# Cleaning Up the Kitchen Sink: On the Consequences of the Linearity Assumption for Cross-Country Growth Empirics.

Francisco Rodríguez[1]
Department of Economics
Wesleyan University
First Version: September 7, 2005
This Version: January 19, 2006

**Abstract:** Existing work in growth empirics either assumes linearity of the growth function or attempts to capture non-linearities by the addition of a small number of quadratic or multiplicative interaction terms. Under a more generalized failure of linearity or if the functional form taken by the non-linearity is not known *ex ante*, such an approach is inadequate and will lead to biased and inconsistent OLS and instrumental variables estimators. This paper uses non-parametric and semi-parametric methods of estimation to evaluate the relevance of strong non-linearities in commonly used growth data sets. Our tests decisively reject the linearity hypothesis. A preponderance of our tests also rejects the hypothesis that growth is a separable function of its regressors. Absent separability, the approximation error of estimators of the growth function grows in proportion to the number of relevant dimensions, substantially increasing the data requirements necessary to make inferences about the growth effects of regressors. We show that appropriate non-parametric tests are commonly inconclusive as to the effects of policies, institutions and economic structure on growth.

[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. Alex Julca, Jomo K.S., José Antonio Ocampo, Codrina Rada, Sanjay Reddy, Roberto Rigobón, Dani Rodrik, Eric Verhoogen, Eduardo Zambrano and seminar participants at Harvard University, the United Nations and Wesleyan University were extremely helpful. Reyes Rodríguez provided first-rate research assistance. All errors remain the authors' pure responsibility.

## 1. Introduction

During the past fifteen years, a voluminous body of literature has been written using the linear growth regression framework to study the effect of variables as diverse as fiscal policy, the rule of law, the incidence of malaria and a country's past colonial regime on economic performance. The papers in this literature commonly study the coefficient on a variable or a subgroup of variables of interest in a regression where per capita GDP growth is the dependent variable and the regressors include a list of country-specific controls such as log of initial per capita GDP, investment rates, the stock of human capital and the growth rate of population.

Growth regressions have become an ubiquitous form of policy analysis. Empirical work in this literature is often geared towards reaching (or rebutting) a conclusion that a certain variable of interest – say a particular economic policy or one of a variety of institutional arrangements – is harmful or beneficial for growth. It is not uncommon for research in this area to conclude with phrases such as "We find clear evidence that the institution and policy variables play a significant role in determining economic growth."[2] Even the widespread practice of inspection of partial scatter plots and correlations between growth and policies is, in essence, the use of a growth regression framework.

Considerable work has been carried out attempting to deal with the more serious econometric problems arising in this literature. Particular attention has been paid to problems of robustness (Levine and Renelt, 1992, Sala-i-Martin, 1997, Sala-i-Martin, Doppelhoffer and Miller, 2004) and endogeneity (Frankel and Romer, 1999, Acemoglu, Johnson and Robinson, 2001).

The issue of functional form has received somewhat less attention. The most common approach has been to explore non-linearities with respect to the variable of interest, assuming linearity in the remaining regressors (Barro, 1996, Banerjee and Duflo, 2003). A number of authors have introduced elements of non-parametric estimation to consider more general non-linearities (Liu and Stengos, 1999, Kalatzidakis et. al., 2001). However, the approach is usually concentrated on understanding the effects of non-linearity in a particular dimension rather than studying the implications of a more general breakdown of the assumption. An alternative approach has been to study models of parameter heterogeneity (Durlauf and Johnson, 1995, Durlauf, Kourtellos and Minkin, 2001) in which countries are characterized by different linear models. These exercises commonly make strong assumptions as to the form that the underlying heterogeneity or non-linearity takes.[3] Despite these explorations, the standard workhorse regression model is still that of the linear regression framework. For example, in Sala-i-Martin, Doppelhoffer and Miller's (2004) recent Bayesian exploration of robustness issues, all of the approximately 89 million regressions studied are linear.

[2] This particular phrase is taken from DeGregorio and Lee (2004).

[3] There is a subtle distinction between non-linearities and parameter heterogeneity. In principle, parameter heterogeneity can refer to all countries having a linear model but with different parameters, whereas non-linearities occur when all countries share a common model with non-linear effects. However, when the heterogeneity depends on one or more variables (as in the cases studied by the authors cited) then parameter heterogeneity and non-linearities share the same mathematical representation.

There are two obvious limitations of using a linear framework to evaluate the growth effects of the variables often considered in this type of analysis. In the first place, a linear framework rules out the possibility that the effect of a change in the variable of interest may differ according to the level taken by the variable. Consider for example the hypothesis that that increases in tariff rates do not have much of an adverse effect on growth when starting from an initial level of relative openness, but that completely isolating an economy from world trade can be very harmful to a country's capacity to sustain adequate living standards. Such a hypothesis would find no place in a linear growth regression. In the second place, the linear framework rules out the possibility that the effect of certain variables may depend on the levels of other variables (i.e., that the effect of openness may depend on whether the economy's initial comparative advantage lies in manufactures or in agricultural goods).

If existing non-linearities were easy to identify and limited to one or two variables, it would perhaps be feasible to deal with them through the use of simple parametric devices such as the inclusion of quadratic and multiplicative interaction terms. This is indeed the solution commonly adopted in the literature when such failures are detected. But what if the failure is more generalized? What if the non-linearity is more complex than what can be captured by a set of simple quadratic and linear interaction terms, and its form is unknown, so that we cannot put *a priori* assumptions on its functional forms? As I discuss below, if this is the case then most of the regressions currently estimated suffer from misspecification bias, making the type of inferences commonly drawn from their estimation invalid. Furthermore, the data requirements of estimating non-linear unknown functions can be quite demanding and far outstrip the availability of data in currently existing data sets.

This problem is more than a theoretical curiosum. A systematic exploration of the theoretical foundations of the linear growth specification reveals that the set of assumptions necessary to justify fitting a linear function to the data is so restrictive as to practically make the linear specification the true theoretical curiosum. I suggest that the starting framework for an exploration of the growth evidence should be a specification that allows for a general set of interactions between the set of potential production function shifters.

The importance of interactions among different dimensions of potential regressors has become the focus of recent attention in the academic literature. In a recent paper, Hausmann, Rodrik and Velasco (2004) point out that the Theorem of the Second Best would lead us to expect 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 models. They also present a discussion of a number of cases in which similar policies appear to have had very different growth effects and suggest that they may be due to the fact that the countries faced different binding constraints on economic growth.[4]

The relevance of this discussion has not been lost among applied economists and policymakers. In a recent comprehensive appraisal of the results of a decade of economic reforms published by the World Bank, the role of interactions between

[4] See also Hausmann and Rodrik's (2005) more in-depth discussion of the Salvadoran case within this framework.

policies, institutions and economic structure is not only recognized but made to play a central role. In their words:

> "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…Countries with remarkably different policy and institutional frameworks – Bangladesh, Botswana, Chile, China, Egypt, India, Lao PDR, Mauritius, Sri Lanka, Tunisia and Vietnam – have all sustained growth in GDP per capita incomes above the U.S. long-term growth rate of close to 2 percent a year." (World Bank, 2005, p. 12)

If these two examples are any indication of the thrust of present theoretical and policy discussions, then the linear growth framework seems to be out of sync with them. On the other hand, before one gets carried away theorizing about the multiple potential interactions in the growth process, it may make sense to explicitly test the linear framework against more general alternative hypotheses in order to empirically evaluate the relevance of claims of strong non-linearities in the data.

The purpose of this paper is to present such a set of tests. The basic idea is to use semi-parametric and semi-nonparametric estimation of a generalized growth regression that allows for any type of interactions between potential production function shifters such as economic policies, institutions and economic structure. Concretely, I will test the hypothesis of linearity against more general non-linear hypotheses. I will also present tests evaluating whether a non-linear but separable specification, in which cross-variable interactions are negligible, is consistent with the data, as well as a discussion of the appropriate tests for evaluating the effects of these variables in a high-dimensional non-linear framework.

The rest of the paper proceeds as follows. Section 2 lays the theoretical groundwork. Section 2.1 discusses the theoretical underpinnings of the linear growth regression and establishes the assumptions that are necessary to obtain a specification that is linear in the production function shifters. Section 2.2 discusses the econometric effects of failure of these assumptions, and particularly of attempting to estimate a non-linear function by fitting a linear specification. Section 3 then shifts to empirical analysis, discussing the data set used (section 3.1) and then going on to present the tests of linearity (section 3.2), separability (section 3,3), and monotonicity (section 3.4). Section 4 concludes.

## 2. Theoretical Framework

### 2.1. Is there a theoretical basis for the Kitchen Sink Regression?

In this section I discuss the theoretical basis for the linear growth regression. This regression, often referred to as a "Barro" regression because of the deep influence of Robert Barro's 1991 *Quarterly Journal of Economics* article, was proposed almost simultaneously by several other authors including Mankiw, Romer and Weil (1992) and Sala-i-Martin (1991). It consists of a growth regression that is linear in the log of initial GDP, some measures of investment or the stock of 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:

$$\gamma_Y = \alpha_0 + \alpha_1 \ln y_{t-1} + \alpha_2 s_k + \alpha_3 s_h + \alpha_4 n + \beta Z \tag{2.1}$$

where $\gamma_Y$ is the rate of per capita GDP growth, $y_{t-1}$ is initial GDP, $s_k$ and $s_h$ refer respectively to the rate of investment in physical and human capital, *n* is the rate of population growth and ***Z*** is a vector of potential production function shifters that commonly includes measures of policies, institutions, and structural characteristics.

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 (2.1) is not surprising. For obvious reasons, I will not discuss this voluminous literature here; the reader is referred to Durlauf, Johnson and Temple ( forthcoming) 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.

