Who Benefits from Financial Development? New Methods, New Evidence*

Daniel J. Henderson†, Chris Papageorgiou♯, and Christopher F. Parmeter‡

†State University of New York at Binghamton
♯International Monetary Fund
‡University of Miami

July 7, 2011

Abstract

This paper takes a fresh look at the impact of financial development on economic growth by using recently developed generalized kernel methods that allow for heterogeneity in partial effects, nonlinearities, and endogenous regressors. Our results suggest that while the positive impact of financial development on growth has increased over time, it is also highly nonlinear with more developed nations benefiting while low-income countries do not benefit at all. This finding contributes to the ongoing policy debate as to whether low-income nations should scale up financial reforms.

Keywords: Country heterogeneity, financial development, growth, nonlinearities, nonparametric regression, irrelevant variables.

1 Introduction

Empirical evidence indicating that the development of the financial sector of a country greatly facilitates its economic growth is abundant (e.g., Demirguc-Kunt and Maksimovic, 1998; King and

*†Daniel J. Henderson, Department of Economics, State University of New York, Binghamton, NY 13902-6000, 607-777-4480, Fax: 607-777-2681, E-mail: djhender@binghamton.edu. ♯Chris Papageorgiou, Corresponding Author, Strategy, Policy and Review Department, International Monetary Fund, Washington, DC 20431. Phone: 202-623-6107, Fax: 202-589-6107, E-mail: cpapageorgiou@imf.org. ‡Christopher F. Parmeter, Department of Economics, University of Miami, Coral Gables, FL, 33124. Phone: 305-284-4397, Fax: 305-284-2985, E-mail: cparmeter@bus.miami.edu. We thank Mark Steel for comments that improved the manuscript, Steven Durlauf and Ross Levine for sharing data as well as Patricio Venezuela for excellent research assistance. The views expressed in this study are the sole responsibility of the authors and should not be attributed to the International Monetary Fund, its Executive Board, or its management.

Levine, 1993; Jayarathe and Strahan, 1996; Rajan and Zingales, 1998, Beck, Levine and Loayza, 2000; Carlin and Meyer, 2003). The broad consensus emerging from the vast amount of work is that improving the operating financial environment and mitigating the level of financial regulation can result in a significant growth performance (see e.g., Levine 2005). However, many countries display underdeveloped financial sectors (Rajan and Zinagles, 2003).

We add to this literature by addressing several empirical issues that have been levied against growth regressions in general to provide a robust perspective on how financial development affects economic growth. Moreover, while many of our key results can be linked to theoretical models explicating a positive link between the level of financial development and economic growth, we uncover an ambiguous effect for countries with extremely limited financial development, suggesting a threshold type effect similar to that found in Aghion, Howitt and Mayer-Foulkes (2005). [1]

One way of empirically assessing the importance of financial development on economic growth would be to examine its impact in the context of other growth determinants using Bayesian Model Averaging (BMA). Since Brock and Durlauf (2001) and Fernandez, Ley and Steel (2001a,b) incorporated BMA in growth regressions the model averaging methodology has been making its mark as a constructive tool in growth empirics.[2] To date, studying the impact that financial development has on growth has has yet to be included in a BMA growth analysis. However, as shown by Ciccone and Jarocinski (2010), the promise of model averaging methods depends crucially on further research seeking answers to concerns related to the sensitivity of results. In this paper we take an alternative approach by examining the empirical content of financial development within a model of economic growth using the most up-to-date nonparametric instrumental variable methods.

The use of nonparametric methods to analyze financial (Ongena and Smith, 2001; Boudoukh, Richardson, Shen and Whitelaw, 2007; Ludvigson and Ng, 2007; Yung, Colak and Wang, 2008; Duchin, Osbas and Sensoy, 2010) and growth (Liu and Stengos, 1999; Durlauf, Kourtellos and

[1]This ambiguous effect has also been touched upon in the work of Deidda and Fattouh (2002) and Rioja and Valev (2004 a,b) though in an admittedly *ad hoc* and strictly linear context.

[2]Model averaging in growth empirics has become common practice; see Sala-i-Martin, Doppelhofer and Miller (2004), Ley and Steel (2007), Durlauf, Kourtellos and Tan (2008), Masanjala and Papageorgiou (2008), and Ciccone and Jarocinski (2010), just to name a few.

Minkin, 2001; Maasoumi, Racine and Stengos, 2007; Henderson, Papageorgiou and Parmeter, forthcoming) data is becoming increasingly popular. It has been recognized that misspecification of functional form can have both a detrimental impact on policy prescriptions as well as general understanding of the underlying economic structure. We add to both of these strains within the applied nonparametric literature. Specifically, this paper is the first to account for endogeneity in growth regressions while in a fully nonparametric framework. More generally, this is only one of a handful of applications of nonparametric instrumental variable estimators in economics.

The methodology that we deploy to produce robust conclusions about the financial development-growth nexus stems from recent advances in the nonparametric regression literature. Contemporary techniques have shown how to improve estimator efficiency in the face of continuous and discrete regressors. Individual and joint tests have been developed for significance testing of regressors. Tests of appropriate parametric specification are available as well. Additionally, we are able to handle the presence of irrelevant variables that are mistakenly included in an empirical analysis. Finally, in this paper we consider a newly developed estimator that can handle endogenous regressors via instrumental variable techniques. In addition to being the first paper that we know of to apply this technique, we further augment this estimator by considering estimation with discrete regressors, developing a bootstrap procedure to construct confidence intervals for our estimates as well as proposing a novel method to pick instruments in a nonparametric framework. This ensemble of modifications enables us to assemble an empirical study that is robust to functional form misspecification, admits heterogenous partial effects across covariates, and provides valid inferential statements.

Employing these state-of-the-art estimators, we find that financial development impacts growth in a highly nonlinear fashion. Our main results are twofold: a) it is shown that, on average, the impact of financial development on growth has increased over time and b) most importantly, however, more developed nations benefited most favorably from financial development while low-income countries did not show any benefits from improvements in their financial sectors.

The results of our study can be compared to several existing studies. In the Rajan and Zingales (1998) context, our results suggest that low-income nations are likely to grow from sources that do not require financial deepening. However, as they make their way to emerging markets, the role of finance becomes ever more important. In the context of Rostow's (1960) stages of development, this paper provides fresh evidence that in the initial stages of development, finance may not be a key determinant of a country's ability to grow but in later stages could play a crucial part. Finally, in the context of the ongoing policy debate regarding the relevance, timing and effect of structural reforms, our findings suggest rethinking whether low-income nations should scale up financial reforms early in the reform process if at all. This result echos the sentiments of Acemoglu, Johnson, Querubín and Robinson (2008) who find that reforms in developing countries often fail to provide the desired effect.

The remainder of this paper is organized as follows: Section 2 presents an intuitive description of our nonparametric kernel regression methods. Section 3 briefly discusses the data used for this study while Sections 4 and 5 reports our parametric and nonparametric results, respectively. Conclusions and potential extensions are discussed in Section 6. Technical aspects of the estimation and inference procedures used are contained in the appendix.

2 Empirical Methodology

Nonparametric kernel regression is becoming an increasingly popular method of estimation in applied economic milieus. The main perceived benefit is that it allows for consistent estimation when the underlying functional form of the regression function and/or errors are unknown. While this is true, there are many other benefits which may prove to be just as useful in our context. In this section we will discuss nonparametric regression, and address the issue of bandwidth selection which can expose irrelevant covariates and detect linearity of others. Finally, we will introduce nonparametric methods which can handle instrumental variables.

2.1 Estimation

Arguably the most popular regression model in the growth literature is the linear parametric model

$$y_i = \alpha + \beta x_i + u_i, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where y_i is our left-hand-side variable (output growth), x_i is a vector of q regressors, α and β are unknown parameters to be estimated and u_i is the random disturbance. Estimation of this model requires that all relevant regressors are included in x_i and the functional form is correctly specified. However, when either of these two assumptions do not hold, the estimates the model produces will most likely be inconsistent. While non-linear functional forms are possible in a parametric framework, the data generating process still must be assumed *a priori*.

Kernel methods have the ability to alleviate many of the restrictive assumptions necessary in the parametric framework. Consider the nonparametric regression model

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

where $m(\cdot)$ is an unknown smooth function and the remaining variables are the same as before. Here, $m(\cdot)$ is interpreted as the conditional mean of y given x. Note that in the (linear) parametric setting above it is implicitly assumed that $E(y_i|x_i) = \alpha + \beta x_i$. Further note that the linear model is a special case of our nonparametric estimator and thus, if the true data generating process is indeed linear, then the nonparametric estimator will give results consistent with that model.

One popular method for estimation of the unknown function is by local-constant least-squares (LCLS) regression. The LCLS estimator of the unknown function is given as

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} y_i \prod_{s=1}^{q} K\left(\frac{x_{si}-x_s}{h_s}\right)}{\sum_{i=1}^{n} \prod_{s=1}^{q} K\left(\frac{x_{si}-x_s}{h_s}\right)}, \tag{3}$$

where $\prod_{s=1}^{q} K\left(\left(x_{si}-x_{s}\right)/h_{s}\right)$ is the product kernel and h_s is the smoothing parameter (bandwidth) for a particular regressor x_s (see Pagan and Ullah 1999). The intuition behind this estimator is that it is simply a weighted average of y_i. It is also known as a local average, given that the weights change depending upon the location of the regressors. We estimate the conditional mean function by locally averaging those values of the left-hand-side variable which are 'close' in terms of the values taken on by the regressors. The amount of local information used to construct the average is controlled by the bandwidth. We give a formal description of this estimator in Appendix A (also see Li and Racine, 2006).

2.2 Bandwidth selection

It is believed that the choice of the continuous kernel function matters little in the estimation of the conditional mean (see Härdle 1990) and that selection of the bandwidths is the most salient factor when performing nonparametric estimation. As indicated above, the bandwidth controls the amount by which the data are smoothed. Large bandwidths will lead to large amounts of smoothing, resulting in low variance, but high bias. Small bandwidths, on the other hand, will lead to less smoothing, resulting in high variance, but low bias. This trade-off is well known in applied nonparametric econometrics, and the 'solution' is most often to resort to automated determination procedures to estimate the bandwidths. Although there exist many selection methods, we utilize the popular least-squares cross-validation (LSCV) criteria. This method has been studied extensively and additionally has known desirable properties regarding the smoothing out of irrelevant variables. Specifically, the bandwidths are chosen to minimize

$$CV\left(h\right)=\frac{1}{n}\sum_{j=1}^{n}\left(y_{j}-\widehat{m}_{-j}(x_{j})\right)^{2}, \tag{4}$$

where $\widehat{m}_{-j}(x_j)$ is the commonly used leave-one-out estimator of $m(\cdot)$. Obviously, as the sample size grows and/or the number of regressors increases, computation time increases dramatically.

