SEMIPARAMETRIC APPLICATIONS IN ECONOMIC GROWTH

by

Mustafa Koroglu

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy
in
Economics

Guelph, Ontario, Canada

ABSTRACT

SEMIPARAMETRIC APPLICATIONS IN ECONOMIC GROWTH

Mustafa Koroglu
University of Guelph, 2017

Advisor:
Professor Thanasis Stengos

This dissertation consists of three essays that deals with estimation of semiparametric regression methods in macroeconomic context.

Chapter 1 introduces the building-blocks of the non-/semiparametric regression methods. A literature review is provided to support the estimation methodologies employed in the subsequent chapters. I survey some nonparametric estimation techniques, including (i) the local least squares kernel estimator; (ii) nonparametric series estimator; (iii) estimation of nonparametric models with endogeneity; and (iv) nonparametric estimation of panel data models. I also survey different bootstrapping methods for nonparametric regression methods.

In Chapter 2 we consider a spatial Durbin model with unknown functional-coefficients and nonparametric spatial weights. We apply series approximation method to estimate the unknown functional coefficients and spatial weighting functions via a nonparametric two-stage least squares (or 2SLS) estimation method. We illustrate proposed estimation method to re-examine national economic growth by augmenting the conventional Solow economic growth convergence model with unknown spatial interactive structures of the national economy, as well as country-specific Solow parameters, where the spatial weighting

functions and Solow parameters are allowed to be a function of geographical distance and the countries' openness to trade, respectively.

In Chapter 3 I re-investigate the relationship between public debt and economic growth and try to expose nonlinearity in this link through using an endogenous smooth coefficient approach. I find some evidence of parameter heterogeneity in the debt-growth link that may be governed by the institutional quality of countries. My results show a significant negative effect of public debt on economic growth for the countries with the lowest democracy score and high democracy score.

DEDICATION

To Feyza, my beautiful wife, you are the sunshine of my life...

And to Berat Kemal and Lina Betul, my little lion king and my little princess...

ACKNOWLEDGEMENTS

I owe this thesis to the large number of people. First and foremost, I am sincerely grateful to Yiguo Sun for her endless guidance and encouragement when I started my research in applied nonparametric econometrics with her. I would like to thank her for her patience, constant support, and availability while working on our published paper.

I would like to offer my special thanks to Thanasis Stengos as being more than a supervisor for me. Without his guidance, encouragement, and utmost generosity throughout my years in the Ph.D. program, it would have been impossible for me to complete this thesis. I am truly indebted to him for all the advice and the way of econometric thinking that I learned over the past couple of years.

I wish to thank Alex Maynard for encouraging me to think about the economic reasoning of any findings in my research. I learned a lot from him the way of teaching an econometric course. My sincere thanks go to Miana Plesca for her insightful comments that helped me to understand the fundamentals of empirical econometrics. I always remember her warm approach to her students.

I wish to thank Kurt Annen for attending his lectures in advanced microeconomics that certainly helped me a lot to be well-prepared for the comprehensive exams. I am also thankful to James Atsu Amegashie as he inspires me to ask an economic intuition of any research question of interest. I would like to thank Francis Tapon for his very kind approach by allowing me to contact with his network. I will also remember David Prescott's unique teaching approach in an introductory econometrics course when I was a teaching assistant for him.

I have had many great friends while pursing a Ph.D. at the University of Guelph. Sergiy Pysarenko is one of the best friends that I have had in my life. I have had the pleasure of studying with him in the first two years of our program. I also learned a lot from his way of approaching economic questions and how to deal with them using critical thinking. I hope that we will be in contact even after the years of graduation.

I would like to thank to all those in my family that have supported me over the years. Especially, without my mother and father's love, understanding, and continuous support, it would not have been possible to complete this thesis.

My son, Berat Kemal, and my daughter, Lina Betul, are the little ones that I have motivated everyday to complete the thesis working harder and faster. As being a Montessori child, Berat has an endless curiosity of exploring the nature and everything. Having lots of fun over the past two and half years have made stressful graduate school years easy for me. Having a little princess in my life gives me more energy to work on my future research projects.

Feyza, you are the princess of my heart. I love you so much, and I certainly could not have completed my Ph.D. without you with me. The words cannot express how I appreciate all you have done for me as you have had to deal with my stresses on a daily basis. With our two wonderful kids, I look forward to have many many lovely years with you.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Empirical research in economics typically aims at to investigate relationship between the mean of Y and an explanatory variable X in the following form:

$$E[Y|X] = f(X), \tag{1.1}$$

where $f(\cdot)$ is a function that describes this link. If we assume that $f(X) = \alpha + X\beta$, then it is called that the mean of Y is related X linearly. The linear regression model is one of the most popular tool in data analysis that has a simple functional form allowing a researcher easily interpret the coefficient estimates for policy-relevant questions. It has also a well-established least-squares theory so that a data analyst may conduct statistical inference based upon this linear specification. However, strong linearity assumption may not approximate the actual relationship between the variables of interest, and, therefore, may result in inconsistency in the estimates and thus misleading inference.

Nonparametric regression methods are used to avoid this misspecification problem assuming not on the structure of the regression relationship, but rather on the characteristics of the structure. This modeling approach essentially let the data show the researcher appropriate functional form. However, nonparametric methods have a problem with dimensionality in terms of having less precise estimation as the dimension of explanatory variable vector increases. This is known in the literature as the "curse of dimensionality" and can be mitigated by the use of semiparametric regression techniques that have the flexibility of nonparametric regression models as well as retain the interpretability of the linear models. In this dissertation I study the varying coefficient models that assume regressors enter the model linearly, but allow the coefficients to vary with respect to covariates. This approach is especially useful in empirical economic research by means of avoiding unnecessary assumptions about the dependence between policy variables and exposing heterogeneity in the response of economic units, e.g., countries, individuals, and firms etc.

In the second chapter, I survey some nonparametric estimation techniques, including (i) the local least squares kernel estimator; (ii) nonparametric series estimator; (iii) estimation of nonparametric models with endogeneity; and (iv) nonparametric estimation of panel data models. I also survey different bootstrapping methods for nonparametric regression methods.

The third chapter, "Functional-Coefficient Spatial Durbin Models with Nonparametric Spatial Weights: An Application to Economic Growth", written jointly with Yiguo Sun, considers a spatial Durbin model with unknown functional coefficients and nonparametric spatial weights. This modeling approach is free from inconsistency related to the ad hoc selection of weight matrices and thus provides more reliable inference. The chapter shows how to estimate unknown coefficient functions and spatial weights and illustrates them by applying to the 91 non-oil countries from 1960-1995. We find that economic growth rate of country of interest has a negative spatial lag effect of neighbouring countries's GDP per capita growth rate and initial GDP per capita. Our results also show that conventional Solow variables have significant average direct impact on economic growth at their predicted signs, but insignificant average indirect impact on economic growth.

The fourth chapter of this dissertation investigates the relationship between public debt and economic growth and tries to expose nonlinearity in this link through using a novel and recently developed semiparametric estimation method that accounts for endogeneity in the model. I find some evidence of parameter heterogeneity in the debt-growth link, which may be governed by the institutional quality of countries. My results show a significant negative effect of public debt on economic growth for the countries with the lowest democracy score and high democracy score and this effect magnifies almost by double, on average, for the former group of countries. As sensitivity analysis, our main results are robust to other measures of democracy and additional control variables. Our core results change if we exclude two outliers from the data, Guyana and Nicaragua. Negative and significant effect of public debt on growth is found only for high democracy score countries.

Chapter 2

Survey of Non-/Semiparametric Regression Methods

2.1 Introduction

Nonparametric regression approach has been increasingly adopted by applied researchers since last decade. The main reason behind its popularity lies in its flexibility, one of the fundamental aspects of statistical models, on analyzing unknown regression function. Furthermore, computational programming improvement on nonparametric technique over the last 10 years and its alternative visual interpretations make it draw more and more practitioners in recent years; see Hayfield and Racine (2008) for **R**-programming package for the nonparametric techniques. For some recent economic applications in nonparametric econometrics with graphical interpretations, see Stengos et al. (2009), Kumbhakar and Sun

(2012), and Delgado et al. (2014), among many others.

Advantages of parametric approach in terms of being computationally ease and parsimonious interpretation of policy-related coefficient estimates depends on correctly specified model. However, the fact that economic theory suggests a set of variables relevant in explaining a policy question, but rarely dictates a specific functional form of the relationship between those variables. An incorrectly specified parametric model leads to serious misspecification bias, which cannot be reduced only by large samples, and, thus, results in misleading inference (Scott, 1992, p.33). Nonparametric approach, on the other hand, makes fewer assumptions on regression functions, therefore, is believed to better describe the underlying process that generated the data. Stoker (1992) intuitively states that nonparametric econometric model specifies a connection between a process, which is attributed, by the model itself, to the economic agents' "systematic" responses, also interpreted as "predictable" behaviors, and the observed data. The following definition makes a clear distinction between the two approaches.

Definition 1 (Parametric family of models) Let Ω be a space, $\mathcal{A}$ a σ-algebra of subsets of Ω, and $\mathcal{P}$ a probability measure defined on $\mathcal{A}$, then the triplet $(\Omega, \mathcal{A}, \mathcal{P})$ is said to be a probability space (Dhrymes, 1989).

Remark 1 A set of probability measure $\mathcal{P}_\theta$ is a known probability measure with an unknown parameter θ belongs to *finite* dimensional parameter space, Θ; i.e., $\theta \in \Theta \subseteq \Re^d$, where d is the dimension.

Remark 2 If a set of probability measure $\mathcal{P}_\theta$ is unknown with an unknown parameter θ

belongs to *infinite* dimensional parameter space, then it constitutes a nonparametric family of models.

In nonparametric method, researcher chooses an appropriate function space to which regression function is believed to belong. Following the lines of explanation in Eubank (1988, p.3), this choice is motivated by regularity conditions including smoothness assumptions imposed on unknown regression function. Note that the "qualitative" restrictions on regression function enables researcher to let data determine the form of regression curve. On the other hand, this flexibility of nonparametric method has a price to be paid in terms of loss of efficiency, higher dimensionality problem and slower convergence rates.

2.2 Least Squares Kernel Estimators

We start with an introduction to kernel estimators as a building block of nonparametric technique and continue with a discussion of assumptions typically made. We refer to this set of assumptions and their extensions for the other nonparametric estimators as we will review in the subsequent sections.

Assume that we have a collection of independently and identically distributed (i.i.d.) observations $\{(Y_i, X_i)\}_{i=1}^n$ realized from a joint probability density function (pdf) $f(y, x)$. We can write unknown regression relationship between Y and X as

$$Y_i = g(X_i) + \epsilon_i, i = 1, ..., n, \tag{2.1}$$

where g(·) is an unspecified function and ϵ_i's are observation errors. If we believe that g(·), true regression function, is smooth, i.e., a differentiable function up to some degree, then we can use a local average of data near a point x, rather than at a point x, to construct an estimator of g(·). Formally, we can write the estimator of g as

$$\hat{g}(x) = \sum_{j=1}^{n} w_j(x, X_j; h) Y_j, \tag{2.2}$$

which is called a *linear smoother* and the way we obtain this estimator is commonly called *smoothing* in the literature. Note that w_j is a weight assigned to each Y_j calculated for each X_j in the $h-$neighborhood of x and h is called the smoothing parameter that determines the size of the local neighborhood.

If we assume $\{w_j\}_{j=1}^n$ is a sequence of nonnegative weights and $\sum_{j=1}^n w_j(x, X_j; h) = 1$ for each x, then Equation 2.2 can be obtained from a minimization of locally weighted least squares problem:

$$\min_a \sum_{j=1}^{n} (Y_j - a)^2 K((x - X_j)/h), \tag{2.3}$$

where $K(\cdot)$ denotes a symmetric kernel weight function. The solution of this problem is $\hat{a} \equiv \hat{g}(x) = \frac{\sum_{i=1}^n Y_i K((x-X_i)/h)}{\sum_{i=1}^n K((x-X_i)/h)}$, which is known as the Nadaraya-Watson kernel estimator, and w_i in (2.2) can be explicitly written as $\frac{K((x-X_i)/h)}{\sum_{j=1}^n K((x-X_j)/h)}$, which is the weight attached to Y_i for each $i = 1, ..., n$. Note that $K(\cdot)$ assigns specific weights to each X_j depending on its closeness to the point at which we estimate the unknown function.

Now, we will introduce the derivation of the Nadaraya-Watson estimator using kernel density estimation; for a well-known reference in this subject, see Silverman (1992). The following result, taken from Li and Racine (2007), allows us to use the conditional mean of Y given $X = x$ as an interpretation of $g(x)$, i.e., $g(x) = E[Y_i|X_i = x]$.

Theorem 1 Let $\mathcal{G}$ denote the class of Borel measurable (or continuous) functions having finite second moment. Assume that $g(x) \equiv E(Y|X = x)$ belongs to $\mathcal{G}$, and that $E(Y^2)$ is finite. Then $E(Y|X)$ is the optimal predictor of Y given X, in the following mean squared error (or MSE) sense:

$$E(Y|X) = arg \min_{r(\cdot)\in\mathcal{G}} E\{[Y - r(X)]^2\}. \tag{2.4}$$

From the definition of the conditional expectation, we have

$$E(Y|X = x) = \int yf(y|x)dy = \frac{\int yf(x,y)dy}{f(x)} \equiv g(x), \tag{2.5}$$

where $f(y|x)$, $f(x,y)$, and $f(x)$ denote the conditional pdf of Y given X, the joint pdf of (Y, X), and the marginal pdf of X, respectively. Denoting h as the bandwidth parameter for X, we have

$$\hat{g}(x) = \frac{\int y\hat{f}(x,y)dy}{\hat{f}(x)} = \frac{\sum_{i=1}^{n} K((X_i - x)/h)Y_i}{\sum_{i=1}^{n} K((X_i - x)/h)}, \tag{2.6}$$

where the second equality is shown from some algebraic manipulations.

Notice that the Nadaraya-Watson kernel estimator in Equation (2.6) is a solution of local polynomial problem of degree 0 given in (2.3). Equation (2.6) is, therefore, known as the local constant kernel estimator. Local linear least squares estimator provided below, on the other hand, has some important desired properties compared to the Nadaraya-Watson kernel estimator. It was first introduced by Stone (1977) and Cleveland (1979), while further the theoretical background of this estimator was established by Stone (1980, 1982), Fan (1992, 1993), Fan and Gijbels (1992), and Ruppert and Wand (1994).

Before proceeding to the local linear estimator, it should be noted that local modeling approach is theoretically originated from a Taylor approximation. Following Adams (1989, p.279), behavior of a function g(·), or a curve, near a point (x_0,g(x_0)) can be best described by the tangent line[1], in general Taylor polynomial of degree n, to the function at x=x_0 provided that g(·) is a smooth function (Adams, 1989, Theorem 11, p.290). More intuitively, if a function satisfies smoothness conditions, its value at a specific point can be "reasonably well" approximated by using its evaluations at neighboring points (Yatchew, 1998, p.677).

Definition 2 The linearization, or linear approximation, of the function g about $x = x_0$ is the function $L(x)$ defined by $L(x) = g(x_0) + g^{'}(x_0)(x - x_0)$ (Adams, 1989, p.279).

Therefore, the local linear least squares kernel estimator of g(·), which is $\hat{g}(\cdot)$, is calculated from a minimization of a kernel-weighted objective function:

$$(\hat{g}(x_0), \hat{g}^{'}(x_0)) = arg \min_{g(x_0), g^{'}(x_0)} \sum_{i=1}^{n} [Y_i - g(x_0) - g^{'}(x_0)(X_i - x_0)]^2 K((X_i - x_0)/h). \quad (2.7)$$

[1] It should be noted that low-order polynomials is more practical for local fitting given the true regression function is smooth (Stone (1977, Corollary 4, p.601); Scott (1992, p.221); Fan and Gijbels (1996, p.59)).

Let $\delta(x_0) = [g(x_0), g^{'}(x_0)]$. Then, the solution to the problem (2.7), is given by

$$\tilde{\delta}(x_0) = (\mathcal{X}^T \mathcal{K} \mathcal{X})^{-1} \mathcal{X}^T \mathcal{K} Y, \tag{2.8}$$

where $\mathcal{X}$ is a $n \times 2$ matrix having $(X_i^T, X_i^T(X_i - x_0))$ as its i^{th} row and $\mathcal{K}$ is a $n \times n$ diagonal matrix with the i^{th} diagonal element being $K((X_i - x_0)/h)$.

Several remarks can be drawn from the local linear kernel estimator and generally from the local polynomial fitting. We would like to discuss these remarks along with the consistency result of this estimator.

An evaluation of an estimator is an important part of the theoretical analysis as there might be several alternative estimators for the econometric model. Most widely used performance measure of an estimator is the mean integrated squared error (MISE) interpreted as one of the global error criteria, among others[2]. Note that L_2-norm referred to the integrated squared error (ISE), which depends on the particular sample realizations, whereas MISE, the average of ISE over the given data points, provides sufficient information as an error criteria (Scott, 1992, p.38). A relevant research question answered by the large sample properties of an estimator is that how fast an estimator converges to the true regression curve. This is mainly explained by the decreasing rate of mean squared error (MSE) of an estimator at a specific point with the sample size. The two of the mostly used stochastic convergence concepts for an estimator are given in the following definitions (Davidson

[2]Other methods include L_∞ and L_1 norms explained in Scott (1992, p.38) and Caselle and Berger (2002, p.330). Specifically, these norms including L_2 norm is calculated as $sup_x|\hat{g}(x) - g(x)|$, $\int |\hat{g}(x) - g(x)|dx$, and $\int [\hat{g}(x) - g(x)]^2 dx$, respectively.

(2004), p.38).

Definition 3 (Convergence in probability) If X is a random variable, and for all $\epsilon > 0$

$$lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$$

X_n is said to converge in probability to X.

Definition 4 (Convergence in mean square) If

$$lim_{n\to\infty} E(X_n - X)^2 = 0$$

X_n is said to converge in mean square to X.

Note that since convergence in mean square implies convergence in probability, the former is commonly used in the literature due to its analytical simplicity and easy to interpretation[3] (Caselle and Berger, 2002, p.349).

We follow Theorem 1 in Fan (1993) for pointwise asymptotic property of the local linear kernel estimator.

Assumption 1 $g(\cdot)$ has a bounded second derivative.

Assumption 2 The conditional variance $\sigma^2(x) = Var(Y|X = x)$ is bounded and continuous.

[3] $E(X_n - X)^2 = E(X_n^2 - 2X_nX + X^2) = E(X_n^2) - 2XE(X_n) + X^2 = Var(X_n) + E(X_n)^2 - 2XE(X_n) + X^2 = Var(X_n) + (Bias(X_n))^2$, where the third equality is due to $Var(X_n) = E(X_n^2) - E(X_n)^2$, and the fourth equality comes from $(Bias(X_n))^2 = E(X_n)^2 - 2XE(X_n) + X^2$. This states that mean squared error of an estimator equals to the summation of the variance and the bias square of an estimator.

Assumption 3 $f(x_0)$ is continuous and bounded away from zero.

Assumption 4 The kernel $K(\cdot)$ is a bounded and continuous density function satisfying following moment conditions: (i) $\int_{-\infty}^{\infty} K(u)du = 1$, (ii) $\int_{-\infty}^{\infty} uK(u)du = 0$, (iii) $c_K = \int_{-\infty}^{\infty} u^2 K(u)du = \alpha \neq 0$, and (iv) $d_K = \int_{-\infty}^{\infty} K(u)^2 du < \infty$.

Theorem 2 Under assumptions 1-4, if $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then for $x_0 \in (a_0, b_0)$ the local linear kernel estimator $\hat{g}(x)$ has the MSE

$$E(\hat{g}(x_0) - g(x_0))^2 = \frac{1}{4}(c_K g^{''}(x_0))^2 h^4 + \frac{d_K}{nh}\frac{\sigma^2(x_0)}{f(x_0)} + o(h^4 + \frac{1}{nh}). \qquad (2.9)$$

Note that first two terms of the right of the equality in (2.9) are the approximate expressions for the bias square and the variance of the estimator, respectively. Assumption 1 is the smoothness condition imposed on the unknown regression function, $g(\cdot)$. Higher degree of differentiability of $g(\cdot)$ reduces the order of the bias term without changing the order of the variance term. The bounded conditional variance in Assumption 2 is commonly imposed in nonparametric literature in order to make the derivation of convergence rates less complicated. Assumption 4(i) ensures that $K(\cdot)$ is a probability density function, 4(ii) is the symmetry condition, and 4(iv) assures us that $\hat{g}(x_0)$ has finite asymptotic variance.

The bandwidth parameter h plays an important role in nonparametric estimation affecting the order of bias square and variance term in Equation (2.9). The first condition on the bandwidth parameter, $h \to 0$, ensures shrinking local neighborhood as the sample size increases, which, therefore, results in a reduction in bias. The second condition,

$nh \to \infty$, ensures that the number of observations in the local neighborhood grows such that the variance term in (2.9) declines to zero.

We can rewrite the mean squared error of an estimator $\hat{g}$ at a point x_0 as

$$E(\hat{g}(x_0) - g(x_0))^2 = O(h^4) + O(\frac{1}{nh}). \quad (2.10)$$

By Theorem A.7.(ii) in Li and Racine (2007) we can restate Equation (2.10) as the convergence of an estimator at a specific point to the true function value.

$$\hat{g}(x_0) = g(x_0) + O(h^2) + O_p(\frac{1}{\sqrt{nh}}). \quad (2.11)$$

Optimal rate of convergence is achieved when the bias and variance terms in (2.11) converge to zero at the same rate, i.e., $O(h^2) = O_p(\frac{1}{\sqrt{(nh)}})$. Therefore, an optimal rate of convergence for $\hat{g}(x_0)$ is achieved when $h = O(n^{-1/5})$. It follows that the MSE of an estimator $\hat{g}$ at a point x_0 has an optimal rate of convergence of $O(n^{-4/5})$. Therefore, Equation (2.11) can be rewritten as

$$\hat{g}(x_0) = g(x_0) + O(n^{-2/5}) + O_p(n^{-2/5}). \quad (2.12)$$

Since convergence in MSE implies convergence in probability (Li and Racine, 2007, Theorem A.3.), Equation (2.12) can be explicitly stated that an estimator $\hat{g}(x_0)$ is "$n^{2/5}$ consistent" if $n^{2/5}(\hat{g}(x_0) - g(x_0)) = O_p(1)$. Intuitively, the value $2/5$ gauges how quickly the difference between $\hat{g}(x_0)$ and $g(x_0)$ vanishes as the sample size increases[4].

It should be noted that the local linear kernel estimator has an advantage over Nadaraya-Watson and Gasser-Muller kernel estimators as its convergence rate is the same at both interior and boundary points of the density of X. This is explicitly seen from the bias square term in Equation (2.9), which does not involve the derivative of the density f; see Fan (1992) and Ruppert and Wand (1994) for detail information.