Equation (2.1) 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’s (1991) augmented Solow model. It is useful to recall this derivation. Let output $Y_t$ be a Cobb-Douglas function of human and physical capital $H_t$ and $K_t$:

$$Y_t = AK^{\alpha}{}_t H^{\beta}{}_t . \tag{2.2}$$

Letting lower case letters denote units per worker, “hatted” lower case letters units per effective labor $AL_t$ and assuming constant rates of accumulation for physical and human capital $s_k$ and $s_l$, the model’s equations can be written as:

$$\hat{y}_t = \hat{k}^{\alpha}{}_t \hat{h}^{\beta}{}_t \tag{2.3}$$

$$\dot{\hat{k}}_t = s_k \hat{y}_t - (n + g + \delta)\hat{k}_t \tag{2.4}$$

$$\dot{\hat{h}}_t = s_h \hat{y}_t - (n + g + \delta)\hat{h}_t \tag{2.5}$$

Setting $\dot{\hat{k}}_t = \dot{\hat{h}}_t = 0$ gives us the steady-state solution to the model:

$$\ln(\hat{y}_{ss}) = \frac{\alpha}{1-\alpha-\beta}\ln(s_k) + \frac{\beta}{1-\alpha-\beta}\ln(s_h) - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta) \tag{2.6}$$

From (2.3), the rate of growth of $\hat{y}$ can be written as $\alpha \frac{d \ln \hat{k}_t}{dt} + \beta \frac{d \ln \hat{h}_t}{dt}$. Approximating each of these terms around the steady state gives us:

$$\gamma_{\hat{Y}} = \frac{d \ln \hat{y}(t)}{dt} = -(n+g+\delta)(1-\beta-\alpha)\ln(\hat{y}/\hat{y}_{ss})\,. \qquad (2.7)$$

Equation (2.7) is a first-order linear differential equation in ln( $\hat{y}_t$). Solving it gives:

$$\ln \hat{y}_t = e^{-\lambda t} \ln \hat{y}_0 + (1-e^{-\lambda t})\ln \hat{y}_{ss}\,, \qquad (2.8)$$

from which we can derive the approximated expression for the growth rate between $\hat{y}_0$ and $\hat{y}_t$ as:

$$\ln(\hat{y}_t/\hat{y}_0) = -(1-e^{-\lambda t})\ln \hat{y}_0 + (1-e^{-\lambda t})\ln \hat{y}_{ss}. \qquad (2.9)$$

Empirically, we do not observe growth rates in units of effective labor. Therefore, in order to have an empirically estimable equation, we must put (2.9) in terms of per capita output $y_t$:

$$\ln(y_t/y_0) = -(1-e^{-\lambda t})\ln y_0 + (1-e^{-\lambda t})\ln \hat{y}_{ss} + \ln(A_t - A_0) + (1-e^{-\lambda t})\ln A_0. \qquad (2.10)$$

Finally, substituting for $\hat{y}_{ss}$ from (2.6) leads to:

$$\ln(y_t/y_0) = -(1-e^{-\lambda t})\ln y_0$$
$$+(1-e^{-\lambda t})\left[\frac{\alpha}{1-\alpha-\beta}\ln s_k + \frac{\beta}{1-\alpha-\beta}\ln s_h - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta)\right] + g + (1-e^{-\lambda t})\ln A_0.$$
(2.11)

There is an alternative formulation which can be arrived at by substituting for $\ln s_h$ from the steady-state condition for human capital accumulation, which is:

$$\ln(y_t/y_0) = -(1-e^{-\lambda t})\ln y_0 + (1-e^{-\lambda t})\left[\frac{\alpha}{1-\alpha}\ln s_k + \frac{\beta}{1-\alpha}\ln h_{ss} - \frac{\alpha}{1-\alpha}\ln(n+g+\delta)\right]$$
$$+ g + (1-e^{-\lambda t})\ln A_0.$$
(2.12)

Note that I have normalized the time interval *t* to 1 in the derivation of these last two equations. Equations (2.11) and (2.12) are linear in ln$y_o$, ln$s_k$, ln$s_h$ (or ln$h_{ss}$), ln($n+g+\delta$), $g$ and ln$A_O$ and would thus be estimable as a linear equation if we had observations on all of these variables.

I now concentrate on equation (2.12) (the derivation for equation (2.11) is analogous). Mankiw, Romer and Weil (1992) assume that $g$ and $\delta$ are constant and equal across countries and that differences in the initial level of technology vary randomly according to:

$$\ln(A_0) = \ln(A) + \varepsilon_i \qquad (2.13)$$

with $\varepsilon_i$ 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:

$$\gamma_t = A_0 + A_1 \ln y_0 + A_2 \ln s_k + A_3 \ln h_{ss} + A_4 \ln(n+g+\delta) + \eta_i. \qquad (2.14)$$

to the data. This is, indeed, what Mankiw, Romer and Weil do.

Equation (2.12) would also seem to open the door to a more general approach. As Mankiw, Romer and Weil 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 $\varepsilon_i$ is correlated with any of the regressors in (11) or (12), its omission would bias the estimated coefficients. Even if omitted variable bias is unimportant, equations (11) and (12) seem to offer a ready framework to evaluate the effect of multiple measures of policies, institutions and economic structure on growth. One simply needs to reason that the variables of interest could affect the growth rate by shifting the production function, so that we could rewrite (2.2) as:

$$Y_t = A(Z) K^{\alpha}{}_t L^{\beta}{}_t \qquad (2.15)$$

,where $Z$ is a vector of our potential explanatory variables. The above derivation would then follow except for the fact that $A_0$ would now be replaced by $A(Z_0)$ and that $g=\ln(A_t/A_0)$ is now $g(Z_0, Z_t) = \ln[A(Z_t)/A(Z_0)]$. Equation (2.12) would become:

$$\ln(y_t / y_0) = -(1-e^{-\lambda t}) \ln y_0 + (1-e^{-\lambda t}) \left[ \frac{\alpha}{1-\alpha} \ln s_k + \frac{\beta}{1-\alpha} \ln h_{ss} - \frac{\alpha}{1-\alpha} \ln(n + g(Z_t, Z_0) + \delta) \right]$$
$$+ g(Z_t, Z_0) + (1-e^{-\lambda t}) \ln A(Z_0). \qquad (2.16)$$

It is important to note that, as it stands, (2.16) 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 $\ln A(Z) = \beta Z$. Additionally, one needs to assume that the growth rate of $A$ over time is the same for all countries, that is, that $g_i(Z_0, Z_t) = \overline{g}\ \forall i$. Given these assumptions, (2.16) reduces to:

$$\ln(y_t / y_0) = -(1-e^{-\lambda t}) \ln y_0 + (1-e^{-\lambda t}) \left[ \frac{\alpha}{1-\alpha} \ln s_k + \frac{\beta}{1-\alpha} \ln h_{ss} - \frac{\alpha}{1-\alpha} \ln(n + \overline{g} + \delta) \right]$$
$$+ \overline{g} + (1-e^{-\lambda t}) \beta Z_0. \qquad (2.17)$$

, leading to the conventional "kitchen sink" specification often used in applied work:

$$\gamma_t = A_0 + A_1 \ln y_0 + A_2 \ln s_k + A_3 \ln h_{ss} + A_4 \ln(n+g+\delta) + \chi_1 Z_{10} + ... + \chi_n Z_{n0} + \eta_i .\text{[5]} \quad (2.18)$$

Summarizing, what this discussion has shown is that the following assumptions are necessary in order for specification (2.18) to be valid:

A1. Savings rates for human and physical capital are constant.

A2. Production is Cobb-Douglas in physical and human capital.

A3. The economy is sufficiently close to the steady state for (2.9) to be a valid approximation.

A4. All countries have the same level of growth of *A(Z)*, i.e.: $g_i(Z_0, Z_t) = \bar{g} \; \forall i$ .

A5. The log of production is linear in all of its shifters, i.e.: $\ln A(Z) = \beta Z$ .

Assumptions A1 and A2 have been the source of an extensive discussion in the literature. Replacing constant savings rates by the solution to an intertemporal optimization problem was a focus of the early growth literature going back to Ramsey's (1928) seminal contribution. Barro and Sala-i-Martin (2004) show that a log-linearized version of the Ramsey optimization problem will lead to a first-order linear differential equation like (2.7), but with a more complex expression for the parametric convergence coefficient. Relaxing the Cobb-Douglas specification leads to non-linearities in $s_h$,$s_k$ and $h_{ss}$ and has been explored empirically, among others, by Liu and Tsengos (1999) and Massanjala and Papageorgiou (2004). Failure of the Taylor approximation leads to differing convergence rates across countries as discussed, among others, by Dowrick (2004).

Assumption A4 seems at first (and at second and third) glance bizarre. 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, taken by Mankiw, Romer and Weil, is to see *g* as capturing only the effects of technological change, which is assumed to be public and available to all countries, while $A(Z_o)$ is held to be fixed at its initial level. Research exploring the failure of this hypothesis often looks at varying rates of diffusion of technologies across countries (Coe and Helpman, 1995, Coe, Helpman and Hoffmaister 1997). 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. Given the pervasive use of equation (2.18) to draw inferences used in the process of policy reform, it is ironic that its theoretical foundation actually restricts these policies to be time invariant.

I know of no systematic treatment of the effects of failure of assumption A5. This is surprising, given that, unlike A1-A4, A5 is almost completely atheoretical. There is no reason why one would expect variables as diverse as economic policies, institutions and structural characteristics to have separable, linear effects on the log of the production function. Indeed, to the extent that one sees the "production-shifting" effect of the *Z* variables on the production function as reflecting the efficiency effects of relaxing different distortions, the Theorem of the Second-Best tells us that there is no reason to

[5] Equation (2.1) also differs from this regression in that it uses levels of $s_k$, $h$ and $n$ as regressors instead of the terms in equation (2.17). Regrettably, these easily corrigible flaws are pervasive in most applied work.

expect that relaxing one distortion would lead to an increase in efficiency when another distortion is present; in other words, it tells us that the effects of distortions on efficiency are unlikely to be separable.

