$, is constant for all x. Thus, not only will the linear model's result be inconsistent, but it is also ignorant of the fact that the true partial effect varies with x. Even worse, the partial effect could take both positive and negative values. Implementing a policy based on results from the linear model when the true technology is quadratic could lead to unintended consequences for a particular group (say, Sub-Saharan Africa), for example, a detrimental instead of positive impact for a group within the population.

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, ε_i is the additive error which has expectation zero, w_i is a vector composed of the 'Solow' variables, initial income, physical capital savings rate, human capital savings rate, and the joint depreciation term on both types of

capital,[6] while z_i is a vector of a given 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 (1) becomes increasingly difficult as more growth theories are constructed. 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 (1).

Given that we generally do not know the true data generating process, we have a few options: First, we can simply hope that the true model is linear. Given that this is only one possibility out of an infinite number of possibilities, this may be a bit naive. Second, we can fit higher order polynomials as well as use interaction terms. This is a promising approach, but given the number of possibilities, it is difficult to model all of these without quickly running out of degrees of freedom. Finally, third, we can let the data tell the form of the technology. This is the approach we take.

Now, consider a general growth specification taking the unknown form

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

where x_i is the union of w_i and z_i. $m(\cdot)$ is the unknown smooth growth process (conditional mean of g given x). It does not assume that the variables enter in linearly or that they are separable from one another. 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 $q_c \times 1$ vector of continuous regressors (for example, initial income, capital savings rate, % ice free coast), x_i^u is a $q_u \times 1$ vector of regressors that assume unordered discrete values (geographic regions, OECD membership), and x_i^o is a $q_o \times 1$ vector of regressors that assume ordered discrete values (time). An additive, mean zero error is captured through u_i.

4.1 Nonparametric regression

In this section we consider two nonparametric regression methods: Local-Constant Least-Squares (LCLS) and Local-Linear Least-Squares (LLLS). Neither estimator requires functional form assumptions for the conditional mean nor do they require assumptions for the distribution of the error term. The additional benefit of LCLS that we exploit here is its ability to detect irrelevant regressors when automated bandwidth selection is used. The additional benefit of LLLS that we exploit here is its ability to detect linearity of regressors when automated bandwidth selection

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

is used. In what follows we explain each estimation method as well as how to determine when variables are relevant or enter linearly.

The basic idea behind LCLS is to calculate a weighted average of the left-hand-side variable. This estimate of the conditional mean is also known as a local average. It is the average of g local to a point x. We estimate the conditional mean function by locally averaging those values of the left-hand-side variable which are "close" in terms of the values taken on by the regressors. The amount of local information used to construct the average is controlled by the bandwidth (smoothing parameter). The unknown function (conditional mean) is estimated by connecting the (locally averaged) point estimates over a range of x.

Our LCLS estimate of the conditional mean in (2) at a specific point x is given by

$$\hat{m}(x) = \left(\mathbf{i}' K(x)\, \mathbf{i}\right)^{-1} \mathbf{i}' K(x)\, g, \tag{3}$$

where $g \equiv (g_1, g_2, \ldots, g_n)'$, $\mathbf{i}$ is a $n \times 1$ vector of ones and $K(x)$ is a diagonal n matrix of kernel weighting functions for mixed continuous and categorical data with bandwidth parameter vector h (Li and Racine, 2007).

The bandwidths, by affecting the degree of smoothing, are not just a means to an end; they provide some indication of how the left-hand-side variable is affected by the regressors. Hall, Li and Racine (2007) show that with LCLS, when the bandwidth on any regressor reaches its upper bound, the regressor is essentially smoothed out. Specifically, when the bandwidth reaches its upper bound, the kernel function becomes a constant. It is obvious in (3) that if the kernel function for a particular regressor is a constant, it can be pulled out of each term and they will cancel one another out. In other words, it is as if the variable never entered the model in the first place.

Our second nonparametric estimation procedure employed is LLLS. In short, LLLS performs weighted least-squares regressions around a point x with weights determined by a kernel function and bandwidth vector. Again, more weight is given to observations in the neighborhood of x. This is performed over the range of x and then the unknown function is estimated by connecting the point estimates. An added benefit is that if indeed the true functional form is linear, the LLLS estimator nests the OLS estimator when the bandwidth is very large.

Specifically, taking a first-order Taylor expansion[7] of (2) around x, yields

$$g_i \approx m(x) + (x_i^c - x^c)\beta(x^c) + \varepsilon_i, \tag{4}$$

[7]The Taylor expansion is only taken for the continuous variables.

where $\beta(x^c)$ is defined as the partial derivative of $m(x)$ with respect to x^c. The LLLS estimator of $\delta(x) \equiv \binom{m(x)}{\beta(x^c)}$ is given by

$$\widehat{\delta}(x) = \left(X'K\left(x\right)X\right)^{-1} X'K\left(x\right)g, \tag{5}$$

where X is a $n \times (q_c + 1)$ matrix with ith row being $(1, (x_i^c - x^c))$ and $K\left(x\right)$ is the same as in (3). Note that here we obtain a fitted value and derivative estimate (for each regressor) for each x^c. This allows us to observe (potential) heterogeneity in the partial effects. 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 the conditional mean when the 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 or even semiparametric procedures (Li and Racine, 2007).

Hall, Li and Racine (2007) also show what happens when the bandwidth reaches its upper bound in LLLS estimation. For ordered and unordered regressors, a bandwidth equal to the upper bound again smooths the regressor out. However, for continuous regressors, when the bandwidth reaches its upper bound, the variable enters in linearly. From (5) we can see that when the bandwidth approaches its upper limit and the kernel function is constant, it cancels out; then we are left with the familiar OLS estimator. Hence, automatic bandwidth selection criteria can show whether or not a continuous variable enters in linearly.

4.2 Cross-validatory bandwidth selection

Estimation of the bandwidths (h) 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) 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. They also show that the procedure has the ability to detect whether continuous variables enter in linearly in

the LLLS case.

For continuous regressors, in the LCLS case, a bandwidth equal to the upper bound implies that the variable is irrelevant. In the LLLS case, a bandwidth equal to the upper bound determines that the variable enters in linearly. The upper bound for the bandwidth on a continuous regressor in either case is infinity. This is impossible to observe in practice. However, when using a Gaussian kernel function, any bandwidth in excess of two standard deviations of the regressor gives essentially equal weight to all observations. In other words, in the local-constant setting, the local average with respect to that variable is actually a global average of the left-hand-side variable and hence the regressor (essentially) has no impact on the conditional mean. In the local-linear setting, all observations are given equal weight and hence the estimate is (essentially) linear. Thus, we follow the suggestion of Hall, Li and Racine (2007) and use two standard deviations of the regressor as the bound for relevance/linearity. Thus, if any bandwidth on a continuous regressor exceeds two standard deviations of its associated variable, we conclude that it enters in an irrelevant fashion (in the local-constant setting) or linearly (in the local-linear setting).