However, it is highly recommended that a bandwidth selection procedure is used as opposed to a rule-of-thumb selection, especially in the presence of discrete data as no rule-of-thumb selection criteria exists.

As an aside, we note that an even simpler bandwidth selection procedure, the ‘ocular’ method, is not appropriate once the number of covariates is larger than two. As the number of regressors exceeds two, visual methods to investigate the fit of the model are cumbersome and uninstructive. With a large dimension for the number of regressors, it is suggested that cross-validation techniques be used as opposed to either ocular or rule-of-thumb methods.

2.2.1 Irrelevant regressors

The bandwidths, by affecting the degree of smoothing, are not just a means to an end; they also provide some indication of how the left-hand-side variable is affected by the regressors. For instance, Hall, Li and Racine (2007) show that with LCLS, when the bandwidth on any regressor reaches its upper bound, the regressor is essentially ‘smoothed out’. In other words, the cross-validation procedure determines the bandwidths which (out-of-sample) predict the left-hand-side variable the best, and thus chooses weights such that irrelevant variables have no impact on the prediction of the left-hand-side variable.

An obvious question is whether or not in practice, these observations actually hit their upper bounds. For the continuous variables this is apparent. Computationally, no cross-validation procedure can give bandwidths equal to their upper bound of infinity. Thus a decision must be made on how large a bandwidth must be until it is considered irrelevant. Hall, Li and Racine (2007) suggest that when the bandwidth exceeds a few standard deviations of the regressor, that the variable be deemed irrelevant. For discrete regressors, their upper bounds are quite obtainable. However, we may also want to deem regressors irrelevant when they are ‘close’ to their upper bounds. In practice it is preferable to use a formal test. This is the approach we take.

2.2.2 Detecting linearity

In the local-linear least-squares (LLLS) framework discussed in Appendix A, as the bandwidth on a continuous regressor becomes large, the weight given to each observation becomes equal. In other words, as $h \to \infty$, the implication is that the regressor enters linearly. The logic is the following: as the bandwidth becomes infinitely large, the local-linear regression fits a linear line using all the points in the neighborhood of x . When all the points are used, then the line is the same for any x in the sample. Hence the estimate is linear.

This emphasizes the importance of obtaining a separate bandwidth for each regressor. If one regressor enters linearly we would expect that the cross-validation procedure should select large values of h for that regressor and relatively small values for regressors that enter nonlinearly. In practice, Li and Racine (2004) suggest that any bandwidth which is more than two standard deviations of the regressor be deemed a variable that enters linearly.

However, linearity in a particular regressor does not mean that we should necessarily switch to a semiparametric model for the sake of efficiency. In the multi-variable case, it may be that there are important interactions between the 'linear' regressor and the remaining variables in the model, implying that the partial effect of the 'linear' regressor may still vary across x. Moreover, linearity should be more formally assessed, when feasible, using statistical tests. There is no formal statistical test for linearity of a specific regressor, but there are methods to test for specific parametric structure which will be employed later.

2.3 Endogenous regressors

Although nonparametric methods have been applauded for various reasons, one criticism of them is that they do not easily handle endogenous regressors in applied settings. The nonparametric literature is not ignorant of this issue and researchers have been developing estimators which can handle endogenous regressors (e.g., Newey and Powell, 2003; Hall and Horowitz, 2005; Darolles, Fan, Florens and Renault, 2010). The problem with most of these estimators is that they require very specific assumptions on the data and/or the model. They also often use spline methods which makes it difficult to include discrete regressors. Recently, Su and Ullah (2008) developed a nonparametric estimator which can handle endogenous regressors in a kernel framework. Unfortunately, applied researchers who use nonparametric methods are largely unaware of this estimator. As far as we are aware, this is the first paper to attempt to apply the Su and Ullah (2008) estimator.

To develop the intuition of the Su and Ullah (2008) estimator consider the generic regression model in (2) but including a single endogenous regressor

$$\begin{aligned} y_i &= m(x_i, z_{1i}) + \varepsilon_i \\ x_i &= g(z_i) + u_i \end{aligned}$$

where y_i is the dependent variable, $m(\cdot)$ is the unknown smooth function of interest, x_i is the endogenous regressor, $z_i = (z_{1i}, z_{2i})$ where z_{1i} and z_{2i} are $d_1 \times 1$ and $d_2 \times 2$ vectors of instrumental variables respectively, $g(\cdot)$ is an unknown smooth function of the instruments z, and u and ε are disturbances. We assume that $E\left(u|z\right) = 0$, and $E\left(\varepsilon|z, u\right) = E\left(\varepsilon|u\right)$. These assumptions are more general than the strict requirement of z being independent of (u, ε) and allow both u and ε to be heteroscedastic.

In the standard case where $m(\cdot)$ and $g(\cdot)$ are known up to a finite number of parameters, we can simply use two-stage least-squares (or an appropriate nonlinear method). However, in practice, neither of these functions are generally known and misspecification of either function will likely lead to inconsistent estimates. LLLS estimation of $m(\cdot)$ is feasible based on the following insight:

$$\begin{aligned} E\left(y|x, z, u\right) &= m(x, z_1) + E\left(\varepsilon|x, z, u\right) \\ &= m\left(x, z_1\right) + E\left(\varepsilon|x - g\left(z\right), z, u\right) \\ &= m\left(x, z_1\right) + E\left(\varepsilon|z, u\right) \\ &= m\left(x, z_1\right) + E\left(\varepsilon|u\right). \end{aligned}$$

It thus follows by the law of iterative expectations that

$$w(x, z_1, u) \equiv E\left(y|x, z, u\right) = m(x, z_1) + E\left(\varepsilon|u\right).$$

This additive structure provides consistent estimates of $m(x, z_1)$ up to an additive constant ($E\left(\varepsilon|u\right)$) without further restrictions, following the procedure outlined in Su and Ullah (2008); a further simplification can be achieved if one is further willing to assume that $E(\varepsilon) = 0$, which we describe below.

1. Obtain a consistent estimate of $g(\cdot)$ by running a local-constant regression of x on z with kernel function $K_{1,\gamma}(\cdot)$ and bandwidth vector γ_1. Denote the estimates of the unknown function as $\widehat{h}(z)$ and obtain the residuals $\widehat{u}_i = x_i - \widehat{h}(z_i)$, for $i = 1, 2, \ldots, n$.

2. Obtain a consistent estimate of $w(\cdot)$ by running a local-linear regression of y on x, z_1, and $\widehat{u}$ with kernel function $K_{2,\gamma}(\cdot)$ and bandwidth vector γ_2. Denote the estimates of the unknown function as $\widehat{w}(x, z_1, u)$.

3. Assuming that $E(\varepsilon) = 0$, we can obtain a consistent estimate of $m(\cdot)$ as

$$\widehat{m}(x, z_1) = n^{-1} \sum_{i=1}^{n} \widehat{w}(x, z_1, \widehat{u}_i),$$

where $\widehat{w}\left(x, z_1, \widehat{u}_i\right)$ is the counterfactual estimate of the unknown function obtained using the bandwidths from the local-linear regression in step 2. Notice that what this last step is doing is estimating the value of the function $\hat{w}(x, z_1, \cdot)$ for every value of $\hat{u}$ and then averaging. Thus, the estimator consists of two estimations stages and then a final step consisting of counterfactual estimation to average out the error term, since we have assumed that is has mean zero.

The derivatives of $m(\cdot)$ can be obtained similarly as

$$\widehat{m}'(x, z_1) = n^{-1} \sum_{i=1}^{n} \widehat{w}'(x, z_1, \widehat{u}_i),$$

where $\widehat{w}'(x, z_1, \widehat{u}_i)$ is the counterfactual derivative of $w(\cdot)$.

For implementation we follow the suggestion of Su and Ullah (2008) and use a local-polynomial least-squares kernel estimator in each of the first two steps. Specifically, we choose the local-linear least-squares estimator for its relative ease of estimation and precision. For more on local-polynomial regression in general, the reader is directed to Fan and Gijbels (1996). We then take their suggestion to extend the estimator to the case where the data in (x, z) may be a mixture of continuous and discrete variables, thus connecting their estimator to the methodology of Racine and Li (2004).

The basic idea of this estimator is similar to two-stage least-squares (2SLS). The first stage requires a nonparametric regression of the endogenous regressor on all the exogenous variables. The second stage is somewhat different in that it requires a nonparametric regression of the y variable on each of the regressors in the model, *including* the endogenous regressor (not a predictor of it) and the residuals from the first stage. A third stage is then required (marginal integration) so that we can ensure that the mean zero assumption of the error term holds.

As suggested in the conclusion Su and Ullah (2008), we extend their estimator to handle discrete regressors. We do so by employing discrete kernels as outlined Racine and Li (2004). In addition, we consider a novel way to choose instruments in a nonparametric framework. Specifically, recall from Section 2.2.1 that when the bandwidth on a regressor hits its upper bound in a LCLS regression, that variable is said to not be a predictor of the left-hand-side variable. If this is the case, we can use a LCLS regression of y on all the potential regressors in order to determine which variables are unrelated to y. Once these have been determined, and their correlation with the endogenous regressor checked, we can use them in our first stage regression as instruments. Alternatively, we could employ other instruments standard in the literature, such as lagged values of the regressors.

We note that our primary analysis will ignore the potential endogeneity of financial development. The reason for this is two-fold. First, as our results in Section 5.4 will show, the qualitative results of our experiment are not dependent on whether or not we control for endogeneity. Second, the instruments we propose require constraints on the data which would leave us with a smaller sample of countries if we focused on the IV results solely. The reduced number of countries comes at the additional price of mostly coming from developing countries and it is often those countries which we have the most interest in studying.

3 Data

In this section we briefly discuss the data used in our estimation. Our data come from two sources. The typical growth variables motivated by the Solow (1956) model are from Durlauf, Kourtellos and Tan (2008) and include per capita real GDP, investment defined as the ratio of average investment to GDP, education defined as the average percentage of working age population (population between the age of 15 and 64) in secondary education, and the average growth rate of the working age population.[3]

The financial development variables used are from the Financial Structure Dataset (revised version September 14, 2006) based on Beck, Demirgüç-Kunt and Levine (2000). The dataset improves on previous efforts by presenting information on the public share of commercial banks and by introducing indicators of the size and activity of non-bank financial institutions. The dataset is particularly novel in that it allows for a comprehensive assessment of the development, structure and performance of the financial sector in growth empirics. They can also be used to analyze the public share of commercial banks, by introducing the implications of financial structure for

[3]The Durlauf, Kourtellos and Tan (2008) data set contains data for the traditional Solow model (initial income, investment rate, human capital, population growth) as well as variables that compose several of the contending growth theories being debated today: fractionalization, institutions, demographics, geography, religion, and macroeconomic policy.

economic indicators of the size and activity of non-bank financial growth. [4] Although we have considered several proxies for financial development, our baseline results are based on the following four variables:

- $DBACBA$ defined as " Deposit Money Bank Assets / (Deposit Money + Central) Bank Assets"
- $DBAGDP$ defined as " Deposit Money Bank Assets / GDP"
- $PCRDGDP$ defined as " Private Credit by Deposit Money Banks / GDP"
- $BDGDP$ defined as " Bank Deposits / GDP"

where deposit money banks, include financial institutions that have " liabilities in the form of deposits transferable by check or otherwise usable in making payments" (Beck, Demirgüç-Kunt, and Levine, 2000, pp. 4), assets refers to total domestic financial intermediation that the respective intermediary performs, and private credit captures the financial intermediation with the private non-financial sector. It is worth noting that the first proxy of financial development, which equals the ratio of deposit money banks assets and the sum of deposit money and central bank assets (DBACBA), has been used most extensively in the literature with the pioneering contributions of King and Levine (1993a,b).

