Growth Empirics without Parameters*

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 Economics
University of Miami

February 10, 2011

Abstract

Recent research on growth empirics has 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 handle nonlinearities as well as select relevant variables. We show that inclusion of nonlinearities is necessary for determining the empirically relevant variables and uncovering key mechanisms of the growth process. We also show how nonparametric methods can sometimes point towards the correct parametric specification. Each of these points are demonstrated by considering specific growth theories.

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

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

*We thank the editor Antonio Ciccone, two anonymous referees, Mike Delgado and Thanasis Stengos for insightful comments as well as participants of seminars given at Louisiana State University, the University of Nevada, Las Vegas, the 2009 Meeting of the Midwest Econometric Group at Purdue University, the 2009 North American Summer Meeting of the Econometric Society at Boston University, as well as the 2010 World Congress of the Econometric Society at Shanghai Jiao Tong University. We 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.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, Low Income Countries Strategy Unit, Strategy, Policy and Review 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 Economics, University of Miami, Coral Gables, FL 33124-6520. Phone: 305-284-4397, Fax: 305-284-2985, E-mail: cparmeter@bus.miami.edu.

1 Introduction

The vast number of theories put forth that attempt to explain economic growth has led to an empirical conundrum known as "theory-openendedness" (Brock and Durlauf 2001). Theory-openendedness suggests that while numerous theories may indeed explain growth of an economy's output, no particular 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 data being used to empirically test growth models. Such techniques allow empiricists to focus on the variables which produce (statistically) robust insights regarding economic growth.

Most model selection and averaging studies assume a linear growth process so that functional form uncertainty can be avoided - see e.g., Fernandez, Ley and Steel (2001), Sala-i-Martin, Doppelhofer and Miller (2004) and Durlauf, Kourtellos and Tan (2008), DKT hereafter. However, an emerging theme in the empirical growth 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 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 nonlinearities in the growth process as well as select 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 display heterogeneous partial effects. This is important in three respects. First, it solidifies the growing consensus in the empirical growth literature that empirical growth models should be based on functional forms that go beyond traditional linear in variables 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 fail to reject the specification of policy theory when including quadratic and simple interaction terms versus a nonparametric alternative. Third, this paper outlines an approach for determining potential nonlinearities which may subsequently guide model selection. We then demonstrate how our approach can unravel important insights that parametric growth models may overlook by focusing on specific growth theories (geography, demography and policy).

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 from which our results stem. Section 5 presents the main results. Section 6 concludes. The Appendix includes a set of Monte Carlo simulations to show how well the nonparametric methods deployed here work for the sample sizes used in our empirical investigations.

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) and Leamer and Leonard (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 regressions. 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 produces posterior distributions (which reflect revised beliefs about the underlying statistical process in the face of model uncertainty) for each parameter. BMA has recently become widely accepted as a way of accounting for model uncertainty, notably in regression models for identifying the determinants of economic growth.[2]

Although the model averaging methodology has been making its mark as a constructive tool in growth empirics it is not without drawbacks, notably, sensitivity in BMA estimates when many potential explanatory variables are considered (which is the norm rather than the exception in this literature which often includes well over 50 variables). For example, in a recent contribution Ciccone and Jarocinski (2010) take a specific approach to this problem by examining how BMA estimates change when using slightly different sources of the GDP series (such as PWT version 6.2 vs. version 6.1). They find that when considering a large number of regressors, relatively small measurement error and data revisions in international GDP estimates result in substantial instability of estimates that are "... too large for agnostic growth empirics."

But other aspects related to the sensitivity documented in Ciccone and Jarocinski (2010) have also been flagged. For example, Kraay and Tawara (2010) use BMA to document partial correlations between disaggregated governance indicators and related outcome variables. These authors find instability across both outcomes and levels of disaggregation in the set of indicators identified by BMA as effective determinants of outcomes. A potential source of the instability in the results may

[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 become common practice; see Sala-i-Martin, Doppelhofer and Miller (2004), Kourtellos, Tan and Zhang (2007), Ley and Steel (2007), Masanjala and Papageorgiou (2008), DKT and Ciccone and Jarocinski (2010), just to name a few.

reflect multicollinearity problems in the individual models considered by BMA. As argued by Kraay and Tawara, a consequence of having nearly collinear regressors in finite samples is that parameter estimates are highly sensitive to small changes in model specification. Thus unlike Ciccone and Jarocinski's work showing sensitivity in BMA results due to measurement error and data revisions, Kraay and Tawara's research points to sensitivity arising from model specification; see Doppelhofer and Weeks (2009) and Ley and Steel (2007, 2009a) for formal treatments of jointness of regressors in BMA.

Another important aspect regarding BMA's sensitivity involves the choice of prior structure. Specifically, the implementation of BMA is subject to the challenge that it requires prior distributions over all parameters in all models and the prior probability of each model must also be specified. But what is the "correct" choice of prior for a given parameter or model? As Pesaran, Schleicher and Zaffaroni (2009) eloquently state "... [model averaging] is subject to its own form of uncertainty, namely the choice of the space of models to be considered and their respective weights. It is therefore important that applications of model averaging techniques are investigated for their robustness to such choices." Indeed recent work by Ley and Steel (2009b), Zeugner and Feldkircher (2009) and Eicher, Papageorgiou and Raftery (2011), venture in this direction. These papers compare several candidate default parameter priors that have been advocated in the literature and alternative model priors and show that results in typical BMA growth applications involving several dozens of regressors can be highly sensitive to the choice of priors.

Finally, endogeneity of regressors in growth estimation is a challenging problem which has started to occupy much of the recent work in the literature. The endogeneity challenge can increase dramatically when considering a very large number of models, as in model averaging exercises. An emerging line of research examines the combination of model averaging and Instrumental Variables (IV) models using panel data; pioneering work on this direction includes Tsangarides (2004), DKT, Eicher, Lenkoski and Raftery (2009) and Moral-Benito (2010a,b). Correcting for potential endogeneity concerns in BMA analysis results in findings that differ quite drastically from OLS-based BMA results reflecting once again instability in results obtained when averaging over a very large number of models.

