

2013

Environmental Kuznets Curve In Water Pollution: A Semiparametric Approach

Mahesh Pandit

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations



Part of the [Agricultural Economics Commons](#)

Recommended Citation

Pandit, Mahesh, "Environmental Kuznets Curve In Water Pollution: A Semiparametric Approach" (2013). *LSU Doctoral Dissertations*. 2647.

https://digitalcommons.lsu.edu/gradschool_dissertations/2647

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

ENVIRONMENTAL KUZNETS CURVE IN WATER POLLUTION: A SEMIPARAMETRIC APPROACH

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Agricultural Economics and Agribusiness

by
Mahesh Pandit
B.S., Tribhuvan University, 2004
M.S., Tribhuvan University, 2006
M.S., Louisiana State University, 2010
M.ApSt., Louisiana State University, 2013
August 2013

To my late mother

Ambika Devi Pandit

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to my major professor Dr. Krishna P. Paudel for his invaluable guidance to complete this dissertation. He always inspired and challenged me to write a dissertation that can contribute and further the knowledge in the field of environmental Kuznets curve. My gratitude is also extended to the committee members Drs. Ashok K. Mishra, Jeffrey M Gillespie, and R Carter Hill for their encouragement and kind help. I would also like to thank the Graduate School Dean's representative Dr. Terrence R. Tiersch for his comments and suggestions.

My deepest gratitude goes to my parents (Tek Nidihi Pandit and late Ambika Devi Pandit), and my wife Saraswati Pandit Paudel. My wife has shown great strength and support for this achievement. Without her love and devotion, my study could not have been accomplished. I cannot forget my cutey baby Samip Pandit inspiring me to succeed in this long journey. I like to thank my mother-in-law Bishnu Maya Paudel for helping me to study at the most critical time of my life.

Back in Nepal, my appreciation goes to all my brothers, sisters, sisters-in-laws, nephews and nieces. I especially appreciate the contributions of Drs. Murali Adhikari and Laxmi Paudel who helped and encouraged me throughout and prior to my study at Louisiana State University. My appreciation is extended to the families of Sheshkanta Buba, Mohan Dai, Pradip Dai, Dipak Dai, and to Suniti Bhauju.

I would like to thank Bryan Gottshall and Brian Hilbun for editing my dissertation. I would also like to thank Kelly Hodgson and C.-Y. Cynthia Lin for data, John You for helping with the program code related to Seemingly Unrelated Partial Linear Regression (SUPLR) model, Huizhen Niu for assistance with the GIS part of my dissertation research. Finally, I would like to express special thanks to all the members of the Nepalese community at LSU for their friendship, help, and encouragement.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
ABSTRACT	ix
CHAPTER 1. INTRODUCTION	1
1.1 Theoretical EKC model	2
1.1.1 Static Models	4
1.1.2 Dynamic models	6
1.2 EKC Policy	7
1.3 Objectives	9
CHAPTER 2. A SURVEY OF SEMIPARAMETRIC REGRESSION METH- ODS USED IN ENVIRONMENTAL KUZNETS CURVE ANALYSIS ..	11
2.1 Introduction	11
2.2 Existing EKC model	12
2.3 Recent Advances in Semiparametric Model	14
2.4 Model Consistent Specification Test	16
2.5 Semiparametric Estimation of the EKC	18
2.6 Conclusions	28
CHAPTER 3. ENVIRONMENTAL KUZNETS CURVE: STOCK AND FLOW WATER QUALITY PARAMETERS	30
3.1 Introduction	30
3.2 Pollution Sources	32
3.2.1 Nitrogen	32
3.2.2 Phosphorous	32
3.2.3 Dissolved Oxygen	32
3.2.4 Mercury	33
3.3 EKC Literature and Water Quality	33
3.4 Methods	36
3.4.1 Parametric Model	37
3.4.2 Semiparametric Model	38
3.4.3 Model Specification Test	38
3.5 Simulation Study	40
3.6 Data	50
3.7 Results	54
3.7.1 Parametric Model	54
3.7.2 Semiparametric Model	54
3.7.3 Model Specification Test	56
3.7.4 Nitrogen	56
3.7.5 Phosphorous	57
3.7.6 Dissolved oxygen	58

3.7.7	Mercury	59
3.8	Conclusions	60
CHAPTER 4. ENVIRONMENTAL KUZNETS CURVE FOR WATER QUALITY PARAMETERS AT THE GLOBAL LEVEL: SEMIPARAMETRIC AND NONPARAMETRIC APPROACHES		63
4.1	Introduction	63
4.2	Methods	66
4.2.1	Parametric Methods	67
4.2.2	Nonparametric Methods	67
4.2.3	Semiparametric Methods	68
4.2.4	Model Specification Test	69
4.3	Data	70
4.4	Results	72
4.4.1	Model Specification Test	72
4.4.2	Parametric Results	74
4.4.3	Nonparametric Results	76
4.4.4	Semiparametric Results	76
4.4.5	Fecal Coliform	81
4.4.6	Mercury	85
4.4.7	Arsenic	85
4.4.8	Lead	86
4.4.9	Dissolved Oxygen	86
4.4.10	COD	86
4.4.11	BOD	87
4.4.12	Total Coliform	87
4.5	Conclusions	87
CHAPTER 5. CONCLUSIONS		89
REFERENCES		101
APPENDIX A. SEEMINGLY UNRELATED PARTIAL LINEAR REGRESSION ESTIMATION PROCEDURE		102
APPENDIX B. WATER QUALITY PARAMETERS		104
B.1	Nitrogen	104
B.2	Phosphorous	104
B.3	Dissolved Oxygen	104
B.4	Chemical Oxygen Demand	105
B.5	Biological Oxygen Demand	105
B.6	Mercury	105
B.7	Cadmium	106
B.8	Lead	106
B.9	Arsenic	106

B.10 Fecal Coliform	107
B.11 Total Coliform	107
APPENDIX C. R CODE FOR CHAPTER 2	108
C.1 Simulation	108
C.2 Empirical Simulation	113
C.3 Empirical Estimation	119
APPENDIX D. R CODE FOR CHAPTER 3	127
VITA	137

LIST OF TABLES

2.1	Smoothing Approach and Literature	14
2.2	Existing Published Studies That Have Used Semiparametric Techniques in Environmental Kuznets Curve Estimation	20
3.1	Simulation Results for Parameters Estimated Parametrically in a Semiparametric Model	42
3.2	Simulation Results of Estimated Error Variance of SUPLR Model	43
3.3	Empirical Simulation Results for Parameters Estimated Parametrically in a Semiparametric Model	48
3.4	Empirical Simulation Results of Estimated Error Variance of Model	48
3.5	Summary Statistics of Variables	52
3.6	Test for Presence of Fixed Effect	54
3.7	Model Specification Test between Parametric and Semiparametric Model	57
3.8	Seemingly Unrelated Regression Results	58
4.1	EKC Literature by Types of Pollutions	64
4.2	Summary Statistics of Variables	73
4.3	Model Specification Test for Model 1	75
4.4	Model Specification Test for Model 2	75
4.5	Cross Validation Score for Model 1	76
4.6	Cross Validation Score for Model 2	76
4.7	Estimated Coefficient from Parametric Model 1	77
4.8	Estimated Coefficient from Parametric Model 2	81

LIST OF FIGURES

1.1	Environmental Kuznets Curve	2
3.1	Partial Regression Plots of y , $n = 100$	44
3.2	Partial Regression Plots of y , $n = 200$	45
3.3	Partial regression plots of y , $n = 500$	46
3.4	Partial Regression Plots of y from Empirical Simulation	49
3.5	Louisiana Parishes and Water Sub-segments	51
3.6	Partial plot of per capita income from semiparametric model	55
4.1	Partial Regression Plot from Nonparametric Model 1.	78
4.2	Partial Regression Plot from Nonparametric Model 2.	80
4.3	Partial Regression Plot from Semiparametric Model 1.	82
4.4	Partial Regression Plot from Semiparametric Model 2.	84

ABSTRACT

The relationship between pollution and per capita income generally appears as an inverted U-shaped curve. This inverted U-shaped curve is known as the environmental Kuznets curve (EKC). The shape of the curve, however, is very sensitive to the data, location and pollutant considered in the analysis. Since the early 1990s, there has been an exponential growth in the number of empirical studies in this field, but many refute the inverted U-shaped nature of the curve for pollutants across different time periods and geographical regions. This has generated an increased interest in developing a more flexible functional form for model specification and estimation.

Our observation is that existing EKC studies have not fully utilized the advances in semiparametric and nonparametric panel econometrics. In order to identify an appropriate functional form between environmental quality and economic growth, we surveyed recent developments in econometrics specifically related to nonparametric and semiparametric models for panel data. We proposed a seemingly unrelated partial linear model (SUPLR) to address potential correlation between pollutants. Simulation study shows that the SUPLR model performs well for our data set. We examined the EKC relationship between water quality indicators (nitrogen, phosphorous, dissolved oxygen and mercury) and income at the watershed level, using environmental quality data from 53 parishes in Louisiana. Additionally, we explored the income-pollution relationship using Global Environment Monitoring System (GEMS) data from sixty eight countries. We found that the relationship followed an inverted U-shaped curve for nitrogen and dissolved oxygen and a cubic shape for mercury. At the global level, an inverted U-shaped relationship is found for three pollutants (dissolved oxygen, fecal coliform and coliform), a cubic relationship is found for three pollutants (mercury, chemical oxygen demand and biological oxygen demand) and an L-shaped relationship is observed for two pollutants (arsenic and lead). Model specification tests suggest that a semiparametric model is better specified to study the income-pollution relationship.

CHAPTER 1.

INTRODUCTION

The environmental Kuznets curve (EKC) hypothesis states that the relationship between pollution and per capita income generally appears as an inverted U-shaped curve as shown in Figure (1.1). The notion presented by the EKC hypothesis is that pollution grows rapidly in the early stages of a country's industrialization because high priority is given to increased production, and people are more interested in income than with environmental concerns (i.e., green production practices, reducing pollutants in industry, etc.). Additionally, at the height of industrialization, environmental quality is considered a luxury good. As a country advances beyond the industrialization phase into an economy that is primarily service dominated, people's demand for environmental quality increases. Further, at that stage people are willing to pay for better water/air quality. EKC studies have been conducted on air pollution, water pollution, deforestation, toxic substances, waste, and energy-related variables. Empirical studies have either refuted or failed to reject the EKC hypothesis.

The Environmental Kuznets curve (EKC) is an empirical phenomenon showing how some pollutants increase and then decrease with rising per capita income. Original thought of the EKC was coined by [Grossman and Krueger \(1991, 1995\)](#) and [Shafik and Bandyopadhyay \(1992\)](#) from the study of economic growth and environmental quality during the North American Free Trade Agreement debate of the 1990s. They connected their findings with the production economy to show the existence of the EKC hypothesis. They stated that the EKC is the result of scale, technique and composition effects. An increase in current production leads to an increase in pollution which is subsequently called the *scale effect*. Increased adoption of more *efficient technologies* decreases pollution. An increase in economic growth shifts the economy from a manufacturing base to one that is more service oriented in its scope. This is commonly referred to as the *composition effect*. Hence, if technique and

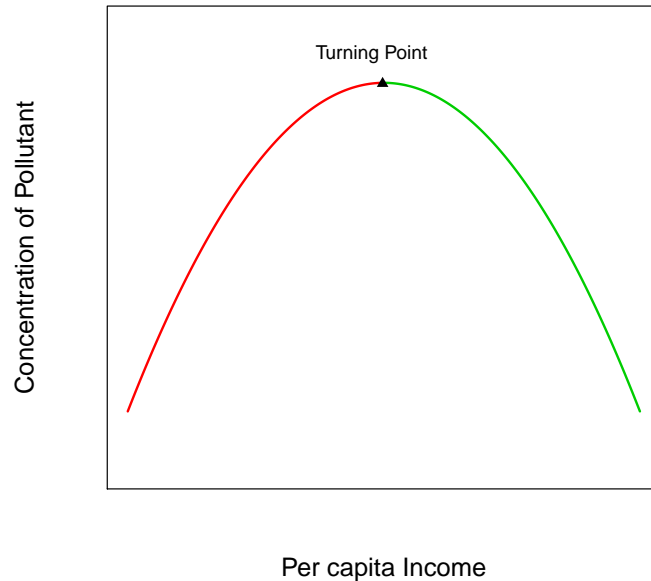


Figure 1.1. Environmental Kuznets Curve

composition effects are higher than the scale effect over a particular time period, the EKC relationship is established.

The shape of the curve, however, is very sensitive to the data period, location and pollutant considered in the analysis (Harbaugh, Levinson, and Wilson, 2002). Since the early 1990s, a number of empirical studies have been conducted, but many of these studies have refuted the inverted U-shape of the curve, indicating that a more flexible functional form is required to examine the EKC hypothesis. The debate on EKC hypothesis has settled to some extent but the research direction is moving toward developing a theoretical model to understand the EKC hypothesis and estimating the empirical model using a flexible model specification.

1.1 Theoretical EKC model

There are two strands of theoretical literature in EKC: one draws its theoretical underpinnings from growth theory (dynamic optimization model) while the other bases its rationale on

ideas drawn from static utility maximization theory. We provide a summary of representative articles covering both strands of literature as well as demonstrate the essence of the two approaches.

Many researchers have proposed different theories behind the EKC hypothesis. [Lopez \(1994\)](#) described the inverted U-shaped relation as a production function. He showed that as the substitution elasticity between conventional input and pollution falls, then the relative curvature of income in the utility function falls and the inverted U-shaped relationship gets established. This suggests that firms pay an increasing price for pollution while it is less costly to reduce pollution by changing the production technology to an environmentally-friendly one. On the other hand, non-homothetic¹ preference implies that consumers are willing to give up additional consumption in order to receive a better environmental quality. [McConnell \(1997\)](#) studies the role of income elasticity of demand for environmental quality and came to the conclusion that it is not the income elasticity of demand for environmental quality that shapes EKC.

[John and Pecchenino \(1994\)](#) used an overlapping generation model and provided a theoretical explanation for the inverted U-shaped correlation between environmental quality and income. They concluded that “the relationship between growth and the quality of the environment is complex.” [Andreoni and Levinson \(2001\)](#) used a Cobb-Douglas utility function to explain the income and pollution relationship. They proposed that utility depends on consumption and pollution, and pollution depends on consumption levels and pollution control efforts. They suggested that an inverted U-shaped EKC relationship occurs if there are increasing returns to scale in terms of the pollution control effort. This case is likely due to many factors such as population growth and technological changes.

[Kelly \(2003\)](#) developed an EKC from a stock externalities perspective. According to this author, the marginal benefit of pollution control and the marginal cost of pollution control

¹A monotone preference relation \succeq on $X = R_+^L$ is homothetic if all indifference sets are related by proportional expansion along rays; that is if $x \sim y$ then $\alpha x \sim \alpha y$ for any $\alpha \geq 0$ ([Mas-Colell, Whinston, and Green, 1995](#)).

rise with income over the growth path. If the marginal benefit rises faster than the marginal costs, the emission-income relationship has a negative slope for a given level of income and vice-versa. Recently, [Brock and Taylor \(2010\)](#) extended the Solow growth model in the EKC framework, also known as the Green Solow Model. Due to diminishing returns, development begins with rapid economic growth, and emissions rise with the output growth, but fall with ongoing technological progress. At first, emission of pollutants overwhelms the impact of technological progress and emission levels rise. As countries mature and approach a balanced growth path, the impact of economic growth is overwhelmed by the impact of technological progress and emission levels decline. So diminishing returns and technological progress are responsible for generating the inverted U-shaped EKC.

1.1.1 Static Models

Following [Andreoni and Levinson \(2001\)](#), let us consider that an individual maximizes utility from consumption of private good denoted C , and pollution P . The utility function is given as below.

$$U = U(C, P) \tag{1.1}$$

Where $\frac{\partial U}{\partial C} = U_C > 0$ and $\frac{\partial U}{\partial P} = U_P < 0$. Since consumption, ‘ C ’ designates normal goods and ‘ P ’ designates non-normal goods, U is quasi-concave in C and $-P$. Pollution enters in as a byproduct from the consumption of goods, and individuals allocate resources to reduce pollution or prevent it from happening. Let us denote the resources spent cleaning the environment by E . Hence, pollution is a function of consumption and environmental effort.

$$P = P(C, E) \tag{1.2}$$

where $P_c > 0$ and $P_E < 0$. Further, suppose that M is total income available to spend on either C or E . Hence the resource constraint is given by

$$C + E = M \quad (1.3)$$

To illustrate, let us consider simple utility and pollution functions as given below,

$$U = C - zP \quad z > 0 \quad (1.4)$$

$$P = C - C^\alpha E^\beta \quad \alpha, \beta > 0 \quad (1.5)$$

where z represents the marginal disutility from pollution. The second term ($C^\alpha E^\beta$) in equation (1.5) represents ‘abatement’. Maximizing the utility function (1.4), subject to the constraint (1.3) yields the optimal solution for C and P as follows.

$$C^* = \frac{\alpha}{\alpha + \beta} M \quad \text{and} \quad E^* = \frac{\beta}{\alpha + \beta} M \quad (1.6)$$

Substituting the optimal value of C^* and E^* from equation (1.6) into equation (1.4), the optimal pollution is given as

$$P^*(M) = \frac{\alpha}{\alpha + \beta} M - \left(\frac{\alpha}{\alpha + \beta} \right) M^{\alpha + \beta} \quad (1.7)$$

Differentiating this equation with respect to M yields

$$\frac{\partial P^*}{\partial M} = \frac{\alpha}{\alpha + \beta} - (\alpha + \beta) \left(\frac{\alpha}{\alpha + \beta} \right) M^{\alpha + \beta - 1} \quad (1.8)$$

The equation (1.8) indicates that when $\alpha + \beta > 1$, the pollution level P^* follows an inverted U-shape curve with respect to income. This is the condition for increasing returns to scale. [Kijima, Nishide, and Ohyama \(2010\)](#) explains this relation as: ‘For low income (M) the consumption level is also low, and the increasing return of abatement indicates that the effect from the abatement effort has little impact on environmental quality. At this condition, the representative agent does not want to spend much money on abatement, and so the pollution

level rises with an increase in income. In contrast, for a sufficiently high level of income, a high level of consumption causes the agent much disutility from pollution. In fact, the impact of abatement on utility value is higher due to the increasing return, and the agent optimally spends more resource on abatement. Thus pollution levels decreases with higher level of income. Hence combining these two conditions implies the existence of EKC.'

1.1.2 Dynamic models

John and Pecchenino (1994) developed an overlapping generation models with two periods of time. According to John and Pecchenino (1994) a person allocates their income between consumption and abatement efforts for two periods of time. Let w_t represent the wage of a person for generating t . Utility at current period t is a function of consumption and environmental quality of the later period and is given as

$$U_t = U(c_{t+1}, E_{t+1}) \quad (1.9)$$

where, c_t = consumption at period t , E_t = environmental quality at period t . A higher value of E represents better environmental quality. The environmental quality holds the following dynamics equation.

$$E_{t+1} - E_t = -bE_t - \beta c_t + \gamma m_t, \quad (1.10)$$

where m_t is the investment in environmental maintenance and improvement, and b , β , and γ are positive constant. Let the production function is given as $Y_t = F(K_t, N_t)$, where Y is output, K is capital stock, and N is the labor. Assuming first-order homogenous production function, the output per capita can be expressed as $y = f(k_t)$. Here $k_t = K_t/N_t$.

John and Pecchenino (1994) shown the equations for dynamic equilibrium path.

$$r_t = f'(k_t) - \delta = r(k_t) \quad (1.11)$$

$$w_t = f(k_t) - k_t, \quad f'(k_t) = w_t(k_t), \quad (1.12)$$

$$U_1(c_{t+1}, E_{t+1})(1 + r_t + 1) - \gamma U_2(C_{t+1}, E_{t+1}) = 0 \quad \text{or} \quad m_t = 0, \quad (1.13)$$

$$k_{t+1} = s_t \quad (1.14)$$

where U_i represents the partial derivative with respect to the i^{th} argument, δ is depreciation rate, r_t is interest rate at period t , and s_t is the saving amount of generation t .

Assume that economy starts with a little capital only. In that scenario, firms do not have enough capital to spend on environmental pollution abatement, i.e. $m_t = 0$. Hence, environmental quality deteriorates initially. After a certain period of time, as capital stock accumulates and the income rises, firms are more willing to pay for enhancing environmental quality and investing in more environmentally friendly production processes. Due to this phenomenon, the income and pollution relationship exhibits an inverted U-shape curve.

1.2 EKC Policy

The inverted U-shaped relationship between economic growth and environmental quality reveals that sufficient economic growth is one possible solution in the abatement of environmental pollution. This is an important motivation that leads us to examine the EKC hypothesis. If this is true, we are led to ask the question do environmental problems reduce automatically with the rise in per capita income? Alternatively speaking, do people start caring about environment once they become richer? Empirical studies have shown that there is no unique answer to this question, because the results are very sensitive to the particular pollutant considered, the time period, and geographical location to name a few. Thus, economic growth does not control environmental quality itself automatically (see more details in [Vincent, 1997](#); [Criado, 2008](#)). This answer leads to another question and that is whether or not if environmental policies are needed in order to improve environmental quality? [Grossman and Krueger \(1995\)](#) and [Dasgupta et al. \(2002\)](#) suggest that

improvement in environmental quality comes through environmental regulations. Effective policy significantly reduce environmental degradation and the environmental cost of growth (Panayotou, 1997). To illustrate, an increase in economic growth changes the preference and environmental regulation that leads to change in production (Tsurumi and Managi, 2010). Environmentally friendly regulations play a significant role in improving environmental quality. Tsurumi and Managi (2010) suggested a tradeoff between economic growth and environmental quality depends on the technique effects. The magnitude of the technique effect is important as to implementing environmental policy, and stringent environmental regulation leads to an improvement in environmental quality. If the technique effect is not sufficient to reduce environmental degradation, environmental regulations are required as to reduce pollution. Developing countries ignore their environmental problem until they are further along their industrial development path and have become wealthier; however, these countries should consider formulating regulations at a less stringent level in the beginning and then ratchet up those regulations as their economy matures (Carson, 2010). Carson (2010) concluded that since environmental regulation and abatement efforts are required to control environmental degradation, an optimal time for abatement and policy should be determined. Further, Stern (2004) suggested that new innovation is needed to be adopted in high income countries before it is adopted in low income countries so as to improve environment quality. However, a ‘one size fits all’ approach for finding a solution for all countries would not work, so heterogenous technologies that are country specific would be needed.

Empirical studies have shown that there exists a cubic shaped relation for some pollutants, implying a chance of further degradation of environmental quality after improvement. We believe that this might be due to the following consequence: as a country becomes wealthier, the demand for industrial products rises. This higher demand subsequently raises the production of goods and, as a result, increases with it the emission of pollutants as a byproduct of the production process. Unless alternative technology is invented, depending upon the extant environmental regulations and the condition of the environment, as per

capita consumption of resources increases so do pollution levels. This is evident with electricity consumption in developed countries and the related by-product of the generation of that electricity, i.e., air pollutants. Returning back to motivation of EKC studies, economic growth might be solution for environmental degradation, but it might not be true for stock pollutant because of irreversibility and catastrophic impact on the environment.

1.3 Objectives

In this dissertation, the focus is on three different objectives related to developing a water pollution EKC using a semiparametric model. The first objective of this dissertation is to survey recent developments on nonparametric and semiparametric methods and their use in EKC literature. We will show potential improvement that can be achieved in EKC estimation using the most recently developed techniques. We provide a tabular summary of estimation methods, data and variables used, test statistics used to compare functional forms of EKC, and major findings.

The second objective of this dissertation is to examine the existence of an EKC for water quality at the local level. We apply a flexible model based on the method suggested in the current literature. The water quality indicators used are nitrogen (N), phosphorous (P), dissolved oxygen (DO)² and Mercury (Hg). Three of these pollutants (N, P, DO) are flow pollutants and one (Hg) is a stock pollutant. We expect that concentrations of nitrogen and phosphorous increase with economic growth. As states/countries become wealthier, concentrations of pollutants will decline after a certain level of income is attained, so we expect an inverted U-shaped relationship for these pollutants. In contrast, due to high pollution, the concentration of oxygen in water will decrease. As the level of dissolved oxygen decreases, it become harder for aquatic animals to get the oxygen they need to survive. Hence, we expect that the DO will decrease (i.e., pollution increases) at first with

²Dissolved oxygen is the amount of oxygen that is present in the water. Low amount of oxygen indicates high pollution levels in water. High concentration of DO in water is good! When we talk about DO being a flow pollutant, we mean not having sufficient amounts of DO in water.

economic growth; however, people are more concerned about water quality above a certain level of income, and concentration of oxygen in water rises (i.e., pollution decreases). Due to this reason, we again expect an inverted U-shaped curve for DO. Mercury is a heavy metal and stock pollutant. Unless a specific plan is developed, the concentration will continue to increase in water bodies.

The third objective of this dissertation is to examine the EKC hypothesis for water pollutants at the global level. Is the behavior of the income-pollutant relationship at the global level consistent with the relationship observed at the local level? This dissertation examines the EKC hypothesis for pollutants at the global level based on data available from the Global Environment Monitoring System (GEMS). This objective focuses on four types of water pollutants: heavy metals (nickel, mercury, arsenic, cadmium, and lead), pathogenic contamination (fecal coliform and total coliform), oxygen regimes (dissolved oxygen (DO), chemical oxygen demand (COD), and biological oxygen demand (BOD)), and nutrients (nitrate).

CHAPTER 2.

A SURVEY OF SEMIPARAMETRIC REGRESSION METHODS USED IN ENVIRONMENTAL KUZNETS CURVE ANALYSIS

2.1 Introduction

This chapter reviews the literature on the effects of economic growth on environmental quality using semiparametric and nonparametric methods. Since the mid-1990s, research has been conducted to examine the existence of the EKC on different types of environmental quality indicators such as air, water, forest and energy consumption. The literature on these subjects is continuously growing with various findings that are inconsistent with the traditional belief that an inverted U-shaped relationship holds for all pollutants.

One of the current and important debates on EKC research is the use of a functional form implemented to examine the environmental quality and economic growth relationship. During the 1990s, parametric models with polynomial specifications were generally used (e.g. [Grossman and Krueger, 1995](#)). A parametric model required distribution assumptions as to estimate relevant model parameters. If the distributional assumption was not valid, the inferences drawn from the wrong model were inconsistent, biased and inefficient. Generally speaking, the true relationship between variables is unknown. From an econometric perspective, a complex model or flexible model is required to extract more information from data, so the use of nonparametric or semiparametric models has begun to emerge in the EKC literature (see [Millimet, List, and Stengos, 2003](#); [Paudel, Zapata, and Susanto, 2005](#); [Paudel and Poudel, 2013](#)).

Many researchers focus only on the effects of economic growth on environmental pollution to examine the EKC hypothesis (e.g. [Grossman and Krueger, 1991, 1995](#)). Other factors such as population density, political freedom, and farmland also play important roles to

determine the concentration of pollutants. If these important variables are omitted in model specification, the results obtained will not be consistent due to omitted variable bias. [Phu \(2003\)](#), [Roy and van Kooten \(2004\)](#), [Paudel, Zapata, and Susanto \(2005\)](#), [Van and Azomahou \(2007\)](#), and [Lin and Liscow \(2013\)](#) used additional variables other than income in their models to partial out the effects of these variables so that they could establish a more accurate relationship between economic growth and environmental quality.

This chapter contributes to the literature primarily for the following reasons: First, we describe how EKC hypotheses are examined in empirical studies. Second, we provide a detailed review of the recent developments in semiparametric econometric methods and how these advances are implemented in the empirical EKC literature. We then discuss existing model specification test statistics and additional variables used in the EKC literature. The details provided here should be beneficial in shaping the direction of future studies on the EKC.

2.2 Existing EKC model

The most general parametric panel model specification used in the EKC literature is a polynomial form equation with two or three degrees for income. According to [Stern \(2004\)](#), the polynomial model in the EKC is specified as follows.

$$P_{it} = \gamma_i + \phi_t + \beta_1 y_{it} + y_{it}^2 \beta_2 + y_{it}^3 \beta_3 + x_{it} \alpha + \epsilon_{it} \quad i = 1, \dots, n \quad t = 1, \dots, T, \quad (2.1)$$

where the first two terms are the intercept parameters of two-way fixed effects for individuals (such as county or state or countries) and times. The intercept parameters control location and time specific factors in the panel data model, respectively. In some cases, researchers have used only a one-way effect model, arguing that country or geographic effects are constant. In that case, either $\gamma_i = 0$ or $\phi_t = 0$. P_{it} represents pollution level for the individual county or watershed i at time t . Pollutants are usually measured in concentration. y_{it} is the measure of

economic growth and is usually measured in per capita income or per capita gross domestic product (GDP). β_1 , β_2 , and β_3 are associated coefficients for y_{it} , y_{it}^2 and y_{it}^3 , respectively. If a quadratic model is used, then β_3 is restricted to zero (i.e. $\beta_3 = 0$). Variable x_{it} includes other factors that affect the pollution emission such as population density, farm crop land, and political freedom; and ϵ_{it} is a contemporaneous error term that can take different structures according to model specification.

By definition, the EKC hypothesis implies that the relationship between income and pollution emissions is nonlinear. Sometimes, it is difficult to parameterize a nonlinear relationship with a parametric specification. In this case, a nonparametric or semiparametric model may be more useful than a parametric model, as the former does not require any distributional assumptions. In the EKC literature, many researchers have employed nonparametric or semiparametric model specifications with an economic growth variable entered as a nonparametric component and other variables entered as parametric components. The model, which contains both parametric and nonparametric components, is a semiparametric model. A semiparametric partially linear regression model ([Robinson, 1988](#); [Millimet, List, and Stengos, 2003](#); [Bertinelli and Strobl, 2005](#)) is specified as

$$P_{it} = \gamma_i + \phi_t + g(y_{it}) + x_{it}\alpha + \epsilon_{it} \quad i = 1, \dots, n \quad t = 1, \dots, T, \quad (2.2)$$

where $g(\cdot)$ is some unknown smooth function. Other parameters are defined as in equation (2.1). The nonparametric component can extract more information from the data about the curvature of the regression at any specific value of y .

In the EKC literature, we found two different approaches to estimate the smoothness of a function. The two approaches are kernel smoothing and spline smoothing, both of which have been extensively used by researchers. Table (2.1) gives examples of EKC literature differentiated by smoothing technique used in semiparametric models. A more flexible smoothing technique is also used in the EKC literature. [Van and Azomahou \(2007\)](#) uses

Table 2.1. Smoothing Approach and Literature

Smoothing Approach	Literature
Kernel smoothing	Millimet, List, and Stengos (2003) , Stern (2004) , Paudel, Zapata, and Susanto (2005) , Poudel, Paudel, and Bhattarai (2009) , and Li (2011)
Spline smoothing	Phu (2003) , Criado (2008) , Criado (2008) , Luzzati and Orsini (2009) , Zanin and Marra (2012) , and Kim (2013)

the smooth coefficient model as proposed by [Li et al. \(2002\)](#). The smooth coefficient model is specified as follows:

$$y_{it} = g(y_{it}) + x_i' \alpha(y_{it}) + \epsilon_{it}. \quad (2.3)$$

where all the representations are the same as above. The semiparametric model (2.2) is nested in this model and can be obtained from a restriction $\alpha(y_{it}) = \alpha$.

2.3 Recent Advances in Semiparametric Model

Given the debate on the functional form used to examine the EKC hypothesis, we were interested to search for recent developments on nonparametric and semiparametric methods that can be used to examine the EKC hypothesis. Semiparametric regression combines parametric and nonparametric regressions, which are found to be better than running only parametric or nonparametric regressions ([Paudel, Zapata, and Susanto, 2005](#); [Pandit, Paudel, and Mishra, 2013](#)). Semiparametric regression relaxes the distribution assumption of a parametric model and reduces the curse of dimensionality associated with a nonparametric method. The semiparametric regression method is used in various subject areas. The two approaches used to smooth variables using nonparametric and semiparametric regression methods are kernel smoothing and spline smoothing of a variable entered as a nonparametric component. Spline is a parametric approach of fitting a nonlinear model, whereas kernel smoothing is a locally weighted average regression method.

The nonparametric and semiparametric statistical methods have been used in economic research since the 1960s, but have only gained widespread use since the early 1990s. Since that time, new development of estimation procedures have been constantly evolving. One of the most used semiparametric models is a semiparametric partial linear model that was developed by [Robinson \(1988\)](#). [Li et al. \(2002\)](#) generalized the model proposed by [Robinson \(1988\)](#) and generated a semiparametric smooth coefficient model using local least squares with a kernel function. This model is more flexible than the partial linear model.

Many variables (such as gender, location, etc.) in economic models are also categorical or binary variables. It is easy to perform statistical analysis if all variables are continuous; however, mixed data containing continuous and categorical variables are tedious to manipulate in a semiparametric regression model compared to a parametric regression model. Many authors have proposed new methodologies to account for mixed variables in the semiparametric model. For example, [Racine and Li \(2004\)](#) proposed a new methodology of nonparametric regression estimation to include both categorical and continuous variables in a semiparametric model. Using kernels along with the cross validation method for smoothing parameters, they showed that the proposed estimator performs much better than the conventional nonparametric estimators in the presence of mixed data. Furthermore, multivariate-based distributions used in economic research is another difficulty in the semiparametric estimation procedure. To account for this phenomenon, [Chen and Fan \(2006\)](#) suggest a Copula-based semiparametric stationary Markov model characterized by a parametric copula and a nonparametric marginal distribution. A Copula serves as a heuristic in constructing a multivariate regression and represents general types of dependence.