Merging data on the typical growth variables with data on these four financial development variables obtains an unbalanced 5-year non-overlapping panel dataset. The total number of country-year observations is 676 from 101 countries starting in 1960 and ending in 2000. The entire dataset used in the empirical estimation is available from the authors upon request.

[4] See Beck, Demirgüç-Kunt, and Levine (2000) and http://go.worldbank.org/X23UD9QUX0 for detailed description of the sources and construction of these different indicators.

4 Parametric estimates

Table 1 presents OLS estimates of three linear growth models which form the basis of our comparison for nonparametric analysis. We estimate a bevy of standard growth regressions. Models I, II, and V pool the data and include region and time fixed effects, Models III and VI employ least-squares dummy variables to incorporate country specific fixed effects and models IV and VII use a two-way panel structure to account for country and time specific effects. Our primary financial variable, $ln(DBACBA)$ is positive and significant in all four panel specifications. However, when used in the pooled regression as a stand along variable for financial market development, it is insignificant. Our panel results for initial income are to be expected as past studies have found that the 'convergence coefficient' is much larger when switching from cross-sectional methods to panel (Islam, 1995; Casselli, Esquivel and Lefort, 1997; Durlauf, Johnson and Temple, 2005, sec. VI(ii)). As expected, none of the additional proxies for financial market development are significant at even the 10% level suggesting that $\ln(DBACBA)$ is the best suited of our measures to attempt to uncover correlations.

It is interesting to point out that the perceived importance of human capital is heavily dependent on model choice. In our pooled cross-section results, we have significant positive values, but when we employ panel methods we lose significance and sign. Previous studies have found similar impacts and have also attempted to explain why/when human capital may be expected to have positive and significant impacts on growth (Benhabib and Spiegel, 1994). We note that our measure of human capital is evolving slowly over time and as such estimation methods which use fixed effects are likely to wipe-out almost all of the expected impact human capital is likely to have in a cross-sectional setting.

Before moving further, we note that the addition of the extra financial market proxies in the pooled setting (Model V) suggest that our main financial measure, $\ln(DBACBA)$ is positive and significant. None of the other financial measures are significant. This suggests that the countries we dropped, all developing countries, may have an influence on the results we found in Model II, similar to the findings of Rioja and Valev (2004a) that suggest at low-levels of development,

Table 1: OLS and Panel estimates for several parametric specifications.

Variable	Model I	Model II	Model III	Model IV	Model V	Model VI	Model VII
$ln(Y_0)$	-0.043	-0.048	-0.183	-0.143	-0.051	-0.188	-0.138
	(0.000)	(0.000)	(0.016)	(0.017)	(0.000)	(0.023)	(0.021)
$ln(School)$	0.033	0.035	0.020	-0.001	0.036	0.000	-0.013
	(0.003)	(0.001)	(0.016)	(0.019)	(0.009)	(0.024)	(0.026)
$ln(Inv)$	0.044	0.400	0.058	0.125	0.035	0.054	0.131
	(0.000)	(0.001)	(0.018)	(0.018)	(0.013)	(0.024)	(0.024)
$ln(Pop\ Growth)$	-0.074	-0.076	-0.007	-0.006	-0.133	-0.076	-0.106
	(0.052)	(0.045)	(0.054)	(0.053)	(0.003)	(0.061)	(0.058)
$OECD$	-0.010	-0.009	—	—	-0.023	—	—
	(0.020)	(0.020)	—	—	(0.027)	—	—
$ln(DBACBA)$	—	0.024	0.075	0.039	0.066	0.135	0.105
	—	(0.140)	(0.022)	(0.022)	(0.002)	(0.034)	(0.034)
$ln(DBAGDP)$					-0.019	-0.043	-0.027
					(0.452)	(0.039)	(0.040)
$ln(PCRDBGDP)$					-0.003	0.015	0.005
					(0.882)	(0.029)	(0.029)
$ln(BDGDP)$					0.033	0.048	0.021
					(0.088)	(0.028)	(0.028)
Region/Time	Yes	Yes	No	No	Yes	No	No
Individual	No	No	Yes	No	No	Yes	No
Individual/Time	No	No	No	Yes	No	No	Yes
$\bar{R}^2$	0.258	0.260	0.251	0.170	0.266	0.247	0.171
n	676	676	676	676	528	528	528

financial market amelioration impedes, or at least does not improve, growth. The addition of these types of countries to increase our sample size for Models I-IV could decide whether or not our results are significant.

5 Nonparametric estimates

We break up our nonparametric results into various sections to highlight the importance of each benefit. We first discuss the implications of our bandwidth estimates for several specifications involving our financial market proxies. When then discuss a set of baseline estimates from our nonparametric regression model. Next we discuss the partial effects for specific groups of countries as well as how the estimates evolve over time. Finally, we discuss the implications of controlling for endogeneity in our nonparametric model.

5.1 Bandwidths

Prior to discussing the partial effects, Table 2 presents both local-constant and local-linear cross-validated bandwidths for three distinct growth models. Following our discussion of the financial development variables in Section 3 we consider a Solow-type model with no financial variables, a model with only one financial proxy as well as a model that incorporates all of our available financial proxies.

The first column in Table 2 gives the upper bound for the bandwidth for each regressor. We choose to list two times the standard deviation instead of the upper bound of infinity for the continuous regressors as this upper bound is infeasible to obtain in practice. Although we use the rule-of-thumb analysis suggested by Hall, Li and Racine (2007) for a first-round analysis, we also use formal tests below. The second column of numbers are the bandwidths from the LCLS regression. We again note that when a bandwidth (for either type of regressor) reaches its upper bound in the LCLS framework, that variable is deemed irrelevant. Here we see that the corresponding bandwidths for the schooling and population growth variables are very large. This suggests that both of these variables may be irrelevant in predicting output growth. This may be expected because we often find these variables to be insignificant in standard parametric growth regressions. At the same time, the bandwidth for OECD hits its upper bound. This is also expected as we have additionally included a variable for region.

The bandwidths for Model I which are obtained by LSCV of the LLLS regression are given as the third column of numbers in Table 2. Recall that for the LLLS regression, when a continuous regressor hits its upper bound, it is considered to enter the regression linearly (holding all else constant). Here we see that the bandwidth for the population growth variable is well beyond two times the standard deviation of the regressor and thus it is very possible that this variable enters in linearly. However, recall that this linearity is not synonymous with constant partial effects. There

Table 2: Bandwidths for various growth models. UB is the upper bound for a given regressor. In the case of the continuous regressors, the upper bound is two times the standard deviation of that variable. For the categorical variables, the upper bound is the true upper bound.

	Model I			Model II			Model III		
Variable	UB	LCLS	LLLS	UB	LCLS	LLLS	UB	LCLS	LLLS
$ln(Y_0)$	2.178	0.188	0.821	2.178	0.223	0.758	2.106	0.612	0.280
$ln(School)$	1.510	17622920	0.249	1.510	0.337	0.220	1.362	12.93	277623
$ln(Inv)$	1.220	0.482	0.624	1.220	0.351	0.573	1.158	0.291	1034410
$ln(Pop\ Growth)$	0.352	298507	19112	0.352	2337505	99781	0.338	42826	207124
$ln(DBACBA)$				0.704	0.422	0.405	0.628	0.188	0.251
$ln(DBAGDP)$							1.558	2249826	30455
$ln(PCRDBGDP)$							1.686	10651825	0.511
$ln(BDGDP)$							1.382	1996485	0.212
OECD	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.000	0.500
Region	0.900	0.100	0.698	0.900	0.487	0.727	0.900	0.678	0.863
Time	1.00	0.931	0.913	1.000	0.930	0.730	1.000	0.255	1.000
n		676	676		676	676		528	528

could be important interactions which would prohibit us from switching to a semiparametric model. Finally, we see similar results for the bandwidths for the categorical regressors. The OECD variable is deemed irrelevant while both the region and time variables are below their upper bounds.

In the single finance variable (Model II) case, the bandwidths reveal three salient points. First, the results from the LSCV procedure of the LCLS estimator show that the bandwidth for ln($Pop.Growth$) far exceeds two times its standard deviation. Since the relevance of the variable disappears as the bandwidths approach infinity, this suggests that population growth is irrelevant in predicting output growth. Second, the bandwidth on the OECD variable equals its upper bound of 0.5. This shows that this variable plays no role in the prediction of output growth. Again, this result should not be surprising here because we have a separate variable for region. When region is excluded from the analysis, the OECD variable is deemed relevant. Third, the remaining bandwidths are much smaller than their respective upper bounds, implying that these variables are relevant in the model.

For the bandwidths selected via LSCV for the LLLS estimator, only population growth has a bandwidth which is larger than two times its standard deviation. This suggests that this variable enters the model linearly. This is somewhat common when the LCLS bandwidths suggest the variables are irrelevant. Further, note that the remaining continuous variables all have bandwidths that are relatively small and thus a simple linear model would not be suggested here. This includes the financial development proxy. This is in contrast to the result found in Ketteni, Mamuneas, Savvides, and Stengos (2007) who suggest that this variable enters linearly. One possible reason the results differ is that their model did not allow for interactions with the financial proxy. We reject the linear model via the Hsiao, Li and Racine (2007) test[5] (p-value = 0.001).[6] In other words, assuming a linear parametric model would most likely lead to inconsistent estimates and incorrect inference.

For the model with four finance variables (Model III), the bandwidths differ somewhat in terms of their implications. First, the LCLS bandwidths for the three additional finance variables are shown to be far larger than two times their standard deviation. What this says is that each is irrelevant in the prediction of output growth. Another way to interpret this is that the first finance variable sufficiently explains the variation in the left-hand-side variable and thus the other finance variables are not necessary for prediction purposes. A second difference is that now the bandwidth on $\ln(School)$ is now larger than twice its standard deviation, and as the Solow model suggests, this variable may be irrelevant as well. Third, the LCLS bandwidth for OECD is now equal to 0.000. This implies that observations with different values of this covariate are given (essentially) no weight in the estimation; the kernel reduces to an indicator function. In other words, the estimation is equivalent to dividing the sample into OECD and non-OECD samples and performing separate nonparametric regressions.[7]

LSCV of the LLLS estimator now provides several bandwidths which are much larger than two times their standard deviations. Here, $\ln(School)$, $\ln(Inv)$, $\ln(Pop.Growth)$ and $\ln(DBAGDP)$

[5] A detailed description of the test is given in Appendix C.