Ultimately, the validity of any of these assumptions will be given by their usefulness in explaining the data. Despite its lack of theoretical appeal, assumption A5 may prove useful if it allows us to account for existing differences in growth rates across countries in a parsimonious way. To lay such a claim, however, the patterns in the data must be consistent with those predicted by the theory. I turn to exploring whether this is the case in section 3. To do this I will test (2.18) against the more general semi-parametric form:

$$\gamma_t = A_0 + A_1 \ln y_0 + A_2 \ln s_k + A_3 \ln h_{ss} + A_4 \ln(n+g+\delta) + f(Z_{10},\ldots Z_{n0}) + \eta_i . \quad (2.19)$$

The next section studies the empirical tenability of A5 through a battery of semiparametric and semi-nonparametric tests of the hypothesis embodied in equation (2.18) against the more general alternative (2.19). Before going on, however, it makes sense to discuss the econometric implications of (2.19) actually being the true data generating process.

## 2.2. The Econometric Effects of Throwing In the Kitchen Sink

### 2.2.1. The Effects of Misspecification Bias

Estimating (2.18) if (2.19) is the true function will lead to misspecification bias. And is analogous to imposing the invalid restriction that the nonlinearity is not present. Its implications can be easily seen within the framework of omitted variable bias.

To see this, rewrite (2.19) as:

$$\gamma_t = A_0 + A_1 \ln y_0 + A_2 \ln s_k + A_3 \ln h_{ss} + A_4 \ln(n+g+\delta) + C_1 Z_{10} + \ldots + C_n Z_{1n} + h(Z_{10},\ldots Z_{p0}) + \eta_i . \quad (2.20)$$

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 (2.18) by OLS is therefore the same as estimating (2.20) subject to the restriction that $h(Z_{10},\ldots,Z_{p0}) = 0$, that is, inappropriately excluding the non-linear term from the regression. The limit in probability of the OLS estimators $\hat{A} = \{\hat{A}_0, \hat{A}_1, \hat{A}_2, \hat{A}_4\}'$ and $\hat{C} = \{\hat{C}_1,\ldots\hat{C}_n\}'$ will be:

$$p\lim \hat{A} = \{A_0 \ldots A_4\}'$$
$$p\lim \hat{C} = C + (Var(Z)^{-1})Cov(Z_0, h(Z_{10},\ldots,Z_{p0})) \quad (2.21)$$

Even if $f(.)$ is independent of $\ln y_o$, $\ln s_k$, $\ln s_h$ 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 gives us the average effect of changing the explanatory variable over the sample of countries. If this were true, it would imply that the linear estimator may not be a poor guide to evaluating the expected effects of changes in policies or institutional and structural reforms: even if we cannot recover the expected effect of these changes for a given country, we may still be able to inform the policymaker of the expected effect of making such a change over all countries. In this vein, Helpman (2004) has argued that:

> "The way trade policy affects an economy's growth rate depends on the country's characteristics, such as the type of products it trades on foreign markets or the human-capital intensity of its import-competing sectors. Nevertheless, empirical studies do not provide estimates of the growth effects of trade policies conditioned on these characteristics. Therefore *estimates that exploit cross-country variations are best interpreted as average effects of trade policies on growth*, similarly to the estimates of the effects of trade volumes in growth that were discussed above." (p. 73, emphasis added).

Is this conjecture correct? The following proposition establishes that, as a general fact, the conditions necessary to ensure that the linear estimator provides a meaningful average effect are quite restrictive.

**Proposition 1.** Let $y=f(x)$ be an arbitrary nonlinear function of the vector $x=\{x_1,\dots x_n\}\in\Re^n$, where $x$ is distributed according to the joint density function $f(x)$ with mean $\mu$ and variance-covariance matrix $Var(x)$. Let $\beta=\{\beta_0,\dots\beta n\}$ be the vector of coefficients on $\{1,x\}$ of the linear projection of $y$ on $x$, $L(y|1,x)$. Then $\beta_i=E(\partial f(x)/\partial x_i)$ only if (i) $f(x)$ is symmetric, i.e.: $E(x_i-\mu_i)^3=0$ for all $i$ (ii) $f(x)$ is mesokurtic, i.e.:$(E(x_i-\mu_i)^4)/(E(x_i-\mu_i)^2)^2=3$ for all $i$.

**Proof:** See Appendix

The linear estimator will return the average partial derivative only if the joint density of the explanatory variables is symmetric and mesokurtic. While these conditions are satisfied by the normal distribution, they are jointly satisfied by few other distributions. More importantly, they do not appear to be satisfied by the data commonly used in cross-country growth regressions.

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 $h(Z_{10},\dots,Z_{p0})$. 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. The next example makes this clear.

**Example.** Let $A_0=\dots=A_4=C_2=\dots=C_p=0; h(.)=Z_1^2$. (2.20) now collapses to:

$$\gamma_t=C_1Z+D_1Z^2+\eta_i. \qquad (2.25)$$

Note that we have omitted the intercept to keep the algebra as simple as possible. We have also dropped unnecessary subscripts on the only $Z$ variable. We assume there is a simple structural relationship between the endogenous variable and the candidate instrument:

$$Z_i = X_i + u_i \,. \tag{2.26}$$

Our assumptions on the DGP are $Cov(u,\eta) \neq 0, Cov(X,u) = Cov(X,\eta) = 0$. The OLS estimator (without a constant) will be:

$$C_1 = \frac{\sum_i X_i \gamma_i}{\sum_i X_i Z_i} = C_1 + D_1 \frac{\sum_i X_i Z_i^{\,2}}{\sum_i X_i Z_i} + \frac{\sum_i X_i \eta_i}{\sum_i X_i Z_i} \,. \tag{2.27}$$

Taking probability limits gives us:

$$p\lim \hat{C}_1 = C_1 + D_1 \frac{E(X^3) + E(X)\sigma^2{}_\eta}{E(X^2)} \,. \tag{2.28}$$

This expression depends on the first through third moments of $X$ and does not appear to admit a simple intuitive interpretation.□

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. To this list of requirements one should add the assumption that the dependent variable depends linearly on the endogenous variable.

### 2.2.2. Non-linearities and the Curse of Dimensionality

A second effect of attempting to estimate (2.18) when (2.19) is true is that the rate of convergence associated with the probability limits calculated in (2.21) can become much slower. Succinctly put, the basic problem is that estimating a non-linear function is much more demanding in terms of data than estimating a linear function, because we must sample it at many more points to be certain of its shape.[6] If there were no sampling error, we would need only two observations to fit a linear function $y = \beta_0 + \beta_1 x$ to the data. However, if the function is a non-linear function of the form $f(x,\beta)$ with β a k-dimensional vector, then even in the absence of sampling error we will need at least *k* points to infer β.

What if, as is the case in non-parametric estimation, we do not know the functional form taken by $f(x)$? Suppose all we know is that it has a bounded first derivative. For concreteness and without loss of generality, let $f(x)$ be defined on the unit interval and the bound on its first derivative to be *L*. If *f* is sampled at *n* equidistant points and is approximated by the closest point at which we have an observation, then the approximation error cannot exceed *L/2n*, which is of order *O(1/n)*. In other words, the approximation error goes to zero at a rate *O(1/n)*.

What if *f* is a function of two variables? Suppose that we have the same *n* observations as before, evenly sampled over the domain. Then the average distance

[6] The following discussion closely follows Yatchew (2004).

between two points will be $1/n^{1/2}$ instead of $1/n$. The average approximation error will now be $O(1/n^{1/2})>O(1/n)$. The argument can be extended to any dimension: if $f(.)$ is a function of $k$ variables then the approximation error will be $O(1/n^{1/k})$. In other words, the order of the approximation error grows with dimensionality at a rate that is proportional to the number of observations.

To illustrate, suppose that we have 100 observations, which is close to the average number of observations commonly used in growth regressions. Then the approximation error for a one-dimensional function will be $O(0.01)$. If the function is two-dimensional, however, it will be $O(0.1)$ and if it is three-dimensional it will be $O(0.21)$. Having 100 observations to estimate a one-dimensional relationship is tantamount to having 10 observations ($100^{1/2}$) to estimate a two-dimensional specification and to having 4.64 observations ($100^{1/3}$) to estimate a three-dimensional specification. To be consistent, a researcher should place the same faith on a regression estimate of a general non-linear function in three dimensions that is run with 100 observations than she should put on a correctly specified linear regression that was run with 5 observations. This result is known as the *curse of dimensionality* in the literature on non-parametric econometrics and it underlines the difficulty in making appropriate inferences about unknown non-linear functions with few observations.

More generally, let $d$ denote the dimensionality of $x$ and $g(x)$ be differentiable up to the $m$-th derivative. Then the optimal rate at which a non-parametric estimator can convege to the true regression function is (Stone 1980):

$$\int\left[\hat{f}(x)-f(x)\right]^2 dx = O\left(\frac{1}{n^{2m/(2m+d)}}\right). \tag{2.23}$$

One possible mechanism for attenuating the curse of dimensionality is the use of additively separable specifications. For example, suppose that we can split $Z$ in $I$ disjoint subsets such that the maximum dimension of each of these subsets is $l<d$ and where $f(Z)=f_1(Z_{I_1})+...+f_n(Z_{I_n})$. Then it can be established (see Hastie and Tibshirani 1990) that the rate of convergence of the optimal estimators is $O_p\left(\frac{1}{n^{2m/(2m+l)}}\right)$, representing a significant improvement over (2.23). One could hope that this result would allow us to make valid inferences in growth empirics without falling into the curse of dimensionality. This is the route taken, for example, by Liu and Tsengos (1999). I discuss this approach in section 3.3.

## 3. Empirical Evidence

The argument set out in the preceding section can be briefly summarized as follows: (i) The theoretical basis for the linear kitchen-sink growth regression is quite tenuous. It does **not** emerge directly from a linear approximation to the steady-state, but rather requires directly assuming that the log of the production function is linear in the variables of interest, i.e. that it takes the form $Y_t=\left[\exp(\beta_1 Z_1+...+\beta_1 Z_n)\right]F(K_t,H_t)$. (ii) Assuming that the valid regression equation is linear when it is not leads to misspecification bias which is akin to that caused by restricting all non-linear terms to having coefficients equal to zero when that is not the case, causing the estimated coefficients to be biased and inconsistent. (iii) Unless $f(Z)$ is an additively separable function, the data requirements necessary to precisely estimating it may far outstrip existing data availability.