Although BMA has its disadvantages, we feel that it is important to highlight the advantages it has over our approach. Even though BMA models in the literature typically use linear models, there is no logical necessity for this. It is possible to include nonlinear models as well (albeit they must be specified a priori). A correctly specified BMA exercise would be a preferable model selection tool to what we propose in this paper. The second main benefit of BMA is that for a given sample size it is able to handle many more variables than we are able to incorporate given the curse of dimensionality problem we face. In other words, we cannot include 50 potential regressors in a typical growth regression. We must arbitrarily choose sets of variables and then check to see what variables are valid within that subset of variables. To put it another way, BMA covers a much larger model space.

In sum, the promise of model averaging methods in shedding new light on important growth features depends crucially on further research seeking answers to concerns related to the sensitivity of results. This paper takes another approach to this problem by considering the effects of nonlinearities in choosing among different growth theories.

2.2 Incorporating nonlinearities in model uncertainty

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 influenced by the pioneering contribution of Durlauf and Johnson (1995). Liu and Stengos (1999) consider an additive partially 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). We are only aware of two papers which consider model uncertainty in a growth regression context while allowing for nonlinearities: Kalaitzidakis, Mamuneas and Stengos (2000) and Minier (2007a). The former paper uses EBA, as in Levine and Renelt (1992), but allows for nonlinearities by setting up the growth regression in a partially linear framework. The latter introduces nonlinearities and heterogenous partial effects, but still considers models which are linear in parameters.

All of these papers have uncovered the existence of significant nonlinearities across an array of variables within cross-country growth regressions. Although these studies are able to relax functional form assumptions and lessen the curse of dimensionality, their consistency still depends on restrictive assumptions. As an alternative, Maasoumi, Racine and Stengos (2007) 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 the focus of Maasoumi, Racine and

Stengos (2007), but exploit their methodology to determine which growth theories empirically 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 allow for 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 nonparametric methods 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 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]

An important aspect of our data is that it is as large as possible with respect to countries and time for the given theories we consider. Our largest dataset spans 1960-2000 (8 time periods) and includes 98 countries for a total of 731 observations, nearly triple that employed in DKT. We note that all of our datasets are unbalanced both with respect to countries and time. This construction is employed to maximize our sample size for each theory. While we certainly could use a balanced dataset, doing so would decrease our sample by almost 50 percent.

Next we briefly review several key features of the DKT dataset 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 policy. At least two variables for each theory are used in our statistical analysis. The depen-

[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 PWT. These results are also available upon request.

dent 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 literature choosing empirical proxies as follows: **(1) Solow growth variables**: is measured by 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) 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 Köeppen-Geiger classification system for climate zones (*Köeppen-Geiger*) and a geographic accessibility/isolation variable, the percentage of a country's land area within 100km of an ice-free coast based on Gallup, Sachs and Mellinger (1999) (*% Ice Free Coast*). **(5) Fractionalization**: is measured by linguistic fractionalization as constructed by Easterly and Levine (1997) and Alesina et al. (2003) (*Language*) and a measure of "the degree of tension within a country attributable to racial, nationality, or language divisions" from the International Country Risk Guide (*Ethnic Tension*). **(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

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

parametric modelling. That is, instead of using regional and time dummies, we introduce two discrete variables. The first variable contains the region to which a country is deemed to belong to and the second denotes the time period under investigation. In our data we follow the regional country classification of the World Bank while our categorical variable for time is constructed based on the time interval. An additional benefit of the nonparametric method is that our categorical variables are allowed to interact with the continuous regressors allowing 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 $E(y_i|x_i = x)$. The general regression model, with an additive mean zero random error, is written as

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

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 OLS 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 relationship 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 (log-linear) form:

$$g_i = \beta_0 + \beta' w_i + \gamma' z_i + \varepsilon_i, \tag{1}$$

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,[6] initial income, physical capital saving rate, human capital saving rate and the joint depreciation term on both types of capital,[7] while z_i is a vector of a given length that contains variables associated with several alternative growth theories. Uncertainty over the exact choice of 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 w_i and z_i vectors, but to date most 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 naïve. 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, we can let the data tell the form of the relationship. 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) 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.

[6]When we discuss the Solow variables, initial income, population growth, investment and human capital, we assume it is taken to mean the logarithm of these respective variables. We generally omit the word logarithm hereafter for simplicity.

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

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 us to assume a specific distribution for the error term.[8] 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 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 (locally) 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. 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

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

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 (Gaussian kernel) and discrete (Li-Racine kernel) data with bandwidth vector $h = (h_1, h_2, \ldots, h_{q_c})$ for the continuous variables and bandwidth vector $\lambda = (\lambda^u, \lambda^o) = (\lambda_1^u, \ldots, \lambda_{q_u}^u, \lambda_1^o, \ldots, \lambda_{q_o}^o)$ for the discrete regressors (Li and Racine, 2007). Note here that instead of using a single bandwidth for all regressors we instead have a bandwidth for *each* covariate, both continuous and discrete.

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

[8]The lack of a specific functional form for the conditional mean and the lack of a specific distribution for the error term does not imply that our estimation and inference do not require assumptions. For example, theory requires that the conditional mean is twice continuously differentiable. For the case of the error term, the paper which proves the properties of detecting irrelevance and linearity with bandwidth selection (Hall, Li and Racine 2007) assumes in its proof that the error term has finite moments of any order. In the Appendix we show that even with errors which generate this moment assumption the methods described here still perform admirably. The exact assumptions necessary for our estimation and inference can be found in Li and Racine (2007).

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[9] of (2) around x, yields

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

where $\beta(x^c)$ is defined as the partial derivative of $m(x)$ with respect to x^c.[10] 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. Estimates of categorical variables are obtained separately. Specifically, they are estimated as the counterfactual difference in the conditional mean when switching from one value of the categorical regressor to another. 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 linearly. From (5) we can see that when the bandwidth

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

[10] Our later discussion often refers to our gradient estimates as if they are causal effects. We want to emphasize here that we do not take into account the potential endogeneity of any regressor. This should be taken into consideration when interpreting the results of our study.