[6] This test result and all others were computed using 399 bootstrap replications

[7] While this possibility is feasible, the conflict with our earlier regression results which suggests that OECD is irrelevant means that this result should be considered suspect. When we obtain differing results between LCLS and LLLS bandwidths for discrete regressors we generally defer to the LLLS bandwidths as LLLS estimation is more reliable in practice. We still need the LCLS regression estimates however, as only LCLS has the ability to detect prediction relevance for continuous regressors.

each have bandwidths which suggest linearity. However, recall that linearity is not synonymous with homogeneous effects of the covariates. The variables may enter linearly, but there may also exist important interactions which simple, linear in parameters models may not account for. Also, not all the variables enter linearly (specifically the two main variables of interest, $\ln(Y_0)$ and $\ln(DBACBA)$) and thus a simple linear specification, even with interactions, may not be appropriate. Again, this is confirmed by the Hsiao, Li and Racine (2007) test which rejects the parametric model (p-value = 0.000). Finally, note that the OECD and time bandwidths each hit their upper bounds of 0.5 and 1.00 respectively. This suggests that in this model neither is important in the prediction of output growth.

In sum, examination of the bandwidths suggest that some variables enter the model nonlinearly and some variables enter the model linearly. While the first assumption made in the majority of empirical analyses, linearity, receives the most attention, heterogeneity may be just as problematic. At this point it is premature to determine which specification is more appropriate, but the model with four financial variables suggests that the additional three of them are irrelevant, whereas in both models II and III, the bandwidths suggest that $\ln(DBACBA)$ is relevant. We now turn to the actual results, as well as formal statistical tests.

5.2 Baseline estimates

Table 3 presents LLLS estimates of the partial effects for the continuous covariates across all three models.[8] We present nonparametric estimates corresponding to the 25^{th}, 50^{th}, and 75^{th} percentiles of the estimated parameter distributions (labelled $Q1$, $Q2$, and $Q3$) along with their corresponding bootstrap standard errors.

The results for the Solow model are listed first. The partial effects on the initial income variable are negative and significant for the first quartile and median. Perhaps more interesting than the significance is the large amount of variation in the partial effects. Note that the result at the first

[8] We choose to present the LLLS estimates as this estimator has more desirable properties as compared to the LCLS estimator. See Li and Racine (2006) for more details.

Table 3: LLLS quartile coefficient estimates from the three growth models. Bootstrap standard errors are in parentheses below each estimate.

	Model I			Model II			Model III		
Variable	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
$ln(Y_0)$	-0.086	-0.043	-0.014	-0.106	-0.058	-0.019	-0.135	-0.094	-0.037
	(0.014)	(0.018)	(0.022)	(0.054)	(0.048)	(0.029)	(0.058)	(0.030)	(0.048)
$ln(School)$	-0.042	0.007	0.054	-0.046	0.023	0.074	-0.030	0.018	0.057
	(0.038)	(0.034)	(0.039)	(0.094)	(0.098)	(0.053)	(0.078)	(0.110)	(0.033)
$ln(Inv)$	0.056	0.082	0.116	0.023	0.062	0.104	0.040	0.057	0.0910
	(0.030)	(0.025)	(0.021)	(0.050)	(0.057)	(0.072)	(0.023)	(0.022)	(0.018)
$ln(Pop\ Growth)$	-0.205	-0.108	-0.017	-0.271	-0.132	-0.040	-0.310	-0.112	0.003
	(0.115)	(0.055)	(0.044)	(0.133)	(0.128)	(0.121)	(0.254)	(0.064)	(0.052)
$ln(DBACBA)$				0.016	0.069	0.122	0.018	0.114	0.226
				(0.102)	(0.068)	(0.096)	(0.195)	(0.098)	(0.210)
$ln(DBAGDP)$							-0.125	-0.047	0.009
							(0.100)	(0.075)	(0.051)
$ln(PCRDBGDP)$							-0.028	0.014	0.063
							(0.060)	(0.078)	(0.051)
$ln(BDGDP)$							-0.024	0.043	0.111
							(0.154)	(0.105)	(0.070)
R^2		0.630			0.772			0.839	

quartile is nearly six times smaller than the partial effect at the third quartile. Further, the results for the schooling variable are insignificant as expected and the results for the investment variable are generally positive and significant. For the last variable, population growth, the partial effects at the first quartile and median are negative and significant, while the partial effect at the upper quartile is negative, but insignificant. Finally, the $R^2 = 0.630$ is nearly three times as large as in the corresponding parametric model. This suggests that there may be a lot to gain in terms of prediction by employing the nonparametric model.

The signs for Model II are as expected. First, the partial effects on $\ln(Y_0)$ are negative but insignificant at the median and $Q3$. Second, the partial effects on $\ln(School)$ are small and insignificant for all three metrics. Next, the partial effects on $\ln(Inv)$ are all positive, but again insignificant across metrics. Fourth, the partial effects on $\ln(Pop.Growth)$ are negative and insignificant. Finally, for $\ln(DBACBA)$, the estimates are positive, but insignificant across measures.

Although we note that generally the signs are similar to those of parametric studies and the nonparametric model I, we find that many of our results are insignificant. The insignificant results are alarming at first glance. However, as we will soon witness, the significance of a particular partial effect depends upon the characteristics of the country. Here we point out that there is significant variation in the parameter estimates for a single variable. Figure 1 shows kernel density estimates for the parameter estimates across countries for each variable. For instance, the partial effect at the third quartile for the finance variable is nearly eight times larger than at the first quartile. In addition, we formally test that the finance variable is irrelevant in the estimation of the dependent variable by employing the Lavergne and Vuong (2002) test.[9] We reject the null that the finance variable is irrelevant using either the bandwidths from the LCLS regression (p-value = 0.000) or the LLLS regression (p-value = 0.000).

The results for Model III differ modestly. The partial effects for the same five variables in Model II are qualitatively similar. A key difference is the change in significance of $\ln(Y_0)$ and $\ln(Inv)$ at the median. The additional three finance covariates offer little in terms of statistical or economic significance. Here we see that none of the partial effect estimates are significant. Again, this is not surprising given the findings of the previous sub-section. From this, there appears to be no reason to include the additional finance variables outside of $\ln(DBACBA)$. There is mixed evidence from the Lavergne and Vuong (2000) joint significance test of the three additional finance variables. We fail to reject the null that they are irrelevant using the LCLS regression bandwidths (p-value = 0.282), but reject the null using the LLLS regression bandwidths (p-value = 0.008).[10] These findings may still be intuitive as the LCLS bandwidths can show irrelevance while the LLLS bandwidths can show linearity.

A traditional parametric analysis usually draws to a close at this point. Fortunately, in our case, we have a partial effect for each variable, for each country in each time period. Therefore we

[9] A detailed description of the test is given in Appendix D.

[10] Finally, we test the null that all four finance variables are irrelevant. We fail to reject the null that they are jointly irrelevant using the bandwidths from the LCLS regression (p-value = 0.124), but reject the null of joint significance with bandwidths from the LLLS regression (p-value = 0.000). The former result is peculiar because this null was rejected with a single variable. This lack of power is likely due to the relatively small sample. Our model with all four finance variables is nearly 20% smaller than our model with just $\ln(DBACBA)$.

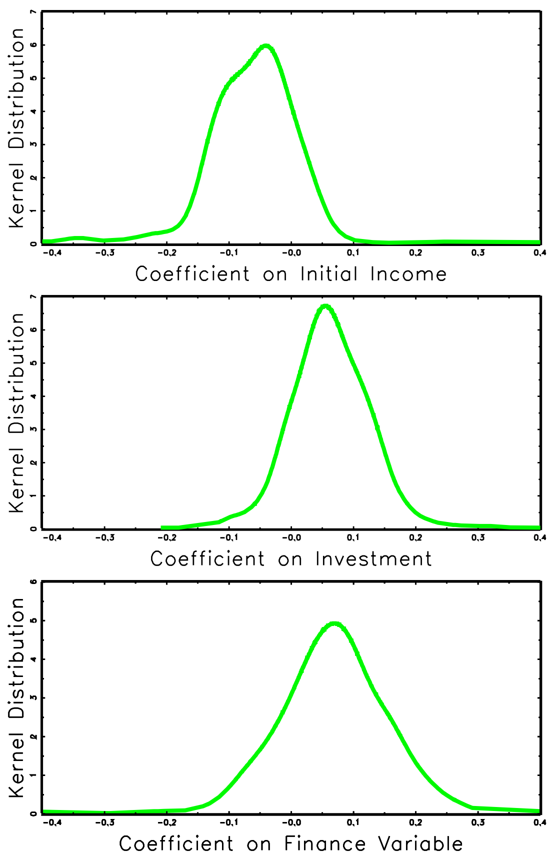

Figure 1: Kernel Density Plots of LLLS Estimated Coefficients from Model II

can examine further the partial effects across suitably defined groups.

5.3 Partial effects for particular groups

Even though we found that many of the results were insignificant for the single finance variable model (Model II), we choose to focus on this particular model for several reasons. First, the Lavergne and Vuong (2000) test was unable to reject the null that the finance variable $\ln(DBACBA)$ was irrelevant. Second, the same test failed to reject the null that the three additional finance variables were irrelevant in the LCLS case. This was further emphasized by their respective bandwidths in Table 2. Third, given the curse of dimensionality in nonparametric regression, it is important to try to limit the number of continuous regressors to only those that are relevant in predicting the left-hand-side variable. Finally, when we include the other three variables, we get a substantial reduction in the sample size.

Table 4 gives the nonparametric estimates corresponding to the median of the distributions for the estimated LLLS partial effects for every continuous variable in model II across splits on every variable's median in the model. That is, we find the median partial effect of the logarithm of initial income, say, for all countries with population growth above the sample median and then do the same for all countries with population growth below the sample median.

Table 4: Median coefficient of LLLS estimates from Model II for each continuous regressor across specific groups. Bootstrap standard errors are in parentheses below each estimate.