In the end, of course, the problems posed by these facts are empirical. Whether Assumption A5 is an adequate characterization of the data or not is a question that can ultimately be answered by the data itself. Therefore, a relevant starting point for answering this question is to ask whether the linearity hypothesis explicit in A5 can be rejected in existing cross-country data sets. That is the first task taken up in this section.

Assuming that one can reject the linear specification (as I will argue can very easily be done) then one must ask to what extent non-linearities can be accounted for through an additively separable specification, and thus the curse of dimensionality can be attenuated, or, on the other hand, to what extent the data says that the phenomenon under study is inherently multi-dimensional and can only be captured through non-separable specifications (implying that we must suffer the full effects of the curse). That is the second task that I will take up in this section.

Lastly, I address the issue of how much can be recovered in terms of evidence from the existing data given pervasive problems of multidimensionality. I will argue that the relevant and useful analogue to a significance test on a linear coefficient is a monotonicity test. To that effect, I will present a set of tests of the monotonicity of growth as a function of a set of commonly used indicators of economic policies, institutions and structural variables. I will argue that these tests point in the direction of a severe complexity of existing interactions which is very much at odds with the idea that there are policies, institutions or varieties of structural transformation that are unequivocally good or bad for all countries.

### 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 1975-00 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 recent contributions.[7] 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 long-run growth, and most 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.[8] 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

[7] 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).

[8] 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 (forthcoming) for a discussion.

variables. I restrict the period of estimation to 1975-00 both because the WDI data is only available for that time period and because this is also the period over which many of the right-hand side variables are available. Results are broadly similar for the 1960-00 PWT series.

Estimation starts out from the semi-linear growth equation (2.19):

$$\gamma_t = A_0 + A_1 \ln y_0 + A_2 \ln s_k + A_3 \ln h_{ss} + A_4 \ln(n + g + \delta) + f(Z_{10},...Z_{n0}) + \eta_i. \qquad (2.19)$$

One of the most commonly expressed preoccupations in the literature about this regression regards the endogeneity bias that may arise out of the inclusion of endogenous variables such as *n* or *s*. To this effect it is common to omit some of these variables from the estimated specification (see, for example, Barro, 1999), thus trading off a reduction in endogeneity bias against an increment in omitted-variable bias. Following this tradition, I present four variants of estimates of (3.1), all of which use the conditional convergence control for log($y_{t-1}$) but which differ in their progressive inclusion of the remaining variables.

As our production-function shifters *Z*, I use 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 is provided in Table 1.

As is common in the literature, I estimate (2.19) 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, 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 the emphasis is placed on the fraction of specifications for which a given hypothesis is rejected.

### 3.2 Linearity

Econometrically, the linearity assumption is the simplest one to test in that it can easily be done within the framework of parametric estimation. The basic reason is that, under the hypothesis that growth is a linear function of its determinants, no non-linear terms should be significant. Therefore, one can evaluate this hypothesis by estimating a standard linear growth regression and adding on a series approximation to estimate the residual non-linearity:

$$\gamma_y = \alpha_0 + \alpha_1 \ln(y_{t-1}) + \alpha_2 \ln h_{t-1} + \alpha_3 \ln(n+g+\delta) + \alpha_4 s_k + \alpha_5 z_p + \alpha_6 z_i + \alpha_7 z_s + p(z_p, z_i, z_s) + \varepsilon_i \quad (3.3)$$

where $z_p, z_i, z_s$ stand, respectively, for the policy, institutional and structural indicators and $p(z_p, z_i, z_s)$ is the series approximation. Under the null hypothesis of linearity, the coefficients on the non-linear terms should have no effect. If we find that they are jointly significant, this will imply rejection of the linearity hypothesis. If we do not find them to be significant, it may always be the case that the terms that we were including did not accurately capture the non-linearities actually present in the data, and we would need to look for more general non-parametric methods to approximate these non-linearities.

A first approach to this issue is presented in Table 2. In it I present the result of estimating (2.19) through ordinary least squares using a Taylor expansion 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 (3.2). 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.

As we can clearly see, the rejection of linearity is quite strong. In the Penn World Tables Data, linearity is rejected between 90.4 and 92.8 percent of the time, depending on which of the production-function controls are included. With the World Bank data, the rejection of linearity is even stronger, with rejection occurring between 96.8 and 99.2 percent of the time. 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 (2.19). The significance of any non-linear terms in *X* can only enhance, and in no way weaken, the case against the linear specification.

### 3.3 Separability

The results of the preceding section tell us that the linearity assumption can easily be rejected in existing cross-country data sets, but leave us little clue as to the form of the actual non-linearity. A common approach in the literature to the estimation of data is to include simple quadratic and interaction terms to approximate existing non-linearities. In this vein, it is common to see theoretical arguments about why the relationship between growth and a potential determinant may be increasing for certain values of the variable – or of other covariates – and decreasing for others followed by estimation of the regression with non-linear terms or multiplicative interactions. Usually linearity in the rest of the variables in the specification is maintained. For example, Barro (1999) argues that democracy can be good for growth at low levels but starts being harmful for growth alter a certain threshold.

The flaws in such an approach are evident. In the first place, it requires accepting the assumption that growth is linear in all other covariates in the regression even when we are trying to discern whether it is non-linear in the variable or interactions of interest. Second, even if such an assumption were to hold, the non-linearity in the variable of interest can be much more complex than what can be captured by a quadratic term or a multiplicative interaction.

In contrast, the approach taken here is to start out from a very general form of the non-linearity and to study the form that it takes and the restrictions that can be imposed on it without excessive loss of fit. As discussed in section 2, 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: for three dimensions and two bounded derivatives, for example, the rate of convergence under additive separability is $O_p\left(n^{-4/5}\right)$ as compared with $O_p\left(n^{-4/7}\right)$ in the non-separable case and $O_p\left(n^{-1}\right)$ in the parametric case. Indeed, given sufficient smoothness in the additively separable terms, the rate of convergence of the additively separable estimator can be made arbitrarily close to the parametric rate of convergence. Therefore a logical starting point is to test for additive separability of the growth function.

I start out by testing additive separability of *f(Z)* in the partially linear specification (2.19). This is done by studying the capacity of the additively separable specification:

$$\gamma_y = \alpha_0 + \alpha_1 \log(y_{t-1}) + \alpha_2 \ln h_{t-1} + \alpha_3 \ln(n+g+\delta) + \alpha_4 s_k + f_p(z_p) + f_i(z_i) + f_s(z_s) \tag{3.4}$$

to account for variations in the data. In order to do this, I carry out four 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 in a simple 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 sampling theory associated with this test is well-known and need not be discussed here.

The results of the F-test for significance of the interaction terms on a the Taylor polynomial specification are shown in Table 3. Rejection rates for the separability hypothesis at the 5% level of significance oscillate between a range of 54.4-65.6% for the Penn World Tables data and 72.8-76.8% for 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. It is thus desirable to attempt estimation of these models via semi-parametric or non-parametric methods that can allow our estimates of the $f_i(.)$ functions to be as complex as possible. The next three 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, that is, 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:

$$\gamma_y = \alpha X + u_0 + \sum_{i=1}^{3} b_i z_i + \sum_{i=1}^{3}\sum_{j=1}^{3} c_{ij} z_i z_j + \sum_{i=1}^{3}\sum_{j=1}^{J_n} \left\{ u_{ij}\cos(jk'_i z) + v_{ij}\sin(jk'_i z) \right\}, \quad (3.5)$$

where I have written the parametric part of the equation compactly as $\alpha X$. 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 $N$. We can obtain a restricted estimator $\beta_r$ 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:

$$HW = \frac{\frac{1}{N}\frac{\sum_{i=1}^{N} e^2_{ri} - \sum_{i=1}^{N} e^2_{ui}}{\sum_{i=1}^{N} e^2_{ui}} - M}{(2M)^{1/2}} \to N(0,1), \quad (3.6)$$

where $M$ is the number of terms in the Fourier expansion. Table 4 shows the median F-Statistic, p-values, and number of rejections of the null of separability using the Hong-White test. For this estimation, I have chosen $J=1$ and $K^*=2$, which gives us $M=28$ and a saturation ratio that varies between .25-.40, a range typical of the applied literature (see Chalfant and Gallant, 1985 and Pagan and Ullah, 2004). The results in Table 4 show much lower frequency of rejections of separability than in the polynomial expansion. The percentage of rejections of separability are in the range of 24.0-36.8% for the Penn World Tables data and 21.6-36.8% for the World Bank data. The probability of rejection appears to increase when more linear terms are added to the specification. These results are somewhat discouraging in that they point neither towards a consistent rejection nor to a consistent non-rejection of separability.

The choice of $M$ in Table 4 is actually the lowest number consistent with allowing the Flexible Fourier Form to nest a non-separable hypothesis. This is symptomatic of the fact that estimation of high-dimensional non-parametric functions requires a high number of observations, which is nothing else than the curse of dimensionality in a different guise. In order to explore the effect of this restriction, we can try to raise the number of $M$, running the risk of having somewhat high saturation ratios. Table 5 shows the results of doing this by raising $J$ to 2, giving us saturation ratios between .41 and .66, which are near the higher end of the range of what is found in the applied literature. It can be seen that shifting to the higher saturation ratio significantly raises the rejection rate to levels as high as 88.0%.