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.

It is important to note that while we consider two estimation procedures, most of the results of this study can be performed using solely LCLS. LCLS can be used to determine relevance as well as estimate the partial effects of the model. In other words, we can handle nonlinearities and variable selection simultaneously, but this approach would not be ideal for at least two reasons. First, one property that LCLS does not have is detecting linearity. This is of particular interest in this application. Second, it is well known (e.g., see Fan and Gijbels 1996) that LCLS is less desirable than LLLS in terms of estimation. Hence, we will conduct our study in two steps, not out of necessity, but for improved performance. The first step will be to use LCLS to determine relevance. In the second step, we will take the variables that survive the first step (for each theory considered) to determine linearity as well as to examine partial effects and conduct inference.

4.2 Cross-validatory bandwidth selection

Estimation of the bandwidths (h) is typically the most critical 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.[11] 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 regressor enters the model (essentially) in a linear fashion. 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, either for LCLS or LLLS, 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 three separate theories (geography, policy, demography). These results will demonstrate how nonparametric methods can be used to deepen our understanding of growth theory. Separately, we also provide Monte Carlo evidence that the nonparametric model selection methods work well for the sample sizes used in

[11] The use of two standard deviations is based on the kernel, not the distribution of the data. For instance, even if the underlying variable was excessively skewed, the use of any standard, second-order kernel would render this variable irrelevant from the viewpoint of smoothing.

our empirical investigations (see the 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)[12] 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). We also present the model fit (square correlation coefficient between the fitted and actual growth rates) 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 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 have estimated bandwidths which are smaller than our benchmark threshold 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.[13] 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 bandwidth on time is now larger and does

[12] A wild bootstrap is necessary when the errors are heteroskedastic. For more discussion see Cameron and Trivedi (2005).

[13] Again note that we are checking whether or not the variables enter linearly in logarithmic form.

not (essentially) run the cross-sections separately. While we see this as more intuitive, it does not imply that the coefficients are constant over time. Instead, it suggests that past and future observations are being used 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 regional Solow theory (Region). We take the Solow model and add a regional indicator (Temple 1998a). 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 result of Temple (1998a) 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 additionaly models we consider is generally 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 of inclusion for human capital to be 0.019. Note 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 policy theories, initial income enters in linearly. The linearity of investment appears to hold both in the Solow and policy 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 generally statistically insignificant in each 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 policy have permeated the literature in recent years and generated academic debate.[14] 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. Table 1 shows that the fertility rate and reciprocal of life expectancy at age one are both relevant predictors of growth. Tables 2 shows that both these variables enter the model in a nonlinear fashion. The results here suggest that (for a portion of the sample) 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 measure jumps above 0.75. We examine this theory in more detail later.

The geography theory has the largest goodness-of-fit measure (0.86). 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 inclusion probability for the Köeppen-Geiger measure, but not for % ice free coast. It appears that the relevance of the Köeppen-Geiger measure and the irrelevance of % ice free coast are robust to nonlinearities. Therefore, in Table 2, we have three continuous regressors: investment, initial income and the Köeppen-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

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

the Köeppen-Geiger variable was robust to potential nonlinearities, the perceived 'importance' of the geography model, in terms of posterior probability in the BMA setting, does not seem to be. Specifically, DKT give a low posterior probability for the theory, while, as mentioned earlier, our goodness-of-fit measure is highest for the geography theory.[15] For this and other reasons, we examine this theory further in the next sub-section.

According to the 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. Our local-constant bandwidths show that each of these variables are relevant. Further, each of the Solow variables are relevant within this theory. Turning to the local-linear bandwidths, we see that each of the Solow variables are 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 variables that fall under the policy nomenclature is that this model 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 heterogeneity. It appears that a nonlinear parametric model along the lines of that specified in HLR2 could lead to a consistent (efficient) parametric model. We also examine this theory in more detail below.

As with the policy model, in the fractionalization theory it is shown that each of the regressors are relevant. Further, both additional variables (ethnic tension and language) enter nonlinearly. Here we see the opposite result for linearity as only the human capital measure enters linearly according to its bandwidth. Also different from the policy column, we reject the null that the model is linear. The goodness of fit measure is quite large here and nonlinearities could be an explanation for this type of improvement. For example, in 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-even split in language (Belgium). To be fair this result is not new, as an existing strand of the literature has often considered nonlinearities in the effect of ethnic diversity on development outcomes and/or used measures of polarization to address the idea that intermediate levels of diversity may be especially harmful – see, e.g. Easterly and Levine (1997), Collier and Hoeffler (1998), Temple (1998b), Block

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

(2001) and Montalvo and Reynal-Querol (2005), among others.

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, their posterior inclusion probability for institutions is 0.96.

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 inclusion 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 be based on fewer observations than is required for reliable results.[16] Again, even though we have argued for use of these methods, they do not dominate other approaches (for example, BMA) in all dimensions.

In summary, we have presented evidence to suggest that each individual growth theory empirical model has nonlinearities and heterogeneous partial effects. Thus, we suggest that future research focusing exclusively on any of these individual theories consider nonlinear impacts of the proxy variables. Next we demonstrate the benefits of our approach by focusing further on three theories with relatively large data coverage and with a high R^2.

5.2.1 Geography

Our analysis so far has shown that the geography theory fits the data well, is nonlinear and relatively parsimonious (recall results from Tables 1 and 2). 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 Köeppen-Geiger). Specifically, we take the vector of partial effects for each continuous regressor and then take the mean, median ($Q2$), first ($Q1$) and second ($Q3$) quartiles of this vector. Below each estimate we give the corresponding (wild) bootstrapped standard error for that particular partial effect.

