TESTING SERIAL CORRELATION IN PARTIALLY LINEAR ADDITIVE MODELS

JIN YANG

M.Phil

The Hong Kong Polytechnic University

2014

The Hong Kong Polytechnic University
Department of Applied Mathematics

Testing Serial Correlation in Partially Linear Additive Models

Jin Yang

A thesis submitted in partial fulfilment of
the requirements for the degree of Master of Philosophy

May 2014

CERTIFICATE OF ORIGINALITY

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgment is made in the text.

_____(Signed)

_____Jin Yang_____(Name of student)

Abstract

This thesis proposes procedures for testing serial correlation in the partially linear additive models without and with errors in variables, which include the partially linear models and additive models as their special cases.

For the partially linear additive models without errors, an empirical-likelihood-based procedure is developed based on the profile least-squares method. It is shown that the proposed test statistic is asymptotically chi-square distributed under the null hypothesis of no serial correlation. Then the rejection region can be constructed using this result. It is noted that the procedures are not only for testing zero first-order serial correlation, but also for testing higher-order serial correlation.

For the partially linear additive models with errors, the methods based on the profile least-squares is invalid because of the existence of the errors in variables. By a corrected profile least-squares approach, another empirical-likelihood-based procedure is developed. The asymptotic properties are investigated, based on which the rejection region can be easily constructed.

Extensive simulation studies were conducted to assess the finite sample properties of the proposed procedures' sizes and powers.

Acknowledgments

First and foremost, I would like to express my deepest and sincerest gratitude to my supervisors, Dr. Liu Chun-ling Catherine and Prof. Wong Heung, for their support, encouragement, patience and guidance during my graduate studies. They had been so kind and patient that kept me on the academic track. Without their help, this thesis could not have been completed.

I would show my appreciation to Prof. Wei Chuan-hua, the collaborator for my two pieces of work to appear. He gave me many useful suggestions and comments on my research when he visited our department. I have benefited a lot from him.

Furthermore, I would like to thank all the research students in the Department of Applied Mathematics, for the inspiring academic discussions and useful assistance over the past two years.

I gratefully acknowledge The Hong Kong Polytechnic University for the financial support during the entire period of my candidature. I also express my gratitude to the supporting staffs in Department of Applied Mathematics for their kind help.

Last but not the least, I would like to express my special thanks to my parents and my friends for their love, encouragement and support.

Contents

Abstract. **iv**

Acknowledgment. **v**

Chapter 1 Introduction **2**

1.1 Partially Linear Additive Models 2

1.2 Serial Correlation 6

1.3 Empirical Likelihood 8

1.4 Outline of the Thesis 10

Chapter 2 Testing Serial Correlation in Partially Linear Additive Models **12**

2.1 Introduction 12

2.2 Test Statistic and Its Properties 13

2.2.1 Profile Least-squares Estimation 13

2.2.2 Test Statistic and Its Properties 18

2.3 Simulation Studies 20

2.4 Proof of the Theorem 21

Chapter 3 Testing Serial Correlation in Partially Linear Additive Models with Errors in Variables **38**

3.1 Introduction 38

3.2 Test Statistic and Its Properties 39

3.2.1 Corrected Profile Least-squares Estimation 43

3.2.2 The Empirical Log-likelihood Ratio Statistic 45

3.3 Simulation Studies 46

3.4 Proof of the Theorem 48

Chapter 4 Conclusion and Future Work **60**

4.1 Conclusion 60

4.2 Future Work 61

References **62**

List of Tables

2.1 Size and power for $H_0 : \rho = 0$ when $\varepsilon_i = \rho\varepsilon_{i-1} + \mu_i$

2.2 Size and power for $H_0 : \rho = 0$ when $\varepsilon_i = \rho u_{i-1} + \mu_i$.

2.3 Size and power for $H_0 : \alpha_1 = \alpha_2 = 0$ when $\varepsilon_i = \alpha_1\varepsilon_{i-1} + \alpha_2\varepsilon_{i-2} + \mu_i$.