The next two tests that I discuss depend on direct estimation of the additively separable specification (3.4) by semi-parametric methods and analysis of its residuals. There are a number of econometric methods available for estimating additively separable functions. In what follows, I will use the results of estimating (3.4) through marginal integration using local polynomial estimation[9]. 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:

$$U = \frac{1}{\lambda^{d/2} n} \sum_i \sum_{j \neq i} \left(y_i - \hat{f}_r(x_i)\right)\left(y_j - \hat{f}_r(x_j)\right) \prod_{k=1}^{3} K\left(\frac{x_{jk} - x_{ik}}{\lambda}\right) \quad (3.7)$$

, 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 $2\sigma^4 \int p^2(x) \int K^2(u)$ [10]. The *differencing* test (Yatchew, 1988, 2003) is a goodness of fit test that consists in comparing the restricted variance estimator to the variance estimator that can be obtained from first-differencing the data. The logic of this approach comes from the fact that, if care is taken to ensure that the data is purged of the parametric effect and appropriately ordered in the non-parametric dimension(s), then the non-parametric effect should be negligible as long as the distance between different observations in the non-parametric dimension is small. In that case, and as long as the restricted variance estimator converges sufficiently rapidly under the null, then

$$Y = (mn)^{1/2} \left(\frac{s^2_r - s^2_{diff}}{s^2_{diff}}\right) \to N(0,1) \quad . \quad (3.8).$$

Tables 6 and 7 report the results for residual regression and differencing estimators of separability respectively. I have set the bandwidth λ to 0.7 and the order of differencing to 1. Rejection rates remain high for other bandwidth choices, even with considerable over-smoothing (Table A-1). Data are ordered according to Yatchew's (1988) bin-restricted nearest-neighbor algorithm.[11] As one can see, the rejection rates are moderately high. For the residual regression test, these oscillate between 60.8% and 75.2% for the Penn World Tables data and between 52.0 % and 69.6% for the World

---

[9] This is carried out using the **intestpl** command in XploRe. See (Fan, Haerdle and Mammen (1996) and Sperlich and Zelinka (2003) for a discussion. The statistical properties of marginal integration estimators tend to be simpler than those of Hastie and Tsibirani's (19xx) backfitting algorithm (Pagan and Ullah, 1999, p.137-9).

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

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

[11] Given the small number of observations, one could worry that differencing will not adequately purge the non-parametric effect, as the distance between nearby observations is likely to be high. As long as the parametric effect is adequately purged, however, this tends to inflate the estimates of $s^2_{diff}$ and thus to bias our results towards non-rejection of separability.

Bank data. In the case of Yatchew's differencing test, rejection rates are very high: between 72.8% and 98.4% for the Penn World Tables data and between 79.2% and 91.2% for the World Bank data.

In sum, the four proposed tests of separability give contrasting results. Differencing tests tend to almost unequivocally reject separability and so do the preponderance of residual regression and Taylor series expansion tests. Only in the case of Fourier expansion tests is there is a tendency to not reject separability more often than not, but this result is heavily dependent on the choice of terms for the expansion: the higher-order Fourier expansion tends to present very high rejection rates of the separability hypothesis, suggesting that the lower-order expansion may not be able to accommodate existing complexities in the data.

The fact that these tests give contrasting results should not be surprising, given that the sampling theory behind them is completely asymptotic but the samples used, which range from N=70 to N=111, are quite small. Not surprisingly, these tests have been found to differ in their small-sample properties in applied Monte-Carlo simulations (see Hong and White, 1995). Furthermore, small-sample biases aside, they may differ in their power to reject the alternative, and these differences in turn may depend on the true form that the alternative may take in reality. 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 just one.[12]

### 3.4 Monotonicity

Taken at face value, the results of the previous subsection are somewhat troubling. It has been shown that there is little theoretical reason to expect growth to be a separable function of production-function shifters; if anything, results based on the Theorem of the Second Best would seem to imply the opposite. Section 3.3 shows that the hypothesis of non-separability is very hard to discount empirically. But we have also seen that the curse of dimensionality implies that the informational requirements for precisely estimating higher-dimensional functions are way above what is available in existing data sets.

One may be tempted to close the argument here by claiming to have demonstrated that cross-country growth empirics simply lack the data necessary to estimate the complex processes which they are attempting to study. This conclusion would coincide with the view of a number of critics of the literature, who have tended to dismiss cross-country regressions as a futile exercise that attempts to reach broad-ranging conclusions based on the comparison of a reduced number of widely divergent cases.[13]

---

[12] Our current research is focused on understanding the effect of small-sample biases on the properties of the alternative tests presented in this context.

[13] See, for example, Bhagwati and Srinivasan, 2000.

However, before we throw away the baby with the bath-water, it may be useful to consider more carefully the full implications of the curse of dimensionality for cross-country growth empirics. As stated in section 2.2, the curse of dimensionality refers to the rate of convergence of estimators $\hat{f}(.)$ of a general parametric function $f(.)$, that is, of the rate at which $E(\hat{f}(.) - f(.))$ can be said to tend to zero. In essence, the curse of dimensionality tell us that we cannot expect to precisely estimate the true function $f(.)$, because in order to do that we would need to know the form that this function takes for many different values of *z*, and when *z* is a higher-dimensional vector and we have a reduced number of observations, there are many values of *z* which we are likely to have very little or no information on. This does not mean that we cannot estimate $f(.)$ with greater confidence locally in areas in which we do have greater amount of information. It also does not imply that we cannot make some reasonable inferences about what $f(.)$ does **not** look like.

In order to make these arguments more concrete, Figure 3.1 shows an example of a three-dimensional scatter plot of growth as a hypothetical function of two policy determinants based on simulated data, with a limited number of observations (N=75). Obviously, there is a lot that we cannot know about the true form of the underlying function. Some areas of the graph are so sparsely populated so as to make inference about the expectation of growth impossible without very strong parametric assumptions. For example, as plotted there are no data points with high values of Policy 2 and low values of Policy 1, so that there is very little that we can say about the expected value of growth for a country that adopted such a policy combination. However, that is not the case for low values of policies 1 and 2. Furthermore, there are some things that can be safely inferred about the form of the growth function given his data. For one, it is quite clear from the picture that, at least for some levels of policies, growth is **not** a monotonically decreasing function in either policy 1 or policy 2. It would be very hard to argue, after having seen this picture, that policy 1 and policy 2 are unequivocally bad for growth.

The preceding argument captures the intuition of the tests that I will present in the remainder of this section. Through them, I will try to ask whether the growth data is consistent or not with certain hypotheses about the form of the *f*(.) function. The basic idea is to attempt to present a more consistent non-parametric alternative to the parametric significance tests that are commonplace in cross-country regression analysis. 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:

$$H_0 : (z_1,...z_m) \geq (z'_1,...z_m') \rightarrow f(z_1,...z_m) \geq f(z'_1,...z'_m) \quad (3.9)$$ [14]

[14] There are a number of non-parametric tests of significance in the literature that consist in studying the results of imposing the restriction that the regression is independent of the variable in question (e.g., Gozalo, 1995). A problem in using such a test for policy analysis would be that it can conclude that a certain variable can be significant as a result of its having positive effects over some ranges and negative effects over other ranges, without allowing us to determine the ranges over which either effect dominates. A monotonicity test, in contrast, tells us

against the alternative:

$$H_1 : \exists (z_1,...z_m),(z'_1,...z_m') \,|\, \{(z_1,...,z_m) > (z_1,...,z_m), f(z_1,...z_m) \le f(z'_1,...z'_m)\}. \quad (3.10)$$

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 $HW$, $U$ and $Y$ 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

$$\frac{\partial \hat{f}}{\partial z_i}(.) \ge 0. \quad (3.11)$$

The sum of squared residuals is minimized subject to (3.11) 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 and differencing tests, however, the issue is a bit more complex, as (3.10) cannot so easily be summarized as one restriction. The explicit parametric representation of the Fourier series expansion allows summarizing (3.11) via explicit calculation of the first derivative. In the residual regression and differencing case, there is no such easy expedient and we must impose (3.11) locally at every point. Since this is not feasible computationally, as it would imply solving a non-linear optimization problem subject to an infinity of constraints, I approximate it by dividing the $Z_p \times Z_i \times Z_s$ space into one-thousand cubes of volume $0.1^3$=0.001 and constraining $\hat{f}_r(.)$ to be monotonic between any two locally adjoining cubes. This still requires estimation subject to a very high number of constraints (5.920). [15] In practice, however, the actual number of binding constraints is much smaller, making estimation computationally feasible. [16]

Table 8 displays the results of the Fourier expansion tests. As we can see in the upper panel, the data seems to have very little capacity to reject the "conventional wisdom" view that distortionary policies are bad for growth, market-preserving institutions are good for growth and a more modernized economic structure is good for growth. The hypotheses are tested separately.[17] Rejection rates for all of the three hypotheses are invariably in the single digits, with a maximum rate of 9.6% for the

---

whether we could expect a certain policy to always be good or at least not harmful for growth. Obviously, one could also carry out local monotonicity tests, but these are likely to be much more demanding in terms of data requirements.

[15] Each cube has seven adjoining cubes that are characterized by one policy being strictly higher, except for the 270 cubes that are on the upper boundaries, which need only be compared with three adjoining cubes.

[16] 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.

[17] Joint testing of the three hypotheses yiedls very simlar results.

structural variables. For no policy is it the case that there are more rejections than non-rejection of the monotonicity hypothesis.

Unlike the case of standard significance tests in linear regressions, the tests shown in this and the next two tables treat the hypothesis that a certain variable is beneficial – or detrimental – to growth as the null hypothesis. By construction, this is a necessity of the logical structure of the tests: in order for the specifications to be appropriately nested, the unrestricted specification must correspond to that in which any non-linear form is possible, while the restricted specification is the one in which only monotonic functional forms are acceptable. Therefore, in the absence of sufficient information, the tendency of the tests will be to not reject the null hypothesis. One likely effect of the curse of dimensionality may be to make just about any hypothesis difficult to reject in very small data sets. In other words, it may be relatively simple to fit *many* highly non-linear functional forms, both monotonic and non-monotonic, to a data set with relatively few observations.

The lower panel of Table 8 explores this possibility by looking at the rejection tests for an alternative set of hypotheses that would stand in complete contrast with the conventional wisdom view, consisting in the hypothesis that the functions in question are monotonic with the contrary sign than specified in the conventional wisdom hypothesis. I call this the **contrarian view**. Thus, in the policy arena, the contrarian view would be that increasing policy-induced distortions, say by raising tariff rates, is monotonically good for growth. In the case of institution, the contrarian hypothesis will take the form of a statement that growth is a monotonically decreasing function of institutional improvements such as improvements in the rule of law. The relevant question becomes whether the data has enough information so as to reject this clearly unorthodox view.