In addition to a mixed model as developed by [Racine and Li \(2004\)](#), [Li, Racine, and Wooldridge \(2009\)](#) developed a nonparametric estimation procedure for treatment effects models which can include categorical and continuous variables. They show that their method is capable of performing better than the conventional nonparametric method. Details on the kernel-based estimation procedure for categorical variables can be found in [Racine \(2011\)](#).

Recently, [Ma and Racine \(2013\)](#) and [Nie and Racine \(2012\)](#) also developed a spline-based nonparametric regression model which includes both continuous and categorical variables. In addition to handling a mixed model in the spline based semiparametric regression model, [Ma and Racine \(2013\)](#) proposed estimating using an additive regression spline model.

All of these models can handle categorical variables and require at least one continuous variable. However, [Li, Ouyang, and Racine \(2013\)](#) developed a categorical semiparametric coefficient model that can handle all categorical variables in a nonparametric component in a semiparametric model. We also observed a rapid growth in literature that uses nonparametric and semiparametric models using panel data. Detailed discussion on a semiparametric model using panel data is found in [Ullah and Roy \(1998\)](#) and [Ai and Li \(2008\)](#). [Griffin and Steel \(2010\)](#) proposed a Bayesian fully nonparametric regression estimation procedure from a combination of Bayesian nonparametric density estimation and a nonparametric regression model. Copulas are usually used to fit the multivariate distribution. Most recently, [Qian and Wang \(2012\)](#) developed a semiparametric panel data model using a first differencing method based on the marginal integration of a locally linear smoothed higher-dimensional function.

2.4 Model Consistent Specification Test

Appropriate nonparametric model specification test statistics are necessary to compare nonparametric and semiparametric models. We reviewed model specification tests used to compare parametric, nonparametric and semiparametric models in the EKC literature in this section. In the 1980s, the nonparametric technique for model specification was first suggested by [Ullah \(1985\)](#) and [Robinson \(1988\)](#). Many studies have proposed test statistics to compare nonparametric or semiparametric versus parametric models ([Delgado and Stengos, 1994](#); [Fan and Li, 1996](#); [Zheng, 1996](#); [Hong and White, 1995](#)). All of these test statistics are used in the EKC study. For example the test statistic developed by [Hong and White \(1995\)](#) was used by

Paudel, Zapata, and Susanto (2005). This test statistic is based on the covariance between the residual from the parametric and discrepancy between the parametric and nonparametric fitted values. The decision is made based on the asymptotic normal distribution, so it does not address the non-linearity of the data. Li and Wang (1998) developed test statistics to test a parametric partial linear model against a semiparametric partial linear model. Because this test is based on the wild bootstrap technique, it performs better than the test statistics that depend on the assumption of an asymptotic normal distribution. We observed that this method is fairly common in the EKC literature to compare parametric and nonparametric or semiparametric models (e.g. Millimet, List, and Stengos, 2003; Roy and van Kooten, 2004; Azomahou, Laisney, and Van, 2006; Poudel, Paudel, and Bhattarai, 2009; Phu, 2010).

Semiparametric model estimation techniques such as the kernel method have been used to construct consistent model specification tests. Robinson (1988) tested the suitability of parametric vs. semiparametric regression models using such a process. Similarly, Hardle and Mammen (1993) suggest the use of the wild bootstrap procedure. Further, semiparametric test statistics are also used to check endogeneity of variables by some researchers. Blundell and Duncan (1998) introduced a specification testing procedure for determining the endogeneity of variables by implementing semiparametric methods in an income-consumption relationship using British family expenditure survey data.

Li et al. (2002) introduced a more flexible semiparametric model as well as test statistics to check model specification. The test statistics developed by Li et al. (2002) are used by Van and Azomahou (2007). All of these test statistics mentioned above have a drawback that they do not work when there are categorical variables that entered as a nonparametric component. Hsiao, Li, and Racine (2007) developed new test statistics which overcome this drawback. Using simulation results, they found that the proposed test has a significant advantage over other conventional frequency-based kernel tests. The test statistics developed by Hsiao, Li, and Racine (2007) are used by Paudel and Poudel (2013) in their EKC paper.

2.5 Semiparametric Estimation of the EKC

In this section, we will discuss how semiparametric models have been used in the EKC literature. A summary of journal articles which have used semiparametric models in the EKC study is provided in Table (2.2). The table provides author, year of publication, type of additional variables included in the model other than income, type of parametric and semiparametric model and model specification test used, and their major finding including turning points (TP) if they found the existence of an EKC in their research. Table (2.2) shows that the use of semiparametric method in the EKC literature is increasing. Recently, the use of semiparametric models by researchers has increased. Generally, the parametric models estimated in the EKC are of quadratic and cubic forms.

[Millimet, List, and Stengos \(2003\)](#) used a flexible semiparametric model to study the existence of the EKC. They tested the existence of the EKC for sulfur dioxide (SO_2) and nitrogen oxide (NO_x) emissions from 1929-1994 using a panel data set at the U.S. state-level. They considered a fixed effect cubic model as a parametric model. Spline smoothing¹ and [Robinson \(1988\)](#) partial linear models are used as a semiparametric model. Income is entered as a nonparametric variable in the semiparametric model. As expected, they found the existence of the EKC for sulfur dioxide and nitrogen oxide. They used the [Zheng \(1996\)](#) and [Li and Wang \(1998\)](#) model specification test to compare the results from parametric and semiparametric models. The model specification tests show the semiparametric model performs better than parametric model. This suggests that a semiparametric model is a more flexible model compared to the parametric model.

[Phu \(2003\)](#) used an additive partial linear model developed by [Hastie and Tibshirani \(1990\)](#) which is a spline based semiparametric model. He used data on protected areas in 89 countries to examine the EKC hypothesis on protected areas. In addition to per-capita GDP, he considered other factors such as trade, population density, education and political

¹Although [Millimet, List, and Stengos \(2003\)](#) used spline as a parametric model, spline smoothing is parametric approach of estimating a nonparametric model ([Ruppert, Wand, and Carroll, 2003](#)).

institutions. These variables were parametrically entered in the semiparametric model. [Phu \(2003\)](#) found that there was no existence of an EKC in the protected area. To compare parametric and nonparametric model specifications, he computed gain statistics developed by [Hastie and Tibshirani \(1990\)](#). Test results show that the semiparametric model performs better than a parametric model.

[Roy and van Kooten \(2004\)](#) also examined the existence of the EKC for three non-point source air pollutants: (a) carbon monoxide (CO), (b) nitrogen oxide (NO_x), and (c) ozone (O₃) using adjusted partial linear models allowing heteroskedasticity ([Robinson \(1988\)](#)). [Li and Wang \(1998\)](#) test statistics were used to compare a quadratic model against the semiparametric model. Compared to the previous literature, they used a log of income in their model. They used linear, quadratic, and cubic models and found that income is very sensitive to model specification. They found no existence of the EKC for these pollutants, which is also consistent with findings of [Millimet, List, and Stengos \(2003\)](#) for NO_x. As with the previous research, they used [Li and Wang \(1998\)](#)'s model specification test and found that the semiparametric model was better compared to the quadratic model.

[Bertinelli and Strobl \(2005\)](#) estimated the relationship between pollutants (sulfur oxide (SO₂) and carbon dioxide (CO₂)) using [Robinson \(1988\)](#) partial linear regression using 108 and 122 cross country observations for SO₂ and CO₂, respectively. In contrast to previous literature, [Bertinelli and Strobl \(2005\)](#) found interesting results that there exists a linear relationship between these pollutants and income. This implies that no EKC exist for these pollutants. The linear hypothesis was tested against the semiparametric model using a method suggested by [Ullah \(1985\)](#). The bootstrap procedure suggested by [Lee and Ullah \(2001\)](#) is used to obtain the standard error and the standard error is used to find the significance of the test statistic. They failed to reject the null of a linear relationship between income and pollution.

Table 2.2. Existing Published Studies That Have Used Semiparametric Techniques in Environmental Kuznets Curve Estimation

Literature	Types of Models Used	Additional Variables	Model Specification Test	Findings and Turning Points (TP)
Millimet, List, and Stengos (2003)	<i>Parametric:</i> Two-way fixed effects, cubic <i>Semiparametric:</i> Robinson (1988)	NO	Zheng (1996) and Li and Wang (1998)	EKC existed for SO ₂ and NO _x PS-SO ₂ : \$16,417, FS-NO _x : \$8,657, PS-NO _x \$10,570
Phu (2003)	<i>Parametric:</i> OLS <i>Semiparametric:</i> Hastie and Tibshirani (1990)	trade, population density, education and political institution	Gain statistics developed by Hastie and Tibshirani (1990)	No EKC for protected areas
Roy and van Kooten (2004)	<i>Parametric:</i> Linear and cubic models <i>Semiparametric model:</i> Robinson (1988) , Stock (1989) , and Kniesner and Li (2002)	Population density, % minorities, % unemployed, % labor in manufacturing, % with high school etc.	Li and Wang (1998)	EKC exists for NO _x , does not exists for CO and O ₃
Bertinelli and Strobl (2005)	<i>Parametric:</i> Quadratic Fixed effects <i>Semiparametric:</i> Robinson (1988)	No	Ullah (1985)	linear relationship between pollutant and income, i.e. No EKC existed for CO ₂ , SO ₂
Paudel, Zapata, and Susanto (2005)	<i>Parametric:</i> Fixed and random effects panel	Weighted income, Population density	Hong and White (1995)	EKC existed for nitrogen and dissolved oxygen but not for phosphorous

Table 2.2. Contd.

Literature	Types of Models Used	Additional Variable Used	Model Specification Test	Findings and Turning Points (TP)
	<i>Semiparametric</i> : Robinson (1988)			TP-N:\$12,993
Azomahou, Laisney, and Van (2006)	<i>Parametric</i> : Within cubic panel estimation	No	Li and Wang (1998)	EKC yes for CO ₂ in a parametric model, no for CO ₂ in a nonparametric model TP CO ₂ : \$13,258
Van and Azomahou (2007)	<i>Semiparametric</i> : Wand and Jones (1995) , Linton and Nielsen (1995) <i>Parametric</i> : Fixed and random effect panel	Trade, population growth rate, population density, literacy rate, political institution	Li et al. (2002)	EKC does not exist for deforestation
Criado (2008)	<i>Semiparametric</i> : Smooth coefficient model by Li et al. (2002) <i>Parametric</i> : Cubic panel fixed effects <i>Semiparametric</i> : Wood (2006) approach	No	V-test, Yatchew (2003) 's pooling test	EKC existed for CH ₄ , CO, CO ₂ , NMVOC TP CH ₄ : \$17,300; CO:\$16,800; CO ₂ : \$16,400; NMVOC: \$17,200
Luzzati and Orsini (2009)	<i>Parametric</i> : Fixed effects panel	No	No	EKC exists in energy consumption

Table 2.2. Contd.

Literature	Types of Models Used	Additional Variable Used	Model Specification Test	Findings and Turning Points (TP)
	<i>Semiparametric</i> : Generalized additive			TP for energy consumption: Low income countries:none; Middle income countries: \$ 57,500; High income countries: \$18500; Other countries: \$9,000
Poudel, Paudel, and Bhattarai (2009)	<i>Parametric</i> : Fixed effect	Forestry, Population density, Illiteracy, income weight	Li and Wang (1998)	EKC exists with N shaped
Phu (2010)	<i>Semiparametric</i> : Robinson (1988) <i>Parametric</i> : Cubic	Coal share, petroleum and gas share, time trend	Li and Wang (1998)	No EKC on energy consumption
Li (2011)	<i>Semiparametric</i> : panel semiparametric partially linear model <i>Parametric</i> : Quadratic Fixed effects	No	Average mean square error, Bootstrap confidence band	OECD-countries support for EKC
Zanin and Marra (2012)	<i>Semiparametric</i> : B-spline <i>Parametric</i> : OLS	No	restricted likelihood ration test (RLRT) by Crainiceanu and Ruppert (2004)	Existence of EKC for CO ₂ , France and Switzerland (U shape), Austria-N, Denmark, M shaped
	<i>Semiparametric</i> : Additive mixed model,penalized regression spline			

Table 2.2. Contd.

Literature	Types of Models Used	Additional Variable Used	Model Specification Test	Findings and Turning Points (TP)
Chiu (2012)	<i>Parametric</i> : OLS	Population density, trade openness, political freedom	F-version LM, and Pseudo LR test	\$3,021 and \$3,103
Kim (2013)	<i>Semiparametric</i> : Panel smooth transition regression (PSTR) of González, Teräsvirta, and Dijk (2005) <i>Parametric</i> : quadratic and cubic <i>Semiparametric</i> : kernel based semiparametric model	No	Upper Confidence bands (UCB)	EKC exists for SO ₂ and CO ₂

Note: FS= Full sample (1929-1994); PS= Partial sample (1985-1994)

EKC was also tested at the local level for water pollutants by [Paudel, Zapata, and Susanto \(2005\)](#). They estimated an EKC for nitrogen (N), phosphorus (P) and dissolved oxygen (DO) at the watershed level for 53 parishes for the period of 1985-1998 using the data collected by the Department of Environmental Quality. One-way and two-way fixed and random effects with quadratic and cubic model were estimated as parametric models. Using the [Hausman \(1978\)](#) test, they found that the fixed effect models was better than the random effects model. Like the previous literature, they also used the [Robinson \(1988\)](#) partial linear model as a semiparametric model. A method suggested by [Hong and White \(1995\)](#) was used to compare the parametric model against a semiparametric model. As expected, they found that the semiparametric model captured nonlinearity better than the quadratic and cubic models. They observed mixed results on the existence of EKC, i.e., the EKC exists for nitrogen but not for phosphorus and dissolved oxygen.

[Azomahou, Laisney, and Van \(2006\)](#) studied the empirical relationship between CO₂ emission and economic development using panel data from 100 countries over the period 1960-1996. They investigated the relationship using a cubic parametric model and a nonparametric model, and found that the parametric model shows an inverted U-shape relation but the nonparametric model does not support this shape. They used test statistics suggested by [Li and Wang \(1998\)](#) to compare results obtained from parametric and semiparametric models and observed that the null of correct parametric model is rejected in favor of the nonparametric model.

A forest is an indicator of environmental quality, because it helps to sequester CO₂ from air. Deforestation can cause serious environmental damage. [Van and Azomahou \(2007\)](#) investigated the relationship between deforestation and economic growth with a panel data set of 59 developing countries over the period 1972-1994 using parametric and semiparametric models. They estimated quadratic and cubic fixed and random effects models. They compared fixed effects versus random effects using the Hausman test. The test favored a random effects model contradictory to the finding from [Paudel, Zapata, and Susanto](#)

(2005). They used a smooth coefficient model suggested by [Li et al. \(2002\)](#), which is a more flexible model than the models used by previous researchers (e.g. Robinson's model). Using this model, they found that there is no EKC for deforestation. However, they found that the other variables (e.g. population density, political institutions) considered in the model have significant effects on deforestation. They tested the robustness between parametric and semiparametric models using test statistics proposed by [Li et al. \(2002\)](#) and found that a parametric model is preferred against the semiparametric model.

In general, many researchers have used panel data to study the pollution-GDP relationship. However, they assume the temporal (stability of the cross-sectional regressions over time) and spatial (stability of the cross-sectional regressions over individual units) homogeneity assumption of the panel data. [Criado \(2008\)](#) questioned these assumptions on model estimation and proposed a nonparametric poolability test of [Yatchew \(2003\)](#) in the EKC to avoid functional misspecification. [Criado \(2008\)](#) used a balanced panel of 48 Spanish provinces over the 1990-2002 time period to examine an EKC for air pollutant emission: methane (CH_4), carbon-monoxide (CO), carbon-dioxide (CO_2) and non-methanic volatile organic compounds (NMVOC). His findings indicate that the temporal poolability assumption holds in the Spanish provinces for three pollutants (CH_4 , CO, and CO_2), but spatial homogeneity does not hold for all four pollutants. The pooled nonparametric regression suggests existence of an EKC.

The use of a semiparametric model is not only used in the analysis of air and water pollutants. It is also tested over all types of EKC hypotheses. [Luzzati and Orsini \(2009\)](#) used a semiparametric model suggested by [Wood \(2006\)](#) to examine an EKC hypothesis on absolute energy consumption and gross domestic product (GDP) per capita for 113 countries over the period 1971-2004. They used both parametric fixed and random effect models as parametric models. They found the existence of an EKC for energy consumption.

Likewise, the previous research of [Poudel, Paudel, and Bhattarai \(2009\)](#) used a semiparametric model to examine an EKC for CO_2 using data from 15 Latin American countries.

They used quadratic and cubic one way fixed and random effects models as parametric models and Robinson's partial linear model. They used a test statistic suggested by [Li and Wang \(1998\)](#) for model specification and found that parametric model specification is rejected in favor of semiparametric specification. Their main finding was that they observed 'N' shaped income-CO₂ relation shape for Latin American countries.

[Phu \(2010\)](#) examines existence of EKC on per capita energy consumption using data from the Energy Information Administration (EIA) that includes a balanced panel of 158 countries and territories for the period of 1980-2004. He estimated both parametric and semiparametric models to study the energy pollution relationship. The model used by him is more general than the model used by [Luzzati and Orsini \(2009\)](#). He did not find the existence of EKC on energy consumption. This finding is contradictory to the finding from [Luzzati and Orsini \(2009\)](#). The test statistic suggested by [Li and Wang \(1998\)](#) is used to compare parametric versus semiparametric models with the result that a semiparametric model is suitable for their data.

[Li \(2011\)](#) proposed a flexible nonparametric approach to study the existence of EKC on sulfur emissions from hard coal, brown coal, petroleum, and mining activities from most of the countries of the world over the 1850-1990 period. She also used B-spline smoothing on semiparametric model and found mixed results between OECD and non-OECD countries. The results show that an EKC exists in OECD countries, but not in non-OECD countries. A correctly specified model produces the least average mean squared error (AMSE), and is usually used to test for goodness-of-fit statistics. [Li \(2011\)](#) used average mean squared error to compare parametric versus nonparametric specifications. The smallest AMSE value for a semiparametric model implies that the semiparametric model is better for this data.

A recent study by [Zanin and Marra \(2012\)](#) used a penalized spline regression method to examine the existence of an EKC for carbon dioxide (CO₂) using data from 10 developed countries. The results were mixed. The penalized spline method is more general than the spline regression used in the previous literature. In a penalized spline smoothing method, the

smoothing parameter is selected automatically, so it is more reliable than spline or B-spline regression. They observed an EKC with an inverted U-shape for France and Switzerland, an ‘N’ shaped for Austria, an inverted ‘L’ shaped for Finland and Canada, and an ‘M’ shaped relation for Denmark. They used a restricted likelihood ratio test (RLRT) suggested by [Ruppert, Wand, and Carroll \(2003\)](#) to compare robustness between semiparametric and parametric models. This test statistic is equivalent to testing the presence of random effects for spline regression coefficients. The random effect parameterizes the deviations of a smooth function from a given linear term ([Zanin and Marra, 2012](#)). The results suggest that the parametric (quadratic or cubic) model is not adequate to capture non-linearity between pollution and income.

[Chiu \(2012\)](#) also studied an EKC hypothesis in deforestation using data from 52 developing countries over the 1972-2003 period. They used a panel smooth transition regression (PSTR) model. Their results support the EKC hypothesis that, with an increase in real income, deforestation increases initially, and after reaching a certain income level, declines. They used an F-version of the likelihood ratio test and a pseudo likelihood ratio test to check model specification. [Chiu \(2012\)](#) found existence of an EKC hypothesis for deforestation.

A recent study by [Kim \(2013\)](#) studied the relationship between air pollution (NO_x and SO_2) emissions and per capita income from 1929-1994 to estimate an EKC model. These data are the same used by [Millimet, List, and Stengos \(2003\)](#). [Kim \(2013\)](#) used a kernel-based semiparametric model. He proposed a Uniform Confidence Band (UCB) for the nonparametric component $g(\cdot)$ to test parametric model specifications against the nonparametric model. According to this test statistic, if the nonparametric 95% upper confidence band contains a parametric estimate then we fail to reject the null hypothesis that the parametric specification is correct. They observed that the null of parametric model specification is rejected in favor of a semiparametric model. They also observed the existence of an EKC for sulfur dioxide and carbon dioxide.

2.6 Conclusions

This survey chapter emphasizes recent developments in the semiparametric econometric method and the recent use of these developments on EKC related studies. From these studies, we found that the partial linear model developed by [Robinson \(1988\)](#) and its extensions are mostly used to test the EKC hypothesis. We observed that many researchers used kernel based partial linear models in EKC (e.g. [Millimet, List, and Stengos, 2003](#); [Paudel, Zapata, and Susanto, 2005](#); [Poudel, Paudel, and Bhattarai, 2009](#); [Azomahou, Laisney, and Van, 2006](#)). Other researchers also used an alternative of kernel regression which is known as spline smoothing. Further, we found various forms of spline smoothing based semiparametric models are used in EKC literature. For example, [Millimet, List, and Stengos \(2003\)](#) used spline smoothing. [Phu \(2003\)](#) used an additive partial linear model suggested by [Hastie and Tibshirani \(1990\)](#), [Luzzati and Orsini \(2009\)](#) used a spline additive model suggested by [Wood \(2006\)](#). These types of models provide the best mean squared fit as well as prevent overfitting, an important concern in nonparametric smoothing. There are different flexible types of splines used in nonparametric regression. B-spline and P-spline smoothing are more flexible than a simple spline. P-spline is the most flexible method, where an optimum smoothing is determined by the data itself. These flexible B-spline and P-spline models are used by [Li \(2011\)](#) and [Zanin and Marra \(2012\)](#) in EKC studies, respectively.

Various advances in econometrics that capture non-linearity are still absent in the EKC literature. Although many authors have used additional variables in addition to income, these additional variables are mainly included in parametric form (see [Phu, 2003](#); [Paudel, Zapata, and Susanto, 2005](#); [Van and Azomahou, 2007](#)). It is likely that these variables may have nonlinear effects too. We need to investigate whether these variables should enter parametrically or nonparametrically in a model. This type of approach is used by [Pandit, Paudel, and Mishra \(2013\)](#) in off-farm labor supply decisions by farm operators and their spouses.

Racine and Li (2004) suggested a nonparametric estimation procedure which admits both continuous and categorical variables, which is absent in previous EKC literature. Further, Hsiao, Li, and Racine (2007) relaxed model specification tests by Li and Wang (1998) which also admits both continuous and categorical variables. Small samples are commonly used in EKC literature, so the usual model specification tests are not valid for the small sample size. The specification test suggested by Hsiao, Li, and Racine (2007) use the bootstrap method to derive significance level and therefore work well with a finite-sample. We observed that spline based semiparametric models are frequently used in recent EKC literature, but the authors have not included categorical variables entered as nonparametric components. Recent papers by Nie and Racine (2012) have proposed nonparametric spline regression for mixed data, which can be used in future EKC research. Lin and Liscow (2013) observed that the reduced form model used to examine the EKC hypothesis has endogeneity problem, a semiparametric instrumental regression developed by Darolles et al. (2011), Horowitz (2011) and Santos (2012) can be used in the EKC studies. Another development that can be used in the EKC studies is a dynamic panel semiparametric model which has been missing so far.

In order to identify an appropriate functional form between environmental quality and economic growth, we reviewed advanced literature in econometrics specifically related to nonparametric and semiparametric models. Then, we explained how the new developments have been used in EKC literature. We observed that there is still an ongoing debate about the use of econometric specification in EKC analyses. We found that recent studies have focused on relaxing distributional assumptions using nonparametric and semiparametric models. Existing studies have indicated a semiparametric model is better compared to a parametric model. Hence, the EKC hypothesis can be analyzed more accurately using recent econometrics advances in nonparametric/semiparametric models. Future research should consider using a more flexible form of econometric modeling.

CHAPTER 3.

ENVIRONMENTAL KUZNETS CURVE: STOCK AND FLOW WATER QUALITY PARAMETERS

3.1 Introduction

The debate on the existence of EKC continues for various pollutants across different geographical regions. The use of watershed level data provides information at the micro level which has been absent in most of the EKC literature. The aggregated national level data in water quality may not be appropriate to test existence of EKC as water quality varies greatly from one watershed to another watershed. Water pollution occurs when the pollutants are discharged directly or indirectly into water sources such as lakes, rivers, oceans, and aquifers. Water pollutants commonly emanating from non-point sources are known as flow pollutants whereas pollutants that continue to build up rather than dissolve are stock pollutants. Generally speaking, flow pollutants come from non-point sources and stock pollutants come from point sources.

As we observed in the previous chapter, several authors ([Millimet, List, and Stengos, 2003](#); [Paudel, Zapata, and Susanto, 2005](#); [Poudel, Paudel, and Bhattarai, 2009](#); [Zapata and Paudel, 2009](#); [Paudel and Poudel, 2013](#)) have found that a parametric model is not sufficient to capture non-linearity between pollutant and income, suggesting a need to include a nonparametric form of income in the regression model. These studies have found that semiparametric forms perform better than parametric forms in the specification test.

Previous literature has examined the EKC hypothesis in many pollutants using separate equations. For example, [Paudel, Zapata, and Susanto \(2005\)](#) studied three water pollutants (N, P and DO), and [Criado \(2008\)](#) studied four pollutants (CH₄, CO, CO₂ and NMVOC). The turning points in pollution-income relationship in these studies are also estimated using

a single equation panel data model for each pollutant with an assumed functional form of income as an explanatory variable for a pollutant.

Water may get polluted from more than one pollutant at the same time or they may come from similar sources, i.e. pollutants may be correlated to each other. N, P, and DO come from agricultural sources and stock pollutants come from operations that have association with agriculture. For example, chemical fertilizers may increase both flow (N and P) and stock (Hg) pollutants. However, previous researchers did not consider the potential correlation among pollutants in their EKC studies. In other words, the researchers do not consider the covariance of the error terms across different pollutants. A single equation estimation method may not be sufficient to examine the true relationship between income and pollutant.

Our study addresses four issues that have been raised but not sufficiently addressed by earlier studies. First, we jointly estimate stock and flow pollutants to determine if the EKC exists in both. Second, we use a seemingly unrelated partial linear regression (SUPLR) panel data model. Third, we use watershed level data (disaggregated data) on water pollution collected from Louisiana. Finally, we utilize a semiparametric model specification and test whether a semiparametric model performs better than a parametric model.

This chapter proceeds as follows. We provide brief reviews of recent studies done on water quality-income relationship. Then, we describe the pollutants studied in this essay. In the next section, we present econometric methodology and simulation study conducted to examine finite sample performance. We describe the data and pollutants used in the essay in the next section. The results section describes the parameter coefficients and other pertinent information obtained from the selected model. The last section concludes and provides some policy thoughts.

3.2 Pollution Sources

3.2.1 Nitrogen

Nitrogen (N) pollution in water comes from various sources. The major sources of nitrogen pollution are agricultural land, aquaculture, and livestock. Nitrogen comes from leaching and runoff from chemical fertilizers and manure. Nitrogen also comes into the surface/ground water from septic tank leakage. Other sources of nitrogen pollution in water include urban storm runoff, industry, and fossil fuel combustion.

3.2.2 Phosphorous

Phosphorous (P) is commonly found in soil particles. When soil particles are disturbed due to agricultural operations, landslides, and erosion, phosphorous gets released into water. Like nitrogen, use of chemical fertilizer and runoff from manure used for agriculture are also a major sources of phosphorous pollution. Other sources of phosphorous include sewage treatment plant discharge, storm water runoff and failing septic tanks.

3.2.3 Dissolved Oxygen

Dissolved oxygen (DO) is the amount of oxygen dissolved in water. DO is required by aquatic plants for respiration. As dissolved oxygen levels decrease, it becomes harder for aquatic animals to get sufficient oxygen they need to survive. Water gets oxygenated from the atmosphere as well as through photosynthesis from aquatic plants. Oxygen level decreases in water due to high temperature. DO is not a pollutant, but it is used as a parameter to measure pollution level in water. Nutrient pollution is a major cause of oxygen reduction in water. High nutrient levels in water cause excess growth of aquatic plants, which absorb oxygen during their decomposition phase. Because of this process, oxygen level can decrease significantly in water.

3.2.4 Mercury

Mercury (Hg) is also released into the water and atmosphere from various human activities. Mercury is found in rocks and coal. Coal is used for various purposes, but primarily for generating electricity. When coal is burned, mercury is released into the air and ultimately drops into the water or land and gets to waterbodies. According to the U.S. Environmental Protection Agency ¹ (EPA), coal burning power plants are the largest human-caused source of mercury emissions. Other sources of mercury include burning hazardous wastes, breaking mercury products, mercury spills, and improper treatment and disposal of products or wastes containing mercury. Mercury has been found in agricultural fertilizer. Application of chemical fertilizers and industrial wastewater disposal releases mercury directly into the soil or water (see [Zheng et al., 2008](#); [Zhao and Wang, 2010](#)). Cattle breeding products can also contain some amount of mercury ².

3.3 EKC Literature and Water Quality

Examination of an EKC hypothesis for water quality parameters began simultaneously with the emergence of the concept of EKC. [Grossman and Krueger \(1995\)](#) examined the existence of an EKC for eleven water pollutants (dissolved oxygen, biological oxygen demand (BOD), chemical oxygen demand, nitrates, fecal coliform, total coliform, lead, cadmium, arsenic, mercury and nickel). They used Global Environment Monitoring System (GEMS) data in their study. Using the GEMS data they found the existence of an inverted U-shaped relation for many water pollutants such as DO and BOD. After this study, many researchers examined the EKC hypothesis on water quality with different data, location and methods.

[Gergel et al. \(2004\)](#) used sediment data of phosphorous, cadmium, chromium, copper, lead and sulfur from Lake Mendota in Dane County, Wisconsin from 1900 to 2000. Using

¹<http://www.epa.gov/hg/about.htm>

²see <http://www.lenntech.com/periodic/elements/hg.htm#ixzz2VM5VqGbK>

quadratic and cubic model specifications, they found existence of an EKC only for chromium. When examining the EKC at the local level, such as county or watershed, it is most likely to have effect on the neighboring county or watershed, i.e., spatial correlation. In the presence of the correlation, the estimated ordinary least square estimates are biased. The issue of spatial correlation was addressed by [Paudel, Zapata, and Susanto \(2005\)](#) using the weighted income of neighboring parishes. [Paudel, Zapata, and Susanto \(2005\)](#) used disaggregated watershed level water data from 53 Louisiana Parishes from the years 1985-1998. [Paudel, Zapata, and Susanto \(2005\)](#) found the existence of an EKC for nitrogen and dissolved oxygen. The quadratic and cubic specifications used in the previous literature are restrictive, so they relaxed the assumed functional form by using a semiparametric model. They also compared the parametric model with the semiparametric model and found that the semiparametric model performs well compared to a parametric model to capture an income-pollution relationship. Researchers have also incorporated additional variables that impact the income pollution relationship. To illustrate, [Paudel and Schafer \(2009\)](#) added a social capital index on data used by [Paudel, Zapata, and Susanto \(2005\)](#) to examine the effect of social capital on the income pollution relationship. Using a parametric and a spatial regression model, they found that social capital plays a significant role for water pollution parameters. Specifically, they found a U-shaped relationship between water quality parameters (nitrogen, phosphorous and levels of dissolved oxygen) and social capital.

Several other approaches have been used to examine the income-water pollution relationship. For example, [Gassebner, Lamla, and Sturm \(2011\)](#) also studied the effect of income and BOD using Extreme Bound Analysis (EBA) as suggested by [Leamer \(1983\)](#) and [Levine and Renelt \(1992\)](#). Using panel data from 120 countries over the time period 1960-2001, they examined EKC for BOD and found the presence of an EKC. [Clement and Meunie \(2010\)](#) introduced concepts of social inequality into EKC hypotheses. They examined the relationship between social inequality and organic water pollution using panel data (fixed and dynamic panel data) models. Using data from 83 transitioning and developing

countries over the period 1988-2003, they found that an increase in inequality causes water pollution to increase for developing countries. The relationship was found to be uncertain for transitioning countries.

Previous studies have shown the existence of regional effects on EKC for water quality. For example, [Lee, Chiu, and Sun \(2010\)](#) studied an EKC using the general method of moment (GMM) approach for BOD, using data from 97 countries over the time period of 1980-2001. Their major finding was that there is a regional difference between EKCs for water pollution. They did not find the existence of an EKC at the global level, but found the existence of an EKC for BOD specific regions. They found the existence of an EKC for America and Europe but not for Asia, Africa or Oceania. [Orubu and Omotor \(2011\)](#) studied per capita income and environmental degradation measured by suspended particles and organic pollutants using data from thirteen African countries for the period 1990-2002. Using cubic and quadratic model specifications, they found the existence of an EKC for suspended particles, and rising pollution of organic pollutants as per capita income increases. Their results also indicated that the turning points for these parameters are lower than the turning points found in previous literature.

[Thompson \(2012\)](#) included water abundance in the EKC model. He analyzed the relationship between water abundance and water quality measured by BOD using data from 38 developed and developing countries. Using a pooled mean group (PMG) estimation procedure proposed by [Pesaran and Smith \(1995\)](#), he found the existence of an EKC for BOD. Specifically this article found that water abundance affects turning points. [Lin and Liscow \(2013\)](#) raised the problem of endogeneity in the income pollution model. That is, a third variable, such as cultural or geographical factors, jointly causes both economic growth and environmental degradation. They used total debt as an instrument for GDP and estimated quadratic and cubic models using the instrumental variable approach applied to GEMS data. They considered the same water quality indicators used by [Grossman and Krueger \(1995\)](#),

and found that there is an EKC relationship between income and water quality indicators for seven of the eleven water quality parameters tested.