We note that there is substantial variation in the marginal effects. This suggests that assuming homogeneous effects across the sample is incorrect. For the initial income variable, the interquartile range is approximately 0.0129. In other words, as the gradients measure the percentage change in the growth rate with respect to a particular regressor, the absolute difference between the first

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

and third quartile is $-0.0026 - (-0.0155) = 0.0129$ or 1.29 percentage points. That same value is 0.0102 ($= 0.0354 - 0.0252$) for investment. In comparison, the interquartile range for the Köeppen-Geiger measure is much larger ($0.1190 - (-0.0198) = 0.1388$), thus providing further evidence of heterogeneity. Table 3 also shows 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 Köeppen-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 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 of the distribution. In Figure 1, we examine the kernel density estimates of the vector of partial effects for each continuous regressor. Here we can see the entire spread of the estimates. It is worth noting that we find a significant percentage of the partial effects of initial income with a positive sign. As expected, we find that most of the mass for the partial effects of investment is to the right of zero implying that additional investment is shown to systematically increase output growth. That being said, we again note that this effect is quite heterogeneous across the sample. Finally, although we find our Köeppen-Geiger variable to be generally insignificant in Table 3, we see that there is substantial variation in the partial effect with some mass to both the left and right of zero.

We now return to the positive partial effects for the Köeppen-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, taken literally, this implies that a certain percentage of the observations are positive and significant reflecting that some countries with relatively cold climates could be worse off, other things equal. It is possible that after controlling for investment, the nonlinearities in the various controls, as well as regional effects, that the expected negative effect does not emerge as strongly as expected.

To look for an explanation for the change in sign, we split the estimated partial effects into countries with above and below median levels of the Köeppen-Geiger measure (0.20). In Figure 2, we plot the kernel density estimates of partial effects on the Köeppen-Geiger measure for each group in one panel. For those countries with above median Köeppen-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 Köeppen-Geigen values, we see that the density is shifted to the right.[17]

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 (1996) 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). Higher initial income, higher investment and lower values for the Köeppen-Geiger measure are associated with more negative partial effects of initial income. While theory suggests that none of these variables play a role in determining the converence rate in the vicinity of the steady state, most would suspect that countries with these attributes converge faster.

5.2.2 Demography

We now turn our focus to the demography theory as it has a large number of observations and the goodness-of-fit measure is relatively high.[18] The results for the Solow variables are as expected. The partial effects of initial income are generally negative, those of investment are mostly positive and those of population growth show both signs and are generally insignificant. For the demography variables, fertility is mostly insignificant whereas the reciprocal of life expectancy is shown to have a negative and significant effect on growth for 140 of the 715 observations in our sample. This result may be surprising at first but in fact it is consistent with the current state of the existing literature. While there is compelling microeconomic evidence that health is important for economic outcomes (see, e.g. Strauss and Thomas 1998), the macroeconomic evidence has been mixed. On the one hand, early empirical work by Gallup and Sachs (2001) and more recently by Lorentzen, McMillan and Wacziarg (2008) find a large positive effect. On the other hand, the findings of Acemoglu and

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

[18]Here we omit presenting the results formally to save space. They are available from the authors upon request.

Johnson (2007) and Weil (2007) question a strong impact of health on growth.

We extend our analysis to explore whether a particular group of countries is associated with the one fifth of observations significantly benefiting from improved life expectancy. Table 5 is analogous to Table 4 except now we are examining the partial effects of the reciprocal life expectancy variable based on splits of the continuous regressors. The first point worth making here is that the standard errors are quite large as compared to other partial effects in the paper. That being said, we find (expected) negative partial effects across the columns for below median initial income, above median population growth, above median fertility and above median reciprocal life expectancy. Each of these cases can be considered as proxies for low-income countries. We take these results to suggest that increases in life-expectancy would benefit less developed countries more than developed countries. Future work focusing solely on observations which produce negative and significant results could lead to a better understanding of what conditions are necessary to obtain economic gains from increases in life expectancy.

5.2.3 Policy

The final model we study in more detail is the policy model. This model is of interest for a variety of reasons. First, as DKT found, the policy (their macro) model had a high posterior probability and two of the three variables used as proxies (government consumption and inflation) also had posterior inclusion probabilities very close to 1. Second, the HLR test was unable to reject either parametric model. These two pieces of information suggest that a deeper look at how policy variables influence growth is warranted.

To avoid being repetitive, we look at a subset of the results for the policy theory. Analogous to Table 3, in Table 6 we present the quartile and mean values of the estimated coefficients for each of the continuous variables in the policy growth regression. The associated standard errors are listed underneath each estimate. The table suggests that there is some dispersion in the estimated impact that any given variable has on growth across country/time. That being said, the interquartile range for each of the estimated marginal effects for each variable is substantially smaller than what we saw with the geography theory. This does not necessarily imply that there is no variation in the partial effects, but it does suggest that a linear model is not far off the mark. Here we see that our nonparametric estimates are not substantially different from what we would expect to find in a typical parametric model.

This result suggests that this model may be best represented by a simple parametric model with a few quadratic terms as well as a few interactions. Future research on this theory should focus on trying to find an effective and efficient parametric model.

6 Conclusion

This paper uses recently developed methods for nonparametric regression to investigate potential nonlinearities in the growth process and to select relevant variables. This is beneficial as a variable might be wrongly dropped, or wrongly included, because of omitted nonlinearities. Further, estimation and inference from an improperly specified model could lead to incorrect policy prescriptions.

Our main findings are twofold. First, our perceptions on economic growth are necessarily linked to the types of empirical models used to link growth determinants and output. This suggests a more careful consideration of individual growth models and theories in future research since a linear parametric model may mislead one in both the direction of believing the theory is true when model misspecification is left unaccounted for 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 the consequences of said nonlinearities.

Second, in a specific context, we focused on three separate theories: geography, demography and policy. Further analysis of the geography theory revealed that this theory was heavily nonlinear and the simple parametric models traditionally considered missed substantial heterogeneity in the partial effects. Our study of the demography theory provided suggestive evidence that less-developed countries would gain the most from increases in life expectancy. Finally, our analysis of the policy theory showed that a simple parametric model may be appropriate as we found little heterogeneity in the partial effects.

A particularly promising line of future research is to investigate whether the nonparametric models used in this paper might translate into parametric models with several regimes, such as Durlauf and Johnson (1995) and Hansen (2000)-type threshold-based parametric models. As shown in this paper, when a given parametric model cannot be ascertained to be statistically valid via testing, partial effects based on given levels of the remaining variables (see e.g., Table 4) can be insightful. These splits could actually link with threshold effects. That is, if distinct differences based on these splits exist, then this may suggest that a parametric type threshold model could in fact be valid.