The results shown in the lower panel of Table 8 are quite discouraging in this respect. Rejection rates of the contrarian view for policies oscillate in the range of 35.2-39.2%, while for institutions they are slightly higher (34.4-48.8%) but still well below one-half. Only in the case of economic structure do we find some cases in which the contrarian view is rejected for the median specification, but even here on average more than four out of every ten specifications cannot reject the hypothesis that a less modern economic structure is always good for growth. Although rejections of the contrarian view are much more common than those of the conventional wisdom, it is still the case that in the majority of specifications one can refute neither the hypothesis that they are monotonically beneficial for growth nor the hypothesis that they are monotonically harmful for growth.

The tendency for greater rejection of the contrarian hypothesis disappears once we consider the residual regression and differencing tests on the partially linear specification, which show rejection rates in the low single digits for both the conventional wisdom and the contrarian view. The results of these tests, shown in Tables 9 and 10, can be succinctly summarized as follows: there appears to be nothing in the data that allows these tests to refute the hypothesis of monotonicity in either direction for any of the production function shifters.

Given the high dimensionality of the function estimated, and the small number of observations available, these results are not surprising. Taken together, these two facts imply that it is very easy to incorporate many restrictions into a non-linear function without suffering much loss of fit. Given the rejection of additive separability and the curse of dimensionality discussed above, these results are little more than expression of the very reduced informational content that is available in existing data once we accept the general non-linear specification. The curse of dimensionality has come back to haunt us, and it appears to be here to stay.

## 4. Concluding Comments

This paper brings a discouraging message to those interested in carrying out policy analysis within the growth regression framework. In essence, I have argued that (i) the theoretical basis for the linear kitchen-sink regression is tenuous; (ii) there are considerable risks from misspecification bias that come from using such a specification when it is not valid; (iii) the data strongly supports the hypothesis that a linear specification is not valid; (iv) there may be too little information in existing data sets to allow us to appropriately make the type of inferences about the growth effects of particular policies or strategies for economic and structural reforms that the profession has become used to drawing.

The basic reason behind these results is that, once we recognize the true multi-dimensionality of the growth process, existing data is clearly insufficient to allow us to understand it in a statistical sense. It is one thing to try to distinguish between the hypothesis that openness is equally good for all countries and the hypothesis that openness is equally bad (or equally irrelevant) for all countries than trying to distinguish among a broad set of potential hypotheses that allow for complex interactions between openness 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 unlikely to be feasible 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.

Does this mean that the empirical analysis of growth data sets is a worthless endeavor? I do not think so. Actually, this paper has shown that one can use existing data sets to make non-trivial inferences about the growth process. The tests presented in section 3 present a decisive rejection of the linearity hypothesis. We do not seem to be in a world where any country can expect to have the same effect from a proportionate change in a particular policy, institution or structural characteristic irrespective of its starting level. Furthermore, the preponderance of the evidence seems to weigh against the hypothesis of separability: we do not appear to be in a world in which the effect of a particular policy does not depend on the state of institutions or the economy's structural characteristics. These conclusions are in themselves very important: they show that we do not live in a simple world, where the same rules can be use to design growth strategies in China and in Chile. In the dimension of policy, institutional and structural effects, the world does not seem to be very flat. Rather, it appears to be a pretty rocky place.

The problem, in my appraisal, comes from trying to make the data say more than it can. It may simply be too much to ask the post-war growth experience to tell us what the effects of government spending on economic growth are if we know that such effects are likely to depend in a complex way on everything else that is going on in the economy.

It is rather like trying to tell the shape of the universe from looking at the sky. Perhaps the most that we can reasonably expect to learn is that it's full of stars.

## 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.

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.

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

Coe, David T., and Elhanan Helpman (1995) "International R&D Spillovers." *European Economic Review*, 39(5), pp.859-87.

Coe, David T., Elhanan Helpman and Alexander W. Hoffmaister (1997) "North-South R&D Spillovers," *Economic Journal*, 107, 134-149.

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.

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

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

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.

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

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.

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

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.

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

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

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

Sala-i-Martin, Xavier (1991) *On Growth and States*. PhD Dissertation, Harvard University.

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” REporduced: MDTech.

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

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

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 Applkied Econometrician*. Cambridge, UK: Cambridge University Press.

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.

**Table 1: Variable Description**

| | |
|---|---|
| **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); Data from World Bank (2004) |
| 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 Data Set. |
| 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 2: Linearity Tests, Taylor Polynomial Expansions**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| Penn World Tables, 1975-00 | | | | |
| Median F-Statistic | 4.90 | 4.92 | 4.62 | 5.01 |
| Median P-Value | 0.00 | 0.00 | 0.00 | 0.00 |
| Number significant (/125) | 116 | 113 | 114 | 114 |
| Percent Significant (5%) | 92.8% | 90.4% | 91.2% | 91.2% |
| World Bank, 1975-03 | | | | |
| Median F-Statistic | 5.00 | 4.72 | 4.74 | 4.71 |
| Median P-Value | 0.00 | 0.00 | 0.00 | 0.00 |
| Number significant (/125) | 123 | 124 | 124 | 121 |
| Percent Significant (5%) | 98.4% | 99.2% | 99.2% | 96.8% |

Reported results refer to conventional F-test of the null hypothesis that all coefficients on non-linear terms in a $3^{rd}$-order polynomial expansion are equal to zero. Number and percent significant are calculated using a significance level of 5%.

**Table 3: Separability Tests, Taylor Polynomial Expansions**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| Penn World Tables, 1975-00 | | | | |
| Median F-Statistic | 2.62 | 2.09 | 2.20 | 2.40 |
| Median P-Value | 0.01 | 0.04 | 0.03 | 0.02 |
| Number significant (/125) | 82 | 68 | 72 | 75 |
| Percent Significant (5%) | 65.6% | 54.4% | 57.6% | 60.0% |
| World Bank, 1975-03 | | | | |
| Median F-Statistic | 2.70 | 2.90 | 2.75 | 2.88 |
| Median P-Value | 0.01 | 0.01 | 0.01 | 0.01 |
| Number significant (/125) | 91 | 92 | 96 | 96 |
| Percent Significant (5%) | 72.8% | 73.6% | 76.8% | 76.8% |

Reported results refer to conventional F-test of the null hypothesis that all coefficients involving inter-variable interactions in a $3^{rd}$-order polynomial expansion are equal to zero. Number and percent significant are calculated using a significance level of 5%.

**Table 4: Hong- White Separability Tests, Flexible Fourier Form, J=1, K*=2**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| Penn World Tables, 1975-00 | | | | |
| Median F-Statistic | -0.02 | 0.60 | 0.59 | 0.87 |
| Median P-Value | 0.51 | 0.27 | 0.28 | 0.19 |
| Number significant (/125) | 30 | 41 | 46 | 46 |
| Percent Significant | 24.0% | 32.8% | 36.8% | 36.8% |
| World Bank, 1975-03 | | | | |
| Median F-Statistic | -0.07 | 0.89 | 0.87 | 0.84 |
| Median P-Value | 0.53 | 0.19 | 0.19 | 0.20 |
| Number significant (/125) | 27 | 41 | 45 | 46 |
| Percent Significant | 21.6% | 32.8% | 36.0% | 36.8% |

Number and percent significant are calculated using a significance level of 5%.

**Table 5: Hong- White Separability Tests, Flexible Fourier Form, J=2, K*=2**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| Penn World Tables, 1975-00 | | | | |
| Median F-Statistic | 3.22 | 4.15 | 4.83 | 5.38 |
| Median P-Value | 0.00 | 0.00 | 0.00 | 0.00 |
| Number significant (/125) | 87 | 102 | 108 | 110 |
| Percent Significant | 69.6% | 81.6% | 86.4% | 88.0% |
| World Bank, 1975-03 | | | | |
| Median F-Statistic | 2.37 | 3.76 | 3.92 | 4.37 |
| Median P-Value | 0.01 | 0.00 | 0.00 | 0.00 |
| Number significant (/125) | 74 | 107 | 105 | 106 |
| Percent Significant | 59.2% | 85.6% | 84.0% | 84.8% |

Number and percent significant are calculated using a significance level of 5%.

**Table 6: Residual Regression Tests, Marginal Integration Estimates of Additively Separable Function**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| Penn World Tables, 1975-00 | | | | |
| Median F-Statistic | 2.85 | 2.19 | 2.03 | 2.63 |
| Median P-Value | 0.00 | 0.01 | 0.02 | 0.00 |
| Number significant (/125) | 94 | 78 | 76 | 86 |
| Percent Significant | 75.2% | 62.4% | 60.8% | 68.8% |
| World Bank, 1975-03 | | | | |
| Median F-Statistic | 2.73 | 1.87 | 1.69 | 2.14 |
| Median P-Value | 0.00 | 0.03 | 0.05 | 0.02 |
| Number significant (/125) | 87 | 69 | 65 | 74 |
| Percent Significant | 69.6% | 55.2% | 52.0% | 59.2% |

Number and percent significant are calculated using a significance level of 5%.

**Table 7: Differencing Tests, Marginal Integration Estimates of Additively Separable Function**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| Penn World Tables, 1975-00 | | | | |
| Median F-Statistic | 18.98 | 17.99 | 7.90 | 33.27 |
| Median P-Value | 0.00 | 0.00 | 0.00 | 0.00 |
| Number significant (/125) | 118 | 118 | 91 | 123 |
| Percent Significant | 94.4% | 94.4% | 72.8% | 98.4% |
| World Bank, 1975-03 | | | | |
| Median F-Statistic | 13.54 | 18.13 | 9.86 | 34.01 |
| Median P-Value | 0.00 | 0.00 | 0.00 | 0.00 |
| Number significant (/125) | 112 | 99 | 100 | 114 |
| Percent Significant | 89.6% | 79.2% | 80.0% | 91.2% |

Number and percent significant are calculated using a significance level of 5%.