2.4 Size and power for $H_0 : \alpha_1 = \alpha_2 = 0$ when $\varepsilon_i = \alpha_1 u_{i-1} + \alpha_2 u_{i-2} + \mu_i$.

3.1 Size and power for $H_0 : \rho = 0$ when $\varepsilon_i = \rho\varepsilon_{i-1} + \mu_i$

3.2 Size and power for $H_0 : \rho = 0$ when $\varepsilon_i = \rho u_{i-1} + \mu_i$.

3.3 Size and power for $H_0 : \alpha_1 = \alpha_2 = 0$ when $\varepsilon_i = \alpha_1\varepsilon_{i-1} + \alpha_2\varepsilon_{i-2} + \mu_i$.

3.4 Size and power for $H_0 : \alpha_1 = \alpha_2 = 0$ when $\varepsilon_i = \alpha_1 u_{i-1} + \alpha_2 u_{i-2} + \mu_i$.

Chapter 1

Introduction

1.1 Partially Linear Additive Models

When investigating relationship between a response Y and some covariates $\mathbf{X}$, the classical linear regression model is frequently used, which assumes a linear form for the regression function. The corresponding theory has been investigated deeply, where details can be found in a classical monograph by Rao (1973). Despite its long history, the classical linear regression model is still in the focus of modern statistical study; See, for example, Fan and Li (2001), Zou (2006), Fan and Lv (2008) and so on.

In spite of the wide applications, the classical linear regression model has some limitations. For example, the linear regression function is too restrictive and does not allow the nonlinear effects of the covariates and so on. To overcome these limitations, statisticians proposed many more flexible regression models to model the relationship of Y and $\mathbf{X}$, such as the generalized linear models by Nelder

and Baker (1972), partially linear models by Speckman (1988), additive models by Friedman and Stuetzle (1981), single-index models by Hardle *et al.* (1993), partially linear single-index models by Carroll *et al.* (1997) and so on. In the partially linear models, the conditional mean of the response is assumed to depend on some covariates $\mathbf{X}$ parametrically and some other covariates $\mathbf{Z}$ nonparametrically. Usually, the effects of $\mathbf{X}$ (e.g., treatment) are of major interest, while the effects of $\mathbf{Z}$ (e.g., confounders) are nuisance parameters. This model provides a nice trade-off between model interpretability and flexibility. Additive models are a popular and flexible class of nonparametric regression models, which assume that the conditional mean function can be represented as the sum of several scalar nonparametric functions of each components of the covariates. Because they allow multidimensional smoothing to reduce to a sequence of one-dimensional smoothing steps, additive models allow analysis of multidimensional problems which would be arduous or even impossible to approach with "full-dimensional" nonparametric methods. They also maintain the ease of interpretation of univariate nonparametric smooths, since the estimates of the component functions can be plotted separately.

Combining the advantages of the partially linear models and additive models, the partially linear additive models (PLAMs, by Liang *et al.* (2008)) assume that

$$Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + m_1(Z_1) + \cdots + m_q(Z_q) + \varepsilon, \tag{1.1}$$

where Y is a scalar response variable, $\mathbf{X}$ is a $l \times 1$ vector of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_l)^{\mathrm{T}}$ is a vector of l-dimensional unknown parameters, $Z_1, Z_2, \cdots, Z_q$ are univariate observed covariates, $m_1(\cdot), \cdots, m_q(\cdot)$ are unknown smooth functions, and ε is the random error with zero mean and finite variance σ^2. To ensure

identifiability of the nonparametric functions, we assume that $E\{m_k(Z_k)\} = 0$ for $k = 1, 2, \cdots, q$. Without loss of generality, we also assume that both Y and $\mathbf{X}$ have been centered about their respective means. As noted by many statisticians, PLAMs allow an easier interpretation of the effect of each variable and are preferable to completely nonparametric additive models, because PLAMs combine both parametric and nonparametric components when it is believed that the response variable depends on some variables in a linear way but is nonlinearly related to the remaining independent variables.