Optimal global bandwidth can be chosen by using the least-squares cross-validation method for the mean integrated squared error (MISE). Note that among other bandwidth selection techniques, this method is most frequently used in practice as it is mainly a data-driven approach. The optimal bandwidth is calculated as

$$h_{opt} = \left(\frac{d_K \int_{-\infty}^{\infty} f^{-1}(x)\sigma^2(x)dx}{c_K^2 \int_{-\infty}^{\infty} [g''(x)]^2 dx} \right)^{1/5} n^{-1/5}. \tag{2.13}$$

It is noteworthy that both the bias square of an estimator in Equation (2.9) and the optimal bandwidth in (2.13) depends on the second derivative of the true function $g(\cdot)$. Specifically, a large magnitude of $|g''(x_0)|$ means a rapid change of g near x_0, which necessitates that only data points close to x_0 provide "good" information about $g(x)$. In other words, observations far from x_0 cause biases in an estimator since they are less similar observations. In this case, smaller bandwidth can be used to reduce the bias. On the other hand, if $|g''(x_0)|$ is small in magnitude, then the bandwidth should then be widened to include relevant obser-

[4]As a comparison, $\sqrt{n}$ is the well-known convergence rate of an estimator for the population value in the parametric settings.

vations, which decreases the variance. Therefore, optimal bandwidth h_{opt} given in (2.13) optimizes the bias-variance trade-off in Equation (2.9). For further interpretations on the balance between bias square and variance of an estimator, see Ruppert and Wand (1994) and Chu and Marron (1991).

It should be noted that Theorem 2 deals with one-dimensional case. If X variable has a higher dimension, the order of the variance term increases without affecting the order of the bias term, which results in a so-called *curse-of-dimensionality* problem. In this case, parameters are estimated less accurately since less information is used in each local estimation, holding sample size fixed. Larger datasets, higher degree of differentiability of the function and other model specifications such as semiparametric model are among the remedies of this problem.

For further information on the topic, see Fan and Gijbels (1996) for a comprehensive discussion of the local linear estimator; Eubank (1988) for excellent interpretations of the properties of the kernel estimators; and Yatchew (1998) for an overview of the nonparametric regression technique with deep insights on the large sample properties of the local linear estimator.

2.3 Nonparametric Series Estimators

Another popular and mostly used technique in nonparametric estimation is the method of sieves, which essentially uses approximating functions such as polynomials and splines to recover unknown functions. Using these functions sieve spaces can be constructed *to approximate infinite-dimensional parameter spaces* in which unknown functions from non-/semiparametric models lie. Major advantages of series methods relative to the kernel regression estimation are its computational convenience and the ease of imposing certain restrictions including additive separability and varying coefficient structure on the unknown conditional expectation form, $E[Y|X]$. Particularly, series methods, known as the linear sieves, enable the researcher to estimate unknown function for all data point in the domain at the same time via one single regression. For an introduction to the series methods, see Li and Racine (2007) and for large sample properties of the series estimators, see Chen (2007) and Hansen (2014).

One important feature of the series methods is that the number of series terms used in the approximation of the unknown function increases with the sample size. Intuitively, this flexibility makes the series methods to be used as a nonparametric technique and this can be seen as an important difference with classical parametric regression, which assumes fixed, finite-dimensional parameter spaces (Chen, 2007; Hansen, 2014). Chen (2007) provides technical details on functions that can be well approximated by the sieves and specifies various types of sieves that can be used for the approximation. Following definition classifies such functions.

Definition 5 A real-valued function g on a d-dimensional compact set $\mathcal{X}$ satisfies a

Hölder condition for some $\gamma \in (0, 1]$ if there is a positive number c such that

$$|g(x) - g(y)| \leq c|x - y|^{\gamma}$$

for all $x, y \in \mathcal{X}$ and $||$ denotes the Euclidean norm.

If a function is continuously differentiable up to an order and its each respective derivative satisfies a Hölder condition, then this function is said to belong to the Hölder class of functions. This class of functions can be well approximated by the linear sieves, which is a linear combination of the finitely many basis functions denoted by $\{\phi_s(x)\}_{s=1}^{K}$, where $K = K_n$ denotes the number of basis functions used in the approximation process. Commonly used basis functions are polynomials, trigonometric polynomials, univariate splines, and wavelets for bounded support; and Hermite polynomials and Laguerre polynomials for unbounded support, where the latter is used in our two empirical applications (Chen, 2007). Formally, a function g can be approximated by the series terms as:

$$g^{*}(x) = \sum_{s=1}^{K} \beta_s \phi_s(x), \tag{2.14}$$

where β_s, $s = 1, ..., K$, are the unknown coefficients of each basis function in the approximation. A series estimator is then obtained by estimating β_s's using OLS regression method if there is no endogeneity in the model. In order to be consistent in using the same notations as in the previous chapters, we denote a vector of approximating functions

$$\Phi_K(x) = [\phi_1(x), \phi_2(x), ..., \phi_K(x)]^T. \tag{2.15}$$

Then, the series estimator of $g(x)$ is calculated as

$$\hat{g}(x) = \Phi_K(x)^T\hat{\beta}, \qquad \hat{\beta} = (P^TP)^-P^TY, \qquad P_K = [\Phi_K(X_1), ..., \Phi_K(X_n)]^T, \tag{2.16}$$

where $(\cdot)^-$ is any generalized inverse and $Y = (Y_1, ...Y_n)^T$ is a response variable.

We now follow Newey (1997) for the asymptotic properties of the series estimator. For the consistency result, the following conditions will be imposed.

Assumption 1 $\{X_i, Y_i\}_{i=1}^n$ are i.i.d. and $Var(Y|X = x)$ is bounded on $\mathcal{X}$, the compact support of X.

Assumption 2 For every K there is a nonsingular $K \times K$ constant matrix B such that, for $P_K(x) = B\Phi_K(x)$,

(i) the smallest eigenvalue of $E[P_K(X_i)P_K(X_i)^T]$ is bounded away from zero uniformly in K;

(ii) there is a sequence of constants $\zeta_0(K)$ satisfying the condition $\sup_{x\in\mathcal{X}} ||P_K(x)|| \leq \zeta_0(K)$, where $K = K(n)$ such that $\zeta_0(K)^2K/n \to 0$ as $n \to \infty$;

(iii) there exists $\alpha > 0$ and β_K such that $\sup_{x\in\mathcal{X}} |g(x) - \Phi_K(x)^T\beta_K| = O(K^{-\alpha})$;

(iv) as $n \to \infty$, $K \to \infty$ and $K/n \to 0$.

Assumption 2 imposes nonsingular linear transformations of $\Phi_K(x)$ using matrix B, which can be interpreted as the normalization of the basis functions. Assumption 2(i) imposes a nonsingular second moment matrix of basis functions. Non-singularity necessitates having non-zero eigenvalues of this expectation as the determinant of this matrix equals to the multiplication of eigenvalues. Assumption 2(ii) restricts the magnitude of the basis functions along with a condition on the size of the series terms, K, i.e., $\zeta_0(K)^2K/n = o(1)$. Assumption 2(i)-(ii) are used in establishing the convergence of the sample second moment matrix of the approximating functions, $P_K^T P_K/n$, to the expectation of it. Assumption 2(iii) is a condition on the approximation error under the strong norm L_∞, i.e., *sup*-norm. It states that the supremum of the series approximation error on $\mathcal{X}$ is of order $K^{-\alpha}$, where α is related to the smoothness of $g(x)$ and the dimension of X. Intuitively, approximation error is of smaller order as more series terms are included in the approximation process, i.e., $K \rightarrow \infty$. Note that this relation can be also viewed as the reduction in the bias of the estimation as K gets larger. As an analogy, in the linear regression model, the variation in the response variable can be better explained if more and more independent variables are added in the regression. But, this has a drawback as the consistency of an estimator requires a balanced bias square and variance. Adding more explanatory variables into the regression increases the variance of an estimator, which is because of an increase in multicollinearity with additional variables. Therefore, it can be said that K has a similar role as the bandwidth parameter h had in the kernel methods. Lastly, Assumption 2(iv) ensures the consistency of the function $g(x)$ depending on the size of the basis, K, and the sample size, n.

The following theorem establishes the consistency of $\hat{g}(x)$.

Theorem 3 Under Assumptions 1 and 2, we have

(i) $\int [g(x) - \hat{g}(x)]^2 dF(x) = O_p(K/n + K^{-2\alpha})$.

(ii) $n^{-1} \sum_{i=1}^{n} [g(X_i) - \hat{g}(X_i)]^2 = O_p(K/n + K^{-2\alpha})$.

(iii) $\sup_{x \in \mathcal{X}} |g(x) - \hat{g}(x)| = O_p(\zeta_0(K)(\sqrt{K}/\sqrt{n} + K^{-\alpha}))$, where $F(\cdot)$ is the CDF of X.

Theorem 3(i) is the result for the mean integrated squared error as the cumulative distribution function of X is used as a weight for each value of X. Theorem 3(ii) provides the convergence rate for the sample mean squared error of $\hat{g}$. Both convergence rate can be stated as $O_p(K/n) + O(K^{-2\alpha})$, where the terms correspond to the variance and the bias squared, respectively. Note that the variance term has an order of K/n in probability as the randomness comes from the estimator, whereas the bias squared term is deterministic as it is derived from the approximation error $g(x) - \Phi_K(x)^T \beta_K$. Recall that the optimal convergence rate is achieved when the two terms, variance and squared bias, go to zero at the same rate. Therefore, when K diverges to infinity at a rate of $n^{1/(1+2\alpha)}$, the convergence of $\hat{g}$ to g in the L_2-norm will be at a rate of $n^{-\alpha/(1+2\alpha)}$. Uniform convergence rate provided in Theorem 3(iii), on the other hand, will not attain the bound of Stone (1982). This is because the *supremum* of the absolute distance between the estimator and the true function for all $x \in \mathcal{X}$ is of order that depends on the term $\zeta_0(K)$.

For power series and regression splines, Newey (1997) imposes primitive conditions in order to hold Assumption 2.

Assumption 3 (i) The support of X is a Cartesian product of compact connected intervals on which X has a probability density function that is bounded away from zero. (ii) $g(x) = E[Y|X]$ is continuously differentiable of order s on the support of X.

Assumption 4 The support of X is $[-1, 1]^d$.

Assumption 3(i) is a necessary condition for both power series and splines to satisfy Assumption 2(i) and (ii). Degree of the differentiability of the unknown function $g(x)$ given in Assumption 3(ii) is needed to be specific about the **uniform** approximation rate, which is determined by α in Assumption 2 (iii). Particularly, following Newey (1997), under Assumption 3(ii) the rate of approximation error bound is $\alpha = s/d$, where d is the dimension of X (The proof of this equality is provided in Lorentz (1986, Theorem 8, p.90).

The support of X given in Assumption 4 should be known along with Assumption 3(i) for splines in order to hold Assumption 2 (i) and (ii). Power series, on the other hand, do not necessitate this restriction.

Notice that Theorem 3(i) and (ii) are enough to express the consistency of the series estimator. However, the ultimate goal is to show the attainability of the uniform convergence rate of the estimator to the bound of Stone (1982). It should be noted that the convergence rate of Newey (1997) is obtained under the conditional moment restriction of Assumption 1. de Jong (2002), on the other hand, provides an improvement on Newey (1997)'s result imposing a conditional fourth moment condition $E[(Y_i - E[Y_i|X_i])^4|X_i] \leq C$ for some constant C, but unable to attain optimal rate convergence of Stone (1982) (de Jong, 2002, Assumption 4). The convergence rate of a series estimator under the extra conditional moment condition is established as

$$\sup_{x \in \mathcal{X}} |g(x) - \hat{g}(x)| = O_p(\zeta_0(K)((log(n)/n)^{1/2} + K^{-\alpha}),$$

provided that $K/log(n) \to \infty$ as $n \to \infty$. In a recent paper, Chen and Christensen (2015) establish the optimal uniform convergence rate for splines and wavelet least squares estimators under a weaker unconditional moment condition $E[|\epsilon_i|^{2+(d/p)}] < \infty$ for i.i.d. data and weakly dependent regressors. In that paper, $\{\epsilon_i\}_{i=1}^n$ is assumed to be a strictly stationary martingale difference sequence and d and p are the usual notations for the number of regressors and degree of smoothness of an unknown function, respectively.

2.4 Smooth Coefficient Semiparametric Models

Besides its flexibility in estimating an unknown regression curve, nonparametric method has some weaknesses, particularly, less interpretability of the regression function estimate and reduction in statistical precision. The latter is exacerbated in multidimensional setting as the fact that only few observations can be used in the regression estimate in the local neighborhood of point of interest. Furthermore, it also depends on the number of series terms in the approximation of an unknown function, which is discussed in detail in the next section. From the statistical point of view, a sharp decrease in the accuracy of the estimate is expressed by the slow convergence rates. Scott (1992) provides an intuitive explanation to the high dimensionality problem naming it as the “paradox of neighborhoods in higher dimensions". Because of the sparsity of data in a multidimensional setting, a local neighborhoods mean almost surely *empty*, whereas a non-empty neighborhoods does not mean that it is *local*. This is statistically formulated as the difficulty of balancing bias-variance trade-off in higher dimensions without very large sample sizes. Semiparametric models, on the other hand, are used to mitigate this problem by reducing the effective dimension size, while having the flexibility of the pure nonparametric models. These models also retain the interpretability of the parametric regression models. See Horowitz and Lee (2002) for a good discussion on some commonly used semiparametric models and their applications.

Much development on smooth coefficient semiparametric models has been made after the work of Cleveland et al. (1992) and Hastie and Tibshirani (1993). Smooth coefficient models enable a researcher to interpret the coefficient estimates in a parsimonious way as it is linear in regressors, but the coefficients are unspecified functions of other covariates satisfying some regularity conditions. Furthermore, it has a wide range of applicability if the researcher wants to investigate an interaction between regressors and some covariates without a priori considerations, but at the same time does not want to assume fully parametric specification, which might result in modeling bias and, thus, misleading inference if not correctly specified.

In an i.i.d. setting, we follow Li et al. (2002a) for large sample distributions of the local least squares kernel estimator of the standard varying coefficient models. Semiparametric analogues to Equation (1.1) in the form of smooth coefficient model is:

$$Y_i = \alpha(Z_i) + X_i^T \beta(Z_i) + \epsilon_i \equiv \mathcal{X}_i^T \delta(Z_i), i = 1, 2, ..., n, \tag{2.17}$$

where $\delta(Z_i) = (\alpha(Z_i), (\beta(Z_i)^T))^T$ is a vector of smooth unspecified functions of Z_i and Z_i is assumed to be a scalar random covariate for ease of exposition. Based on estimation method employed by Li et al. (2002a), local constant kernel estimate of $\alpha(z)$ and $\beta(z)$ are calculated using the observations in the interval $[z - h, z + h]$. It is similar to Equation (2.6), but now we are using Z covariate as a variable that the weight function is constructed on. Asymptotic results of the estimator is established under the following assumptions. Let $f(x, z)$ and $f(z)$ denote the joint density function of (X, Z) and the marginal density function of Z, respectively.

2.5 Endogeneity Problem in Nonparametric Regression Estimation

Endogeneity, originally arose from the context of simultaneous relationship between response variable and explanatory variables, is an important problem that should be addressed correctly in order to obtain consistent estimation. In a nonparametric regression, dealing with endogeneity requires identification and estimation of the structural model rather than estimation of only the conditional expectation, which we have discussed in detail so far in the previous sections.

The estimation methodology of nonparametric models with endogeneity includes the instrumental variable approach (see Newey and Powell (2003) and the control function approach; for a comparison of the two approaches, see Blundell and Powell (2003)). Note that we followed the latter approach for the identification and estimation of functional coefficients in the smooth coefficient model investigated in my second empirical paper. The former is employed in my first study.

2.5.1 *Fully nonparametric models with endogeneity*

In contrast to the identification of parameters in the standard linear simultaneous equations model, the parameters of interest in a nonparametric model are the unknown functional form of the conditional expectations and unknown conditional density functions. A lack of prior information on the structural form of the relationship between response variable and explanatory variables is the key to employ nonparametric conditions for identification. Nevertheless, knowledge of characteristics of the structure has to be part of the prior information in the nonparametric models in order to establish consistency result. Roehrig (1988) provides an identification result for a nonparametric structural model under the independence assumption of errors and instruments. Newey and Powell (2003) weaken this restriction by imposing zero conditional mean of the errors given the instruments, which is essentially to strengthen the unconditional moment restriction of parametric simultaneous equations model. In Newey et al. (1999) identification condition is imposed with an additive structure using the control function approach. We consider fully nonparametric system of equations model following Newey et al. (1999).

$$Y = m(X, Z_1) + \epsilon, \qquad X = \Pi(Z) + u, \tag{2.18}$$

$$E[\epsilon|u, Z] = E[\epsilon|u], \qquad E[u|Z] = 0,$$

where X is an endogeneous variable with a dimension d_X, Z_1 is a d_{11} dimensional subvector of $Z = (Z_1, Z_2)$, which has a dimension d_1, $\Pi(Z)$ is a d_X dimensional vector of functions of instruments. Identification strategy of Newey et al. (1999) is based on assumptions given in Model (2.18). Formally, for $b(u) = E[\epsilon|u]$, Model (2.18) implies an additive structure in Equation (2.19), which is equivalent to the conditional mean restriction given in Model (2.18), provided that b is an unknown function. Then, we have

$$E[Y|X, Z] = m(X, Z_1) + E[\epsilon|X, Z] = m(X, Z_1) + E[\epsilon|u, Z] \tag{2.19}$$

$$= m(X, Z_1) + b(u) \equiv h(W), \qquad W = (X', Z_1', u')'.$$

It is noted that since u is identified, the identification of m under Equation (2.18) is the same as the identification of the additive component, m, in Equation (2.19). Newey et al. (1999) provide that absence of functional relationship between (X, Z_1) and u is a sufficient condition for $m(X, Z_1)$ to be identified up to an additive constant. Furthermore, Newey et al. (1999) generalize the rank condition for parameters of linear structural simultaneous equation given in the next lemma.

Lemma 1 If $m(X, Z_1)$, $b(u)$, and $\Pi(z)$ are differentiable, the boundary of the support of (Z_i, u_i) has zero probability, and with probability one, the rank of $\partial\Pi(z)/\partial z_2$ is d_X, then $m(x, z_1)$ is identified.

In a recent survey, Chen and Qiu (2016) discuss the advantages of using series approximation methods in the case of endogeneity in the nonparametric models. They emphasize that series methods is one of the common approaches to obtain the asymptotic variance of the final-stage estimator in the control function approach. Specifically, in this estimation methods, the residuals from the first step of estimation is used to control for the endogeneity of the regressors; see Newey et al. (1999) and Pinkse (2000) for further details on the estimation procedure. We next provide the results for the mean-square and uniform convergence rates of the series estimator following Newey et al. (1999). Let $\mathcal{W} = \{w : \tau(w) = 1\}$.

Assumption 1 $\{Y_i, X_i, Z_i\}, (i = 1, 2, ...)$ is i.i.d. and $var(X|Z)$ and $var(Y|X)$ are bounded.

Assumption 2 Z is continuously distributed with density that is bounded away from zero on its support, and the support of Z is a cartesian product of compact, connected intervals. Also, W is continuously distributed and the density of W is bounded away from zero on $\mathcal{W}$, and $\mathcal{W}$ is contained in the interior of the support of W.

Second part of the Assumption 2 ensures that there is no functional relationship between (X, Z_1) and u, which is the main idea for the identification of an unknown function, m.

Therefore, under Assumption 2, by Theorem 2.2 of Newey et al. (1999), identification of m holds.

Assumption 3 $\Pi(Z)$ is continuously differentiable of order s_1 on the support of Z, and $m(X, Z_1)$ and $b(u)$ are Lipschitz and continuously differentiable of order s on $\mathcal{W}$.

Assumption 4 For power series, $(K^3 + K^2 L)[(L/n)^{1/2} + L^{-s_1/d_1}] \to 0$; for splines, $(K^2 + KL^{1/2})[(L/n)^{1/2} + L^{-s_1/d_1}] \to 0$.

Assumption 3 imposes smoothness conditions on unknown functions in order to control the bias of the series estimators. Moreover, the degree of differentiability s_1 and s control the rate of approximation of unknown functions. Recall Assumption 3 in Newey (1997) that for power series and splines, the rate of approximation for $m(X, Z_1)$ and $b(u)$ will be $O(K^{-s/d})$, where d is the dimension of (X, Z_1) for polynomials. Lastly, Assumption 4 is the regularity condition on the growth rate of the number of series terms K and L in each step of the series estimation. Next theorem provides the consistency of the estimator proposed in Newey et al. (1999).

Theorem 5 Under Assumptions 1-4, setting $q = 1/2$ for splines and $q = 1$ for power series, then

(i) $\int \tau(w)[\hat{m}(w) - m(w)]^2 dF(w) = O_p((K/n) + K^{-2s/d} + L/n + L^{2s_1/d_1})$ and

(ii) $\sup_{(w \in \mathcal{W})} |\hat{m}(w) - m(w)| = O_p(K^q[(K/n)^{1/2} + K^{-s/d} + (L/n)^{1/2} + L^{-s_1/d_1}])$.

2.5.2 *Varying coefficient models with endogeneity*

We now turn our attention to the functional coefficient model and the identification problem associated with unknown coefficient functions. Following Cai et al. (2006), we first consider a smooth coefficient IV model, in which endogeneity arises in a parametric part.

$$\begin{cases} Y = g(X, Z_1) + \epsilon, \\ g(X, Z_1) = \sum_{j=0}^{d} g_j(Z_1) X_j, \\ E(\epsilon | \mathcal{Z}) = 0, \end{cases}$$

with the same dimensional specification as in the previous model, and $\{g_j(\cdot)\}$ are the unknown structural functions of interest. The reduced form is then defined as

$$E(Y|\mathcal{Z}) = \sum_{j=0}^{d} g_j(Z_1) E(X_j|\mathcal{Z}) = \sum_{j=0}^{d} g_j(Z_1) \pi_j(\mathcal{Z}),$$

where $\pi_j(\mathcal{Z}) = E(X_j|\mathcal{Z})$ is an unknown function for each $j = 1, 2, ..., d$.

For an identification of $g(X, Z_1)$, Cai et al. (2006) first assume an identification of $\pi_j(\mathcal{Z})$s, $j = 1, 2, ..., d$, in the sense that the uniqueness of the conditional expectations implies an identification of these functions up to an additive constant. Cai et al. (2006) provides the same sufficient condition as in Lemma 1 for the identification of $g(\cdot)$, which mainly states that the dimension of $\mathcal{Z}$ is at least as large as the dimension of $\{(X, Z_1)\}$.

Cai et al. (2006) propose a two-stage estimation procedure, in which the local linear kernel estimation is employed. Following assumptions provide regularity conditions for the large sample distribution results of local linear kernel estimators when an endogeneity arises in the model. Let $\xi = X - E(X|\mathcal{Z})$ and $\eta = Y - E(Y|\mathcal{Z})$. Define $\sigma_\eta^2 = E(\eta^2|\mathcal{Z})$, $\Sigma_\xi(\mathcal{Z}) = E(\xi\xi^T|\mathcal{Z})$, and $\sigma_{\eta\xi}(\mathcal{Z}) = E(\eta\xi|\mathcal{Z})$. $\Omega_0(z_1) = E(\pi(\mathcal{Z})\pi^T(\mathcal{Z})|Z_1 = z_1)$. h_1 and h_2 are the bandwidth parameters in the first and second step of the estimation, respectively.

Assumption 1 For the regularity conditions on the kernel functions, see Assumption 4 on page 4.

Assumption 2 $\{(X_i, Y_i, \mathcal{Z}_i)\}, (i = 1, 2, ...)$ is i.i.d. and the fourth moment of ξ and η exists.

Assumption 3 As $n \to \infty$, $nh_1^{d_1}/\text{log}n \to \infty$, $h_1 \to 0$, $nh_2^{d_{11}} \to \infty$, $h_2 \to 0$.

Assumption 4 $\mathcal{Z}$ has a uniformly continuous density function bounded away from zero on its support, and the support of $\mathcal{Z}$ is a compact set.