A recent article by [Paudel and Poudel \(2013\)](#) used a stock pollutant (mercury) to test its effects on flow water quality parameters (Nitrogen, Phosphorus and Dissolved Oxygen), but they did not find mercury to have an effect on these flow pollutants. They used a semiparametric model in their study using data from [Paudel, Zapata, and Susanto \(2005\)](#). Another recent study by [Farzin and Grogan \(2013\)](#) examined EKC for 24 water quality indicators using data from 1993-2006 in California. They used data obtained from the U.S. Environmental Protection Agency. In addition to per capita income, they considered other socio-economic variables that affect water pollution. Social factors such as education, ethnic composition, land use, population density, and water area are correlated with many water quality parameters.

Although these studies have addressed many issues on EKC hypotheses for economic growth and water quality, they did not consider potential correlation among pollutants. Our study is the first to consider both stock and flow pollutants, and test an EKC hypothesis using seemingly unrelated semiparametric models. This is also the first study which considers disaggregated watershed level data in Louisiana for the period covering 1985-2006.

3.4 Methods

Both fixed and random effect models have been used to examine the existence of environmental Kuznet curves for different pollutants ([Paudel, Zapata, and Susanto, 2005](#)). In this essay, we use data collected from 53 parishes in Louisiana for the 1985-2006 period. Although it is reasonable to consider a fixed effects model (given parishes are fixed), we also estimate a random effects model. The descriptions provided below are for the fixed effects model.

3.4.1 Parametric Model

Numerous econometric models have been estimated to test the existence of an EKC. Typically, researchers use a reduced-form model in which pollution (pollutant concentration or pollution per capita) is a quadratic or cubic function of income and a linear form of other factors that affect pollution. We have balanced panel data, since pollution information is collected from each parish for every year between 1985 and 2006. Let the pollution P_{it} from parish i at time t satisfy a linear model with an intercept that is specific to parish i given by

$$P_{it} = \mathbf{y}'_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\alpha} + \gamma_i + \epsilon_{it} \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (3.1)$$

where, $\mathbf{y}_{it} = (y_{it}, y_{it}^2)$ if we consider quadratic, and $\mathbf{y}_{it} = (y_{it}, y_{it}^2, y_{it}^3)$ if we consider cubic model; y_{it} is per capita GDP; $\boldsymbol{\beta}$ is the vector representing the parameter for corresponding variables. \mathbf{x} represents variables that affect pollution other than per capita income such as population density, total crop acres; $\boldsymbol{\alpha}$ are parameters corresponding to these factors; γ_i is fixed effect; ϵ_{it} is i.i.d with a zero mean and a finite variance σ_ϵ^2 .

We have four water pollutants that need to be analyzed. We can estimate the EKC based equation 3.1 for each pollutant separately if errors from each equation are not contemporaneously correlated or if explanatory variables are the same for each equation. Since the variables that affect stock and flow pollutants are different, we use a Seemingly Unrelated Regression (SUR) model to incorporate contemporaneous errors from each equation. Suppose there is a set of M equations for each pollutant (e.g. N, P, DO, Hg), then it can be written in a SUR panel data model as shown in equation (3.2)

$$P_{jit} = Y_{jit}\boldsymbol{\delta}_j + X_{jit}\boldsymbol{\alpha}_j + \Gamma_j + \epsilon_{jit} \quad j = 1, \dots, M \quad i = 1, \dots, N \quad t = 1, \dots, T \quad (3.2)$$

where P_{jit} is stack vector for concentration of pollutant j in parish i in time t . Y_{jit} is the stack matrix of quadratic or cubic of per capita income as defined above and X_{jit} is

the stack matrix of other factors that affect pollutant j in equation 3.1, ϵ_{jit} are random vectors with a zero mean and $\sum_{\epsilon} \otimes I_{NT}$ variances. If the assumption that the covariance of residuals between M equations is not zero, the parameter estimated by joint equations are asymptotically more efficient than the parameters estimated from an individual equation [Baltagi \(1980\)](#).

3.4.2 Semiparametric Model

In this section, we consider the following semiparametric panel data model with fixed effects

$$P_{it} = g(y_{it}) + x_{it}\alpha + \gamma_i + \epsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (3.3)$$

where, $g(\cdot)$ is an unknown smooth function, and all other symbols are same as in the previous section. This partially linear model with fixed effects can be estimated using the first difference as described in [Li and Racine \(2007\)](#). The semiparametric SUR model can be written as follows:

$$P_{jit} = G_j(y_{it}) + X_{jit}\alpha_j + \Gamma_{ji} + \epsilon_{jit} \quad j = 1, \dots, M, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad (3.4)$$

Where $G(\cdot)$ is an unknown smooth function for M system of equations, Γ_{ji} represents fixed effects for i^{th} parish in j^{th} equation. [You, Zhou, and Chen \(2013\)](#) have provided an estimation procedure of a multivariate partial linear model. In the multivariate partial linear model, all the explanatory variables are common for each equation. We extended the work by [You, Zhou, and Chen \(2013\)](#) for the SUPLR model. The detailed estimation procedure for the SUPLR model is provided in the Appendix (A).

3.4.3 Model Specification Test

The true model specification is never known. Therefore, alternative model specifications should be considered in practice. Contributions on semiparametric modeling of the environ-

mental Kuznets curve hypothesis (Paudel, Zapata, and Susanto, 2005; Bertinelli and Strobl, 2005; Roy and van Kooten, 2004; Millimet, List, and Stengos, 2003) suggest the specification of a semiparametric partial linear regression (PLR) model such as in Robinson (1988). The model is flexible in capturing non-linearity between environmental quality and per capita income, and it minimizes the tradeoff between variance and bias (Härdle, 1990). Consistent with preliminary parametric diagnostics, the panel data model is specified as a fixed effects SUR model, which can be rewritten as in equation (3.1). The model specification test consists of testing a parametric model in equation (3.1) against a semiparametric specification in the equation (3.3). We used test statistics suggested by Hsiao, Li, and Racine (2007). Assume that the parametric model is correctly specified. Then, the null and alternative hypotheses are:

H_0 : Parametric Model

H_1 : Nonparametric/Semiparametric Model

The test statistic purposed by Hsiao, Li, and Racine (2007) is

$$\hat{J}_n = \frac{n(\hat{h}_1 \dots \hat{h}_q)^{1/2} \hat{I}_n}{\sqrt{\hat{\Omega}}}, \quad (3.5)$$

where h is bandwidth and

$$I_n = n^{-2} \sum_i \sum_{i \neq j} \hat{\epsilon}_i \hat{\epsilon}_j K_{\gamma,ij} \quad (3.6)$$

where $K_{\gamma,ij}$ is product of continuous and discrete variables.

$$\Omega = \frac{2(\hat{h}_1 \dots \hat{h}_q)}{n^2} \sum_i \sum_{i \neq j} \hat{\epsilon}_i \hat{\epsilon}_j W_{h,ij}^2 L_{\lambda,y}^2 \quad (3.7)$$

$$\hat{\epsilon}_i = y_i - g(x_i, \hat{\beta}) \quad (3.8)$$

where $W_{h,ij}$ and $L_{\lambda,y}$ are kernel functions for continuous and discrete variables, respectively. J_n is distributed $N(0,1)$ under the null hypothesis. J_n test diverges to $-\infty$ if H_0 is false.

Thus, we reject the null hypothesis if the J_n value is lower than the critical value from a normal distribution.

3.5 Simulation Study

We examined the performance of a Seemingly Unrelated Partially Linear Regression (SUPLR) model using a Monte Carlo experiment. We first ran a simulation to demonstrate the finite sample performance of an SUPLR estimator proposed in equation (3.4). We also ran empirical simulations to determine whether this model performs well with our data set. We simulated data sets from a two dimensional SUPLR model. The data are generated from the following seemingly unrelated partially linear regression model.

$$p_1 = X_1\beta_{11} + X_3\beta_{13} + g_1(y) + \epsilon_1 \quad (3.9)$$

$$p_2 = X_2\beta_{21} + X_3\beta_{23} + g_2(y) + \epsilon_2 \quad (3.10)$$

In our data, the explanatory variables are nonnegative, right skewed and independent, so we generated independent variables as $X_1 \sim 0.3 \times \chi_1^2$, $X_2 \sim \chi_1^2$, $X_3 \sim |N(0, 1)|$. In order to check the sensitivity of estimation procedure, we chose different signs and values for the parameters. We set $\beta_{11} = 1.5$, which represents a small coefficient, $\beta_{13} = 5$ implies a large parameter value. $\beta_{21} = -2$ and $\beta_{23} = 2$ represent negative and positive values, respectively. A nonlinear relation between pollution and income is represented by *sin* and *cosine* functions. These functions cannot be exactly approximated by quadratic and cubic models, which are commonly used to show nonlinear relationship between income and pollution. We specified different nonlinear relations for each equation as given below:

$$g_1(.) = 2 \sin(2\pi.)$$

$$g_2(.) = \cos(1.5\pi.).$$

We allowed correlation of errors of the two equations, such that the errors follows a bivariate normal distribution with $\epsilon = (\epsilon_1, \epsilon_2)' \sim N(0, \Sigma)$, $\Sigma = \sigma_{ij}^2$. We assumed errors are homoskedastic in each equation with $\sigma_{11}^2 = \sigma_{22}^2 = 1$, and correlated across equations as shown with $\sigma_{12}^2 = 0.3$ or 0.6 or 0.9 . Thus, $\sigma_{12}^2 = 0.3$ represents low correlation between error terms in two equations and $\sigma_{12}^2 = 0.9$ indicates high correlation between error terms in two equations. Here, variable X_1 is present only in the first equation, variable X_2 is present only in the second equation and variable X_3 is present in both equations. In this simulation, we drew samples of different sizes, viz, $n = 100, 200$ and 500 , and estimated the SUPLR model. In each case, we repeated the simulation 1000 times. The Gaussian kernel function defined in equation (3.11) is used to fit the nonparametric component of the SUPLR.

$$K_h(y) = \frac{1}{h\sqrt{2\pi}} \exp\left\{\frac{-(y)^2}{2h^2}\right\} \quad (3.11)$$

The Cross-Validation bandwidth selection method is used to find an optimal bandwidth (h) for each estimation.

For a given sample size and error correlation, we calculated the average of estimated parameters, average of asymptotic standard error, standard deviation (SD) of estimated parameters and rejection rate of the 5% test of the estimated parameters. These statistics are summarized in Table (3.1). This table shows that the averages of estimated parameters are close to the true parameter values for all sample sizes. This result indicates that the estimated parameters of the SUPLR model are asymptotically unbiased. As sample size N increases, the coefficients are closer to the true parameter values indicating the consistency of estimated parameters. The rejection rates from a 5% t-test are close to 0.05, hence the approximate normality of the estimators is very good. The average of the asymptotic standard error is close to the SD of the estimated parameters. These statistics are also consistent and unbiased for the high correlation of errors. We also estimated average and standard deviation of error variance (i.e. σ_{11}^2 , σ_{12}^2 and σ_{22}^2). The summary statistics for these

Table 3.1. Simulation Results for Parameters Estimated Parametrically in a Semiparametric Model

No. Obs	σ_{12}	Statistics	Equation 1		Equation 2	
			$\beta_{12} = 1.5$	$\beta_{13} = 5$	$\beta_{22} = -2$	$\beta_{23} = 2$
n=100	$\sigma_{12}=0.3$	$\hat{\beta}_{ij}$	1.498	4.988	-1.993	1.998
		SD	0.267	0.176	0.077	0.180
		SE	0.246	0.168	0.075	0.171
		REJECT	0.055	0.066	0.056	0.066
	$\sigma_{12}=0.6$	$\hat{\beta}_{ij}$	1.489	4.988	-1.991	1.995
		SD	0.261	0.177	0.081	0.175
		SE	0.266	0.178	0.079	0.179
		REJECT	0.052	0.053	0.053	0.047
	$\sigma_{12}=.9$	$\hat{\beta}_{ij}$	1.494	4.984	-1.992	1.993
		SD	0.260	0.170	0.076	0.168
		SE	0.263	0.177	0.079	0.179
		REJECT	0.054	0.044	0.053	0.032
n=200	$\sigma_{12}=0.3$	$\hat{\beta}_{ij}$	1.502	4.991	-1.993	1.997
		SD	0.172	0.122	0.054	0.123
		SE	0.170	0.117	0.052	0.118
		REJECT	0.061	0.061	0.063	0.059
	$\sigma_{12}=0.6$	$\hat{\beta}_{ij}$	1.497	4.997	-1.997	1.997
		SD	0.184	0.118	0.052	0.125
		SE	0.173	0.120	0.052	0.120
		REJECT	0.062	0.046	0.051	0.064
	$\sigma_{12}=.9$	$\hat{\beta}_{ij}$	1.501	4.996	-1.995	1.996
		SD	0.174	0.121	0.054	0.118
		SE	0.178	0.122	0.053	0.123
		REJECT	0.050	0.047	0.052	0.044
n=500	$\sigma_{12}=0.3$	$\hat{\beta}_{ij}$	1.498	4.999	-1.998	1.995
		SD	0.108	0.074	0.033	0.074
		SE	0.107	0.074	0.032	0.074
		REJECT	0.053	0.051	0.057	0.052
	$\sigma_{12}=0.6$	$\hat{\beta}_{ij}$	1.499	4.998	-1.999	1.999
		SD	0.108	0.078	0.033	0.077
		SE	0.107	0.075	0.032	0.075
		REJECT	0.055	0.059	0.057	0.050
	$\sigma_{12}=.9$	$\hat{\beta}_{ij}$	1.495	5.000	-1.999	2.001
		SD	0.105	0.074	0.033	0.074
		SE	0.108	0.076	0.033	0.076
		REJECT	0.048	0.050	0.055	0.044

Note. SD refers standard deviation, SE represent asymptotic standard error of estimated parameters, and REJECT denotes rejection rate of 5% test.

Table 3.2. Simulation Results of Estimated Error Variance of SUPLR Model

No. obs.	σ	Statistics	σ_{11}	σ_{12}	σ_{22}
n=100	$\sigma_{12} = 0.3$	$\hat{\sigma}_{ij}$	0.977	0.296	1.011
		SD	0.089	0.040	0.104
	$\sigma_{12} = 0.6$	$\hat{\sigma}_{ij}$	1.025	0.606	1.053
		SD	0.101	0.064	0.113
	$\sigma_{12} = 0.9$	$\hat{\sigma}_{ij}$	1.092	0.927	1.117
		SD	0.130	0.102	0.148
n=200	$\sigma_{12} = 0.3$	$\hat{\sigma}_{ij}$	0.983	0.303	1.002
		SD	0.039	0.019	0.057
	$\sigma_{12} = 0.6$	$\hat{\sigma}_{ij}$	1.0155	0.611	1.029
		SD	0.051	0.036	0.066
	$\sigma_{12} = 0.9$	$\hat{\sigma}_{ij}$	1.064	0.927	1.080
		SD	0.093	0.067	0.094
n=500	$\sigma_{12} = 0.3$	$\hat{\sigma}_{ij}$	0.993	0.302	0.997
		SD	0.014	0.008	0.015
	$\sigma_{12} = 0.6$	$\hat{\sigma}_{ij}$	1.009	0.607	1.011
		SD	0.021	0.016	0.022
	$\sigma_{12} = 0.9$	$\hat{\sigma}_{ij}$	1.033	0.912	1.036
		SD	0.039	0.029	0.039

error variances are given in the Table (3.2). The estimated error variances are also close to the true error variance. They are also consistent and unbiased for high correlation. Hence, these estimated parameters are asymptotically unbiased and consistent.

Performance of nonparametric estimation is examined using partial regression plots of a variable entering nonparametrically into the semiparametric model. The partial regression plots, by number of samples and error variance, are provided in Figures (3.1-3.3). These plots provide curves of assumed functional forms $g_j(y), (j = 1, 2)$ and estimated curve $\hat{g}_j(y)$. If both curves are close to each other, the nonparametric estimates are unbiased and consistent. Figures (3.1-3.3) show that estimated nonparametric estimates are close to the assumed functional forms. The estimated nonparametric components are closer as the sample size n increases (see Figure 3.3). Hence, the seemingly unrelated semiparametric partial linear model performs well in a finite sample.

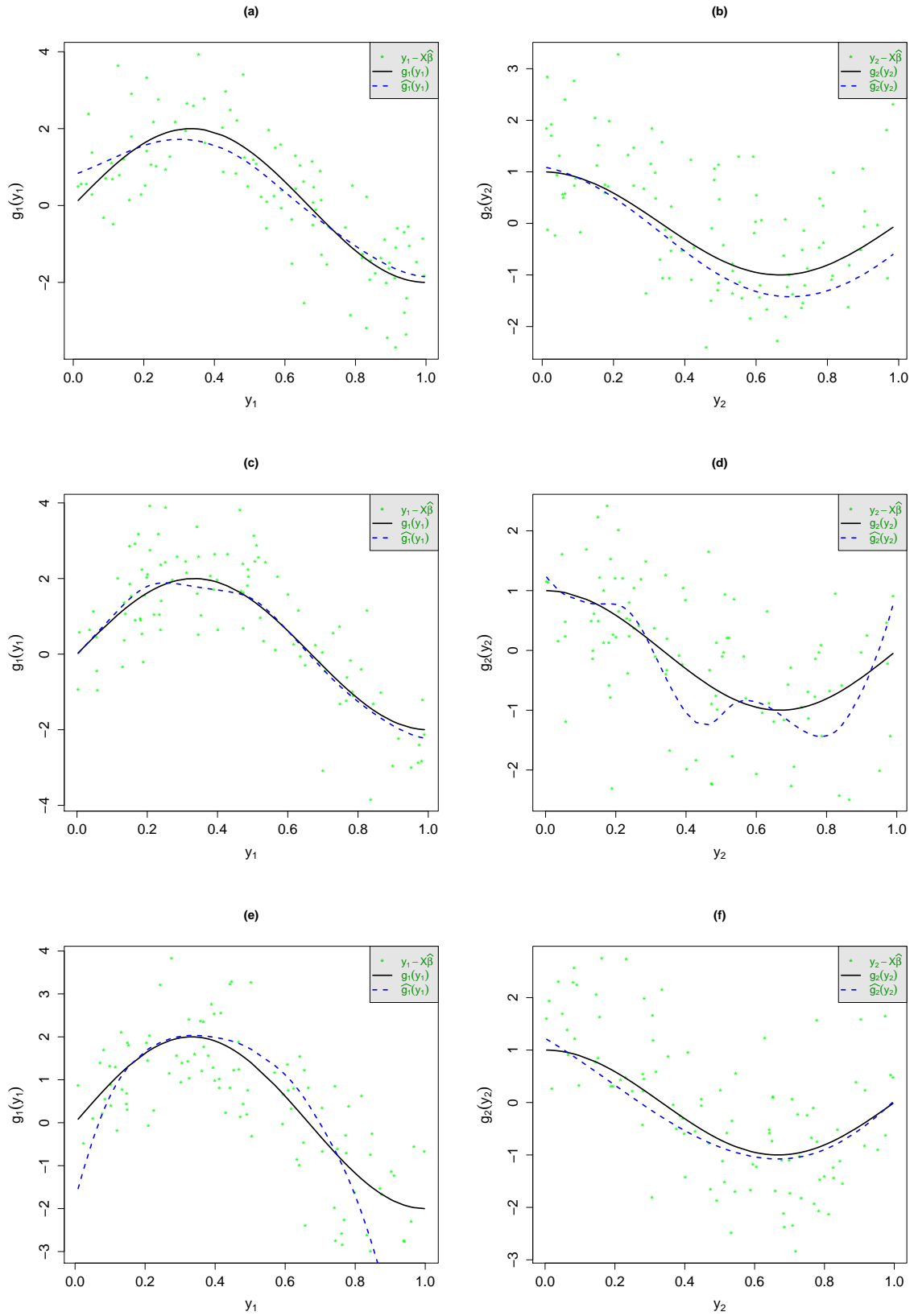


Figure 3.1. Partial Regression Plots of y , $n = 100$
 Note: $\sigma_{12} = 0.3(a, b), 0.6(c, d), 0.9(e, f)$

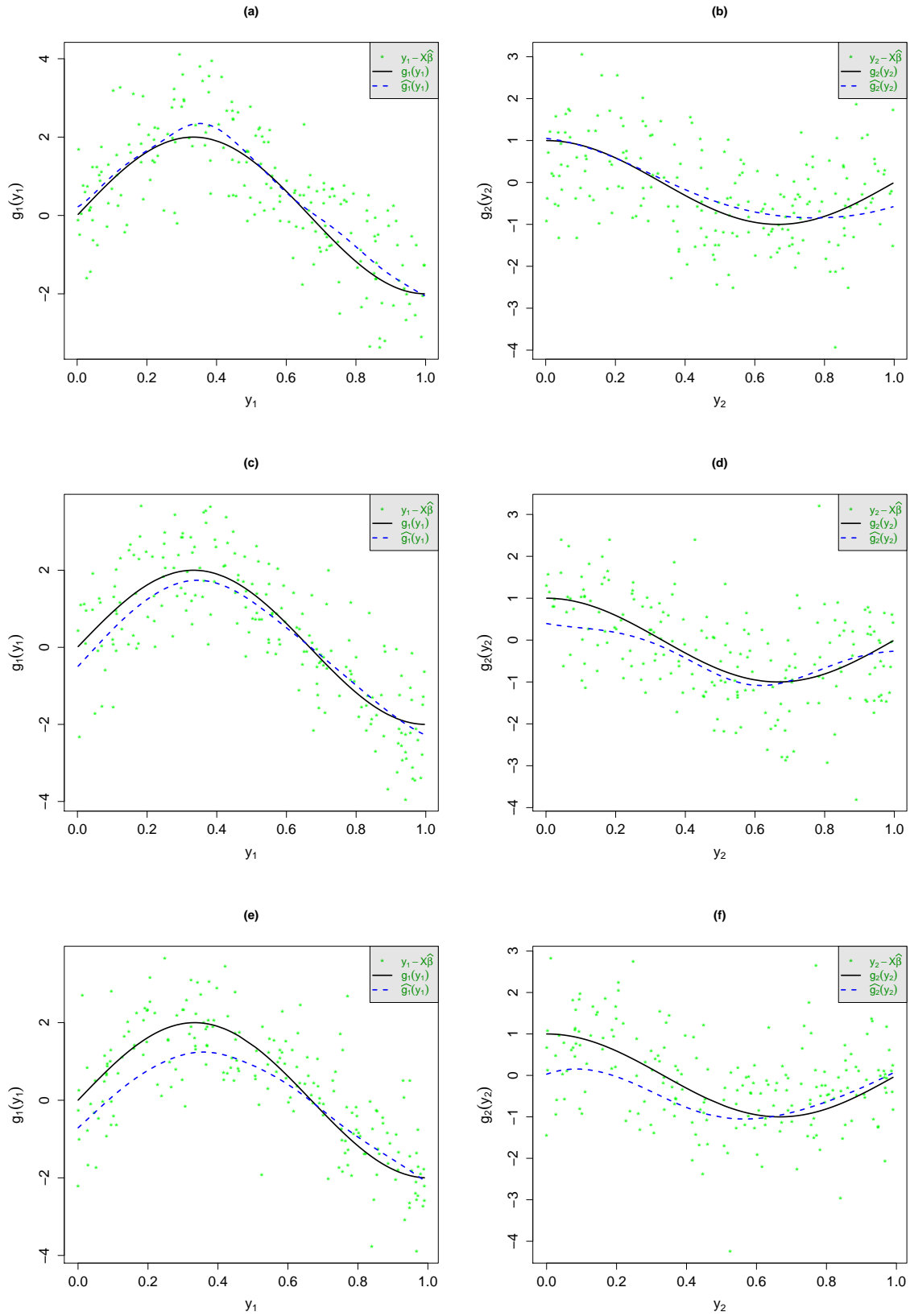


Figure 3.2. Partial Regression Plots of y , $n = 200$
 Note: $\sigma_{12} = 0.3(a, b), 0.6(c, d), 0.9(e, f)$

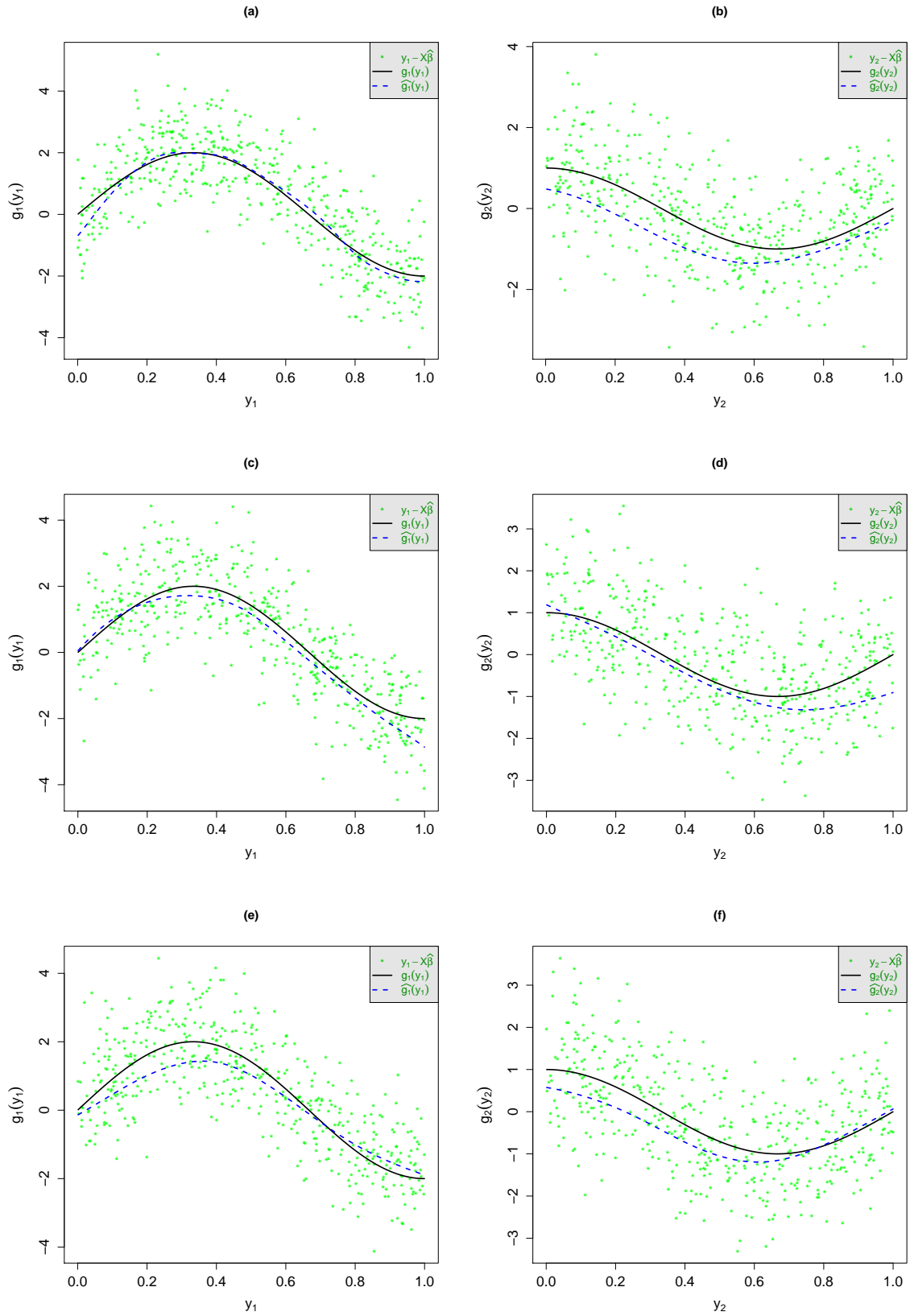


Figure 3.3. Partial regression plots of y , $n = 500$
 Note: $\sigma_{12} = 0.3(a, b), 0.6(c, d), 0.9(e, f)$.

In order to check whether this method is appropriate for empirical estimation of the pollution data, we generated values for two pollutants (p_1, p_2) using per capita income (y) , farm land area (X_1) , number of permits issued to point sources (X_2) and population density (X_3) . The values of X_1 , X_2 and X_3 come from the original data. Per capita income is entered nonparametrically and population density is entered parametrically in both equations. The first equation includes farm land area, whereas the second equation includes number of permits. The first equation represents nitrogen, phosphorous and dissolved oxygen, and the second equation represents mercury in this empirical model. The two models are expressed as follows.

$$p_1 = X_1\beta_{11} + X_3\beta_{13} + g_1(y) + \epsilon_1 \quad (3.12)$$

$$p_2 = X_2\beta_{21} + X_3\beta_{23} + g_2(y) + \epsilon_2 \quad (3.13)$$

where $\beta_{11} = 1.5$, $\beta_{13} = 1$, $\beta_{21} = 2$, $\beta_{23} = -1$, $g_1(\cdot) = 2 \sin(2\pi\cdot)$, $g_2(\cdot) = 1.5 \times \cos(1.5\pi\cdot)$, $\epsilon = (\epsilon_1, \epsilon_2)' \sim N(0, \Sigma)$, $\Sigma = \sigma_{ij}^2$ with $\sigma_{11}^2 = \sigma_{22}^2 = 1$ and $\sigma_{12}^2 = 0.3$ or 0.6 or 0.9 .

The number of observations is equal to 1166. For given error variance, average of estimated parameters, average of asymptotic standard error (SE), and standard deviation (SD) of estimated parameters, a test with a rejection rate of 5% is calculated. These statistics are provided in Table (3.3). The average value of estimated parameters is close to the true parameters, indicating that the parametric estimates of the SUPLR model are unbiased. The average rejection rates from the 5% t-test are close to 0.05, thus the approximate normality of the estimators is very good. Further, SE and SD are also close. Thus the estimated parameters are unbiased and consistent. The estimated error variance is close to the true error variance as shown in Table (3.4). Hence, error variance are also unbiased and consistent.

In order to check performance of the nonparametric estimation, we plotted true functional form and estimated functional form as shown in Table (3.4) for low to high error variance or error correlation. These figures show that the estimated nonparametric estimates are very

Table 3.3. Empirical Simulation Results for Parameters Estimated Parametrically in a Semiparametric Model

Covariance	Statistics	Equation 1		Equation 2	
		$\beta_{11} = 1.5$	$\beta_{13} = 1$	$\beta_{22} = 2$	$\beta_{23} = -1$
$\sigma_{12} = 0.3$	$\hat{\beta}_{ij}$	1.500	0.998	1.990	-1.002
	SD	0.121	0.050	0.217	0.051
	SE	0.111	0.050	0.190	0.049
	REJECT	0.071	0.056	0.086	0.050
$\sigma_{12} = 0.6$	$\hat{\beta}_{ij}$	1.493	0.998	1.969	-1.002
	SD	0.114	0.053	0.217	0.052
	SE	0.112	0.052	0.190	0.051
	REJECT	0.056	0.059	0.089	0.055
$\sigma_{12} = .9$	$\hat{\beta}_{ij}$	1.499	0.999	1.983	-1.002
	SD	0.117	0.053	0.216	0.053
	SE	0.113	0.052	0.191	0.051
	REJECT	0.053	0.046	0.084	0.053

Note: $N = 1166$

Table 3.4. Empirical Simulation Results of Estimated Error Variance of Model

Covariance	Statistics	σ_{11}	σ_{12}	σ_{22}
$\sigma_{12} = 0.3$	$\hat{\sigma}$	1.006	0.304	0.997
	SD	0.043	0.031	0.041
$\sigma_{12} = 0.6$	$\hat{\sigma}$	1.021	0.600	0.999
	SD	0.042	0.034	0.042
$\sigma_{12} = 0.9$	$\hat{\sigma}$	1.042	0.893	1.005
	SD	0.042	0.040	0.043

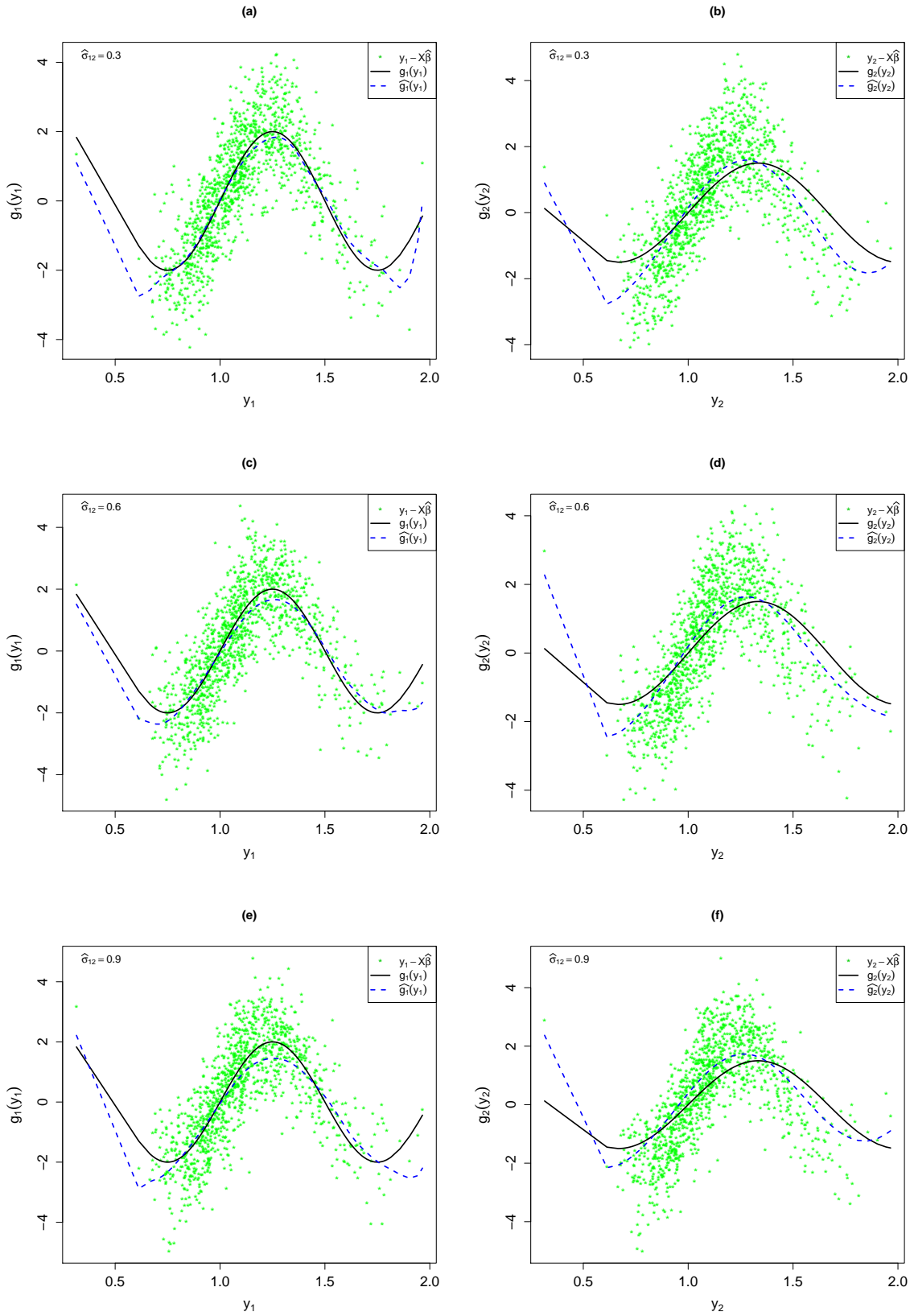


Figure 3.4. Partial Regression Plots of y from Empirical Simulation

close to the true functional forms for both low and high correlation. These simulation results show that the SUPLR model performs well in finite samples as well as for pollution data.