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 $x_i^o \neq x^o$ or $x_i^u \neq x^u$. 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 $x_i^o = x^o$ and $x_i^o \neq x^o$. In this case, the variable is completely smoothed out (and thus does not impact the estimation results). For both unordered and ordered discrete variables, the upper bound is unity. See Hall, Li and Racine (2007) for further details.

5 Empirical Results

Our first goal is to examine the Solow growth variables and use these results as a baseline when additional theories are investigated. Specifically, we will use nonparametric methods to determine the relevance of each regressor and whether or not it enters the model linearly. From there we will examine which of the individual growth theories are nonlinear by testing our nonparametric model versus both linear and nonlinear parametric specifications. We will then briefly discuss the

findings of each theory. Afterwords, we will examine in detail results stemming from the geography growth theory. These results will showcase how nonparametric methods can be used to deepen our understanding of growth theory. Finally, we provide Monte Carlo evidence that the nonparametric model selection methods work well for the sample sizes used in our empirical investigations (see Appendix).

Our bandwidths for the sample, across all theories, are presented in Tables 1 and 2. The bandwidths in the first table come from estimation by LCLS. Here we can observe which variables are relevant and which are irrelevant. The second table gives the bandwidths from the LLLS estimator. Here we can determine which variables enter the model linearly. Just below the separation in Table 2 we give the p-values (399 wild bootstrap replications) of a consistent test of model misspecification (Hsiao, Li and Racine, 2007) for both a linear (HLR1) and nonlinear parametric specification (HLR2), similar to those found in Maasoumi, Racine and Stengos (2007). Finally, we present the model fit for each theory in the final row of the second table.

5.1 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 growth. We first note that human capital is smoothed out. Specifically, the asterisk in Table 1 for the LCLS bandwidth on human capital (19955251) signifies that it is larger than two times the standard deviation of human capital ($7.852 = 2 \times 3.9258$). We remove this variable when we estimate the Solow model via LLLS given that this method cannot automatically remove irrelevant continuous variables. All other variables are significantly small and thus are considered to be relevant in terms of the estimation of output growth. A point worth noting here is that the bandwidth on time is zero to four decimal places. What this implies is that each cross-section can (essentially) be treated separately. This result suggests that there are significant differences across time in this model.

Turning to Table 2, we see that there are nonlinearities occurring in both population growth and initial income. The nonlinearities in initial income are in accord with the findings of Durlauf, Kourtellos, and Minkin (2001). Aside from a handful of studies, most growth researchers ignore any type of nonlinear structure either between or across these variables. We also see that investment has a bandwidth that is more than twice the size of its standard deviation. The cross next to the bandwidth denotes this. Finally, we note that the time variable is now larger and does not

(essentially) run the cross-sections separately. This is more intuitive as we believe that past and future observations would be of help when estimating a unit in a particular year as countries evolve over time. The conflict between these two tables should be pointed out here. Whereas both estimators are consistent, in finite samples small differences between these estimators will exist. That being said, LLLS is generally shown to outperform LCLS in relatively small samples. Hence, we use this estimation method to conduct inference and assess model fit.

The second column of numbers in Tables 1 and 2 correspond to the region theory. We take the Solow model and add a regional indicator as in Temple (1998). The first thing to note here is that in Table 1, in addition to human capital, population growth is smoothed out. The region variable here may pick up population growth differences across regions as well as other regional differences. Turning to the results of Table 2 we see that all variables enter the model nonlinearly. One possible explanation for this difference between models may be omitted variable bias. The Solow model may be too simple to adequately describe the growth process. We note in passing that just the inclusion of regional effects improves the model's fit, bumping up the pseudo-R^2 from 0.46 to 0.52. This is similar to the results of Temple (1998) who found that there were significant regional impacts on output growth. Looking across the columns of either table we note that region is never smoothed away. However, it is also important to note that the bandwidth on region in the other models is generally significantly different from zero. Therefore, there likely exist important interactions between region and the continuous variables entering the models that are not captured with simple intercept shifts.

When looking at the Solow variables across theories it is interesting to note that human capital is smoothed out in most settings. This is not necessarily surprising as DKT show the posterior probability for human capital is 0.019. We mention in passing that both investment and initial income are each relevant across all theories. Similarly, DKT have posterior inclusion probabilities near one for each of these regressors. 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 depends upon the model. In the demography and macroeconomic theories, initial income enters in linearly. The linearity of investment appears to hold both in the Solow and macroeconomic theories. Thus we again confirm that in general both investment and initial income are relevant predictors of economic growth. The perceived linearity depends upon the model. It is interesting to note that human capital enters linearly in all models for which it is relevant. That being said,

further examination shows that this variable is not statistically significant in any of these theories.

5.2 Estimating alternative theories

While examining the impact of the Solow variables on economic growth is interesting and insightful, much of the recent focus on economic growth has focused on alternative explanations aside from factor accumulation and initial conditions. Theories such as geography, institutions and macroeconomic policy have permeated the literature in recent years and created quite a stir among academics.[8] To determine how each 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.

The third column of numbers in Tables 1 and 2 correspond to the demography theory. The first table shows that the fertility rate and reciprocal of life expectancy at age one are both relevant predictors of growth. The second table shows that both of them enter the model in a nonlinear fashion. The results here suggest that after region and time effects have been controlled for, increasing investment in health should lead to higher growth. We also note that the demography theory provides a considerable improvement in fit over the basic Solow model. Specifically, the R^2 measures jumps above 0.75.