Split/Variable	$ln(Y_0)$	$ln(School)$	$ln(Inv)$	$ln(Pop\ Growth)$	$ln(DBACBA)$
> median(ln(Y_0))	-0.084	0.046	0.070	-0.093	0.080
	(0.071)	(0.123)	(0.027)	(0.025)	(0.034)
< median(ln(Y_0))	-0.035	-0.007	0.054	-0.220	0.046
	(0.037)	(0.088)	(0.039)	(0.045)	(0.031)
> median(ln($School$))	-0.087	0.048	0.075	-0.101	0.079
	(0.062)	(0.069)	(0.025)	(0.071)	(0.077)
< median(ln($School$))	-0.035	-0.010	0.049	-0.191	0.042
	(0.035)	(0.167)	(0.080)	(0.045)	(0.047)
> median(ln(Inv))	-0.089	0.049	0.090	-0.083	0.075
	(0.035)	(0.064)	(0.057)	(0.027)	(0.039)
< median(ln(Inv))	-0.034	-0.018	0.042	-0.224	0.058
	(0.031)	(0.229)	(0.031)	(0.046)	(0.042)
> median(ln($Pop.Growth$))	-0.036	-0.010	0.048	-0.213	0.046
	(0.040)	(0.052)	(0.036)	(0.035)	(0.037)
< median(ln($Pop.Growth$))	-0.085	0.049	0.075	-0.100	0.083
	(0.152)	(0.092)	(0.047)	(0.039)	(0.029)
> median(ln($DBACBA$))	-0.080	0.041	0.069	-0.092	0.088
	(0.027)	(0.053)	(0.056)	(0.275)	(0.087)
< median(ln($DBACBA$))	-0.041	0.002	0.055	-0.197	0.046
	(0.034)	(0.078)	(0.041)	(0.093)	(0.091)

5.3.1 Initial income

The results across different covariates are interesting, but the main purpose of most growth studies is to examine the partial effect estimate on the initial income variable. In most studies, a single partial effect, the partial effect estimate, is obtained for this variable and its sign determines whether or not convergence exists across the sample. Here we obtain a separate partial effect for each country in each time period. Thus, we can examine the partial effects among pre-specified groups.

The results for the first column of numbers correspond to the partial effect of the initial income variable. The partial effects are negative and significant at the median for those observations where the initial income was above the median, the level of investment was above the median and the finance variable was above the median. It is not surprising that when the level of schooling is above the median, the median partial effect on initial income is smaller at the median than when the level of schooling is below the median. However, this partial effect is insignificant. At the same time, when the level of population growth was above the median, the partial effect on the initial income variable was smaller on average than the partial effect on initial income for an observation with population growth below the median. Figure 2 shows the plots of each of these distributions of parameter estimates. In each case we can see that the means of the distributions differ as well as the variances. The differences are confirmed by the Li (1996) test[11] for equality of two unknown distributions. In each case we reject the null that the distributions of the comparisons are equal. Specifically, each test has a p-value which is zero to three decimal places.

The results are also broken down for the initial income partial effect by geographical region in Table 5. Recall that the results from Table 3 showed that the partial effects were significant only at the first and second quartiles. Thus, it makes sense that some groups may have insignificant results. Here we see that the partial effects are significant at the median solely for OECD and North African/Middle Eastern countries. The other groups of countries have insignificant results for the median partial effect for their particular group.

[11] A detailed description of the test is given in Appendix E.

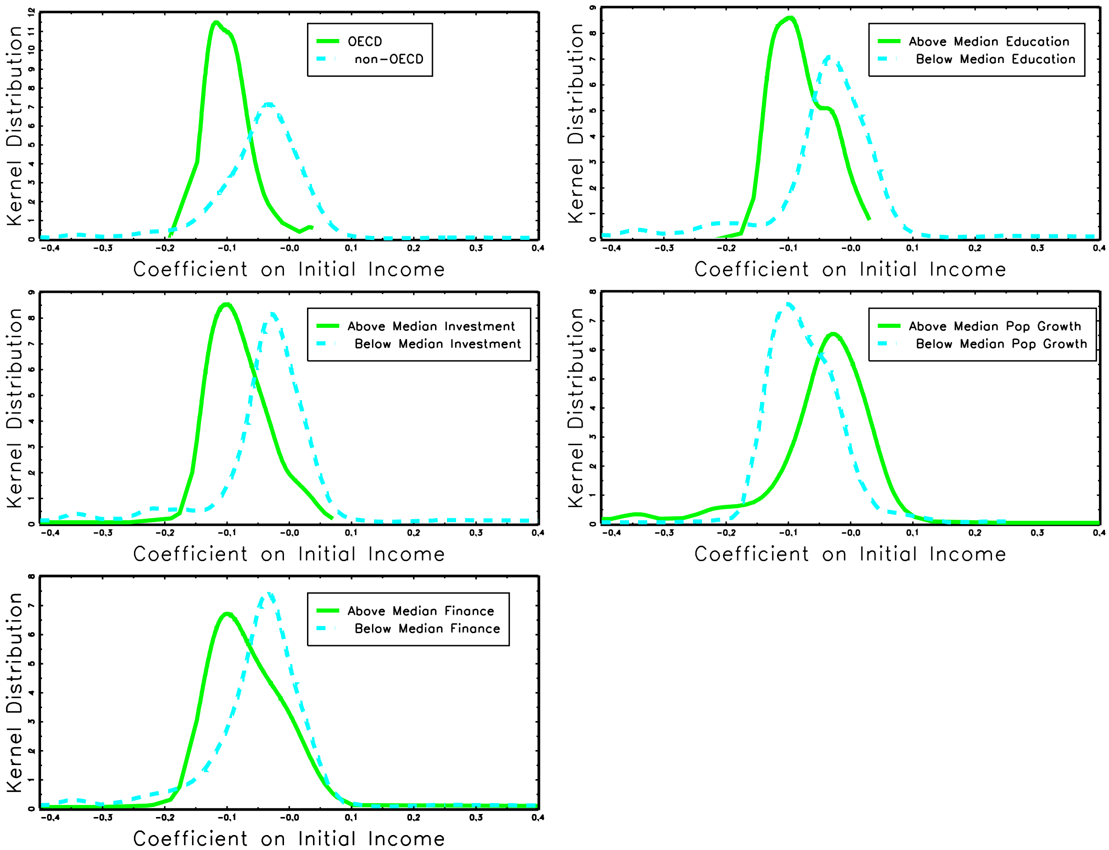

Figure 2: Comparison of LLLS Estimated Coefficients of Initial Income from Model II.

5.3.2 Conditioning variables

Regardless of the split, Table 4 shows us that the median marginal effect of additional average schooling is insignificant. This is not surprising since our bandwidth estimates from this model suggest that $\ln(School)$ is 'smoothed out'. This is not the case when looking at the median marginal effects for either $\ln(Inv)$ or $\ln(PopGrowth)$. We see that positive and significant impacts related to capital investment appear for countries with above median initial income and years of schooling. Splits based according to median population growth and financial market access reveal no impact at the median for the logarithm of investment.

Focusing our attention on population growth impacts on the growth process we see negative and significant impacts almost across the board. For initial income, investment and population growth, the median effect of population growth has a significant adverse affect on economic growth above and below the median with the above median estimates being smaller for initial income and investment. The opposite impact is found with respect to the population growth split. Countries with above median population growth witness a more than double impact on economic growth for an increase in the rate of increase in the population as opposed to those countries currently below the median level of population growth; a truly Malthusian effect.

For countries with above median levels of financial market access, we see that all three of our conditional variables from the Solow model are insignificant at the median. This is not suggestive, however, that there exist no interactions between the classical Solow conditioning variables and financial market access; the median marginal effect of $\ln(DBACBA)$ is positive and significant for above median levels of investment and below median levels of population growth. What these results do imply however is that across a broad range of splits, the median impact of the additional Solow controls have varied insights. This suggests that a one-size-fits all linear regression is inappropriate for modelling the growth process.

5.3.3 Financial development proxy

The results for the breakdown of the financial development variable are in the final column of Table 4. Here we see that those observations where the initial income is above the median have larger returns to the finance variable on average. The partial effect is significant at the median for the higher initial income group and insignificant for the lower initial income group. This is similar to the result found in Deidda and Fattouh (2002). The same magnitude differences occur for observations with above median levels of schooling and investment, but the median effect is insignificant for the median breakdown by schooling and marginally significant for the median result from the investment breakdown. The opposite result is true for those with above median population growth rates. There is a significant partial effect on the finance variable at the median for countries which have lower population growth rates. Finally, we looked at the median partial effect on the finance proxy for observations where $\ln(DBACBA)$ was above the median. We find this result to be larger than the median result for observations where the financial variable was below the median, but neither are significant. Thus, we are unable to confirm or refute the finding in Rioja and Valev (2004a) that the former is larger.

These variations are visually seen in Figure 3. The differences are confirmed by the Li (1996) test. Again, each test has a p-value which is zero to three decimal places. The main departure from the previous case is that the (unreported) first quartile for the 'lesser' groups is negative in each case. For all scenarios, the negative partial effects are insignificant. However, what this shows us is that in each of these comparisons, while the partial effects for the 'greater' groups are generally positive and significant at the median, over a quarter of the partial effects are (insignificantly) negative for the 'lesser' groups. *In other words, for a large portion of the sample, the impact of financial development does not have a significant impact on output growth.*

Table 5 breaks down the partial effects on the financial variable by geographical region. First, we see a similar pattern in that now OECD countries have a larger effect of the finance variable as opposed to non-OECD countries. Further, the effect is significant for OECD economies at the median, but insignificant for non-OECD economies. *This means that developed economies are able to exploit financial markets for economic gain while lesser developed countries are not in a position to exploit their financial markets either because they are in a primitive stage of development or they lack the proper financial infrastructure.* The same seems to hold true for each of the other groupings. Each are positive, but insignificant at their median values.

