Wesleyan Economic Working Papers

http://repec.wesleyan.edu/

N^o: 2006-004

Cleaning Up the Kitchen Sink: Growth Empirics When the World Is Not Simple

Francisco Rodríguez

May 17, 2007

WESLEYAN UNIVERSITY

Department of Economics
Public Affairs Center
238 Church Street
Middletown, CT 06459-007

Tel: (860) 685-2340
Fax: (860) 685-2301
http://www.wesleyan.edu/econ

Cleaning Up the Kitchen Sink:
Growth Empirics When the World Is Not Simple

Francisco Rodríguez[1]
Department of Economics
Wesleyan University
First Version: September 7, 2005
This Version: May 17, 2007

Abstract: This paper explores the relevance of unknown nonlinearities for growth empirics. Recent theoretical contributions and case-study evidence suggest that non-linearities are pervasive in the growth process. I show that the postwar data provide strong evidence in favor of generalized non-linearities. I provide two alternative mechanisms for making inference about the effects of production-function shifters on growth that do not make *a priori* assumptions about functional form: monotonicity tests and average derivative estimation. The results of these tests point towards a greater role for structural variables and a smaller role for policy variables than the linear model.

JEL Categories: O40, O47
Keywords: Economic Growth, Cross-Country Growth Regressions, Non-linearities, Non-parametric econometrics.

---

[1] The author is grateful to the Department of Economic and Social affairs of the United Nations for its financial support. María Ángela Parra provided extensive and invaluable collaboration in the early stages of writing this paper. Suggestions by María Eugenia Boza, Donald Davis, Ricardo Hausmann. Chang-Tai Hsieh, Alex Julca, Aart Kraay, Jomo K.S., Norman Loayza, Edward Miguel, José Antonio Ocampo, Codrina Rada, Sanjay Reddy, Roberto Rigobón, Dani Rodrik, Cameron Shelton, Eric Verhoogen, Eduardo Zambrano and seminar participants at Amherst College, Harvard University, the United Nations, the World Bank and Wesleyan University were extremely helpful. Reyes Rodríguez provided first-rate research assistance. All errors remain the authors' pure responsibility.

> The supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.
>
> *Albert Einstein (1933, pp.10-11)*

1. Introduction

How important are non-linearities in the growth process? Over the past fifteen years, a consensus appears to have emerged among policymakers and applied economists that the growth effects of economic reforms are likely to depend on a country's initial conditions in a complex way. In the words of a recent World Bank study:

> "To sustain growth requires key functions to be fulfilled, but there is no unique combination of policies and institutions for fulfilling them…different polices can yield the same result, and the same policy can yield different results, depending on country institutional contexts and underlying growth strategies." (World Bank, 2005, p. 12)

Case studies of country and regional experiences often emphasize the relevance of structural and institutional conditions for determining the effects of policy reforms. For example, Sachs and Woo (1994) and Qian (2000) have argued that China's large and unproductive agricultural commune system allowed gradual price liberalization to generate high productivity growth rates that did not materialize in Eastern European economies, where most employment was initially concentrated in state-owned enterprises. Maloney (2006) contends that the lack of institutions that facilitate the adoption and creation of new technologies is key for an understanding of Latin America's inability to convert natural resource rents into economic growth. Ocampo (2004) argues that the effect of market reforms in Latin America varied broadly depending on countries' initial level of development, geographic proximity to the United States, and size of external and fiscal debt overhangs.

Recent contributions to the theoretical growth literature have also focused attention on the role of non-linearities in the growth process. Hausmann, Rodrik and Velasco (2005) use Lipsey and Lancaster's (1956) Theorem of the Second Best to argue that the reduction of a particular distortion may have very different effects on welfare and growth depending on the initial levels of other distortions. Their theoretical examples illustrate the potentially complex interactions that can arise even in relatively simple growth models. Aghion and Griffith (2005) present a theoretical argument for why the link between innovation and competition – crucial to many endogenous growth models – should be non-linear, and provide extensive microeconomic evidence that such nonlinearities effectively characterize firm-level manufacturing data. Non-linearities are of course not new to growth theory, but what distinguishes the more recent literature from prior contributions (e.g., Murphy, Sgleifer and Vishny, 1989, Azariades and Drazen, 1990) is its insistence that these results are not only theoretically interesting but vital for making sense of the empirical evidence.

This paper asks two related questions about the consequences of generalized non-linearities in the growth process. First, to what extent is there evidence that these non-linearities are present in commonly used cross-country growth data? Our answer to this question is strongly affirmative: we find that in a very high number of specifications the linear model can be comfortably rejected, while a fraction of specifications also reject the hypothesis of separability (low-dimensionality). Second, what are the appropriate mechanisms to make inferences on this data without making recourse to special parametric assumptions? That is, what can we conclude from the growth data if we effectively treat the functional form characterizing existing non-linearities as *unknown* instead of as given by theory? Here our message is cautiously optimistic. While the estimation tools we explore – additive separability, monotonicity tests and average derivative estimation – enable us to reach some interesting conclusions about the effects of policies, institutions and structure on the growth process, these conclusions are necessarily more limited than those that we could reach if we were in a linear world.

This is far from the first paper to empirically analyze the relevance of non-linearities in the growth process. Indeed, there is a vast empirical literature that studies the existence of non-linear effects in the context of growth regressions. Some illustrative examples are Barro (1996) on democracy, Liu and Stengos (1999) on education, Banerjee and Duflo (2003) on inequality, DeJong and Ripoli (2006) on tariffs and Chang, Kaltani, and Loayza (2006) on trade ratios. Almost invariably, the approach of these papers is to explore non-linearities with respect to the dimension of interest through parametric or non-parametric methods while assuming linearity in the remaining regressors. Kalatzidakis et. al. (1999) have studied higher-dimensional non-linearities through non-parametric estimation, but their interest was in testing the robustness of inferences about the linear part of the specification to non-linearities in a set of auxiliary variables. Durlauf and Johnson (1995) and Durlauf, Kourtellos and Minkin (2001) have studied more generalized non-linearities using a model of parameter heterogeneity in which different sets countries are characterized by different linear models. These exercises commonly make strong assumptions as to the form that the underlying heterogeneity takes. Despite these explorations, the standard workhorse regression model is still that of the linear regression framework. [2]

My approach differs from the above contributions in that I center on studying the effect of *unknown, generalized* non-linearities. More concretely, I will argue that it is unreasonable to make *a priori* assumptions about the functional form through which a set of variables enter into the growth function. Doing so presumes that we know much more than we actually do about the data generating process. I will also argue that it makes sense to treat these non-linearities as possibly high-dimensional, instead of concentrating on the dimension of interest while assuming that the rest of

[2] For example, Sala-i-Martin, Doppelhoffer and Miller's (2004) propose a Bayesian Averaging of Classical Estimates methodology to study the robustness of results previously found in the literature. Among the 67 potential explanatory variables that they consider, only one (inflation) is treated non-linearly through a quadratic term.

the model is linear. Given these assumptions, I will ask, what can we conclude about the effects of these variables on the growth process?

The rest of the paper proceeds as follows. In Section 2 I analyze the theoretical basis for the linear growth regression and discuss the likely effects of misspecification bias in this framework. Section 3 briefly discusses the data set and presents tests for non-linearities and non-separabilities. It then discusses several alternatives for drawing inferences about the nature of the growth function when there are unknown and generalized non-linearities. Section 4 concludes and suggests directions for further research.

2. Theoretical Considerations

2.1. Who Threw In the Kitchen Sink? The Theoretical Basis of the Kitchen Sink Regression

Our starting point is a brief exposition of the theoretical foundations for the linear growth regression. This regression, often referred to as a "Barro" regression because of the deep influence of Robert Barro's path-breaking 1991 *Quarterly Journal of Economics* article. It consists of a regression where economic growth is the dependent variable and the specification is linear in the log of initial GDP, some measures of investment in physical and human capital, population growth, and a set of "production function shifters" that commonly includes policy, institutional and structural controls. Formally, the specification often looks like:

$$\gamma_Y = \alpha_0 + \alpha_1 \ln y_{t-1} + \alpha_2 s_k + \alpha_3 s_h + \alpha_4 n + \beta Z \qquad (1)$$

where γ_Y is the rate of per capita GDP growth, y_{t-1} is initial GDP, s_k and s_h refer respectively to the rates of investment in physical and human capital, n is the rate of population growth and $\boldsymbol{Z}$ is the vector of potential production function shifters.

Given the ease of running this regression with readily available data sets and the obvious interest of exploring whether a particular set of policies, institutions or structural variables are harmful or beneficial for growth, the proliferation of applied work using equation (1) is not surprising. For obvious reasons, I will not discuss this voluminous literature here; the reader is referred to Durlauf, Johnson and Temple (2005) for a recent comprehensive survey. It suffices to note for our purposes that this analysis tends to take the form of varying the subset of variables included in Z and using conventional significance tests to evaluate the effect of potential determinants on economic growth.

Popular opinion to the contrary, the linear growth regression is not a purely *ad-hoc* specification. Its analytical foundations were elaborated early on in the literature and, to my knowledge, were first presented systematically in Mankiw, Romer, and Weil (1992, henceforth MRW). The MRW specification, however, differs from equation (2) in a number of important respects, so it is useful to briefly revisit it.

The MRW model can be fully described by a Cobb-Douglas production function that maps inputs of physical capital, human capital, and raw labor into output and one accumulation equation for each type of capital:

$$Y_t = K^{\alpha}{}_t H^{\beta}{}_t \left(A_t L_t\right)^{1-\alpha-\beta}, \qquad (2)$$

$$\dot{\hat{k}}_t = s_k \hat{y}_t - (n+g+\delta)\hat{k}_t, \quad (3)$$
$$\dot{\hat{h}}_t = s_h \hat{y}_t - (n+g+\delta)\hat{h}_t. \quad (4)$$

where Y_t denotes output, K_t the stock of physical capital, H_t the stock of human capital, A_t is a productivity shift term, s_i the rate of investment in factor i (i={K,L}), δ the depreciation rate, and "hatted" variables denote quantities per units of effective labor (e.g.: $\hat{y}_t = \frac{Y_t}{AL_t}$). n and g denote the proportional growth rates of L_t and A_t and are taken as exogenous to the system in the MRW formulation.

Substituting equation (2) into (3) and (4) gives us a system of two non-linear differential equations in $\hat{k}$ and $\hat{h}$ which can in principle be solved for numerically given initial conditions $\hat{k}_0$ and $\hat{h}_0$. An analytical approximation to this solution can be arrived at by linearizing the growth rates of physical and human capital near their steady states, given by $\dot{\hat{k}}_t = \dot{\hat{h}}_t = 0$. Doing this allows us to express the rate of growth of $\hat{y}$ as:

$$\dot{\hat{y}} = \alpha \frac{d \ln \hat{k}_t}{dt} + \beta \frac{d \ln \hat{h}_t}{dt} = -(n+g+\delta)(1-\beta-\alpha)\ln(\hat{y}/\hat{y}_{ss}), \quad (5)$$

where the second equality follows from linearly approximating the growth rates of $\hat{k}$ and $\hat{h}$ and

$$\ln(\hat{y}_{ss}) = \frac{\alpha}{1-\alpha-\beta}\ln(s_k) + \frac{\beta}{1-\alpha-\beta}\ln(s_h) - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta), \quad (6)$$

denotes the log of the steady-state level of income.