References

[1] Acemoglu, D. and S. Johnson, 2007. "Disease and Development: The Effect of Life Expectancy on Economic Growth," *Journal of Political Economy*, 115: 925-985.

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

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

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

[5] Block, S. A., 2001. "Does Africa Grow Differently?" *Journal of Development Economics,* 65: 443-467.

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

[7] Cameron, A. C. and P. K. Trivedi, 2005. *Microeconometrics,* Cambridge University Press, Cambridge.

[8] Ciccone, A. and M. Jarocinski, 2010. "Determinants of Economic Growth: Will Data Tell?" *American Economic Journal - Macroeconomics,* 2(4): 222-246.

[9] Collier, P. and A. Hoeffler, 1998. "On the Economic Causes of Civil War," *Oxford Economic Papers*, 50: 563-573.

[10] 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.

[11] Doppelhofer, G. and M. Weeks, 2009. "Jointness of Growth Determinants," *Journal of Applied Econometrics,* 24: 209-244.

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

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

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

[15] Easterly, W. and R. Levine, 1997. “Africa’s Growth Tragedy: Policies and Ethnic Divisions,” *Quarterly Journal of Economics*, 112(4): 1203-1250.

[16] Eicher, T. S., A. Lenkoski and A. E. Raftery, 2009. “Bayesian Model Averaging and Endogeneity Under Model Uncertainty: An Application to Development Determinants,” Working Paper no. 94, Center for Statistics and the Social Sciences, University of Washington.

[17] Eicher, T. S., C. Papageorgiou and A. E. Raftery, 2011. “Default priors and predictive performance in Bayesian model averaging, with application to growth determinants” *Journal of Applied Econometrics*, 26(1): 30-55.

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

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

[20] Gallup, J. L., J. D. Sachs and A. D. Mellinger, 1999. “Geography and Economic Development,” *International Regional Science Review*, 22(2): 179-232.

[21] Gottardo, R. and A. E. Raftery, 2007. “Bayesian Variable and Transformation Selection: A Unified Approach,” Statistics Department, University of Washington, Technical Report 570.

[22] 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.

[23] Hansen, B. E., 2000. “Sample Splitting and Threshold Estimation,” *Econometrica* 68: 575-603.

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

[25] 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.

[26] 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.

[27] 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.

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

[29] Kaufmann D., A. Kraay and M. Mastruzzi, 2005. "Governance Matters IV: Governance Indicators for 1996-2004," World Bank Policy Research no. 3630.

[30] 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.

[31] 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.

[32] Kraay, A. and N. Tawara, 2010. "Can Disaggregated Indicators Identify Governance Reform Priorities?" World Bank Policy Research Working Paper no. 5254.

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

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

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

[36] Leamer, E. E. and H. Leonard, 1983. "Reporting the Fragility of Regression Estimates," *Review of Economics and Statistics*, 65(2): 306-317.

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

[38] 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.

[39] Ley, E. and M. F. J. Steel, 2009a. "Comments on Jointness of Growth Determinants," *Journal of Applied Econometrics,* 24: 248-51.

[40] Ley, E. and M. F. J. Steel, 2009b. "On the effect of prior assumptions in Bayesian Model Averaging with applications to growth regression," *Journal of Applied Econometrics*, 24: 651-674.

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

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

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

[44] Lorentzen, P., J. McMillan and R. Wacziarg, 2008. "Death and Development," *Journal of Economic Growth*, 13: 81-124.

[45] 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.

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

[47] Masanjala, W. and C. Papageorgiou, 2008. "Rough and Lonely Road to Prosperity: A reexamination of the sources of growth in Africa using Bayesian Model Averaging," *Journal of Applied Econometrics,* 23: 671-682.

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

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

[50] Montalvo, J. G. and M. Reynal-Querol, 2005. "Ethnic Diversity and Economic Development," *Journal of Development Economics*, 76(2): 293-323.

[51] Moral-Benito, E., 2010a. "Panel Growth Regressions with General Predetermined Variables: Likelihood-Based Estimation and Bayesian Averaging," working paper, CEMFI.

[52] Moral-Benito, E., 2010b. "Determinants of Economic Growth: A Bayesian Panel Data Approach," *Review of Economics and Statistics*, forthcoming.

[53] Pesaran, M.H., C. Schleicher and P. Zaffaroni, 2009. "Model Averaging in Risk Management with an Application to Futures Markets," *Journal of Empirical Finance,* 16: 280-305.

[54] 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.

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

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

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

[58] Soares, R. R., 2007. "On the Determinants of Mortality Reductions in the Developing World," *Population and Development Review*, 33(2): 247-287.

[59] Strauss, J. and D. Thomas, 1998. "Health, Nutrition and Economic Development," *Journal of Economic Literature,* 36: 766-817.

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

[61] Temple J., 1998b. "Initial Conditions, Social Capital and Growth in Africa," *Journal of African Economies*, 7(3): 309-347.

[62] Tsangarides, C. G., 2004. "A Bayesian Approach to Model Uncertainty," IMF Working Paper, no. 04/68.

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

[64] Zeugner S. and M. Feldkircher, 2009. "Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging," IMF Working Paper, no. 09/202.

Table 1: Bandwidths using Local-Constant Regression and Penn World Table 6.1

Variable	Solow	Region	Demo.	Geo.	Policy	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	.	.	.	.
Keppen-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*

Bandwidths obtained using LSCV as described in the text for the local-constant nonparametric regression. A bandwidth with a * next to it indicates that this variable is smoothed out of the regression.

Table 2: Bandwidths using Local-Linear Regression and Penn World Table 6.1

Variable	Solow	Region	Demo.	Geo.	Policy	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	.	.	.	.
Keppen-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

Bandwidths obtained using LSCV as described in the text for the local-linear nonparametric regression. A bandwidth with a $^+$ next to it indicates that this variable enters the regression in a linear fashion.