The next theory, geography, has the largest goodness-of-fit measure (0.8582). This is perhaps surprising given that so many variables are smoothed out. We also note that the sample size is smaller than that of the previous theories. Here we see that in the first table, population growth, human capital and % ice free coast all have bandwidths greater than two times their standard deviations. It is important to note that DKT find a large posterior probability from the Koeppen-Geiger measure, but not for % ice free coast. It appears that the relevance of this variable is robust to nonlinearities. Therefore, in Table 2, we have three continuous regressors: investment, initial income and the Koeppen-Geiger measure. Of note is the result from this table that shows that each of the variables enter nonlinearly. Of the competing (extended) models, this is the only one to do so. This fact is also apparent by the p-values from the HLR tests which reject the parametric models. Whereas we stated that the relevance of the Koeppen-Geiger variable was robust to potential nonlinearities, the perceived 'value' of the geography model does not seem to

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

be. DKT give a low posterior probability for the theory, while, as mentioned earlier, our goodness-of-fit measure is highest for the geography theory.[9] For this and other reasons, we examine this theory further in the next sub-section.

According to the macroeconomic policy model, output growth is said to depend upon the level of openness of a country, net government consumption as well as the inflation rate. Each of these measures are proxies for the 'health' of the economy. Our local-constant bandwidths show that each of these variables are relevant. Further, each of the Solow models are relevant with this theory. Turning to the local-linear bandwidths, we see that each of the Solow variables is assumed to enter linearly. Further, the HLR test fails to reject the null of linearity. This is the only theory where we fail to reject the parametric model(s). It can be argued that a reason the model averaging papers point to macroeconomic policy is that it is correctly specified. Even though we fail to reject the null, we see that several bandwidths in Table 2 are relatively small. These nonlinearities and interactions still allow for significant heterogeneity. It appears that a nonlinear parametric model along the lines of that specified in HLR2 could lead to a consistent (efficient) parametric model. This is an avenue that deserves further attention.

As with the macroeconomic policy model, in the fractionalization theory, we see that each of the regressors are relevant. Further, we see that both additional variables (ethnic tension and language) enter nonlinearly. Here we see the opposite result for linearity as only the human capital measure enters in linearly according to its bandwidth. Also different from the macroeconomic policy column, we reject the null of linearity. Further, we see that the goodness of fit measure is quite high here. Nonlinearities could be an explanation for this type of improvement. For example, in most linear regressions, linguistic fractionalization has a constant partial effect. This implies that larger levels of fractionalization are always worse for an economy (assuming a negative coefficient). However, it can be argued that uniformity (South Korea) or high fractionalization (the United States) may be preferable to a near split in language (Belgium).

Finally, our setting for studying institutions uses eight proxy variables, of which five are deemed relevant and three of these enter nonlinearly. This theory also has the third highest goodness-of-fit measure. DKT note that when they consider fundamental growth theories separate from the proximate growth theories, that their posterior probability for institutions is 0.96.

[9] The reader should be careful when comparing the fit of our model versus DKT. In DKT geography is compared simultaneously with the other theories. Further, our sample size is much larger than theirs.

What is desirable at this stage is a model combining all of our competing growth theories to see which variables are robust predictors. However, in a nonparametric setting the addition of additional predictors decreases the ability of the methods to provide sound insights. Thus, while we feel confident analyzing individual growth theories for nonlinearities, a combined nonparametric regression would contain far fewer observations than one could reliably feel confident.[10]

In summary, we have presented evidence to suggest that each model has nonlinearities and heterogenous partial effects. Thus, we suggest that future research focusing exclusively on any of these individual theories 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 techniques give us. We have also shown that the macroeconomic policy variable may be well estimated with a relatively simple nonlinear parametric model.

5.3 Geography

In this sub-section, we demonstrate the benefits of our approach by focusing on a particular growth theory: geography. Our analysis so far has shown that this model fits the data well, it is nonlinear and it is relatively parsimonious (recall results from Tables 1 and 2). Further, investigating the geographic growth regression is interesting given that it is nearly immune to policy decisions (aside from starting a war to acquire more land).

In Table 3 we give a summary of the LLLS partial effect estimates for each of the continuous regressors included in the model (initial income, investment and Koeppen-Geiger). Here we note that there is substantial variation in the partial effects. This suggests that assuming homogenous effects across the sample is incorrect. For the initial income variable, the interquartile range is approximately 0.0129, that same value is 0.0102 for investment. In comparison, the interquartile range for the Koeppen-Geiger measure is much larger (0.1388) thus providing evidence for more heterogeneity. Regarding significance, we can see that the partial effect of initial income is significant at the first quartile and median, but insignificant at the mean and upper quartile. Investment is significant at each point whereas we get predominantly insignificant results for each number associated with Koeppen-Geiger in Table 3. In other words, although we found the variable to be relevant in the prediction of growth, we only find the upper quartile of the partial effects to be

[10] In the Appendix we use Monte Carlo simulation methods to guide us on how reliable our methods are in practice.

significantly different from zero. The (uncommon) positive and significant result will be discussed further below.

Although informative, these descriptive statistics can only tell us what happens at particular points. In Figure 1 we examine the kernel density estimates of the partial effects for each continuous regressor. Here we can see the entire spread of the estimates. One interesting point is that we find a significant percentage of the partial effects of initial income to be positive. As expected, we find that most of the mass for the partial effects of investment are significantly to the right of zero. In other words, additional investment is shown to systematically increase output growth. That being said, we again note that this effect is heterogenous across the sample. Finally, although we found it to be generally insignificant in Table 3, we see that there is substantial variation in the partial effect of Koeppen-Geiger with some mass to both the left and right of zero.

We now return to the positive partial effects for the Koeppen-Geiger variable which warrant further explanation. The conventional wisdom is that more tropical climates have lower levels of output growth. DKT find a negative and insignificant value. However, here we see that a large percentage of the coefficients are positive. Further, a certain percentage of the observations are positive and significant. While we fully believe that a larger percentage of a country's land area classified as tropical in a place that is already tropical may hurt output growth, there is some reason to believe that relatively cold climates could be helped by warmer temperatures. To examine this we broke the estimated partial effects up into countries with above and below median levels of the Koeppen-Geiger measure (0.20). In Figure 2, we plot the kernel density estimates of partial effects on the Koeppen-Geiger measure for each group in one panel. For those countries with above median Koeppen-Geiger measures, the density is centered slightly to the left of zero. We see both positive and negative partial effects and the results are generally insignificant. At the same time, for those countries with below median Koeppen-Geigen values, we see that the density is shifted to the right.[11] Here we find generally insignificant estimates as well, but some evidence that colder countries may be better off with warmer temperatures.

The careful reader will note that we state that "generally" we find insignificant results for the Koeppen-Geiger measure. However, there were some interesting significant findings. For example, we found that for many years, Argentina, Ireland, Poland, and the United Kingdom all had positive and significant marginal effects on the Koeppen-Geiger measure. Note that each of these countries

[11] The Li (1996) test rejects equality of these densities at the 1% level.

have Koeppen-Geiger values below the median. At the same time we found that Bolivia, Mexico and Peru would benefit from lower percentages of tropical land area. In other words, these three countries had primarily negative and significant partial effects on the Koeppen-Geiger measure.