Equation (5) is a first-order linear differential equation in $\ln(\hat{y}_t)$. We can solve it and express the solution in terms of the growth rate between $\hat{y}_0$ and $\hat{y}_t$ as:

$$\ln(\hat{y}_t / \hat{y}_0) = -(1-e^{-\lambda t})\ln \hat{y}_0 + (1-e^{-\lambda t})\ln \hat{y}_{ss}, \quad (7)$$

where $\lambda = (n+g+\delta)(1-\alpha-\beta)$ denotes the rate of convergence. This expression can be rewritten in terms of observables as:

$$\ln(y_t / y_0) = -(1-e^{-\lambda t})\ln y_0 + (1-e^{-\lambda t})\left[\frac{\alpha}{1-\alpha-\beta}\ln s_k + \frac{\beta}{1-\alpha-\beta}\ln s_h - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta)\right]$$
$$+ g + (1-e^{-\lambda t})\ln A_0. \quad (8)$$

This equation is linear in lny_o, lns_k, lnh_{ss}, ln(n+g+δ), g and lnA_o and would thus be estimable by linear methods if we had observations on all of these variables. Since we do not observe g nor A_o this is not yet possible unless we make an assumption about their behavior. MRW assume that g is constant and equal across countries and that differences in the initial level of technology vary randomly according to:

$$\ln(A_0) = \ln(A) + \varepsilon_i \quad (9)$$

with ε_i representing a country-specific shock. Given these assumptions as well as a value for the common $g+\delta$, equation (2.12) can be estimated by fitting the linear regression:

$$\gamma_t = A_0 + A_1 \ln y_0 + A_2 \ln s_k + A_3 \ln s_h + A_4 \ln(n+g+\delta) + \eta_i. \qquad (10)$$

to the data. This is, indeed, what MRW do.

Equation (10) also appears to open the door to a more general approach. As MRW note, "the *A(0)* term reflects not just technology but resource endowments, climate, institutions and so on." If differences across countries are not simply randomly distributed but ε_i is correlated with any of the regressors in (11) or (12), the resulting least squares coefficients would be contaminated by omitted variable bias. Even if this source of bias is unimportant, equations (8)-(10) seems to offer a ready framework to evaluate the effect of multiple measures of policies, institutions and economic structure on growth, by replacing A_O with a fuller, more detailed specification of potential production function shifters. Most of the modern growth empirics literature can be interpreted as doing precisely this.

However, it is interesting to note what happens if one tries to replace A_t by a more general function $A(\mathbf{Z}_t)$, where $\mathbf{Z}_t$ is a vector of potential explanatory variables such as economic policies, institutions, and structural characteristics, which are often thought to affect growth by affecting the capacity of the economy to transform inputs into GDP. In that case we would rewrite (2) as:

$$Y_t = K^{\alpha}{}_t H^{\beta}{}_t \left(A(Z_t)L_t\right)^{1-\alpha-\beta}. \qquad (11)$$

The above derivation would then follow except for the fact that A_O would now be replaced by $A(Z_O)$ and $g=\ln(A_t/A_O)$ by $g(Z_0, Z_t) = \ln[A(Z_t)/A(Z_0)]$. Equation (8) would become:

$$\ln(y_t/y_0) = -(1-e^{-\lambda t})\ln y_0 + (1-e^{-\lambda t})\begin{bmatrix}\dfrac{\alpha}{1-\alpha-\beta}\ln s_k + \dfrac{\beta}{1-\alpha-\beta}\ln s_h \\ -\dfrac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g(Z_t,Z_0)+\delta)\end{bmatrix}$$
$$+ g(Z_t, Z_0) + (1-e^{-\lambda t})\ln A(Z_0). \qquad (12)$$

What is interesting about equation (12) is that it is *not* a linear equation in the components of *Z*. In order to make it into a linear function of *Z* one would need to add in two additional assumptions. In the first place, one needs to assume that the log of *A(Z)* is linear in the production function shifters, i.e., that

$$A(Z) = Z_1^{\beta_1} \ldots Z_n^{\beta_n}. \qquad (13)$$

Additionally, one needs to assume that the growth rate of *A* over time is the same for all countries, that is, that

$$g_i(Z_0, Z_t) = \overline{g} \;\forall i. \qquad (14)$$

Only given these assumptions is it that (12) reduces to the traditional kitchen-sink specification:

$$\ln(y_t / y_0) = -(1-e^{-\lambda t})\ln y_0 + (1-e^{-\lambda t})\left[\frac{\alpha}{1-\alpha-\beta}\ln s_k + \frac{\beta}{1-\alpha-\beta}\ln h_{ss} - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+\bar{g}+\delta)\right]$$
$$+\bar{g} + (1-e^{-\lambda t})\beta Z_0. \qquad (15)$$

How reasonable are these two additional assumptions? Equation (13) can be defended along the lines that it is essentially a translation of the Cobb-Douglas assumption to the determinants of productivity. If we are willing to make such a simplification in order to describe the relationship between physical inputs and outputs, it may not be too much of an additional stretch to assume that it can also describe the relationship between output and production function shifters. However, it should be noted that while the Cobb-Douglas assumption can be directly evaluated either by microeconomic evidence or by analysis of the behavior over time of national factor shares,[3] there is no evident way to so this with the assumption embodied in equation (13). Indeed, this assumption directly rules out the second-best effects advocated by Hausmann, Rodrik, and Velasco (2004).

Equation (14) embodies an assumption that is fundamentally at least as, if not considerably more, problematic. Why would one expect all countries to have the same rate of change in *A(Z)* if they differ in the fundamental *Z*'s? One possible line of defense is to see *g* as capturing only the effects of technological change, which is assumed to be public and available to all countries, while $A(Z_0)$ is held to be fixed at its initial level. This leaves unanswered the questions raised by the terms of *Z* that have no relation to technological diffusion. While the assumption that they are time-invariant may be adequate for thinking about some production function shifters such as economic geography and perhaps institutions, it is much less useful if one wants to understand the effect of variables like economic policies, institutional reform or structural change.[4]

Given how untenable the assumption contained in equation (14) seems, is there any other way to salvage the linear specification in growth empirics? An alternative is to limit the evolution of the production-function shifters contained in *Z* to other trajectories than those given by equation (14). One possibility is to treat each Z_i as evolving towards its own long-run equilibrium Z_i*, so that the model can effectively be linearized near the full steady state, which is characterized by the fact that not only the stocks of human and physical capital, but also the institutional, political, and structural production-function shifters, are near their rest points. For example, we could write the production function as:

$$Y_t = K^{\alpha}{}_t H^{\beta}{}_t (AL_t)^{1-\alpha-\beta} \prod_{i=1}^{n} Z_i^{\theta_i}, \qquad (16)$$

[3] See, for example, Gollin (2002), Bernanke and Gurkaynak(2002) and Rodríguez and Ortega (2006).
[4] The almost universal neglect of this non-linearity in the growth literature has been previously called attention to by Durlauf et al. (2005, p. 580).

where we have used equation (13). Equations (3) and (4) would still characterize the evolution of the stocks of human and physical capital, but equation (14) would now be replaced by the set of *n* differential equations:

$$\dot{Z}_i = -\varphi_i Z_i \ln(Z_i / Z_i^*),\ i = \{1,\ldots,n\} \quad (17)$$

where $\varphi_i > 0$ and Z_i* denotes the long-run equilibrium of Z_i. In this model, we sue *g* to denote the proportional growth rate of the purely technological term A_t and assume that it is common across countries. In this case equations (5) and (6) will now be replaced by:

$$\dot{\hat{y}} = \alpha \frac{d \ln \hat{k}_t}{dt} + \beta \frac{d \ln \hat{h}_t}{dt} = -(n+g+\delta)(1-\beta-\alpha)\ln(\hat{y}/\hat{y}_{ss}) + \sum_{i=1}^{n}\left[(n+g+\delta)-\varphi_i\right]\theta_i \ln(Z_i / Z_i^*)$$

, (18)

and

$$\ln(\hat{y}_{ss}) = \frac{\alpha}{1-\alpha-\beta}\ln(s_k) + \frac{\beta}{1-\alpha-\beta}\ln(s_h) - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta) + \frac{1}{1-\alpha-\beta}\sum_{i=1}^{n}\theta_i \ln Z^*_i$$

, (19)

In the special case where $\varphi_i = n+g+\delta$ equation (18) again reduces to an ordinary differential equation in *ln(y)*, which can be solved as before leading to the linear expression for the growth rate:

$$\ln(y_t / y_0) = (e^{\lambda t} - 1)\ln y_0 + (1-e^{\lambda t})\frac{\alpha+\beta}{\alpha+\beta-1}\ln(n+g+\delta) + (1-e^{\lambda t})\frac{\alpha}{\alpha+\beta-1}\ln s_k$$
$$+ (1-e^{\lambda t})\frac{\beta}{\alpha+\beta-1}\ln s_h - (1-e^{\lambda t})\frac{1}{\alpha+\beta-1}\sum_{i=1}^{n}\theta_i \ln Z^*_i \quad (20)$$

The assumption that the evolution of the production-function shifters in **Z** converges to a steady-state level can thus rescue the linearity assumption. There are, however, a number of shortcomings with this approach. The first one is that it is by no means clear that (17) is an adequate way to think about the evolution of the variables that we commonly include in the set of production function shifters. To take one example, the literature on institutions and economic development has emphasized the relevance of multiple equilibria, path-dependence, and lock-in effects in institutional evolution.[5] It is hard to square the complex dynamics that this literature often attributes to institutional development with the process described by (17). A second problem is that, even if we consider (17) an appropriate mechanism to capture the dynamics of institutional change, it requires thinking about steady states as long-run equilibria in institutions, policies, and economic structure. The time dimension necessary for these variables to reach long-run equilibrium may be much different from what we may consider reasonable for factors of production, and the assumption that countries are generally close to their steady states may consequently be made less appealing. A third problem is that the conventional specification will only hold if $\varphi_i = n+g+\delta$. If $\varphi_i \neq n+g+\delta$, the rate of convergence to the steady state will not be constant but will depend on the distance of the *Z* variables from their steady state

[5] See North (1990), Arthur (1994).

levels. The resulting growth equation will be linear, but it will be very different from the conventional specification as it will depend both on the steady-state levels of Z and on their initial levels. [6]

The discussion in this section has highlighted the fact that the MRW derivation of the linear growth regression does not in general extend to the case in which aggregate productivity depends on a set of production function shifters. The linear specification is only correct when the time path characterizing the evolution of these shifters obeys a set of strong restrictions; otherwise estimation by linear methods will be subject to misspecification bias. We turn to considering the effects of these possible biases.

2.2. Cooking with a Dirty Sink: The Effects of Misspecification Bias under Omitted Non-linearities

2.2.1. Misspecification Bias

Estimating (10) if (12) is the true function will lead to misspecification bias, and is analogous to imposing the invalid restriction that a nonlinearity is not present when it is. Its implications can be easily seen within the framework of omitted variable bias, where we assume for simplicity that (14) holds and that nonlinearities in the growth function come exclusively from a failure of (13).