3.6 Data

Disaggregated data on nitrogen (N), phosphorus (P), dissolved oxygen (DO) (primarily flow pollutants), and mercury (Hg)(stock pollutant) concentration in water from Louisiana watersheds were used in this study. The value of these water quality parameters for each watershed was obtained from the Louisiana Department of Environmental Quality (LDEQ)³. Since each parish contains portions of several watersheds (see Figure 3.5), a weighted arithmetic mean⁴ was used to measure the level of water pollutant concentration for a given parish. The data consist of observations from 53⁵ parishes of Louisiana during 1985-2006.

We used per capita income as a measure of economic growth. Per capita income captures the endogenous characteristics of economic growth or all the factors of economic growth i.e. industrialization, urbanization and other development factors (Shafik, 1994). Per capita income for each Louisiana parish was obtained from the Bureau of Economic Analysis (BEA)⁶. Income is adjusted by CPI (1982-1984=100) as to convert all values into real dollars.

To account for the effect of adjacent parishes on individual parish pollution levels i.e. spillover effect, we calculated the queen contiguity matrix. This matrix considers all adjacent parishes within Louisiana and contiguous counties from adjacent states. Using this matrix, we obtained average income by summing the per capita income of the adjacent parishes/county for each year and dividing the total income by the number of contiguous parishes/counties. This average income is used as a weighted income variable to measure

³The data can be accessed from this link <http://www.deq.louisiana.gov/portal/tabid/2739/Default.aspx>.

⁴Weighted arithmetic mean of pollution P_i for i^{th} parish = $\sum_{j \in i} p_j \times A_{ji}/A_i$, where j represents j^{th} parish, A_{ji} represents area of j^{th} watershed in i^{th} parish, and A_i is total area of i^{th} parish.

⁵Data from eleven parishes (Bienville, Claiborne, Concordia, Evangeline, Iberia, Red River, Sabine, St. Bernard, Vernon, West Baton Rouge, and West Feliciana) are not available. These parishes are highlighted yellow in Figure (3.5).

⁶Per capita income and total population data is available from <http://www.bea.gov>.



Figure 3.5. Louisiana Parishes and Water Sub-segments

(Note: sub-segments are shown in the background)

spillover effects in the empirical model. A similar approach has been used by [Paudel, Zapata, and Susanto \(2005\)](#).

It is hypothesized that more populated areas are likely to be more concerned about the environmental quality than less populated areas. Higher population density indicates the generation of higher amounts of waste and higher levels of water pollution as well. Similarly, lower population density has just the opposite meaning (i.e., lower amounts of waste and lower levels of water pollution). High population density is therefore likely to have positive

Table 3.5. Summary Statistics of Variables

Variable	Variable definition		Mean	Std. Dev.	Min.	Max.
<i>Dependent variables</i>						
N	Nitrogen (mg/l)	overall	0.5344	0.6203	0.0000	8.3579
		between		0.2323	0.1975	1.0794
		within		0.5760	-0.5099	8.1465
p	Phosphorous (mg/l)	overall	0.1959	0.1583	0.0000	3.2832
		between		0.0836	0.0667	0.3788
		within		0.1349	-0.0423	3.1817
do	Dissolved Oxygen (mg/l)	overall	6.0898	1.8862	0.0000	9.7342
		between		1.1066	3.3489	8.0318
		within		1.5348	-0.4731	10.4533
hg	Mercury (μ g/l)	overall	0.0929	0.2493	0.0000	7.5500
		between		0.0628	0.0096	0.4511
		within		0.2414	-0.3582	7.1918
<i>Explanatory variables</i>						
pinc	CPI adj. per capita income (US \$10000/year)	overall	1.1124	0.2288	0.3155	1.9650
		between		0.1785	0.8574	1.5704
		within		0.1452	0.0505	1.5970
pinc2	Income square	overall	1.2898	0.5481	0.0996	3.8611
		between		0.4321	0.7506	2.5072
		within		0.3422	-0.5729	2.7573
pinc3	Income cube	overall	1.5590	1.0397	0.0314	7.5871
		between		0.8157	0.6698	4.0668
		within		0.6538	-1.2382	5.3676
wpinc	Weight income (US \$10000/year)	overall	0.7381	0.0754	0.6012	1.0581
		between		0.0446	0.6561	0.8725
		within		0.0610	0.6229	0.9619
popden	Population density (person /1000 sq. mile)	overall	0.1180	0.1812	0.0040	0.9203
		between		0.1823	0.0049	0.8535
		within		0.0142	-0.1668	0.2197
area	Farm land area (area in 10,000 acres)	overall	7.2073	9.7486	0.0200	37.0000
		between		9.6261	0.1999	30.4018
		within		2.0110	-2.2279	17.7316
permit	Number of facility sources that emits mercury pollution (SIC code: 20-40)	overall	1.4408	2.6217	0.0000	17.0000
		between		1.7117	0.0000	7.8182
		within		1.9991	-6.3774	11.5317

Note: Data is balanced panel with N=1166, n=53 and T=22.

or negative sign in the regression. Previous studies have used population density as one of the important factors that affect pollution. Population density is also considered in our reduced form model. Total population in each parish by year is also obtained from the BEA. Then, the population density is calculated by dividing the population in a parish by its corresponding parish area.

The number of point sources in each parish also plays an important role in water pollution, so we included the number of point sources as an important factor for stock pollution. We identified point sources by the number of permit holders in a given area. The data about permit compliance systems is obtained from the Better Assessment Science Integrating point Non-point Sources (BASINS) which is available from United States Environmental Protection Agency (EPA)⁷. Then we identified selected facilities that release mercury. According to the [MCRIA Council \(1999\)](#), mercury releasing facilities are those facilities whose standard industrial classification (SIC) code is between 20-40. Thus, we only used these facilities to count the number of point sources (permit) operating in each parish by year.

Farmland produces flow pollutants through the application of fertilizer and manure, which subsequently leaches into waterways. We consider farmland in each parish, measured by acres, as an important factor in water pollution. The farmland areas are obtained from the National Agricultural Statistics Service (NASS) quick stats⁸. This data provides the total acres of farmland planted in each parish every year.

Table (3.5) provides summary statistics for four water pollutants and independent variables. The water pollutants nitrogen (N), phosphorous (P) and dissolved oxygen (DO) are measured in milligrams per liter (ml/l) of water, mercury (Hg) is measured in micrograms per liter ($\mu\text{g/l}$), per capita income is in \$1,000 US dollars in real value, the population density is measured as the number of persons living per 1000 sq. miles, farm crop area is farm acres measured in 10,000 acre units, and permit represents the number of permits issued to point source polluters. Table (3.5) shows that the range of income is from \$3,155 to \$19,650 and

⁷<http://water.epa.gov/scitech/datait/models/basins/index.cfm>

⁸<http://quickstats.nass.usda.gov/>

average income is \$5,344. Population density ranged from a minimum of 4 people per square mile to a maximum of 920 people per square mile. The mean crop farm land in each parish is found to be 72,073 acres. Descriptive statistics show that the average number of point sources are 1.44 by parish per year.

3.7 Results

3.7.1 Parametric Model

Before describing the final results, we performed various tests related to panel data to find an appropriate functional form. The F-statistics for testing the joint significance of the individual heterogeneity (Fixed Effects) is given in Table 3.6 for all equations. The test statistic values are large and the p-values are less than 0.05 for all equations (nitrogen, phosphorous, dissolved oxygen and mercury) in both the quadratic and cubic models. This result strongly implies the presence of individual heterogeneity in our data.

Table 3.6. Test for Presence of Fixed Effect

Model	Nitrogen	Phosphorous	Dissolved Oxygen	Mercury
Quadratic	4.030 (0.000)	7.080 (0.000)	15.190 (0.000)	2.150 (0.000)
Cubic	4.030 (0.000)	7.100 (0.000)	15.110 (0.000)	2.180 (0.000)

Note: This test is conducted using F-test.

The values given in parenthesis are p-values.

3.7.2 Semiparametric Model

Following existing research in EKC's, we included income nonparametrically in a semiparametric model. Other remaining variables are entered parametrically in the semiparametric model. We used a kernel smoothing technique in a nonparametric model and we used a cross validation method to select an optimal bandwidth. Then, we estimated local linear semi-

parametric models. The parameter estimates of variables entering as parametric components in the semiparametric model are given in Table (3.8). We use a partial regression plot to see the effect of different variables (i.e per capita income) entering nonparametrically in the semiparametric model. The fitted partial regression plot for per capita income is shown in Figure (3.6).

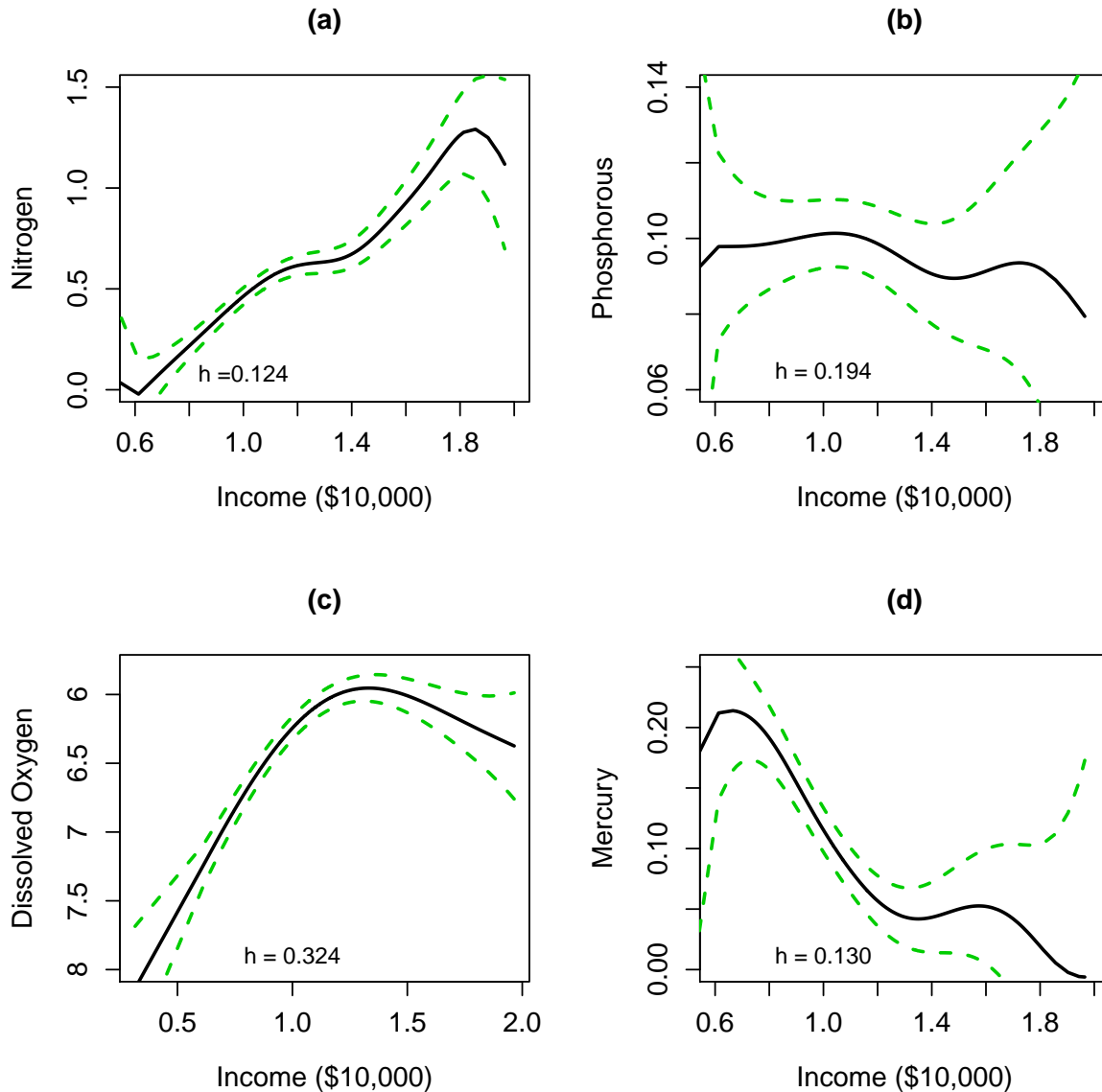


Figure 3.6. Partial plot of per capita income from semiparametric model
 Note: green color represents 95% pointwise confidence interval

3.7.3 Model Specification Test

The one way quadratic and cubic fixed effect parametric models are compared against the SUPLR model using test statistics suggested by [Hsiao, Li, and Racine \(2007\)](#). The null and alternative hypothesis considered in this model specification are as follows:

H_0 : Parametric model

H_1 : Semiparametric model

The estimated test statistics \hat{J}_n and their corresponding p -value for all conditions are reported in Table (3.7). The test statistics asymptotically follow the normal distribution $N(0, 1)$ under H_0 . [Hsiao, Li, and Racine \(2007\)](#) suggested to use the bootstrap method for approximating a finite samples null distribution of the CV-based test statistic \hat{J}_n . The p value is calculated based on the bootstrap standard error of the test statistic \hat{J}_n . Since the p -values of estimated J_n test statistics are very small for nitrogen, phosphorous, and dissolved oxygen, we rejected the null hypothesis that the parametric model is correctly specified. This finding suggests a semiparametric model is a better specified alternative for the three pollutants. In contrast, the p -values for estimated test statistics for mercury are large for both quadratic and cubic models. Hence, we fail to reject the null hypothesis that parametric model specification is correct for mercury. Since we found that the semiparametric model is significant for nitrogen, phosphorous and dissolved oxygen, we interpret results from the semiparametric model for N, P, and DO. The estimated coefficients from the parametric model are used to interpret results for mercury.

3.7.4 Nitrogen

Our results indicate that the concentration of nitrogen in water initially increased and then decreased with an increase in per capita income as shown in Figure 3.6 (a). This figure shows an approximately cubic relationship between the concentration of nitrogen pollution and per capita income. Based on this figure, we infer that the turning point for nitrogen is about \$18,000 per capita. This implies that as per capita income increases beyond this

Table 3.7. Model Specification Test between Parametric and Semiparametric Model

Pollutant	Quadratic vs Semiparametric	Cubic vs semiparametric
	J_n	J_n
Nitrogen	1.4102 (0.000)	0.6377 (0.000)
Phosphorus	20.2556 (0.000)	2.9840 (0.000)
Dissolved Oxygen	8.3317 (0.000)	7.9169 (0.000)
Mercury	-1.6698 (0.814)	-1.6742 (0.815)

income level, nitrogen pollution in the water begins to decrease. This turning point is higher than the turning point estimated by [Paudel, Zapata, and Susanto \(2005\)](#). They found the turning point for N at \$13,000, based on data from 1985 to 1998.

The estimated parameter of variables entering parametrically in the semiparametric model are shown in Table (3.8). The table shows that estimated parameters for weighted income are negative and significant, indicating that an increase in average income of the adjacent parishes decreases the level of nitrogen pollution. This result is consistent with the hypothesis that there is a spillover effect of income on nitrogen pollution in water. We also estimated the effects of population density and farmland area on all three flow-pollutants. According to the semiparametric model, the parameter estimates for population density are negative and significant, indicating an increase in population implies a decrease in nitrogen pollution. In contrast, we found a positive relationship between farmland area and nitrogen pollution.

3.7.5 Phosphorous

The estimated effect of per capita income on phosphorous pollution is shown in Figure 3.6 (b). The estimated curve shows that as per capita income increases to \$11,000, phosphorous pollution in water increases. We see a decrease in phosphorous pollution with further increases in per capita income up to \$14,000. However, the estimated curve shows the

Table 3.8. Seemingly Unrelated Regression Results

Variables	Nitrogen	Phosphorous	Dissolved Oxygen	Mercury
Income				1.3539 (0.086)
Income square				-1.6136 (0.030)
Income cube				0.4736 (0.028)
Weight income	-1.1818 (0.000)	0.0414 (0.435)	2.2100 (0.004)	0.1650 (0.307)
Population density	-1.1378 (0.000)	0.0172 (0.445)	7.6416 (0.000)	0.4372 (0.284)
Farm land area	0.0042 (0.000)	0.0025 (0.000)	0.0733 (0.000)	
Permit				0.0029 (0.046)

Note: The results for N, P, and DO are from the SUPLR model and the results for mercury are from the cubic model. Value in the parenthesis are P-values, Test of contemporaneous correlation result $LM = 22.884$, Correlation matrix =

$$\left(\begin{array}{c|ccc} & N & P & DO \\ \hline P & 0.1585 & & \\ DO & 0.0137 & -0.0638 & \\ Hg & -0.0349 & 0.01093 & -0.0015 \end{array} \right)$$

pollution increases again up to a per capita income of \$17,000 and then decreases. This finding suggests that there is no distinct EKC in the case of phosphorous. Consistent with the results of nitrogen, the coefficient of farmland area is positive and significant at the 10% level of significance. Therefore, an increase in farmland increases phosphorous pollution in the water.

3.7.6 Dissolved oxygen

A low oxygen level in water is one indicator of pollution. The EKC hypothesis for water implies that the amount of dissolved oxygen in water decreases at first and then increases

again after a certain level of income. The estimated nonparametric curve for dissolved oxygen is shown in the Figure 3.6 (c). As expected, the estimated curve shows that as per capita income increases to \$11,000, the amount of dissolved oxygen in the water falls and then rises with further increases in per capita income. The estimated coefficient of variables entering parametrically is given in Table (3.8). The estimated coefficient of weighted income is positive and significant. Therefore, an increase in the per capita income of neighboring parishes increases the level of oxygen, which is consistent with our expectation that improved economic conditions help to reduce water pollution. The estimated coefficients for population density and farmland are also positive and strongly significant. Thus, higher population helps to reduce water pollution. Our results also indicate that more farmland helps to increase oxygen levels in water. This result is inconsistent with real world observations and is not consistent with our expectations.

3.7.7 Mercury

The model specification test shows that we fail to reject null hypothesis that the cubic parametric model is correctly specified. The estimated coefficient of variables are given in Table (3.8). The coefficient for per capita income, per capita square and per capita cubic are all significant at the 5% level of significance as shown in Table (3.8). Since the cubic term is significant for the mercury equation, we describe the model parameters from the cubic model. According to these parameter estimates, the turning points for mercury are \$14953 and \$19117. This cubic shape EKC implies that it is likely to increase mercury pollution again after it declines. As a parish improves in its' per capita income, there is subsequently more demand for industrial products. Consider the case of electricity consumption. As the demand for industrial products goes up, the need for electricity increases as well. Increased electricity production from coal fired electricity generating plants may lead to higher mercury pollution. The estimated parameter for number of permits is positive and significant in the cubic model. This result suggests that if the number of point sources that emit mercury

pollution increases, mercury pollution in the water rises. This finding is consistent with our expectation.

3.8 Conclusions

In this chapter, we extended the multivariate partially linear regression model proposed by [You, Zhou, and Chen \(2013\)](#) to a seemingly unrelated partially linear model. This is an extension of a usual seemingly unrelated linear regression into the semiparametric model. A simulation study shows that the SUPLR model is unbiased and consistent in finite samples. The empirical simulations verify that the SUPLR model performs well for water pollution data. Hence, we estimated SUPLR models to examine EKC hypothesis for four water pollution parameters (N, P, DO and Hg). Model specification tests indicated that the SUPLR model performs better than a parametric model for N, P, and DO. Cubic models work well for mercury pollution.

We used disaggregated data to determine the existence of the environmental Kuznets curve for four major pollutants (N, P, DO and Mercury) in Louisiana. The SUPLR models were estimated to address the correlation between the four water pollution parameters. This study indicates an existence of parish level heterogeneity on the income-pollution relationship. We found that an inverted U shaped EKC exists for dissolved oxygen, and a cubic shaped EKC exists for nitrogen and mercury. Although the time dimension of the panel sample is small, our results suggest a need to continually assess policy effectiveness for pollution control as income increases in the state.

Other factors such as population density, farmland, spillover effects and the number of factories also affect water quality. Improvements in the economic conditions of neighboring parishes/counties have positive effect on environmental quality, as they reduce nitrogen pollution and increase oxygen levels in the water. Dense populations improve environmental quality as people demand better environmental quality with the rise in income. An increase

in farmland area in a parish does not improve water quality. Over the time period 1985-2006, Louisiana showed continuous growth in its per capita GDP. At the same time, contribution to statewide GDP from agricultural industry has shrunk while industries' contribution has been increasing.

An important question that can be asked here is, whether or not economic progress is the panacea to environmental quality improvement? Our results show that a higher level of economic growth decreases nitrogen pollution and increases dissolved oxygen levels in waterbodies. These indicate flow pollutant levels decrease with increases in per capita income. For stock pollution, economic growth improves pollution levels up to a certain threshold level but after that level is reached, pollution levels begin to rise again. Therefore, our conclusion is that higher economic growth could be the solution for the flow pollutants, but not for the case of stock pollutants. Likewise [Grossman and Krueger \(1995\)](#) and [Dasgupta et al. \(2002\)](#), environmental regulations are perhaps needed to control environmental degradation especially for stock pollutants.

One may question whether or not economic growth alone is responsible for water quality improvement in Louisiana. While we cannot disagree on other factors causing water quality improvement, we cannot disentangle and quantify the induced effect such as the effects of water quality regulations over the study period. Although water quality regulations specifically related to nonpoint source pollution (flow pollutants studied here) do not exist, several incentive structures do exist that have been implemented in an effort to improve water quality. These programs which have been implemented by the USDA/NRCS include the Environmental Quality Incentive Program (EQIP) and the Conservation Reserve Program (CRP). The environmental benefit of these programs are well documented ([Feather, Hellerstein, and Hansen, 1999](#); [Paudel et al., 2008](#)). In addition to these programs, Total Maximum Daily Load (TMDL) implemented by the U.S. Environmental Protection Agency has helped to improve water quality. [Paudel and Schafer \(2009\)](#) indicated social capital in Louisiana over the period 1985-1998, as measured by the existence of various clubs and

social groups, exhibited a steady increase. Social capital also exerts pressure on improving environmental quality. It is very likely that all these factors have worked in unison to improve water quality thus reducing the negative impacts stemming from flow pollutants in Louisiana.

CHAPTER 4.

ENVIRONMENTAL KUZNETS CURVE FOR WATER QUALITY PARAMETERS AT THE GLOBAL LEVEL: SEMIPARAMETRIC AND NONPARAMETRIC APPROACHES

4.1 Introduction¹

The Environmental Kuznets Curve (EKC) is a relationship between income and pollution which is hypothesized to have an inverted U-shape. The idea of a Kuznets curve with an inverted U-shape stems from previous work in income equality ([Kuznets, 1955](#)). The EKC hypothesis states that as income increases pollution goes up initially but after some income level pollution declines. The point at which pollution level is the highest is called the turning point.

Research on the validity, application, and measurement of the EKC has been expanding rapidly for several types of pollution as shown in Table (4.1). Critics have challenged both the findings and policy implications of these studies ([Dasgupta et al., 2002](#); [Stern, 2004](#)). Research has found a positive relationship between CO₂ and per capita income (a flow by some definitions), in place of any inverted (or non-inverted) curve. More specifically, the EKC holds not for any specific pollutant but rather for different pollutants, in different ways, depending on the choice of the pollutant, study area, and time period.

Traditionally, in the EKC relationship, the dependent variable is pollution level and the independent variables are income and various polynomial specifications of income, primarily those of quadratic and cubic forms. Several authors ([Millimet, List, and Stengos, 2003](#); [Paudel, Zapata, and Susanto, 2005](#); [Paudel and Schafer, 2009](#); [Zapata and Paudel, 2009](#)) have refuted the parametric forms and suggested a need to include a nonparametric form of

¹Some portions of this chapter draw material from the manuscript with Pandit as a coauthor. The reference to that article is: [Paudel, Lin, and Pandit \(2011\)](#).

Table 4.1. EKC Literature by Types of Pollutions

Pollution type	Literature
Air pollution	List and Gallet (1999); Heerink, Mulatu, and Bulte (2001); Bruvoll and Medin (2003); Cole (2004); Deacon and Norman (2006); Merlevede, Verbeke, and Clercq (2006); Akbostanci, Turut-Asik, and Tunc (2009); Kim (2013)
Water pollution	Jha and Murthy (2003); Paudel, Zapata, and Susanto (2005); Paudel and Schafer (2009); Paudel and Poudel (2013); Lin and Liscow (2013)
Deforestation	Heerink, Mulatu, and Bulte (2001); Rodriguez-Meza, Southgate, and Gonzalez-Vega (2004); Barbier (2004); Culas (2007); Van and Azomahou (2007)
Hazardous waste and toxin	Gawande, Berrens, and Bohara (2001); Rupasingha et al. (2004)
Energy consumption	Luzzati and Orsini (2009); Phu (2010)
Carbon dioxide	Cavlovic et al. (2000); Dasgupta et al. (2002); Copeland and Taylor (2004); Plassmann and Khanna (2006); Azomahou, Laisney, and Van (2006); Paudel and Schafer (2009); Melenberg, Vollebergh, and Dijkgraaf (2011); Stern (2004); Poudel, Paudel, and Bhattarai (2009); Azomahou, Laisney, and Van (2006); Kim (2013)

income in the regression model. These nonparametric or semiparametric regression models were found to perform better than parametric forms in specification tests.

EKC relationship may be observed because of social capital, political rights and civil liberties (Paudel and Schafer, 2009; Paudel and Poudel, 2013; Lin and Liscow, 2013). Some economists suggest that these are very important omitted variables. For example, Grossman and Krueger (1995) speculate that “the strongest link between income and pollution in fact is via an induced policy response”, and that these policies are, in turn, induced by popular demand. According to this line of reasoning, impoverished countries, at first, have so little development that they have high environmental quality. Then, countries’ environments degrade as they develop and become richer. Finally, they reach a point at which environmental quality is poor enough and the people are rich enough that they begin

to desire to pay for improvements in environmental quality. At this point, they begin to demand changes from their government, and environmental degradation decreases. Similarly, [Dasgupta and Mäler \(1995\)](#) indicate that political rights and civil liberties are important components in protecting environmental rights. [Barrett and Graddy \(2000\)](#) find that, for many pollution variables, “political reforms may be as important as economic reforms in improving environmental quality worldwide” (p. 433). However, they also find an absence of significant results for some pollution variables, which suggests that something other than an induced policy response may be affecting pollution levels. [Lin and Liscow \(2013\)](#) found that political institutions have a significant effect on environmental quality for five of the eleven water pollutants examined. [Torras and Boyce \(1998\)](#) hypothesized that changes in the distribution of power underlie the EKC relationship, and find that literacy, political rights and civil liberties have particularly strong effects on environmental quality in low-income countries. [Farzin and Bond \(2006\)](#) develop and estimate an econometric model of the relationship between several local and global air pollutants and economic development while allowing for critical aspects of the sociopolitical-economic regime of a state.

A related concept to political institutions that may need to be accounted for in the EKC relationship is social capital. Social capital is defined as shared norms, trust, and social networks that facilitate coordination and cooperation for mutually beneficial collective action. [Paudel and Schafer \(2009\)](#) and [Paudel et al. \(2011\)](#) include a social capital index in the EKC model. An example of social capital is “most people can be trusted”. The relationship between economic growth and trust as a measure of social capital has been studied by many authors. For example, [Zak and Knack \(2001\)](#) developed a general equilibrium growth model using trust and found that trust significantly influenced growth rate. [Dincer and Uslander \(2010\)](#) found that there is positive relationship between trust and growth. According to [Dincer and Uslander \(2010\)](#), the GDP increases by 0.5% for every 10% point increase in trust. In addition, it is found that trust is an important factor for high environmental quality. To illustrate, [Rosser and Rosser \(2006\)](#) indicates that high environmental quality depends on

the levels of trust within the society. Social capital such as trust applied to environmental stewardship impact on national environmental performance ([Grafton and Knowles, 2004](#)).

Researchers have used population density, democracy, political rights, openness of countries, etc. as additional variables in the model. [Israel and Levinson \(2004\)](#) use a different tactic in their attempt to discover the political mechanisms of the EKC. They try to extrapolate people's marginal willingness to pay (MWTP) for environmental protection from international survey data obtained from the World Value Survey. They found little relationship between the MWTP and economic development. This suggests that neither technological nor institutional constraints explain the inverted-U shaped pollution-income path or that their data were inadequate.

Although the literature on estimating the environmental Kuznets curve is growing fast and becoming very sophisticated in terms of empirical methodology used, hitherto articles in the EKC literature have not properly addressed the properties of categorical variables in the model. One problem that arises in incorporating political rights, civil liberties and trust variables or any other categorical, ordered or binary variables in a semiparametric or nonparametric regression is that those cannot be treated as continuous variables. We analyze the relationship between water quality and per capita income at the global level for the years 1980-1998. We identify the roles played by political rights, civil liberties, and trust in determining water quality.

4.2 Methods

We are interested in identifying how different types of water pollutants relate to income, civil liberties, political rights and trust². The effects of income and other factors that affect water quality can be expressed in a regression model. Let P represents pollution in a country, y

²A theoretical basis for the EKC can be found in recent papers by [Brock and Taylor \(2010\)](#) and [Acemoglu et al. \(2010\)](#). Our focus is on the empirical model.

represents per capita GDP of that country and X represents other factors. We used the following method to estimate the effect of y and X on P .

4.2.1 Parametric Methods

Generally, the EKC relationships among these variables are studied using a parametric model with an income variable regressed in a polynomial form (quadratic or cubic) and other factors are in linear forms. The parametric regression model is given as in equation (4.1).

$$P = Y\beta + X\alpha + \epsilon \quad (4.1)$$

where, $Y = (y, y^2)$ if we consider a quadratic, and $Y = (y, y^2, y^3)$ if we consider a cubic model; y is per capita GDP; β is the vector representing parameter for corresponding variables. $X = (x_1, x_2, \dots, x_p)$ represents factors that affect pollution other than per capita income such as civil liberties, political rights and trust; α are parameters corresponding to these factors; ϵ is i.i.d. with zero mean, finite variance σ_ϵ^2 . Parametric regression equation (4.1) is estimated using least squares estimation procedures.

4.2.2 Nonparametric Methods

Parametric methods put a priori restrictions on how the relationship should look in empirical research. One of the alternatives in relaxing the assumption of parametric methods is to utilize either nonparametric or semiparametric regression techniques that allow more flexibility in modeling. In addition, nonparametric estimates are more robust and detect structures which sometimes remain undetected by traditional parametric estimation techniques. Although the semiparametric or nonparametric method is tedious in terms of computing resources, this method is used by many researchers ([Schmalensee, Stoker, and Judson, 1998](#); [List and Gallet, 1999](#); [Millimet, List, and Stengos, 2003](#); [Roy and van Kooten, 2004](#); [Paudel, Zapata, and Susanto, 2005](#); [Bertinelli and Strobl, 2005](#); [Azomahou, Laisney,](#)

and Van, 2006; Van and Azomahou, 2007; Criado, 2008; Luzzati and Orsini, 2009). The nonparametric regression model is given in equation (4.2),

$$P = g(y) + \sum_j^p g_j(x_j) + \epsilon \quad (4.2)$$

where, $g(\cdot)$ is an unknown smooth function for y , i.e. income, and $g_j(\cdot)$ is the unknown function for other factors such as civil liberties, political rights and trust. Variables civil liberties and political rights are ordinal, and trust is a categorical variable. Thus we need an estimation procedure that can address both ordinal and categorical variables. Recently, Ma and Racine (2013), Nie and Racine (2012) and Ma, Racine, and Yang (2011) have developed a nonparametric estimation procedure to address ordinal and categorical variables in a nonparametric model. We used a method suggested by them to estimate a nonparametric model given in equation (4.2)³.