5.3.4 Time variation

In Table 6 we present the median partial effects for each of the continuous variable as in Table 4, but here the groups correspond to the years under consideration. Most of the Solow variables do not seem to show a very strong trend. *However, it is obvious that the median partial effect is increasing*

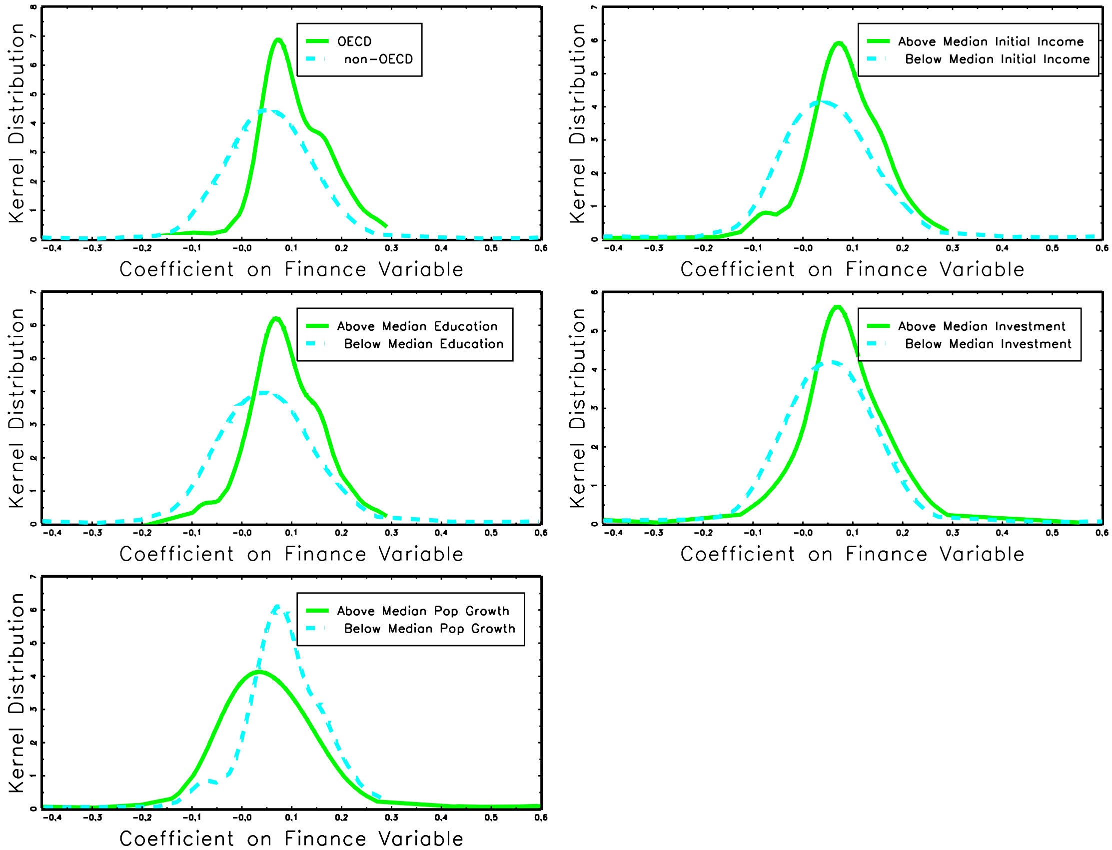

Figure 3: Comparison of LLLS Estimated Coefficients of DBACBA from Model II.

with time for the finance variable. The same holds true for the (unreported) upper and lower quartiles. Notice that the partial effect in 1960 is roughly four times smaller (and insignificant) at the median than the partial effect in 2000. This shows that the returns to financial markets have been increasing over time.

These large changes in the impact of financial development over time correspond to a similar impact that economic globalization has had over the same time period. Having access to a well developed financial market has also provided businesses with unabridged access to world markets over time. These connections can dramatically impact the economic outcomes of countries. Thus, while financial development appears to impact economic growth for specific types of countries, it also appears that having the ability to exploit a financial market is paramount.

Table 5: Median coefficient of LLLS estimates from Model II for each continuous regressor for specific groups of countries. Bootstrap standard errors are in parentheses below each estimate.

Classification	$ln(Y_0)$	$ln(School)$	$ln(Inv)$	$ln(Pop\ Growth)$	$ln(DBACBA)$
OECD	-0.100	0.055	0.075	-0.056	0.093
	(0.053)	(0.191)	(0.057)	(0.032)	(0.025)
Non-OECD	-0.041	0.000	0.055	-0.186	0.053
	(0.070)	(0.073)	(0.049)	(0.054)	(0.454)
Sub-Saharan Africa	-0.036	0.016	0.049	-0.149	0.042
	(0.052)	(0.157)	(0.042)	(0.078)	(0.063)
North Africa/Middle East	-0.127	0.077	0.157	-0.096	0.110
	(0.045)	(0.082)	(0.043)	(0.115)	(0.129)
Asia	-0.043	-0.019	0.048	-0.247	0.052
	(0.064)	(0.143)	(0.049)	(0.152)	(0.143)
Latin America	-0.023	-0.010	0.054	-0.360	0.019
	(0.044)	(0.125)	(0.053)	(0.128)	(0.075)

5.4 Instrumental variables

Uncovering potential causation instead of solely correlation is deemed important when trying to determine what drives the wealth of nations.[12] In this sub-section we employ the estimator of Su and Ullah (2008), which is described in detail in Appendix F. We note that we also follow the suggestion of Su and Ullah (2008) and extend their estimator such that it can incorporate discrete covariates.

We focus on Model II and allow for the financial development variable, $\ln(DBACBA)$, to be endogenous. Here we consider a novel way to obtain instruments in a nonparametric framework. Recall that the bandwidth from the LCLS estimation technique determines whether or not a right-hand-side variable is a relevant predictor of the left-hand-side variable. For example, in Model III (Table 2), we saw that four variables were smoothed out of the regression: $\ln(Pop\ Growth)$, $\ln(DBADGP)$, $\ln(PCRDBGDP)$ and $\ln(BDGDP)$. We argue that each of these variables do

[12]That being said, Durlauf, Johnson, and Temple (2005) argue that any growth study is likely to be flawed by regressor endogeneity and the instruments used to 'correct' for endogeneity are themselves likely flawed. See Ashley (2009) for methods to determine the impact of using a flawed instrument in an IV setting.

Table 6: Median coefficient of LLLS estimates from Model II for each continuous regressor for each time period. Bootstrap standard errors are in parentheses below each estimate.

Time Period	$ln(Y_0)$	$ln(School)$	$ln(Inv)$	$ln(Pop\ Growth)$	$ln(DBACBA)$
1960	-0.062	0.023	0.093	-0.077	0.030
	(0.069)	(0.172)	(0.054)	(0.064)	(0.068)
1965	-0.052	0.035	0.091	-0.111	0.025
	(0.034)	(0.059)	(0.035)	(0.044)	(0.059)
1970	-0.058	0.017	0.085	-0.061	0.037
	(0.040)	(0.092)	(0.040)	(0.031)	(0.163)
1975	-0.070	0.011	0.079	-0.121	0.073
	(0.031)	(0.124)	(0.064)	(0.042)	(0.026)
1980	-0.072	0.009	0.065	-0.174	0.077
	(0.032)	(0.309)	(0.031)	(0.035)	(0.065)
1985	-0.044	-0.009	0.062	-0.224	0.066
	(0.067)	(0.111)	(0.050)	(0.050)	(0.153)
1990	-0.046	0.036	0.047	-0.180	0.090
	(0.060)	(0.092)	(0.059)	(0.062)	(0.026)
1995	-0.047	0.023	0.012	-0.164	0.097
	(0.031)	(0.096)	(0.067)	(0.050)	(0.025)
2000	-0.089	0.085	0.039	-0.083	0.139
	(0.094)	(0.172)	(0.042)	(0.045)	(0.045)

not affect the left-hand-side variable, but are correlated with $\ln(DBACBA)$, the variable which we believe is potentially endogenous. Given that we have a single endogenous regressor, our new model will be over-identified $(4 > 1)$.

For our first stage estimation, we consider a regression of $\ln(DBACBA)$ on each of the variables which were deemed irrelevant in Model III as well as each of the other remaining right-hand-side variables from that model. The bandwidths for this LCLS regression can be found in Table 7. Note that the bandwidths on the regressors $\ln(Pop\ Growth)$, $\ln(DBADGP)$, $\ln(PCRDBGDP)$ and $\ln(BDGDP)$ are now each smaller than twice their standard error and we argue that each is a relevant instrument for $\ln(DBACBA)$.

Given that we believe that we have had a successful first stage regression, we move to the second stage. Unlike standard 2SLS regressions in the parametric framework, the Su and Ullah

Table 7: Bandwidths for both stages of our IV estimation. UB is the upper bound for a given regressor. In the case of the continuous regressors, the upper bound is two times the standard deviation of that variable. For the categorical variables, the upper bound is the true upper bound. $\widehat{u}$ is the residual from the first stage which is required in the second stage of the Su and Ullah (2008) estimation method.

	Stage I		Stage II	
	LHS Variable - $ln(DBACBA)$		LHS Variable - $ln(g_y)$	
Variable	UB	LCLS	UB	LLLS
$ln(Y_0)$	2.106	0.013	2.106	0.860
$ln(School)$	1.362	0.258	1.362	2815529
$ln(Inv)$	1.158	0.151	1.158	626484
$ln(Pop\ Growth)$	0.338	0.088	0.338	6.162
$ln(DBACBA)$			0.628	0.763
$ln(DBAGDP)$	1.558	0.127		
$ln(PCRDBGDP)$	1.686	0.141		
$ln(BDGDP)$	1.382	0.290		
$\widehat{u}$			0.0018	167777
OECD	0.500	0.189	0.500	0.5
Region	0.900	0.417	0.900	0.750
Time	1.000	0.448	1.000	0.894
n		528		528

(2008) estimator does not replace the endogenous regressor with its fitted value from the first stage. Instead, the residual from the first stage is included in the second stage estimation along with the potentially endogenous regressor. Table 7 also gives the bandwidths for the second stage LLLS regression.[13] The bandwidths suggest that the logarithms of initial income and our finance variable both enter nonlinearly. At the same time, we see that the logarithms of schooling, investment and population growth enter linearly.

The resulting marginal effects can be found in the second panel of Table 8. For a more fair comparison we also ran Model II with the same set of observations and report the summary of the marginal effects in the first panel of Table 8. It is obvious from first glance at the table that the qualitative results do not change substantially. The main difference is the number of significant results. In contrast from parametric IV estimators which necessarily have larger standard

[13]We choose LCLS in the first stage to determine relevance of the instruments and LLLS in the second stage in order to obtain more reliable estimates for our primary objects of interest, the marginal effects.

errors than their least-squares counterparts, the standard errors in a nonparametric IV estimation procedure can either be larger or smaller than when estimating without instruments. It appears that the IV estimator has taken out substantial variation in the estimates. When we look at individual groups, we see that now each group shows significant benefits from increases in physical capital, but still only developed countries benefit from increased financial development.

Table 8: LLLS quartile coefficient estimates from the IV regression (third step) as well as analogous estimates without using instruments. Bootstrap standard errors are in parentheses below each estimate.

	Model II			IV Estimation		
Variable	Q1	Q2	Q3	Q1	Q2	Q3
$ln(Y_0)$	-0.139	-0.073	-0.004	-0.136	-0.101	-0.031
	0.076	0.020	0.027	0.036	0.031	0.037
$ln(School)$	-0.027	0.028	0.089	-0.026	0.025	0.048
	0.030	0.049	0.064	0.033	0.020	0.032
$ln(Inv)$	-0.006	0.034	0.087	0.046	0.058	0.090
	0.050	0.024	0.087	0.018	0.022	0.021
$ln(Pop\ Growth)$	-0.328	-0.122	0.017	-0.231	-0.134	-0.040
	0.216	0.096	0.103	0.061	0.100	0.063
$ln(DBACBA)$	-0.009	0.091	0.184	-0.004	0.049	0.123
	0.054	0.123	0.146	0.042	0.034	0.056
R^2		0.841			0.643	
n		528			528	

6 Conclusion

This paper has shed new light on the impact of financial development on economic growth using measures of financial development. Specifically, it makes two contributions: First, we uncover a highly nonlinear relationship between finance and growth that has been masked primarily due to the linearity assumption imposed by the parametric methodology employed by most existing studies. Using recently developed nonparametric methods, we show that although the relationship is significantly positive and becoming stronger across time for middle- and high-income countries, it is non-existent or playing only a small role in determining growth in low-income countries. This finding is consistent with thinking about development in stages (as in Rostow, 1960) in which financial development starts emerging as a key determinant of growth in later stages of development when market imperfections are less severe and institutions less constraining. Second, our findings are relevant to the policy discussions regarding the type and sequencing of structural reforms in developing countries. Specifically, they would suggest that financial reforms will be most beneficial if they are introduced later in the reforms agenda and at a point where they can benefit other potential sources of growth. Both of these issues are areas that warrant future research.