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
Keppen-Geiger	0.0601	-0.0198	0.0588	0.1190
	0.0507	0.0258	0.0426	0.0590

Partial effects are tabulated as the estimated derivatives from the local-linear regression using the bandwidths obtained in Table 2 for the Geography (Geo) column. The estimate in the table for a particular continuous regressor represents the mean, median ($Q2$), first ($Q1$) or third ($Q3$) quartile of the vector of partial effects for that particular regressor. Beneath each estimate is the corresponding (wild) bootstrapped standard error.

Table 4: Partial effects of initial income across various groups for the 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 Keppen-Geiger	-0.0089	-0.0175	-0.0066	-0.0002	0.0000
	0.0027	0.0025	0.0047	0.0064	
Below Median Keppen-Geiger	-0.0116	-0.0153	-0.0137	-0.0085	
	0.0025	0.01153	0.0039	0.0057	

Partial effects are tabulated as the estimated derivatives from the local-linear regression using the bandwidths obtained in Table 2 for the Geography (Geo) column. Results are tabulated based on splits of the data indicated by the first column of this table. The estimate in the table for a particular continuous regressor represents the mean, median ($Q2$), first ($Q1$) or third ($Q3$) quartile of the vector of partial effects (based on the split) for that particular regressor. Beneath each estimate is the corresponding (wild) bootstrapped standard error.

Table 5: Partial effects of (the reciprocal of) life expectancy (at age 1) across various groups for the demography model

Variable	Mean	$Q1$	$Q2$	$Q3$	Li Test
Above Median Initial Income	-0.0569	-3.867	0.3607	4.2816	0.0000
	1.1738	1.4480	1.5164	7.6324	
Below Median Initial Income	-1.9860	-3.3731	-2.1264	-0.4583	
	1.4830	6.7051	3.0681	4.8863	
Above Median Population Growth	-1.6618	-3.3378	-2.1313	-0.3470	0.0000
	1.8149	1.5295	3.0252	3.2255	
Below Median Population Growth	-0.3767	-4.0400	0.2161	4.3002	
	3.0755	5.3138	1.1626	8.5103	
Above Median Investment	-0.3980	-4.1268	-0.5829	3.9685	0.0000
	3.2796	1.1339	1.5994	1.7870	
Below Median Investment	-1.6439	-3.2734	-1.8776	0.3938	
	9.0246	1.4725	9.1179	8.7375	
Above Median Fertility Rate	-2.4022	-3.4623	-2.2887	-1.0647	0.0000
	2.2499	5.0324	7.4060	1.1401	
Below Median Fertility Rate	0.3657	-3.3283	1.1866	4.9003	
	1.5155	7.745	7.4900	1.2758	
Above Median Life Expectancy	-2.1598	-3.4175	-2.1537	-0.9021	0.0000
	8.4173	1.7896	1.7824	1.2064	
Below Median Life Expectancy	0.12276	-3.6379	0.8454	4.6098	
	4.6533	1.8152	6.7051	7.9399	

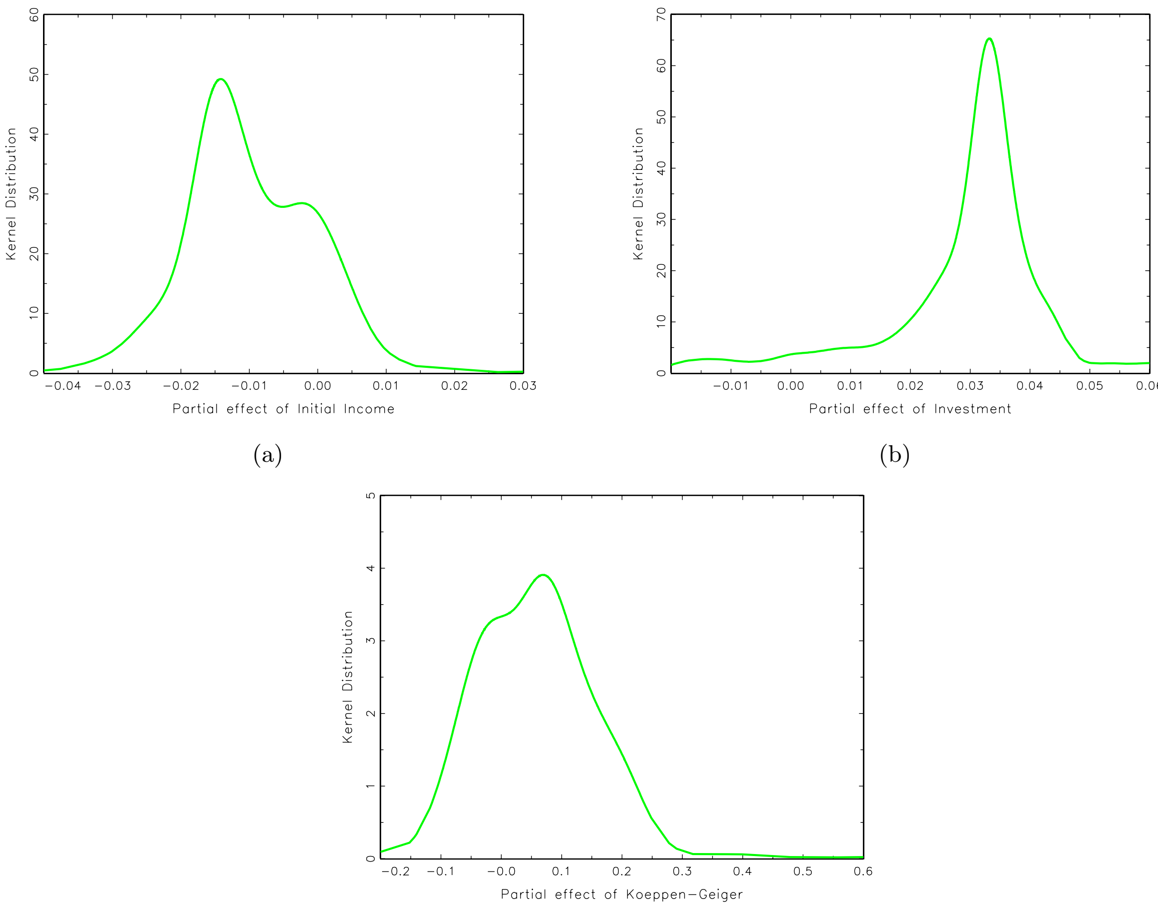
Partial effects are tabulated as the estimated derivatives from the local-linear regression using the bandwidths obtained in Table 2 for the Demography (Demo.) column. Results are tabulated based on splits of the data indicated by the first column of this table. The estimates in the table for a particular continuous regressor represents the mean, median ($Q2$), first ($Q1$) or third ($Q3$) quartile of the vector of partial effects (based on the split) for that particular regressor. Beneath each estimate is the corresponding (wild) bootstrapped standard error. Recall that a negative partial effect implies that increases in life- expectancy will lead to increases in economic growth as life is defined as the inverse of life expectancy at age one.