4.2.3 Semiparametric Methods

When there are a large number of observations and explanatory variables, nonparametric methods encounter a problem known as the curse of dimensionality. A semiparametric method can correct the weaknesses of the parametric and nonparametric methods because it balances the pros and cons of the parametric and nonparametric methods (Pandit, Paudel, and Mishra, 2013). Like nonparametric methods, the nonparametric components in a semiparametric method are distribution free, so a strong assumption of the functional form is not required. The semiparametric regression model also refers to an additive or generalized additive model (GAM). A semiparametric regression model contains both nonparametric and parametric components and is expressed as in equation (4.3).

$$P = g(y) + X\alpha + \epsilon \quad (4.3)$$

³A ‘crs’ R package is available to estimate the nonparametric model which contains both categorical and continuous variables. See Racine and Nie (2012) for ‘crs’ package manuel.

First term $g(\cdot)$ is an unknown smooth function for variables entering nonparametrically, and the second term $X\alpha$ is the component for variables entering parametrically. A penalized smoothing spline estimation procedure is used to estimate equation (4.3). Parametric model matrix X also includes a column of ones for the intercept variable, and α is a parameter vector. In our case, y is a pollution variable and X denotes a matrix of independent variables such as civil liberties, political rights and trust. Table (4.2) includes summary statistics of these variables. The vector y represents variable economic growth whose functional form cannot be specified. These variables enter the model nonparametrically. In equation (4.3), the variable X is assumed to have a linear effect. This model can be analyzed by using a penalized likelihood maximization procedure suggested by [Wood \(2006\)](#), [Hastie and Tibshirani \(1990\)](#), and [Ruppert, Wand, and Carroll \(2003\)](#).

4.2.4 Model Specification Test

Existing studies have proposed several test statistics to compare the suitability of different functional forms ([Hong and White, 1995](#); [Fan and Li, 1996](#); [Zheng, 1996](#)). We used the likelihood ratio or contrasting deviance test suggested by ([Ruppert, Wand, and Carroll, 2003](#), p. 168) to test the parametric versus semiparametric model. The null and alternative hypotheses are

H_0 : Parametric Model

H_1 : Semiparametric Model

The log likelihood ratio (LR) test or contrasting deviance statistic is

$$LR = -2(L_0 - L_1), \tag{4.4}$$

where L_0 is the log likelihood of the parametric model and L_1 is the log likelihood of the semiparametric model. The test statistics under the null hypothesis follow an approximate χ^2 distribution, and the degrees of freedom equal the difference in the number of parameters across the two models. If the observed LR value falls within the upper tail of a chi-square

distribution, then we conclude that the null hypothesis of the parametric model specification should be rejected (Ruppert, Wand, and Carroll, 2003). Since this test statistic cannot be used for the nonparametric model suggested by Ma, Racine, and Yang (2011), we used a cross validation (CV) score to compare the nonparametric model with the parametric and semiparametric model. The model which has a smaller CV value is better (Racine and Nie, 2012).

4.3 Data

Water pollution data comes from the Global Environment Monitoring System (GEMS) Water Dataset, which consists of triennial surveys of water quality statistics from 1979 to 1999 from sixty-eight developed and developing countries⁴. The GEMS data set consist of over 70,000 observations of dozens of different types of water pollution, providing a substantive amount of data on varied measures of water quality. Each data point consists of the average over three years of one or more data point from one of GEMS/water's hundreds of sites around the world. This data set also has several drawbacks. First, the variety of measures seems conducive to a study that fails to appreciate the unique dynamics that govern each different pollutant and takes data as numbers without a great deal of meaning. Second, the data can be rather spotty; providing observations in all seven triennial surveys for cases in only a few countries. If we construct panel data, we face very few observations useful for analysis. We mitigate this problem by choosing the conventional data, and analyze the pollution-income relationship for each pollutant assuming homogeneity across these countries.⁵

⁴The countries used in this research are Argentina, Australia, Austria, Bangladesh, Belgium, Bolivia, Brazil, Cambodia, Canada, Chile, China, Colombia, Cuba, Denmark, Ecuador, Egypt, Fiji, Finland, France, Germany, Ghana, Greece, Guatemala, Hong Kong, Hungary, India, Indonesia, Iran, Ireland, Israel, Italy, Japan, Jordan, Kenya, Korea, Laos, Lithuania, Luxembourg, Malaysia, Mali, Mexico, Morocco, Netherlands, New Zealand, Norway, Pakistan, Panama, Peru, Philippines, Poland, Portugal, Russian Federation, Senegal, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Tanzania, Thailand, Tunisia, Turkey, Uganda, United Kingdom, United States, Uruguay, Vietnam and Zaire.

⁵We understand the drawback of assuming homogeneity across these different countries. Unfortunately, lack of sufficient data for all seven triennial years for all 68 countries prevented us from forming convergence groups and running regression models for each convergence group as has been done by Panopoulou and

This chapter focuses on four types of water pollutants: heavy metal (nickel, mercury, arsenic, cadmium, lead), pathogenic contamination (fecal coliform, total coliform), oxygen regime (dissolved oxygen (DO), chemical oxygen demand (COD), biological oxygen demand (BOD)) and nutrients (nitrate)⁶. The sources of these water quality parameters are provided in Appendix (B). All data are in the form of concentrations of mg/l except for the mercury data, which is in the form of $\mu\text{g}/\text{l}$ and the coliform data, which is in the form of measured count/100 ml. The year assigned to each data point is the middle of the three years. To this data, we added data on gross domestic product (GDP) and per capita purchasing power parity in constant 2000 international dollars from the World Development Indicators (WDI). For data on political mechanisms, we use indices on political rights (PR) and civil liberties (CL) from Freedom House. Each index varies from 1 to 7, with 1 meaning the most political rights or civil liberties. For example, the United States has a 1 in each category in all years, Indonesia, which has recently been in the middle of the range, and China, which has 7 in both categories for most years. Freedom House attempts to use a methodology not bound by culture, but instead uses standards drawn from the Universal Declaration of Human Rights (Freedom House, 2010). Political rights measure factors like the fairness of the electoral process, the degree of political pluralism and participation, and the presence of a non-corrupt and transparent government (Freedom House, 2010). Civil liberties measure freedom of expression and beliefs, the ability to associate, the rule of law, and the degree of individual autonomy. The mean of the political rights variables is lower than that for civil liberties, which implies that political rights are more prevalent in many countries than civil liberties are. The data on trust (TR) is obtained from the World Value Survey (WVS)⁷, the measure of trust used herein is the frequency of respondents in each country agreeing that

Pantelidis (2009) in the case of CO_2 pollutant. This could be a subject of further research provided sufficient data are available. Such studies could help us to formulate more policy related pollution control as has been done by Mazzanti and Musolesi (2011).

⁶Although existence/nonexistence of an EKC for some of these pollutants for different time periods and different sets of counties has been established, change in the data period, and additional variables in the regression may give different results. This is exactly the point raised by Harbaugh, Levinson, and Wilson (2002).

⁷The full data set can be obtained from <http://www.wvsevsdb.com/wvs/WVSDData.jsp?Idioma=I>

‘most people can be trusted’ against the alternative that ‘you can’t be too careful in dealing with people’. This data consists of the value 2 for the first category and 1 for the second category. Summary statistics of the data used are presented in Table 4.2. Most pollutants exhibit a large range in values and a high standard deviation.

4.4 Results

We estimated parametric, nonparametric and semiparametric models of the income-pollution relationship, which are given in equations (4.1), (4.2), and (4.3), respectively. The detailed results of each model are given in separate sections. In addition to economic growth (measured by GDP), we used civil liberties, political rights and trust variables as a measure of social capital. Trust variable are not available for all countries where political rights are available, so we defined two types of models below.

$$\text{Model 1: Pollution} = f(\text{GDP, CL, PR}) \quad (4.5)$$

$$\text{Model 2: Pollution} = f(\text{GDP, TR}) \quad (4.6)$$

Where pollution is a function of GDP, civil liberties, and political rights for Model 1. Pollution is a function of GDP and trust for model 2. Due to data limitations, we estimated Model 2 only for BOD, DO, and fecal coliform. Model 1 is estimated for all nine pollutants.

4.4.1 Model Specification Test

Results obtained from both quadratic and cubic parametric models are compared utilizing the method suggested by [Ruppert, Wand, and Carroll \(2003\)](#). This method does not allow us to compare these models when categorical variables are entered nonparametrically. In our study, civil liberties and political rights are ordinal variables, and trust is a categorical variable. We used a method suggested by [Racine and Nie \(2012\)](#) to compare the nonparametric model with the semiparametric and parametric models. According to [Racine and Nie \(2012\)](#), a

Table 4.2. Summary Statistics of Variables

Variable		Mean	Std. Dev.	Min	Max	Observations
Mercury	overall	0.3422	0.7287	0.0000	5.0000	N = 129
	between		0.4774	0.0000	2.8723	n = 39
	within		0.4739	-2.1684	2.4699	T-bar = 3.31
Arsenic	overall	0.0126	0.0479	0.0000	0.4283	N = 80
	between		0.0812	0.0000	0.4283	n = 27
	within		0.0054	-0.0083	0.0334	T-bar = 2.96
Cadmium	overall	0.0163	0.0952	0.0000	0.7875	N = 137
	between		0.0362	0.0000	0.1373	n = 39
	within		0.0868	-0.1210	0.6668	T-bar = 3.51
Lead	overall	0.0237	0.0694	0.0000	0.5000	N = 113
	between		0.1260	0.0000	0.5000	n = 29
	within		0.0181	-0.0315	0.1289	T-bar = 3.90
Fecal coliform	overall	2.8792	12.0671	0.0000	89.6496	N = 223
	between		9.1655	0.0000	44.8729	n = 56
	within		8.2337	-41.8483	51.1483	T-bar = 3.98
Total coliform	overall	3.5691	10.3190	0.0000	62.0000	N = 125
	between		8.8917	0.0000	49.3333	n = 41
	within		6.8787	-16.5596	40.1443	T-bar = 3.05
Dissolved oxygen	overall	8.6062	5.2346	0.1500	84.6667	N = 282
	between		2.5040	4.1259	20.0561	n = 67
	within		4.5718	-7.9899	73.2168	T-bar = 4.21
COD	overall	23.7759	39.0176	0.5000	393.4000	N = 164
	between		32.8090	1.4616	184.8000	n = 51
	within		27.7668	-104.7735	268.9265	T-bar = 3.22
BOD	overall	3.8184	7.9852	0.1500	74.3333	N = 226
	between		5.6525	0.6378	30.0000	n = 55
	within		6.2553	-20.1817	48.1517	T-bar = 4.11
GDP	overall	11.6042	9.4384	0.4977	40.1655	N = 265
	between		9.1230	0.4986	33.8326	n = 60
	within		1.6839	0.6757	17.9371	T-bar = 4.42
Civil liberties	overall	3.0803	1.8891	1.0000	7.0000	N = 274
	between		1.8109	1.0000	7.0000	n = 64
	within		0.5785	1.2231	5.2231	T-bar = 4.28
Political rights	overall	2.9380	2.0930	1.0000	7.0000	N = 274
	between		1.9850	1.0000	7.0000	n = 64
	within		0.7890	-0.0620	6.2237	T-bar = 4.28
Trust	overall	1.6875	0.4709	1.0000	2.0000	N = 32
	between		0.4035	1.0000	2.0000	n = 21
	within		0.3111	1.0208	2.3542	T-bar = 1.52

model with a low cross validation score⁸ is preferred. The results of the likelihood ratio or deviance model specification tests are given in Tables (4.3) and (4.4). Tables (4.3) and (4.4) give null and alternative hypotheses, deviance value, degree of freedom and p-value of the test statistics. If the p-values are less than 10% level of significance, it indicates that the null of the parametric model is rejected, and we accept the alternative hypothesis that the semiparametric model is correctly specified. The results show that the semiparametric model is correctly specified for fecal coliform, mercury, arsenic, lead, and COD compared to the quadratic model. The semiparametric model is correctly specified for mercury, arsenic, and lead compared to the cubic model. We fail to reject the cubic model for fecal coliform, cadmium, COD, BOD and total coliform. Similarly, we fail to reject a quadratic model for DO. Table (4.4) shows that the semiparametric model is preferred for BOD and DO compared to the quadratic model. The cubic model fails to reject for BOD and fecal coliform. The cross validation (CV) score is given in Table (4.5) and Table (4.6) for Model 1 and Model 2 respectively. Based on the CV score, we found that, except for arsenic, nonparametric model is preferred for all water quality parameters.

4.4.2 Parametric Results

We present the results of the relationship between water quality and per capita GDP (linear, square, cubic forms), political rights and civil liberties. For brevity, results are presented only if those variables are significant in the parametric results. The estimated coefficients of Model 1 and Model 2 are provided in Tables (4.7) and (4.8) respectively. The quadratic model is a polynomial model of GDP with two degrees and the cubic model is a polynomial of GDP with 3 degrees.

⁸The cross validation score represents optimal leave one out cross validation score. The leave one out cross validation for parametric model is equivalent to $1/N \sum_i \epsilon_i^2 / (1 - h_{ii})^2$, where ϵ is residual from the model and h_{ii} is diagonal elements of hat matrix.

Table 4.3. Model Specification Test for Model 1

Pollutants	H_0	H_1	Deviance	DF	P-value
Fecal coliform	Quadratic	Semiparametric	0.0000	0.0001	0.0008
	cubic	Semiparametric	16.9010	1.0001	0.7405
Mercury	Quadratic	Semiparametric	6.4785	5.5790	0.0415
	cubic	Semiparametric	5.6000	4.5790	0.0435
Arsenic	Quadratic	Semiparametric	0.0607	6.5624	0.0000
	cubic	Semiparametric	0.0544	5.5624	0.0000
Cadmium	Quadratic	Semiparametric	0.0000	0.0000	-
	cubic	Semiparametric	0.0041	1.0000	0.5346
lead	Quadratic	Semiparametric	0.0845	5.0776	0.0001
	cubic	Semiparametric	0.0897	6.0776	0.0001
Oxygen	Quadratic	Semiparametric	1.0001	1.4002	0.8317
	cubic	Semiparametric	0.0000	0.0001	0.0008
COD	Quadratic	Semiparametric	7068.5000	1.7019	0.0756
	cubic	Semiparametric	7157.6000	2.7019	0.1647
BOD	Quadratic	Semiparametric	1.0003	18.7280	-
	cubic	Semiparametric	0.0000	0.0003	0.6082
Coliforms	Quadratic	Semiparametric	0.0000	0.0000	-
	cubic	Semiparametric	12.0360	1.0000	0.7474

Note: H_0 = null hypothesis, H_1 = alternative hypothesis, and DF = degree of freedom

Table 4.4. Model Specification Test for Model 2

Pollutant	H0	H1	Deviance	P-value
BOD	Quadratic	Semiparametric	11.400	0.097
	cubic	Semiparametric	2.990	0.246
DO	Quadratic	Semiparametric	-11.400	0.098
	cubic	Semiparametric	2.998	0.246
Fecal coliforms	Quadratic	Semiparametric	-22.662	0.115
	cubic	Semiparametric	10.831	0.268

Table 4.5. Cross Validation Score for Model 1

Pollutants	Quadratic	Cubic	Semiparametric	Nonparametric
Fecal coliform	165.6900	167.3900	165.6900	147.1333
Mercury	0.6251	0.6274	0.6197	0.5181
Arsenic	0.0027	0.0027	0.0020	0.0023
Cadmium	0.0118	0.0119	0.0118	0.0096
lead	0.0046	0.0046	0.0040	0.0036
DO	32.8830	33.1670	32.8830	30.0379
Cod	1730.7000	1756.1000	1740.8000	1524.7674
Bod	76.6370	77.3570	76.6360	70.6316
Coliforms	131.8200	134.2800	131.8200	113.0197

Table 4.6. Cross Validation Score for Model 2

Pollutant	Quadratic	Cubic	semiparametric	Nonparametric
BOD	8.2235	8.1981	8.0483	6.6566
DO	6.7018	7.1178	6.7034	5.8712
Fecal coliforms	23.552	23.873	23.2770	7.5438

4.4.3 Nonparametric Results

In the nonparametric model, we entered all variables nonparametrically. As in the previous sub-section, we used GDP, civil liberties and political rights as explanatory variables for Model 1 and GDP and trust as explanatory variables for Model 2. We plotted the effect of all variables on different pollutants in Figure (4.1) and (4.2) for Models 1 and 2, respectively.

4.4.4 Semiparametric Results

Following previous research in EKC, we entered Gross Domestic Product (GDP) nonparametrically in a semiparametric model. Civil liberties and political liberties for Model 1 and trust for Model 2 are entered parametrically in the semiparametric models. A penalized spline semiparametric regression model is estimated for each pollutant. The estimated parameters of variables entering parametrically in the semiparametric model are given in Table (4.7) and Table (4.8) for Model 1 and Model 2 respectively. We use a partial regression plot to see the effect of the variable entering as a nonparametric component (i.e. GDP) as

Table 4.7. Estimated Coefficient from Parametric Model 1

Pollutant	Model	GDP	GDP ²	GDP ³	Civil liberties	Political rights
Fecal coliform	Quadratic				-17.551[5] (0.064)	6.6454 (0.035)
	Cubic				-17.935[1] (0.061)	6.65467[1] (0.035)
	Semiparametric				-17.550 [1] 0.064	6.6454[6] 0.035
Arsenic	Cubic	-0.02674 0.021	0.01554 0.049	-0.00285 0.074	0.037[6] (0.062)	
	Semiparametric				-0.121[1] (0.011) -0.0508[2] (0.069) 0.051[3] (0.046) 0.103[4] (0.000) 0.089[5] (0.000) 0.044[6] (0.011)	
Lead	Quadratic					-0.055[6] (0.026)
	Cubic				-0.1203[1] (0.090)	-0.0536[6] (0.029)
	Semiparametric				-0.084[2] (0.045)	0.117[1] (0.094) 0.079[4] (0.023)
Bod	Quadratic					-4.991[6] (0.012)
	Cubic					-5.028[6] (0.012)
	Semiparametric					-4.991[6] (0.012)
Coliforms	Quadratic				-5.559[6] (0.062)	
	Cubic				-5.586[6] (0.062)	
	Semiparametric				-5.559[6] (0.062)	

Note: Only significant parameter values are shown in the table. For the semiparametric model, only parametric variables have estimated coefficients. P-values are shown inside parentheses. The number in the bracket indicate the level of civil liberties or political rights.

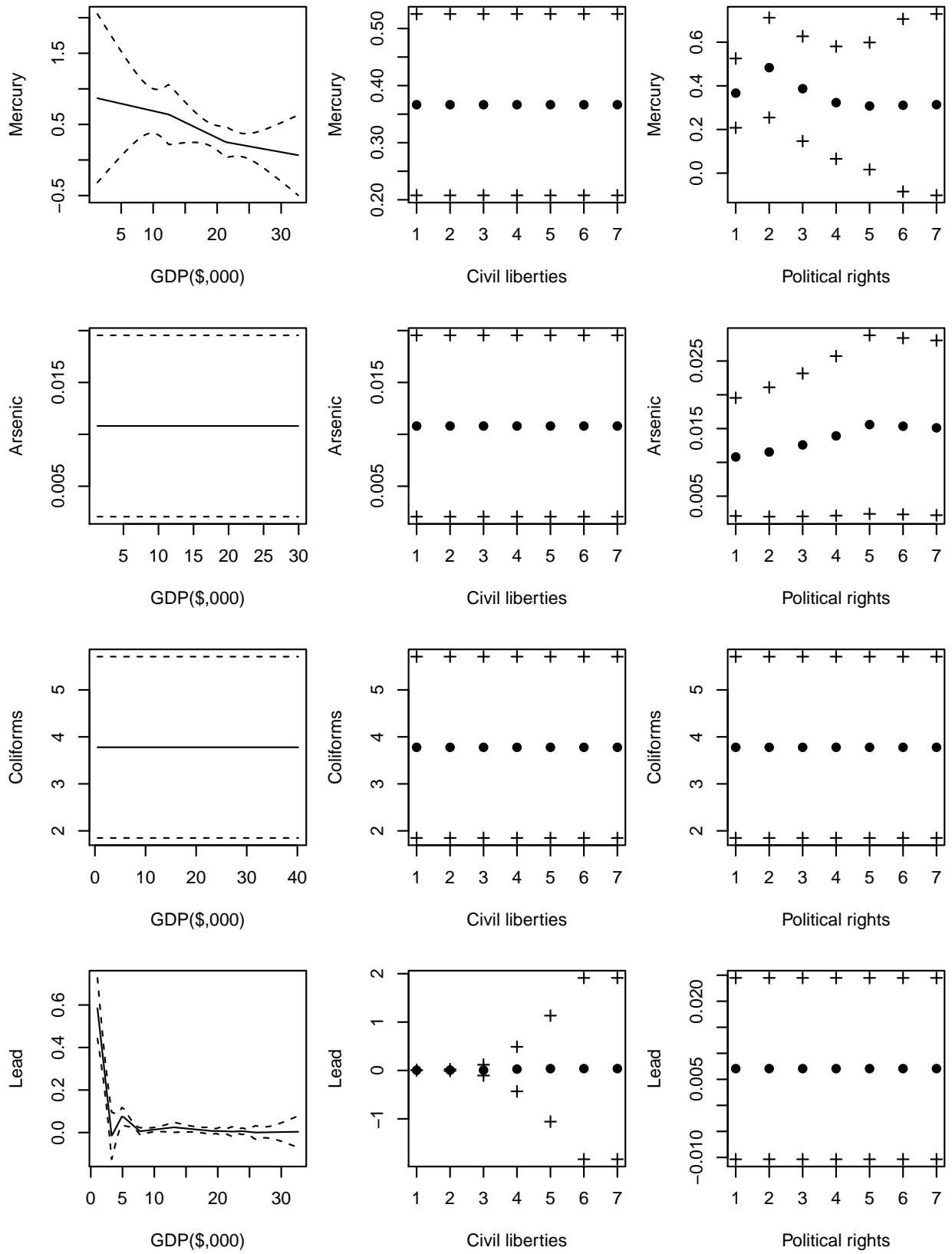


Figure 4.1. Partial Regression Plot from Nonparametric Model 1.

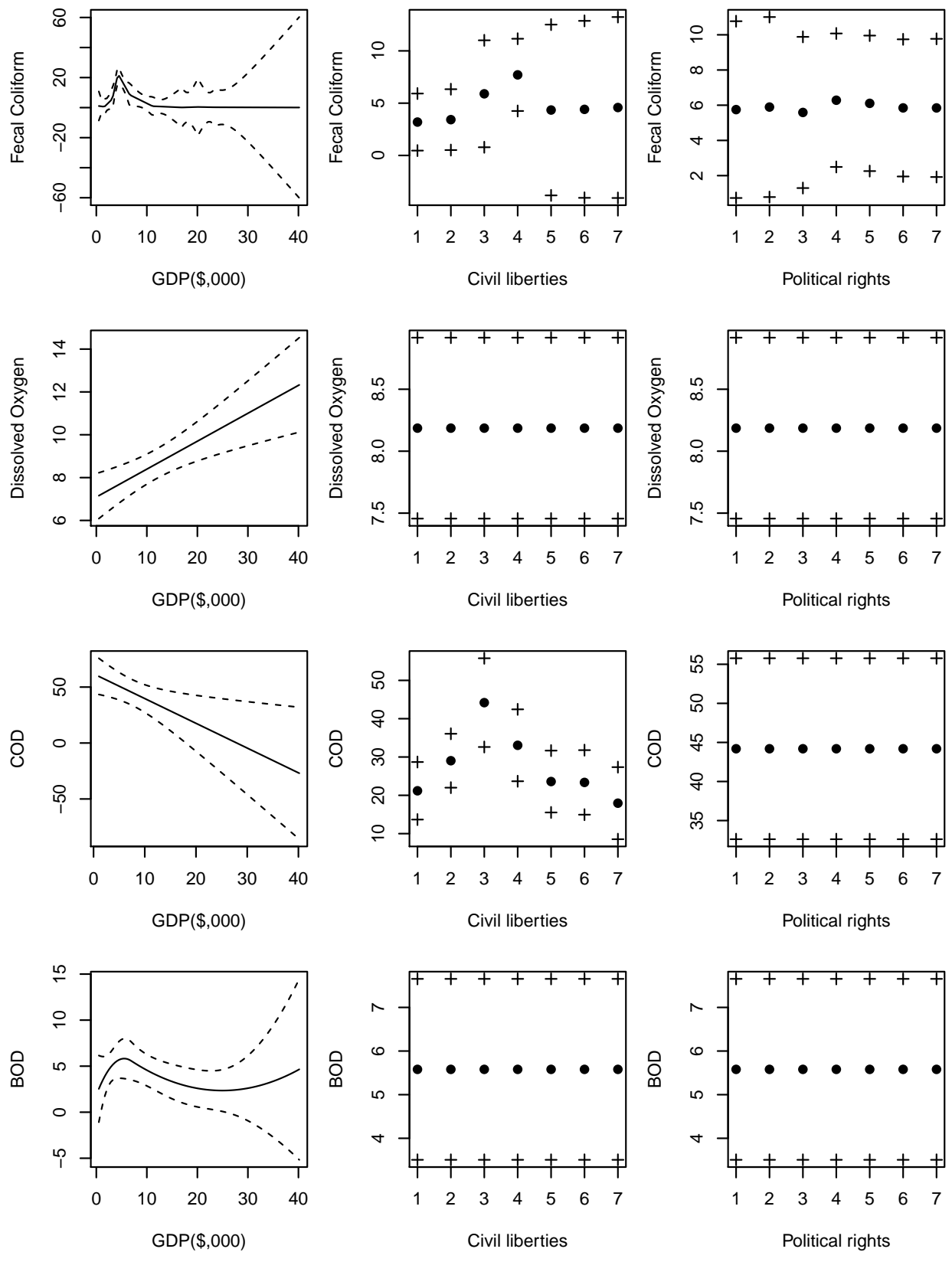


Figure 4.1. Contd.

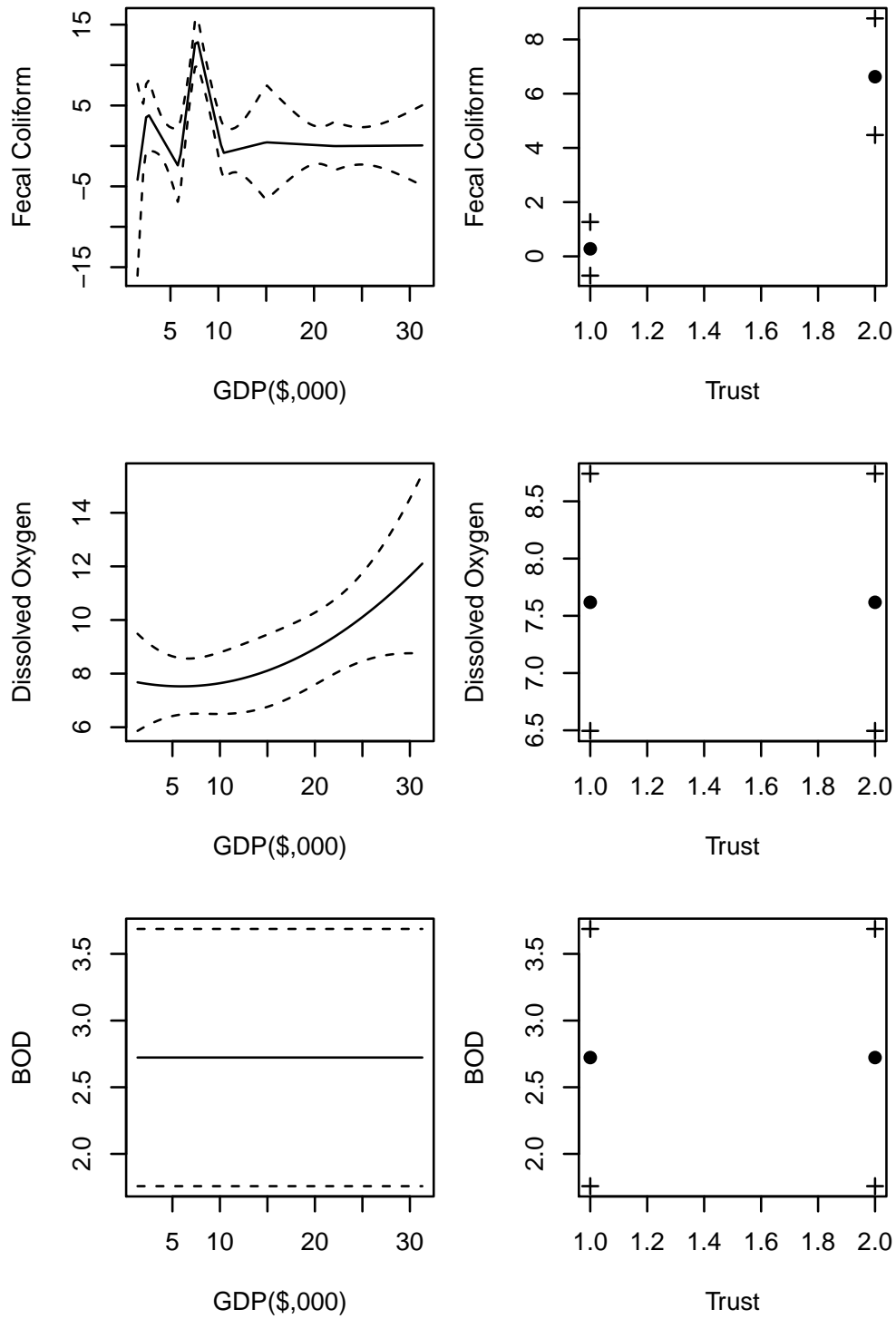


Figure 4.2. Partial Regression Plot from Nonparametric Model 2.

Table 4.8. Estimated Coefficient from Parametric Model 2

Pollutant	Model	GDP	GDP ²	GDP ³	Trust
Fecal coliform	Quadratic	0.061 (0.772)	-0.079 (0.340)		3.044 (0.058)
	Cubic	1.083 (0.118)	-0.916 (0.084)	0.182 (0.077)	2.630 (0.070)
	Semiparametric				2.866 (0.173)
DO	Quadratic	0.019 (0.912)	0.028 (0.613)		-0.890 (0.259)
	Cubic	-0.329 (0.463)	0.311 (0.349)	-0.061 (0.358)	-0.721 (0.339)
	Semiparametric				-0.260 (0.786)
BOD	Quadratic	0.062 (0.733)	-0.029 (0.628)		-0.331 (0.809)
	Cubic	0.717 (0.217)	-0.574 (0.190)	0.119 (0.181)	-0.805 (0.604)
	Semiparametric				-0.260 (0.786)

well as parametric component in the semiparametric model. The fitted nonparametric curves are given in Figures (4.3) and (4.4) for Model 1 and Model 2, respectively.

4.4.5 Fecal Coliform

For fecal coliform, we found that either the semiparametric or cubic models perform better than a quadratic model. However, we did not find GDP-related variables to be significant. The semiparametric plot (Figure 4.1) shows that there is an inverted U- shaped relationship between fecal coliform and GDP up to GDP \$1,000, and there is linear relationship for higher values of GDP. This figure also shows that pollution of fecal coliform is higher for low and high values of civil liberties but high for medium values of civil liberties. This indicated the presence of an EKC for fecal coliform with respect to civil liberties. When we use trust in our model results show the existence of an EKC for fecal coliform as shown in Figure (4.4).

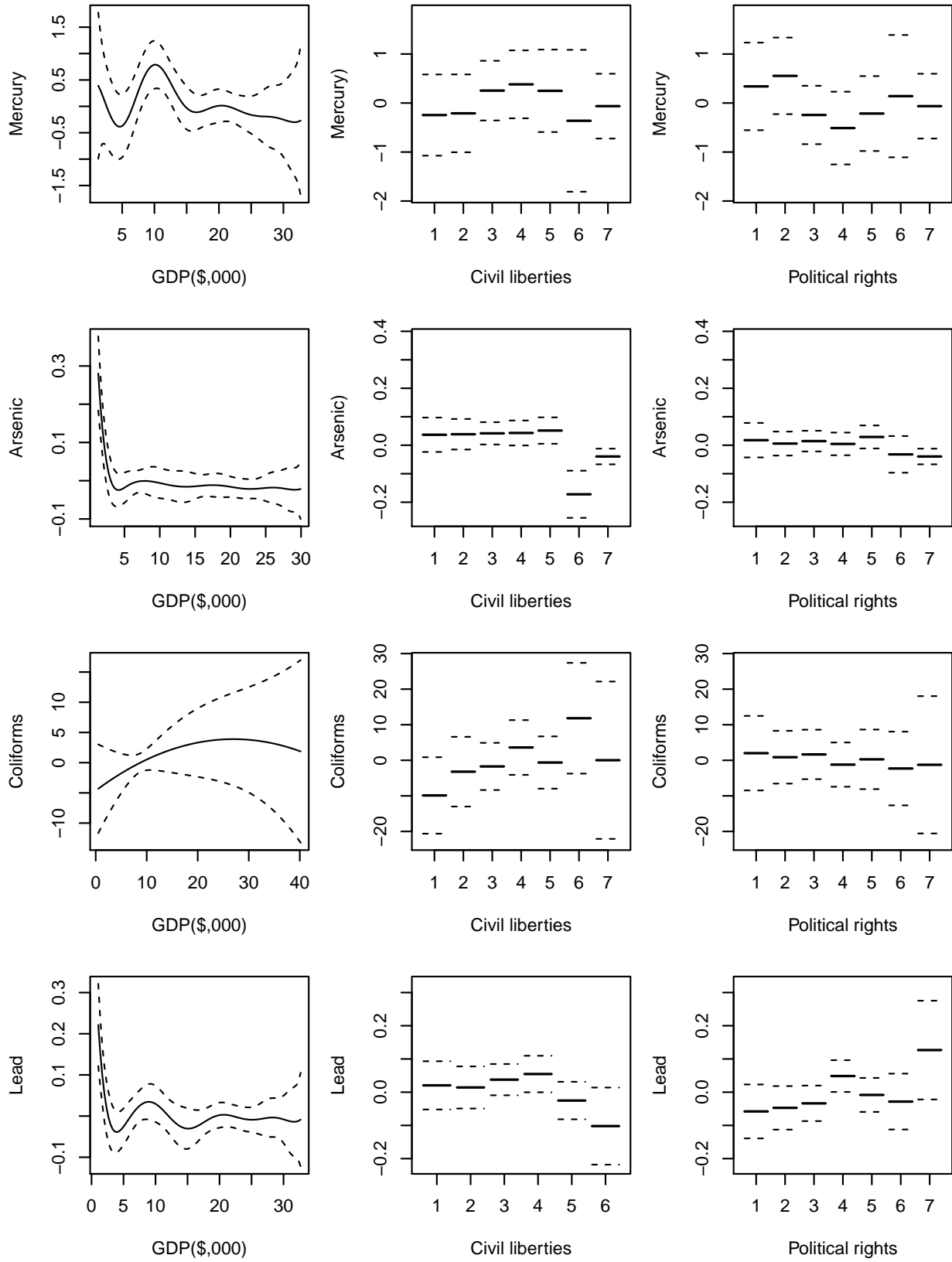


Figure 4.3. Partial Regression Plot from Semiparametric Model 1.