To see this, rewrite (12) as:

$$\gamma_t = A_0 + A_1 \ln y_0 + A_2 \ln s_k + A_3 \ln h_{ss} + A_4 \ln(n+g+\delta) + C_1 Z_{10} + \ldots + C_n Z_{1n} + h(Z_{10},\ldots Z_{p0}) + \eta_i . \quad (21)$$

Where $h(Z_{10} \ldots Z_{p0}) = \left(f(Z_{10},\ldots,Z_{p0}) - C_1 Z_{10} - \ldots - C_n Z_{p0} \right)$.Estimating (21) by OLS is therefore the same as invalidly imposing the restriction that $h(Z_{10},\ldots,Z_{p0}) = 0$. The limit in probability of the OLS estimators $\hat{C} = \{\hat{C}_1,\ldots \hat{C}_n\}'$ will be:

$$p\lim \hat{C} = C + (Var(Z)^{-1})Cov(Z_0, h(Z_{10},\ldots,Z_{p0}))$$

Even if *f(.)* is independent of lny_o, lns_k, lns_h and ln($n+g+\delta$), all of our estimates of C_i will be inconsistent estimators of $\hat{C} = \{\hat{C}_1,\ldots \hat{C}_n\}'$ unless *h(.)* is independent of *Z* (that is, unless *g(.)* is linear in *Z*). It is impossible to predict the sign of this bias unless we know the sign of the covariance of *Z* with the omitted term. There is thus no reason to believe that our estimated $\hat{C}_i$s will be accurate indicators of the linear effects of changing a variable.

Is there a meaningful interpretation to the linear estimator? Some authors have suggested that the linear estimator provides the average effect of changing the explanatory variable over the sample of countries. For example, Helpman (2004) has argued that “estimates that exploit cross-country variations are best interpreted as average effects of trade policies on growth,” while Temple (2000) writes that “growth

[6] Details of this derivation are available upon request.

regressions are best thought of as picking up an average effect of schooling." If this were true, the linear estimator may not be a poor guide to evaluating the expected effects of changes in policies or institutional and structural reforms.

Regrettably, this conjecture is not correct. The linear estimator in a misspecified non-linear model will usually not converge to the average partial effect. To the best of my knowledge, this point was first made by White (1980), who established that the linear estimator of an arbitrary non-linear function will converge to the *linear approximation* of the function. The linear approximation is the closest linear function in the least-squares norm to the non-linear function. The properties of this approximation are in themselves the subject of a body of mathematical literature known as approximation theory and will generally depend on the distribution of the explanatory variable (see e.g., Rice, 1964 or Rivlin, 1969). In particular, the coefficients of an OLS regression will only converge to the average partial effect of an arbitrary non-linear function when the explanatory variable is normally distributed, as the following proposition establishes.

Proposition 1. Let y be generated by the true model $y_i=f(x_i)+\varepsilon_i$, i=1...n, where $f(x_i)$ is an arbitrary nonlinear measurable function of $x_i \in \Re^1$, and x_i is distributed according to the distribution function H(x) with mean normalized to 0 and variance σ_x^2. Let $E(\varepsilon_i)=0$ and $E(\varepsilon_i^2)=\sigma_\varepsilon^2<\infty, E(x_i\varepsilon_i)=0$, $E(f(x_i)\varepsilon_i)=0$ and $E(f(x_i)^2)=\sigma_f^2<\infty$. Let $\beta=\{\beta_0,\beta_1\}$ be the vector of coefficients from an OLS regression of y on {1,x} Then $\beta_1 \xrightarrow{a.s.} E\left(\frac{\partial f(x)}{\partial x}\right)$ for any function f(x) only if H(x) is the normal distribution.

Proof: See Appendix

We can illustrate this result through a simple example. Suppose we try to estimate the function $f(x)=x^2$ through a linear estimator. The partial derivative is $f'(x)=2x$ and the expected partial derivative will be $E(\partial f(x)/\partial x)=2E(x)=2\mu$. Table 1 shows the results of a basic Monte Carlo simulation in which we have used the quadratic function *f(x)* as the data generating process but have estimated it using OLS. Column (1) shows the results of estimating it when we draw *x* from a normal distribution (thus satisfying the conditions of Proposition 1) while column (2) shows the results of estimating it with a non-normal distribution with exactly the same mean and standard deviation. In particular, column (2) uses the standard log-normal distribution:

$$f(x)=\frac{e^{-((\ln x)^2/2\sigma^2)}}{x\sigma\sqrt{2\pi}}$$

with shape parameter σ=1. Note that the mean and standard deviation of this distribution are, respectively, $\mu_X=e^{.5}\approx 1.65$ and $\sigma^2_X=(e^2-e)^{.5}\approx 4.67$. We also set the mean and standard deviation of the normally distributed variable in column (1) respectively to $e^{.5}$ $(e^2-e)^{.5}$. The first of these assumptions ensures that in both cases the expected partial derivative is $E(\partial f(x)/\partial x)=2\mu\approx 3.30$. As Table 1 shows, in the case of the normal distribution, which is symmetric and mesokurtic, the average slope estimator of the linear regression in 1000 replications is 3.28, very close to the expected partial derivative. In the non-symmetric case, in contrast, the average linear coefficient estimate is 10.64, substantially higher than the expected partial derivative.

Figure 1 provides some intuition as to why non-normality can generate such serious biases. Here we maintain the same comparison as in Table 1 but instead illustrate the behavior of the estimator for a large *N* of 10,000 in both the normal and log-normal case. As the second panel in the Figure shows, the asymmetry of the log-normal distribution makes observations with large values relative to the mean much more frequent than observations with low values relative to the mean. For a given mean (and thus a given mean effect), these outliers have n inordinate effect on the least square estimator as they are heavily penalized, tilting the least squares line upwards relative to the expected marginal effect.

Proposition 1 suggests a strategy of verifying whether the explanatory variables often used as production-function shifters in growth empirics satisfy the conditions of normality. If this is the case, the linear estimates will allow us to recover the average partial effects. Table 3 shows the results of skewness, kurtosis, and joint normality tests of 12 common explanatory variables in growth empirics (variable definitions are presented in Table 2). These are the same variables we will use in our empirical analysis of section 3. According to these standard tests, in all twelve cases we reject the normality hypothesis, and in eleven of them we do so with p-values lower than 0.01. Commonly used explanatory variables in growth empirics appear not to satisfy the conditions established by Proposition 1 for the linear model to yield adequate estimator of the average effects.

The characterization of the misspecification bias arising from ignored non-linearities in (2.21) as a special case of omitted variable bias may lead us to think about dealing with it through the use of instrumental variables. Regrettably, this will generally not be possible. The reason is that is any candidate instruments that is correlated with Z is also likely to be correlated with $h(Z_{10},\ldots,Z_{p0})$. Since the misspecified regression treats *h(.)* as part of the disturbance term, our instrument will be correlated with the residual in the second-stage regression, rendering it invalid. For the same reason, ignored non-linearities will generally make instruments that would be valid in a linear framework yield biased estimates. This result is illustrated in columns (3) and (4) of Table 1, which illustrate the effect of using two stage least squares to estimate a quadratic relationship when the right hand side variable is endogenous **and** we have an exogenous instrument for it. Essentially, the misspecification bias carries over to instrumental variables estimation, producing a coefficient that is similarly biased in relation to the expected partial derivative.

This result is particularly important in light of recent interest in the use of instrumental variables techniques to study the causes of cross-country variations in growth and income. Recent contributions (Frankel and Romer, 1999, Acemoglu, Johnson, and Robinson, 2001) tend to argue for the validity of their instruments based on their exogeneity, their causal relationship with the endogenous variable, and their lack of a direct effect on the dependent variable. However, these exercises generally do not discuss the effect of omitted nonlinearities on the validity of their estimates. To the extent that there exist plausible reasons to believe that the relation between trade, institutions, and development, is characterized by strong nonlinearities, these exercises may yield seriously distorted estimates of the average effects growth effects of their explanatory variables.

2.2.2. Non-linearities and the Curse of Dimensionality

Despite the fact that the possibility of omitted non-linearities can shed doubt of the meaningfulness of linear estimates, the literature has seen relatively few efforts at seriously going beyond the linear model. As mentioned in the introduction, the bulk of existing work examines nonlinearities in a particular dimension by using simple quadratic or multiplicative interactions on a restricted dimension, conditional on the rest of the specification being linear. One of the reasons for this is the widespread perception that adequately dealing with unknown nonlinearities through non-parametric techniques requires a much higher number of observations than is commonly available to growth researchers.

The central result that gives reason for this skepticism is known as the *curse of dimensionality*. As established in the pioneering work of Robert Stone (1980), it puts severe limits on the rate of convergence of a non-parametric estimator to the true non-linear function. If *f(x)* is the true non-linear function, *d* is the dimensionality of *x* and *g(x)* is differentiable up to the *m*-th derivative, then the optimal rate at which a non-parametric estimator can converge to the true regression function is (Stone 1980):

$$\int \left[\hat{f}(x) - f(x) \right]^2 dx = O_P \left(\frac{1}{n^{2m/(2m+d)}} \right). \qquad (22)$$

For example, with m=2 and d=3, then convergence occurs at a rate $O_P(n^{-4/7})$, considerably slower than the parametric $O_P(n^{-1})$. A data set of 100 observations will be as informative under these assumptions as a data set with 14 observations in a linear parametric setting.

A number of remarks are in order regarding this result. In the first place, while this result places strong limitations on the set of inferences that we can make about *f(x)*, it does not offer a justification for the estimation of a misspecified linear regression. Unless we believe that we have enough information about the true nature of the model so as to impose *a priori* restrictions on functional form, equation (22) should be taken not as a measure of the limitations of the methods of nonparametric estimation, but rather as an indicator of the limitations of our data in uncovering the true growth function.

More importantly, (22) places strong limitations on estimating *f(x)* with precision at any particular value of *x*. But attempting to do this would be quite a tall order. While the specific form of *f(x)* near a particular $x=x_o$ can certainly be of interest in the growth context, it is definitely not all that is of interest about *f(.)*. More commonly, we are interested in particular *characteristics* of *f(.)*, such as whether it is approximately linear, monotonically increasing or decreasing, or has a particular "average" behavior. We may be able to make such statements with much more precision than we can make statements about the shape of *f(.)* at any particular point.

This reasoning, in essence, is the basis for our empirical exercise. In what follows, we will use several methods to attempt to discern information as to the shape of *f(x)* without making strong *a priori* assumptions about its functional form. Without negating the limitations placed on us by Stone's result, we proceed to ask the positive question: given existing data, what reasonable inferences can we make with respect to the form of the growth function if we treat it as unknown?

3. Empirical Results

We now turn to our empirical analysis. A logical starting point is to test whether the linearity hypothesis is an adequate characterization of the growth data commonly used in cross-country empirical exercises. Despite the fact that we have argued that the linear specification does not emerge naturally from basic growth theory, it may still be true that it works adequately as a characterization of the world. If this were the case, an argument could be put forward for its use - in the spirit of Friedman (1953) - as an useful if unrealistic simplification of the growth process. As we will see, the evidence in this case speaks very strongly against the hypothesis of linearity, casting doubt on this line of reasoning.

We will then turn to asking whether the evidence is consistent with other characterizations of the growth process. In particular, we will present non-parametric tests of the hypotheses of separability and monotonicity. Separability is relevant because it may allow for attenuation of the curse of dimensionality by the use of additively separable models. Monotonicity tests address the issue of whether one can expect certain policy, institutional or structural reforms to be beneficial or damaging to all countries. We will also provide average derivative estimates of the growth function that do not depend on functional form assumptions. These estimates provide a useful characterization of the expected effect of changes in a production-function shifter, given uncertainty as to the underlying functional form.