These results were determined via contemporary nonparametric regression and inference results. These methods allow a researcher to avoid functional form misspecification, construct heterogenous parameter estimates and can invoke dimension reduction avenues easily. Taken together these methods represent the current apex of conducting a robust nonparametric analysis. We further add to the literature by exploiting recent advances in nonparametric methods which allow us to use instruments. In addition to conducting nonparametric IV estimation, we present a novel approach to obtaining instruments in a nonparametric framework.

Future extensions of this work involve tracking continuing changes over time as more data (and countries) become available for inclusion. It would also be worth thinking about other types of groups to construct when considering the magnitude/significance of the estimated partial effects on both initial income and financial development. Applying these methods to other growth theories would also add to the discussion of model specification and parameter heterogeneity that has become a popular topic in the growth empirics literature.

References

[1] Acemoglu, D., S. Johnson, P. Querubín and J. A. Robinson (2008). "When Does Policy Reform Work? The Case of Central Bank Intelligence," *Brookings Papers on Economic Activity* 1,351-418.

[2] Aghion, P., P. Howitt, and D. Mayer-Foulkes (2005). "The Effect of Financial Development on Convergence: Theory and Evidence," *Quarterly Journal of Economics* 120, 173-222.

[3] Aitchison, J., and C.B.B. Aitken (1976). "Multivariate Binary Discrimination by Kernel Method," *Biometrika* 63, 413-420.

[4] Ashley, R.A. (2009). "Assessing the Credibility of Instrumental Variables Inference with Imperfect Instruments via Sensitivity Analysis," *Journal of Applied Econometrics 24,* 325-337.

[5] Beck, T., A. Demirgüç-Kunt, and R. Levine (2000). "A New Database on Financial Development and Structure," *World Bank Economic Review* 14, 597-605.

[6] Beck, T., R. Levine, and N. Loayza (2000). "Finance and the Sources of Growth," *Journal of Financial Economics* 58, 261-300.

[7] Benhabib, J. and M. Spiegel (1994). "The Role of Human Capital in Economic Development Evidence from Aggregate Cross-Country Data," *Journal of Monetary Economics* 34, 143-173.

[8] Boudoukh, J., M. Richardson, Y. Shen and R. F. Whitelaw (2007). "Do Assest Prices Reflect Fundamentals? Freshly Squeezed Evidence from the OJ Market," *Journal of Financial Economics* 83, 397-412.

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

[10] Carlin, W. and C. Meyer (2003). "Finance, Investment and Growth," *Journal of Financial Economics* 69, 191-226.

[11] Casselli, F., G. Esquivel and F. Lefort (1997). "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics," *Journal of Economic Growth* 1, 363-389.

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

[13] Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2010). "Nonparametric Instrumental Regression," *Econometrica* forthcoming.

[14] Deidda, L. and B. Fattouh (2002). "Non-linearity Between Finance and Growth," *Economics Letters* 74, 339-345.

[15] Demirgüç-Kunt, A. and Maksimovic (1998). "Law, Finance, and Firm Growth," *Journal of Finance* 53, 2107-2137.

[16] Duching, R., O. Osbas and B. A. Sensoy (2010). "Costly External Finance, Corporate Investment, and the Subprime Mortgage Credit Crisis," *Journal of Financial Economics* 97, 418-435.

[17] Durlauf, S.N., P. Johnson, and J. Temple (2006). "Growth Econometrics," in *Handbook of Economic Growth*, P. Aghion and S.N. Durlauf (Eds.), Amsterdam, North Holland, Volume 1A, 555-677.

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

[19] Durlauf, S., Kourtellos, A. and Tan, C. (2008). "Are any Growth Theories Robust?" *Economic Journal* 118, 329-46.

[20] Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*, London, Chapman and Hall/CRC.

[21] Fan, Y., and A. Ullah (1999). "On Goodness-of-Fit Tests for Weakly Dependent Processes Using Kernel Methods," *Journal of Nonparametric Statistics* 11, 337-360.

[22] Fernandez, C., E. Ley, and M. F. J. Steel, (2001a.) "Benchmark priors for Bayesian model averaging," *Journal of Econometrics* 100, 381-427.

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

[24] Hall, P., and J. L. Howowitz (2005). "Nonparametric methods for inference in the presence of instrumental variables," *Annals of Statistics* 33, 2904-2929.

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

[26] Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge, Cambridge University Press.

[27] Henderson, D.J., C. Papageorgiou, and C.F. Parmeter (forthcoming). "Growth Empirics without Parameters," *Economic Journal.*

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

[29] Jayarathe, J. and P. Strahan (1996). "The Finance-Growth Nexus: Evidence from Bank Branch Deregulation," *Quarterly Journal of Economics* 111, 639-670.

[30] Ketteni, E., T. Mamuneas, A. Savvides, and T. Stengos (2007). "Is the Financial Development and Economic Growth Relationship Nonlinear?" *Economics Bulletin* 15, 1-12.

[31] King, R.G. and R. Levine (1993a). "Finance and Growth: Schumpeter Might Be Right," *Quarterly Journal of Economics* 108, 717-738.

[32] King, R.G. and R. Levine (1993b). "Finance, Entrepreneurship, and Growth: Theory and Evidence," *Journal of Monetary Economics*, 32, 513-542.

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

[34] Levine, R. (2005). "Finance and Growth: Theory and Evidence," in *Handbook of Economic Growth*, P. Aghion and S.N. Durlauf (Eds.), Amsterdam, North Holland, Volume 1A, 865-934.

[35] Levine, R., N. Loayza, and T. Beck (2000). "Financial Intermediation and Growth: Causality and Causes," *Journal of Monetary Economics* 46, 31-77.

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

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

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

[39] Li, Q., and J. Racine (2006). *Nonparametric Econometrics: Theory and Practice*, Princeton, Princeton University Press.

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

[41] Loayza, N.V. and R. Ranciere (2006). "Financial Development, Financial Fragility, and Growth," *Journal of Money, Credit, and Banking* 38, 1051-1076.

[42] Ludvigson, S. C. and S. Ng (2007). "The Empirical Risk-Return Relation: A Factor Analysis Approach," *Journal of Financial Economics* 83, 171-222.

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

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

[45] Newey, W.K. and J.L. Powell (2003). "Instrumental Variable Estimation of Nonparametric Models," *Econometrica* 71, 1565-1578.

[46] Ongena, S. and D. C. Smith (2001). "The Duration of Bank Relationships," *Journal of Financial Economics* 61, 449-475.

[47] Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics*, Cambridge, Cambridge University Press.

[48] Racine, J., J. Hart, and Q. Li (2006). "Testing the Significance of Categorical Predictor Variables in Nonparametric Regression Models," *Econometric Reviews* 25, 523-544.

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

[50] Rajan, R. and L. Zingales (1998). "Financial Dependence and Growth," *American Economic Review* 88, 559-586.

[51] Rioja, F. and N. Valev (2004a). "Does One Size Fit All?: A Reexamination of the Finance and Growth Relationship," *Journal of Development Economics* 74, 429-447.

[52] Rioja, F. and N. Valev (2004b). "Finance and the Sources of Growth at Various Stages of Economic Development," *Economic Inquiry* 42, 127-140.

[53] Rostow, W.W. (1960). *The Stages of Economic Growth: A Non-Communist Manifesto*, Cambridge: Cambridge University Press.

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

[55] Solow, R. (1956). "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics* 70, 65-94.

[56] Su, L. and A. Ullah (2008). "Local Polynomial Estimation of Nonparametric Simultaneous Equations Models," *Journal of Econometrics* 144, 193-218.

[57] Wang, M.C., and J. Van Ryzin (1981). "A Class of Smooth Estimators for Discrete Estimation," *Biometrika* 68, 301-309.

[58] Yung, C., G. Colak and W. Wang (2008). "Cycles in the IPO Market," *Journal of Financial Economics* 89, 192-208.

Not Intended For Publication

A Technical Appendix

A.1 Generalized kernel regression

The discussion above assumes that all the regressors are continuous. This assumption is not reasonable for most economic data sets. Previously, in the presence of 'mixed' data, kernel users had to resort to semiparametric methods. Often authors would assume that the categorical regressors entered the model linearly for ease of estimation. Fortunately, recent advances in nonparametric estimation allow for estimation of both continuous and discrete (order and unordered) variables. In a series of papers, Li and Racine (2004) and Racine and Li (2004) show that the unknown function can include both types of data. The nonparametric model in (2) is rewritten as

$$y_i = m(x_i^c, x_i^u, x_i^o) + u_i, \qquad i = 1, 2, \ldots, n, \tag{5}$$

where x_i^c is a vector of continuous regressors (for example, initial income), x_i^u is a vector of unordered categorical regressors (for example, geographic region) and x_i^o is a vector of ordered categorical regressors (for example, year). All other variables are as previously described.