Assumption 5 $\{g_j^{''}(\cdot)\}$, $\sigma_\eta^2(\cdot)$, $\Sigma_\xi(\cdot)$, and $\sigma_{\eta\xi}(\cdot)$ are continuous at the point z_1. Moreover, $\{\pi_j^{''}(\cdot)\}$ are bounded and uniformly continuous and satisfy the Lipschitz condition.

Assumption 6 $\Omega_0(z_1)$ is positive definite.

Assumption 7 $E|X_j|^\gamma < \infty$ for some $\gamma > 2$.

Assumption 8 $h_1 = o(h_2)$.

Assumptions 1-5 are the standard conditions in the literature for the local polynomial estimation methods. Assumption 6 is a necessary and sufficient identification condition, so that by Corollary 1 in Cai et al. (2006) identification holds. Assumptions 7 and 8 are required for the two-stage estimation method in standard nonparametric regression. Assumption 8 imposes undersmoothing in the first stage of the estimation procedure, so that the bias from the first stage can be ignored.

Theorem 6 Under Assumptions 1-8,

$$\hat{g}(z_1) - g(z_1) = bias_g(z_1) + o_p(h_2^2) + O_p((nh_2^m)^{-1/2}), \qquad where$$

$$bias_g(z_1) = \frac{h_2^2}{2} tr\{\mu_2(L) g_j^{''}(z_1)\}.$$

The consistency of the final stage estimator is not affected by the bias of the first stage estimator as $bias_g(z_1)$ depends only on the final stage bandwidth h_2 and the second derivative of $\{g_j(\cdot)\}$.

2.6 Nonparametric Panel Data Regression Models

Following a recent survey by Sun et al. (2014), our aim, in this section, is to provide a brief overview of the estimation methods proposed for nonparametric panel data regression models, where the conditional mean of dependent variable is unknown to the researcher. We consider a nonparametric panel data model

$$Y_{i,t} = m(X_{i,t}) + u_{i,t}, \qquad i = 1, ..., n, t = 1, ..., T, \tag{2.20}$$

where $Y_{i,t}$ and $X_{i,t}$ are a scalar and $p \times 1$ vector of real-valued random variables, respectively, $m(\cdot)$ is unknown smooth function, and $u_{i,t}$ is the disturbance term possessing the one-way error component structure, $u_{i,t} = \mu_i + \epsilon_{i,t}$. Specifically, μ_i represents the cross sectional heterogeneity, for which it can be treated as fixed or random, and $\epsilon_{i,t}$ is the error term satisfying the standard assumptions. Note that this error structure includes only time-invariant component μ_i, which is assumed to be independently and identically distributed with mean zero and variance σ_μ^2. We first introduce estimation methodology of nonparametric panel data models with random effects.

2.6.1 *Random effects panel data models*

The random effects panel data models assume $E(u_{i,t}|X_{i,t}) = 0$ for all i and t for large n and small T. This assumption says that unobserved heterogenous differences across cross sectional units are uncorrelated with the regressors. The correlations between the error terms $u_{i,t}$ is the main point that we need to account for in the estimation methods. Note that the correlations between two cross sectional units at the same and different time periods are ruled out in this model; see Hill et al. (2008). The correlation of a cross sectional unit at different time periods is not equal to zero and expressed in (nT) by (nT) conditional covariance matrix $V = E(uu^T|X) = diag(V_1, ..., V_n)$ with $T \times T$ matrix $V_i = E(u_i u_i^T|X_i)$.

The local linear kernel estimation method is employed to estimate the unknown conditional mean regression model given in (2.20). Recall from Section 2 that Taylor expansion provides a linear approximation to the unknown function, so that Model (2.20) can be written as

$$Y_{i,t} \approx m(x) + m^{'}(x)(X_{i,t} - x) + u_{i,t}, \tag{2.21}$$

where we are interested in estimating $m(x)$ for each x in its support. Denoting $\beta(x) = [m(x), m^{'}(x)]^T$, the local linear weighted least squares problem is

$$\min_{\beta(x)}[Y - Z(x)\beta(x)]^T W(x)[Y - Z(x)\beta(x)], \tag{2.22}$$

where the i^{th} row vector of $nT \times 2$ matrix $Z(x)$ is $(1, X_{i,t} - x)$ and $W(x)$ is a $nT \times nT$ kernel-based weight matrix. The local linear weighted least squares estimator $\hat{\beta}$ solves the problem (2.22) and can be written as $\hat{\beta}(x) = [Z(x)^T W(x) Z(x)]^{-1} Z(x)^T W(x) Y$. When $W(x) = K_h(x)$, $\hat{\beta}(x)$ becomes the local linear least squares estimator, for which the weight matrix has the same role as in the previous sections assigning a positive weight to $X_{i,t}$ only if $|X_{i,t} - x| \leq ch$ for some positive constant c, where h is the bandwidth satisfying $h \to 0$ as $n \to \infty$. It is the crucial point that the possible correlation between $u_{i,t}$ and $u_{i,s}$ for $t \neq s$ can be accounted for when the weight matrix is constructed as it is set $W(x) \equiv V^{-1/2} K_h(x) V^{-1/2}$ as in Ullah and Roy (1998) or $W(x) \equiv K_h^{1/2}(x) V^{-1} K_h^{1/2}$ as in Lin and Carroll (2000), where V is the covariance matrix. In other words, the latter estimator ignores the within-group correlation in errors, which may result in misleading inference, but still asymptotically more efficient than the former obtained by using the weight matrix suggested in Lin and Carroll (2000). However, Wang (2003) and Welsh et al. (2002) demonstrated that kernel and smoothing spline estimators for the model having a correlation structure of the errors is more efficient than the independent structure. Ruckstuhl et al. (2000) proposed a two-step estimator, which has asymptotically smaller variance than the local linear least squares estimator under some conditions. Su and Ullah (2007) showed that the two-step estimator is asymptotically more efficient than the local linear least squares estimator under the condition that the bandwidth in the first step has a faster convergence rate than the bandwidth in the second step, which has an optimal rate.

We would like to note that the papers discussed so far assume a known or pre-determined covariance matrix of the errors, which, results in an infeasible estimator as the covariance structure is often unknown. Henderson and Ullah (2005) proposed a feasible version of the estimators of Lin and Carroll (2000) and Ullah and Roy (1998). Fan et al. (2007) and Qu and Li (2006) estimate the covariance matrix prior to the estimation of unknown regression function, where the former follows two-step estimation procedure and the latter approximates the unknown correlation matrix by basis functions. A recent paper by Yao and Li (2013) propose a simultaneous estimation of covariance matrix and $m(x)$ using Cholesky decomposition and profile least squares techniques.

2.6.2 *Fixed effects panel data models*

In the fixed effects panel data models, the error term $u_{i,t}$ given in Model (2.20) is allowed to be correlated with the regressors $X_{i,t}$. This is, in fact, because of the possible correlation between unobserved cross-sectional fixed effects μ_i and the regressors. μ_i's maintain the same distributional assumption given previously. Considering the same error structure of $u_{i,t}$ as $u_{i,t} = \mu_i + \epsilon_{i,t}$, there are two more cases depending on whether $\epsilon_{i,t}$ is correlated with the regressors or not. We first discuss the estimation procedure under the independence assumption between $\epsilon_{i,t}$ and $X_{i,t}$. The first paper dealing with fixed effects in nonparametric panel data models is Henderson et al. (2008). They estimate unknown function $m(x)$ applying a first-differencing method and backfitting algorithm; i.e., an iterative nonparametric kernel estimator. A simpler estimator was proposed by Sun et al. (2009) for the fixed effects panel data varying coefficient models. Their estimator is a nonparametric version of a least squares dummy variable estimator in parametric panel data analysis. Recently, Lin et al. (2014) proposed a consistent test statistic for the null hypothesis of a linear functional form against a nonparametric alternative under the fixed effects panel data model framework using a non-iterative kernel estimator motivated by Sun et al. (2009). We, therefore, rewrite model (2.20) as in the following matrix form, which incorporates all individual-specific effects as dummy variables.

$$Y = m(x) + D\mu + U, \tag{2.23}$$

where $D = I_n \otimes e_m$ is an $(nm) \times n$ matrix with main diagonal blocks being e_m and e_m is the m-dimensional vector of ones. Su and Ullah (2006) assume $\sum_{i=1}^{n} \mu_i = 0$ for the identification of this model. However, Sun et al. (2009) show that i.i.d. structure of μ_i's with mean zero and variance σ_μ^2 is a weaker condition that provides an identification of this model in an asymptotic case. Recently, Lin et al. (2014) show that above condition can be dropped for large T of order $O(n^{1/4})$ and large n. Two-step estimation procedure has been proposed to solve the following optimization problem.

$$\min_{\mu, m(\cdot)} (Y - m(X) - D\mu)^T K_h(x)(Y - m(X) - D\mu). \tag{2.24}$$

In the first step, μ will be estimated assuming that $\sum_{i=1}^{n} \mu_i = 0$ and $m(\cdot)$ has a known functional form. Notice that $K_h(x)$ ensures the locality of nonparametric estimation. In other words, $\hat{\mu}$ is obtained using the data points close to x. The first order derivative of problem (2.24) with respect to μ gives

$$D^T K_h(x)[Y - m(X) - D\hat{\mu}(x)] = 0, \tag{2.25}$$

which yields

$$\hat{\mu}(x) = (D^T K_h(x) D)^{-1} D^T K_h(x)[Y - m(X)]. \tag{2.26}$$

In order to estimate $m(\cdot)$, we need to remove the fixed effects term μ from Model (2.24). Therefore, a new weight matrix will be constructed by specifying $S_h(x) = M_h(x)^T K_h(x) M_h(x)$ and $M_h(x) = I_{n\times m} - D(D^T K_h(x)D)^{-1} D^T K_h(x)$, where $I_{n\times m}$ denotes an identity matrix of dimension nm by nm. Then, we replace μ in (2.24) by $\hat{\mu}(x)$, so that we obtain the following weighted least squares problem.

$$\min_{m(x)} [Y - m(x)]^T S_h(x)[Y - m(x)]. \tag{2.27}$$

The local constant kernel estimator then obtained as

$$\tilde{m}(x) = n^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \omega_{i,t} Y_{i,t}, \tag{2.28}$$

where $\omega_{i,t} = \lambda_{i,t} / \sum_{t=1}^{T} \lambda_{i,t}$ and $\lambda_{i,t}$ is the $(i,t) - th$ element of S_h matrix, for which $K_h(x) = ((X_{i,t} - x)/h)$; for further details, see Sun et al. (2014) and Lin et al. (2014).

2.7 Nonparametric Bootstrap Methods

This section provides underlying conceptual and theoretical framework of bootstrap methods, which has a wide application in nonparametric regression context in recent years. Bootstrap method is a data resampling technique to obtain standard errors and other statistics as well as to construct confidence intervals; for an introduction to the bootstrapping method, see the monograph of Efron and Tibshirani (1993). How accurate an estimator is generally assessed by approximating its distribution asymptotically; i.e., evaluating an estimator's limiting results as the sample size goes to infinity. But, generally, how large a sample should be to satisfy an estimator's asymptotic distribution is not clear and assessed by estimator's finite sample performance using simulation studies such as Monte Carlo methods. However, these methods rely on randomly generated variables based on the distributional assumption of the variables in the data. The bootstrap method, on the other hand, is a data-based simulation method, which allows to mimic the data closely, so that provides more accurate approximation than the asymptotic theory does in practical applications. Horowitz (1997) well explains that an asymptotic approximation has an error of order of $O(n^{-1})$, whereas the bootstrap makes an error of size $o(n^{-1})$, which indicates a smaller order than n^{-1}, and as a result, an error from the bootstrap approximation decreases to zero faster with an increasing sample size than an error does from the asymptotic theory.

We will now discuss three different ways of bootstrapping summarized in Härdle and Mammen (1991). The first way is the näive approach based on resampling the pairs $\{(X_i, Y_i)\}_{i=1}^{n}$ according to the following model.

Model 1: *The pairs* (X_i, Y_i) *are independent, identically distributed random variables with* $E(\epsilon_i|X_i) = 0$. *Then* $m(x) = E(Y_i|X_i = x)$.

This approach can be seen inappropriate as the bootstrap bias is asymptotically zero, which is not desirable since the nonparametric estimators have intrinsic biases; see our discussion on the bias-variance trade-off in the nonparametric regression estimation in Section 2. The second approach is known as the residual-based bootstrapping, which mimics the random nature of the following model.

Model 2: *The* ϵ_i's *are independent, identically distributed random variables with* $E(\epsilon_i) = 0$. *The* X_i are deterministic.

This method of bootstrapping was used in our first empirical application, Koroglu and Sun (2016). We were interested in calculating confidence intervals for the average direct and indirect impacts of the regressors in the SDM model. However, since we have intercept term and slope coefficient curve estimates as of dimension $4.n$ as well as estimated spatial weight matrices, G_n and M_n, as of dimension $2.n.(n-1)$, where the diagonals of the matrices are zero, in that model, it would not be feasible to simulate a joint distribution with this huge dimension. Therefore, we employed this second procedure to obtain bootstrap estimates of the confidence intervals of the averages of the respective impacts, for which the bootstrap method provides well approximation. It should be noted that the bootstrap sample in the resampling step in Koroglu and Sun (2016) is obtained using oversmoothed bandwidth to account for the bias *implicitly*. Härdle and Bowman (1988) uses the optimal bandwidth, which has an order of $n^{-1/5}$, in both the resampling and the estimation step, which necessitates an *explicit* estimation of the bias of the bootstrap estimator. Following Härdle and Mammen (1991), we impose a condition on bandwidth parameters h and g such as $h \sim n^{-1/5}$, $g \to 0.$ and $g/h \to \infty$, which indicates that g has a slower convergence rate than the optimal bandwidth h. It should be noted that residual resampling method needs error structure to be homoscedastic. Moreover, this method of bootstrapping is not robust when there is a variability in the error variance; i.e., the case of heteroscedasticity (Wu (1986)). Identical distribution assumption of errors, ϵ_i's, is relaxed in the following model.

Model 3: *The* ϵ_i's *are independent random variables with* $E(\epsilon_i) = 0$. *The* X_i are *deterministic. The distribution of the errors may depend on the design variable.*

The bootstrap procedure that provides a good approximation to the random structure of Model 3 is known as the wild bootstrap, which was first studied by Wu (1986) and further developed by Mammen (1993). Härdle and Mammen (1993) used the wild bootstrap to compare nonparametric and parametric regression fits. Cao-Abad (1991) derives the convergence rate of the wild bootstrap estimator of nonparametric regression model. We employed the wild bootstrapping in our second study to calculate robust standard errors of the coefficient curve estimates as well as mean estimates reported in Table 4.1. The main idea here is to draw each bootstrap residual, call ϵ_i^*, from a two point distribution $\hat{G}_i$, which satisfies the following moment conditions:

$$E(\epsilon_i^*) = 0, \qquad E(\epsilon_i^{*2}) = \hat{\epsilon}_i^2, \qquad \text{and} \qquad E(\epsilon_i^{*3}) = \hat{\epsilon}_i^3.$$

The distribution is defined as $\hat{G}_i = \gamma\delta_a + (1-\gamma)\delta_b$, where γ, a, and b are the parameters and δ_a and δ_b represent point measures at a and b, respectively. Then, these parameters at each point X_i is algebraically calculated as $a = \hat{\epsilon}_i(1-\sqrt{5})/2$, $b = \hat{\epsilon}_i(1+\sqrt{5})/2$ and $\gamma = (5+\sqrt{5})/10$, which satisfy above moment conditions (Härdle and Marron, 1991). The same procedure will be followed as in Koroglu and Sun (2016) to obtain new bootstrap sample using oversmoothed bandwidth.

2.8 Conclusion

In this paper, I survey some estimation methodologies in nonparametric econometrics, including (i) the local least squares kernel estimator; (ii) nonparametric series estimator; (iii) estimation of nonparametric models with endogeneity; and (iv) nonparametric estimation of panel data models. I also survey different bootstrapping methods for nonparametric regression methods. Cai et al. (2009) give a survey on the recent developments of nonparametric estimation and testing of regression functions with various data examples.

Chapter 3

Functional-Coefficient Spatial Durbin Models with Nonparametric Spatial Weights: An Application to Economic Growth

3.1 Introduction

Ever since the seminal work of Mankiw et al. (1992), there has been a significant amount of empirical work studying variation in economic growth rates across countries. Particularly, more and more economists have paid attention to the impacts of economic interaction and spillover effects on the regional and national economy in the past two decades;

see, e.g., Backus et al. (2008) for taxation and the global allocation of capital, Baltagi et al. (2008) for cross-border foreign direct investment decisions, Ertur and Koch (2007, 2011) for economic growth models with worldwide interactions, Cassette et al. (2013) for country interactions in discretionary fiscal policy and citetbru03 for an overview of empirical studies of strategic interaction among governments over environmental standards and public expenditures. In the meanwhile, econometric theory in parametric spatial regression models has been introduced and well developed to analyse the spatial and economic externalities; for a detailed survey on parametric spatial econometric models, see Cliff and Ord (1981), Anselin (1988), LeSage and Pace (2009) and Anselin and Florax (2011). This paper joins the others to examine the impact of cross-country economic externalities on national growth through a Solow growth model augmented with economic externalities.

The role of spatial dependence in regional economic growth has received substantial attention in the empirical growth literature in the recent decade; see, e.g., Abreu and Groot (2005) for a survey on economic growth and space. It has been recognized that a nation's per capita GDP growth rate is affected not only by its own values of determinants, such as savings, population growth rate and initial level of income, but also by its neighbouring nations' per capita GDP growth rates and the values of these determinants. For example, Ertur and Koch (2007) developed a theoretical growth model with spatial externality resulting from technological interdependence among economies and proposed a spatially-augmented Solow economic growth model yielding a conditional convergence equation with heterogeneous Solow parameters. Note that heterogeneous Solow parameters are also supported by similar studies with no spatial interactions; see, e.g., Durlauf et al. (2001), Mamuneas et al. (2006) and Kourtellos (2011).

Fitting a spatial Durbin model (SDM) using data from 91 non-oil regions/and countries for the period from 1960 to 1995, Ertur and Koch (2007) found positive and significant spatial dependence across these economies together with predicted signs for all coefficients. However, their study suffers from two potential problems. First, the parametric SDM requires researchers to pre-determine the non-stochastic spatial dependence structure among economies before estimating parameters appearing in the model, and the misspecified spatial interactive relations can incur inconsistent estimation and misleading inference. Second, the subsampling method may not be the best way of studying heterogeneous Solow parameters. In this paper, working on the same dataset used in Ertur and Koch (2007), we therefore aim to re-examine the spatial spillover effects of economic growth, while estimating in a nonparametric way the true spatial dependence structure among economies and allowing the Solow parameters to vary with respect to the trade openness of an economy.

Specifically, we propose a functional-coefficient spatial Durbin model with nonparametric spatial weights and estimate the unknown spatial weights and coefficient curves via a series approximation approach by a nonparametric two-stage least squares method. Based on the first-step consistent estimator, we then construct a second-step estimator for the unknown functional coefficients, which is oracle efficient in the sense that the limit distribution of the second-step estimator is the same regardless of whether the spatial weights are known. Moreover, we give our inference on spatial dependence through average direct and indirect impact values with standard errors calculated from the bootstrap method.

The remainder of this paper is organized as follows. Section 2 introduces our proposed semiparametric spatial Durbin model. Section 3 presents our estimation methodology. Section 4 reports results from a small Monte Carlo simulation to examine the finite sample performance of our proposed estimators. Section 5 gives our empirical results. Section 6 concludes.

3.2 Model

In the empirical economic growth literature, DeLong and Summers (1991), to the best of our knowledge, is the first study to investigate spatial correlation taking geographical distance into account. Using a sample of 61 countries for the period from 1960 to 1985, they find no significant spatial correlation in their sample. Moreno and Trehan (1997), on the other hand, augment Mankiw et al. (1992)'s model with a spatial interactive term and find highly significant spillover effects between geographical neighbours, and they argue that using a border dummy variable instead of a spatial lag term neglects the influence of neighbour countries that do not have a common border with the country of interest; relevant literature includes Chua (1993), Ades and Chua (1997), and Barro and Salai-Martin (1995). Moreover, Ertur et al. (2006) provide strong evidence of spatial dependence in economic convergence processes among European regional economies. Using 155 European regions for the period 1988 to 2000, Basile (2008) also finds some evidence of spatial spillovers across countries.

The question of how to measure spatial interactive relations between any pair of spatial units is answered by defining a neighbourhood set for each spatial unit according to some selected relevant variables. For example, Cliff and Ord (1981) specify spatial weights for spatial unit i as the ratio of the length of the common border between units i and j to the geographical distance between them, $g_{ij} = b_{ij}^{\beta}/d_{ij}^{\alpha}$, with some parameters $\alpha > 0$ and $\beta > 0$. A common approach in practice is to use only distance-based weights with a decay parameter α; i.e., the spatial weight from unit j on unit i is defined as $g_{ij} = g\left(d_{ij}, \alpha\right)$, where $g\left(\cdot\right)$ is a known function and α is a parameter to be estimated. The popularly-used distance function includes an inverse power function $g_{ij} = 1/d_{ij}^{\alpha}$ or a negative exponential function $g_{ij} = exp(-\alpha d_{ij})$ (e.g., Murdoch et al. (1993) for some $\alpha > 0$. Moreover, if a cut-off distance is not used, then a non-sparse spatial weight matrix is constructed. This implies that every region is a neighbour of other regions, but the spatial weights depreciate as the distance between two regions increases. For recent studies on spatial weights, see LeSage (2014) and LeSage and Pace (2014). Moreover, the term "neighbours" can also refer to contiguities defined by economic distances (e.g., Case et al. (1993); Pinkse et al. (2002); Conley and Ligon (2002)) and social networks Lee (2007b).

As the functional form $g\left(\cdot\right)$ is unknown in practice, to avoid misspecifying the wrong spatial weighting function, one can alternatively estimate the unknown spatial weighting function $g\left(\cdot\right)$ from the data via a nonparametric series estimation method. Compared to the parametric spatial modelling approach, the nonparametric approach enables researchers to impose less restrictive assumptions on the spatial weight function; see Pinkse et al. (2002) and Sun (2016) for details. Alternatively, Ahrens and Bhattacharjee (2015) proposed to estimate the unknown spatial weights via the Lasso estimation method when the unknown spatial weights matrix is sufficiently sparse.