To get a better understanding of the impact of the results as well as to further understand the nature of the heterogeneity in the partial effects, we further dissect the partial effects of initial income. Whereas we split the partial effects solely on the regressor in Figure 2, here we split the partial effects based on several criteria to observe the behavior of the partial effects of initial income. Specifically, we look at differences in the estimated partial effect densities for initial income across splits along the median for the three continuous variables. We see that these results are again indicative of parameter heterogeneity. The Li test rejects equality of the estimated densities for every split in the table (all p-values are zero to four decimal points) suggesting potential interactions between initial income and the other variables used in the model (see Table 4 and Figure 3). We see that higher initial income, higher investment and lower values for the Koeppen-Geiger measure are associated with more negative partial effects of initial income. These results are in line with expectations that countries with these attributes converge faster. These results here are in line with conventional thought. The combination of these results show how nonparametric estimation is fruitful for studying economic growth. When models are nonlinear and/or contain parameter heterogeneity, a simple linear model cannot tease out these insights.

In summary, this sub-section focused on a specific theory: geography. We showed that this model fits the data well and that there is strong evidence of nonlinearities. We found substantial heterogeneity in the partial effects as well as ranges for which some effects were significant while others were insignificant for a particular regressor. Finally, we split the partial effects for both the initial income variable and Koppen-Geiger measure arbitrarily to try to attempt what explains heterogeneity in their respective partial effects. These splits uncovered new insights regarding the nonlinear effects of tropical land on growth therefore challenging the common view that tropics are *globally* detrimental to growth.

6 Conclusion

This paper uses recently-developed methods for nonparametric regression to investigate potential non-linearities in the growth process at the same time as selecting relevant variables. An innovation,

compared to related papers, is that the information obtained from automated bandwidth selection can be used to select the variables that have some explanatory power. This allows the consideration of non-linearities and variable selection simultaneously, which is highly attractive. The more common approach of considering them individually could well lead to errors. For example, whether or not a variable appears in an estimated model may be contingent on the functional form that applies elsewhere in the model; a variable might be wrongly dropped, or wrongly included, because of omitted non-linearities.

Our main findings are twofold. First, generally, our perceptions on economic growth are necessarily linked to the types of models used to link growth determinants and output. This suggests a more careful consideration of individual growth models and theories going forward since a linear parametric model may mislead one in both the direction of believing the theory is true when model misspecification is left unaccounted or towards rejection of a theory given that the nonlinearity of the theory is missed by a linear model. We argue that empirical studies of economic growth should attempt to determine the robustness of variables to nonlinearities in their model as well and the consequences of said nonlinearities.

Second, in a specific context, our focus on geography as a growth theory revealed that this theory was heavily nonlinear and the simple parametric models considered missed substantial heterogeneity in the partial effects. This heterogeneity in our geography determinant was important because it suggested that a generic effect did not exist. Moreover, several important interactions between the common Solow variables and our geography determinant were uncovered. Interactions between initial economic conditions and geographic conditions had differing effects on growth, again indicating that simple linear growth models are incapable of revealing the intricate link between factors of accumulation and geographic conditions.

References

[1] Acemoglu, D., S. Johnson, and J. A. Robinson, 2001. "The Colonial Origins of Comparative Development," *American Economic Review*, 91: 1369-1401.

[2] Alesina, A., A. Devleeschauwer, W. Easterly, S., Kurlat, and R. Wacziarg, (2003). "Fractionalization," *Journal of Economic Growth*, 8(2): 155-194.

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

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

[5] Casselli, F., 2005. "Accounting for Cross-Country Income Differences," in *Handbook of Economic Growth*, P. Aghion and S.N. Durlauf (Eds.), Amsterdam, North Holland, Volume 1A, 678-742.

[6] Ciccone, A. and M. Jarocinski, forthcoming. "Determinants of Economic Growth: Will Data Tell?," *American Economic Journal - Macroeconomics*.

[7] Djankov, S., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer, 2002. "The Regulation of Entry," *Quarterly Journal of Economics*, 117(1): 1-37.

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

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

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

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

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

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

[14] Gottardo, R. and A. E. Raftery, 2007. "Bayesian Variable and Ttransformation Selection: A Unifed Approach," Statistics Department, University of Washington, Technical Report 570.

[15] 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.

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

[17] Hoeting, J. A., A. E. Raftery and D. Madigan, 2002. "Bayesian Variable and Transformation Selection in Linear Regression," *Journal of Computational and Graphical Statistics*, 11: 485-507.

[18] 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.

[19] Johnson, S., W. Larson, C. Papageorgiou, and A. Subramanian, 2009. "Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates," NBER Working Paper no. 15455.

[20] 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.

[21] Kaufmann D., A. Kraay, and M. Mastruzzi, 2005. "Governance Matters IV: Governance Indicators for 1996-2004," World Bank Policy Research #3630.

[22] 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.

[23] Kourtellos, A., C. M. Tan and X. Zhang, 2007. "Is the Relationship Between Aid and Economic Growth Nonlinear?," *Journal of Macroeconomics*, 29(3): 515-540.

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

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

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

[27] 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.

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

[29] 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.

[30] Li, Q., 1996. "Nonparametric Testing of closeness between two unknown distribution functions," *Econometric Reviews*, 15: 261-274.

[31] Li, Q., and J. S. Racine, 2007. *Nonparametric Econometrics: Theory and Practice*, Princeton University Press: Princeton, New Jersey.

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

[33] 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.

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

[35] Minier, J., 2007a. "Institutions and Parameter Heterogeneity," *Journal of Macroeconomics*, 29(3): 595-611.

[36] Minier, J., 2007b. "Nonlinearities and Robustness in Growth Regressions," *American Economic Review, Papers and Proceedings*, 97(2): 388-392.

[37] Rodrik, D., A. Subramanian, and F. Trebbi, 2004. "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development," *Journal of Economic Growth*, 9: 131-165.

[38] Sachs, J. D., 2003. "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income," NBER Working Paper no. 9490.

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

[40] 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.

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

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

Table 1: Bandwidths using Local Constant Regression and Penn World Table 6.1. A bandwidth with a * next to them indicate that this variable is smoothed out of the regression.