Table 6: Partial effects for all continuous regressors for the policy model

m	*Mean*	*Q1*	*Q2*	*Q3*
Initial Income	−0.0133	−0.0167	−0.0144	−0.0047
	0.0036	0.0038	0.0035	0.0034
Population Growth	−0.0081	−0.0301	−0.0128	0.0074
	0.0088	0.0126	0.0179	0.0198
Human Capital	0.0007	−0.0004	0.0001	0.0017
	0.0008	0.0009	0.0008	0.0008
Investment	0.0248	0.0175	0.0250	0.0282
	0.0041	0.0042	0.0030	0.0039
Openness	0.0075	−0.0007	0.0069	0.0156
	0.0062	0.0170	0.0053	0.0065
Net Govt. Cons.	−0.1171	−0.1345	−0.1053	−0.0878
	0.0328	0.0287	0.0249	0.0266
Inflation	−0.0004	−0.0006	−0.0004	−0.0002
	0.0004	0.0003	0.0005	0.0005

Partial effects are tabulated as the estimated derivatives from the local-linear regression using the bandwidths obtained in Table 2 for the Policy column. The estimates in the table for a particular continuous regressor represents the mean, median ($Q2$), first ($Q1$) or third ($Q3$) quartile of the vector of partial effects for that particular regressor. Beneath each estimate is the corresponding (wild) bootstrapped standard error.

Figure 1: Estimated marginal effects for each continuous regressor for the geography theory



(a)

(b)

(c)

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

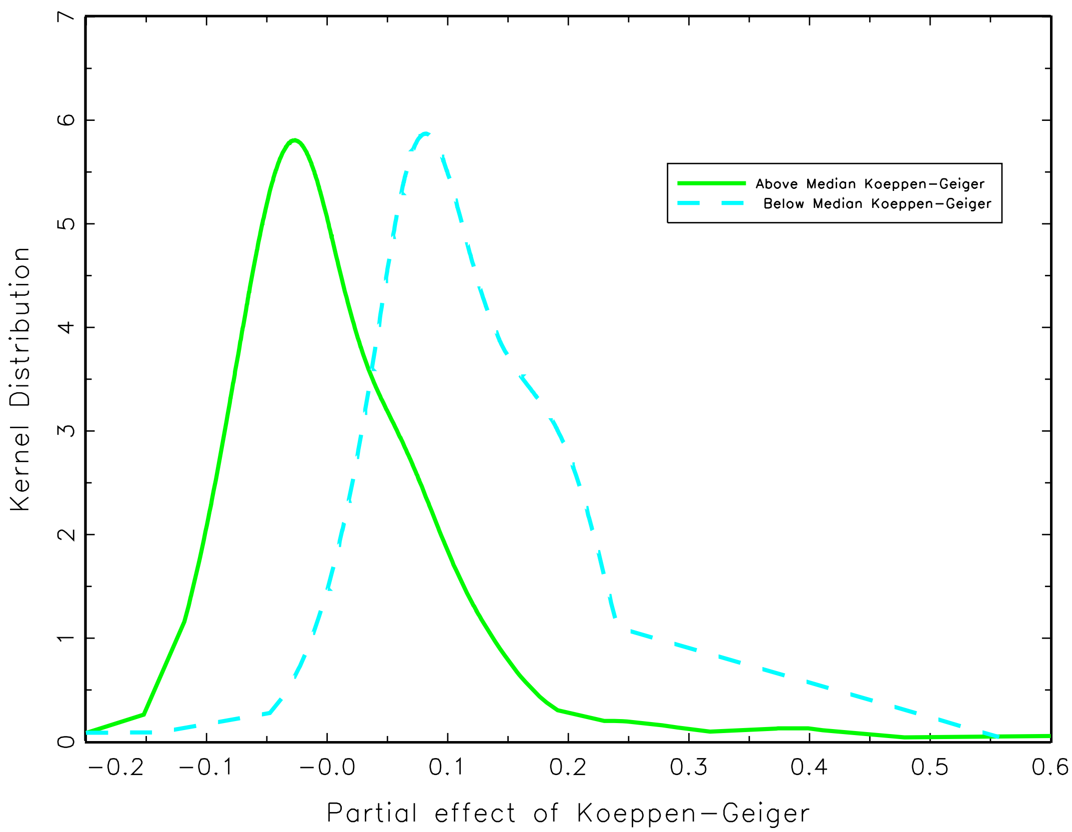
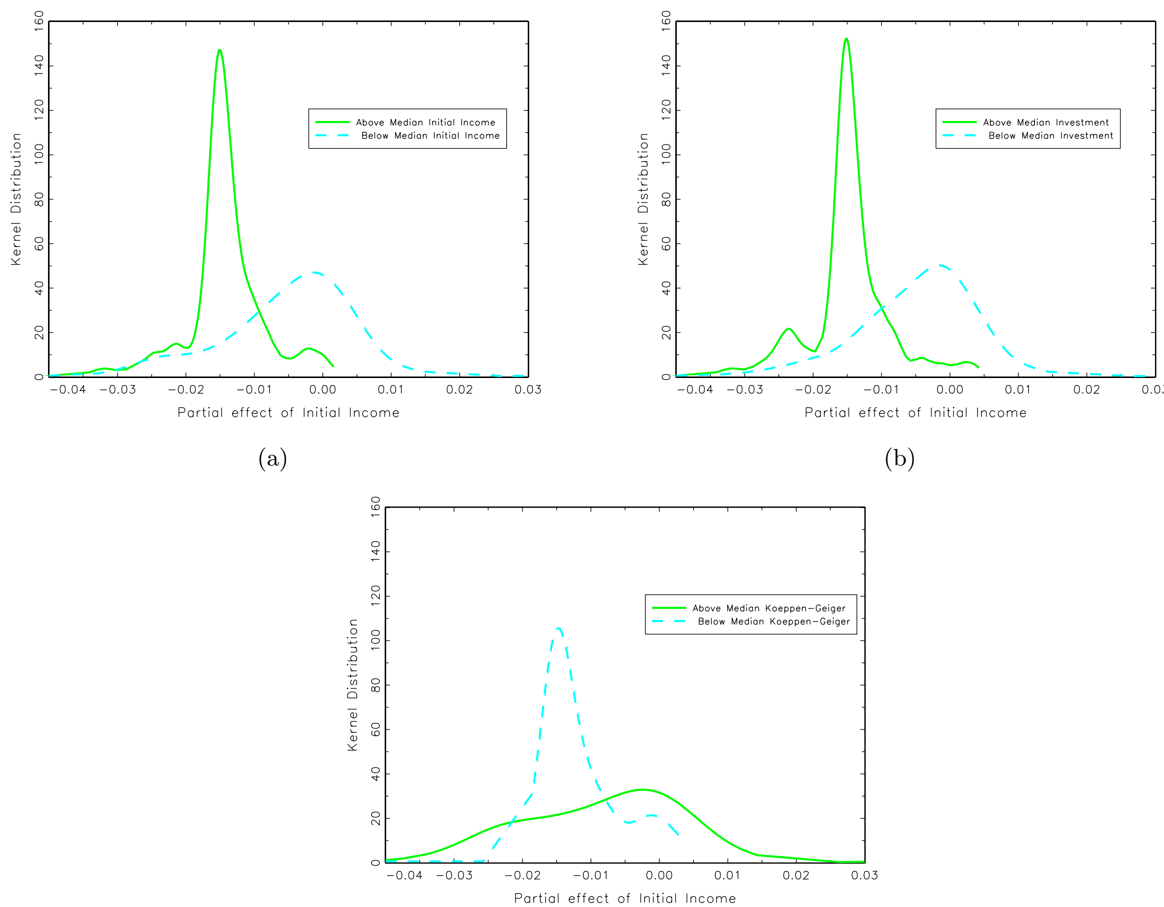