Ertur and Koch (2007) derive a theoretical Solow economic growth model augmented with global technological interdependence. They then approximate their theoretical model by a parametric spatial Durbin model via a linearization procedure and calculate the spatial weights from both the inverse power function and the exponential function of geographic distances with $\alpha = 2$ for the sake of robustness, as the true spatial weighting function is unknown. As the linearization and selected parametric spatial weights may result in a model misspecification problem, in this paper, in order to better approximate the theoretical model of Ertur and Koch (2007), we therefore propose a semiparametric growth model that extends parametric model of Ertur and Koch (2007) by allowing nonparametric spatial weights, as well as varying Solow coefficients. Specifically, our proposed functional-coefficient spatial Durbin model with nonparametric spatial weights is given by:

$$Y_i = \sum_{j \neq i} g(Z_{ij}) Y_j + \sum_{t=1}^{p} \sum_{j \neq i} m_t(Z_{ij}) X_{tj} + X_i^T \theta(D_i) + u_i, i = 1, ..., n, \tag{3.1}$$

where Y_i is a scalar dependent variable, $X_i = [X_{1i}, \ldots, X_{pi}]^T$ is a $p \times 1$ vector, D_i is a continuous scalar random variable and Z is a non-stochastic spatial covariate with $Z_{ii} = 0$ and $Z_{ij} > 0$ for $i \neq j$. Moreover, $g(\cdot)$, $m_t(\cdot)$ for $t = 1, ..., p$ and $\theta(\cdot)$ are all unknown measurable smooth functions with $g(0) = 0$ and $m_t(0) = 0$, $t = 1, .., p$. In Model (3.1), X_i has a functional coefficient depending on D_i, and the unknown spatial weights are a function of non-stochastic geographic distance Z_{ij} . The first two terms in the right-hand side of Equation (3.1), $\sum_{j \neq i} g(Z_{ij})Y_j$ and $\sum_{j \neq i} m_t(Z_{ij})X_{tj}$ for each $t = 1, ..., p$, are called the spatial lag of the dependent variable and the spatially-lagged exogenous variables, respectively. In the spatial econometrics literature, this kind of model specification is referred to as the spatial Durbin model, where (X_i, D_i, u_i) are independently distributed across i while only the dependent variable Y_i is dependently distributed across spatial units. The detailed data information on X_i, D_i, Y_i and Z_{ij} is delayed to Section 5. If the model includes only the spatial lag of the dependent variable, it is called a pure spatial autoregressive (SAR) model. Basile et al. (2014) introduced the spatial autoregressive semiparametric geoadditive models to account for spatial dependence, spatial unobserved heterogeneity and unknown functional curves of regressors simultaneously, where the spatial autoregression is represented by a pre-determined spatial lag term of the dependent variable.

Let G_n and $M_{n,t}$ be $n \times n$ unknown spatial weight matrices with its $(i, j)^{th}$ element being

$g_{ij} = g(Z_{ij})$ and $m_{t,ij} = m_t(Z_{ij})$, respectively, for $t = 1, ..., p$, $i = 1, ..., n$, $j = 1, ..., n$. We then obtain a reduced form of Model (3.1) written in matrix form:

$$Y = (I_n - G_n)^{-1} [\sum_{t=1}^{p} M_{n,t} X_t + mtk\{X, \theta(D)\} + U], i = 1, ..., n, \tag{3.2}$$

where $Y = [Y_1, \ldots, Y_n]^T$, $X_t = [X_{t1}, \cdots, X_{tn}]^T$ and $U = [u_1, \ldots, u_n]^T$ are all $n \times 1$

vectors and $mtk\{X, \theta(D)\}$ is an $n \times 1$ vector with the i^{th} element equal to $X_i^T \theta(D_i)$. Furthermore, I_n is an $n \times n$ identity matrix.

Let $\lambda_i(A_n)$ be the i^{th} eigenvalue of an $n \times n$ matrix A_n, $\rho(A_n) = \max_{1 \leq i \leq n} |\lambda_i(A_n)|$ and $\|A_n\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$ and $\|A_n\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|$ be the respective row and column norm of A_n. Furthermore, $C > 0$ is a finite positive number that takes different values at different appearances. Below, we impose some regularity conditions on Model (3.1).

Assumption A1: (i) $\{Y_i\}$ is generated from Model (3.1), and $\{(X_i, D_i)\}$ is independently distributed with finite second moments; (ii) $g(\cdot)$, $m_t(\cdot)$ and $\theta(\cdot)$ are all uniformly bounded up to their respective p^{th}-order derivatives for some $p > 2$; (iii) $\{u_i\}$ is an independent sequence with zero mean, $E[u_i | X_i = x, D_i = d] = 0$ and $E[u_i^2 | X_i = x, D_i = d] = \sigma_i^2(x, d) > 0$ for all i and $(x, d) \in R^p \times R$, and $\sup_{(x,d) \in R^p \times R} \max_{1 \leq i \leq n} E[|u_i|^{2+\delta} | X_i = x, D_i = d] \leq C < \infty$ for some $\delta > 0$ and a positive constant C.

Assumption A2: (i) There exist a positive integer N and a constant $c_G \in (0, 1)$, such that for all $n > N$, $\rho(G_n) \leq c_G$; (ii) $\|G_n\|_j \leq C < \infty$, $\|M_{n,t}\|_j \leq C < \infty$ for all t and $\left\|(I_n - G_n)^{-1}\right\|_j \leq C < \infty$ for $j = 1$ and ∞ and some finite value $C > 0$.

Assumption A1 (i) states that the explanatory variables (X_i, D_i) are independent, while the dependent variable Y_i exhibits spatial dependence; and Assumption A1 (iii) allows the error term, u_i, to be independent with heteroskedasticity, and the bounded higher order moment is required for deriving the limiting normal distribution of the proposed estimator. By Seber (2008, p. 421), Assumption A2 (i) ensures that $I_n - G_n$ is a non-singular matrix with

$(I_n - G_n)^{-1} = \sum_{j=0}^{\infty} G_n^j$, which implies that $\{Y_i\}$ is spatially stationary. In addition, we have $\max_{1\leq i,j\leq n} |g_{ij}| \leq \rho(G_n) < 1$ by Properties 4.66 and 4.67 in Seber (2008, p. 68). It is ready to show that $n^{-1}Y^TY = O_p(1)$, and $E(Y_i^2|X_1 = x_1, \ldots, X_n = x_n, D_1 = d_1, \ldots, D_n = d_n)$ is continuously differentiable and uniformly bounded under Assumptions A1 and A2. In addition, Assumption A2 (ii) is a regularity condition (see, e.g., Assumption 1 in Su (2012)), and it holds if the spatial weight function, $g(z)$, decreases to zero for large z and:

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n}\sum_{j=1}^{n} I(Z_{ij} \in \mathcal{Z}) < \infty \text{ for any fixed bound set } \mathcal{Z}, \tag{3.3}$$

where the indicator function $I(\mathcal{A}) = 1$ if event $\mathcal{A}$ holds, and zero otherwise.

Note that a parametric SDM is given by $Y_i = \rho\sum_{j\neq i} w_{ij}Y_j + \delta\sum_{t=1}^{p}\sum_{j\neq i} p_{t,ij}X_{tj} + X_i^T\theta_0 + u_i$, $i = 1, ..., n$, where W_n and $P_{n,t}$ are spatial weight matrices with their respective $(i,j)^{th}$ element equal to w_{ij} and $p_{t,ij}$. Therefore, if the parametric spatial Durbin model holds true, the spatial weight matrices G_n and $M_{n,t}$ in Model (3.1) are equivalent to ρW_n and $\delta P_{n,t}$ in the spatial Durbin model, respectively. From an estimation and econometric modelling viewpoint, the normalization of spatial weight matrices in the spatial Durbin model is used to identify the spatial multiplier parameters (ρ, δ), but this is not necessary in our proposed Model (3.1). Therefore, allowing nonparametric spatial weights saves us from applying an *ad hoc* spatial weight matrix normalization procedure as in the parametric SDM.

3.3 Estimation Methodology

If the dependent variable exhibits spatial autocorrelation, it must be accounted for by incorporating the spatially-lagged dependent variable into the model. If this variable is not included in the model, there would be an omitted variable type specification error due to the fact that unobserved factors may have a direct effect on the response variable. Moreover, the presence of the spatial lag-dependent variable in the model results in a simultaneity bias problem. This can be seen explicitly from the reduced form Model (3.2). From Model (3.2), we see $E(G_nYU^T) = G_n(I_n - G_n)^{-1}E\left(UU^T\right)$ is a non-zero matrix, so that the spatial lag of the dependent variable, G_nY, is correlated with the error term, U, which, therefore, results in an endogeneity problem in Model (3.1). Therefore, the ordinary least squares (or OLS) estimator would be biased and inconsistent. Moreover, the term $(I_n - G_n)^{-1}$ in Equation (3.2) explains that region i is affected not only by its own determinants, but also by its neighbouring regions' values. This has been called a global interaction effect in Ertur and Koch (2007) and LeSage (2014). Another source of the spatial endogeneity problem is due to the endogeneity of spatial covariate in the model. Recently, Kelejian and Piras (2014) and Sun (2016) estimated the spatial panel data model and the SAR model with an endogenous spatial weight matrix in a nonparametric way, respectively. Moreover, Qu and Lee (2015) proposed estimators for the parametric SAR model with an endogenous spatial covariate. This paper only deals with the endogeneity of the spatial lag of the dependent variable, as our spatial weight matrix is non-stochastic.

The endogeneity problem can be addressed by using the maximum likelihood estimation (or MLE) method, as well as the instrumental variable (or IV) approach. Ord (1975) was the first to examine the MLE of SAR models. He proposed to use the eigenvalues of the spatial weights matrix to alleviate the computational complexity of the MLE method in large sample sizes. Lee (2004) derived the large sample properties of the quasi-MLE without a normality assumption on error terms, while Bao and Ullah (2007) obtained the second order bias of the maximum likelihood estimator for spatial autoregressive models. As the (quasi-) maximum likelihood estimator can be computationally difficult in moderate or large-sized samples, Kelejian and Prucha (1998) proposed a two-stage least squares (or 2SLS) estimator for a SAR model with spatial autoregressive errors, while Lee (2003) proposed an asymptotically-optimal 2SLS estimator. As the spatial weights in Model (3.1) are unknown, the 2SLS estimation methods derived in Kelejian and Prucha (1998) and Lee (2003) are not feasible; we therefore use a series approximation method to recover the unknown spatial weight function and estimate all unknown functions via a nonparametric 2SLS (or NP2SLS) estimation method. For an overview of the sieve estimation method, see Chen (2007).

Specifically, we approximate the unknown weighting functions $g(\cdot)$ and $m_t(\cdot)$, $t = 1, ..., p$, and the vector of functional coefficients $\theta(\cdot)$ by series expansions:

$$g^*(z) = \sum_{l=1}^{L_n} \alpha_l \phi_l(z), \tag{3.4}$$

$$m^*(z) = [m_1^*(z), ..., m_p^*(z)]^T = [\sum_{l=1}^{L_n} \gamma_{1l}\phi_l(z), \sum_{l=1}^{L_n} \gamma_{2l}\phi_l(z), ..., \sum_{l=1}^{L_n} \gamma_{pl}\phi_l(z)]^T, \quad (3.5)$$

and:

$$\theta^*(d) = [\theta_1^*(d), ..., \theta_p^*(d)]^T = [\sum_{l=1}^{L_n} \beta_{1l}\phi_l(d), \sum_{l=1}^{L_n} \beta_{2l}\phi_l(d), ..., \sum_{l=1}^{L_n} \beta_{pl}\phi_l(d)]^T, \quad (3.6)$$

respectively, where $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{L_n})^T$, $\gamma_t = (\gamma_{t1}, \gamma_{t2}, ..., \gamma_{tL_n})^T$ and $\beta_t = (\beta_{t1}, \beta_{t2}, ..., \beta_{tL_n})^T$ for $t = 1, ..., p$ are all $L_n \times 1$ vectors of unknown coefficients, $\{\phi_j(\cdot)\}_{j=1}^{L_n}$ is a sequence of square integrable orthonormal basis functions over the interval $[0, \infty)$ and L_n denotes the number of basis functions. The following assumption regulates the sparseness of the weight matrix and the smoothness of unknown functions.

Assumption A3: (i) There exists a positive constant sequence $\{v_n\}$, such that:

$$\max_{1\le l\le L_n, 1\le i\le n} \sum_{j\ne i}^{n} |\phi_l(Z_{ij})| \le Cv_n \text{ and } \max_{1\le l\le L_n, 1\le j\le n} \sum_{i=1}^{n} |\phi_l(Z_{ij})| \le Cv_n. \quad (3.7)$$

(ii) There exist $L_n \times 1$ vectors α, γ_t, and β_t for $t \in \{1, ..., p\}$, such that:

$$\max_{1\le i\le n} \sum_{j\ne i} |g(Z_{ij}) - \alpha^T \Phi_{L_n}(Z_{ij})| = O(L_n^{-\zeta}) \quad (3.8)$$

$$\max_{1\leq t\leq p}\max_{1\leq i\leq n}\sum_{j\neq i}|m_t(Z_{ij})-\gamma_t^T\Phi_{L_n}(Z_{ij})|=O(L_n^{-\zeta}) \tag{3.9}$$

and:

$$\sup_{d\in R}\max_{1\leq t\leq p}|\theta_t(d)-\beta_t^T\Phi_{L_n}(d)|=O(L_n^{-\zeta}),$$

respectively, for some $\zeta > 2$ as $L_n \to \infty$, where $\Phi_{L_n}(z) = [\phi_1(z), \phi_2(z), ..., \phi_{L_n}(z)]^T$ is an $L_n \times 1$ vector of basis functions.

It is not necessary to know the exact order of υ_n in Assumption A3 (i), and the consistency of our proposed estimator does not require $\upsilon_n \equiv 1$, as assumed in Pinkse et al. (2002, Assumption (vi)). From approximation theory in mathematics, Assumption A1 (ii) is a necessary condition for Assumption A3 (ii). However, (3.8) and (3.9) also require spatial units expanding sparsely as more spatial units are included, for example when (3.3) holds true; and the consistency of our proposed estimator relies on increasing domain asymptotic theory. Moreover, we use Laguerre polynomial series to approximate the unknown functions, as it is one of the common choices for series expansions when a function has a domain over $[0, \infty)$ (Chen, 2007, p. 5574). In addition, L_n acts as a smoothing parameter that increases slowly with the sample size. In other words, it is required to have $L_n \to \infty$ and $L_n/n \to 0$ as $n \to \infty$. An introduction of series estimation methods in a nonparametric framework can be found in Chapter 15 of Li and Racine (2007).

Now, we approximate Model (3.1) by:

$$Y_i \approx \sum_{l=1}^{L_n} \alpha_l \sum_{j \neq i} \phi_l(Z_{ij}) Y_j + \sum_{t=1}^{p} \sum_{l=1}^{L_n} \gamma_{tl} \sum_{j \neq i} \phi_l(Z_{ij}) X_{tj} + \sum_{t=1}^{p} \sum_{l=1}^{L_n} \beta_{tl} \phi_l(D_i) X_{ti} + u_i, i = 1, ..., n. \tag{3.10}$$

To derive our first-step estimator, we rewrite Model (3.10) in matrix form as follows:

$$Y \approx V_n \xi + U, \tag{3.11}$$

where we denote the i^{th} row vector of an $n \times [(2p+1)L_n]$ matrix V_n by:

$$V_{n,i}^T = [(\sum_{j \neq i} \Phi_{L_n}(Z_{ij}) Y_j)^T, (\sum_{j \neq i} \Phi_{L_n}(Z_{ij}) X_{1j})^T, ..., (\sum_{j \neq i} \Phi_{L_n}(Z_{ij}) X_{pj})^T,$$
$$(\Phi_{L_n}(D_i) X_{1i})^T, ..., (\Phi_{L_n}(D_i) X_{pi})^T]$$

and a $[(2p+1)L_n] \times 1$ vector of parameters $\xi = [\alpha^T, \gamma_1^T, \gamma_2^T, ..., \gamma_p^T, \beta_1^T, \beta_2^T, ..., \beta_p^T]^T$.

The specification of the instrumental variable matrix is of great importance to obtain a consistent estimator. Since the number of endogenous variables increases with the number of approximating functions, L_n, it is intuitively appealing to instrument the endogenous variables, $\sum_{j \neq i} \phi_l(Z_{ij}) Y_j$, $l = 1, ..., L_n$, by $(\sum_{j \neq i} \Phi_{L_n}(Z_{ij}) X_{t_1 j}) X_{t_2 i}$ and $\sum_{j \neq i} \Phi_{L_n}(Z_{ij}) D_j$, $t_1, t_2 \in \{1, 2, ..., p\}$ for $p > 1$ as in our empirical application; see, e.g., Anselin and Bera (1998). Since $X_{t_1 j} X_{t_2 i}$ and D_j are exogenous and relevant in predicting Y_j, we would expect the proposed instrumental variables to serve as valid instruments for $\sum_{j \neq i} \Phi_{L_n}(Z_{ij}) Y_j$.

Therefore, we define the i^{th} row vector of an $n \times [(2p+2)L_n]$ instrumental matrix Q_n as:

$$Q_{n,i}^T = [((\sum_{j \neq i} \Phi_{L_n}(Z_{ij})X_{1j})X_{1i})^T, (\sum_{j \neq i} \Phi_{L_n}(Z_{ij})X_{1j})^T, ..., (\sum_{j \neq i} \Phi_{L_n}(Z_{ij})X_{pj})^T,$$
$$(\Phi_{L_n}(D_i)X_{1i})^T, ..., (\Phi_{L_n}(D_i)X_{pi})^T, (\sum_{j \neq i} \Phi_{L_n}(Z_{ij})D_j)^T].$$

We then can estimate ξ from (3.11) by the 2SLS estimation method. Note that we do not pursue optimal instrument variables in this paper due to the complexity of this approach in the semiparametric setup and the fact that the oracle efficiency of the second-step estimator of $\theta(\cdot)$ does not rely on the use of optimal instruments in the first-step estimation.

To ensure the existence of our 2SLS estimator, we assume that the exogenous regressors matrix X_n, the instrumental variables matrix Q_n and $V_n^T Q_n (Q_n^T Q_n)^{-1} Q_n^T V_n$ all have full column rank. Moreover, for the relevance of the instruments, we assume that $E[Q_n^T V_n]$ has a full column rank. Otherwise, we can remove linearly-dependent terms as long as the number of instruments in Q_n is more than the number of endogenous variables L_n plus the number of exogenous regressors $2p$. Lee (2007a) argues that the 2SLS estimator would be inconsistent if (X_i, D_i) are both irrelevant in predicting $\{Y_i\}$. Therefore, throughout this paper, we assume that X and D contain relevant variables in predicting $\{Y_i\}$ and $\theta(\cdot)$ takes non-zero values over any non-empty interval, so that there is no need to use the quadratic moments as additional orthogonal relations, as suggested in Lee (2007a). Our empirical application in this paper satisfies this assumption by both economic theory and empirical findings observed from the economic growth literature.

To construct a consistent estimator for $g(\cdot)$, $m_t(\cdot)$ and $\theta_t(\cdot)$, $t = 1, ..., p$, we consider the following nonparametric 2SLS objective function:

$$\min_{\xi} \left[Q_n^T(Y - V_n\xi)\right]^T \left[Q_n^T(Y - V_n\xi)\right]. \tag{3.12}$$

The nonparametric 2SLS estimator of ξ solves (3.12) and is given by:

$$\hat{\xi} = [V_n^T Q_n (Q_n^T Q_n)^{-1} Q_n^T V_n]^{-1} V_n^T Q_n (Q_n^T Q_n)^{-1} Q_n^T Y,$$

and hence, the corresponding nonparametric 2SLS estimators [1] of unknown functions are given by:

$$\hat{g}(z) = \sum_{l=1}^{L_n} \hat{\alpha}_l \phi_l(z), \tag{3.13}$$

$$\hat{m}(z) = [\hat{m}_1(z), ..., \hat{m}_p(z)]^T = [\sum_{l=1}^{L_n} \hat{\gamma}_{1l}\phi_l(z), \sum_{l=1}^{L_n} \hat{\gamma}_{2l}\phi_l(z), ..., \sum_{l=1}^{L_n} \hat{\gamma}_{pl}\phi_l(z)]^T, \tag{3.14}$$

and

$$\hat{\theta}(d) = [\hat{\theta}_1(d), ..., \hat{\theta}_p(d)]^T = [\sum_{l=1}^{L_n} \hat{\beta}_{1l}\phi_l(d), \sum_{l=1}^{L_n} \hat{\beta}_{2l}\phi_l(d), ..., \sum_{l=1}^{L_n} \hat{\beta}_{pl}\phi_l(d)]^T. \tag{3.15}$$

Next, we propose a second-step estimator for the functional coefficients, $\theta(d)$, using

[1] The asymptotic properties of the series estimator without an endogenous variable follows from Andrews (1991) and with some extensions from Newey (1997). Moreover, we expect the asymptotic properties of our proposed estimators closely follow Sun (2016).

the local linear regression approach. We would expect the local linear estimate of $\theta(d)$, $\tilde{\theta}(d)$, to have an improvement over the first-step estimator, $\hat{\theta}(d)$. Sun (2016) considered a semiparametric spatial autoregressive model that has a mathematical representation of Model (3.1) with $M_{n,t} = 0$ for all $t = 1, \ldots, p$ and has recently shown that the local linear estimator of $\theta(\cdot)$ can be oracle efficient under some regularity conditions in the sense that its limiting distribution does not depend on whether or not the spatial weights are known. This is a general result from non-/semi-parametric additive models. As the unknown functions $(g(Z_{ij}), m_1(Z_{ij}), \ldots, m_p(Z_{ij}))$ and $\theta(D_i)$ enter Model (4.1) additively, we expect that $\tilde{\theta}(d)$ is oracle efficient, as well. We assume that as $n \to \infty$, $h \to 0$, $nh \to \infty$ and $nh^5 \to c \in (0, \infty)$, where h is the bandwidth, which controls the size of the local neighbourhood around an interior point d. Moreover, let $K(\cdot)$ be a kernel function, which assigns more weights to the data closer to point d, satisfying: (i) $\int K(a)da = 1$; (ii) $K(a) = K(-a)$; and (iii) $\int a^2 K(a)da > 0$.

The estimation procedure for $\tilde{\theta}(d)$ is given as follows:

(i) We replace $g(z)$ and $m_t(z)$ in (3.1) by $\hat{g}(z)$ and $\hat{m}_t(z)$, respectively, and treat $\hat{Y}_i = Y_i - \sum_{j \neq i} \hat{g}(Z_{ij}) Y_j - \sum_{t=1}^{p} \sum_{j \neq i} \hat{m}_t(Z_{ij}) X_{tj}$ as the dependent variable.

(ii) Applying the first-order Taylor series expansion of $\theta(D)$ around d, $\theta(D) \approx \theta(d) + \theta'(d)(D - d)$, we calculate the local linear estimator from a minimization of a kernel-weighted objective function:

$$\left(\tilde{\theta}(d), \tilde{\theta}'(d)\right) = \arg \min_{\theta(d), \theta'(d)} \sum_{i=1}^{n} [\hat{Y}_i - X_i^T \theta(d) - X_i^T \theta'(d)(D_i - d)]^2 K((D_i - d)/h) \quad (3.16)$$

where $\tilde{\theta}(d)$ estimates $\theta(d)$ and $\tilde{\theta}'(d)$ estimates $\theta'(d)$, the first order derivative of $\theta\,(d)$.

For a complete treatment of local linear estimator, see Fan and Gijbels (1996). As the mathematical proofs of the consistency of the first-step estimator of $\left(\hat{g}(z), \hat{m}_1(z), \ldots, \hat{m}_p(z), \hat{\theta}(d)^T\right)^T$ and the limiting result of the second-step estimator $\tilde{\theta}(d)$ closely follow those given in Sun (2016), the proofs are omitted from the paper.

3.4 Monte Carlo Simulations

In this section, we present the results from a very small Monte Carlo simulation study to assess the finite-sample properties of our estimators and more simulation results can be obtained from the authors upon request. We generate the data from the following regression model:

$$Y_i = \sum_{j \neq i} g(Z_{ij}) Y_j + \sum_{j \neq i} m(Z_{ij}) X_j + X_i \exp(-\left(4D_i - 1\right)^2) + u_i, i = 1, 2, ..., n, \quad (3.17)$$

where we randomly draw $u_i \sim$ i.i.d.$N(0, 0.5)$, $D_i \sim$ i.i.d.$U[0,1]$ and $X_i = 0.5D_i + \eta_i$ with $\eta_i \sim$ i.i.d.$N(0,1)$ independent of $\{u_i\}$. For the exogenous variable, Z, we first randomly generate n observations from the $U[0, R_n]$ distribution with $R_n = 0.001n^{1.6}$, by which we control the sparseness of spatial units. Then, we calculate Z_{ij} as the absolute distance between observations i and j. The specification of spatial weight functions requires that $g(\cdot)$ and $m(\cdot)$ are both decreasing and non-negative functions. We therefore set $g(z) = m(z) = 0.01exp(-z/0.01)$ for $z > 0$ with $g(0) = 0$ and $m(0) = 0$. The random variables,

u_i, D_i and Z_i, are all mutually independent.