**Table 8: Fourier Expansion Tests of Monotonicity (PWT Data)**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| **Conventional Wisdom** | | | | |
| **Policies** | | | | |
| Median F-Statistic | -2.00 | -1.90 | -1.50 | -1.62 |
| Median P-Value | 0.98 | 0.97 | 0.93 | 0.95 |
| Number significant (/125) | 6 | 9 | 7 | 9 |
| Percent Significant | 4.8% | 7.2% | 5.6% | 7.2% |
| **Institutions** | | | | |
| Median F-Statistic | -2.37 | -2.20 | -1.84 | -1.91 |
| Median P-Value | 0.99 | 0.99 | 0.97 | 0.97 |
| Number significant (/125) | 6 | 5 | 7 | 4 |
| Percent Significant | 4.8% | 4.0% | 5.6% | 3.2% |
| **Structure** | | | | |
| Median F-Statistic | -2.52 | -2.36 | -2.15 | -2.09 |
| Median P-Value | 0.99 | 0.99 | 0.98 | 0.98 |
| Number significant (/125) | 8 | 12 | 12 | 10 |
| Percent Significant | 6.4% | 9.6% | 9.6% | 8.0% |

Number and percent significant are calculated using a significance level of 5%.

**Table 8: Fourier Expansion Tests of Monotonicity (PWT Data) (Continued)**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| | | **Contrarian View** | | |
| **Policies** | | | | |
| Median F-Statistic | 0.61 | 0.90 | 0.70 | 0.76 |
| Median P-Value | 0.27 | 0.18 | 0.24 | 0.22 |
| Number significant (/125) | 45 | 49 | 44 | 44 |
| Percent Significant | 36.0% | 39.2% | 35.2% | 35.2% |
| **Institutions** | | | | |
| Median F-Statistic | 0.31 | 1.56 | 0.63 | 0.87 |
| Median P-Value | 0.38 | 0.06 | 0.27 | 0.19 |
| Number significant (/125) | 43 | 61 | 52 | 52 |
| Percent Significant | 34.4% | 48.8% | 41.6% | 41.6% |
| **Structure** | | | | |
| Median F-Statistic | 2.81 | 2.18 | 1.62 | 1.44 |
| Median P-Value | 0.00 | 0.01 | 0.05 | 0.08 |
| Number significant (/125) | 75 | 68 | 62 | 62 |
| Percent Significant | 60.0% | 54.4% | 49.6% | 49.6% |

Number and percent significant are calculated using a significance level of 5%.

**Table 9: Residual Regression Tests of Monotonicity (PWT Data)**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| **Conventional Wisdom** | | | | |
| **Policies** | | | | |
| Median F-Statistic | -1.44 | -1.45 | -1.70 | -1.74 |
| Median P-Value | 0.92 | 0.93 | 0.96 | 0.96 |
| Number significant (/125) | 0 | 0 | 0 | 0 |
| Percent Significant | 0.0% | 0.0% | 0.0% | 0.0% |
| **Institutions** | | | | |
| Median F-Statistic | -1.44 | -1.42 | -1.54 | -1.58 |
| Median P-Value | 0.93 | 0.92 | 0.94 | 0.94 |
| Number significant (/125) | 3 | 3 | 2 | 2 |
| Percent Significant | 2.4% | 2.4% | 1.6% | 1.6% |
| **Structure** | | | | |
| Median F-Statistic | -1.55 | -1.48 | -1.75 | -1.72 |
| Median P-Value | 0.94 | 0.93 | 0.96 | 0.96 |
| Number significant (/125) | 0 | 0 | 0 | 0 |
| Percent Significant | 0.0% | 0.0% | 0.0% | 0.0% |

Number and percent significant are calculated using a significance level of 5%.

**Table 9: Residual Regression Tests of Monotonicity (PWT Data) ( Continued)**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| | | **Contrarian View** | | |
| **Policies** | | | | |
| Median F-Statistic | -1.42 | -1.41 | -1.61 | -1.62 |
| Median P-Value | 0.92 | 0.92 | 0.95 | 0.95 |
| Number significant (/125) | 2 | 1 | 0 | 0 |
| Percent Significant | 1.6% | 0.8% | 0.0% | 0.0% |
| **Institutions** | | | | |
| Median F-Statistic | -1.32 | -1.27 | -1.38 | -1.51 |
| Median P-Value | 0.91 | 0.90 | 0.92 | 0.93 |
| Number significant (/125) | 4 | 3 | 2 | 2 |
| Percent Significant | 3.2% | 2.4% | 1.6% | 1.6% |
| **Structure** | | | | |
| Median F-Statistic | -1.34 | -1.30 | -1.54 | -1.60 |
| Median P-Value | 0.91 | 0.90 | 0.94 | 0.95 |
| Number significant (/125) | 4 | 3 | 3 | 2 |
| Percent Significant | 3.2% | 2.4% | 2.4% | 1.6% |

Number and percent significant are calculated using a significance level of 5%.

**Table 10: Differencing Tests of Monotonicity (PWT Data)**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| **Conventional Wisdom** | | | | |
| **Policies** | | | | |
| Median F-Statistic | -4.92 | -4.57 | -4.48 | -4.43 |
| Median P-Value | 1.00 | 1.00 | 1.00 | 1.00 |
| Number significant (/125) | 0 | 0 | 0 | 0 |
| Percent Significant | 0.0% | 0.0% | 0.0% | 0.0% |
| **Institutions** | | | | |
| Median F-Statistic | -4.57 | -4.50 | -4.48 | -4.20 |
| Median P-Value | 1.00 | 1.00 | 1.00 | 1.00 |
| Number significant (/125) | 0 | 0 | 0 | 0 |
| Percent Significant | 0.0% | 0.0% | 0.0% | 0.0% |
| **Structure** | | | | |
| Median F-Statistic | -4.62 | -4.69 | -4.64 | -4.38 |
| Median P-Value | 1.00 | 1.00 | 1.00 | 1.00 |
| Number significant (/125) | 0 | 0 | 0 | 0 |
| Percent Significant | 0.0% | 0.0% | 0.0% | 0.0% |

Number and percent significant are calculated using a significance level of 5%.

**Table 10: Differencing Tests of Monotonicity (PWT Data) (Continued)**

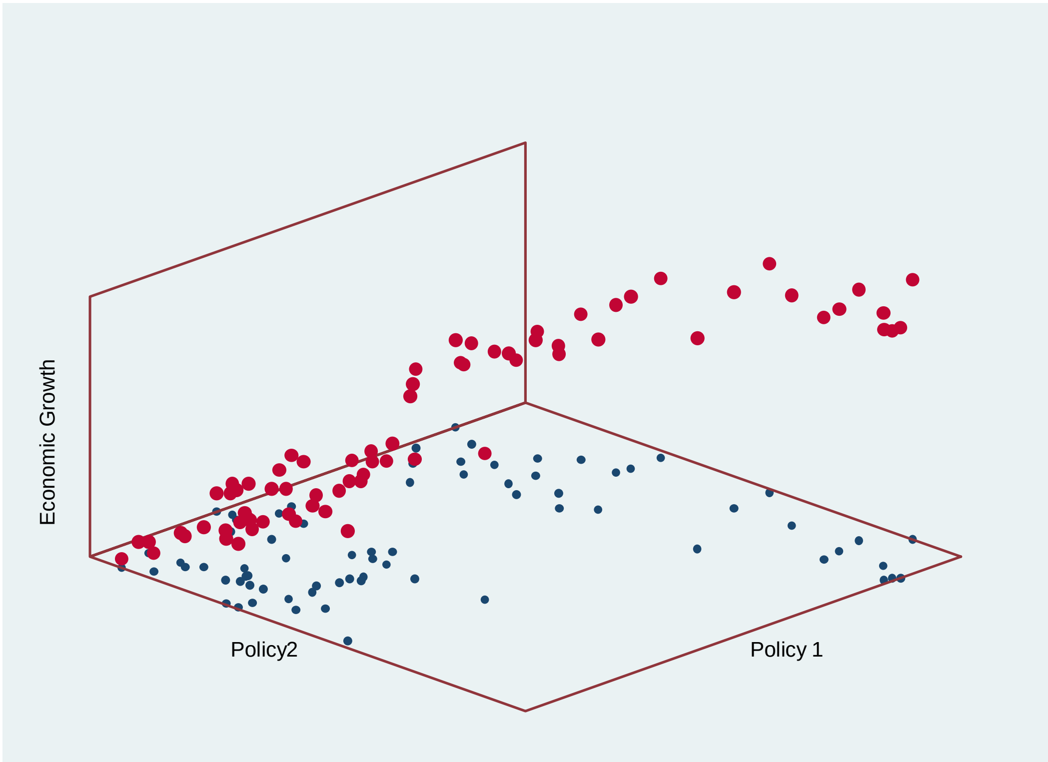
| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Controls* | *1975 GDP* | *1975 GDP, Schooling* | *1975 GDP, Schooling, Investment Rate* | *1975 GDP, Schooling, Investment Rate, Population Growth* |
| | | **Contrarian View** | | |
| **Policies** | | | | |
| Median F-Statistic | -4.60 | -4.66 | -4.43 | -4.13 |
| Median P-Value | 1.00 | 1.00 | 1.00 | 1.00 |
| Number significant (/125) | 1 | 0 | 0 | 1 |
| Percent Significant | 0.8% | 0.0% | 0.0% | 0.8% |
| **Institutions** | | | | |
| Median F-Statistic | -4.50 | -4.59 | -4.22 | -4.26 |
| Median P-Value | 1.00 | 1.00 | 1.00 | 1.00 |
| Number significant (/125) | 1 | 0 | 0 | 0 |
| Percent Significant | 0.8% | 0.0% | 0.0% | 0.0% |
| **Structure** | | | | |
| Median F-Statistic | -4.59 | -4.60 | -4.32 | -4.20 |
| Median P-Value | 1.00 | 1.00 | 1.00 | 1.00 |
| Number significant (/125) | 0 | 0 | 0 | 0 |
| Percent Significant | 0.0% | 0.0% | 0.0% | 0.0% |

Number and percent significant are calculated using a significance level of 5%.