Our framework will take as its baseline the partially linear specification derived in section 2.1:

$$\gamma_t = A_0 + A_1 \ln y_0 + A_2 \ln s_k + A_3 \ln h_{ss} + A_4 \ln(n+g+\delta) + f(Z_{10},\ldots Z_{n0}) + \eta_i. \qquad (23)$$

In other words, we will adopt the Mankiw-Romer-Weil specification augmented with a general non-linear term in production-function shifters. Our baseline specification will treat these regressors as fixed, although we will consider the implications of treating them as time-varying in section 3.2.2 This specification will be tested against restricted versions of (12) where *f(.)* is constrained to be linear, separable, or monotonic.[7]

3.1 The Data

Our analysis will use a standard cross-sectional data set of economy-wide measures of growth and its potential determinants for the 1960-03 period. Despite the recent expansion of use of panel-data methods in cross-country growth analysis, I restrict attention to the cross-sectional framework for several reasons. First, the cross-sectional approach is still broadly used and characterizes some of the most relevant

[7] A commonly expressed preoccupation in the literature regards the endogeneity bias that may arise out of the inclusion of endogenous variables such as *n* or *s* in this specification. One partial solution to this problem is to omit some of these variables from the estimated specification (see, e.g., Barro, 1999), thus trading off a reduction in endogeneity bias against an increment in omitted-variable bias. We have repeated all of our empirical exercises using alternative subsets of linear controls. This change has little effect on our results. The working paper version of this paper provides details of those estimations.

recent contributions.[8] Second, relevant methodological questions remain about the applicability of the panel data approach to study questions of long-term economic growth. For example, it is not clear that segmenting the data into ten or five-year intervals is appropriate when the phenomenon of interest is development over long periods of time, and most existing panel methods used require the introduction of fixed effects, impeding the analysis of the effect of potential growth determinants, such as institutions or geography, which exhibit little or no variation over time.[9] Third, the theory behind the specification tests presented in this section is not at this moment fully developed for its application to a panel context.

I use Penn World Tables (PWT) and World-Bank PPP-adjusted per capita GDP Growth Rates from the *World Development Indicators* (WDI) as our dependent variables. WDI data is available for the 1975-03 period, while PWT data is available for the 1960-00 period. Given that a number of explanatory variables are not available for the sixties or early seventies, the former period may be more adequate – for this reason I also present regression results where the PWT data is restricted to the 1975-00 period.

As production-function shifters *Z*, I use combinations of twelve commonly used production function shifters, as well as three summary indicators made up of subgroups of these. The sample attempts to cover the three key dimensions that have played relevant roles in the analysis of growth empirics: policies, institutions and economic structure. To measure policy distortions, I use government consumption as a percent of GDP, the average tax on imports and exports, the log of one plus the inflation rate and the log of the black market premium. To capture the role of institutions, I introduce four commonly used indicators: a measure of the rule of law, a measure of political instability, an index of economic freedom, and an index of the effectiveness of government spending. In the list of structural measures of the level of social development and economic modernization of nations, I use the share of primary exports in GDP in 1975, the rate of urbanization, the ratio of liquid liabilities to GDP, and the average years of life expectancy. I also use three summary indicators of each of these three dimensions, made up by simple normalized averages of the relevant indicators. A full description of the variables and their sources is provided in Table 2.

As is common in the literature, I estimate (23) with a restricted subset of the variables available in the data set in order to economize on degrees of freedom and reduce possible problems of multicollinearity arising from the fact that some of indicators may be capturing what is essentially the same phenomenon. Given that the results can be sensitive to the choice of indicators, in the next two subsections I present estimates for all possible specifications with one policy indicator, one institutional variable and one measure of structural characteristics. In other words, each specification is estimated 125 times and I concentrate on the fraction of specifications for which a given hypothesis is rejected. This approach can be justified

---

[8] Some examples are Frankel and Romer (1999), Acemoglu, Johnson and Robinson (2000), and Sala-i-Martin, Doppelhoffer and Miller (2004). The first two articles use a levels specification, whereas the third uses the growth specification that we reproduce here. For a recent critique of the levels approach, see Sachs (2005).

[9] Standard random effects estimators require the random effect to be uncorrelated with the residual, which is by construction not the case in a growth regression. See Durlauf, Johnson and Temple (2005) for a discussion.

in terms of Bayesian model uncertainty in the spirit of Sala-i-Martin (1997) and Sala-i-Martin, Doppelhoffer, and Miller (2004).

3.2 Linearity

The null hypothesis of linearity can be tested within the framework of traditional OLS estimation. Under the hypothesis that growth is a linear function of its determinants, non-linear terms should not enter significantly into the regression. A simple test of linearity can be carried out by augmenting the linear specification with a series approximation and testing for the joint excludability of the higher-order terms. We estimate

$$\gamma_y = \alpha_0 + \alpha_1 \ln(y_{t-1}) + \alpha_2 \ln h_{t-1} + \alpha_3 \ln(n+g+\delta) + \alpha_4 s_k + \alpha_5 z_p + \alpha_6 z_i + \alpha_7 z_s + p(z_p, z_i, z_s) + \varepsilon_i \quad (24)$$

where z_p, z_i, z_s stand, respectively, for the policy, institutional and structural indicators and $p(z_p, z_i, z_s)$ groups the higher-order components of the series approximation. We test the null hypothesis that all terms in *p(.)* are jointly zero,

A first approach to this issue is presented in Table 4. In it I present the result of estimating (24) through ordinary least squares using 2nd and 3rd order Taylor expansions in (z_p,z_i,z_s) as *p(.)*. The table reports the median F-Statistic and associated P-value for rejection of the null hypothesis that the non-linear terms are jointly zero, as would be implied by (3.1). It also reports the percentage of specifications (out of the 125 regressions generated by alternative combinations of the *z* variables) for which the null hypothesis is rejected.

The results of Table 2 present relatively strong rejections of the linearity hypothesis. Approximately 3/4 of the specifications that use a 2nd order Taylor expansion and more than 90% of the specifications that use a 3rd order Taylor expansion reject the linear specification. Note that the fact that the rejections become stronger as the order of the polynomial is increased is not automatic: even though the R-squared of a regression increases whenever higher order terms (or, for that matter, any additional variables) are included, this is not true for F-tests for joint significance, which can and often do become weaker when new variables are added. This pattern suggests that the higher-order polynomial terms may be playing an important role in explaining growth differences across countries. These rejections remain as strong when we use more general non-parametric methods to approximate possible non-linearities.

It should also be evident that whether or not the assumptions necessary to ensure partial linearity of the growth function are valid is of little relevance for this exercise. Establishing the relevance of the non-linear terms in *f(Z)* is sufficient to establish the lack of validity of the linear specification (10). The significance of any non-linear terms in the rest of the explanatory variables can only enhance, and in no way weaken, the case against the linear specification.

3.2.1 The "True" MRW regression

As we have pointed out in section 2, the linear specification requires both linearity of productivity in the production function shifters (Equation 13) *and* equal growth rates of productivity (Equation 14). We now explore the possibility that the rejections of linearity that we have just derived are due to incorrectly imposing the restriction of equal productivity growth rates. In other words, we now estimate:

$$\gamma_y = \alpha_0 + \alpha_1 \ln(y_{t-1}) + \alpha_2 \ln h_{t-1} + \alpha_3 \ln(n + g_t + \delta) + \alpha_4 s_k + \alpha_5 z_p + \alpha_6 z_i + \alpha_7 z_s + p(z_p, z_i, z_s) + \varepsilon_i \quad (25)$$

where the country-specific growth rate of productivity in both equations is:

$$g_t = \alpha_5 (z_{pt} - z_{pt-1}) + \alpha_6 (z_i - z_{it-1}) + \alpha_7 (z_{st} - z_{st-1}) . \quad (26)$$

We now test *p(.)=0* while allowing for differences across countries in *g(.)*. The restricted form is non-linear but is a *known* parametric functional form. If we had confidence in the fact that the rejections of linearity come from the fact that *p(.)=0* in (25), we could resort to estimating it with parametric nonlinear methods.

Table 5 shows the results of testing *p(.)=0* in (25) for our 125 specifications. We estimate both regressions through nonlinear least squares and show estimates with both a second order and third order Taylor expansion to approximate the nonlinear terms. For the purposes of equation (25), t-1 corresponds to the 1970-1974 period and t corresponds to the 1999-04 period.[10] The results of Table 5 show that relaxing the restriction of equal productivity rates does not help the MRW model. Indeed, rejections of linearity of *A(.)*are even stronger than those of the MRW linear specification.[11] These tests thus strongly support the hypothesis that the failure of linearity has something to do with the failure of the assumption that productivity is linear in the production function shifters and not just the uniform productivity growth assumption.

3.3 Separability

The results of the preceding section tell us that the linearity assumption can be rejected in existing cross-country data sets, but leave us little clue as to the form of the actual non-linearity. In whatfollows, I start out from a very general form of the non-linearity and study the restrictions that can be imposed on it without excessive loss of fit. One key issue in estimation of such a non-linear multi-dimensional function is whether it can be taken to be additively separable. Under additive separability, the rate of convergence of the optimal estimator increases significantly in comparison with the non-separable case, as it collapses to the one-dimensional non-parametric rate of convergence (Hastie and Tibshirani, 1990).

I start out by testing additive separability of *f(Z)* in our partially linear specification. This is done by testing the additively separable specification:

[10] The index of government effectiveness and the economic freedom index are only available for the end of the period, so we set their growth rates to zero. We also set the growth rate of the tariff indicator to zero because the number of countries that have observations for both the initial and final periods is just 29. For rule of law, we use 1980-84 as the initial period and 1995-1999 as the final period. For the black market premium, we use 1995-99 as the final period.

[11] The reason for this is that the restricted regression is not necessarily more flexible than the linear MRW equation. The choice between the non-linear and the linear versions of the MRW model is essentially a choice between two non-nested parametric specifications.

$$\gamma_y = \alpha_0 + \alpha_1 \log(y_{t-1}) + \alpha_2 \ln h_{t-1} + \alpha_3 \ln(n+g+\delta) + \alpha_4 s_k + f_p(z_p) + f_i(z_i) + f_s(z_s)$$

(27)

against the more general model of (24). In order to do this, I carry out three different specification tests that are broadly used in the literature on estimation of non-parametric and semi-parametric methods. These are briefly described in what follows.

The first test consists of a joint F-test for the significance of the interaction terms in a Taylor series approximation. This test uses the same regression used to test for non-linearity in subsection 3.2 but tests the restriction that all coefficients in terms that include multiplicative interactions of z_i variables are zero. The results of the F-test for significance of the interaction terms on the Taylor polynomial specification are shown in the first column of Table 6. Rejection rates for the separability hypothesis at the 5% level of significance oscillate between 50.4% and 76.0 %, with the lowest value corresponding to the 1960-00 PWT data and the highest to the World Bank data.

The basic problem with Taylor polynomial tests of separability is that they may lead to overrejection of the null by not taking account of the full level of potential complexity of the $f_i(.)$ functions. By restricting these functions to be third (or n-) order polynomials, we may end up attributing to the separable interaction terms part of the variation that actually arises from the complex non-linearities in each of the $f_i(.)$ functions. The next two tests address this issue in different ways.