We consider a sample size $n \in \{100, 200, 400\}$. The number of replications is 1000 for each n in the Monte Carlo experiments. Moreover, we set $L_n = 1, 2, 3$ for each sample size, respectively. In the second-step estimation of coefficient functions, we select the bandwidth via a cross-validation method and use the Gaussian kernel function. To measure the performance of the estimators, we compute the root mean squared errors (or RMSEs) for each simulation. In Table 3.1, we report the averages of the RMSEs computed over 1000 repetitions, where $\hat{g}(\cdot)$, $\hat{m}(\cdot)$ and $\hat{\theta}(\cdot)$ denote the NP2SLS estimators of $g(\cdot)$, $m(\cdot)$ and $\theta(\cdot)$, respectively; $\tilde{\theta}(\cdot)$ is the second-step estimator of $\theta(\cdot)$, and $\vec{\theta}(\cdot)$ is the local linear estimator of $\theta(\cdot)$, while $g(\cdot)$ and $m(\cdot)$ are known. Furthermore, we also estimate the average direct impact (ADI) and the average indirect impact (AII) and report their corresponding RMSEs in Table 3.1. Specifically, we first obtain the reduced form model from (3.17):

$$Y = (I_n - G_n)^{-1}[M_n X + \theta(D) \circ X + U], \tag{3.18}$$

where Y, X, $\theta(D)$ and U are all $n \times 1$ vectors, and "$\odot$" denotes the Hadamard multiplication. Then, the expected marginal effect of X is given by the following $n \times n$ matrix:

$$\frac{\partial E(Y|X, D)}{\partial X} \equiv (I_n - G_n)^{-1}[M_n + diag\{\theta(D)\}] = S(G_n, M_n, D)$$

from which we obtain $ADI = n^{-1} tr\{S(G_n, M_n, D)\}$ and $AII = n^{-1}\mathbf{i}_n' S(G_n, M_n, D)\mathbf{i}_n - ADI$; see LeSage and Pace (2009), where $\mathbf{i}_n$ is the $n \times 1$ vector of ones and $diag\{\theta(D)\}$ is a diagonal matrix. Replacing the two unknown spatial weight matrices and $\theta(D)$ by their estimates, we obtain the estimates for ADI and AII .

Table 3.1: Average RMSEs.

n	$\hat{g}(\cdot)$	$\hat{m}(\cdot)$	$\hat{\theta}(\cdot)$	$\tilde{\theta}(\cdot)$	$\vec{\theta}(\cdot)$	ADI	AII
100	0.2109	0.0734	0.3535	0.2159	0.1557	0.1061	0.4386
200	0.1901	0.0593	0.2579	0.1680	0.1168	0.0761	0.2913
400	0.0500	0.0160	0.2013	0.1001	0.0859	0.0432	0.2989

From Table 3.1, we observe that there is a decrease in the RMSEs for all three estimators as the sample size increases in each design. Moreover, the second-step estimator always performs better than the nonparametric 2SLS estimator. The relative ratios of the RMSEs of the second-step estimator and $\vec{\theta}(\cdot)$ generally reduce as the sample size increases. Therefore, our simulation results support the consistency of our proposed estimators. As for the ADI and AII, we also see an overall decreasing pattern in the RMSEs as the sample size increases, where the AII is less accurately estimated than the ADI, as the former is calculated from $n(n-1)$ terms and the latter is calculated from n elements only.

3.5 Empirical Application

Monte Carlo simulations results given in Section 4 support the consistency of our proposed estimation method. We are now in a position to re-investigate cross-country growth patterns. We want to evaluate the impact of a country's initial income, savings rate, population growth rate and openness, as well as neighbour countries' economic growth spillovers on a country's economic growth rate. We follow Ertur and Koch (2007) in using a sample of 91 countries listed in Mankiw et al. (1992), which is the Heston-Summers data taken from Penn World Table 6.1. Consider the following conditional convergence Solow growth model [2]:

$$\begin{aligned} Y_i &= \sum_{j\neq i} g(Z_{ij})Y_j + \sum_{j\neq i} m(Z_{ij})\ln(y60_j) + \theta_1(lopen_i) + \theta_2(lopen_i)\ln(y60_i) \\ &\quad +\theta_3(lopen_i)\ln(s_i) + \theta_4(lopen_i)\ln(n_i + 0.05) + u_i, \end{aligned} \tag{3.19}$$

for $i = 1, 2, ..., 91$, where Y_i is the i^{th} country's average growth rate of real GDP per capita between 1960 and 1995, $y60_i$ is the i^{th} country's initial real GDP per capita in 1960, s_i is the i^{th} country's average saving rate, n_i is the i^{th} country's average growth rate of working-age population (ages between 15 and 64), $lopen_i$ is a scalar development index of a country defined as the logarithm of the i^{th} country's average ratio of total imports plus exports over its real GDP over the period from 1960 to 1995, and Z_{ij} is the great-circle distance between i^{th} and j^{th} countries' capitals [3].

We approximate $g(\cdot)$, $m(\cdot)$ and $\theta_i(\cdot)$, $i = 1, 2, 3, 4$, using the Laguerre polynomials with $L_n = 2$. Moreover, a cross-validation selected bandwidth, h_{opt}, is calculated as 0.5285. We obtained $\rho(\hat{G}_n) = 0.041$, which suggests a spatial stationarity in the data. The

[2] Since the sample size is less than 100, we include only one spatially-lagged exogenous variable, $M_n ln(y60)$, to have better finite sample estimation accuracy. Moreover, the reason behind this choice is that spatial lag effects from the savings rate and the population growth rate were not found significant in Ertur and Koch (2007, Table IV on p. 1051).

[3] We follow Ertur and Koch (2007) in calculating the variable Z_{ij}:

$$Z_{ij} = radius * \arccos[cos\,(|long_i - long_j|)\cos{(lat_i)}\cos{(lat_j)} + \sin{(lat_i)}\sin{(lat_j)}],$$

where $radius$ is taken as the Earth's radius and lat_i and $long_i$ are the latitude and longitude for country i, respectively.

distribution of the estimated residuals is approximately normal as the q-q plot of estimated residuals is close to linear. The coefficient estimates, $\tilde{\theta}_1(\cdot)$, $\tilde{\theta}_2(\cdot)$, $\tilde{\theta}_3(\cdot)$ and $\tilde{\theta}_4(\cdot)$ are presented in Figure 3.1, where the solid lines with circles display the second-step estimates. The dashed lines represent the estimates of the spatially-augmented Solow growth model of Ertur and Koch (2007) using the inverse power spatial weight function, which we include as a baseline. We interpret Figure 3.1 as follows. First, in Figure 3.1b, we see that there is a negative relation between the initial level of income and the economic growth rate, except for Mauritius, Hong Kong, Zambia, Cameroon and Singapore, which confirms a conditional β-convergence hypothesis. Moreover, we observe that $\tilde{\theta}_2(\cdot)$ is increasing in openness, which, however, results in a gradually declining degree of convergence. In addition, we see that the nonparametric model reveals slightly weaker conditional economic growth convergence as compared to the parametric model.

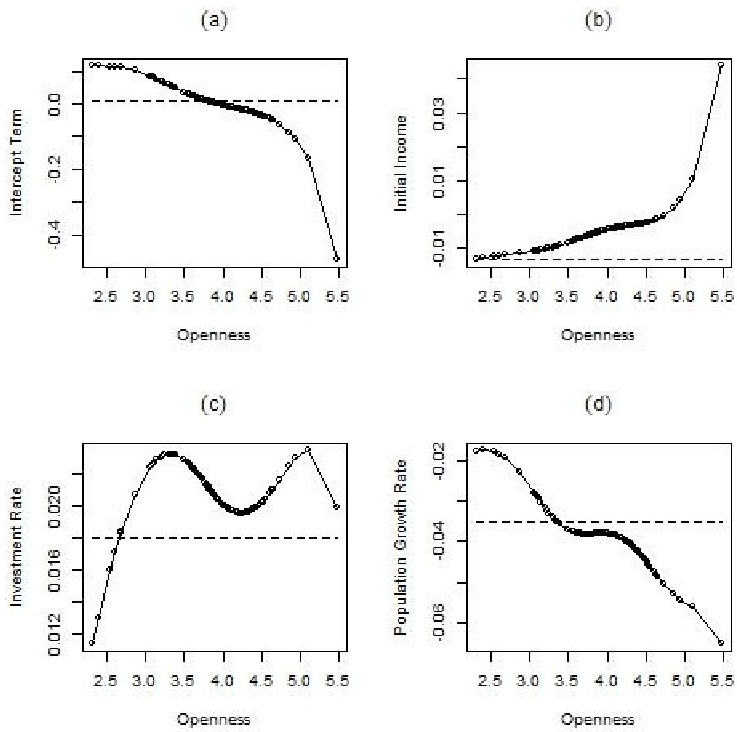


Figure 3.1: Estimated coefficient curves.

Second, in Figure 3.1c, we see that $\tilde{\theta}_3(\cdot)$ exhibits a positive, but not a monotonic, relation between the real investment rate and the real GDP per capita growth rate. Our estimate of the coefficient of the investment rate fluctuates as the trade openness of countries increases. For the economies with a trade openness higher than 15% of GDP, our result indicates that the nonparametric model sees stronger positive impact of the investment rate on the real GDP per capita growth rate than the parametric model does. Third, in Figure 3.1d, it is observed that the population growth rate has a negative impact on the real GDP per capita growth rate. For the countries whose trade openness ranges between 29% and 65%, our estimates for the coefficient of the population growth rate are relatively flat.

Moreover, we note that the magnitude of the negative effect of the population growth rate is getting larger as the trade openness of countries increases, especially when the trade openness is over 65% of GDP. Overall, Figure 3.1 can be interpreted as the fact that an open economy suffers from higher negative impact of the population growth rate, but at the same time takes the advantage of high initial real GDP per capita.

Next, due to the cross-country interactions through spatial weights, the functional coefficient estimates have a different interpretation than the one obtained from the non-spatial model. In order to correctly interpret these estimates, we rewrite the estimated model in a reduced from as follows:

$$\begin{aligned} Y &= (I_n - \hat{G}_n)^{-1}[\hat{M}_n \ln(y60) + \tilde{\theta}_1(lopen) + \tilde{\theta}_2(lopen) \odot \ln(y60) \\ &\quad + \tilde{\theta}_3(lopen) \odot \ln(s) + \tilde{\theta}_4(lopen) \odot \ln(n + 0.05) + \tilde{U}], \end{aligned} \tag{3.20}$$

where $\tilde{U}$ is the $n \times 1$ vector of residuals. Then, from Equation (3.20), the marginal effects are given by the following $n \times n$ matrices:

$$\frac{\partial E(Y|y60, s, n)}{\partial \ln(y60^T)} \equiv (I_n - G_n)^{-1}[M_n + diag\{\theta_2(lopen)\}]$$

$$\frac{\partial E(Y|y60, s, n)}{\partial \ln(s^T)} \equiv (I_n - G_n)^{-1} diag\{\theta_3(lopen)\}$$

$$\frac{\partial E(Y|y60, s, n)}{\partial \ln(n^T + 0.05)} \equiv (I_n - G_n)^{-1} diag\{\theta_4(lopen)\}$$

where we define $diag\{a\}$ as an $n \times n$ diagonal matrix with the elements of an $n \times 1$ vector, a, on the main diagonal. Following LeSage and Pace (2009), we label the diagonal elements of each matrix given above as the direct impacts and off-diagonal elements as the indirect impacts.

In Table 3.2, we report the estimated average direct impact (ADI) and average indirect impact (AII) of the explanatory variables, where the latter can be easily defined as the difference between average total impact [4] and the average direct impact. Average direct and indirect impacts from the parametric model of Ertur and Koch (2007) are denoted as ADI_{EK} and AII_{EK}, respectively. The interpretation of Table 3.2 is as follows. Firstly, we observe that a 1% increase in the real initial GDP per capita of an economy, holding other factors fixed, results in a decrease by 0.5% in its own real GDP per capita growth rate. However, this change increases the rest of the economies' economic growth rates by 0.01% on average due to the spatial dependence. From another point of view, a 1% increase in all of the regions'/nations' initial real GDP per capita speeds up this economy's real GDP per capita growth by 0.01%. This result indicates a positive spillover effect of the initial level of income. Secondly, a 1% increase in this economy's real investment rate increases its own real GDP growth rate by 2.08% on average. However, this change slows down the rest of the nations' real GDP growth by 0.08% on average. Thirdly, we see that a 1% increase in the population growth rate of this economy retards its own economic growth by 3.8% on average, but helps to improve the rate of economic growth of the rest of the countries by

[4] As it is stated by LeSage and Pace (2009), average total impact can be expressed in two different ways, however, which give the same numerical results. The first viewpoint states an influence from a change in the initial real GDP per capita of an economy on all of the regions, while the second viewpoint states an impact of changes in the initial real GDP per capita of the entire economy on a region/nation.

0.16% on average.

Table 3.2: Average direct and indirect impact estimates.

	ln(y60)	**ln(s)**	**ln(n + 0.05)**
ADI	−0.0050 (−0.0122, −0.0005)	0.0208 (0.0089, 0.0341)	−0.0381 (−0.0782, −0.0054)
ADI_{EK}	−0.0119 (−0.0159, −0.0078)	0.0184 (0.0139,0.0229)	−0.0336 (−0.0585, −0.0094)
AII	0.0001 (−0.0077, 0.0056)	−0.0008 (−0.0174, 0.0190)	0.0016 (−0.0347, 0.0289)
AII_{EK}	0.0140 (0.0052, 0.0244)	−0.0018 (−0.0169, 0.0124)	0.0275 (−0.0321, 0.0860)

Note: A 95% bootstrap percentile confidence interval is given in the parenthesis.

Fourthly, when comparing our results to the results from Ertur and Koch (2007), we find that both the nonparametric and the parametric model give almost the same average direct effects of the initial per capita income, investment rate and population growth rate on the economic growth rate on average and that both models result in the same signs in the average direct and indirect effects. However, the AII values from the nonparametric model are much smaller than the results from the parametric model in absolute value, especially for the initial per capita income and the population growth rate. This is not surprising, as the parametric model assumes that all of the spatial weights take non-negative values, while our nonparametric spatial weights are estimated from the data without such a restriction. Although it is popular practice to assume non-negative spatial weights, this is an assumption imposed without support from econometric or economic theory. For example, trade treaties and monetary policies are both double-edged swords that may bring opposite impacts to different national economies, and non-negative spatial weights may not be able to capture the opposite interactions among different economies. As both parametric SDM of Ertur and Koch (2007) and our proposed semiparametric SDM approximate the unknown true relationships in their own best capacity, however, our model imposes less restrictions than the parametric SDM and is believed to bring a better fit to the data and more reliable inference. Although the numbers are different, both models give the same sign in estimated direct and indirect effects. Overall, we observe that the average direct and indirect effects can take opposite signs, and the effect of the former is much stronger than that of the latter in absolute magnitude.

LeSage and Pace (2009, p.39) explain how to obtain the standard errors for the ADI and AII estimates via a simulation method. In the parametric setup, as the spatial weight matrices are known, theoretically, one can apply the delta method to obtain the standard errors, and the simulation method tends to provide at best an approximation as one does not know the exact distribution of the estimated coefficients in finite samples; however, this method is feasible as the average direct and indirect impacts only depend on a finite number of unknown parameters. As our proposed semiparametric model contains both unknown spatial weights and unknown coefficient curves, the simulation method would involve simulating from a joint distribution with dimension equal to $2n\,(n-1)$ (two spatial weight matrix estimates) plus $4n$ (four coefficient curve estimates) or 16,744 in our empirical application. Therefore, the simulation method is infeasible for our empirical interest here. As both the ADI and AII are in the form of sample averages and it is well known that the bootstrap method can be used to estimate the sample average and its standard error well (e.g., Efron and Tibshirani, 1993), we decide to report our bootstrap estimates of the confidence intervals for the ADI and AII. It should be noted that although both ADI and AII are in the form of sample averages, they depend on nonparametric estimators. Therefore, we follow the bootstrapping method proposed for nonparametric models summarized in Section 7 of Chapter 2; see Härdle and Mammen (1991) for details. Another point to be mentioned is that the methods of bootstrapping summarized in Chapter 2 are for fixed design regressors. Since our model in this paper includes stochastic regressors, namely the spatially lagged dependent variable, bootstrapping for the model with random regressors could be used to obtain the confidence intervals.

Below, we explain a residual-based bootstrap method to test whether the ADI and AII are significantly different from zero at the 5% significance level. We use the nonparametric bootstrap percentile method to construct a 95% confidence interval. Following Härdle and Marron (1991) and Sun (2006), we first estimate the functional coefficients using an over-smoothed bandwidth, which tends to zero at a slower speed than the optimal bandwidth. Then, we obtain estimated residuals. The rest of the bootstrap procedure is given below.

1. Resample the estimated residuals and obtain the bootstrap errors, U^b.

2. Calculate:

$$Y^b = (I_n - \hat{G}_n)^{-1}[\hat{M}_n \ln(y60) + \tilde{\theta}_1^*(lopen) + \tilde{\theta}_2^*(lopen) \odot \ln(y60) + \tilde{\theta}_3^*(lopen) \odot \ln(s) + \tilde{\theta}_4^*(lopen) \odot \ln(n + 0.05) + U^b],$$

where $\tilde{\theta}_k^*(\cdot)$, $k = 1, 2, 3, 4$ are coefficient estimates using a larger bandwidth than the optimal bandwidth. Call ($\left\{\ln(y60_i), \ln(s_i), \ln(n_i + 0.05), Y^b\right\}_{i=1}^n$ the bootstrap sample.

3. Estimate Model (3.19) from the bootstrap sample, and record $\tilde{\theta}_1^b(\cdot)$, $\tilde{\theta}_2^b(\cdot)$, $\tilde{\theta}_3^b(\cdot)$ and $\tilde{\theta}_4^b(\cdot)$ the bootstrap estimates of the functional coefficients.

4. Calculate the bootstrap value of average direct and indirect impacts of the explanatory variables, ADI^b and AII^b, respectively.

5. Repeat Steps 1 to 4 999 times.

6. Find the 0.025th and 0.975th empirical percentile of the 999 bootstrap values of ADI and AII and the point estimates given in Table 2 to establish the 95% bootstrap percentile confidence interval.

The confidence intervals are reported in Table 3.2. We see that the ADI values are statistically significantly different from zero at the 5% significance level. Moreover, we find that there is no significant effect on average from neighbouring countries per capita initial income, savings rate and population growth rate on economic growth rate of the country of interest. The same inference is obtained for the parametric model, except that the parametric model sees a significant average indirect impact of initial per capita income. Note that insignificant AIIs do not imply that the indirect impact from economy i on economy j is insignificant for all (i, j).

Figure 3.2 presents estimated spatial weighting functions. We plot both estimated spatial weighting functions, $\hat{g}(\cdot)$ and $\hat{m}(\cdot)$, for the geographic distances ranging from zero to 20 in 100 km, as the estimated spatial weights in the absolute value have an average of 5.069 $\times$ 10^{-7} and 3.893 $\times$ 10^{-9}, respectively, when $z > 20$. In Figure 3.2a,b, firstly, we see negative spatial weights, which greatly contradicts traditional parametric spatial regression models, which often assume non-negative spatial weights. Negative spatial interactions are indeed common in practice, especially in social networks; see Bhattacharjee and Holly (2011, 2013) for strategic interactions within the monetary policy committee of the Bank of England. Secondly, both spatial weight functions are not strictly monotonic and exhibit convexity among nations that are not very far apart, concavity among nations with moderately far distances and a zero spatial weight function among nations that are far away. Moreover, the spatial weight functions take bigger absolute values among nations with smaller distance apart and smaller absolute values among far-away nations, which implies a relatively larger economic interaction among nearby nations than among far-away nations. Lastly, in Figure 3.2b, we observe positive estimated spatial weights when the distance ranges between 0.229 and 1.894, which correspond to 22.9 and 189.4 km, respectively. For the distances greater than 189.4 km, negative spatial weights are getting closer to zero as the distance between two countries increases. As the turning point 1.894 is really small and occurs for countries with a small area, our results imply that spatial interaction is very strong and different between two nearby small countries with small areas than between two countries with longer distances when at least one country has a large area.

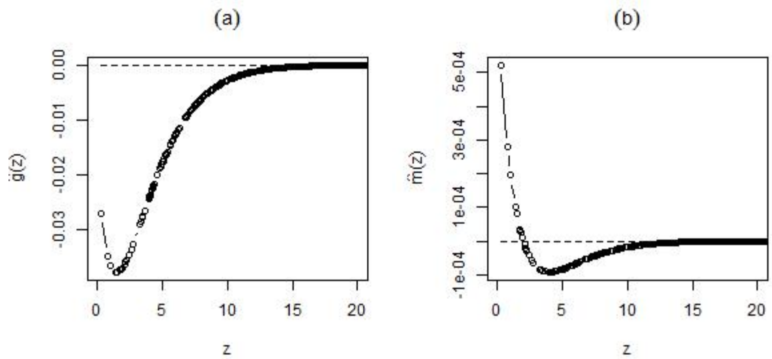


Figure 3.2: Estimated spatial weighting functions.

3.6 Conclusions

We employ a spatial Durbin model combined with the nonparametric spatial weighting functions, as well as the unknown functional coefficients to estimate the augmented Solow growth model with a sample of 91 countries over the period 1960 to 1995. We find a negative spatial lag effect of neighbouring country's GDP per capita growth rate and initial GDP per capita on the economic growth rate of country i. These effects are declining in magnitude as the geographical distance between the two countries increases. Finally, allowing coefficients as a function of trade openness of a country enables us to see the true country-specific effect of each determinant of economic growth. Moreover, we find significant average direct impact from each production factors. However, our findings show that the average indirect impact of these variables is insignificant at the 5% significance level.

As a future research, spatial lag of investment rate may explain another spillover effect across countries. As the initial income is predetermined by construction, among other variables, investment rate might also cause endogeneity problem in the spatial regression model. Both spatial effect and endogeneity of investment rate will be controlled for using nonparametric spatial weights.

Chapter 4

Growth and Debt: An Endogenous Smooth Coefficient Approach

4.1 Introduction

In the aftermath of the recent global financial crisis, government debt has increased substantially across the world. For advanced economies, public debt-to-GDP ratio has risen on average from about 66% in 2007 to 105% by the end of 2015. Particularly, Greece, Ireland, Japan, Portugal, Spain, and the United Kingdom, comparable to others, have experienced a rapid increase in public debt ratio between the years 2008 and 2012. A growing concern behind these facts is that countries may not achieve debt sustainability implying higher vulnerability to economic and financial crisis (Cecchetti et al., 2010). In fact, over the last two centuries there are twenty financial crisis followed by debt build-ups periods, which lasted more than a decade and are associated with lower growth than during other periods (Reinhart et al., 2012). Therefore, a relevant policy question is centered on the long-term growth effects of high public debt.