Variable	Solow	Region	Demo	Geo	Macro	Frac	Inst
Population Growth	0.1021	470531*	0.2797	386408*	0.1328	0.0838	0.1023
Investment	0.5076	0.4509	0.5306	0.5742	0.3898	0.3978	0.2998
Human Capital	19955251*	7230883*	4894096*	1653682*	2.3827	5.1338	20633303*
Initial Income	0.9395	1.0423	1.1007	1.1413	0.7347	0.6233	0.4590
Time	0.0000	0.7459	0.7377	0.7114	0.8012	0.7311	0.4264
Region	.	0.0165	0.0391	0.0220	0.2538	0.1121	0.2000
Fertility	.	.	0.9638	.	.	.	.
Life Expectancy	.	.	0.0038	.	.	.	.
Koeppen-Geiger	.	.	.	0.7057	.	.	.
% Ice Free Coast	.	.	.	3016761*	.	.	.
Openness	.	.	.	.	0.3979	.	.
Net Govt. Cons	.	.	.	.	0.0539	.	.
Inflation	.	.	.	.	2.6982	.	.
Language	.	.	.	.	.	0.1798	.
Ethnic Tension	.	.	.	.	.	0.1585	.
Exec. Constraints	.	.	.	.	.	.	0.2555
Exprop. Risk	.	.	.	.	.	.	461711*
KKZ96	.	.	.	.	.	.	0.3902
Eviction	.	.	.	.	.	.	0.1162
Civil Liberties	.	.	.	.	.	.	0.3641
Bur. Quality	.	.	.	.	.	.	0.1636
Political Rights	.	.	.	.	.	.	814386*
Rule of Law	.	.	.	.	.	.	3010408*
# of Countries	98	98	96	92	94	85	60
# of Observations	731	731	715	691	532	562	409

Table 2: Bandwidths using Local Linear Regression and Penn World Table 6.1. A bandwidth with a $^+$ next to them indicate that this variable enters the regression in a linear fashion.

Variable	Solow	Region	Demo	Geo	Macro	Frac	Inst
Population Growth	0.0612	.	0.1055	.	710656.8^+	0.3226	424078.4^+
Investment	2095.654^+	0.5575	0.5223	0.3075	983401.6^+	0.4769	1.0568
Human Capital	.	.	.	.	7881216^+	13564807^+	4090271^+
Initial Income	0.8777	5.3597	117891.5^+	0.9865	878412^+	1.5261	1.0556
Time	0.8007	0.8174	0.8395	0.6454	0.9553	0.7701	0.8634
Region	.	0.0169	0.2243	0.1047	0.7670	0.1266	0.2196
Fertility	.	.	0.6697	.	.	.	.
Life Expectancy	.	.	0.0063	.	.	.	.
Koeppen-Geiger	.	.	.	0.0741	.	.	.
% Ice Free Coast	.	.	.	.	.	.	.
Openness	.	.	.	.	0.7670	.	.
Net Govt. Cons	.	.	.	.	0.9781	.	.
Inflation	.	.	.	.	5.8763^+	.	.
Language	.	.	.	.	.	0.6181	.
Ethnic Tension	.	.	.	.	.	0.2255	.
Exec. Constraints	.	.	.	.	.	.	0.5524
Exprop. Risk	.	.	.	.	.	.	.
KKZ96	.	.	.	.	.	.	295035.9^+
Eviction	.	.	.	.	.	.	89357.11^+
Civil Liberties	.	.	.	.	.	.	602819.7^+
Bur. Quality	.	.	.	.	.	.	0.2225
Political Rights	.	.	.	.	.	.	.
Rule of Law	.	.	.	.	.	.	.
HLR1 Test	0.003	0.008	0.005	0.010	0.120	0.000	0.013
HLR2 Test	0.000	0.000	0.002	0.025	0.226	0.000	0.003
# of Countries	98	98	96	92	94	85	60
# of Observations	731	731	715	691	532	562	409
R^2	0.4634	0.5200	0.7531	0.8582	0.6224	0.8016	0.7660

Table 3: Partial effects for all continuous regressors for the geography model.

Variable	Mean	$Q1$	$Q2$	$Q3$
Initial Income	-0.0102	-0.0155	-0.0116	-0.0026
	0.0041	0.0056	0.0025	0.0060
Investment	0.0289	0.0252	0.0327	0.0354
	0.0081	0.0073	0.0054	0.0073
Koeppen-Geiger	0.0601	-0.0198	0.0588	0.1190
	0.0507	0.0258	0.0426	0.0590

Table 4: Partial effect of initial income across various groups for the DKT geography model.

Variable	Mean	$Q1$	$Q2$	$Q3$	Li Test
Above Median Initial Income	-0.0143	-0.0160	-0.0147	-0.0122	0.0000
	0.0035	0.0064	0.0045	0.0028	
Below Median Initial Income	-0.0062	-0.0105	-0.0035	0.0010	
	0.0030	0.0045	0.0053	0.0079	
Above Median Investment	-0.0147	-0.0162	-0.0148	-0.0127	0.0000
	0.0026	0.0025	0.0024	0.0025	
Below Median Investment	-0.0057	-0.0101	-0.0032	0.0007	
	0.0025	0.0080	0.0025	0.0025	
Above Median Koppen-Geiger	-0.0089	-0.0175	-0.0066	-0.0002	0.0000
	0.0027	0.0025	0.0047	0.0064	
Below Median Koppen-Geiger	-0.0116	-0.0153	-0.0137	-0.0085	
	0.0025	0.01153	0.0039	0.0057	

Figure 1: Estimated marginal effects for the geography theory.