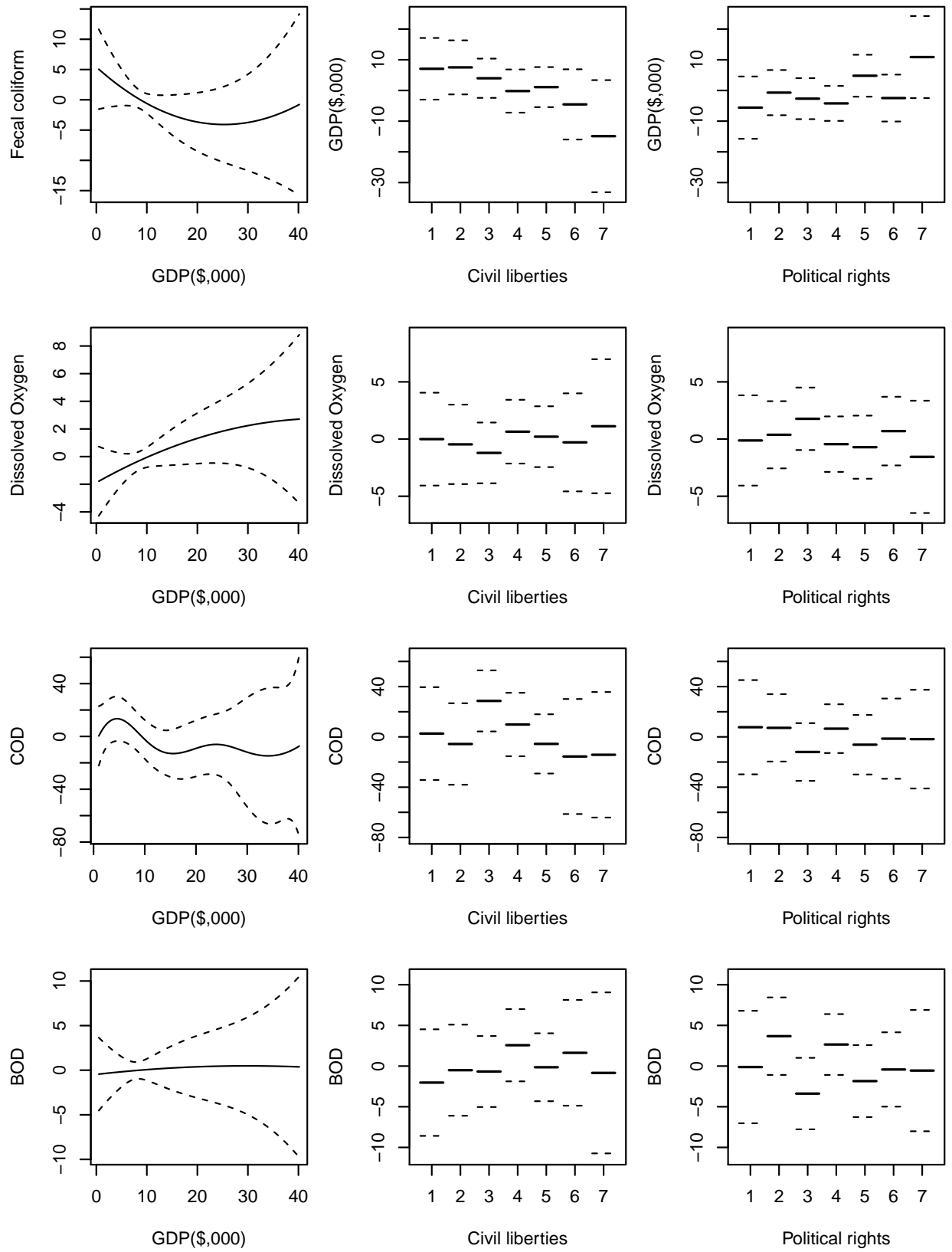


Figure 4.3. Contd.

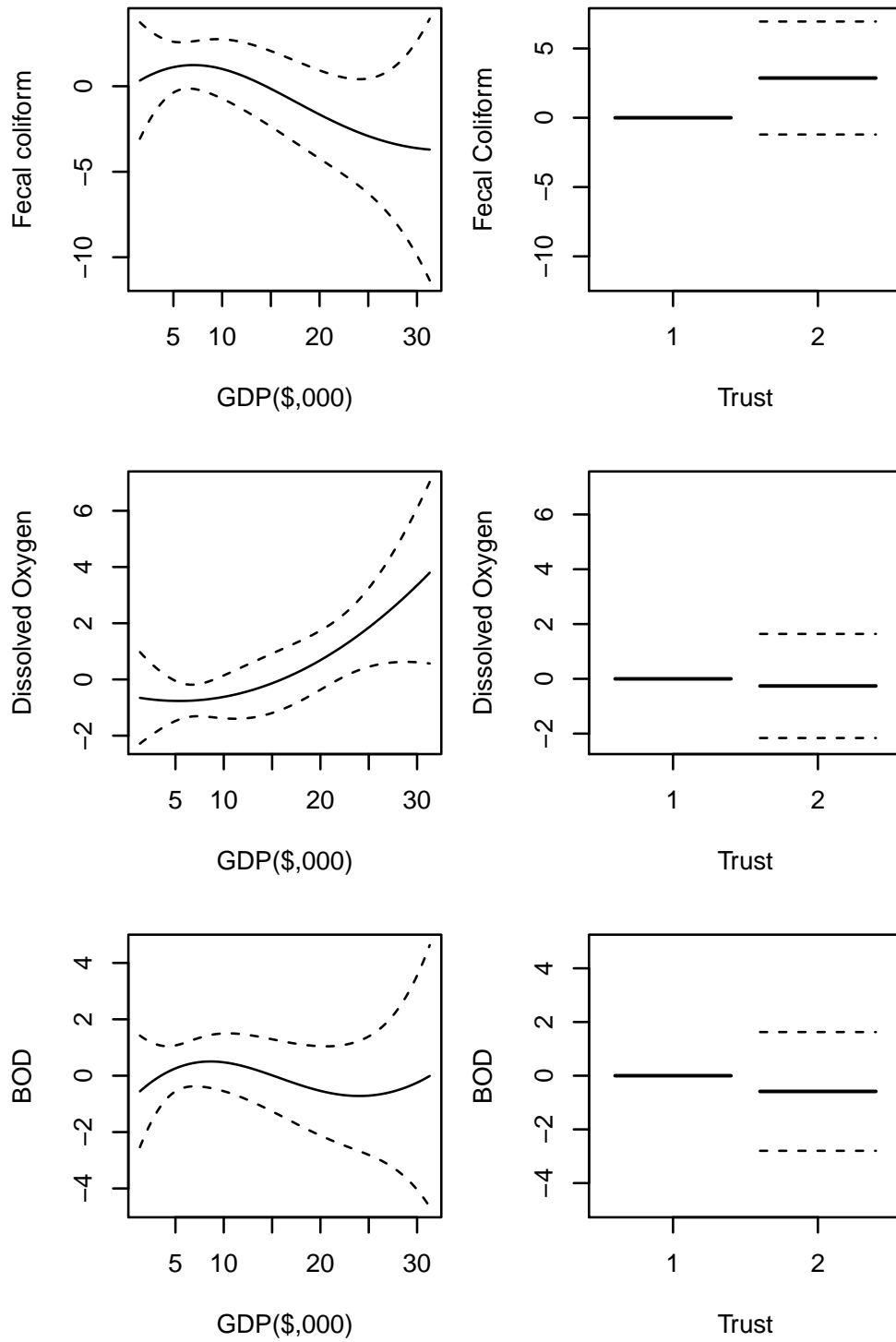


Figure 4.4. Partial Regression Plot from Semiparametric Model 2.

This figure also shows that higher trust among people indicated low levels of fecal coliform pollution. This is also supported by nonparametric results as shown in Figure (4.2).

4.4.6 Mercury

In the case of mercury, the model specification test shows that semiparametric models are better specified compared to a parametric model. The estimated semiparametric model is plotted in Figure (4.3). The figure shows that a cubic relationship exists between mercury and GDP. There are also low levels of mercury pollution for small and high values of civil liberties. Based on the CV score, we can choose a nonparametric model over parametric and semiparametric models. The nonparametric estimation (Figure 4.1) shows similar effects of political rights for mercury pollution. Hence, we conclude that an EKC exists for the relationship between civil liberties and mercury pollution. This is also true for political rights and mercury pollution.

4.4.7 Arsenic

Although all GDP related variables are significant in arsenic, the model specification test results show that the semiparametric model is more appropriate compared to the quadratic and cubic models. The CV score for the semiparametric model is the lowest compared to any other model, so the semiparametric model is preferred compared to parametric and nonparametric models. The estimated semiparametric model (Figure 4.3) indicates that arsenic pollution rapidly decreases with increases in GDP up to the level \$5,000 and becomes constant for higher levels of GDP. This finding indicates the presence of an L shaped EKC. The results also show that the amount of arsenic pollution is low for high values of civil liberties, which tells us that if the country has a high value of civil liberties (i.e. low level of civil liberties), the arsenic pollution will be low.

4.4.8 Lead

Model specification tests indicate a semiparametric model is better for lead. The estimated semiparametric model (Figure 4.3) shows that lead pollution also rapidly decreases with an increase in income up to level \$4,000, and is constant for higher levels of GDP. Thus, there also exists an L shaped EKC for lead. We observed that lead pollution increases as the civil liberties value goes up to 4 and then decreases for higher values. This implies that as the level of civil liberties goes up, lead pollution increases and then decreases for higher levels of civil liberties. This supports the existence of an EKC with respect to civil liberties. In contrast, we find an N shaped relationship between political rights and lead pollution as shown in Figure 4.3.

4.4.9 Dissolved Oxygen

Table (4.3) shows that either quadratic or semiparametric models are preferred for dissolved oxygen. Since the GDP variables are not significant in a quadratic model, we interpret the results from the semiparametric model. The estimated semiparametric model (Figure 4.3) shows that dissolved oxygen has a positive relationship with GDP, indicating a reduction of pollution with increases in income. In Model 2, where the trust variable is included as an explanatory variable, we found that dissolved oxygen decreases with an increase in GDP up to a level of \$5,000 and then increases (Figures 4.2, 4.4). This supports an inverted U-shaped relation between pollution and GDP.

4.4.10 COD

Table 4.3 indicates that either semiparametric or cubic models are preferred for COD. The GDP related variables are not significant for COD. The data generated from the estimated semiparametric model shows that COD increases with an increase in GDP up to \$5,000 and then starts to decrease until \$15,000. Again for higher GDP, the COD are also likely to increase, as shown in Figure (4.3). This indicates an N-shaped relationship between

pollution and income. Based on the CV score, we can also choose a nonparametric model over parametric and semiparametric models. We found an inverted U-shape relationship between COD and civil liberties as shown in Figure (4.1).

4.4.11 BOD

The model specification test shows that either a cubic or semiparametric model is preferred in the case of BOD; however, none of the GDP variables are significant in a cubic model. The CV score for the nonparametric model is small compared to other models (Tables 4.5, 4.6). Hence, we choose a nonparametric model for the BOD. The estimated nonparametric model (Figure 4.1) shows that BOD initially increases with an increase in GDP, and then decreases after \$8,000. With further increases in GDP, the figure shows that BOD also increases. Therefore, we found there to be an N-shaped relationship between BOD and GDP. This result is also similar to what we have found for Model 2 as shown in Figure (4.2).

4.4.12 Total Coliform

We found that a semiparametric model performs better for coliform compared to a parametric model. The estimated semiparametric model is given in Figure (4.3). The figure shows an inverted U-shaped relationship between coliform and GDP. Thus, our results indicate the existence of an EKC for coliform with a turning point of \$17,000. We also found that coliform is small for low and high values of civil liberties and high for medium levels of civil liberties. This result indicates an inverted U-shaped relationship between coliform and civil liberties.

4.5 Conclusions

This study provides an understanding of the relationship between income, civil liberties, political rights and trust with regard to water quality at the global level. Despite data limitations, we found an inverted U-shape relationship for three pollutants (dissolved oxygen,

fecal coliform and coliform), a cubic relationship for three pollutants (mercury, chemical oxygen demand and biological oxygen demand), and an L shaped relationship for two pollutants (arsenic and lead). In general, we found that the coefficients on the political rights and civil liberties are significant in many models.

The test statistics used in this research show that semiparametric and nonparametric models are better than parametric cubic and quadratic models for modeling EKC relationships. We find that a model specification test is important to select an appropriate functional form. According to the results of these model specification tests, there is an important role for semiparametric and nonparametric models in studies of EKCs, as was found by several previous studies cited in this dissertation.

Based on the estimated nonparametric and semiparametric plots, we found that an EKC for the relationship between pollution and civil liberties and political rights. Our results suggest an inverted U-shape curve for fecal coliform, COD, BOD, Mercury, and lead with respect to civil liberties. In contrast, we found a cubic shaped relationship between political rights and pollution. Results suggest that as countries progress towards political rights, water pollution increases at first but then decreases after certain levels of political rights have been attained. However, results indicate a likelihood of increasing water pollution for the highest levels of political rights. In contrast, we find that civil liberties has an inverted U-shaped relationship, meaning that those countries with low and high civil liberties have better water quality. Thus, factors affecting political rights such as the fairness of the electoral process, the degree of political pluralism and participation, and the presence of a non-corrupt and transparent government are beneficial for water quality. Trust among people is also important in the determination of water quality. We found that if there is more trust among people, there will be better water quality. Thus, our results suggest that the mechanisms through which higher income may improve water quality is through the political process, individual expression, and trust among people.

CHAPTER 5. CONCLUSIONS

This dissertation consists of three essays on the EKC for flow and stock water pollutants. The EKC describes the relationship between pollution and economic growth as an inverted U-shape curve. This implies that, as economic growth continues, it increases pollution levels up to a certain point after which pollution declines. Economic growth is measured by income per capita and pollution concentration or pollution per capita are used to measure pollution. Generally, water quality parameters consist of heavy metals (nickel, mercury, arsenic, cadmium, lead), pathogenic contamination (fecal coliform, total coliform), oxygen regime (dissolved oxygen (DO), chemical oxygen demand (COD), biological oxygen demand (BOD) and nutrients (nitrogen, phosphorous). The level of per capita income that maximizes pollutant concentration is defined as the “turning point”.

Previous studies have not presented a unanimous view on the shape or existence of EKCs for flow and stock pollutants. The genesis of discrepancies in findings is a result of data and estimation methods used in EKC literature. The literature suggests continuous search be made for a more flexible model specification so as to examine the EKC hypothesis. On the other hand, researchers are trying to establish the theoretical reasoning and framework as to the existence of EKCs. In this dissertation, parametric assumptions are relaxed in order to assess the shape of the income-pollution relationship.

In chapter two of this dissertation, advanced literature in econometrics specifically related to nonparametric and semiparametric models are reviewed. Surveys indicate that EKC studies have not fully utilized the advances in nonparametric and semiparametric methods. There is still a debate on the use of econometric methods to establish relationships between pollutants and income. Researchers are focusing on relaxing distributional assumptions using nonparametric and semiparametric model. Previous studies have suggested that semiparametric models are a better model specification compared to a parametric model

specification. A survey of literature suggests that future research should continue to use more flexible model specification in EKC modeling. This chapter provides a complete summary of literature and points out future directions for estimating an EKC model using semiparametric and parametric regression models.

In chapter three, EKC hypothesis on four major water quality parameters (N, P, DO and Mercury) are examined at a local level (Louisiana) using panel data available for water pollutants in Louisiana. A seemingly unrelated partial linear regression model (SUPLR) is proposed to address the potential correlation between the water quality parameters. Simulation study shows that the SUPLR model gives unbiased and consistent estimators for both parametric and nonparametric components. Empirical simulations verify that the SUPLR model behave well for Louisiana water quality parameters. Thus, SUPLR model is estimated for all the water quality parameters. The SUPLR model is compared with the parametric models using the model specification test developed by [Hsiao, Li, and Racine \(2007\)](#). The results indicated that SUPLR model is correctly specified for N, P, and DO, and the cubic model is correctly specified for Mercury. Results indicated that EKC was present in nitrogen, dissolved oxygen and mercury, but not in phosphorus. Findings also indicated higher population density reduce water pollution, and higher farmland areas are responsible for increased water pollution. Point sources increase mercury pollution. Although the time dimension of the panel sample is small, our results suggests a need to continually assess policy effectiveness for pollution control as income increases.

In chapter four, the relationship between water quality and income is examined at the global level using data from GEMS. Explanatory variables included in the model are political rights, civil liberties and trust, closely related to social capital in the country. Despite data limitations, an inverted U-shaped relationship was found for three pollutants (dissolved oxygen, fecal coliform and coliform), a cubic relationship is found for three pollutants (mercury, chemical oxygen demand and biological oxygen demand) and an L-shaped relationship is observed for two pollutants (arsenic and lead). In general, the coefficients on the political

rights and civil liberties are significant in many models. Interestingly, results indicate that there is also an inverted U-shaped relationship for civil liberties and pollution for some of the water quality parameters. This means that as a country improves its civil liberties and political rights, the level of pollution increases then reaches a turning point and begins to decrease. At the highest level of the civil liberties and political rights, pollution decreases substantially. Our results suggest that the mechanism through which higher income may improve water quality is through the political process and individual expression. This finding lends support to previous studies which emphasize a close connection between political rights and civil liberties ([Dasgupta et al., 2002](#); [Dasgupta and Mäler, 1995](#)). Trust is also an important determinant of social capital that plays an important role on determining environmental pollution. If trust among people is higher, there is low level of water pollution. Hence, high environmental quality depends partially on the levels of trust within the society.

Overall, there are still shortcomings on the existing estimation methods used in EKC analysis. A more flexible data driven approach like nonparametric or semiparametric models needs to be used in order to find more accurate relationships between economic growth and pollution levels. Since economic growth is not the only factor that affects pollution level, consideration of additional variables will help to avoid an omitted variable bias in a model estimation procedure. This dissertation contributed to the use of semiparametric modeling on empirical studies of the EKC hypothesis. Finally, there is consistent behavior of environmental quality parameters at least for water quality parameters (for nitrogen and dissolve oxygen) at the local and the global level as suggested by this study. Higher economic growth could be the solution for the flow pollutants, but not for the stock pollutants. Heavy metal pollutants are stock in environment so their levels are less likely to reduce until abatement effort are employed. Likewise [Grossman and Krueger \(1995\)](#) and [Dasgupta et al. \(2002\)](#) purport that environmental regulations are needed to control environmental degradation especially for stock pollutants. Further, economic growth is not the only

important factor that determines environmental quality, but other important variables need to be incorporated in future EKC studies.

REFERENCES

- Acemoglu, D., P. Aghion, L. Bursztyn, and D. Hemous. 2010. The Environment and Directed Technical Change. <http://www.economics.harvard.edu/faculty/aghion/files/Environment%20and%20Directed.pdf>. Accessed from the web on December 12, 2010.
- Ai, C., and Q. Li. 2008. *Semi-parametric and Non-parametric Methods in Panel Data Models*. Springer-Verlag Berlin Heidelberg.
- Akbostanci, E., S. Turut-Asik, and G. I. Tunc. 2009. The Relationship Between Income and Environment in Turkey: Is There an Environmental Kuznets Curve? *Energy Policy* 37(3): 861–867.
- Andreoni, J., and A. Levinson. 2001. The Simple Analytics of the Environmental Kuznets Curve. *Journal of Public Economics* 80(2): 269–286.
- Azomahou, T., F. Laisney, and N. P. Van. 2006. Economic Development and CO_2 Emissions: A Nonparametric Panel Approach. *Journal of Public Economics* 90(6-7): 1347–1363.
- Baltagi, B. H. 1980. On Seemingly Unrelated Regressions with Error Components. *Econometrica* 48: 1547–1551.
- Barbier, E. B. 2004. Explaining Agricultural Land Expansion and Deforestation in Developing Countries. *American Journal of Agricultural Economics* 86(5): 1347–1353.
- Barrett, S., and K. Graddy. 2000. Freedom, Growth, and the Environment. *Environment and Development Economics* 5(4): 433–456.
- Bertinelli, L., and E. Strobl. 2005. The Environmental Kuznets Curve Semi-Parametrically Revisited. *Economics Letters* 88(3): 350–357.
- Blundell, R., and A. Duncan. 1998. Kernel Regression in Empirical Microeconomics. *Journal of Human Resources* 33(1): 62–87.
- Brock, W. A., and M. S. Taylor. 2010. The Green Solow Model. *Journal of Economic Growth* 15(2): 127–153.
- Bruvoll, A., and H. Medin. 2003. Factors Behind the Environmental Kuznets Curve: A Decomposition of the Changes in Air Pollution. *Environmental and Resource Economics* 24(1): 27–48.
- Carson, R. T. 2010. The Environmental Kuznets Curve: Seeking Empirical Regularity and Theoretical Structure. *Review of Environmental Economics and Policy* 4(1): 3–23.
- Cavlovic, T. N., K. H. Baker, R. P. Berrens, and K. Gawande. 2000. A Meta-Analysis of Environmental Kuznets Curve Studies. *Agricultural and Resource Economics Review* 29(1): 32–42.

- Chen, X., and Y. Fan. 2006. Estimation of Copula-based Semiparametric Time Series Models. *Journal of Econometrics* 130(2): 307–335.
- Chiu, Y. B. 2012. Deforestation and the Environmental Kuznets Curve in Developing Countries: A Panel Smooth Transition Regression Approach. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroéconomie* 60(2): 177–194.
- Clement, M., and A. Meunie. 2010. Is Inequality Harmful for the Environment? An Empirical Analysis Applied to Developing and Transition Countries. *Review of social economy* 68(4): 413–445.
- Cole, M. A. 2004. Trade, the Pollution Haven Hypothesis and the Environmental Kuznets Curve: Examining the Linkages. *Ecological Economics* 48(1): 71–81.
- Copeland, B. R., and M. S. Taylor. 2004. Trade, Growth, and the Environment. *Journal of Economic Literature* 42(1): 7–71.
- Crainiceanu, C. M., and D. Ruppert. 2004. Likelihood Ratio Tests in Linear Mixed Models with One Variance Component. *Journal of the Royal Statistical Society* 66(1): 165–185.
- Criado, C. O. 2008. Temporal and Spatial Homogeneity in Air Pollutants Panel EKC Estimations: Two Nonparametric Tests Applied to Spanish Provinces. *Environmental and Resource Economics* 40(2): 265–283.
- Culas, R. J. 2007. Deforestation and the Environmental Kuznets Curve: An Institutional Perspective. *Ecological Economics* 61(2-3): 429–437.
- Darolles, S., Y. Fan, J. P. Florens, and E. Renault. 2011. Nonparametric Instrumental Regression. *Econometrica* 79(5): 1541–1565.
- Dasgupta, P., and K. G. Mäler. 1995. *Poverty, Institutions, and the Environmental Resource Base*. In *Handbook of Development Economics*, volume 3A of *Elsevier Science*, chapter 39. Amsterdam.
- Dasgupta, S., B. Laplante, W. Hua, and D. Wheeler. 2002. Confronting the Environmental Kuznets Curve. *Journal of Economic Perspectives* 16(1): 147–168.
- Deacon, R. T., and C. S. Norman. 2006. Does the Environmental Kuznets Curve Describe How Individual Countries Behave? *Land Economics* 82(2): 291–315.
- Delgado, M. A., and T. Stengos. 1994. Semiparametric Specification Testing of Non-Nested Econometric Models. *The Review of Economic Studies* 61(2): 291–303.
- Dincer, O. C., and E. M. Uslander. 2010. Trust and Growth. *Public Choice* 142(1-2): 59–67.
- Fan, Y., and Q. Li. 1996. Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms. *Econometrica* 64(4): 865–890.
- Farzin, Y. H., and C. A. Bond. 2006. Democracy and Environmental Quality. *Journal of Development Economics* 81(1): 213–235.

- Farzin, Y. H., and K. A. Grogan. 2013. Socioeconomic Factors and Water Quality in California. *Environmental Economics and Policy Studies* 15(1): 1–37.
- Feather, P., D. Hellerstein, and L. Hansen. 1999. *Economic Valuation of Environmental Benefits and the Targeting of the Conservation Programs: The Case of the CRP*. U.S. Department of Agriculture Economic Research Service, Washington DC. Agricultural Economics Report No. 778.
- Freedom House 2010. Freedom in the World 2003: Survey Methodology. <http://www.freedomhouse.org/research/freeworld/2003/methodology.htm>. Cited 10 Dec 2010.
- Gassebner, M., M. J. Lamla, and J. E. Sturm. 2011. Determinants of Pollution: What Do We Really Know? *Oxford Economic Papers* 63(3): 568–595.
- Gawande, K., R. P. Berrens, and A. K. Bohara. 2001. A Consumption-Based Theory of the Environmental Kuznets Curve. *Ecological Economics* 37(1): 101–112.
- Gergel, S. E., E. M. Bennett, B. K. Greenfield, S. King, C. A. Overdeest, and B. Stumborg. 2004. A Test of the Environmental Kuznets Curve Using Long-Term Watershed Inputs. *Ecological Applications* 14(2): 555–570.
- González, A., T. Teräsvirta, and D. v. Dijk. 2005. *Panel Smooth Transition Regression Models*. SSE/EFI Working Paper Series in Economics and Finance. No. 604.
- Grafton, R. Q., and S. Knowles. 2004. Social Capital and National Environmental Performance: A Cross-Sectional Analysis. *The Journal of Environment & Development* 13(4): 336–370.
- Griffin, J. E., and M. F. Steel. 2010. Bayesian Nonparametric Modelling with the Dirichlet Process Regression Smoother. *Statistica Sinica* 20(4): 1507.
- Grossman, G., and A. Krueger. 1995. Economic Growth and the Environment. *Quarterly Journal of Economics* 110(2): 353–377.
- Grossman, G. M., and A. B. Krueger. 1991. *Environmental impacts of a North American free trade agreement*. NBER Working Paper No. 3914.
- Harbaugh, W. T., A. Levinson, and D. M. Wilson. 2002. Reexamining the Empirical Evidence for an Environmental Kuznets Curve. *Review of Economics and Statistics* 84(3): 541–551.
- Hardle, W. 1990. *Applied Nonparametric Regression*, volume 27: Cambridge Univ Press.
- Hardle, W., and E. Mammen. 1993. Comparing Nonparametric Versus Parametric Regression Fits. *The Annals of Statistics* 21(4): 1926–1947.
- Hastie, T., and R. Tibshirani. 1990. *Generalized Additive Models*: Chapman and Hall.
- Hausman, J. A. 1978. Specification Tests in Econometrics. *Econometrica*: 1251–1271.

- Heerink, N., A. Mulatu, and E. Bulte. 2001. Income Inequality and the Environment: Aggregation Bias in Environmental Kuznets Curves. *Ecological Economics* 38(3): 359–367.
- Hong, Y., and H. White. 1995. Consistent Specification Testing via Nonparametric Series Regression. *Econometrica* 63(5): 1133–1159.
- Horowitz, J. L. 2011. Applied Nonparametric Instrumental Variables Estimation. *Econometrica* 79(2): 347–394.
- Hsiao, C., Q. Li, and J. S. Racine. 2007. A Consistent Model Specification Test with Mixed Discrete and Continuous Data. *Journal of Econometrics* 140(2): 802–826.
- Israel, D., and A. Levinson. 2004. Willingness to Pay for Environmental Quality: Testable Empirical Implications of the Growth and Environment literature. *Contributions to Economic Analysis & Policy* 3(1): Article 2.
- Jha, R., and K. V. B. Murthy. 2003. An Inverse Global Environmental Kuznets Curve. *Journal of Comparative Economics* 31(2): 352–368.
- John, A., and R. Pecchenino. 1994. An Overlapping Generations Model of Growth and the Environment. *The Economic Journal*: 1393–1410.
- Kelly, D. L. 2003. On Environmental Kuznets Curves Arising from Stock Externalities. *Journal of Economic Dynamics and Control* 27(8): 1367–1390.
- Kijima, M., K. Nishide, and A. Ohyama. 2010. Economic Models for the Environmental Kuznets Curve: A Survey. *Journal of Economic Dynamics and Control* 34(7): 1187–1201.
- Kim, K. H. 2013. Inference of the Environmental Kuznets Curve. *Applied Economics Letters* 20(2): 119–122.
- Kniesner, T. J., and Q. Li. 2002. Nonlinearity in Dynamic Adjustment: Semiparametric Estimation of Panel Labor Supply. *Empirical Economics* 27(1): 131–148.
- Kuznets, S. 1955. Economic Growth and Income Equality. *American Economic Review* 45(1): 1–28.
- Leamer, E. E. 1983. Let's Take the Con out of Econometrics. *The American Economic Review* 73(1): 31–43.
- Lee, C. C., Y. B. Chiu, and C. H. Sun. 2010. The Environmental Kuznets Curve Hypothesis for Water Pollution: Do Regions Matter? *Energy Policy* 38(1): 12–23.
- Lee, T. H., and A. Ullah. 2001. Nonparametric Bootstrap Tests for Neglected Nonlinearity in Time Series Regression Models. *Journal of Nonparametric Statistics* 13(3): 425–451.
- Levine, R., and D. Renelt. 1992. A Sensitivity Analysis of Cross-Country Growth Regressions. *The American economic review* 82(2): 942–963.

- Li, Q., C. J. Huang, D. Li, and T. T. Fu. 2002. Semiparametric Smooth Coefficient Models. *Journal of Business & Economic Statistics* (3): 412.
- Li, Q., and J. Racine. 2007. *Nonparametric econometrics: theory and practice*: Princeton University Press.
- Li, Q., J. S. Racine, and J. M. Wooldridge. 2009. Efficient Estimation of Average Treatment Effects with Mixed Categorical and Continuous Data. *Journal of Business & Economic Statistics* 27(2): 206–223.
- Li, Q., and S. Wang. 1998. A Simple Consistent Bootstrap Test for a Parametric Regression Function. *Journal of Econometrics* 87(1): 145–165.
- Li, Q. I., D. Ouyang, and J. S. Racine. 2013. Categorical Semiparametric Varying-coefficient Models. *Journal of Applied Econometrics* 29: 551–579.
- Li, W. 2011. A nonparametric analysis on the environmental Kuznets curve. *Environmetrics* 22(3): 420.
- Lin, C. Y. C., and Z. D. Liscow. 2013. Endogeneity in the Environmental Kuznets Curve: An Instrumental Variables Approach. *American Journal of Agricultural Economics* 95(2): 268–274.
- Linton, O., and J. P. Nielsen. 1995. A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration. *Biometrika* 82(2): 93–100.
- List, J. A., and C. A. Gallet. 1999. The Environmental Kuznets Curve: Does One Size Fit All? *Ecological Economics* 31(3): 409–423.
- Lopez, R. 1994. The Environment as a Factor of Production: The Effects of Economic Growth and Trade Liberalization. *Journal of Environmental Economics and management* 27(2): 163–184.
- Luzzati, T., and M. Orsini. 2009. Investigating the Energy-environmental Kuznets Curve. *Energy* 34(3): 291–300.
- Ma, S., and J. S. Racine. 2013. Additive Regression Splines with Irrelevant Categorical and Continuous Regressors. *Statistica Sinica* 23: 515–541.
- Ma, S., J. S. Racine, and L. Yang. 2011. *Spline regression in the presence of categorical predictors*. Working paper, McMaster University and Michigan State University.
- Mas-Colell, A., M. Whinston, and J. Green. 1995. *Microeconomic Theory*: Oxford University Press.
- Mazzanti, M., and A. Musolesi. 2011. Carbon Abatement Leaders and Laggards Non Parametric Analyses of Policy Oriented Kuznets Curves. <http://www.bepress.com/feem/paper537/>. Accessed from the web on July 1, 2011.