The relationship between public debt and economic growth is, to the best of our knowledge, still unresolved in both theoretical and empirical literature. Theoretically, the conventional view of public debt is that fiscal deficits in the short-run can have a positive effect on economic growth through stimulating aggregate demand and output, whereas having a potential crowding out effect on private investment in the long run (Elmendorf and Mankiw, 1999). On the other side, a large number of economic growth research papers find some evidence of nonlinearity in the effect of public debt on growth, particularly focusing on threshold levels. The idea is to detect a debt level beyond which economic growth is adversely affected implying a concave (inverted-U shape) relationship between debt and growth. Using a basic nonparametric technique, i.e., a histogram, to investigate correlation between public debt and growth, Reinhart and Rogoff (2010) find a threshold level of 90% for the 20 advanced countries over the period 1945-2009. Their findings are striking in the sense that real mean GDP growth decreases substantially (at about 4%) when public debt is beyond the 90% threshold as compared to other public debt-to-GDP ratios. Moreover, the debt-growth link disappears for the public debt ratios below 90% threshold; see Herndon et al. (2014) for a criticism of Reinhart and Rogoff (2010).

In the empirical growth literature, an extensive amount of studies formally tested the 90% threshold level using different sets of countries, data coverages, and debt measures, as well as using different empirical methods. An important observation gleaned from the recent studies aimed at unveiling the nonlinear relationship between government debt and economic growth is that there is no common finding for the threshold level, except for a small number of research papers, which find a turning point for a public debt-to-GDP ratio at around 90%. As one study in the latter group of papers, Cecchetti et al. (2011) look at a panel of 18 OECD countries (all from advanced economies) for the period 1980-2006. Using least squares dummy variable and threshold estimation within the context of dynamic fixed-effects panel data model, they find a negative relationship between government debt and growth beyond the 85% threshold level, after controlling for other determinants of growth including trade openness, inflation rate, and total dependency ratio (related to ageing). Their approach avoids possible feedback effect from economic growth to public debt using five-year averages of growth, so that regressors are predetermined. Their results suggest that on average, a 10 percentage points increase in public debt-to-GDP ratio is predicted to reduce economic growth by 0.13 percentage points per year. Checherita-Westphal and Rother (2012) study 12 euro area economies from 1970-2008 aiming at to investigate nonlinearity in the debt-growth link by using a quadratic equation in debt. To control for endogeneity of public debt variable, the authors use lagged value of debt and average debt of the other countries in the sample. They find a public debt threshold level in between 90% and 100%, beyond which economic growth is negatively affected. Baum et al. (2013) deal with the endogeneity problem arising from dynamic model specification in their study of 12 euro area countries from 1990-2007/2010. They find a threshold level of public debt-to-

GDP ratio at 95% for the extended period. In a recent publication, Woo and Kumar (2015) look at 38 advanced and emerging economies from 1970-2008. Using several estimation strategies and subsamples, the authors examine nonlinearity in the debt-growth relationship by fitting the data to the dynamic panel regression model. They also find a 90% threshold level, beyond which public debt has a negative and significant effect on economic growth. In a last study that needs to be emphasized, Panizza and Presbitero (2014) account for the potential endogeneity of public debt using the share of foreign currency debt in total public debt as an instrument. Using the same data set and empirical approach of Cecchetti et al. (2011) as well as performing various robustness checks, they find little evidence on the adverse effect of high public debt on future growth in advanced economies.

Many other studies provide evidence of a threshold level of public debt different than 90 percent of GDP as these studies use a different set of countries in their sample. For example, Caner et al. (2010) look at a cross-section of 101 developed and emerging market economies from 1980-2008. Using threshold estimation, they find a turning point of public debt-to-GDP ratio at 77% for the full sample, while this value is lower, at 64% of GDP, for the subsample of developing countries only, after controlling for initial GDP per capita, trade openness and inflation rate. In the Wright and Grenade (2014) study of 13 Caribbean countries from 1990-2012, the authors find a threshold level of 61% of GDP beyond which debt has a negative effect on economic growth and investment. A few other research papers closely replicate Reinhart and Rogoff (2010)'s paper using econometric techniques. For example, using panel smooth transition regression model, Minea and Parent (2012) find a negative and gradually decreasing effect of public debt on growth below the threshold level of 115%. Their finding does, in fact, support the presence of nonlinearity in the effect of debt on growth for the debt-to-GDP ratio above 90%. On the other hand, they find a positive growth effect of debt for the debt level above 115%. Relatedly, using nonlinear threshold models for the same dataset used in Reinhart and Rogoff (2010), Égert (2015) found limited evidence for a negative nonlinear correlation between public debt and growth. The author's findings suggest that a debt threshold level can be lower than 90% of GDP depending on data coverage (in terms of country coverage and time dimension), model specification, and measure of the public debt. Eberhardt and Presbitero (2015) provide strong evidence of different nonlinearities in the debt-growth relationship across 118 countries from 1961-2012 by doing comprehensive analysis of dynamic panel time series estimation. They employ common factor framework to uncover possible heterogeneity in the effect of public debt stock on economic growth through taking into account latent factors of growth and public debt, which include a country's debt composition, macroeconomic policies related to past crises, and institutional framework. They find no evidence for the common threshold effect for all countries in their sample.

The main purpose of the above research and analysis is to reveal a nonlinear relationship between public debt and economic growth depending on the public debt level. In other words, those papers try to expose nonlinear growth effect of high public debt levels. However, this point of view ignores that other variables may govern the debt-growth relationship. Formally testing for several threshold variables including democracy, trade openness, fertility, life expectancy, and inflation rate, among others, Kourtellos et al. (2013) study 82 countries in a 10-year panel from 1980-2009. They employ the structural threshold regression model of Kourtellos et al. (2016) to account for the endogeneity of both the threshold variable and the regressors. The authors find a strong evidence in favor of heterogeneity in the debt-growth relationship in the sense that the effect of public debt on economic growth depends on the institutional quality of a country. Particularly, they find that countries with low institutional quality experience a negative and significant effect of public debt on economic growth, holding other factors fixed, while public debt has a positive but insignificant effect on growth for countries with high institutional quality. Jalles (2011) investigate the impact of democracy and corruption on the external debt-growth relationship in a panel of 72 developing countries from 1970-2005. Using fixed effects and GMM estimation strategies under various model specifications (linear and quadratic terms in debt-to-GDP ratio), they find a negative growth effect of external debt in countries with higher levels of corruption.

A few more observations can be gathered from the past literature on empirical debt-growth nexus. First, the relationship appears be heterogeneous and complex. Second, there might be other factors that potentially contribute to the marginal impacts of regressors on economic growth rates, which implies that heterogeneity in the debt-growth relationship might be with respect to other variables in the model. Third, there is lack of strong evidence of the negative effect of public debt on economic growth for advanced economies. These limitations of the existing debt-growth literature, coupled with the lack of clear theoretical argument on the debt-growth link (in advanced economies), suggests that a flexible approach may be more appropriate for estimating the effect of debt on growth and letting other factors to characterize this relationship. We, therefore, present an augmented conventional Solow economic growth model with public debt-to-GDP ratio and country-specific parameters, which relax the homogeneity assumption of a standard growth regression. Specifically, we model parameters to be a function of one or more covariates including democracy, fertility, and life expectancy, among others. Our approach is also related to the empirical growth studies that use nonparametric and semiparametric models to model parameter heterogeneity in the cross-country growth process. Examples are Liu and Stengos (1999) and Ketteni et al. (2007) for an additive semiparametric partially linear model, Vaona and Schiavo (2007) for a semiparametric partial linear model, Durlauf et al. (2001), Mamuneas et al. (2006), Kourtellos (2011), and Kumbhakar and Sun (2012) for a varying coefficient model and Henderson et al. (2011) for a nonparametric model.

To ensure that our regression model captures heterogeneous effects of variables, we further assume the parameters to be unknown measurable smooth functions. This assumption enables us to use nonparametric techniques, which essentially let the data decide functional form of each parameters. In addition, the coefficient estimates avoid biasedness by the misspecification of parameter heterogeneity, which is in parametric form in existing debt-growth studies. Furthermore, economic theory does not suggest a functional form for the regression model of debt-growth relationship or even for the parameter heterogeneity in the debt-growth link. Therefore, nonparametric techniques permit unknown functions to be governed by country-specific characteristics such as country's initial conditions, state of development variables, institutional quality, and macroeconomic policies playing an indirect role in explaining nonlinear relationship between growth and its determinants across countries and time domain.

We use a recently developed smooth coefficient instrumental variable estimator (Delgado et al., 2015) that assumes linearity in the regressors, but allows parameters vary nonparametrically with respect to a set of covariates. One advantage of this estimation method is to control for endogeneity of covariates in the functional coefficients.

In terms of our findings, we find strong evidence of heterogeneity in the effect of public debt with respect to institutional quality of countries. Our results support Kourtellos et al. (2013), which suggest an adverse effect of public debt on growth for the countries below a particular institutional quality level. However, our results also show that for countries with a democracy score above a critical level, higher public debt level leads to lower economic growth (all else equal). But, this effect is comparably less strong than for the countries with a lowest democracy score. Our findings are robust to using other measures of institutional quality, using alternative covariates in the functional form, controlling other variables in the regression model, and using different subsamples of countries. However, when we exclude the outliers (Guyana and Nicaragua), only countries with high institutional quality has a negative and significant growth effect of public debt. Our results from prediction exercises also suggest that our semiparametric model can better describe the underlying process that generated the data. Our paper therefore contributes to the empirical debt-growth literature from the point of view that explains parameter heterogeneity in the cross-country growth process through fundamental determinants of economic growth proposed by new growth theories.

The remainder of this paper is organized as follows. Section 2 describes our empirical methodology. Section 3 describes our data. In Section 4 we present the empirical results of the paper. In Section 5 we present robustness checks. Section 6 concludes.

4.2 Empirical Methodology

4.2.1 The augmented Solow growth model

In this section, we provide a brief description of a linear Solow growth model augmented with the debt-to-GDP ratio to investigate the impact of country's debt level on its economic growth rate. This model assumes a common regression across countries as well as constant coefficient estimates for all economic variables, which intuitively explains the average effect of the variables.

$$g_i = X_i^T \beta + u_i = \beta_0 + S_i^T \beta_s + \beta_d debt_i + u_i, i = 1, .., n, \tag{4.1}$$

where $X_i = [1, S_i^T, debt_i]^T$ is a $(d_s + 2) \times 1$ vector of regressors consists of a constant term, a d_s dimensional vector of standard Solow growth determinants, including $ln(yin_i)$, the logarithm of the i^{th} country's real GDP per worker in the initial year of each 10-year period; $ln(s_i)$, the logarithm of the i^{th} country's average saving rate; $ln(n_i + 0.05)$, the logarithm of the i^{th} country's population growth plus 0.05; and $ln(sch_i)$, the logarithm of the i^{th} country's average years of secondary and tertiary schooling for male population over 25 years of age, and $debt_i$ which is defined as the i^{th} country's public debt-to-GDP ratio. Moreover, S_i includes a time trend. u_i is an identically independently distributed error term.

4.2.2 An endogenous smooth coefficient model

We consider the following semiparametric varying coefficient model of Delgado et al. (2015) for the augmented Solow growth model:

$$\begin{cases} g_i = \theta_0(Z_i) + \sum_{j=1}^{d_s} \theta_{sj}(Z_i)S_{ji} + \theta_d(Z_i)debt_i + \epsilon_i \\ Z_i = \mu_Z + a_1(E_{i,1}) + a_2(E_{i,2}) + ... + a_p(E_{i,p}) + u_i, i = 1, ..., n, \end{cases} \tag{4.2}$$

$$(i)E[u_i|\boldsymbol{E_i}] = 0$$

$$(ii)E[\epsilon_i|\boldsymbol{E_i}, u_i] = E[\epsilon_i|u_i], i = 1, ..., n,$$

where Z_i is an endogenous variable defined as an additive nonparametric functions of E_{ij}, $j = 1, ..., p$, where $\boldsymbol{E_i} = [E_{i,1}, E_{i,2}, ..., E_{i,p}] = [S_i^T, debt_i, W_i^T]^T$ is a $p \times 1$ vector of continuous variables including a d_w dimensional vector of instrumental variables, W_i^T. $a_t(\cdot)$, $t = 1, ..., p$, $\theta_0(\cdot), \theta_s(\cdot)$, and $\theta_d(\cdot)$ are all unknown smooth measurable functions and

u_i is zero-mean error term.

In Equation (4.2), the object of estimation is structural model that necessitates different identification strategy than standard nonparametric regression, which is used to estimate conditional expectations. Additive separability of Z and conditional mean of ϵ and u given in (i) and (ii) in Equation (4.2) are nonparametric restrictions for identification in this model[1].

After setting $E[\epsilon_i|u_i] \equiv b(u_i)$ and denoting $v_i \equiv \epsilon_i - b(u_i)$ that satisfies $E[v_i|\boldsymbol{E_i}, u_i] = 0$, we can rewrite Model (4.2) as

$$g_i = \theta_0(Z_i) + \sum_{j=1}^{d_s} \theta_{sj}(Z_i)S_{ji} + \theta_d(Z_i)debt_i + b(u_i) + v_i, i = 1, .., n, \tag{4.3}$$

provided that $b(\cdot)$ is an unknown smooth function. Equation (4.3) consists of two additive components, $\theta_0(Z_i)$ and $b(u_i)$, together with the functional coefficient terms, $\sum_{j=1}^{d_s} \theta_{sj}(Z_i)S_{ji}$ and $\theta_d(Z_i)debt_i$. According to Newey et al. (1999), identification of unknown functions in Equation (4.3) is the same as identification in Equation (4.2), as the additive structure of Equation (4.3) is equivalent to conditional mean restriction (assumption (ii)) in Equation (4.2). The sufficient condition for identification of unknown functions in Equation (4.3) is, therefore, assuming no additive functional relationship between Z_i and u_i (see Newey et al. (1999, Theorem 2.1 and 2.2 on page 567-568)).

If we assume that Z and all conditioning variables are exogeneous, then the first equa-

[1] In another paper (Newey and Powell, 2003) conditional mean of disturbances given instruments are assumed to be zero without imposing an additive structure for the endogeneous variables.

tion in (4.2) is a pure varying coefficient model that can be consistently estimated using the nonparametric kernel estimator of Li et al. (2002b); otherwise, this estimator yields a bias in estimation of unknown functional coefficients. Assuming exogeneity of covariates seems to be strong in the present growth application; we, therefore, allow variables representing Z to be endogeneous.

Other than the methodology adopted in this paper, nonparametric estimators for regression models that include endogeneity problem have been proposed in the context of varying coefficient models, for example, Das (2005), Cai et al. (2006), Cai and Li (2008). However, these papers allow for endogeneous variables in the parametric part of a regression. The estimator proposed by Delgado et al. (2015), on the other hand, deals with endogenous variables that appear in the nonparametric part of a smooth coefficient model. This estimator is applicable to the economic studies, where endogeneous variable has a potential interaction effect with the other regressors on response variable. For example, child care use can have a potential indirect effect on students' test scores through a flexible interaction with mother's education, age, and experience, among other regressors (see Bernal and Keane (2011) for a parametric estimation and full description of the regressors and Ozabaci et al. (2014) for an additive nonparametric regression estimation).

To circumvent the endogeneity problem, Delgado et al. (2015) use the control function approach in the estimation of structural function of interest. Since u enters Equation (4.3) as a conditioning variable and it is generally unobserved, Delgado et al. (2015), first, calculate $\hat{u}$ from the regression of Z on $\boldsymbol{E}_i$ using second equation of Model (4.2). Then, they estimate

$\theta(Z_i)$ and $b(\hat{u})$ via sieve approximation approach by an ordinary least squares method. In the third step, they use a local linear regression method to estimate $\theta(Z_i)$ and $\theta'(Z_i)$. They show that their estimator is oracle efficient in the sense that large sample distribution of the estimator is the same regardless of whether the function $b(\cdot)$ is known. It is also noted that third-step estimator is not affected from the errors in the first two steps of estimation. The estimation procedure is given in detail as follows.

In the first step, Delgado et al. (2015) approximate unknown functions $a_1(\cdot)$,...,$a_p(\cdot)$ by series expansions[2]

$$a_m^*(e) = \sum_{l=1}^{L_n} \alpha_{ml}\phi_l(e), \tag{4.4}$$

for $m = 1, ..., p$, where $\alpha_m = (\alpha_{m1}, \alpha_{m2}, ..., \alpha_{mL_n})^T$ is $L_n \times 1$ vector of unknown coefficients, $\{\phi_j(\cdot)\}_{j=1}^{L_n}$ is a sequence of square integrable orthonormal basis functions over the interval $[0, \infty)$, and L_n denotes the number of basis functions. It is noteworthy that Laguerre polynomial series is used to approximate the unknown functions as it is one of the common choices for series expansions when a function has a domain over $[0, \infty)$ (see, e.g., Assumption 1(ii) in Delgado et al. (2015) and Chen (2007) for further details).

The coefficients α_m, $m = 1, ..., p$ in (4.4) can be consistently estimated from the ordinary least squares (or OLS) regression of Z_i on $a_1^*(E_{i,1}), a_2^*(E_{i,2}), ..., a_p^*(E_{i,p})$. Then, the OLS estimator of the unknown function is given by $\hat{a}_m(e) = \sum_{l=1}^{L_n} \hat{\alpha}_{ml}\phi_l(e)$, $m = 1, ..., p$. Fitted values and the residuals from the OLS regression can be calculated as

[2]The authors use B-spline smoothing in the first two steps assuming domain of the basis functions over the closed interval.

$\hat{Z}_i = \hat{\mu} + \hat{a}_1(E_{i,1}) + \hat{a}_2(E_{i,2}) + ... + \hat{a}_p(E_{i,p})$ and $\hat{\epsilon}_i = Z_i - \hat{Z}_i$ for all $i = 1, ..., n$, respectively.

In the second step, using series expansions they approximate unknown functions $\theta(z)$ and $b(\hat{\epsilon}_i)$, respectively, by

$$\theta_k^*(z) = \sum_{l=1}^{L_n} \beta_{kl}\phi_l(z), \qquad \text{and} \qquad b^*(\hat{\epsilon}) = \sum_{l=1}^{L_n} \gamma_l \phi_l(\hat{\epsilon}), \tag{4.5}$$

where $\beta_k = (\beta_{k1}, \beta_{k2}, ..., \beta_{kL_n})^T$ for $k = 0, ..., d_s + 1$, and $\gamma = (\gamma_1, \gamma_2, ..., \gamma_{L_n})^T$ are all $L_n \times 1$ vectors of unknown coefficients. Model (4.3) can be, now, approximated by substituting equalities in (4.5) for $\theta_k(z)$, $k = 0, ..., d_s + 1$, and $b(\hat{\epsilon})$ in (4.3).

$$g_i \approx \sum_{k=0}^{d_s+1} \sum_{l=1}^{L_n} \beta_{kl}\phi_l(z) X_{ki} + \sum_{l=1}^{L_n} \gamma_l \phi_l(\hat{\epsilon}_i) + v_i, i = 1, .., n, \tag{4.6}$$

where residuals $\hat{\epsilon}_i$ is calculated from the first step. The least squares problem is, then, defined as follows:

$$[\hat{\beta}^T, \hat{\gamma}^T]^T = \arg\min_{(\beta,\gamma)} \sum_{i=1}^{n} \left\{ g_i - \sum_{k=0}^{d_s+1} \sum_{l=1}^{L_n} \beta_{kl}\phi_l(z) X_{ki} + \sum_{l=1}^{L_n} \gamma_l \phi_l(\hat{\epsilon}_i) \right\}^2. \tag{4.7}$$

In the third step, Delgado et al. (2015) use the local linear regression approach to estimate the functional coefficients, $\theta(\cdot)$, and its first-order derivatives, $\theta^{'}(\cdot)$. Following Delgado et al. (2015), we assume that unknown function, $\theta(Z)$ is continuously differentiable up to second order so that we can apply a first order Taylor series approximation of $\theta(Z)$

around a given point z, technically by $\theta(Z) \approx \theta(z) + \theta^{'}(z)(Z - z)$. We, further, assume that $K(\cdot)$ to be a kernel weight function assigning more weights to the observations closer to point z, satisfying: (i) $\int K(a)da = 1$, (ii) K(a)=K(-a), and (iii) $\int a^2 K(a)da > 0$.

Replacing $b(\epsilon_i)$ in Equation (4.3) by $\hat{b}(\hat{\epsilon}_i)$ calculated from the second step estimation and treating $\hat{g}_i = g_i - \hat{b}(\hat{\epsilon}_i)$ as a dependent variable, Delgado et al. (2015) show that a consistent estimate of $(\theta(\cdot), \theta^{'}(\cdot))$ can be obtained from a minimization of a kernel-weighted objective function:

$$\min_{\theta(z),\theta^{'}(z)} \sum_{i=1}^{n} [\hat{g}_i - X_i^T \theta(z) - X_i^T \theta^{'}(z)(Z_i - z)]^2 K((Z_i - z)/h), \tag{4.8}$$

where $\theta^{'}(z)$ reflects the partial effects $\partial\theta(z)/\partial z$ and h is the bandwidth controlling the size of the local neighborhood around an interior point z.

Letting $\delta(z) = [\theta(z), \theta^{'}(z)]$, the solution of problem (4.8) is given by

$$\tilde{\delta}(z) = (\mathbf{X}^T \mathbf{K} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{K} \hat{g}, \tag{4.9}$$

where **X** is a $n \times 2(d_s + 2)$ matrix having $(X_i^T, X_i^T(Z_i - z))$ as its ith row and **K** is a $n \times n$ diagonal matrix with the ith diagonal element being $K((Z_i - z)/h)$.

The bandwidth parameter has a particular importance in estimation of non- /semiparametric models as it determines the degree of smoothing. We use a cross-validation method, a data-driven approach, to choose the bandwidth parameter so that the bias-variance tradeoff in the estimation is optimized by using the data itself. We also provide wild-bootstrap standard errors, which are robust to heteroscedasticity, using 399 bootstrap replications (see Härdle and Marron (1991, p.782)). The confidence intervals for each coefficient estimates plotted in Figure 4.1 and 4.2 are constructed from 95% bootstrap percentiles.

We use three goodness-of-fit measures including in-sample R^2, out-of-sample R^2, and average squared predicted error (ASPE). The out-of-sample measures are robust to over-fitting of the model, which, therefore, implies that the model of interest may better describe the underlying process that generated the data. The predictive exercises are based on 1000 bootstrap replications. We use 80 percent of the data to estimate the model parameters and evaluate on the hold-out data. See Hayfield and Racine (2008) and Racine and Parmeter (2013) for further details on out-of-sample prediction exercises.

4.3 Data

We employ the same data set as used in Kourtellos et al. (2013). We have a balanced 10-year period panel dataset covering 82 countries in 1980-1989, 1990-1999, and 2000-2009. This data set does not include any short-run fluctuations as they are smoothed out by taking 10-year averages. The set of functional covariates described below are the threshold variables that resulted in a rejection of the null hypothesis of global linearity in the model in Kourtellos et al. (2013). The variables, in addition to the Solow variables, include *fertility*, the log of the average total fertility rate; *life expectancy*, the log of the average life expectancy at birth; *inflation*, the log of the average inflation rate plus one; *tropics*, the percentage of a country's land area that is classified as tropical or subtropical; *democracy*, a measure of the extent of institutionalized democracy; and *executive constraints*, a measure of the extent of institutionalized constraints on the power of chief executives. We also consider *public debt-to-GDP* and *trade openness*, i.e., the average ratio for each period of exports plus imports to GDP, as a variable that the coefficients of the regressors could vary over. Summary statistics of the variables used in this paper are given in Table 4.2.