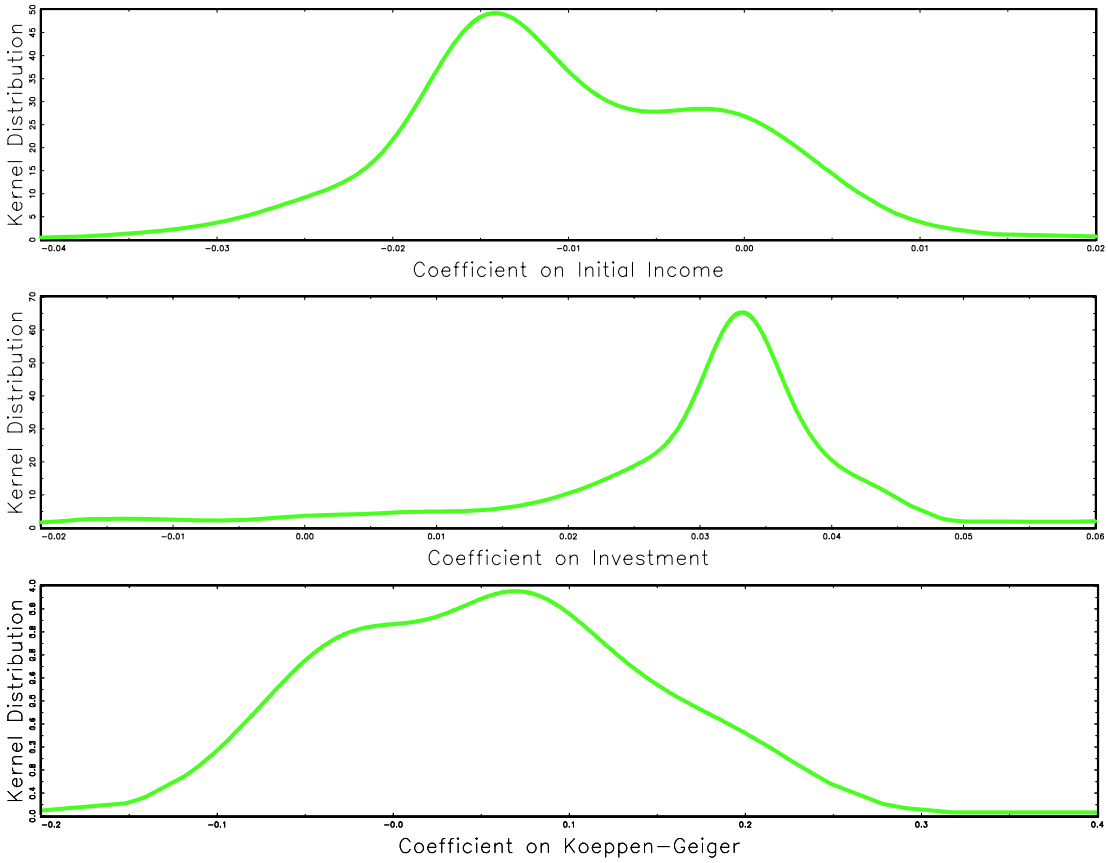


Figure 2: Comparison of estimated marginal effects on Koeppen-Geiger measure for values both above and below the Koeppen-Geiger measure.

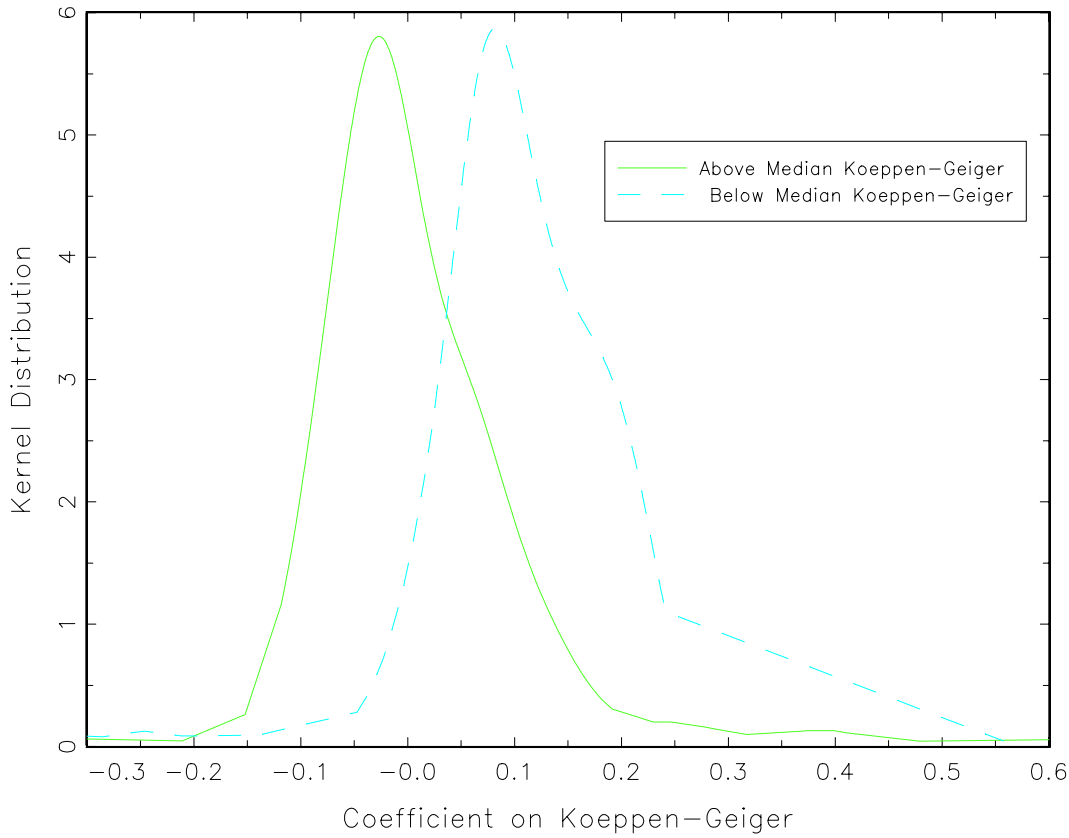
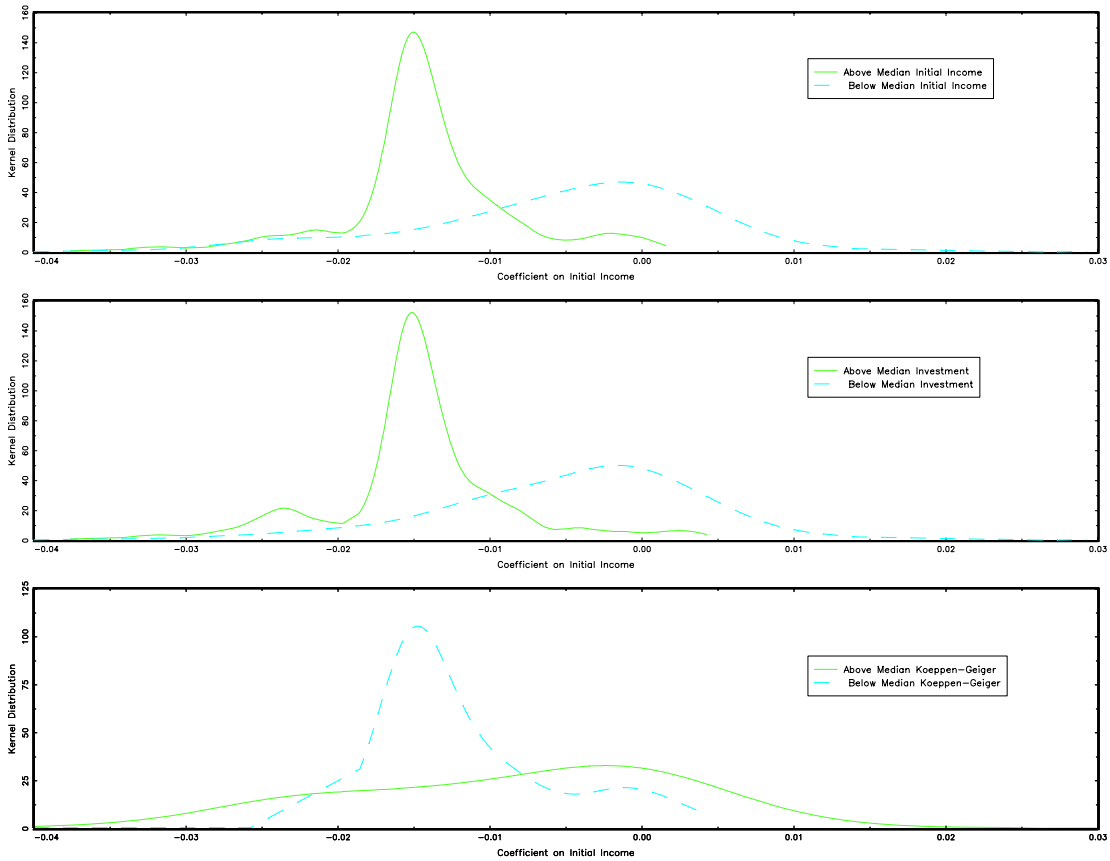