- McConnell, K. E. 1997. Income and the Demand for Environmental Quality. *Environment and Development Economics* 2(4): 383–399.
- MCRIA Council 1999. *Report on the Mercury Contamination Reduction Initiative Advisory Council's Results and Recommendations*: Policy & Planning Division, Minnesota Pollution Control Agency.
- Melenberg, B., H. R. Vollebergh, and E. Dijkgraaf. 2011. Grazing the Commons: Global Carbon Emissions Forever? <http://arno.uvt.nl/show.cgi?fid=113967>. Accessed from the web on July 1, 2011.
- Merlevede, B., T. Verbeke, and M. D. Clercq. 2006. The EKC for SO_2 : Does Firm Size Matter? *Ecological Economics* 59(4): 451–61.
- Millimet, D. L., J. A. List, and T. Stengos. 2003. The Environmental Kuznets Curve: Real Progress or Misspecified Models? *Review of Economics and Statistics* 85(4): 1038–1047.
- Nie, Z., and J. S. Racine. 2012. The crs Package: Nonparametric Regression Splines for Continuous and Categorical Predictors. *The R Journal* 4: 48–56.
- Orubu, C. O., and D. G. Omotor. 2011. Environmental Quality and Economic Growth: Searching for Environmental Kuznets Curves for Air and Water Pollutants in Africa. *Energy Policy* 39(7): 4178–4188.
- Panayotou, T. 1997. Demystifying the Environmental Kuznets Curve: Turning a Black Box into a Policy Tool. *Environment and development economics* 2(4): 465–484.
- Pandit, M., K. P. Paudel, and A. K. Mishra. 2013. Do Agricultural Subsidies Affect the Labor Allocation Decision? Comparing Parametric and Semiparametric Methods. *Journal of Agricultural and Resource Economics* 38(1): 1–18.
- Panopoulou, E., and T. Pantelidis. 2009. Club Convergence in Carbon Dioxide Emissions. *Environmental and Resource Economics* 44(3): 47–70.
- Paudel, K., W. Gauthier, J. Westra, and L. Hall. 2008. Factors Influencing and Steps Leading to the Adoption of Best Management Practices by Louisiana Dairy Farmers. *Journal of Agricultural and Applied Economics* 40(1): 203.
- Paudel, K., B. N. Poudel, D. Bhandari, and T. Johnson. 2011. Examining the Role of Social Capital in Environmental Kuznets Curve Estimation. *Global Journal of Environmental Science and Technology*: 1–11.
- Paudel, K. P., C. Y. C. Lin, and M. Pandit. 2011. *Estimating the Environmental Kuznets Curve for Water Pollutants at the Global Level: Semiparametric and Nonparametric Approaches*. University of California at Davis.
- Paudel, K. P., and B. N. Poudel. 2013. Functional Form of Water Pollutants-Income Relationship under the Environmental Kuznets Curve Framework. *American Journal of Agricultural Economics* 95(2): 261–267.

- Paudel, K. P., and M. J. Schafer. 2009. The Environmental Kuznets Curve under a New Framework: The Role of Social Capital in Water Pollution. *Environmental and Resource Economics* 42(2): 265–278.
- Paudel, K. P., H. Zapata, and D. Susanto. 2005. An Empirical Test of Environmental Kuznets Curve for Water Pollution. *Environmental and Resource Economics* 31(3): 325–348.
- Pesaran, M. H., and R. Smith. 1995. Estimating Long-Run Relationships from Dynamic Heterogeneous Panels. *Journal of econometrics* 68(1): 79–113.
- Phu, N. V. 2003. A Semiparametric Analysis of Determinants of a Protected Area. *Applied Economics Letters* 10(10): 661–665.
- Phu, N. V. 2010. Energy Consumption and Income: A Semiparametric Panel Data Analysis. *Energy Economics* 32: 557–563.
- Plassmann, F., and N. Khanna. 2006. Household Income and Pollution: Implications for the Debate about the Environmental Kuznets Curve Hypothesis. *Journal of Environment and Development* 15(1): 22–41.
- Poudel, B. N., K. P. Paudel, and K. Bhattarai. 2009. Searching for an Environmental Kuznets Curve in Carbon Dioxide Pollutant in Latin American Countries. *Journal of Agricultural and Applied Economics* 41(1): 13–27.
- Qian, J., and L. Wang. 2012. Estimating Semiparametric Panel Data Models by Marginal Integration. *Journal of Econometrics* 167(2): 483–493.
- Racine, J., and Q. Li. 2004. Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data. *Journal of Econometrics* 119(1): 99–130.
- Racine, J. S. 2011. *Nonparametric Kernel Methods for Qualitative and Quantitative Data*, 183–204. CRC Press.
- Racine, J. S., and Z. Nie. 2012. *crs: Categorical Regression Splines*. R package version 0.15-17.
- Robinson, P. M. 1988. Root-N-Consistent Semiparametric Regression. *Econometrica* (4): 931.
- Rodriguez-Meza, J., D. Southgate, and C. Gonzalez-Vega. 2004. Rural Poverty, Household Responses to Shocks, and Agricultural Land Use: Panel Results for El Salvador. *Environment and Development Economics* 9(2): 225–239.
- Rosser, J. B., and M. V. Rosser. 2006. Institutional Evolution of Environmental Management under Global Economic Growth. *Journal of Economic Issues* 40(2): 421–429.
- Roy, N., and G. van Kooten. 2004. Another Look at the Income Elasticity of Non-Point Source Air Pollutants: A Semiparametric Approach. *Economics Letters* 85(1): 17–22.

- Rupasingha, A., S. J. Goetz, D. L. Debertin, and A. Pagoulatos. 2004. The Environmental Kuznets Curve for US Counties: A Spatial Econometric Analysis With Extensions. *Papers in Regional Science* 83(2): 407–424.
- Ruppert, D., P. Wand, and R. Carroll. 2003. *Semiparametric Regression*: Cambridge University Press.
- Santos, A. 2012. Inference in Nonparametric Instrumental Variables with Partial Identification. *Econometrica* 80(1): 213–275.
- Schmalensee, R., T. M. Stoker, and R. A. Judson. 1998. World Carbon Dioxide Emissions: 1950 – 2050. *Review of Economics and Statistics* 80(1): 15–27.
- Shafik, N. 1994. Economic Development and Environmental Quality: An Econometric Analysis. *Oxford Economic Papers* 46: 757–773.
- Shafik, N., and S. Bandyopadhyay. 1992. *Economic Growth and Environmental Quality: Time Series and Cross Country Evidence*: World Development Working Paper WPS 904. Washington, DC: World Bank.
- Stern, D. I. 2004. The Rise and Fall of the Environmental Kuznets Curve. *World Development* 32(8): 1419–1439.
- Stock, J. H. 1989. Nonparametric Policy Analysis. *Journal of the American Statistical Association* 84(406): 567–575.
- Thompson, A. 2012. Water Abundance and an EKC for Water Pollution. *Economics Letters* 117: 423–425.
- Torras, M., and J. K. Boyce. 1998. Income, Inequality, and Pollution: A Reassessment of the Environmental Kuznets Curve. *Ecological Economics* 25(2): 147–160.
- Tsurumi, T., and S. Managi. 2010. Decomposition of the Environmental Kuznets Curve: Scale, Technique, and Composition Effects. *Environmental Economics and Policy Studies* 11(1-4): 19–36.
- Ullah, A. 1985. Specification Analysis of Econometric Models. *Journal of Quantitative Economics* 2: 187–209.
- Ullah, A., and N. Roy. 1998. *Nonparametric and Semiparametric Econometrics of Panel Data*, 579–604. CRC Press, New York: Marcel Dekke.
- Van, N. P., and T. Azomahou. 2007. Nonlinearities and Heterogeneity in Environmental Quality: An Empirical Analysis of Deforestation. *Journal of Development Economics* 84(1): 291–309.
- Vincent, J. R. 1997. Testing for Environmental Kuznets Curves within a Developing Country. *Environment and Development Economics* 2(4): 417–431.
- Wand, M., and C. Jones. 1995. *Kernel Smoothing*: Chapman & Hall.

- Wood, S. 2006. *Generalized Additive Models: An Introduction with R*: Chapman & Hall/CRC.
- Yatchew, A. 2003. *Semiparametric Regression for the Applied Econometrician*: Cambridge University Press.
- You, J., Y. Zhou, and G. Chen. 2013. Statistical Inference of Multivariate Partially Linear Regression Models. *Canadian Journal of Statistics* 41(1): 1–22.
- Zak, P. J., and S. Knack. 2001. Trust and Growth. *The economic journal* 111(470): 295–321.
- Zanin, L., and G. Marra. 2012. Assessing the Functional Relationship between CO_2 Emissions and Economic Development Using an Additive Mixed Model Approach. *Economic Modelling* 29: 1328–1337.
- Zapata, H. O., and K. P. Paudel. 2009. *Functional Form of the Environmental Kuznets Curve*. In *Nonparametric Econometric Methods (Advances in Econometrics)*, volume 25, 471–493. Emerald Group Publishing Limited.
- Zhao, X., and D. Wang. 2010. Mercury in Some Chemical Fertilizers and the Effect of Calcium Superphosphate on Mercury Uptake by Corn Seedlings (*Zea Mays* L.). *Journal of Environmental Sciences* 22(8): 1184–1188.
- Zheng, J. X. 1996. A Consistent Test of Functional Form via Nonparametric Estimation Techniques. *Journal of Econometrics* 75(2): 263–289.
- Zheng, Y., Y. Liu, H. Hu, and J. He. 2008. Mercury in Soils of Three Agricultural Experimental Stations with Long-Term Fertilization in China. *Chemosphere* 72(9): 1274–1278.

APPENDIX A.

SEEMINGLY UNRELATED PARTIAL LINEAR REGRESSION ESTIMATION PROCEDURE

Let us consider partial linear model in equation (3.4) For each given α_j and Γ_{ji} , equation (3.4) can be written as

$$P_{jit} - X_{jit}\alpha_j - \Gamma_{ji} = G_j(y_{it}) + u_{jit} \quad j = 1, \dots, M, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad (\text{A.1})$$

Using the local linear smoothing technique (Fan and Li, 1996), we can obtained the unknown function $G_j(\cdot)$. Specifically, for y in a smallest neighborhood of y_0 , $G_j(y)$ can be approximated by

$$G_j(y) \approx G_j(y_0) + G'_j(y_0)(y - y_0) \equiv a_j + b_j(y - y_0) \quad (\text{A.2})$$

where $G'_j(y) = dG'_j(y)/dy$. This makes the following local least-squares problem: find $[(a_j, b_j), j = 1, \dots, m]$ to minimize

$$\sum_{i=1}^n \sum_{j=1}^M [(P_{jit} - X'_{jit}\alpha_j - \Gamma_{ji}) - (a_j + b_j(y_i - y_0))]^2 K_{h_j}(y_i - y_0), \quad (\text{A.3})$$

where $K(\cdot)$ is a kernel function, h_j is a bandwidth for j^{th} equation and $K_{h_j}(\cdot) = h_j^{-1}K(\cdot/h_j)$. By simple algebra, the solution to equation (A.3) is found to be

$$(\hat{a}_j(\alpha_j), \hat{b}_j\alpha_j) = (\hat{a}_j, \hat{b}_j) = (D'_{y_0}W_{y_0}D_{y_0})^{-1}D'_{y_0}W_{y_0}(P_j - X\alpha_j - \Gamma_j) \quad (\text{A.4})$$

where $D_y = \begin{pmatrix} 1 & (y_1 - y) \\ \vdots & \vdots \\ 1 & (y_n - y) \end{pmatrix}$, $W_{y_j} = \text{diag}(K_{h_j}(y_1 - y), \dots, K_{h_j}(y_n - y))$, $X = (X_1, \dots, X_n)'$ and

$P_j = (P_{j1}, \dots, P_{jn})'$ We can now estimate $G_j(y_0)$ by $\hat{a}_j(\alpha_j) = \hat{a}_j$. Substituting \hat{a}_j into equation (A.2). We obtain

$$\hat{P}_{jit} = \hat{X}'_{jit}\alpha_j + \epsilon_{jit}^* \quad (\text{A.5})$$

where $\hat{P}_j = (\hat{P}_{j1}, \dots, \hat{P}_{jn})' = (I_n - S)P_j$ I_n is the n by n identity matrix $\hat{X} = (\hat{X}_1, \dots, \hat{X}_n)' = (I_n - S)X$, $\epsilon_j^* = (\epsilon_{1j}^* \dots \epsilon_{nj}^*)' = (I_n - S)\epsilon_j$ with

$$S = \begin{pmatrix} (1, 0)(D'_{y_1}W_{y_1}D_{y_1})^{-1}D'_{y_1}W_{y_1} \\ \vdots \\ (1, 0)(D'_{y_n}W_{y_n}D_{y_n})^{-1}D'_{y_n}W_{y_n} \end{pmatrix}, \quad \epsilon_j = (\epsilon_{1j}, \dots, \epsilon_{nj})$$

Using the least squares method we estimate the parameter vector α_j by

$$\hat{\alpha}_j = \left(\sum_i^n \hat{X}_i \hat{X}'_i \right)^{-1} \sum_i^n \hat{X}_i \hat{P}'_{ij} \quad (\text{A.6})$$

This is the profile least squares estimator of α_j . An estimator of the unknown function $G_j(y)$ has the form

$$\widehat{G}_j(y) = (1, 0)(D'_y W_y D_y)^{-1} D'_y W_y (P_j - X_j \widehat{\alpha}_j) \quad (\text{A.7})$$

APPENDIX B.

WATER QUALITY PARAMETERS

B.1 Nitrogen

Nitrogen pollution occurs in the form of nitrate, nitrite and ammonia . They are nitrogen-oxygen chemicals, and can combine with various organic and inorganic compounds. The major sources of nitrogen pollution are fertilizer runoff , incursion from leaking septic tanks, sewage, and erosion of natural deposits. More information about nitrogen pollution can be obtained from <http://water.epa.gov/drink/contaminants/basicinformation/nitrate.cfm>. Infants who ingest water containing nitrogen are likely to suffer serious illness and if left untreated, possibly die.

B.2 Phosphorous

Phosphorous (P) is commonly found in soil particles. When soil particles are disturbed due to agricultural operations, landslides, and erosion, phosphorous gets released into water. Like nitrogen, use of chemical fertilizer and runoff from manure used for agriculture production are also major sources of phosphorous pollution. Other sources of phosphorous include sewage treatment plant discharge, storm water runoff and failing septic tanks.

B.3 Dissolved Oxygen

Dissolved Oxygen (DO) is the amount of oxygen dissolved in water. Low amount of oxygen in water indicates pollution. Usually, high temperature and nutrient pollution causes to decline oxygen levels in water. DO is measured in miligrams per liter. Aquatic animals are most vulnerable to lowered DO levels. Dissolve oxygen is measured by using the Winkler

method or Meter and Probe . Details of this method can be found at <http://water.epa.gov/type/rs1/monitoring/vms52.cfm>.

B.4 Chemical Oxygen Demand

Chemical Oxygen Demand (COD) is the amount of oxygen needed to fully oxidize an organic compound in water. Thus, the COD indirectly measures the amount of organic compounds in water. Usually, COD measures the amount of organic compounds in surface water. COD is measured in milligrams per liter. A detailed procedure of COD measurement is available at <ftp://ftp.nmenv.state.nm.us/www/swqb/UOCP/StudyManuals/WWLabStudyGuide/13.pdf>.

B.5 Biological Oxygen Demand

Biological Oxygen Demand (BOD) is the amount of consumed dissolved oxygen by microorganism to break down organic material present in a water. The main source of the organic material are sewage treatment plants. BOD is measured in milligrams per liter. The measurement procedure to calculate BOD is available at <http://www.epa.gov/region9/qa/pdfs/5210dqi.pdf>.

B.6 Mercury

Mercury is a liquid metal and found in natural deposits such as ores containing other elements. Major sources of mercury are erosion of natural deposits, discharge from refineries and factories, runoff from landfills and runoff from croplands. See details about mercury at <http://water.epa.gov/drink/contaminants/basicinformation/mercury.cfm>. The major health issues due to high contamination is kidney damage.

B.7 Cadmium

Cadmium (Cd) is found in natural deposits such as ores containing other elements. The major sources of cadmium in drinking water are corrosion of galvanized pipes, erosion of natural deposits, discharge from metal refineries, runoff from waste batteries and paints. Basic information about cadmium water pollution is available in United States Environmental Protection Agency website (<http://water.epa.gov/drink/contaminants/basicinformation/cadmium.cfm>). Excess drink of water containing Cadmium causes kidney damage.

B.8 Lead

Lead is also a toxic metal found in natural deposits. The major source for lead for drinking water is through corrosion of plumbing materials and erosion of natural deposits. Ingesting paint chips and inhaling paint dust are the major sources of exposure to lead (see details at <http://water.epa.gov/drink/contaminants/basicinformation/lead.cfm>). The major health effects of this pollutant are kidney related problems and high blood pressure.

B.9 Arsenic

Arsenic is a semi-metallic element. Major sources of arsenic are natural deposit erosion, runoff from orchards, runoff from glass and electronic production (see details at <http://water.epa.gov/lawsregs/rulesregs/sdwa/arsenic/index.cfm>). Major health problems associated with exposure to arsenic are skin damage, circulatory systems complications, and a higher than normal risk of contracting cancer

B.10 Fecal Coliform

Fecal coliform are the coliform bacteria which are originated from intestinal tract of warm-blooded animals. The main sources of Fecal coliform are human and animal wastes. Fecal coliform causes short term effects such as diarrhea, cramps, nausea and headaches. For detailed information visit <http://water.epa.gov/drink/contaminants/basicinformation/pathogens.cfm>.

B.11 Total Coliform

Total coliform indicate coliform bacteria that are present in the environment naturally. Higher concentrations of this bacterial indicates the presence of other potentially harmful bacteria. For detailed information visit <http://water.epa.gov/drink/contaminants/basicinformation/pathogens.cfm>.

APPENDIX C. R CODE FOR CHAPTER 2

C.1 Simulation

```
library(MASS)
library(np)
library(foreign)
setwd("~/Dissertation/Chapter2/Final analysis")
rm(list=ls())
# weight function to calculate nonparametric model
weight<-function(h,n,t1,a){
  s1<-(n*h)^(-1)*1/sqrt(2*pi)*exp(-t1^2/(2*h^2))
  a1<-(n*h)^(-1)*a**%(1/sqrt(2*pi)*exp(-t1^2/(2*h^2)))
  a2<-(n*h)^(-1)*a**%(1/sqrt(2*pi)*exp(-t1^2/(2*h^2))*t1)
  a3=(n*h)^-1*a**%(1/sqrt(2*pi)*exp(-t1^2/(2*h^2))*(t1^2));
  w<-s1*(a**%a3-t1*(a**%a2))/((a**%a1)*(a**%a3)-(a**%a2)^2);
  return(w)
}

#initial value to set up simulation
##sig12=.6

mysim<-function(n,sig12){
#Number of observation
n<-n

#Parametric variable
x1<-.3*rchisq(n,1) #for equation 1
x2<-rchisq(n,1) #for equation 2

#Common variables
x3<-abs(rnorm(n,0,1))

#nonparametric variable
u<-runif(n,0,1)
g1<-2*sin(1.5*pi*u)
g2<-cos(1.5*pi*u)

#Covariave between error term
sig12<-sig12

#seemingly unralated error term
```

```

Sigma <- matrix(c(1,sig12,sig12,1),2,2) #symmetry covariance matrix

#bivariate error matrix
error<-mvrnorm(n, rep(0, 2), Sigma, empirical = TRUE)

#Generate dependent variables
y1<-g1+1.5*x1+5*x3+error[,1]
y2<-g2+(-2)*x2+2*x3+error[,2]

#Generate dependent variables
#y1<-g1+1.5*x1+5*x3+error[,1]
#y2<-g2+(-2)*x1+error[,2]

a<-rep(1,n)
t1<-a%%t(u)-u%%t(a)

#Find optimal bandwidth for each equation
h1<-npregbw(y1~u)$bw
h2<-npregbw(y2~u)$bw
#h1<-0.1096
#h2<-0.1096
#h1 <- npudensbw(u)$bw
#h2<-h1
# weight for fist equation

w1<-weight(h1,n,t1,a)
w2<-weight(h2,n,t1,a)

#combine parametric variable in each equation
X1<-cbind(x1,x3)
X2<-cbind(x2,x3)

#transform each equation
HX1<-X1-t(w1)%%X1
HX2<-X2-t(w2)%%X2

hx1.1=x1-t(w1)%%x1
#hx3.1=x3-t(w1)%%x3

#hx2.2=x2-t(w2)%%x2
#hx3.2=x3-t(w2)%%x3

```

```

hy1=y1-t(w1)%*%y1
hy2=y2-t(w2)%*%y2;

#Obtain parametric coefficient in each equation
b1<-solve(t(HX1)%*%HX1)%*%t(HX1)%*%hy1
b2<-solve(t(HX2)%*%HX2)%*%t(HX2)%*%hy2

#Obtain nonparametric model
g1=t(w1)%*%(y1-HX1)%*%b1
g2=t(w2)%*%(y2-HX2)%*%b2

#res.1<-y1-g1-X1)%*%b1
res.2<-y2-g2-X2)%*%b2
#res<-data.frame(res.1,res.2)

#tmp=sort(u,index.return=TRUE);
#u=tmp$x; idx=tmp$ix

d1=mean((hy1-HX1)%*%b1)^2)
d2=mean((hy2-HX2)%*%b2)^2)
d3=mean((hy1-HX1)%*%b1)*(hy2-HX2)%*%b2))

#s=(t(hx)%*%hx)^-1*d2

#S=[[d1,d3]; [d3, d2]]^-1;
S<-solve(matrix(c(d1,d3,d3,d2), nrow=2, byrow=TRUE))
S

#X<-cbind(x,x1)
#
shg1=S[1,1]^-1*t(w1)%*%(S[1,1]*(y1-X1)%*%b1)+S[1,2]*(y2-g2-X2)%*%b2))
shg2=S[2,2]^-1*t(w2)%*%(S[2,2]*(y2-X2)%*%b2)+S[1,2]*(y1-g1-X1)%*%b1))
#Two stage residual variance
tres.1<-y1-shg1-X1)%*%b1
tres.2<-y2-shg2-X2)%*%b2
tres<-data.frame(tres.1,tres.2)
sigma<-var(tres)

In<-diag(n)
j<-rep(1,n)
M<-j)%*%solve(t(j)%*%j)%*%j

```



```

#se for fist equation
X1.dev<-(In-M)%*%X1
sum_sq_1<-t(X1.dev)%*%X1.dev
var1<-solve(sum_sq_1)*sigma[1,1]
se1<-sqrt(diag(var1))

#se for fist equation
X2.dev<-(In-M)%*%X2
sum_sq_2<-t(X2.dev)%*%X2.dev
var2<-solve(sum_sq_2)*sigma[2,2]
se2<-sqrt(diag(var2))
se2

var1<-var(tres)
b<-as.vector(cbind(t(b1),t(b2)))
se<-as.vector(cbind(se1,se2))
v<-as.vector(var1)
output<-list(coef=b,se=se,sigma=v)
return(output)
}

#-----Simulation with different condition----#

#simulation 1 (n=100, sigma=.3)

sim1_coef<-matrix(0,nrow=1000,ncol=4)
sim1_se<-matrix(0,nrow=1000,ncol=4)
sim1_var<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(100,.3)
  sim1_coef[i,]<-output$coef
  sim1_var[i,]<-output$sigma
  sim1_se[i,]<-output$se
}

#simulation 2 (n=100, sigma=.6)
sim2_coef<-matrix(0,nrow=1000,ncol=4)
sim2_se<-matrix(0,nrow=1000,ncol=4)
sim2_var<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(100,.9)
  sim2_coef[i,]<-output$coef

```

```

    sim2_var[i,]<-output$sigma
    sim2_se[i,]<-output$se
}

#simulation 3 (n=100, sigma=.9)
sim3_coef<-matrix(0,nrow=1000,ncol=4)
sim3_se<-matrix(0,nrow=1000,ncol=4)
sim3_var<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(100,.9)
  sim3_coef[i,]<-output$coef
  sim3_var[i,]<-output$sigma
  sim3_se[i,]<-output$se
}

#simulation 4 (n=200, sigma=.3)
sim4_coef<-matrix(0,nrow=1000,ncol=4)
sim4_se<-matrix(0,nrow=1000,ncol=4)
sim4_var<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(200,.3)
  sim4_coef[i,]<-output$coef
  sim4_var[i,]<-output$sigma
  sim4_se[i,]<-output$se
}

#simulation 5 (n=200, sigma=.6)
sim5_coef<-matrix(0,nrow=1000,ncol=4)
sim5_se<-matrix(0,nrow=1000,ncol=4)
sim5_var<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(200,.6)
  sim5_coef[i,]<-output$coef
  sim5_var[i,]<-output$sigma
  sim5_se[i,]<-output$se
}

#simulation 6 (n=500, sigma=.3)
sim6_coef<-matrix(0,nrow=1000,ncol=4)
sim6_se<-matrix(0,nrow=1000,ncol=4)
sim6_var<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(500,.3)
  sim6_coef[i,]<-output$coef
  sim6_var[i,]<-output$sigma

```

```

    sim6_se[i,]<-output$se
}

#simulation 7 (n=500, sigma=.6)
sim7_coef<-matrix(0,nrow=1000,ncol=4)
sim7_se<-matrix(0,nrow=1000,ncol=4)
sim7_var<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(500,.6)
  sim7_coef[i,]<-output$coef
  sim7_var[i,]<-output$sigma
  sim7_se[i,]<-output$se
}

#simulation 8 (n=500, sigma=.9)
sim8_coef<-matrix(0,nrow=1000,ncol=4)
sim8_se<-matrix(0,nrow=1000,ncol=4)
sim8_var<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(500,.9)
  sim8_coef[i,]<-output$coef
  sim8_var[i,]<-output$sigma
  sim8_se[i,]<-output$se
}

#simulation 9 (n=200, sigma=.9)
sim5c_coef<-matrix(0,nrow=1000,ncol=4)
sim5c_se<-matrix(0,nrow=1000,ncol=4)
sim5c_var<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(200,.9)
  sim5c_coef[i,]<-output$coef
  sim5c_var[i,]<-output$sigma
  sim5c_se[i,]<-output$se
}

save.image("simulation_result_may31.RData")

```

C.2 Empirical Simulation

```

library(MASS)
library(foreign)
#library(np)
#setwd("C:/Users/Mahesh/Documents/Dissertation/Chapter2/June1")

```

```

setwd("Z:/Dissertation/May 31")
rm(list=ls())
# weight function to calculate nonparametric model

weight<-function(h,n,t1,a){
  s1<-(n*h)^(-1)*1/sqrt(2*pi)*exp(-t1^2/(2*h^2))
  a1<-(n*h)^(-1)*a%*(1/sqrt(2*pi)*exp(-t1^2/(2*h^2)))
  a2<-(n*h)^(-1)*a%*(1/sqrt(2*pi)*exp(-t1^2/(2*h^2))*t1)
  a3=(n*h)^-1*a%*(1/sqrt(2*pi)*exp(-t1^2/(2*h^2))*(t1^2));
  w<-s1*(a%*a3-t1*(a%*a2))/((a%*a1)*(a%*a3)-(a%*a2)^2);
  return(w)
}

#Read data
data1<-read.dta("final_data_may16.dta")
attach(data1)
n<-nrow(data1)

#Parametric variable
x1<-area
x2<-permit
#Common variables
x3<-popden

#nonparametric variable
u<-pinc
g1<-2*sin(2*pi*u)
g2<-1.5*cos(1.5*pi*u)
#par(mfrow=c(2,2))
#plot(u,g1)
#initial value to set up simulation
#sig12=.6

mysim<-function(n,sig12){

#Covariave between error term
sig12<-sig12

#seemingly unralated error term
Sigma <- matrix(c(1,sig12,sig12,1),2,2) #symmetry covariance matrix

#bivariate error matrix
error<-mvrnorm(n, rep(0, 2), Sigma)#, empirical = TRUE)

```

```

#Generate dependent variables
y1<-g1+1.5*x1+1*x3+error[,1] #From nitrogen eq.
y2<-g1+2*x2-1*x3+error[,2] #From mercury

#Generate dependent variables
#y1<-g1+1.5*x1+5*x3+error[,1]
#y2<-g2+(-2)*x1+2*x3+error[,2]

a<-rep(1,n)
t1<-a%%t(u)-u%%t(a)

#Find optimal bandwidth for each equation
h1<- npregbw(y1~u)$bw #
h2<- npregbw(y2~u)$bw #
#h1<-0.3996
#h2<-0.3296
#h1 <- npudensbw(u)$bw
#h2<-h1
# weight for fist equation

w1<-weight(h1,n,t1,a)
w2<-weight(h2,n,t1,a)

#combine parametric variable in each equation
X1<-cbind(x1,x3)
X2<-cbind(x2,x3)

#transform each equation
HX1<-X1-t(w1)%%X1
HX2<-X2-t(w2)%%X2

hx1.1=x1-t(w1)%%x1
#hx3.1=x3-t(w1)%%x3

#hx2.2=x2-t(w2)%%x2
#hx3.2=x3-t(w2)%%x3

hy1=y1-t(w1)%%y1
hy2=y2-t(w2)%%y2;

#Obtain parametric coefficient in each equation
b1<-solve(t(HX1)%%HX1)%%t(HX1)%%hy1
b2<-solve(t(HX2)%%HX2)%%t(HX2)%%hy2

```

```

#Obtain nonparametric model
g1=t(w1)%*(y1-HX1%*b1)
g2=t(w2)%*(y2-HX2%*b2)

res.1<-y1-g1-X1%*b1
res.2<-y2-g2-X2%*b2
sigma<-var(data.frame(res.1,res.2))

#tmp=sort(u,index.return=TRUE);
#u=tmp$x; idx=tmp$ix

d1=mean((hy1-HX1%*b1)^2)
d2=mean((hy2-HX2%*b2)^2)
d3=mean((hy1-HX1%*b1)*(hy2-HX2%*b2))

#s=(t(hx)%*hx)^-1*d2

#S=[[d1,d3]; [d3, d2]]^-1;
S<-solve(matrix(c(d1,d3,d3,d2), nrow=2, byrow=TRUE))
S

#X<-cbind(x,x1)
#
shg1=S[1,1]^-1*t(w1)%*(S[1,1]*(y1-X1%*b1)+S[1,2]*(y2-g2-X2%*b2))
shg2=S[2,2]^-1*t(w2)%*(S[2,2]*(y2-X2%*b2)+S[1,2]*(y1-g1-X1%*b1))
#Two stage residual variance
tres.1<-y1-shg1-X1%*b1
tres.2<-y2-shg2-X2%*b2
tres<-data.frame(tres.1,tres.2)
sigma<-var(tres)

In<-diag(n)
j<-rep(1,n)
M<-j%*%solve(t(j)%*%j)%*%j

#se for fist equation
X1.dev<-(In-M)%*%X1
sum_sq_1<-t(X1.dev)%*%X1.dev
var1<-solve(sum_sq_1)*sigma[1,1]
se1<-sqrt(diag(var1))

```

```

#se for fist equation
X2.dev<-(In-M)%*%X2
sum_sq_2<-t(X2.dev)%*%X2.dev
var2<-solve(sum_sq_2)*sigma[2,2]
se2<-sqrt(diag(var2))
#se2
b0<-c(1.5,1,2,-1)
var1<-var(tres)

b<-as.vector(cbind(t(b1),t(b2)))
se<-as.vector(cbind(se1,se2))
v<-as.vector(var1)
tval<-(b-b0)/se
pval<-ifelse(abs(tval)<1.9620076,0,1)
output<-list(coef=b,se=se,sigma=v,tval=tval,pval=pval)
return(output)
}

```

```

#-----Simulation with different condition-----#

```

```

mysim(1166,.6)

```

```

#number of simulation=500

```

```

#simulation 1 (sigma=.3)

```

```

n<-1166

```

```

sim1_coef<-matrix(0,nrow=1000,ncol=4)

```

```

sim1_var<-matrix(0,nrow=1000,ncol=4)

```

```

sim1_se<-matrix(0,nrow=1000,ncol=4)

```

```

sim1_tval<-matrix(0,nrow=1000,ncol=4)

```

```

sim1_pval<-matrix(0,nrow=1000,ncol=4)

```

```

for (i in 1:1000){

```

```

  set.seed(i)

```

```

    output<-mysim(n,.3)

```

```

    sim1_coef[i,]<-output$coef

```

```

    sim1_var[i,]<-output$sigma

```

```

    sim1_se[i,]<-output$se

```

```

    sim1_tval[i,]<-output$tval

```

```

    sim1_pval[i,]<-output$pval

```

```

  }

```

```

#simulation 2 (sigma=.6)
sim2_coef<-matrix(0,nrow=1000,ncol=4)
sim2_var<-matrix(0,nrow=1000,ncol=4)
sim2_se<-matrix(0,nrow=1000,ncol=4)
sim2_tval<-matrix(0,nrow=1000,ncol=4)
sim2_pval<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(n,.6)
  sim2_coef[i,]<-output$coef
  sim2_var[i,]<-output$sigma
  sim2_se[i,]<-output$se
  sim2_tval[i,]<-output$tval
  sim2_pval[i,]<-output$pval
}

#simulation 3 (sigma=.9)
sim3_coef<-matrix(0,nrow=1000,ncol=4)
sim3_var<-matrix(0,nrow=1000,ncol=4)
sim3_se<-matrix(0,nrow=1000,ncol=4)
sim3_tval<-matrix(0,nrow=1000,ncol=4)
sim3_pval<-matrix(0,nrow=1000,ncol=4)
for (i in 1:1000){
  output<-mysim(n,.9)
  sim3_coef[i,]<-output$coef
  sim3_var[i,]<-output$sigma
  sim3_se[i,]<-output$se
  sim3_tval[i,]<-output$tval
  sim3_pval[i,]<-output$pval
}

save.image("dis_empirical_simulation_june1a.RData")
apply(sim1_coef,2,mean)
apply(sim1_coef,2,sd)
apply(sim1_se,2,mean)
apply(sim1_pval,2,mean)
apply(sim1_var,2,mean)
apply(sim2_coef,2,mean)
apply(sim2_coef,2,sd)
apply(sim2_se,2,mean)
apply(sim2_pval,2,mean)
apply(sim2_var,2,mean)
apply(sim2_var,2,sd)
apply(sim3_coef,2,mean)
apply(sim3_coef,2,sd)
apply(sim3_se,2,mean)