4.4 Estimation Results

We present estimates from various model specifications for the augmented Solow growth model and an endogeneous semiparametric smooth coefficient model in Table 4.1. Columns 1-8 show estimates for four homogeneous model specifications from ordinary least squares and two-stage least squares estimation. Since semiparametric models take democracy into account through the functional coefficients, we include democracy as an additional conditioning variable in the standard growth model specifications. Year indicator is another factor that is controlled for in the parametric regression models in columns 1-8. Columns 1-4 show that the OLS estimates for the coefficient of public debt are negative and are significant at the 5% and 10% levels with their values ranging from -0.0058 to -0.0080. The OLS regression in column 3 suggests that a 10 percentage points increase in the debt-to-GDP ratio is, on average, associated with a 0.060% decrease in subsequent 10-year period real GDP growth rate. In columns 5-8, the estimates of public debt coefficient decrease sharply in magnitude as well as lost its statistical significance when we account for an endogeneity using the two-stage least squares regression. Other than public debt variable, only democracy and investment rate are statistically significant with their expected sign in columns 5-8.

Table 4.1: Summary of the results

Variable	OLS				2SLS				SPSCM-IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Intercept	**0.0355**	**0.0258**	-0.0203	-0.0126	0.0174	0.0037	-0.0202	-0.0142	**0.0408**	0.0193	**0.0572**
	0.0143	0.0143	0.0437	0.0450	0.0170	0.0172	0.0459	0.0472	0.0115	0.0385	0.0375
Public Debt	**-0.0080**	**-0.0067**	**-0.0060**	**-0.0058**	-0.0036	-0.0017	-0.0024	-0.0019	**-0.0071**	**-0.0072**	**-0.0028**
	0.0034	0.0033	0.0033	0.0033	0.0040	0.0040	0.0036	0.0035	0.0035	0.0025	0.0011
Democracy		**0.0012**	**0.0014**	**0.0014**		**0.0016**	**0.0023**	**0.0022**	—	—	—
		0.0004	0.0006	0.0006		0.0004	0.0007	0.0007			
Initial Income			-0.0049	-0.0051			-0.0055	-0.0056		**-0.0097**	**-0.0081**
			0.0035	0.0035			0.0032	0.0033		0.0024	0.0024
Investment Rate			**0.0178**	**0.0176**			**0.0193**	**0.0191**		0.0078	**0.0077**
			0.0053	0.0053			0.0054	0.0053		0.0040	0.0039
Population Growth Rate			-0.0111	-0.0102			-0.0045	-0.0039		**-0.0291**	**-0.0281**
			0.0248	0.0248			0.0240	0.0241		0.0141	0.0139
Schooling			0.0050	0.0051			0.0044	0.0045		**0.0090**	**0.0090**
			0.0039	0.0040			0.0039	0.0039		0.0029	0.0030
Inflation Rate				-0.0015				-0.0017			**-0.0053**
				0.0012				0.0013			0.0023
In-Sample R^2	0.0832	0.1211	0.2093	0.2154	0.0721	0.1020	0.1951	0.2008	0.1803	0.3827	0.4150
Out-of-Sample R^2	0.0982	0.1399	0.2684	0.2767	0.0852	0.1196	0.2492	0.2585	0.1187	0.3099	0.3411
ASPE	0.00048	0.00046	0.00044	0.00044	0.00049	0.00048	0.00071	0.00072	0.00049	0.00041	0.00040

1. Semiparametric model specifications allow coefficients to vary with respect to democracy.
2. We use Gaussian kernel function for all semiparametric estimation. The cross-validated bandwidth in column 10 is 1.62.
3. Statistically significant coefficient estimates (at the 5% level) are highlighted in bold.
4. Column 9-11 reports the mean coefficient estimates and their respective standard errors.
5. Out-of-sample R^2 and ASPE report mean of 1000 bootstrap replications.

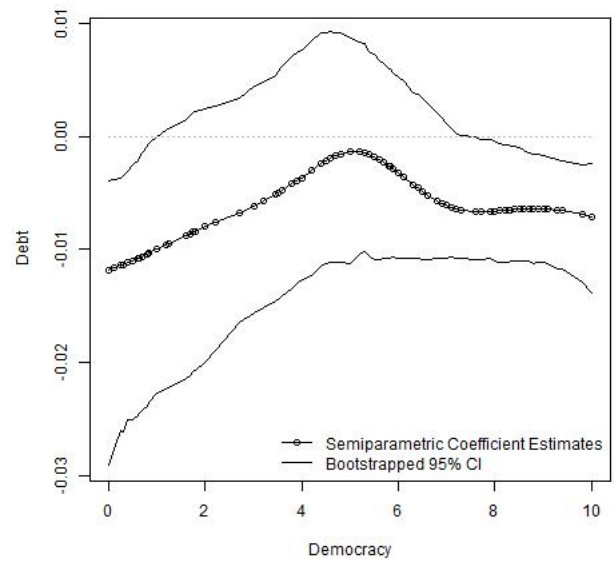
Columns 9-11 reports average of semiparametric smooth coefficient instrumental variable estimates and its standard error. Column 9 and 10 show that the coefficient estimates of public debt are negative and statistically significant at 5% and 1% level with having the values around -0.0071 and -0.0072, respectively. The estimated effect suggests that a 10 percentage points increase in the debt-to-GDP ratio is associated with a 0.072% decrease in subsequent ten-year period real GDP growth rate, on average. Comparing column 10 and 3, we observe that mean value of public debt coefficient estimates from the semiparametric smooth coefficient instrumental variable estimation is almost in agreement with that of from the ordinary least squares estimation. This result suggests that average of ten-years period democracy score of countries may avoid us to have endogeneity problem as it is emphasized in Kourtellos (2011) that institutions are among the medium moving variables comparable to faster moving variables including the conventional Solow variables. Nevertheless, the in-sample goodness of fit of the semiparametric model (38%) is higher than for the parametric model (20%). This comparison holds for all specifications between semiparametric and parametric regression models. We further investigate the model's out-of-sample performance to decide on whether this improvement reflects over-fitting or not. In each semiparametric models in columns 9-11, the out-of-sample R^2 (ASPE) is in general higher (lower) than in the corresponding parametric models. The results indicate that semiparametric smooth coefficient model in column 10 is 7.3% more efficient than the parametric linear model in column 3 in terms of out-of-sample predictive ability, which, therefore, implies that the semiparametric model may better describe the underlying process that generated the data than the parametric model does. One may have a concern that higher-order polynomial terms in the homogeneous model may be sufficient to capture the parameter heterogeneity. This concern can be highlighted by the bias-variance trade-off in both parametric and nonparametric model estimation. Adding polynomial terms in a parametric regression model reduces the bias of the estimates (since more information is used in the estimation), but the parameters are less accurately estimated (i.e., standard errors are larger). Therefore, nonlinearity in the parametric model may be captured at the cost of efficiency. Nonparametric regression model, on the other hand, allows to control the bias-variance trade-off through the choice of bandwidth parameter, which essentially controls the local sample size for the estimation of each point of interest. Furthermore, the bandwidth can be chosen using the data via the cross-validation method. Therefore, nonparametric models, comparable to parametric counterparts, is believed to bring a better fit to the data, and, thus, more reliable inference.

The coefficients on other explanatory variables (initial per capita income, investment rate, population growth rate and average years of schooling) in columns 9-11 are of the predicted sign and significant at conventional levels. Column 11 reports the mean estimates for the semiparametric regression model, which controls for inflation rate as well. All variables have statistically significant coefficient estimates at conventional levels, but the magnitude of the coefficient estimate of public debt decreases more than half as the inflation rate accounts for part of its negative effect on economic growth. This result is, in fact, consistent with the theoretical literature on inflation and economic growth (see Barro (1995) for theoretical background and Vaona and Schiavo (2007) for a semiparametric application). Homogeneous model specifications in column 4 and 8, on the other hand, do not estimate economically significant drop in the growth effect of public debt when inflation rate is included as additional conditioning variable.

Figure 4.1 displays country-specific coefficient estimates for public debt variable from the semiparametric regression model in column 10 of Table 4.1 along with 95% bootstrap confidence intervals[3]. We see that more public debt leads to lower economic growth for

[3]Henderson et al. (2012) suggest to plot gradient estimates in a 45^o plot to expose parameter heterogeneity

Figure 4.1: Estimated coefficient curve for public debt



countries with democracy score less than 1 and greater than 7.6, holding other factors fixed. This result is consistent with the existing literature that suggest that countries with weak institutional quality are the only ones that tend to have an adverse effect of more public debt on growth. However, our results also show that more public debt can be detrimental to growth for countries with strong institutional quality (all else equal). On the other hand, for countries with a democracy score in between 1 and 7.6, public debt has no significant effect on growth. Particularly, the impact of public debt on growth for countries with a median level of democracy score reduces to values around zero, which are therefore economically insignificant as well. This result can be justified by the fact that countries in transition to institutionalized democracy may obey their debt payments, which therefore allows them to

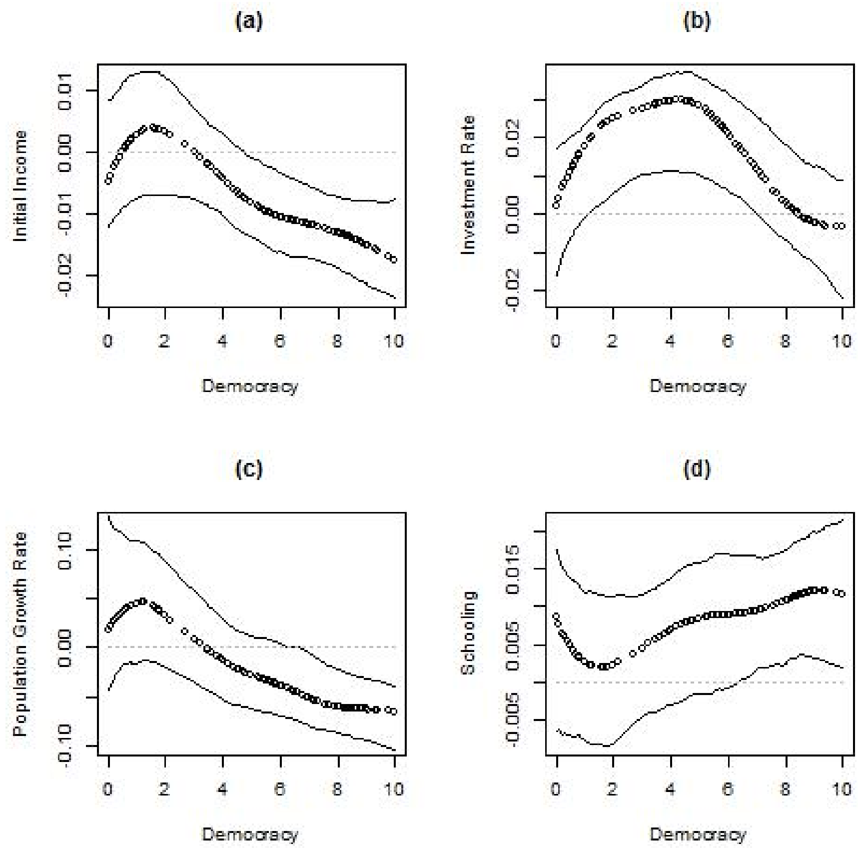
that exists in the estimates. Their suggestion is useful especially when covariate vector is more than one dimension. Since in our model estimation the coefficients vary with respect to only one variable, from the graphical point of view it is better to plot coefficient estimates on a Cartesian coordinate system.

borrow again so that public debt becomes growth neutral for these countries.

The mean coefficient estimates of the countries in the low and high democracy groups are -0.0114 and -0.0069, respectively, while average public debt-to-GDP ratio for the two country groups are 52.46 and 71.52, respectively. This result contradicts traditional parametric regression models, which suggest high public debt and low economic growth relationship beyond various threshold public debt levels. In particular, this result suggests that public debt adversely affects economic growth in countries with weak institutional quality less strong than countries with high institutional quality. Our results also show some evidence of parameter heterogeneity for countries with democracy score less than 1. However, since we use pooled sample that includes three ten-year periods in the estimation, our results in Figure 4.1 does not reflect country characteristics as well as global factors. In fact, the semiparametric regression model in column 10 estimates the same value, -0.0071, for advanced countries with a democracy score of 10. Moreover, thirty seven countries, which have a democracy score of zero, are predicted to have the same growth effect of public debt, i.e., -0.0118.

Our results give some support to the recently shifted focus in the debt-growth nexus that institutions may be one of the main factor that determines the debt-growth link. Our contribution in this paper also puzzles over the complexity of debt-growth relationship for advanced countries. We would like to mention first that we are not able to expose heterogeneity in the effect of public debt on growth within the advanced country group in Figure 1 as there is no variability in the *democracy* for these countries. This result

Figure 4.2: Coefficient estimates for other regressors



essentially highlights the need of other potential candidate variables such as sovereign debt credit ratings that might have a variability within the advanced country group. We defer this for future research. On the other hand, high public debt under the moral hazard context may be one of the possible reason for having significant negative coefficient estimates for advanced countries in Figure 1. Particularly for some of the Euro area countries including Greece, Italy, Portugal, and Spain, the root of the public debt crises is at an excessive risk taking behavior of countries that results from widespread support to the financial system (Allen et al., 2015).

The curves in Figure 4.2 show how democracy affects the coefficients of other conditioning variables. In Figure 4.2a, we find that countries with an institutionalized democracy, a score greater than 4.7, have an increasing significant negative effect of initial income on economic growth, which confirms the conditional β-convergence hypothesis. The curve in Figure 4.2b exhibits a significant positive and an inverse U-shape relationship between the real investment rate and the real GDP per capita growth rate for the countries with a democracy score in between 1.2 and 7. It is observed in Figure 4.2c that higher population growth rate is associated with a slowdown in an economic growth for the countries with a democracy score greater than 6.6. Lastly, schooling in Figure 4.2d has a significant positive effect on growth rate for the countries with a democracy score greater than 6.3. For a better exposition of parameter heterogeneity for these variables, we plot the estimated curves in a larger scale in Figure 4.3 in the appendix. It is clearly seen that for each regressors, except for the investment rate, there is a heterogeneous relationship in the effect of the variable on economic growth rate for the mid- and high-level democracy score countries.

4.5 Robustness Checks

We aim to see the main results obtained from the semiparametric model in the previous section are robust to additional model specifications.

4.5.1 *Alternative measure for democracy*

We examine whether our main results are sensitive to different measures of institutional quality such as executive constraints obtained from the same data source, Polity IV. We find that countries with an executive constraint score less than 2.8 and greater than 5.8 have a significant negative coefficient estimates for public debt variable. This result does not alter the conclusions drawn from our main results; that is, institutional quality is an important factor that governs the effect of public debt on growth.

4.5.2 *Additional control variables*

We estimate semiparametric endogeneous smooth coefficient model that excludes three Solow variables (investment rate, population growth rate and average years of schooling) and adds inflation rate and trade openness. The results show that only the coefficients of public debt and inflation rate are significant and negative at conventional levels. Our main results are the same in this model estimation. We also do the estimation of the regression model that includes government consumption. We continue to find a negative and significant relationship between public debt and real per capita GDP growth rates for the countries with low and high democracy scores. Moreover, estimated coefficient curve for public debt retains the same path as in the model in column 10 of Table 4.1.

4.5.3 *Outliers*

We exclude the countries with the highest average debt-to-GDP ratios to see whether our results may be driven by the outliers. These countries are Guyana and Nicaragua. We find a significant negative effect of public debt on growth for only countries with a democracy score greater than 8.8. Our core results are changed in terms of having insignificant and weaker growth effect of debt for low democracy score countries. We also observe that the negative effect of public debt is getting slightly stronger as the countries' institutional quality rising up by one point (from democracy score of 9 to 10). This result does not depend on the public debt level of countries as the average public debt level for countries having a democracy score greater than 9 and less than 10 is 57% of GDP and that for countries with a democracy score equal to 10 is 56% of GDP. We may wish to conduct exclusion exercise for the highest public debt levels in advanced countries including Japan (public debt level is 178% of GDP on average for the period 2000-2009) and Israel (196% of GDP on average for the period 1980-1989). However, in this case the local linear regression estimation does not account for these countries' lower public-debt-to-GDP ratios that belong to the previous decades, which, therefore, does not reflect our interest of the relationship between debt-growth in the advanced countries.

4.6 Further Research

We further investigate parameter heterogeneity using other potential covariates including fertility rate, life expectancy, inflation rate, trade openness, and government consumption, respectively. We find that mean coefficient estimates for public debt is significant at conventional level for all covariates except for government consumption. However, only at around 7 percent of the coefficient estimates are significant in each of the models that controls for inflation rate and government consumption in the functional coefficients. Furthermore, none of these covariates expose heterogeneous relationship in between debt and growth. One possible explanation behind these results is that these covariates are fast-moving that result in less observations in the local sample size as opposed to medium-moving variables such as democracy.

4.7 Conclusion

We employ a semiparametric smooth coefficient model with an endogenous variable in the nonparametric part to analyze heterogeneous relationship between debt and growth with a sample of 82 countries over the three 10-year periods. Most of the papers in the growth literature have aimed at finding a nonlinearity in the debt-growth relationship depending on the level of public debt. As it is emphasized with evidences in Kourtellos et al. (2013), nonlinear effect of debt on growth depends on a country's institutional quality. In the same vein, our results show some evidence of heterogeneity in the effect of debt on growth for the countries with a democracy score below 1. Our results also show some evidences for the adverse effect of public debt on growth in advanced countries.

Appendix

Figure 4.3: Coefficient estimates for other regressors

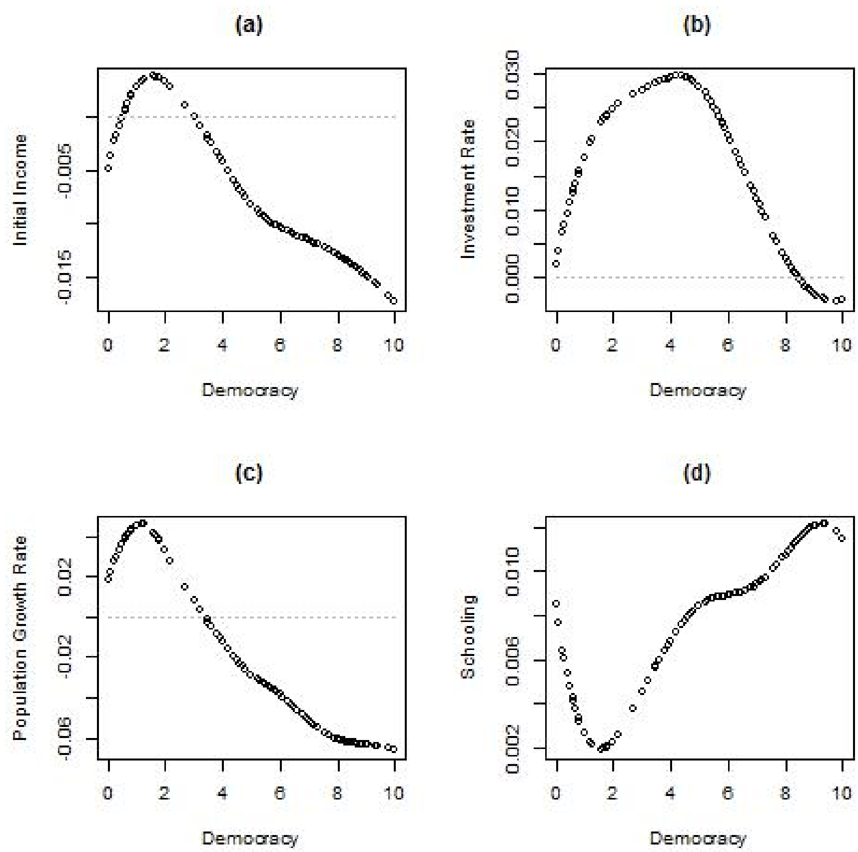


Table 4.2: Summary statistics of the variables in the dataset

Variable	Mean	Std. Dev.	Min	Max
Growth	0.013737	0.022960	-0.099459	0.083383
Initial income	8.423263	1.266566	5.86825	10.71059
Lag of initial income	8.335907	1.232223	5.77992	10.5477
Investment rate	3.046038	0.351779	1.873230	3.891546
Lag of investment rate	3.055552	0.394585	1.743324	4.312730
Population growth rate	-2.711415	0.160957	-3.2289	-2.384709
Lag of population growth rate	-2.690982	0.165420	-3.083584	-2.276809
Schooling	0.598071	0.768791	-2.183513	1.970172
Lag of schooling	0.320655	0.901583	-2.662667	1.901029
Public debt	4.080373	0.610824	2.173561	6.327447
Lag of public debt	3.924271	0.730973	1.116990	6.462404
Fertility	3.616907	1.729945	1.2080	7.7777
Lag of fertility	4.064228	1.888114	1.1660	7.8244
Life expectancy	4.170482	0.175197	3.634336	4.406866
Lag of life expectancy	4.139680	0.176947	3.636147	4.385104
Trade openness	66.5114	36.4878	9.7683	199.8575
Lag of trade openness	61.0066	35.8041	9.6979	180.0895
Democracy	5.742649	3.834012	0.00	10.00
Lag of democracy	5.021545	4.167344	0.00	10.00
Executive constraints	4.958977	2.047979	1.00	7.00
Lag of executive constraints	4.512398	2.332962	1.00	7.00
Government consumption	2.195023	0.439004	1.056177	3.560925
Lag of government consumption	2.192095	0.477742	1.014359	3.694487
Inflation rate	2.298081	1.167341	-1.951826	7.571372
Lag of inflation rate	2.338690	1.193889	-1.459525	8.258299

Table 4.3: List of countries grouped with respect to democracy levels for the coefficient estimates from column 10 of Table 4.1

Negative and Significant			Insignificant
≤ 1	≥ 7.6 & < 9	≥ 9	
Algeria (1980, 1990)	Argentina (2000)	Australia (1980, 1990, 2000)	Argentina (1980, 1990)
Bangladesh (1980)	Bolivia (1990, 2000)	Austria (1980,1990,2000)	Benin (1990, 2000)
Benin (1980)	Botswana (2000)	Belgium (1980, 1990, 2000)	Bangladesh (1990, 2000)
Burundi (1980, 1990)	Brazil (1990, 2000)	Canada (1980, 1990, 2000)	Bolivia (1980)
Cameroon (1980, 1990, 2000)	Chile (1990, 2000)	Costa Rica (1980, 1990, 2000)	Botswana (1980, 1990)
Central African Republic (1980)	Colombia (1980, 1990)	Cyprus (1980, 1990, 2000)	Brazil (1980)
Chile (1980)	Dominican Republic (2000)	Denmark (1980, 1990, 2000)	Burundi (2000)
Cote'd Ivoire (1980, 1990)	Ecuador (1980, 1990)	Finland (1980, 1990, 2000)	Central African Republic (1990, 2000)
Egypt (1980, 1990, 2000)	France (1980)	France (1990, 2000)	Congo Republic (1990)
Gabon (1980, 1990, 2000)	Greece (1980)	Greece (1990, 2000)	Cote'd Ivoire (2000)
Gambia (2000)	Guatemala (2000)	Ireland (1980, 1990, 2000)	Colombia (2000)
Ghana (1980)	India (1980, 1990, 2000)	Italy (1980, 1990, 2000)	Dominican Republic (1980, 1990)
Guyana (1980)	Republic of Korea (2000)	Israel (1980, 1990, 2000)	Ecuador (2000)
Indonesia (1980, 1990)	Lesotho (2000)	Jamaica (1980, 1990, 2000)	Gambia (1980, 1990)
Iran (1980)	Mexico (2000)	Japan (1980, 1990, 2000)	Ghana (1990, 2000)
Kenya (1980, 1900)	Panama (1990, 2000)	Netherlands (1980, 1990, 2000)	Guatemala (1990)
Lesotho (1980)	Paraguay (2000)	New Zealand (1980, 1990, 2000)	Guyana (1990, 2000)
Malawi (1980)	Peru (2000)	Norway (1980, 1990, 2000)	Honduras (1980, 1990, 2000)
Mali (1980)	Philippines (2000)	Portugal (1980, 1990, 2000)	Kenya (2000)
Mauritania (1980, 2000)	Senegal (2000)	Spain (1980, 1990, 2000)	Lesotho (1990)
Morocco (1980, 1990, 2000)	South Africa (1990, 2000)	Sweden (1980, 1990, 2000)	Malaysia (1980, 2000)
Nicaragua (1980, 1990)	Thailand (1990)	United Kingdom (1980, 1990, 2000)	Malawi (1990, 2000)
Niger (1980)	Trinidad & Tobago (1990, 2000)	United States (1980, 1990, 2000)	Mali (1990, 2000)
Panama (1980)	Turkey (1990, 2000)	Uruguay (1990, 2000)	Mexico (1980, 1990)
Paraguay (1980)	Venezuela (1980, 1990)		Nepal (1980, 1990, 2000)
Sierre Leone (1980)			Nicaragua (1990)
Swaziland (1980, 1990, 2000)			Niger (1990, 2000)
Syria (1980, 1990, 2000)			Pakistan (1980, 1990, 2000)
Togo (1980, 1990, 2000)			Papua New Guinea (1980, 1990, 2000)
Tunisia (1980, 1990, 2000)			Paraguay (1990)
Zambia (1980)			Republic of Korea (1980, 1990)
Zimbabwe (1990)			Peru (1980, 1990)
			Sierre Leone (1980, 2000)
			South Africa (1980)
			Sri Lanka (1980, 1990, 2000)
			Thailand (2000)
			Turkey (1980)
			Venezuela (2000)
			Zambia (1990, 2000)
			Zimbabwe (1980, 2000)

Chapter 5

Conclusion

In this dissertation, I have studied semiparametric varying coefficient models in macroeconomic context. In the second chapter, I survey the nonparametric estimation methods that includes the least squares kernel estimator and series estimator. I also survey the endogeneity problem in nonparametric models and the identification conditions specific to that models.