Figure 3: Comparison of estimated marginal effects by various splits for the geography theory



Appendix

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. 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 notice that due to lack of information on certain variables that for any given theory we have samples as small as 409 observations and as large as 731. Therefore we conduct our small sample (balanced panel) analysis using both $n_1 = 59$ and 95 cross-sectional units and $T = 7$ time periods. Our setup is similar to that in Hall, Li and Racine (2007), except that we include more relevant and irrelevant regressors as well as allow for observations to be observed over multiple time periods. Our goal is to generate a data set that is similar to the one used in the empirical section. 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.

For $i = 1, 2, \ldots, n_1$, with $n_1 = 59$ or 95 and $t = 1, 2, \ldots, 7$, we generate three types of random variables: unordered categorical (z_{it}), ordered categorical (w_{it}) and continuous (x_{it}). For each of the three unordered categorical variables $(z_{1i}, z_{2i}, z_{3i}) \in \{0, 1\}$, $Pr[z_{1it} = z_{1it-1}] = \rho_1$, $Pr[z_{2it} = z_{2it-1}] = \rho_2$ and $Pr[z_{3it} = z_{3it-1}] = \rho_3$, where $Pr[z_{1i0} = 1] = 0.62$, $Pr[z_{2i0} = 1] = 0.71$, $Pr[z_{3i0} = 1] = 0.82$. ρ_1, ρ_2, and ρ_3 are set equal to 0.50, 0.70 and 0.90, respectively. Obviously, $Pr[z_{jit} \neq z_{jit-1}] = 1 - \rho_j$ for $j = 1, 2, 3$. In other words, a higher value of ρ indicates a stronger persistence in the unordered categorical variable over time.

We allow each of the ordered categorical variables (w_{1it}, w_{2it}) to take integer values from 0 to 3. They are generated as $Pr[w_{1it} = w_{1it-1}] = \phi_1$ and $Pr[w_{2it} = w_{2it-1}] = \phi_2$. We set $\Pr[w_{1i0} = \ell] = 0.25\ \forall \ell$ and $\Pr[w_{2i0} = 0] = 0.40$ and $Pr[w_{2i0} = \ell] = 0.20$ for $\ell = 1$, 2, and 3. The persistence parameters ϕ_1 and ϕ_2 are set equal to 0.50 and 0.90, respectively. When $w_{sit} \neq w_{sit-1}$ for $s = 1$ or 2, w_{sit} takes one of the other values from 0 to 3 with equivalent probability, $(1 - \phi_s)/3$.

Finally, we consider five continuous variables $(x_{1it}, x_{2it}, x_{3it}, x_{4it}, x_{5it})$ which are generated as $x_{jit} = \varphi_j x_{jit-1} + \nu_{jit}$, where for $j = 1, 2, \ldots, 5$, φ_j is set equal to 0.80, 0.90, 1.00, 1.10, and 1.20, respectively. Note here that x_3 has a unit root. Assuming a zero error ($\nu_{jit} = 0$), values of φ less than one indicate that the regressor is decreasing with time (e.g., population growth) and values of φ in excess of one indicate that the regressor is increasing with time (e.g., human capital

accumulation). Further, x_{ji0} are generated as uniform from one to two and the ν_{jit} are generated as normal with mean zero and variance equal to 0.10. The initial values are drawn so that they exhibit a 0.50 degree of correlation.[12]

We generate y_{it} according to

$$y_{it} = z_{1it} + x_{1it} + x_{2it} + x_{1it} \cdot x_{2it} + \varepsilon_{it}, \tag{DGP 1}$$

or

$$y_{it} = z_{1it} + \sqrt{w_{1it}} \cdot x_{1it} + x_{2it} + x_{1it} \cdot x_{2it} + x_{3it}^2 + \varepsilon_{it}, \tag{DGP 2}$$

where $\varepsilon_{it} = \pi\varepsilon_{it-1} + u_{it}$, $\pi = 0.50$, u_{it} is drawn from a normal distribution with mean zero and variance equal to 0.10 and ε_{i0} is drawn from a t-distribution with five degrees of freedom. In the data set considered in our study, the role of outliers could be critical and to assume normality may under-estimate the practical importance of extreme observations.

In each model there are at least two relevant continuous variables as well as 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 probability.[13]

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 80% of its upper bound (i.e., bandwidths larger than 0.80) 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 then two standard deviations of the regressor we conclude that the continuous variable has been smoothed out of the exercise. For our 399 replications, we note the median, 10th and 90th percentiles of the cross-validated bandwidths.