The second test of separability that I present consists in estimation of $f(Z)$ by a flexible Fourier series approximation. This is essentially a polynomial expansion in quadratic and trigonometric terms. There is an extensive econometric literature studying the properties of these estimators (Gallant, 1982, Geman and Huang, 1982 and Gallant, 1987). The basic benefit of a Fourier approximation is the greater flexibility of the trigonometric expansion to approximate highly non-linear functions. Formally, estimation proceeds by estimating:

$$\gamma_y = \alpha X + u_0 + \sum_{i=1}^{3} b_i z_i + \sum_{i=1}^{3}\sum_{j=1}^{3} c_{ij} z_i z_j + \sum_{i=1}^{3}\sum_{j=1}^{J_n} \left\{ u_{ij} \cos(jk'_i z) + v_{ij} \sin(jk'_i z) \right\},$$

(28)

where I have written the parametric part of the equation compactly as αX. The k'_i are known as *multi-indices* and are vectors whose elements are integers with absolute values summing to a number k^* less than a pre-specified value K^*. Given a value of K^* and J, the parameter vector $\beta_u = \{u_0, b_1 ... b_3, c_{11} ... c_{33}, u_{11} ... u_{3J_n}, v_{11} ... v_{3J_n}\}$ can be estimated by ordinary least squares. The choices of K^* and J are given and are a somewhat arbitrary feature of estimation. In principle the total number of terms in the expansion is supposed to grow with sample size but knowing this is not terribly helpful since it only gives us an order of magnitude and not a specific number of observations. In practice, many authors tend to look to the "saturation ratio", the ratio of the total number of terms in the expansion M to the number of observations N. In practice, saturation ratios between .25-.40 are typical of the applied literature (see Chalfant and Gallant, 1985 and Pagan and Ullah, 2004). We can obtain a restricted estimator β_r by restricting the coefficients on the terms involving interactions between different z

variables to equal zero. Let e_u and e_r denote respectively the residuals from the restricted and unrestricted estimation. Hong and White (1995) have established that under the null hypothesis that the restrictions are valid:

$$HW = \frac{\frac{1}{N}\frac{\sum_{i=1}^{N} e^2{}_{ri} - \sum_{i=1}^{N} e^2{}_{ui}}{\sum_{i=1}^{N} e^2{}_{ui}} - M}{(2M)^{1/2}} \rightarrow N(0,1), \tag{29}$$

where M is the number of terms in the Fourier expansion.

The third separability test is based on analysis of the residuals derived from direct estimation of the additively separable specification (27) by penalized spline estimation[12]. The *residual regression test* (Fan and Li, 1996) consists in estimating a non-parametric function of the residuals from the restricted estimation on the explanatory variables z_i. Under the null hypothesis, these variables should have no explanatory power in the auxiliary regression. Formally, we calculate the U- statistic:

$$U = \frac{1}{\lambda^{d/2} n}\sum_{i}\sum_{j\neq i}\left(y_i - \hat{f}_r(x_i)\right)\left(y_j - \hat{f}_r(x_j)\right)\prod_{k=1}^{3} K\left(\frac{x_{jk} - x_{ik}}{\lambda}\right) \text{`} \tag{30}$$

, which, under the null hypothesis of separability, and as long as the restricted estimator converges sufficiently rapidly, is normally distributed with mean zero and variance $2\sigma^4 \int p^2(x) \int K^2(u)$ [13].

Since our data sets have a relatively small number of observations (between 62 and 96, depending on the precise specification), asymptotic standard errors may lead to erroneous inferences. We therefore construct bootstrapped test statistics based on residual sampling with 100 observations per specification.[14] For the residual regression tests, the bandwidth for the additively separable estimation is set by Generalized Cross-Validation. The bandwidth for the test statistic is set to $\lambda = n^{-1/5}$, thus ensuring that the higher order components of the U-statistic in (30) converge to zero (see Yatchew, 2003, p. 118). However, the results are very similar under alternative bandwidth choices for the test statistic.

The last two columns of Table 6 show the median test statistic, p-values, and number of rejections of the null of separability using respectively the Hong-White test and residual regression test. For the Hong-White test we use $J=1$ and $K^*=2$, which gives us M=28 and a saturation ratio which varies between .29 and .45. The Hong-White test shows a percentage of rejections of separability are in the range of 15.2-31.2%. For the residual regression test, the result point also to low rejection rates, in the range of 12.0-16.8%

[12] This is carried out using the *gam* command in **R** .Bandwidth is chosen by Generalized Cross-Validation.

[13] In practice, the variance term is estimated by

$$U = \frac{1}{\lambda^{2d} n^4}\sum_{i}\sum_{j\neq i}\left(y_i - \hat{f}_r(x_i)\right)^2\left(y_j - \hat{f}_r(x_j)\right)^2\prod_{k=1}^{3} K^2\left(\frac{x_{jk} - x_{ik}}{\lambda}\right).$$

[14] Li and Wang (1998) have found the bootstrap approximation to be superior to the asymptotic one for the residual regression tests.

The three proposed tests of separability give contrasting results. While the Taylor polynomial tests allow us to reject separability in a preponderance of the specifications, Fourier and residual regression tests, which give more flexibility to the separable function, indicate that most specifications are consistent with separability. Even in these cases, the number of specifications in which separability is rejected can be as high as one-third. Furthermore, it is important to note that even though we have addressed small-sample biases through bootstrapping of confidence intervals, these tests may differ in their power to reject the alternative. To a certain extent, this problem is heightened by the curse of dimensionality: in high dimensions and with limited information, one is likely to be able to fit *many* functional forms to the data, including separable and non-separable specifications. The null hypothesis of separability may be difficult to reject not because the world looks particularly separable, but rather because sparsity of data allows the world to be consistent with many views, among which separability is one.

The reading that one gives to these results will determine the approach that one takes to further study of the data. If one takes the evidence as supportive of the additively separable specification, then one should concentrate on the estimates derived from the additively separable estimation in order to understand where the important non-linearities are concentrated. In contrast, if one reads the above results as indicative of non-separabilities, either because the estimators do reject separability in a non-negligible fraction of specifications or because failure to reject it may itself be a consequence of the curse of dimensionality, one would be interested in knowing what one can learn from the growth data without making the separability assumption. We explore each of these approaches in turn.

3.4 Additively Separable Specification

The results of our additively separable estimates are presented in Table 7. For the purposes of this and the next sub-section, we will concentrate on the analysis of a small subset of regressions. This is because we take these sections to be concerned not so much with the robustness of the underlying results (as was the case for the tests of the previous two sections) but rather with the illustration of different possible approaches to the problem of estimating unknown non-linear specifications. In particular, we will discuss the estimate of one regression for each potential production function shifter, in which that variable is paired with the combined indices for the other two dimensions. Thus, for example, the inflation variable is paired with the combined institutions index as well as the combined structure index.

Table 7 presents a comparison of the OLS and additively separable estimates. It is interesting to note that in six cases the additive estimator selects the linear function as the appropriate one (inflation, government consumption, political instability, effectiveness of government spending, financial development, and the combined policy index). It is interesting to note that in six cases the additive estimator selects the linear function as the appropriate one (inflation, government consumption, political instability, effectiveness of government spending, financial development, and the combined policy index). A few others, however, indicate substantial nonlinearities. A few others, however, indicate substantial nonlinearities. Table 7 also shows two useful indicators of the difference between the linear and the additively separable specification. One is the average derivative of the function, as it compares to the linear slope coefficient. In only one of the fifteen cases does the weighted average derivative obtained from the additively separable estimates have a different slope then the OLS estimate, suggesting that the results are broadly similar (in the remaining case, tariffs, the coefficient was not significant in OLS estimation). However, a number of variables see their statistical significance altered: four variables (inflation, government consumption, political instability, and urbanization) lose statistical significance going from OLS to the additively separable estimates, while one (rule of law) now becomes significant. This pattern resembles what we will find in some of the more general specifications presented below, in that they suggest a lesser role for policy variables and a greater role for structural factors than the linear model. The other way in which we can see the differences between both approaches is by looking at the fraction of the sample that will share the same derivative as the linear estimate. For a number of variables, this number is not high. For example, despite the fact that both the OLS and the additively separable derivative estimates for the effect of the index of economic freedom are positive, significant at 1% and similar in magnitude, the additively separable estimation also predicts that 22.5% of countries in the sample would suffer *deteriorations* from growth as a result of increases in their economic freedom. The result is even worse for some of the structural variables such as the urbanization rate and life expectancy, for which even though the slope estimators are positive, respectively 69.2% and 54.4% of countries would see a decline in growth associated with increases in these variables.

3.5 Monotonicity Tests

A logical counterpart to parametric significance tests would be to test whether $f(.)$ is a monotonically increasing (or decreasing) function of its arguments. In other words, we would try to ask of the data the following question: is there evidence that if a country were to carry out policy reform A, we could always expect its growth rate to rise or at the least not to fall with that policy reform? Formally, this is implemented by testing the null hypothesis:

$$H_0 : z_i > z_i', z_{-i} = z_{-i}' \rightarrow f(z_1,...z_m) \geq f(z'_1,...z'_m) \qquad (31)$$

against the alternative:

$$H_1 : \exists (z_1,...z_m),(z'_1,...z_m') \mid z_i > z_i', z_{-i} = z_{-i}', f(z_1,...z_m) < f(z'_1,...z'_m)\}. \qquad (32)$$

We can use the same framework for testing as in the previous subsection by imposing monotonicity as a restriction on the estimated $\hat{f}(.)$ and calculating the HW and U statistics. The only technical issue has to do with the calculation of the restricted estimator. In the case of the Fourier series expansion, we can explicitly calculate the derivative of the series and directly impose the restriction that

$$\frac{\partial \hat{f}}{\partial z_i}(.) \geq 0 \qquad (33)$$

at all observed values of z_i. The sum of squared residuals is minimized subject to (33) to obtain the restricted estimator $\hat{f}_r(.)$ and calculate the HW statistic. This is a non-linear optimization problem subject to an inequality restriction that can be solved numerically. For the case of the residual regression test, however, the issue is a bit more complex. The continuity of the first derivatives of the Fourier representation ensure that imposing (31) on all observations will deliver a reasonably smooth function. No such parametric representation exists for the restricted penalized spline estimator. In order to ensure that the restricted function is reasonably smooth, I divide the $Z_p \times Z_i \times Z_s$ space into one-thousand cubes of volume 0.1^3=0.001, constraining $\hat{f}_r(.)$ to be monotonic between any two locally adjoining cubes as well as between any observed value of z_i and the edges of each cube. This still requires estimation subject to more than six thousand constraints. In practice, however, the actual number of binding constraints is much smaller, making estimation computationally feasible.[15]

Table 8 displays the results of the Fourier expansion and residual regression tests. We have divided our results in tests of what we call the "conventional wisdom" and "anti-conventional wisdom" hypotheses. The conventional wisdom hypothesis reflects the common accepted wisdom about the effect of the candidate variables on growth derived from previous literature. In this case, government distortions are detrimental for growth, while certain institutions and structural changes are beneficial. The anti-conventional wisdom hypothesis reflect the *exactly* opposite view: that distortions are good for growth and market-preserving institutions and structural modernization are bad for growth.

The Fourier monotonicity tests are able to comfortably reject in most cases the anti-conventional view for the institutional and structural variables. Interestingly, they are **not** able to reject it for four out of the five policy variables. Since for these variables it is also not able to reject the conventional-wisdom hypothesis, then the tests are completely uninformative about the effect of these variables on growth.