In the third chapter, we focus on the nonparametric counterpart of spatial regression models proposing the functional-coefficient spatial Durbin model with nonparametric spatial weights. We find that nonparametric model estimation reveals negative spatial weigths as opposed to the common assumption of parametric spatial models, which require an ad hoc selection of spatial weights with non-negative restriction. Our results also support the conditional β-convergence hypothesis, but the convergence declines gradually as the country's trade openness increases.

In the fourth chapter, I re-investigate the public debt-growth relationship using an endogeneous smooth coefficient model. My results support the view of some literature that provides some evidence of negative growth effect of public debt for the countries with weak institutional quality. I also find some evidence of parameter heterogeneity in the debt-growth link for only low democracy score countries.

References

Maria Abreu and Henri L.F. De Groot. Space and growth: A survey of empirical evidence and methods. 2005.

Robert A. Adams. *Calculus: A Complete Course*. Pearson Education Canada Inc., Toronto, Ontario, 1989.

Alberto Ades and Hak B. Chua. Thy neighbor's curse: Regional instability and economic growth. *Journal of Economic Growth*, 2:279–304, 1997.

Achim Ahrens and Arnab Bhattacharjee. Two-step lasso estimation of the spatial weights matrix. *Econometrics*, 3:128–155, 2015.

Franklin Allen, Elena Carletti, Itay Goldstein, and Agnese Leonello. Moral hazard and government guarantees in the banking industry. *Journal of Financial Regulation*, 1:1–21, 2015.

Donald D. K. Andrews. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Journal of Economic Growth*, 59:307–345, 1991.

Luc Anselin. *Spatial Econometrics: Methods and Models.* Kluver Acedemic Publishers, Dordrecht, 1988.

Luc Anselin and Anil K. Bera. Spatial dependence in linear regression models with an introduction to spatial econometrics. In David E. A. Giles, editor, *Handbook of Applied Economic Statistics*, pages 237–289. Marcel Dekker, Inc., New York, 1998.

Luc Anselin and Raymond J.G.M. Florax. *New Directions in Spatial Econometrics*. Springer, Berlin, 2011.

David Backus, Espen Henriksen, and Kjetil Storesletten. Taxes and the global allocation of capital. *Journal of Monetary Economics*, 55:48–61, 2008.

Badi H. Baltagi, Peter Egger, and Michael Pfaffermayr. Estimating regional trade agreement effects on fdi in an interdependent world. *Journal of Econometrics*, 145:194–208, 2008.

Yong Bao and Aman Ullah. Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics*, 137:396–413, 2007.

Robert J. Barro. Inflation and economic growth. Technical report, NBER Working Paper, 1995.

Robert J. Barro and Xavier I. Salai-Martin. *Economic Growth*. New York: McGraw-Hill, 1995.

Robert Basile. Regional economic growth in europe: A semiparametric spatial dependence approach. *Papers in Regional Science*, 87:527–544, 2008.

Robert Basile, Mara Durbn, Romn Mnguez, Jose Mara Montero, and Jess Mur. Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control*, 48:229–245, 2014.

Anja Baum, Cristina Checherita-Westphal, and Philipp Rother. Debt and growth: New evidence for the euro area. *Journal of International Money and Finance*, 32:809–821, 2013.

Raquel Bernal and Michael P. Keane. Child care choices and children's cognitive achievement: The case of single mothers. *Journal of Labor Economics*, 29:459–512, 2011.

Arnab Bhattacharjee and Sean Holly. Structural interactions in spatial panels. *Empirical Economics*, 40:69–94, 2011.

Arnab Bhattacharjee and Sean Holly. Understanding interactions in social networks and committees. *Spatial Economic Analysis*, 8:23–53, 2013.

Richard Blundell and James L. Powell. Endogeneity in nonparametric and semiparametric regression models. In Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, editors, *Advances in Economics and Econometrics, Theory and Applications: Eight World Congress of the Econometric Society*, volume II. 2003.

Zongwu Cai and Qi Li. Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory*, 24:1321–1342, 2008.

Zongwu Cai, Mitali Das, Huaiyu Xiong, and Xizhi Wu. Functional coefficient instrumental variables models. *Journal of Econometrics*, 133:207–241, 2006.

Zongwu Cai, Jingping Gu, and Qi Li. Some recent developments on nonparametric econometrics. In Qi Li and Jeffrey S. Racine, editors, *Nonparametric Econometric Methods (Advances in Econometrics, Volume 25)*, pages 495–549. Emerald Group Publishing Limited, 2009.

Mehmet Caner, Thomas Grennes, and Fritzi Koehler-Geib. Finding the tipping point-when sovereign debt turns bad. Technical report, The World Bank, 2010.

R. Cao-Abad. Rate of convergence for the wild bootstrap in nonparametric regression. *The Annals of Statistics*, 19:2226–2231, 1991.

Anne C. Case, Harvey S. Rosen, and Jr. James R. Hines. Budget spillovers and fiscal policy interdependence. *Journal of Public Economics*, 52:285–307, 1993.

George Caselle and Roger L. Berger. *Statistical Inference*. Duxbury Press, 2002.

Aurlie Cassette, Jrme Creel, Etienne Farvaque, and Sonia Paty. Governments under influence: Country interactions in discretionary fiscal policy. *Economic Modelling*, 30: 79–89, 2013.

Stephen G. Cecchetti, M. S. Mohanty, and Fabrizio Zampolli. The future of public debt: Prospects and implications. Technical report, Bank for International Settlements, 2010.

Stephen G. Cecchetti, M. S. Mohanty, and Fabrizio Zampolli. The real effects of debts. Technical report, Bank for International Settlements, 2011.

Cristina Checherita-Westphal and Philipp Rother. The impact of high government debt on economic growth and its channels: An empirical investigation for the euro area. *European Economic Review*, 56:1392–1405, 2012.

Xiaohong Chen. Large sample sieve estimation of semi-nonparametric models. In James J. Heckman and Edward E. Leamer, editors, *Handbook of Econometrics*, volume 6B, pages 5549–5632. Springer, New York, 2007.

Xiaohong Chen and Timothy M. Christensen. Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics*, 188:447–465, 2015.

Xiaohong Chen and Yin J. J. Qiu. Methods for nonparametric and semiparametric regressions with endogeneity: A gentle guide. Technical report, Yale University, 2016. Cowles Foundation for Research in Economics.

Chih-Kang Chu and James S. Marron. Choosing a kernel regression estimator. *Statistical Science*, 6:404–419, 1991.

William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836, 1979.

William S. Cleveland, Eric Grosse, and William M. Shyu. Local regression models. In John M. Chambers and Trevor J. Hastie, editors, *Statistical Models in S*, pages 309–376. Wadsworth & Brooks, California, 1992.

Andrew D. Cliff and Keith Ord. *Spatial Processes: Models & Applications*. London: Pion, 1981.

Timothy G. Conley and Ethan Ligon. Economic distance and cross-country spillovers. *Journal of Economic Growth*, 7:157–187, 2002.

Mitali Das. Instrumental variables estimators for nonparametric models with discrete endogenous variables. *Journal of Econometrics*, 124:335–361, 2005.

James Davidson. *Econometric Theory*. Blackwell Publishers, 2004.

Robert M. de Jong. A note on “convergence rates and asymptotic normality for series estimators”: Uniform convergence rates. *Journal of Econometrics*, 111:1–9, 2002.

Michael S. Delgado, Nadine McCloud, and Subal C. Kumbhakar. A generalized empirical model of corruption, foreign direct investment, and growth. *Journal of Macroeconomics*, 42:298–316, 2014.

Michael S. Delgado, Deniz Ozabaci, Yiguo Sun, and Subal C. Kumbhakar. Smooth coefficient models with endogenous environmental variables. Technical report, Purdue University, 2015.

J. Bradford DeLong and Lawrence H. Summers. Equipment investment and economic growth. *The Quarterly Journal of Economics*, 106:445–502, 1991.

Phoebus J. Dhrymes. *Topics in Advanced Econometrics*. Springer-Verlag Inc., New York, 1989.

Steven N. Durlauf, Andros Kourtellos, and Artur Minkin. The local solow growth model. *European Economic Review*, 45:928–940, 2001.

Markus Eberhardt and Andrea F. Presbitero. Public debt and growth: Heterogeneity and non-linearity. *Journal of International Economics*, 97:45–58, 2015.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, Inc., London, 1993.

Balázs Égert. Public debt, economic growth and nonlinear effects: Myth or reality? *Journal of Macroeconomics*, 43:226–238, 2015.

Douglas W. Elmendorf and N. Gregory Mankiw. Government debt. In John B. Taylor and Michael Woodford, editors, *Handbook of Macroeconomics*. North Holland, 1999.

Cem Ertur and Wilfried Koch. Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics*, 22:1033–1062, 2007.

Cem Ertur and Wilfried Koch. A contribution to the theory and empirics of schumpeterian growth with worldwide interactions. *Journal of Economic Growth*, 16:215–255, 2011.

Cem Ertur, Julie Le Gallo, and Catherine Baumont. The european regional convergence process, 1980-1995: Do spatial regimes and spatial dependence matter? *International Regional Science Review*, 29:3–34, 2006.

Randall L. Eubank. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, Inc., 1988.

Jianqing Fan. Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87:998–1004, 1992.

Jianqing Fan. Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, 21:196–216, 1993.

Jianqing Fan and Irene Gijbels. Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, 20:2008–2036, 1992.

Jianqing Fan and Irene Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman and Hall/CRC, 1996.

Jianqing Fan, Tao Huang, and Runze Li. Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, 102: 632–641, 2007.

Lung fei Lee. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22:307–335, 2003.

Lung fei Lee. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72:1899–1925, 2004.

Lung fei Lee. Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137:489–514, 2007a.

Lung fei Lee. Identification and estimation of econometrics models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140:333–374, 2007b.

Bruce E. Hansen. Nonparametric sieve regression: Least squares, averaging least squares, and cross-validation. In Jeffrey S. Racine, Liangjun Su, and Aman Ullah, editors, *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, pages 1–42. Oxford University Press, 2014.

Wolfgang Härdle and Adrian W. Bowman. Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, 83:102–110, 1988.

Wolfgang Härdle and Enno Mammen. Bootstrap methods in nonparametric regression. In George G. Roussas, editor, *Nonparametric Functional Estimation and Related Topics*, pages 111–123. Kluwer Academic Publishers, Netherlands, 1991.

Wolfgang Härdle and Enno Mammen. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21:1926–1947, 1993.

Wolfgang Härdle and J. S. Marron. Bootstrap simultaneous error bars for nonparametric regression. *The Annals of Statistics*, 19:778–796, 1991.

Trevor J. Hastie and Robert Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society*, 55:757–796, 1993.

Tristen Hayfield and Jeffrey S. Racine. Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27:1–32, 2008.

Daniel J. Henderson and Aman Ullah. A nonparametric random effects estimator. *Economics Letters*, 88:403–407, 2005.

Daniel J. Henderson, Raymond J. Carroll, and Qi Li. Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, 144:257–275, 2008.

Daniel J. Henderson, Chris Papageorgiou, and Christopher F. Parmeter. Growth empirics without parameters. *The Economic Journal*, 122:125–154, 2011.

Daniel J. Henderson, Subal C. Kumbhakar, and Christopher F. Parmeter. A simple method to visualize results in nonlinear regression models. *Economics Letters*, 117:578–581, 2012.

Thomas Herndon, Michael Ash, and Robert Pollin. Does high public debt consistently stifle economic growth? a critique of reinhart and rogoff. *Cambridge Journal of Economics*, 38:257–279, 2014.

R. Carter Hill, William E. Griffiths, and Guay C. Lim. *Priciples of Econometrics*. John Wiley & Sons, Inc., 2008.

Joel L. Horowitz. Bootstrap methods in econometrics: Theory and numerical performance. In David M. Kreps and Kenneth F. Wallis, editors, *Advances in Economics and Econometrics: Theory and Applications*, pages 188–222. Cambridge University Press, 1997.

Joel L. Horowitz and Sokbae Lee. Semiparametric methods in applied econometrics: Do the models fit the data? *Statistical Modelling*, 2:3–22, 2002.

Joao Tovar Jalles. The impact of democracy and corruption on the debt-growth relationship in developing countries. *Journal of Economic Development*, 36:41–72, 2011.

Harry H. Kelejian and Gianfranco Piras. Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics*, 46:140–149, 2014.

Harry H. Kelejian and Ingmar R. Prucha. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17:99–121, 1998.

Elena Ketteni, Theofanis P. Mamuneas, and Thanasis Stengos. Nonlinearities in economic growth: A semiparametric approach applied to information technology data. *Journal of Macroeconomics*, 29:555–568, 2007.

Mustafa Koroglu and Yiguo Sun. Functional-coefficient spatial durbin models with nonparametric spatial weights: An application to economic growth. *Econometrics*, 4:1–16, 2016.

Andros Kourtellos. Modeling parameter heterogeneity in cross-country regression models. In Olivier de La Grandville, editor, *Economic Growth and Development (Frontiers of Economics and Globalization)*, pages 579–604. Emerald Group Publishing Limited, Bingley, UK, 2011.

Andros Kourtellos, Thanasis Stengos, and Chih Ming Tan. The effect of public debt on growth in multiple regimes. *Journal of Macroeconomics*, 38:35–43, 2013.

Andros Kourtellos, Thanasis Stengos, and Chih Ming Tan. Structural thresold regression. *Econometric Theory*, 32:827–860, 2016.

Subal C. Kumbhakar and Kai Sun. Estimation of tfp growth: A semiparametric smooth coefficient approach. *Empirical Economics*, 43:1–24, 2012.

James LeSage. What regional scientists need to know about spatial econometrics. *The Review of Regional Studies*, 44:13–32, 2014.

James LeSage and R. Kelley Pace. *Introduction to Spatial Econometrics*. Chapman and Hall/CRC, 2009.

James LeSage and R. Kelley Pace. The biggest myth in spatial econometrics. *Econometrics*, 2:217–249, 2014.

Qi Li and Jeffrey S. Racine. *Nonparametric Econometrics*. Princeton University Press, Princeton, 2007.

Qi Li, Dong Li Cliff J. Huang, and Tsu-Tan Fu. Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics*, 20:412–422, 2002a.

Qi Li, Cliff J. Huang, Dong Li, and Tsu-Tan Fu. Semiparametric smooth coefficient models. *Journal of Business and Economic Statistics*, 20:412–422, 2002b.

Xihong Lin and Raymond J. Carroll. Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, 95:520–534, 2000.

Zhongjian Lin, Qi Li, and Yiguo Sun. A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *Journal of Econometrics*, 178:167–179, 2014.

Zhenjuan Liu and Thanasis Stengos. Non-linearities in cross-country growth regressions: A semiparametric approach. *Journal of Applied Econometrics*, 14:527–538, 1999.

Rudolph A. Lorentz. *Approximation of Functions*. Chelsea Publishing Company, New York, N. Y., 1986.

Enno Mammen. Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21:255–285, 1993.

Theofanis P. Mamuneas, Andreas Savvides, and Thansis Stengos. Economic development and the return to human capital: A smooth coefficient semiparametric approach. *Journal of Applied Econometrics*, 21:111–132, 2006.

N. Gregory Mankiw, David Romer, and David N. Weil. A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107:407–437, 1992.

Alexandru Minea and Antoine Parent. Is high public debt always harmful to economic growth? reinhart and rogoff and some complex nonlinearities. Technical report, CERDI, Etudes et Documents, 2012.

Ramon Moreno and Bharat Trehan. Location and the growth of nations. *Journal of Economic Growth*, 2:399–418, 1997.

James C. Murdoch, Morteza Rahmatian, and Mark A. Thayer. A spatially autoregressive median voter model of recreation expenditures. *Public Finance Quarterly*, 21:334–350, 1993.

Whitney K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79:147–168, 1997.

Whitney K. Newey and James L. Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71:1565–1578, 2003.

Whitney K. Newey, James L. Powell, and Francis Vella. Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67:565–603, 1999.

Keith Ord. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70:120–126, 1975.

Deniz Ozabaci, Daniel J. Henderson, and Liangjun Su. Additive nonparametric regression in the presence of endogenous regressors. *Journal of Business & Economic Statistics*, 32:555–575, 2014.

Ugo Panizza and Andrea F. Presbitero. Public debt and economic growth: Is there a causal effect. *Journal of Macroeconomics*, 41:21–41, 2014.

Joris Pinkse. Nonparametric two-step regression estimation when regressors and error are dependent. *The Canadian Journal of Statistics*, 28:289–300, 2000.

Joris Pinkse, Margaret E. Slade, and Craig Brett. Spatial price competition: A semiparametric approach. *Econometrica*, 70:1111–1153, 2002.

Annie Qu and Runze Li. Quadratic inference functions for varying coefficient models with longitudinal data. *Biometrics*, 62:379–391, 2006.

Xi Qu and Lung fei Lee. Estimating a spatial autoregressive model with an endogeneous spatial weight matrix. *Journal of Econometrics*, 184:209–232, 2015.

Jeffrey S. Racine and Christopher F. Parmeter. Data-driven model evaluation: A test for revealed performance. In Aman Ullah, Jeffrey S. Racine, and Liangjun Su, editors, *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, pages 308–345. Oxford University Press, New York, 2013.

Carmen M. Reinhart and Kenneth S. Rogoff. Growth in a time of debt. *NBER*, 2010.

Carmen M. Reinhart, Vincent R. Reinhart, and Kenneth S. Rogoff. Public debt overhangs: Advanced-economy episodes since 1800. *Journal of Economic Perspectives*, 26:69–86, 2012.

Charles S. Roehrig. Conditions for identification in nonparametric and parametric models. *Econometrica*, 56:433–447, 1988.

Andreas F. Ruckstuhl, Alan H. Welsh, and Raymond J. Carroll. Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statistica Sinica*, 10:51–71, 2000.

David Ruppert and Matt Wand. Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22:1346–1370, 1994.

David W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, Inc., 1992.

George A. F. Seber. *A Matrix Handbook for Statisticians*. John Wiley & Sons, Inc., 2008.

Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1992.

Thanasis Stengos, Brennan S. Thompson, and Ximing Wu. The evolution of the conditional joint distribution of life expectancy and per capita income growth. In Qi Li and Jeffrey S. Racine, editors, *Nonparametric Econometric Methods (Advances in Econometrics)*, volume 25, pages 171–191. Emerald Group Publishing limited, 2009.

Thomas M. Stoker. Lectures on semiparametric econometrics. Technical report, CORE Lecture Series, 1992.

Charles J. Stone. Consistent nonparametric regression. *The Annals of Statistics*, 5:595–620, 1977.

Charles J. Stone. Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, 8:1348–1360, 1980.

Charles J. Stone. Optimal rates of convergence for nonparametric regression. *The Annals of Statistics*, 10:1040–1053, 1982.

Liangjun Su. Semiparametric gmm estimation of spatial autoregressive models. *Journal of Econometrics*, 167:543–560, 2012.

Liangjun Su and Aman Ullah. Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters*, 92:75–81, 2006.

Liangjun Su and Aman Ullah. More efficient estimation of nonparametric panel data models with random effects. *Economics Letters*, 96:375–380, 2007.

Yiguo Sun. A consistent nonparametric equality test of conditional quantile functions. *Econometric Theory*, 22:614–632, 2006.

Yiguo Sun. Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *Journal of Econometrics*, 195:134–153, 2016.

Yiguo Sun, Raymond J. Carroll, and Dingding Li. Semiparametric estimation of fixed-effects panel data varying coefficient models. In Qi Li and Jeffrey S. Racine, editors, *Nonparametric Econometric Methods (Advances in Econometrics, Volume 25)*, pages 101–129. Emerald Group Publishing Limited., 2009.

Yiguo Sun, Yu Yvette Zhang, and Qi Li. Nonparametric panel data regression models. In Badi H. Baltagi, editor, *The Oxford Handbook of Panel Data*, pages 285–324. Oxford University Press, 2014.

Aman Ullah and Nilanjana Roy. Nonparametric and semiparametric econometrics of panel data. In Aman Ullah and David E. A. Giles, editors, *Handbook of Applied Economic Statistics*, pages 579–604. Marcel Dekker, Inc., New York, 1998.

Andrea Vaona and Stefano Schiavo. Nonparametric and semiparametric evidence on the long-run effects of inflation on growth. *Economics Letters*, 94:452–458, 2007.

Naisyin Wang. Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, 90:43–52, 2003.

Alan H. Welsh, Xihong Lin, and Raymond J. Carroll. Marginal longitudinal nonparametric regression: Locality and efficiency of spline and kernel methods. *Journal of the American Statistical Association*, 97:484–493, 2002.

Jaejoon Woo and Manmohan S. Kumar. Public debt and growth. *Economica*, 82:705–739, 2015.

Alan Wright and Kari Grenade. Determining optimal public debt and debt-growth dynamics in the caribbean. *Research in Applied Economics*, 6:87–115, 2014.

C. F. J. Wu. Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14:1261–1295, 1986.

Weixin Yao and Runze Li. New local estimation procedure for a non-parametric regression for longitudinal data. *Journal of the Royal Statistical Society*, (Series B) 75:123–138, 2013.

Adonis Yatchew. Nonparametric regression techniques in economics. *Journal of Economic Literature*, 36:669–721, 1998.