Estimation of this model by local-constant least-squares is quite similar. The main departure is in the construction of the product kernel. The product kernel, as the name suggests, is the product of the kernel functions for each variable. Here, one type of kernel function is used for continuous regressors, another is used for unordered discrete regressors and a third is used for ordered discrete regressors. The estimator is given as

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} y_i \left[\prod_{s=1}^{q_c} K\left(\frac{x_{si}^c - x_s^c}{h_s}\right) \prod_{s=1}^{q_u} l^u\left(x_{si}^u, x_s^u, \lambda_s^u\right) \prod_{s=1}^{q_o} l^o\left(x_{si}^o, x_s^o, \lambda_s^o\right)\right]}{\sum_{i=1}^{n} \left[\prod_{s=1}^{q_c} K\left(\frac{x_{si}^c - x_s^c}{h_s}\right) \prod_{s=1}^{q_u} l^u\left(x_{si}^u, x_s^u, \lambda_s^u\right) \prod_{s=1}^{q_o} l^o\left(x_{si}^o, x_s^o, \lambda_s^o\right)\right]}, \tag{6}$$

where $l^u(x^u_{si}, x^u_s, \lambda^u_s)$ is the kernel function for a particular unordered discrete regressor with bandwidth λ^u_s and again shows that our estimate is a weighted average of the y_is. Similarly, $l^o(x^o_{si}, x^o_s, \lambda^o_s)$ is the kernel function for a particular ordered discrete regressors with bandwidth λ^o_s. For the continuous regressors, we choose the Gaussian kernel function

$$K\left(\frac{x^c_{si} - x^c_s}{h_s}\right) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x^c_{si} - x^c_s}{\lambda^c_s}\right)^2\right]; \tag{7}$$

where the bandwidth ranges from zero to infinity. A variation of the Aitchinson and Aitken (1976) kernel function for unordered categorical regressors is given as

$$l^u(x^u_{si}, x^u_s, \lambda^u_s) = \begin{cases} 1 - \lambda^u_s & \text{if } x^u_{si} = x^u_s \\ \frac{\lambda^u_s}{d_s - 1} & \text{otherwise} \end{cases}; \tag{8}$$

where the bandwidth is constrained to lie in the range $[0, (d_s - 1)/d_s]$, where d_s is the number of unique values the unordered variable will take. For example, for the case where the unordered variable is a traditional 'dummy variable', the upper bound will be 0.50. Finally, the Wang and Van Ryzin (1981) kernel function for ordered categorical regressors is given by

$$l^o(x^o_{si}, x^o_s, \lambda^o_s) = \begin{cases} 1 - \lambda^o_s & \text{if } x^o_{si} = x^o_s \\ \frac{1-\lambda^o_s}{2}(\lambda^o_s)^{|x^o_{si} - x^o_s|} & \text{otherwise} \end{cases}, \tag{9}$$

where the bandwidth ranges from zero to unity.

Beyond the benefit of being able to incorporate categorical regressors, Li and Racine (2004) show that the rate of convergence of the conditional mean is *only* dependent on the number of continuous regressors. This is extremely important given the curse of dimensionality that is one of the criticisms levied against the use of nonparametric methods. In essence, the addition of discrete regressors need not require additional observations to achieve the same level of precision as the inclusion of additional continuous regressors would.

Another popular estimation procedure is local-linear least-squares estimator. This estimation procedure generally estimates the unknown function with more precision than the local-constant estimator.[14] Additionally it estimates the marginal effects simultaneously. Again, consider the nonparametric regression model in (2). Taking a first-order Taylor expansion of (2) with respect to x_j^c yields

$$y_i \approx m(x_j) + (x_i^c - x_j^c)\beta(x_j^c) + \varepsilon_i$$

where $\beta(x_j^c)$ is defined as the partial derivative of $m(x_j)$ with respect to x^c. For example, if y and x are both expressed in logarithmic form, then $\beta(x_j)$ is interpreted as an elasticity. The estimator of $\delta(x_j) \equiv [m(x_j) \quad \beta(x_j)]'$ is given by

$$\widehat{\delta}(x) = \left[X'K\left(x\right)X\right]^{-1} X'K\left(x\right)y$$

where $X = [1 \quad (x_i - x)]$ and $K\left(x\right)$ is an $n \times n$ diagonal matrix where the i^{th} diagonal element is $K\left(h^{-1}\left(x_i - x\right)\right)$. The intuition behind this estimator is that it fits a line through x based on the points 'close' to x. This is repeated for each x and the slope and intercept of the lines do not have to be equal for different x. Each of these lines are connected to produce the estimate of the unknown function.

A.2 Inclusion of irrelevant variables

In standard nonparametric regression, it is assumed that the bandwidth for a particular continuous regressor goes to zero as the sample size tends towards infinity. Here, when the variable is irrelevant, the cross-validated smoothing parameters converge in probability to the upper extremities of their respective ranges. In addition to improving prediction, this attenuates the curse of dimensionality by removing these variables from the analysis.

More formally, consider the estimator in (5), but say we add an additional $p > 0$ irrelevant

[14] Precision here is taken to mean less bias with the same amount of variance, resulting in a lower mean square error.

regressors for each particular type of variable. The estimator of the conditional mean becomes

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} y_i \prod_{s=1}^{q_c} K\left(\frac{x_{si}^c - x_s^c}{h_s}\right) \prod_{s=q_c+1}^{q_c+p} K\left(\frac{x_{si}^c - x_s^c}{h_s}\right) \prod_{s=1}^{q_u} l^u\left(x_{si}^u, x_s^u, \lambda_s^u\right)}{\sum_{i=1}^{n} \prod_{s=1}^{q_c} K\left(\frac{x_{si}^c - x_s^c}{h_s}\right) \prod_{s=q_c+1}^{q_c+p} K\left(\frac{x_{si}^c - x_s^c}{h_s}\right) \prod_{s=1}^{q_u} l^u\left(x_{si}^u, x_s^u, \lambda_s^u\right)} \times \frac{\prod_{s=q_u+1}^{q_u+p} l^u\left(x_{si}^u, x_s^u, \lambda_s^u\right) \prod_{s=1}^{q_o} l^o\left(x_{si}^o, x_s^o, \lambda_s^o\right) \prod_{s=q_o+1}^{q_o+p} l^o\left(x_{si}^o, x_s^o, \lambda_s^o\right)}{\prod_{s=q_u+1}^{q_u+p} l^u\left(x_{si}^u, x_s^u, \lambda_s^u\right) \prod_{s=1}^{q_o} l^o\left(x_{si}^o, x_s^o, \lambda_s^o\right) \prod_{s=q_o+1}^{q_o+p} l^o\left(x_{si}^o, x_s^o, \lambda_s^o\right)}. \tag{10}$$

The idea is that the cross-validation procedure will recognize that each of these p irrelevant regressors for each variable type are in fact irrelevant and thus should not be used in the prediction of y. For the continuous regressors, the upper bound is infinity. When this bandwidth takes its upper bound, the kernel function becomes $K\left(\left(x_i^c - x^c\right)/\infty\right) = K(0)$. For the unordered and ordered categorical kernels, the upper bounds are $\left(d_s - 1\right)/d_s$ and unity respectively. A closer examination of (8) and (9) show that when the bandwidth hits its upper bound, the weights given to observations equal to x_s^u and x_s^o, respectively, are equal to the weights when the observations are different from x_s^u and x_s^o, respectively. Therefore, when each of these irrelevant regressors are given their appropriate (upper bound) bandwidth, the kernel functions for the irrelevant regressors in (10) cancel out and we are left with (6), i.e. the second fraction in (10) is 1.

A.3 Consistent specification testing

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

$$T_n^a = \frac{n\sqrt{h_1 h_2 \cdots h_q}\widehat{I}_n}{\widehat{\sigma}_n^a} \sim N(0,1), \tag{11}$$

where

$$\begin{aligned}
\widehat{I}_n^a &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \widehat{u}_i \widehat{u}_j K_{\hat{h},\hat{\lambda}^u,\hat{\lambda}^o}, \\
\widehat{\sigma}_n^{a2} &= \frac{2h_1 h_2 \cdots h_q}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \hat{u}_i^2 \widehat{u}_j^2 K_{\hat{h},\hat{\lambda}^u,\hat{\lambda}^o}^2,
\end{aligned}$$

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

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

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

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

4. Use the bootstrap residuals to compute the test statistic $T_n^{a*} = n\left(h_1 h_2 \cdots h_q\right)^{1/2} \widehat{I}_n^{a*}/\widehat{\sigma}_n^{a*}$, where $\widehat{I}_n^{a*}$ and $\widehat{\sigma}_n^{a*}$ are the same as $\widehat{I}_n^a$ and $\widehat{\sigma}_n^a$ except that $\widehat{u}_i$ is replaced by $\widehat{u}_i^*$.

5. Repeat steps (1-4) a large number (B) of times and then construct the empirical distribution of the B bootstrap test statistics, $\{T_n^{a*}\}_{b=1}^B$. This bootstrap empirical distribution is used to

approximate the null distribution of the test statistic T_n^a. We reject H_0 if $T_n^a > T_{n(\alpha B)}^{a*}$, where $T_{n(\alpha B)}^{a*}$ is the upper α -percentile of $\{T_n^{a*}\}_{b=1}^B$.

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

A.4 Nonparametric significance testing

To determine whether or not a set of variables are jointly significant, we utilize the Lavergne and Vuong (2000) test modified by Racine, Hart and Li (2006) to allow for mixed categorical and continuous data. Consider a nonparametric regression model of the form

$$y_i = m\left(w_i, z_i\right) + u_i.$$

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

$$H_0 : E\left(y|w, z\right) = E\left(y|w\right)$$

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

$$E\left\{uf_w\left(w\right)E\left[uf_w\left(w\right)|x\right]f\left(x\right)\right\}$$

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

$$\widehat{I}_n^b = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} (y_i - \widehat{y}_i) \widehat{f}_w(w_i)(y_j - \widehat{y}_j) \widehat{f}_w(w_j) W(x_i, x, h, \lambda^o, \lambda^u) \tag{12}$$

where $W(x_i, x, h, \lambda^o, \lambda^u) = \prod_{s=1}^{q_c} K\left(\frac{x_{si}^c - x_s^c}{h_s}\right) \prod_{s=1}^{q_u} l^u(x_{si}^u, x_s^u, \lambda_s^u) \prod_{s=1}^{q_o} l^o(x_{si}^o, x_s^o, \lambda_s^o)$ is the product kernel mentioned in Section 2 and

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

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

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

Under the null we have that

$$T_n^b = (nh_1 h_2 \cdots h_q)^{1/2} \widehat{I}_n^b / \widehat{\sigma}_n^b \to N(0, 1)$$

where

$$\widehat{\sigma}_n^{b2} = \frac{2h_1 h_2 \cdots h_q}{n^2} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} (y_i - \widehat{y}_i)^2 \widehat{f}_w(w_i)(y_j - \widehat{y}_j)^2 \widehat{f}_w(w_j) W(x_i, x, h, \lambda^o, \lambda^u)$$

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

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

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

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

The estimated density-weighted bootstrap residual is

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

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

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

A.5 Testing equality of pdfs

To test whether two vectors of data $\{x_i\}_{i=1}^{n_1}$ and $\{z_i\}_{i=1}^{n_2}$ are drawn from the same distribution we employ the Li (1996) test. The Li (1996) test, which tests the null hypothesis $H_0 : f(x) = g(x)$ for all x, against the alternative $H_1 : f(x) \neq g(x)$ for some x, works with either independent or dependent data. The test statistic used to test for the difference between the two unknown distributions (which Fan and Ullah 1999 show goes asymptotically to the standard normal), predicated on the integrated square error metric on a space of density functions, $I(f,g) = \int_x \left(f(x) - g(x)\right)^2 dx$, is

$$T_n^c = \frac{\left(n_1 n_2 h_1 h_2 \cdots h_q\right)^{1/2} \widehat{I}_n^c}{\widehat{\sigma}_n^c} \sim N(0,1), \tag{13}$$

where

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

and

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

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

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

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

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

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