We see from Tables 5 and 6 that the median results suggest that the method is correctly smoothing away irrelevant discrete and continuous variables. For instance, in DGP 1, only z_1, x_1

[12]To generate the initial values we take draws of size n_1 from a 10-dimensional multivariate normal distribution with zero means and variances equal to 1. The covariances are set so that the initial values display positive correlation of 0.5. The five discrete variables are constructed by taking the corresponding draw from the normal and using the quantile transformation based on ϕ or ρ to assign an integer value.

[13]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.

and x_2 are relevant. Table 5 shows that h_{z_1} is the only categorical bandwidth whose median value is well below its upper bound. At the same time, the median bandwidths for x_1 and x_2 in Table 6 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, it is obvious that the ability to smooth away irrelevant regressors is generally enhanced by additional data. Notice that the median bandwidths increase for all irrelevant variables and the 10th percentiles increase. 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.

These results do not suggest that irrelevant variables are always smoothed away, especially in small samples. Table 5 suggests that in some instances (for $n_1 = 59$) the discrete variables are not smoothed out of the model. For example, the results on $\hat{\lambda}_{z_2}$ suggest that while the upper bound is obtained in at least 10% of the simulations, there are also numerous simulations where z_2 was not smoothed away. The results are much better for the continuous variables, however, as in almost all the simulations the irrelevant continuous variables are smoothed away.

One point of concern is the behavior of the estimated bandwidths on the continuous regressor x_3. Recall that x_3 contains a unit root. In DGP 2 the variable is relevant and the results from Table 6 shows that the bandwidth selector correctly shows this. However, in DGP 1, the variable is irrelevant and the bandwidth at the median is roughly equal to two times its standard deviation. Further, we see some decrease in the 10th percentile and median bandwidth when the sample size is increased. Fortunately, in each case the 90th percentile is significantly large. However, we warn here that when the regressor possesses a unit root that the user should be careful interpreting the result. This finding also deserves further research.

Our second assessment involves the model's predictive performance where we generate data, independent from the original draw, from the same DGP of the same size, $n_2 = 413$ or 665. Predictive performance is assessed via $PMSE = 1/n_2 \sum_{j=1}^{n_2} (\hat{y}_j - y_j)^2$. We repeat this process 500 times for each simulation and for each DGP. 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 LCLS cross-validated results. The estimates for the first two models should lead to inconsistent estimates while the second two are consistent estimators. Table 7 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 DGP 1, when $n_1 = 59$, the relative performance is approximately 60% better than both the incorrectly specified linear model with every variable included and the incorrectly specified linear model with only the relevant

variables. Additionally, as the sample size increases, the relative performance of the nonparametric model relative to the correctly specified linear model improves from 28.8% to 24.5%. 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 with bandwidths selected via LSCV 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. The ability to smooth out irrelevant regressors with relatively small samples gives us more confidence in the results from the preceding section.

Table 5: Summary of cross-validated bandwidths for the discrete covariates NP LSCV estimator.

	Median, [10th Percentile, 90th Percentile] of $\hat{\lambda}$				
	$\hat{\lambda}_{z_1}$	$\hat{\lambda}_{z_2}$	$\hat{\lambda}_{z_3}$	$\hat{\lambda}_{w_1}$	$\hat{\lambda}_{w_2}$
$n_1 = 59$					
DGP 1	0.007	0.704	0.695	0.869	0.650
	[0.000,0.027]	[0.399,1.000]	[0.253,1.000]	[0.623,1.000]	[0.381,0.999]
DGP 2	0.050	0.894	0.894	0.280	0.861
	[0.026,0.074]	[0.563,1.000]	[0.491,1.000]	[0.220,0.404]	[0.641,1.000]
$n_1 = 95$					
DGP 1	0.002	0.894	0.790	0.886	0.709
	[0.000,0.021]	[0.400,1.000]	[0.259,1.000]	[0.662,1.000]	[0.410,0.957]
DGP 2	0.039	0.997	1.000	0.303	0.885
	[0.016,0.059]	[0.600,1.000]	[0.536,1.000]	[0.192,0.352]	[0.709,1.000]

Table 6: Summary of cross-validated bandwidths for the continuous covariates NP LSCV estimator. When an estimated bandwidth is very large it is replaced by $\approx \infty$ to denote that it is effectively equal to the asymptotic upper bound.

	Median, [10th Percentile, 90th Percentile] of $\hat{h}$				
	$\hat{h}_{x_1}$	$\hat{h}_{x_2}$	$\hat{h}_{x_3}$	$\hat{h}_{x_4}$	$\hat{h}_{x_5}$
$n_1 = 59$					
DGP 1	0.157	0.171	2.397	2.675	$\approx \infty$
	[0.115,0.189]	[0.128,0.228]	[0.746,$\approx \infty$]	[0.911,$\approx \infty$]	[1.886,$\approx \infty$]
DGP 2	0.245	0.300	0.186	$\approx \infty$	16.155
	[0.188,0.282]	[0.231,0.386]	[0.143,0.252]	[2.252,$\approx \infty$]	[3.261,$\approx \infty$]
$n_1 = 95$					
DGP 1	0.142	0.160	1.812	3.512	7.071
	[0.111,0.162]	[0.117,0.195]	[0.627,$\approx \infty$]	[1.188,$\approx \infty$]	[2.022,$\approx \infty$]
DGP 2	0.213	0.265	0.155	$\approx \infty$	$\approx \infty$
	[0.179,0.256]	[0.226,0.324]	[0.125,0.195]	[3.036,$\approx \infty$]	[5.509,$\approx \infty$]

Table 7: Out-of-sample PMSE performance for parametric and nonparametric models containing irrelevant regressors ($\rho = 0.5$).

	Median, [10th Percentile, 90th Percentile] of PMSE			
	NP-LSCV	PI-ALL	PI-ONLY	PC
$n_1 = 59$, $n_2 = 413$				
DGP 1	0.118	0.297	0.294	0.084
	[0.092,0.158]	[0.215,0.434]	[0.216,0.434]	[0.064,0.125]
DGP 2	0.414	2.977	3.010	0.086
	[0.337,0.529]	[2.098,4.510]	[2.119,4.527]	[0.065,0.125]
$n_1 = 95$, $n_2 = 665$				
DGP 1	0.094	0.245	0.240	0.071
	[0.074,0.122]	[0.183,0.343]	[0.183,0.343]	[0.056,0.095]
DGP 2	0.344	3.296	3.289	0.084
	[0.296,0.422]	[2.228,4.119]	[2.236,4.107]	[0.071,0.112]