Figure 3: Comparison of estimated marginal effects of initial income by various splits for the geography theory



(a)

(b)

(c)

A 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, 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 in a manner which we would expect in a panel growth study. 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 $\Pr[w_{1i0} = \ell] = 0.25\ \forall \ell$, $\Pr[w_{2i0} = 0] = 0.40$ 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. Notice that we have constructed x_3 to have a unit root while x_4 and x_5 are explosive. We choose to construct our irrelevant variables in this fashion for two reasons. First, the theoretical results of Hall, Li and Racine (2007) consider well-behaved, *relevant* covariates. Second, their simulations have covered cases where the *irrelevant* covariates display nonzero correlation with the relevant covariates but are well-behaved. Thus, our simulations can be viewed as a complement to those of Hall, Li and Racine (2007). 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.[19]

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.[20]

[19]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.

[20]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.

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 A1 and A2 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 and x_2 are relevant. Table A1 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 A2 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 A1 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 A2 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. The limited evidence here suggests that when the regressor possesses a unit root that the user should be careful interpreting the result. At the same time, we see that the explosive irrelevant variables ($\varphi > 1$: x_4 and x_5) are correctly smoothed out. It is unclear why the unit root would behave worse than the explosive cases. Clearly, further careful research (both theoretical and empirical) needs to be done before we can make strong claims on this result.

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 on n_1 out-of-sample points is assessed via $PMSE = 1/n_1 \sum_{j=1}^{n_1} (\hat{y}_j - y_j)^2$. 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 estimators for the first two models should lead to inconsistent estimates while the second two are consistent estimators. Table A3 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 (in terms of the ratio of PMSE across the models) from 28.8% to 24.5%. We also mention 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 in the main text.

Table A1: Summary of cross-validated bandwidths for the discrete covariates (NP LSCV estimator)

	Median, [10th Percentile, 90th Percentile] of $\hat{\lambda}$				
	$\hat{\lambda}_{z_1}$	$\hat{\lambda}_{z_2}$	$\hat{\lambda}_{z_3}$	$\hat{\lambda}_{w_1}$	$\hat{\lambda}_{w_2}$
$n = 59$					
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]
$n = 95$					
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]

Median bandwidths for our discrete covariates using LSCV and local-constant kernel regression. The interdecile range of our estimated bandwidths are presented in square brackets beneath the median value.

Table A2: Summary of cross-validated bandwidths for the continuous covariates (NP LSCV estimator)

	Median, [10th Percentile, 90th Percentile] of $\hat{h}$				
	$\hat{h}_{x_1}$	$\hat{h}_{x_2}$	$\hat{h}_{x_3}$	$\hat{h}_{x_4}$	$\hat{h}_{x_5}$
$n = 59$					
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,$\approx\infty$]
$n = 95$					
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,$\approx\infty$]

Median bandwidths for our continuous covariates using LSCV and local-constant kernel regression. The interdecile range of our estimated bandwidths are presented in square brackets beneath the median value. 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.

Table A3: 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]

Median predicted mean square errors (PMSE) for our different estimation methodologies. The interdecile range of our estimated PMSE are presented in square brackets beneath the median value. NP-LSCV refers to our local-constant nonparametric regression estimates with bandwidths obtained via LSCV, PI-ALL refers to a linear in parameters model including all variables (relevant and irrelevant), PI-ONLY is our linear in parameters model that only includes the relevant variables and PC is the correctly specified parametric model.