Such is not the case, however, with most of the variables in the data set. For three of the institutional variables as well as all the structural variables, the data rejects the anti-conventional wisdom view but not the conventional wisdom view. Therefore, the estimates suggest that at least for some countries it is true that increases in these variables will lead to increases in their well-being. Interestingly, there are three variables in which this test rejects both the conventional and the anti-conventional wisdom views. The variables are tariffs, the index of economic freedom, and the effectiveness of government spending. For these variables, the data is inconsistent with the idea that the function is monotonic (which includes the possibility of it being irrelevant, as we test weak monotonicity). In other words, the data tells us that there are some countries that will benefit, and others that will lose, from increases in these variables.

---

[15] All problems were solved using the CONOPT solver in GAMS; the code is available from the author upon request. For the Fourier series expansion, the sum of squared residuals was minimized subject to the explicit constraint that the analytic first-derivative have the prescribed sign. For the residual regression and differencing tests, we first obtain a non-parametric estimate of $\hat{f}(.)$ by penalized thin-plate regression spline estimation using the **gam** command in R; we then calculate the restricted estimate by finding the closest set of points to the fitted function that satisfy the monotonicity constraints in GAMS. See Mammem (1991) for a discussion of this method for construction of monotonic estimates.

The residual regression tests, in contrast, give a view which is much closer to the conventional wisdom. All fifteen tests reject the anti-conventional wisdom view, while in only one case (tariffs) is the conventional wisdom view rejected. Interestingly, in the case of tariffs both the Fourier series and residual regression tests support the hypothesis that this variable has a complex effect on growth which is positive for some countries and negative in others.

3.4 Weighted Average Derivative Estimators

An alternative approach that does not require parametric assumptions is to estimate the weighted average derivatives using the estimation methods suggested by Hardle and Stoker (1989) and Rilstone (1991). These authors have shown that it is possible to derive weighted average derivative estimators (WADE) without making assumptions about the functional form of the estimated function. More importantly, these estimates are root-n consistent and are thus not subject to the curse of dimensionality. What is important to bear in mind is that these estimates, while consistent, are not substitutes of the linear estimator as they only apply to the average over countries. Nevertheless, they can give us a good idea of what the shape of the function is without making arbitrary functional assumptions.

Table 9 shows the results of these regressions. Significance tests are again based on bootstrapped standard errors with 100 observations per equation. A number of facts are interesting. First, a number of variables lose significance when one goes from OLS to the WADE estimators. The average slope estimate on the policy variables falls by 83% and that on the institutional variables declines by 42%, while that on the structural variables increases by 31%. None of the policy variables have average derivatives significantly different from zero, while four institutional and five structural variables do. The corresponding numbers for the OLS estimates are three policy variables, four institutional variables, and five structural variables. The conclusions of the WADE estimators thus deliver a markedly different emphasis than that which arises out of the linear framework, suggesting a greater role for structural variables and a smaller role for policy variables. Note that these results are very similar to those of the Fourier tests in Table 8, which also showed considerably greater uncertainty about the role of the policy variables.

Statistical significance tests on WADE estimators must be interpreted with caution. In a non-parametric setting, the average slope may be positive while the slope faced by an individual policymaker is negative or zero. A positive significant average derivative estimator should not be taken to imply that policymakers can be certain that they will face a local positive derivative. The uncertainty faced by the policymaker is a combination of two types of uncertainties: uncertainty about the average slope of the function, and uncertainty about the local slope that corresponds to his position on a function with a given average slope. The possibility of calculating average derivative estimators should not lead us to overestimate the knowledge that we have about the effects of particular reforms in specific country settings.

4. Concluding Comments

This paper has explored the implications of the linearity assumption for growth empirics. It has argued that the theoretical basis for the linear kitchen-sink growth regression is tenuous, that there are considerable risks from misspecification bias that come from mistakenly imposing such a specification, and that the data strongly support the hypothesis that a linear specification is not valid. I have explored several alternatives to linear analysis in growth empirics, among them additively separable estimation, monotonicity restrictions, and weighted average derivative estimation. These methods serve to deliver valuable conclusions about the growth data that are consistent with less restrictive specifications than the linear one. For example, monotonicity tests suggest that institutional and structural reforms can be beneficial for at least a number of countries, and that an important subset of variables – such as trade policy – have complex effects in which some countries benefit but others lose from moving them in a given direction.

Nevertheless, there are restrictions on the type of conclusions that can be drawn. This is a reflection of the limited amount of data relative to the dimensionality problem. It is one thing to try to distinguish between the hypothesis that a certain policy is equally good for all countries and the hypothesis that it is equally bad (or equally irrelevant) for all countries; it is quite another to try to distinguish among a broad set of potential hypotheses that allow for complex interactions between the policy and a host of country-specific characteristics such as its primary export dependence and the effectiveness of its government spending. In order to do the former one may be able to get away with using a small number of observations; this is much harder to do if one is attempting the latter.

The problem is that if the latter specification is the better reflection of reality, attempting to use the former is likely to lead to results that are at best misleading and at worse meaningless. But the conclusion to be drawn from this inherent complexity is not that the empirical analysis of growth data sets is a worthless endeavor. Indeed, we have shown that it is possible to reach conclusions that place significant restrictions on the shape of the growth function without imposing insensible restrictions. Our analysis does suggest that there is considerable complexity in the growth function, but it also finds that the conventional wisdom view in many cases is closer to reality than its polar opposite. They also provide evidence in favor of a different emphasis than that of the traditional literature, with a greater role for structural and institutional variables and less of an emphasis on policy variables in the design of growth strategies. More generally, it suggests that the focus of the empirical growth literature may be better served by analyzing general, "deeper" hypotheses about the growth process than in trying to reach an exact characterization of its form.

References

Acemoglu, Daron, Simon Johnson and James A. Robinson "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369-1401.

Aghion, Phillipe and Peter Howitt (1998) *Endogenous Growth Theory*. Cambridge, US: MIT Press.

Aghion, Philipe and Steven Durlauf, eds (forthcoming) *Handbook of Economic Growth*. Amsterdam: North Holland.

Arhur, Brian (1994) *Increasing Returns and Path Dependence in the Economy*, Ann Arbor: University of Michigan Press.

Azariades, C. and A. Drazen (1990), 'Threshold externalities in economic development', Quarterly Journal of Economics, CV, 501-526.

Banerjee, Abhijit and Esther Duflo (2003) "Inequality and Growth: What Can the Data Say?" *Journal of Economic Growth* 8:267-99.

Barro, Robert (1991) "Economic Growth in a Cross-Section of Countries," *Quarterly Journal of Economics*, 106(2): 407-443.

Barro, Robert (1996) "Democracy and Growth," *Journal of Economic Growth* 1:1-27.

Barro, Robert (1999) *Determinats of Economic Growth*. Cambridge, US: MIT Press.

Barro, Robert and Xavier Sala-i-Martin (2004) *Economic Growth*. 2nd Edition. Cambridge: MIT Press.

Bernanke, B. S. and Gürkaynak, R. S. (2002), "Is Growth Exogenous? Taking Mankiw, Romer, and Weil Seriously", in B. S. Bernanke and K. Rogoff (eds.) NBER MacroeconomicsAnnual 2001 (Cambridge, MA: MIT Press), 11-57.

Bhagwati, Jagdisjh and T.N. Srinivasan (2000) "Outward Orientation and Development: Are the Revisionists Right?" Discussion Paper No. 806, Economic Growth Center, Yale University.

Chalfant, J. A. and A. R. Gallant (1985) "Estimating Substitution Elasticities with the Fourier Cost Function: Some Monte-Carlo Results," *Journal of Econometrics* 28: 205-222.

Chang, Roberto, Linda Kaltani and Norman Loayza (2005) **"**Openness Can be Good for Growth: The Role of Policy Complementarities" *World Bank Policy Research Working Paper 3763*.

DeGregorio, José and Jhong-Wha Lee (2004) "Growth and Adjustment in East Asia and Latin America," reproduced, Banco Central de Chile.

DeJong, David and Marla Ripoll "Tariffs and Growth: An Empirical Exploration of Coningent Relationships." *The Review of Economics and Statistics,* November 2006, 88(4): 625–640

Dowrick, S. (2004). De-Linearising the Neo-Classical Convergence Model. Recent Developments in Economic Theory. S. Turnovsky, S. Dowrick and S. Pitchford. New York, Cambridge University Press.

Durlauf, S. N. and P. A. Johnson (1995) "Multiple Regimes and Cross-Country Growth Behavior," *Journal of Applied Econometrics* 10:365-84.

Durlauf, S., P. Johnson, and J. Temple, (2005), "Growth Econometrics," in *Handbook of Economic Growth*, P. Aghion and S. Durlauf eds., Amsterdam: North Holland.

Durlauf, S., Kourtellos, A., and Minkin, A. (2001), "The Local Solow Growth Model", European Economic Review, 15, 928-940.

Einstein, Albert (1933) *On the Method of Theoretical Physics*. New York: Oxford University Press.

Fan, Y. and Q. Li (1996) "Consistent Model Specification Tests: Omitted Variables and Semi-parametrtic Functional Forms." *Econometrica* 64:865-890.

Fan, J., Härdle, W. and Mammen, E. (1998). "Direct estimation of low dimensional components in additive models," *Annals of Statistics* 26: 943-971.

Frankel, Jeffrey and David Romer (1999) "Does Trade Cause Growth?" *American Economic Review* 89(3): 379-99.

Friedman, Milton. 1953. "The Methodology of Positive Economics." in Essays in Positive Economics, edited by Milton Friedman. Chicago: University of Chicago Press.

Gollin, Douglas (2002) "Getting Income Shares Right," *Journal of Political Economy* 90: 458-474.

Gallant, A. R. (1985) "Unbiased Determination of Production Technologies," *Journal of Econometrics* 20: 285-323.

Gozalo,P. (1995) "Nonparametric specification Testing with $\sqrt{n}$-Local Power and Bootstrap Critical Values," Working Paper No. 95-21-R, Brown University.

Hardle, W. and T. Stoker (1989) "Investigating Smooth Multiple Regressino by the Method of Average Derivatives," *Journal of the American Statistical Association,* 84 (408): 986-995.

Hastie, T. and R. Tibshirani (1990) *Generalized Additive Models*. London: Chapman and Hall.

Hong, Y. and H. White (1995) “Consistent Specification Testing via Nonparametric Series Regression,”*Econometrica* 63: 1133-1160.

Hausmann, Ricardo, Dani Rodrik and Andrés Velasco (2004) ”Growth Diagnostics.” Reproduced, Harvard University.

Hausmann, Ricardo and Dani Rodrik (2005) “"Self Discovery in a Development Strategy for El Salvador." Reproduced, Harvard University.

Helpman, Elhanan (2004) *The Mystery of Economic Growth*. Cambridge, US: Harvard University Press.

Heston, Alan, Robert Summers and Bettina Aten (2002), Penn World Tables Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October.

Kalatzidakis, Pantelis, Theofanis P. Mamuneas, Andreas Savvides and Thanasis Stengos (1999) “Measures of Human Capital and Nonlinearities in Economic Growth,” *Journal of Economic Growth 6(3): 229-254***.**

Kourtellos, Andros (2003) “Modeling Parameter Heterogeneity in Cross-Country Growth Regression Models” Reproduced: University of Cyprus.

Levine, Ross and David Renelt (1992) “A Sensitivity Analysis of Cross-Country Growth Regressions,” *American Economic Review* 82(4):942-963.