**Table A-1: Sensitivity to Bandwidth Choice, Residual Regression and Differencing Tests of Separability**

| *Equation* | *(1)* | *(2)* | *(3)* | *(4)* |
|---|---|---|---|---|
| *Bandwidth* | *0.25* | *0.5* | *0.75* | *1* |
| *Residual Regression Tests* | | | | |
| Penn World Tables, 1975-00 | | | | |
| Median P-Value | 0.01 | 0.00 | 0.00 | 0.08 |
| Percent Significant | 71.4% | 91.8% | 64.0% | 46.4% |
| Penn World Tables, 1975-00 | | | | |
| Median P-Value | 0.00 | 0.00 | 0.03 | 0.12 |
| Percent Significant | 66.3% | 90.0% | 56.0% | 36.8% |
| *Differencing Tests* | | | | |
| Penn World Tables, 1975-00 | | | | |
| Median P-Value | 0.00 | 0.00 | 0.00 | 0.00 |
| Percent Significant | 100.0% | 99.1% | 89.6% | 68.0% |
| Penn World Tables, 1975-00 | | | | |
| Median P-Value | 0.00 | 0.00 | 0.00 | 0.00 |
| Percent Significant | 100.0% | 100.0% | 88.0% | 63.2% |

Figure 3.1: Simulated Data, Growth and Two Policy Determinants



# Appendix 1: Proofs of Propositions

January 19, 2006

**Proposition 1** *Let* $y = f(x)$ *be an arbitrary nonlinear function of the vector* $x = \{x_1, ...x_n\} \in \Re^n$, *where* $x$ *is distributed according to the joint density function* $f(x)$ *with mean* $\mu$ *and variance-covariance matrix* $Var(x)$. *Let* $\beta = \{\beta_0, ...\beta_n\}$ *be the vector of coefficients on* $\{1, x\}$ *of the linear projection of* $y$ *on* $x$, $L(y|1, x)$. *Then* $\beta_i = E\left(\frac{\partial f(x)}{\partial x_i}\right)$ *only if (i)* $f(x)$ *is symmetric, i.e.:* $E(x_i - \mu_i)^3 = 0$ *for all* $i$ *(ii)* $f(x)$ *is mesokurtic, i.e.:*$\frac{E(x_i-\mu_i)^4}{(E(x_i-\mu_i)^2)^2} = 3$ *for all* $i$.

***Proof.*** *Define* $g(z) = f(z + \mu_x) = f(x)$ *for* $z = x - \mu_x$. *Note that by the chain rule* $\frac{\partial f(x)}{\partial x_i} = \frac{\partial g(x)}{\partial z_i}$.*Note that if* $\gamma = \{\gamma_0, ...\gamma_n\}$ *is the vector of coefficients on* $\{1, z\}$ *in the linear projection of* $y$ *on* $z$, $L(y|1, z)$ *then* $\gamma_i = \beta_i$ *for all* $i = \{1, ...n\}$ *The reason is that* $\gamma$ *solves*

$$\min_{\gamma \in R^n} E\left[(y - \gamma_0 - \sum_{i=1}^{n} z_i\gamma_i)^2\right] = \min_{\gamma \in R^n} E\left[(y - \gamma_0 - \sum_{i=1}^{n} (x_i - \mu_x)\gamma_i)^2\right] \quad (1)$$

$$= \min_{\gamma \in R^n} E\left[(y - \left[\gamma_0 - \sum_{i=1}^{n} \mu_x\gamma_i\right] - \sum_{i=1}^{n} x_i\gamma_i)^2\right] \quad (2)$$

$$= \min_{\beta \in R^n} E\left[(y - \beta_0 - \sum_{i=1}^{n} x_i\beta_i)^2\right]. \quad (3)$$

*The last equality follows because if* $\gamma^*$ *is a solution to (1) (and thus to (2)) then* $\beta^* = \{\gamma_0^* - \sum_{i=1}^n \mu_x\gamma_i^*, \gamma_1^*...\gamma_n^*\}$ *gives the same value of the objective function. But then* $\beta^*$ *must be a solution to (3), because if there were another solution* $\beta' \neq \beta^*$ *then* $\gamma' = \left\{\beta_0' + \sum_{i=1}^n \mu_x\beta_i', \beta_1'...\beta_n'\right\}$ *would be a solution to (1). Let* $\overline{\beta} = \{\beta_1...\beta_n\} = \{\gamma_1...\gamma_n\}$ *denote the collection of the last n components of the coefficient vector for the linear projection of* $y$ *on* $z$ *(or* $x$*). We can find the values of these coefficients using the definition of a linear projection:*

$$\overline{\beta} = (Var(z))^{-1} Cov(y, z). \quad (4)$$

*Since by definition* $E(z)$ *is normalized to zero then* $Cov(g(z), z) = E(yz) = E(g(z)z)$. *Therefore (4) becomes:*

$$\overline{\beta} = (Var(z))^{-1} E(g(z)z) = E\left(\frac{\partial g(z)}{\partial z}\right). \quad (5)$$

*Now let us approximate $f(z)$ by an n-th order Taylor expansion, that is:*

$$g(z) \approx P(g(z)) = \alpha_0 + \sum_{k=1}^{m} \left( \sum_{i_1=1}^{n} ... \sum_{i_k=1}^{n} \alpha_{i_1...i_k} z_{i_1} ... z_{i_k} \right).$$

*where:*

$$\begin{aligned} \alpha_0 &= g(0) \\ \alpha_{i_1...i_k} &= \frac{\partial}{\partial z_{i_1} ... z_{i_k}} g(0). \end{aligned}$$

*Therefore (5) implies:*

$$\begin{aligned} \overline{\beta} &= (Var(z))^{-1} \begin{pmatrix} E\begin{pmatrix} \alpha_0 z_1 \\ + \sum_{k=1}^{m} \left( \sum_{i_1=1}^{n} ... \sum_{i_k=1}^{n} \alpha_{i_1...i_k} z_{i_1} ... z_{i_k} z_1 \right) \end{pmatrix} \\ ... \\ E\begin{pmatrix} \alpha_0 z_n \\ + \sum_{k=1}^{m} \left( \sum_{i_1=1}^{n} ... \sum_{i_k=1}^{n} \alpha_{i_1...i_k} z_{i_1} ... z_{i_k} z_n \right) \end{pmatrix} \end{pmatrix} \quad (6) \\ &= E \begin{pmatrix} \frac{\partial}{\partial z_1} \left( \alpha_0 + \sum_{k=1}^{m} \left( \sum_{i_1=1}^{n} ... \sum_{i_k=1}^{n} \alpha_{i_1...i_k} z_{i_1} ... z_{i_k} \right) \right) \\ \frac{\partial}{\partial z_1} \left( \alpha_0 + \sum_{k=1}^{m} \left( \sum_{i_1=1}^{n} ... \sum_{i_k=1}^{n} \alpha_{i_1...i_k} z_{i_1} ... z_{i_k} \right) \right) \end{pmatrix}. \quad (7) \end{aligned}$$

*Since (**??**) must be true for any $g(z)$ then it must be true for any $\alpha$. Set $\alpha_{i_1...i_k} > 0$ if $m = p$ and $i_1 = ... = i_p = d$ for some $d \in \{1..n\}$, $\alpha_{i_1...i_k} = 0$ otherwise . That is, the polynomial expansion of $g(z)$ is simply a linear function of the p–th order direct (?) moments of $z_d$: $P(g(z)) = \alpha_0 + \alpha_{d...d} z_d^p$ (perhaps because this is itself $g(z)$). Then (**??**) becomes:*

$$\overline{\beta} = (Var(z))^{-1} \begin{pmatrix} E(\alpha_0 z_1 + \alpha_{d...d} z_d^p z_1) \\ ... \\ E(\alpha_0 z_d + \alpha_{d...d} z_d^{p+1}) \\ ... \\ E(\alpha_0 z_n + \alpha_{d...d} z_d^p z_n) \end{pmatrix} = E \begin{pmatrix} \frac{\partial}{\partial z_1} (\alpha_0 + \alpha_{d...d} z_d^p) \\ ... \\ \frac{\partial}{\partial z_d} (\alpha_0 + \alpha_{d...d} z_d^p) \\ ... \\ \frac{\partial}{\partial z_n} (\alpha_0 + \alpha_{d...d} z_d^p) \end{pmatrix}. \quad (8)$$

*Premultiplying by $Var(z)$ and expanding, we get:*

$$\alpha_{d...d} \begin{pmatrix} E(z_d^p z_1) \\ ... \\ E(z_d^{p+1}) \\ ... \\ E(z_d^p z_n) \end{pmatrix} = \begin{pmatrix} \sigma_{11} & ... & \sigma_{1d} & ... & \sigma_{1n} \\ ... & ... & ... & ... & .. \\ \sigma_{d1} & ... & \sigma_{dd} & ... & \sigma_{dn} \\ ... & ... & ... & ... & ... \\ \sigma_{n1} & ... & \sigma_{nd} & ... & \sigma_{nn} \end{pmatrix} \alpha_{d...d} E \begin{pmatrix} 0 \\ ... \\ 0 \\ p z_d^{p-1} \\ 0 \\ ... \\ 0 \end{pmatrix}$$

*that is,*

$$\begin{pmatrix} E(z_d^p z_1) \\ ... \\ E(z_d^{p+1}) \\ ... \\ E(z_d^p z_n) \end{pmatrix} = E \begin{pmatrix} \sigma_{1d} E(p z_d^{p-1}) \\ ... \\ \sigma_{dd} E(p z_d^{p-1}) \\ ... \\ \sigma_{nd} E(p z_d^{p-1}) \end{pmatrix}.$$

*The d-th row of this column equality states that* $E(z_d^{p+1}) = E(z_d^2)E(p z_d^{p-1})$. *Plugging in* $p = 2$ *this gives us* $E(z_d^3) = E(z_d^2) 2E(z_d^1) = 0$, *proving symmetry. Plugging in* $p = 3$ *gives us* $E(z_d^4) = 3\left(E(z_d^2)\right)^2$, *establishing mesokurtosis.* ∎