```



```

apply(sim3_pval,2,mean)
apply(sim3_var,2,mean)
apply(sim3_var,2,sd)

```

C.3 Empirical Estimation

```

#-----#
#Mahesh pandit
#Final dissertation semiparametric code
# May 16, 2013
#-----#
rm(list=ls())
#load required package
library(MASS)
library(np)
library(foreign)
#define working directory
setwd("~/Dissertation/Chapter2/Final analysis")

#Defind required function

#pvalue
pval<-function(x){
  p<-(1-pt(abs(x),1163))*2
  p
}

# weight function to calculate nonparametric model
weight<-function(h,n,t1,a){
  s1<-(n*h)^(-1)*1/sqrt(2*pi)*exp(-t1^2/(2*h^2))

  #a1=(n*h)^-1*ones(1,n)*(1/sqrt(2*pi)*exp(-t1.^2/(2*h^2)));
  a1<-(n*h)^(-1)*a%*(1/sqrt(2*pi)*exp(-t1^2/(2*h^2)))

  #a2=(n*h)^-1*ones(1,n)*(1/sqrt(2*pi)*exp(-t1.^2/(2*h^2)).*t1);

  a2<-(n*h)^(-1)*a%*(1/sqrt(2*pi)*exp(-t1^2/(2*h^2))*t1)

  #a3=(n*h)^-1*ones(1,n)*(1/sqrt(2*pi)*exp(-t1.^2/(2*h^2)).*(t1.^2));
  a3=(n*h)^-1*a%*(1/sqrt(2*pi)*exp(-t1^2/(2*h^2))*(t1^2));

  #w=s1.*(ones(n,1)*a3-t1.*(ones(n,1)*a2))./((ones(n,1)*a1).*...
  #      (ones(n,1)*a3)-(ones(n,1)*a2).^2);

```

```

w<-s1*(a%%a3-t1*(a%%a2))/((a%%a1)*(a%%a3)-(a%%a2)^2);
return(w)
}

#import data
data1<-read.dta("final_data_may16.dta")
#data1$N<-data1$n
attach(data1)

#define number of observation
n<-nrow(data1)
a<-rep(1,n)

#defind t matrix
t1<-a%%t(pinc)-pinc%%t(a)

#Find optimal bandwidth for each equation
h1<- 0.12477152 #npregbw(N~pinc,regtype="ll")$bw
h2<-0.1943902 #npregbw(p~pinc)$bw
h3<-0.32447222 #npregbw(do~pinc)$bw
h4<-0.130 #npregbw(hg~pinc)$bw

# Weight for each four equation
w1<-weight(h1,n,t1,a)
w2<-weight(h2,n,t1,a)
w3<-weight(h3,n,t1,a)
w4<-weight(h4,n,t1,a)

#combine parametric variable in each equation
X1<-cbind(wpinc,popden,area) #Nitrogen
X2<-cbind(wpinc,popden,area) #Phosphorous
X3<-cbind(wpinc,popden,area) #Dissolved oxygen
X4<-cbind(wpinc,popden,permit) #Mercury

#transform independent variables each equation
HX1<-X1-t(w1)%%X1
HX1<-cbind(HX1,model.matrix(~factor(id)-1))
HX2<-X2-t(w2)%%X2
HX2<-cbind(HX2,model.matrix(~factor(id)-1))
HX3<-X3-t(w3)%%X3
HX3<-cbind(HX3,model.matrix(~factor(id)-1))
HX4<-X4-t(w4)%%X4
HX4<-cbind(HX4,model.matrix(~factor(id)-1))

```

```

#transform dependent variables
hy1=N-t(w1)%*%N
hy2=p-t(w2)%*%p
hy3=do-t(w3)%*%do
hy4=hg-t(w4)%*%hg

#Obtain parametric coefficient in each equation
b1<-solve(t(HX1)%*%HX1)%*%t(HX1)%*%hy1
b2<-solve(t(HX2)%*%HX2)%*%t(HX2)%*%hy2
b3<-solve(t(HX3)%*%HX3)%*%t(HX3)%*%hy3
b4<-solve(t(HX4)%*%HX4)%*%t(HX4)%*%hy4

#Obtain nonparametric model
g1=t(w1)%*%(N-HX1)%*%b1
g2=t(w2)%*%(p-HX2)%*%b2
g3=t(w3)%*%(do-HX3)%*%b3
g4=t(w4)%*%(hg-HX4)%*%b4

X1<-cbind(X1,model.matrix(~factor(id)-1))
X2<-cbind(X2,model.matrix(~factor(id)-1))
X3<-cbind(X3,model.matrix(~factor(id)-1))
X4<-cbind(X4,model.matrix(~factor(id)-1))

res.1<-N-g1-X1)%*%b1
res.2<-p-g2-X2)%*%b2
res.3<-do-g3-X3)%*%b3
res.4<-hg-g4-X4)%*%b4

res<-data.frame(res.1,res.2,res.3,res.4)
sigma<-var(res)
cor<-cor(res)

In<-diag(n)
j<-rep(1,n)
M<-j)%*%solve(t(j)%*%j)%*%j

#se for first equation
X1.dev<-(In-M)%*%X1[,1:3]
sum_sq_1<-t(X1.dev)%*%X1.dev
var1<-solve(sum_sq_1)*sigma[1,1]
se1<-sqrt(diag(var1))

```

```

se1
t1<-b1[1:3,1]/se1
p1<-pval(t1)
output1<-cbind(b1[1:3],se1,t1,p1)

#se for fist equation
X2.dev<-(In-M)%*%X2[,1:3]
sum_sq_2<-t(X2.dev)%*%X2.dev
var2<-solve(sum_sq_2)*sigma[2,2]
se2<-sqrt(diag(var2))
se2
t2<-b2[1:3,1]/se2
p2<-pval(t2)
output2<-cbind(b2[1:3],se2,t2,p2)

#se for third equation
X3.dev<-(In-M)%*%X3[,1:3]
sum_sq_3<-t(X3.dev)%*%X3.dev
var3<-solve(sum_sq_3)*sigma[3,3]
se3<-sqrt(diag(var3))
se3
t3<-b3[1:3,1]/se3
p3<-pval(t3)
output3<-cbind(b3[1:3],se3,t3,p3)
output3

#se for fourth equation
X4.dev<-(In-M)%*%X4[,1:3]
sum_sq_4<-t(X4.dev)%*%X4.dev
var4<-solve(sum_sq_4)*sigma[4,4]
se4<-sqrt(diag(var4))
se4
t4<-b4[1:3,1]/se4
p4<-pval(t4)
output4<-cbind(b4[1:3],se4,t4,p4)
output4

d1=mean((hy1-HX1)%*%b1)^2)
d2=mean((hy2-HX2)%*%b2)^2)
d3=mean((hy3-HX3)%*%b3)^2)
d4=mean((hy4-HX4)%*%b4)^2)

```

```

d5=mean((hy1-HX1%%b1)*(hy2-HX2%%b2))
d6=mean((hy1-HX1%%b1)*(hy3-HX3%%b3))
d7=mean((hy1-HX1%%b1)*(hy4-HX4%%b4))
d8=mean((hy2-HX2%%b2)*(hy3-HX3%%b3))
d9=mean((hy2-HX2%%b2)*(hy4-HX4%%b4))
d10=mean((hy3-HX3%%b3)*(hy4-HX4%%b4))

S<-solve(matrix(c(d1,d5,d6,d7,d5,d2,d8,d9,d6,d8,d3,d10,d7,d9,d10,d4),
nrow=4, byrow=TRUE))
S
y1<-N
y2<-p
y3<-do
y4<-hg

shg1=S[1,1]^-1*t(w1)%%(S[1,1]*(y1-X1%%b1)+S[1,2]*(y2-g2-X2%%b2)+
S[1,3]*(y3-g3-X3%%b3)+S[1,4]*
(y4-g4-X4%%b4))

shg2=S[2,2]^-1*t(w2)%%(S[2,2]*(y2-X2%%b2)+S[1,2]*(y1-g1-X1%%b1)+
S[2,3]*(y3-g3-X3%%b3)+S[2,4]*
(y4-g4-X4%%b4))

shg3=S[3,3]^-1*t(w3)%%(S[3,3]*(y3-X3%%b3)+S[3,2]*(y2-g2-X2%%b2)+
S[3,1]*(y1-g1-X1%%b1)+S[3,4]*
(y4-g4-X4%%b4))

shg4=S[4,4]^-1*t(w4)%%(S[4,4]*(y4-X4%%b4)+S[4,1]*(y1-g1-X1%%b1)+
S[4,3]*(y3-g3-X3%%b3)+S[4,2]*
(y2-g2-X2%%b2))

#Two stage residual variance
tres.1<-y1-shg1-X1%%b1
tres.2<-y2-shg2-X2%%b2
tres.3<-y3-shg3-X3%%b4
tres.4<-y4-shg3-X3%%b4

tres<-data.frame(tres.1,tres.2,tres.3,tres.4)
sigma2<-var(tres)

#se for second equation
X1.dev<-(In-M)%%X1[,1:3]
sum_sq_1<-t(X1.dev)%%X1.dev
var1<-solve(sum_sq_1)*sigma2[1,1]

```

```

se1<-sqrt(diag(var1))
se1
t1<-b1[1:3,1]/se1
p1<-pval(t1)
output1<-cbind(b1[1:3],se1,t1,p1)
output1

#se for fist equation
X2.dev<-(In-M)%*%X2[,1:3]
sum_sq_2<-t(X2.dev)%*%X2.dev
var2<-solve(sum_sq_2)*sigma2[2,2]
se2<-sqrt(diag(var2))
se2
t2<-b2[1:3,1]/se2
p2<-pval(t2)
output2<-cbind(b2[1:3],se2,t2,p2)
output2

#se for third equation
X3.dev<-(In-M)%*%X3[,1:3]
sum_sq_3<-t(X3.dev)%*%X3.dev
var3<-solve(sum_sq_3)*sigma2[3,3]
se3<-sqrt(diag(var3))
se3
t3<-b3[1:3,1]/se3
p3<-pval(t3)
output3<-cbind(b3[1:3],se3,t3,p3)
output3

#se for fourth equation
X4.dev<-(In-M)%*%X4[,1:3]
sum_sq_4<-t(X4.dev)%*%X4.dev
var4<-solve(sum_sq_4)*sigma2[4,4]
se4<-sqrt(diag(var4))
se4
t4<-b4[1:3,1]/se4
p4<-pval(t4)
output4<-cbind(b4[1:3],se4,t4,p4)
output4

tmp=sort(pinc,index.return=TRUE);

```

```

pinc=tmp$x; idx=tmp$ix

d01=y1-X1%*%b1;
plot(pinc,d01[idx],type='p',pch='*', col="green")
plot(pinc,g1[idx],col="red")
plot(pinc,shg1[idx],type="l") #this looks more accurate
title('(a)g_1(u)')

d02=y2-X2%*%b2;
plot(pinc,d02[idx],type='p',pch='*', col="green", ylim=c(0,1))
plot(pinc,g2[idx],col="red") #this looks more accurate
plot(pinc,shg2[idx])

d03=y3-X3%*%b3;
plot(pinc,d03[idx],type='p',pch='*', col="green")
plot(pinc,(max(g3)-g3[idx]),col="red") #this looks more accurate
plot(pinc,(max(shg3)-shg3[idx]))

d04=y4-X4%*%b4;
plot(pinc,d04[idx],type='p',pch='*', col="green", ylim=c(0,1))
plot(pinc,g4[idx],col="red",type="l") #this looks more accurate
plot(pinc,shg4[idx],type="l")

#find standard error to plot it

z1=1/(h1*sqrt(2*pi))*exp(-(pinc%*%t(a)-a%*%t(pinc))^2/h1^2)
z2=1/(h2*sqrt(2*pi))*exp(-(pinc%*%t(a)-a%*%t(pinc))^2/h2^2)
z3=1/(h3*sqrt(2*pi))*exp(-(pinc%*%t(a)-a%*%t(pinc))^2/h3^2)
z4=1/(h4*sqrt(2*pi))*exp(-(pinc%*%t(a)-a%*%t(pinc))^2/h4^2)

s1=(z1%*%(a)/n)^-1;
s2=(z2%*%(a)/n)^-1;
s3=(z3%*%(a)/n)^-1;
s4=(z4%*%(a)/n)^-1;

zs1=sqrt(1/(n*h1)*0.2821*d1*s1);
zs2=sqrt(1/(n*h2)*0.2821*d2*s2);
zs3=sqrt(1/(n*h3)*0.2821*d3*s3);
zs4=sqrt(1/(n*h4)*0.2821*d4*s4);

zsz1=sqrt(1/(n*h1)*0.2821*S[1,1]^-1*s1);
zsz2=sqrt(1/(n*h2)*0.2821*S[2,2]^-1*s2);

```

```

zsz3=sqrt(1/(n*h3)*0.2821*S[3,3]^-1*s3);
zsz4=sqrt(1/(n*h4)*0.2821*S[3,3]^-1*s4);

#subplot(1,2,1)
#plot(u, ws1, u,wws1,'--')

pdf("npplot.pdf")
op<-par(mfrow=c(2,2))

#Nitrogen
#g1
plot(pinc,g1[idx],col="black",type="l",xlim=c(0.6,2),ylim=c(0,1.5),xlab="Income
($0,000)",ylab="Nitrogen",lwd=2,cex.lab=1.2,cex.axis=1.2) #main="(a)",
lines(pinc,g1[idx]+2*zs1,col=4,lty=2, lwd=2)
lines(pinc,g1[idx]-2*zs1,col=4,lty=2, lwd=2)
text(1,0.08,"h =0.124")

#Phosphorous
plot(pinc,shg2[idx],col="1",type="l",ylim=c(0.06,.14),xlim=c(0.6,2),xlab="Income
($0,000)",ylab="Phosphorous",lwd=2,cex.lab=1.2,cex.axis=1.2) #main="(b)",
lines(pinc,shg2[idx]+2*zs2,col=4,lwd=2,lty=2)
lines(pinc,shg2[idx]-2*zs2,col=4,lwd=2,lty=2)
text(1,0.065,"h = 0.194")
#Dissolved oxygen
#g1
plot(pinc,-g3[idx],col="1",type="l",xlab="Income ($0,000)",
ylab="Dissolved Oxygen",axes=F,ylim=c(-8,-5.8),lwd=2,cex.lab=1.2) #main="(c)",
axis(1,cex.axis=1.2)
ax<-seq(6,8,.5)
axis(2,-ax,ax,cex.axis=1.2)
box()
lines(pinc,-g3[idx]+2*zs3, col=4,lwd=2,lty=2)
lines(pinc,-g3[idx]-2*zs3,col=4,lwd=2,lty=2)
text(1,-7.9,"h = 0.324")
#Mercury
#g1
plot(pinc,g4[idx],col="1",type="l",ylim=c(0,.25),xlab="Income ($0,000)",
ylab="Mercury",xlim=c(0.6,2),cex.axis=1.2, cex.lab=1.2,lwd=2) #main="(d)",
lines(pinc,g4[idx]+2*zs4,col=4,lwd=2,lty=2)
lines(pinc,g4[idx]-2*zs4,col=4,lwd=2,lty=2)
text(1,0.012,"h = 0.130")

```


APPENDIX D. R CODE FOR CHAPTER 3

```
library(foreign)
library(mgcv)
library(crs)

#-----with civil liberties and political rights-----#
data2<-read.dta("data_all.dta")
data2$fecal_coliform<-data2$fecal_coliform/10000
data2$coliforms<-data2$coliforms/10000
attach(data2)

myplot<-function(pdat,ylab){
  op<-par(mfrow=c(1,3))
  #gdp
  matplot(pdat[[1]][,1],pdat[[1]][,-1],
  xlab="GDP($,000)",ylab=ylab,
  lty=c(1,2,2),col=c(1,1,1),type="l")
  #Cl
  matplot(pdat[[2]][,1],pdat[[2]][,-1],
  xlab="Civil liberties",ylab=ylab,
  pch=c(19,3,3),col=c(1,1,1),type="p")
  #polrts
  matplot(pdat[[3]][,1],pdat[[3]][,-1],
  xlab="Political rights",ylab=ylab,
  pch=c(19,3,3),col=c(1,1,1),type="p")
  op
}

#-----1. fecal_coliform-----#
sem.fecal_coliform<-gam(fecal_coliform~s(gdp,bs='ps',m=3)+ordered(c1)+ordered
(polrts))
par3.fecal_coliform<-gam(fecal_coliform~gdp+gdp_squared+gdp_cubed+ordered(c1)
+ordered(polrts)) #parametric cubed
par2.fecal_coliform<-gam(fecal_coliform~gdp+gdp_squared+ordered(c1)+ordered
(polrts)) #parametric squared
crs.fecal_coliform<-crs(fecal_coliform~gdp+ordered(c1)+ordered(polrts))
pdat.fecal_coliform <- plot(crs.fecal_coliform,mean=TRUE,ci=TRUE,
plot.behavior="data")

#summarize results
summary(sem.fecal_coliform)
```

```

summary(par2.fecal_coliform)
summary(par3.fecal_coliform)
summary(crs.fecal_coliform)

#Model comparision
anova(sem.fecal_coliform,par3.fecal_coliform, test="Chisq")
anova(sem.fecal_coliform,par2.fecal_coliform, test="Chisq")

#plot the model
pdf("np_fecal_coliform.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot(pdat.fecal_coliform,ylab="Fecal Coliform")
dev.off()

pdf("sem_fecal_coliform.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.fecal_coliform,rug=FALSE,xlab="GDP($,000)",ylab="Fecal coliform",
col=c(1,2,2))
termplot(sem.fecal_coliform,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=
c("Civil liberties","Political rights"),ylabs=c("GDP($,000)","GDP($,000)"))
dev.off()

#-----2. mercury-----#
sem.mercury<-gam(mercury~s(gdp,bs='ps',m=3)+ordered(cl)+ordered(polrts))
par3.mercury<-gam(mercury~gdp+gdp_squared+gdp_cubed+ordered(cl)
+ordered(polrts))
#parametric cubed
par2.mercury<-gam(mercury~gdp+gdp_squared+ordered(cl)+ordered(polrts))
#parametric squared
crs.mercury<-crs(mercury~gdp+ordered(cl)+ordered(polrts))
pdat.mercury <- plot(crs.mercury,mean=TRUE,ci=TRUE,plot.behavior="data")

#summarize results
summary(sem.mercury)
summary(par2.mercury)
summary(par3.mercury)
summary(crs.mercury)

#Model comparision
anova(sem.mercury,par3.mercury, test="Chisq")
anova(sem.mercury,par2.mercury, test="Chisq")

#plot the model
pdf("np_mercury.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))

```

```

myplot(pdat.mercury,ylab="Mercury")
dev.off()

pdf("sem_mercury.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.mercury,rug=FALSE,xlab="GDP($,000)",ylab="Mercury",col=c(1,2,2))
termplot(sem.mercury,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=
c("Civil liberties","Political rights"),ylabs=c("Mercury"),"Mercury"))
dev.off()

#-----3. arsenic-----#
sem.arsenic<-gam(arsenic~s(gdp,bs='ps',m=3)+ordered(cl)+ordered(polrts))
par3.arsenic<-gam(arsenic~gdp+gdp_squared+gdp_cubed+ordered(cl)+ordered(polrts))
#parametric cubed
par2.arsenic<-gam(arsenic~gdp+gdp_squared+ordered(cl)+ordered(polrts))
#parametric squared
crs.arsenic<-crs(arsenic~gdp+ordered(cl)+ordered(polrts))
pdat.arsenic <- plot(crs.arsenic,mean=TRUE,ci=TRUE,plot.behavior="data")

#summarize results
summary(sem.arsenic)
summary(par2.arsenic)
summary(par3.arsenic)
summary(crs.arsenic)

#Model comparision
anova(sem.arsenic,par3.arsenic, test="Chisq")
anova(sem.arsenic,par2.arsenic, test="Chisq")

#plot the model
pdf("np_arsenic.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot(pdat.arsenic,ylab="Arsenic")
dev.off()

pdf("sem_arsenic.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.arsenic,rug=FALSE,xlab="GDP($,000)",ylab="Arsenic",col=c(1,2,2))
termplot(sem.arsenic,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=
c("Civil liberties","Political rights"),ylabs=c("Arsenic"),"Arsenic"))
dev.off()

#-----4. cadmium-----#
sem.cadmium<-gam(cadmium~s(gdp,bs='ps',m=3)+ordered(cl)+ordered(polrts))

```

```

par3.cadmium<-gam(cadmium~gdp+gdp_squared+gdp_cubed+ordered(cl)+ordered(polrts))
#parametric cubed
par2.cadmium<-gam(cadmium~gdp+gdp_squared+ordered(cl)+ordered(polrts))
#parametric squared
crs.cadmium<-crs(cadmium~gdp+ordered(cl)+ordered(polrts))
pdat.cadmium <- plot(crs.cadmium,mean=TRUE,ci=TRUE,plot.behavior="data")

#summarize results
summary(sem.cadmium)
summary(par2.cadmium)
summary(par3.cadmium)
summary(crs.cadmium)

#Model comparision
anova(sem.cadmium,par3.cadmium, test="Chisq")
anova(sem.cadmium,par2.cadmium, test="Chisq")

#plot the model
pdf("np_cadmium.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot(pdat.cadmium,ylab="Cadmium")
dev.off()

pdf("sem_cadmium.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.cadmium,rug=FALSE,xlab="GDP($,000)",ylab="Cadmium",col=c(1,2,2))
termplot(sem.cadmium,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=
c("Civil liberties","Political rights"),ylabs=c("Cadmium","Cadmium"))
dev.off()

#-----5. lead-----#
sem.lead<-gam(lead~s(gdp,bs='ps',m=3)+ordered(cl)+ordered(polrts))
par3.lead<-gam(lead~gdp+gdp_squared+gdp_cubed+ordered(cl)+ordered(polrts))
#parametric cubed
par2.lead<-gam(lead~gdp+gdp_squared+ordered(cl)+ordered(polrts)) #parametric squared
crs.lead<-crs(lead~gdp+ordered(cl)+ordered(polrts))
pdat.lead <- plot(crs.lead,mean=TRUE,ci=TRUE,plot.behavior="data")

#summarize results
summary(sem.lead)
summary(par2.lead)
summary(par3.lead)
summary(crs.lead)

```

```

#Model comparision
anova(sem.lead,par3.lead, test="Chisq")
anova(sem.lead,par2.lead, test="Chisq")

#plot the model
pdf("np_lead.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot(pdat.lead,ylab="Lead")
dev.off()

pdf("sem_lead.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.lead,rug=FALSE,xlab="GDP($,000)",ylab="Lead",col=c(1,2,2))
termplot(sem.lead,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=c("Civil liberties",
"Political rights"),ylabs=c("Lead","Lead"))
dev.off()

#-----6. oxygen-----#
sem.oxygen<-gam(oxygen~s(gdp,bs='ps',m=3)+ordered(cl)+ordered(polrts))
par3.oxygen<-gam(oxygen~gdp+gdp_squared+gdp_cubed+ordered(cl)+ordered(polrts))
#parametric cubed
par2.oxygen<-gam(oxygen~gdp+gdp_squared+ordered(cl)+ordered(polrts)) #
parametric squared
crs.oxygen<-crs(oxygen~gdp+ordered(cl)+ordered(polrts))
pdat.oxygen <- plot(crs.oxygen,mean=TRUE,ci=TRUE,plot.behavior="data")

#summarize results
summary(sem.oxygen)
summary(par2.oxygen)
summary(par3.oxygen)
summary(crs.oxygen)

#Model comparision
anova(sem.oxygen,par3.oxygen, test="Chisq")
anova(sem.oxygen,par2.oxygen, test="Chisq")

#plot the model
pdf("np_oxygen.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot(pdat.oxygen,ylab="Dissolved Oxygen")
dev.off()

pdf("sem_oxygen.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.oxygen,rug=FALSE,xlab="GDP($,000)",ylab="Dissolved Oxygen",col=c(1,2,2))

```

```

termplot(sem.oxygen,rug=FALSE,se=TRUE,col.term=1,col.se=1,
xlabs=c("Civil liberties","Political rights"),ylabs=c("Dissolved Oxygen",
"Dissolved Oxygen"))
dev.off()

#-----7. cod-----#
sem.cod<-gam(cod~s(gdp,bs='ps',m=3)+ordered(cl)+ordered(polrts))
par3.cod<-gam(cod~gdp+gdp_squared+gdp_cubed+ordered(cl)+ordered(polrts))
#parametric cubed
par2.cod<-gam(cod~gdp+gdp_squared+ordered(cl)+ordered(polrts))
#parametric squared
crs.cod<-crs(cod~gdp+ordered(cl)+ordered(polrts))
pdat.cod <- plot(crs.cod,mean=TRUE,ci=TRUE,plot.behavior="data")

#summarize results
summary(sem.cod)
summary(par2.cod)
summary(par3.cod)
summary(crs.cod)

#Model comparision
anova(sem.cod,par3.cod, test="Chisq")
anova(sem.cod,par2.cod, test="Chisq")

#plot the model
pdf("np_cod.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot(pdat.cod,ylab="COD")
dev.off()

pdf("sem_cod.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.cod,rug=FALSE,xlab="GDP($,000)",ylab="COD",col=c(1,2,2))
termplot(sem.cod,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=c("Civil liberties",
"Political rights"),ylabs=c("COD","COD"))
dev.off()

#-----8. bod-----#
sem.bod<-gam(bod~s(gdp,bs='ps',m=3)+ordered(cl)+ordered(polrts))
par3.bod<-gam(bod~gdp+gdp_squared+gdp_cubed+ordered(cl)+ordered(polrts))
#parametric cubed
par2.bod<-gam(bod~gdp+gdp_squared+ordered(cl)+ordered(polrts)) #parametric squared
crs.bod<-crs(bod~gdp+ordered(cl)+ordered(polrts))
pdat.bod <- plot(crs.bod,mean=TRUE,ci=TRUE,plot.behavior="data")

```

```

#summarize results
summary(sem.bod)
summary(par2.bod)
summary(par3.bod)
summary(crs.bod)

#Model comparision
anova(sem.bod,par3.bod, test="Chisq")
anova(sem.bod,par2.bod, test="Chisq")

#plot the model
pdf("np_bod.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot(pdat.bod,ylab="BOD")
dev.off()

pdf("sem_bod.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.bod,rug=FALSE,xlab="GDP($,000)",ylab="BOD",col=c(1,2,2))
termplot(sem.bod,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=c("Civil liberties",
"Political rights"),ylabs=c("BOD","BOD"))
dev.off()

#-----9. coliforms-----#
sem.coliforms<-gam(coliforms~s(gdp,bs='ps',m=3)+ordered(cl)+ordered(polrts))
par3.coliforms<-gam(coliforms~gdp+gdp_squared+gdp_cubed+ordered(cl)+ordered
(polrts)) #parametric cubed
par2.coliforms<-gam(coliforms~gdp+gdp_squared+ordered(cl)+ordered(polrts))
#parametric squared
crs.coliforms<-crs(coliforms~gdp+ordered(cl)+ordered(polrts))
pdat.coliforms <- plot(crs.coliforms,mean=TRUE,ci=TRUE,plot.behavior="data")

#summarize results
summary(sem.coliforms)
summary(par2.coliforms)
summary(par3.coliforms)
summary(crs.coliforms)

#Model comparision
anova(sem.coliforms,par3.coliforms, test="Chisq")
anova(sem.coliforms,par2.coliforms, test="Chisq")

#plot the model
pdf("np_coliforms.pdf",height=2, width=6)

```

```

op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot(pdat.coliforms,ylab="Coliforms")
dev.off()

pdf("sem_coliforms.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.coliforms,rug=FALSE,xlab="GDP($,000)",ylab="Coliforms",col=c(1,2,2))
termplot(sem.coliforms,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=
c("Civil liberties","Political rights"),ylabs=c("Coliforms","Coliforms"))
dev.off()

detach(data2)

#-----With trust variable-----#
#fecal coliform
data1<-read.dta("finaldata_chap3_1.dta")
data1$mfecal_coliform<-data1$mfecal_coliform/10000
data1$mcoliforms<-data1$mcoliforms/10000
attach(data1)

myplot1<-function(pdat,ylab){
op<-par(mfrow=c(1,3))
#gdp
matplot(pdat[[1]][,1],pdat[[1]][,,-1],
xlab="GDP($,000)",ylab=ylab,
lty=c(1,2,2),col=c(1,1,1),type="l")
#trust
matplot(pdat[[2]][,1],pdat[[2]][,,-1],
xlab="Trust",ylab=ylab,
pch=c(19,3,3),col=c(1,1,1),type="p")
op
}

#-----fecal coliforms-----#
sem.fecal_coliform1<-gam(mfecal_coliform~s(mgdp,bs='ps',m=3)+factor(trust))
par3.fecal_coliform1<-gam(mfecal_coliform~mgdp+mgdp_squared+mgdp_cubed+
factor(trust)) #parametric cubed
par2.fecal_coliform1<-gam(mfecal_coliform~mgdp+mgdp_squared+factor(trust))
#parametric squared
crs.fecal_coliform1<-crs(mfecal_coliform~mgdp+factor(trust))
pdat.fecal_coliform1 <- plot(crs.fecal_coliform1,mean=TRUE,ci=TRUE,plot.behavior
="data")

#summarize results
summary(sem.fecal_coliform1)

```



```

summary(par2.fecal_coliform1)
summary(par3.fecal_coliform1)
summary(crs.fecal_coliform1)

#Model comparision
anova(sem.fecal_coliform1,par3.fecal_coliform1, test="Chisq")
anova(sem.fecal_coliform1,par2.fecal_coliform1, test="Chisq")

#plot the model
pdf("np_fecal_coliform1.pdf",height=2, width=6)
op<-par(mfrow=c(1,2),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot1(pdat.fecal_coliform1,ylab="Fecal Coliform")
dev.off()

pdf("sem_fecal_coliform1.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.fecal_coliform1,rug=FALSE,xlab="GDP($,000)",ylab="Fecal coliform",
col=c(1,2,2))
termplot(sem.fecal_coliform1,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=
c("Trust"),ylabs=c("Fecal Coliform"))
dev.off()

#-----oxygen-----#
sem.oxygen1<-gam(moxygen~s(mgdp,bs='ps',m=3)+factor(trust))
par3.oxygen1<-gam(moxygen ~mgdp+mgdp_squared+mgdp_cubed+factor(trust))
#parametric cubed
par2.oxygen1<-gam(moxygen ~mgdp+mgdp_squared+factor(trust)) #parametric squared
crs.oxygen1<-crs(moxygen ~mgdp+factor(trust))
pdat.oxygen1 <- plot(crs.oxygen1,mean=TRUE,ci=TRUE,plot.behavior="data")

#summarize results
summary(sem.oxygen1)
summary(par2.oxygen1)
summary(par3.oxygen1)
summary(crs.oxygen1)

#Model comparision
anova(sem.oxygen1,par3.oxygen1, test="Chisq")
anova(sem.oxygen1,par2.oxygen1, test="Chisq")

#plot the model
pdf("np_oxygen1.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot1(pdat.oxygen1,ylab="Dissolved Oxygen")

```

```

dev.off()

pdf("sem_oxygen1.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.oxygen1,rug=FALSE,xlab="GDP($,000)",ylab="Dissolved Oxygen",
col=c(1,2,2))
termplot(sem.oxygen1,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=c("Trust")
,ylabs=c("Dissolved Oxygen"))
dev.off()

#-----BOD-----#
sem.bod1<-gam(mbod~s(mgdp,bs='ps',m=3)+factor(trust))
par3.bod1<-gam(mbod ~mgdp+mgdp_squared+mgdp_cubed+factor(trust))
#parametric cubed
par2.bod1<-gam(mbod ~mgdp+mgdp_squared+factor(trust)) #parametric squared
crs.bod1<-crs(mbod ~mgdp+factor(trust))
pdat.bod1 <- plot(crs.bod1,mean=TRUE,ci=TRUE,plot.behavior="data")

#summarize results
summary(sem.bod1)
summary(par2.bod1)
summary(par3.bod1)
summary(crs.bod1)

#Model comparision
anova(sem.bod1,par3.bod1, test="Chisq")
anova(sem.bod1,par2.bod1, test="Chisq")

#plot the model
pdf("np_bod1.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
myplot1(pdat.bod1,ylab="BOD")
dev.off()

pdf("sem_bod1.pdf",height=2, width=6)
op<-par(mfrow=c(1,3),mai=c(0.6732, 0.5412, 0.1, 0.1))
plot(sem.bod1,rug=FALSE,xlab="GDP($,000)",ylab="BOD",col=c(1,2,2))
termplot(sem.bod1,rug=FALSE,se=TRUE,col.term=1,col.se=1,xlabs=c("Trust"),
ylabs=c("BOD"))
dev.off()

save.image("final_chapter3.RData")

```

VITA

Mahesh Pandit was born in Sundarbazar-2, Lamjung, Nepal. He received his high school education at Shree Adarsha Bal Secondary School, Sundarbazar-4, Lamjung, Nepal. The author graduated from Tribhuvan University, Nepal, where he received a Bachelor of Science degree in Statistics in 2004 and his Master of Science degree in Statistics in 2007. In August 2008, he entered the masters program in Agricultural Economics and Agribusiness at Louisiana State University. He obtained an M.S. degree in Agricultural Economics in August, 2010. He then continued his education by entering the Ph.D degree program in the Department of Agricultural Economics and Agribusiness at Louisiana State University. He obtained a Masters in Applied Statistics from Louisiana State University in May 2013. He has published three journal articles and one research bulletin. Currently, five manuscripts are under review for publication in various academic journals. He has also presented 19 research papers in national and international conferences. He is currently a candidate for the degree of Doctor of Philosophy which will be awarded in August 2013.