Li, Q. and S. Wang (1998) “A Simple Consistent Bootstrap Test for a Parametric Regression Function,”*Journal of Econometrics* 87: 145-165.

Lipsey, R. G. and Kelvin Lancaster, “The General Theory of Second Best”, *The Review of Economic Studies*, Vol. 24, No. 1. (1956 - 1957), pp. 11-32.

Liu, Zhenjuan and Thanasis Tsengos (1999) “Non-linearities in Cross-Country Growth Regressions: A Semiparametric Approach.” *Journal of Applied Econometrics* 14: 527-538.

Maloney, William (2007) “Missed Opportunities: Innovation and Resource-Based Growth in Latin America” in Daniel Lederman and William F. Maloney, editors, *Natural Resources: Neither Curse nor Destiny*. Stanford, CA: Stanford University Press.

Mammem, E. (1991) “Estimating a Smooth-Monotone Regression Function,”*Annals of Statistics*, 19:724-740.

Mankiw, N. Gregory, David Romer and David N. Weil (1992) "A Contribution to the Empirics of Economic Growth." *The Quarterly Journal of Economics*, 107(2): 407-437.

Muphy, K., A. Shleifer and R. Vishny (1989) "Industrialization and the Big Push",*Journal of Political Economy*, October, 1989.

North, Douglass C. (1990) Institutions, Institutional Change and Economic Performance. Cambridge University Press, New York.

Ocampo, José Antonio (2004) "Latin America's Growth and Equity Frustrations During Structural Reforms." *Journal of Economic Perspectives* 18(2), pp. 67-88.

Rodríguez, Francisco and Daniel Ortega (2006) "Are capital shares higher in poor countries? Evidence from Industrial Surveys." Wesleyan Economics Working Paper 2006-023

Pagan, Adrian and Aman Ullah (1999) *Nonparametric Econometrics*. Cambridge, UK: Cambridge University Press.

Qian, Yingyi, (2003) "How Reform Worked in China," in D. Rodrik, ed., In Search of Prosperity: Analytic Narratives of Economic Growth, Princeton, NJ, Princeton University Press.

Ramsey, Frank (1928) "A Mathematical Theory of Saving,"*Economic Journal,* 38: 543-559.

Rice, J. R. (1964) *The Approximation of Functions*. Reading, Mass.: Addison-Wesley Pub. Co.

Rilstone, Paul (1991) "Nonparametric Hypothesis Testing with Parametric Rates of Convergence,"*International Economic Review* 32(1): 209-227.

Rivlin, Theodore (1969) "An Introduction to the Approximation of Functions." Waltham, Mass., Blaisdell Pub. Co.

Rohatgi, V. K. (1976) *An Introduction to Probability Theory and Mathematical Statistics*. New York: John Wiley and Sons.

Sachs, Jeffrey D. and Wing Thye Woo (1994) "Structural Factors in the Economic Reforms of China, Eastern Europe, and the Former Soviet Union," Economic Policy April 1994.

Sachs, J. (2005). "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income." NBER Working Paper 9490.

Sala-i-Martin, Xavier. “I Just Ran 2 Million Regressions.”*American Economic Review*, May (*Papers and Proceedings*), *87*(2), pp.178–83.

Sala-i-Martin, Xavier, Gernot Doppelhoffer and Ronald Miller (2004) “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach” *American Economic Review* September, 812-835.

Sperlich, S. and Jiri Zelinka (2003) “Generalized Additive Models” Reproduced: MDTech.

Stone, Robert (1980) “Optimal Rates of Convergence for Non-Parametric Estimators.”*Annals of Statistics* 8: 1348-1360.

Stoyanov, Jordan (2000) “Krein condition in probabilistic moment problems” *Bernoulli* 6(5):939-949.

Temple, Jonathan (1999) ‘The New Growth Evidence,”*Journal of Economic Literature* 37(1):112-56.

Temple, Jonathan (2000) “Growth Effects of Education and Social Capital in the OECD Countries,” OECD Economics Department Working Paper No. 263, Organization for Economic Co-operation and Development.

Yatchew, Adonis (1988) “Some Tests of Nonparametric Regression Models,” in W. Barnett, H. White and E. Berndt, (eds.) *Dynamic Econometric Modeling: Proceedings of the Third International Symposium in Economic Theory and Econometrics*. Cambridge, UK: Cambridge University Press, pp. 121-135.

Yatchew, Adonis (2003) *Semiparametric Regression for the Applied Econometrician*. Cambridge, UK: Cambridge University Press.

White, Halbert, (1980) "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, vol. 21(1), pages 149-70, February.

World Bank (2005) *World Development Indicators*. Electronic Database. Washington, DC: The World Bank.

World Bank (2005) *Economic Growth in the 1990s: Learning from a Decade of Reform*. Washington: The World Bank.

Appendix: Proof of Proposition 1

Proposition 1. Let *y* be generated by the true model $y_i = f(x_i) + \varepsilon_i, i = 1...n$, where *f(x$_i$)* is an arbitrary nonlinear measurable function of $x_i \in R$ and x_i is distributed according to the distribution function H(x) with mean normalized to 0 and variance $\sigma_x^{\ 2}$. Let $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = \sigma_\varepsilon^{\ 2} < \infty$, $E(x_i\, \varepsilon_i) = 0$, $E(f(x_i)\, \varepsilon_i) = 0$ and $E(f(x_i)^2) = \sigma_f^{\ 2} < \infty$. Let β ={$\beta_0$, β_1} be the vector of coefficients from an OLS regression of y on {1,x}. Then $\beta_1 \overset{a.s.}{\to} E\left(\frac{\partial f(x)}{\partial x}\right)$ for any function f(x) only if H(x) is the normal distribution:

Proof. By White (1980, Theorem2), $\beta \overset{a.s.}{\to} \beta^*$, the vector which uniquely solves $\underset{\beta}{Min} \int [f(x) - x\beta]^2 dH(x)$ .This vector is the coefficient vector on (1, x) in the linear projection of f(x) on (1, x). Particularly,

$$\beta_1^{\ *} = \frac{Cov(x, f(x))}{Var(x)} = \frac{E(xf(x))}{\sigma_x^{\ 2}}$$

where we have used the normalization E(x) = 0. Now let us approximate f(x) by an n-th order Taylor expansion at x = 0, that is:

$$f(x) \approx P(f(x)) \equiv \alpha_0 + \sum_{k=1}^{m} \alpha_n x^n$$

where:

$$\alpha_0 = f(0)$$

$$\alpha_n = \frac{\partial^n}{\partial x^n} f(0).$$

Since the linear projection is a linear operator (see Wooldridge, p.32) then:

$$\beta_1^{\ *} = \sum_{i=1}^{n} \alpha_n \frac{E(x \cdot x^n)}{\sigma_x^{\ 2}}.$$

Therefore $\beta_1^{\ *}$ will equal $E\left(\frac{\partial f(x)}{\partial x}\right)$ for any α_n only if:

$$\beta_1^{\ *} = \frac{1}{\sigma_x^{\ 2}} E(x^{n+1}) = nE(x^{n-1}).$$

Since E(x) = 0, (2) implies $E(x^q) = 0$ for any q > 1 odd. For q ≥ 2 even, repeated substitution gives $E(z^q{}_d) = \sigma_{dd}^{\ q/2}(q-1)(q-3)...3 \cdot 1.$ These are exactly the moments of the normal distribution (see, e.g., Rohatgi, 1976 pp.220-221). Since the normal distribution is M-determinate (Stoyanov, 2000), then H(x) is the normal distribution.

Table 1: Monte Carlo Simulation, OLS and IV estimation of quadratic function with alternative distributional assumptions

	1	2	3	4
Data Generating Process	$y=x^2+\varepsilon$		$y=x^2+\varepsilon+u$ $x=z+u$	
Estimated Equation	$y=a_0+a_1x+e$		$y=a_0+a_1x+e$ $x=b_0+b_1z$	
Method of Estimation	OLS	OLS	IV	IV
Distribution of x	Normal	Log-Normal	Normal	Log-Normal
Mean of x	1.649	1.649	1.649	1.649
Variance of x	2.161	2.161	2.381	2.381
E(dy/dx)	3.297	3.297	3.297	3.297
Replications	1000	1000	1000	1000
$\hat{a}_1$	3.279	10.636	3.274	10.386

Table 2: Data Sources	
Policy Indicators	
1. Trade Policy Openness	$(1+t_m)(1+t_e)$-1, with t_m (t_e) the ratio of import (export) tax revenue in total imports (exports);
2. Log of Black Market Premium	Dollar and Kraay (2002)
3. Government Consumption as a Percentage of GDP	World Bank (2004)
4. Log of (1+Inflation Rate)	World Bank (2004)
5.Summary Policy Indicator	Sum of 1-4, normalized over the unit interval
Institutional Indicators	
6. Rule of Law	Dollar and Kraay (2002)
7. Political Instability	Average Variation in POLITY variable, Polity IV
8. Effectiveness of Government Spending	Glaeser et al. (2004)
9. Economic Freedom Index	Heritage Foundation
10. Summary Institutions Indicator	Sum of 6-9, normalized over the unit interval
Economic Structure Indicators	
11. Share of Primary Exports in Total Exports	World Bank (2004)
12. Urbanization Rate	World Bank (2004)
13. Share of liquid liabilities in GDP	International Monetary Fund (2004)
14, Life Expectancy	World Bank (2004)
15. Summary Structure Indicator	Sum of 10-14, normalized over the unit interval

Table 3: Skewness, Kurtosis and Normality Tests of some Common Explanatory Variables

Explanatory Variable	Number of observations	Skewness p-value	Kurtosis p-value	Joint Normality Test χ^2	Joint Normality Test p-value
Log(1+Inflation)	137	0.00	0.00	129.95	0.00
Black Market Premium	132	0.00	0.00	133.65	0.00
Government Consumption	133	0.00	0.32	11.43	0.00
Tariff Rate	150	0.00	0.00	137.14	0.00
Rule of Law	112	0.06	0.01	11.58	0.00
Political Instability	121	0.00	0.00	103.43	0.00
Economic Freedom Index	137	0.00	0.01	25.06	0.00
Index of government Effectiveness	133	0.02	0.28	6.51	0.04
Primary Exports in 1970	112	0.00	0.47	14.95	0.00
Urbanization Rate	136	0.60	0.00	13.66	0.00
Life Expectancy	137	0.02	0.00	48.79	0.00
Liquid Liabilities/GDP	126	0.00	0.00	118.91	0.00

All values refer to the D'agostino, Balanger, and D'Agostino, Jr. (1990) tests for skewness, kurtosis and normality

Table 4: Linearity Tests, 2nd and 3rd Order Polynomial Expansions

Data Set	Penn World Tables, 1975-00	World Bank, 1975-03	Penn World Tables, 1960-00
Second-Order Polynomial			
Median F-Statistic	3.80	2.575	3.581121
Median P-Value	0.00	0.026	0.0042164
Number significant (/125)	98	75	104
Percent Significant	78.4%	60.0%	83.2%
Third-Order Polynomial			
Median F-Statistic	6.25	4.97	5.514457
Median P-Value	0.00	0.00	2.02E-06
Number significant (/125)	116	119	121
Percent Significant	92.8%	95.2%	96.8%

Reported results refer to conventional F-test of the null hypothesis that all coefficients on non-linear terms in a polynomial expansion are equal to zero. Number and percent significant are calculated using a significance level of 5%.

Table 5: Mankiw-Romer-Weil Nonlinear Specification, Polynomial Tests, Full MRW Specification

Data Set	**Penn World Tables, 1975-00**	**World Bank, 1975-03**	**Penn World Tables, 1960-00**
Second-Order Taylor Expansion			
Median F-Statistic	4.51	2.86	3.71
Median P-Value	0.00	0.02	0.00
Number significant (/125)	107	85	91
Percent Significant	85.6%	68.0%	72.8%
Third-Order Taylor Expansion			
Median F-Statistic	12.80	8.13	6.21
Median P-Value	0.00	0.00	0.00
Number significant (/125)	122	124	120
Percent Significant	97.6%	99.2%	96.0%

Estimators obtained through Nonlinear Least Squares.Reported results refer to conventional F-test of the null hypothesis that all coefficients on non-linear terms in a polynomial expansion are equal to zero. Number and percent significant are calculated using a significance level of 5%.

Table 6: Separability Tests, Series and Non-Parametric Estimators

	Taylor Polynomial	Fourier Series	Residual Regression
Penn World Tables, 1975-00			
Median Statistic	2.46216	1.102	-1.099
Number significant (/125)	74	36	16
Percent Significant (5%)	59.20%	28.80%	12.80%
World Bank, 1975-03			
Median Statistic	3.131997	1.259	-1.095
Number significant (/125)	95	39	21
Percent Significant (5%)	76.00%	31.20%	16.80%
Penn World Tables, 1960-00			
Median Statistic	2.233067	0.286	-1.060
Number significant (/125)	63	19	15
Percent Significant (5%)	50.40%	15.20%	12.00%

Hong-White tests are based on estimation of the HW statistic in equation (29) while residual regression tests are based on estimation of the U-statistic in equation (30). Confidence intervals built using bootstrapped test statistics with residual sampling and 100 observations. HW tests use J=1 and K*=2. Bandwidth choisen by Generalized Cross-Validation for the non-parametric estimate and set to n-.2 for the U-statistic.

Table 7: Comparison of Average Derivatives in OLS and Additively Separable Specifications			
	OLS	Additively Separable Estimation	
	Coefficient	Weighted Average Derivative	Fraction with same sign
Inflation	-0.013*	-0.0153	1.000
Log(Black Market Premium)	-0.0158	-0.026	0.944
Government Consumption	-0.0182**	-0.0101	1.000
Tariffs	0.0166	-0.0034	0.099
Combined Policy Index	-0.0183***	-0.0147**	1.000
Rule of Law	0.0095	0.0162*	0.795
Political Instabiilty	0.0226*	0.018	1.000
Effectiveness of Government Spendin	0.0353***	0.0379***	1.000
Index of Economic Freedom	0.0351***	0.039***	0.775
Combined Institutions Index	0.0335***	0.0478***	1.000
Non-Primary Exports	0.0031	0.0081	0.649
Urbanization	0.0118*	0.0002	0.308
Life Expectancy	0.0395***	0.015*	0.456
Financial Development	0.0242	0.0178	1.000
Combined Structure Index	0.0425***	0.0341***	0.789

OLS: Coefficient comes from linear estimation of equation (27) using the combined indices to proxy for the two other dimensions of production function shifters. All regressions use the World Bank 1975-03 sample. Significance test carried out with conventional t-statistics. ASE: Unweighted average derivative assigns equal weight to every point on the normalized [0,1] scale. Weighted average derivative is weighted by the frequency of country observations at each point. Significance tests based on bootstrapped confidence intervals derived using 100 replications for each regression. Stars denote significance level: *-10%, **-5%, ***-1%.

Table 8: Monotonicity Tests, Fourier Series and Kernel Estimators

	Hong-White Fourier Series Test		Residual Regression Penalized Spline Tests	
	Conventional Wisdom	Anti-Conventional Wisdom	Conventional Wisdom	Anti-Conventional Wisdom
Null hypothesis	Decreasing	Increasing	Decreasing	Increasing
Inflation	0.97	0.55	0.60	0.00
Log(Black Market Premium)	0.32	0.27	0.69	0.00
Government Consumption	0.63	0.31	0.45	0.00
Tariffs	0.04	0.02	0.01	0.00
Combined Policy Index	0.49	0.15	0.52	0.00
Null hypothesis	Increasing	Decreasing	Increasing	Decreasing
Rule of Law	0.30	0.10	0.41	0.00
Political Instabiilty	0.24	0.10	0.17	0.00
Effectiveness of Government Spending	0.02	0.05	0.72	0.00
Index of Economic Freedom	0.06	0.00	0.86	0.00
Combined Institutions Index	0.54	0.00	0.81	0.00
Null hypothesis	Increasing	Decreasing	Increasing	Decreasing
Primary Exports	0.37	0.09	0.61	0.01
Urbanization	0.48	0.06	0.52	0.01
Life Expectancy	0.54	0.00	0.17	0.00
Financial Developmentg	0.11	0.01	0.52	0.00
Combined Structure Index	0.58	0.00	0.67	0.00

Table reports p-values for the null hypothesis of monotonicity. Hong-White tests are based on estimation of the HW statistic in equation (29) while residual regression tests are based on estimation of the U-statistic in equation (30). Confidence intervals built using bootstrapped test statistics with residual sampling and 100 observations. HW tests use J=1 and K*=2. Bandwidth choisen by Generalized Cross-Validation for the non-parametric estimate and set to $n^{-.2}$ for the U-statistic.

Table 9: Average partial derivative estimates, linear and non-parametric estimates.

	Linear Coefficient	Average Derivative
Log(1+Inflation)	-0.013*	-0.0025
	(0.0070)	(0.0037)
Black Market Premium	-0.0158	-0.0033
	(0.0157)	(0.0042)
Government Consumption	-0.0182**	-0.0073
	(0.0073)	(0.0049)
Tariff Rate	0.0166	-0.0008
	(0.0181)	(0.0054)
Policy Index	-0.0183***	0.0037
	(0.0067)	(0.0045)
Rule of Law	0.0095	0.0057
	(0.0065)	(0.0041)
Political Stability	0.0226*	0.0192***
	(0.0119)	(0.0044)
Economic Freedom Index	0.0353***	0.0159***
	(0.0131)	(0.0048)
Index of government Effectiveness	0.0351***	0.0121***
	(0.0090)	(0.0043)
Institutions Index	0.0335***	0.0102**
	(0.0106)	(0.0049)
Non-Primary Exports in 1970	0.0031	0.0104**
	(0.0068)	(0.0047)
Urbanization Rate	0.0118*	0.0139***
	(0.0061)	(0.0043)
Life Expectancy	0.0395***	0.0137**
	(0.0097)	(0.0060)
Liquid Liabilities/GDP	0.0242	0.0206***
	(0.0163)	(0.0053)
Economic Structure Index	0.0425***	0.0177***
	(0.0117)	(0.0053)

OLS: Standard errors in parenthesis. ADE: The semi-parametric estimate is obtained using Xplore's implementation of Hardle and Stoker's (1989) indirect average derivative estimator, with m=1.5. Bootstrapped standard errors, with 200 replications, in parenthesis. Asterisks denote levels of significance: ***-1%, **-5%, *-10%

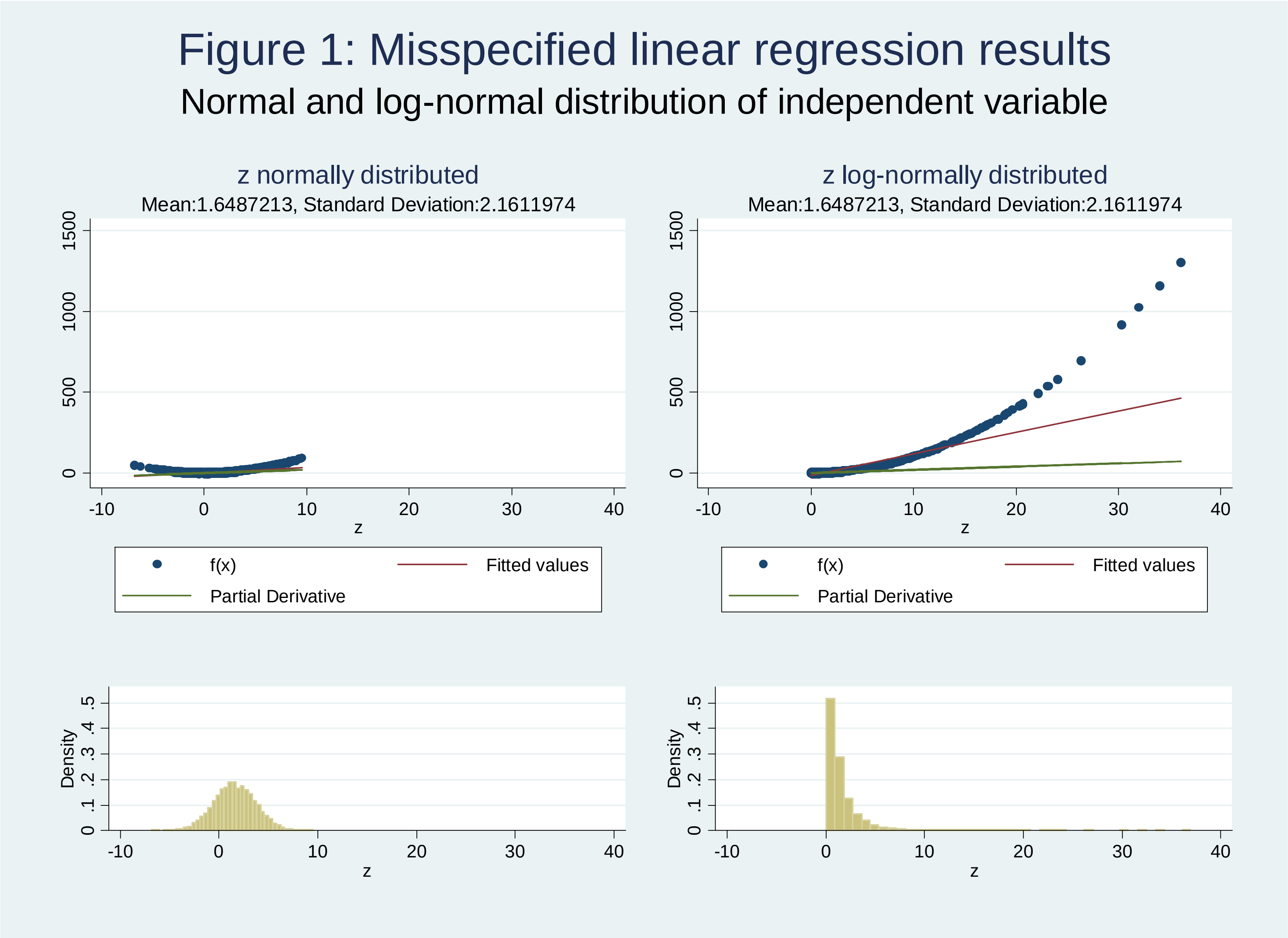

Figure 1: Misspecified linear regression results
Normal and log-normal distribution of independent variable
z normally distributed
Mean:1.6487213, Standard Deviation:2.1611974
z log-normally distributed
Mean:1.6487213, Standard Deviation:2.1611974
0
500
1000
1500
-10
0
10
20
30
40
z
f(x)
Fitted values
Partial Derivative
Density
0
.1
.2
.3
.4
.5