Model (1.1) has been studied well in literature. Several methods for estimating the parametric component $\boldsymbol{\beta}$ and the nonparametric functions $m_k(\cdot), k = 1, 2, \cdots, q$, have been proposed, including the backfitting method of Ospomer and Ruppert (1999), the series approach of Li (2002), the marginal integration method of Manzan and Zerom (2005), and the polynomial splines procedure of Liu (2011). Liang *et al.* (2008) proposed an attenuation-to-correction estimator of the parametric component when the linear covariate $\mathbf{X}$ is measured with additive error, and showed that the estimator is asymptotically normal and requires no undersmoothing. Jiang *et al.* (2007) applied the generalized likelihood ratio method of Fan and Jiang (2005) to test the nonparametric component. Wei and Liu (2012) proposed the restricted profile least-squares estimator for the parametric component. In addition, they proposed profile generalized likelihood ratio statistics for testing problems on the parametric component.

In the practical problems, some covariates are mismeasured frequently. The presence of measurement errors in variables causes biased and inconsistent parameter estimates and leads to erroneous conclusions to various degrees in the statistical

analysis. So the problem of measurement errors is one of the most fundamental problems in statistical inference. And there is a huge literature investigating this problem in various settings under different models. Details can be found in an excellent book by Fuller (1987). Throughout this thesis, we assume that the covariate $\mathbf{Z}$ is measured correctly without errors, and $\mathbf{X}$ is mismeasured. Typically, it is assumed that the error is additive, i.e.

$$\mathbf{V} = \mathbf{X} + \boldsymbol{\eta}, \tag{1.2}$$

where $\mathbf{V}$ is the observed covariate and $\boldsymbol{\eta}$ is the measurement error with mean zero and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$, and independent of $(Y, \mathbf{X}, Z_1, \cdots, Z_q)$. Based on (1.1) and (1.2), we have the following partially linear additive model with errors in variable or partially linear additive errors-in-variables (PLAM-EV) model

$$\begin{cases} Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + m_1(Z_1) + \cdots + m_q(Z_q) + \varepsilon, \\ \mathbf{V} = \mathbf{X} + \boldsymbol{\eta}. \end{cases} \tag{1.3}$$

Obviously, model (1.3) includes a variety of existing models. For example, when $m_1(Z_1) = \cdots = m_q(Z_q) = 0$, model (1.3) becomes the linear EV model which has been studied by Bekker (1986). When $q = 1$, model (1.3) reduces to the partially linear EV model which has been discussed by Cui and Li (1998), Liang *et al.* (1999), and Wand (1999). For model (1.3), Liang *et al.* (2008) proposed a $\sqrt{n}-$consistent estimator of $\boldsymbol{\beta}$ by a corrected profile-least squares method without undersmoothing. Liang *et al.* (2009), Wang *et al.* (2010) and Wei *et al.* (2012) have obtained the confidence region of the parametric component $\boldsymbol{\beta}$ by the empirical likelihood method. Wei and Wang (2012) proposed a novel approach to testing the parametric

components.

1.2 Serial Correlation

Ordinary, the random errors in models (1.1) and (1.3) are assumed to be independent from one observation to the next in the statistical analysis. However, this assumption is often violated in some data set, especially for the business and economic data. When the error terms from different observations are correlated, we say that the error term is serially correlated. Serial correlation happens when the errors associated with a given time period carry over into future time periods. For example, if we predicting the production of food in a region, an overestimate in one year is likely to lead to overestimates in succeeding years. Suppose that ε_i is the error for the ith subject or time. Then the serial correlation can be modelled by a p-th order moving average, denoted by MA(p), i.e.,

$$\varepsilon_i = \mu_i + \alpha_1\mu_{i-1} + \cdots + \alpha_p\mu_{i-p}, \quad \mu_i \quad \text{i.i.d} \quad (0, \sigma^2), \tag{1.4}$$

or by a p-th order autoregression, denoted by AR(p), i.e.,

$$\varepsilon_i = \mu_i + \alpha_1\varepsilon_{i-1} + \cdots + \alpha_p\varepsilon_{i-p}, \quad \mu_i \quad \text{i.i.d.} \quad (0, \sigma^2), \tag{1.5}$$

where α_i satisfies the stationary condition that the roots of equation $\alpha(\mu) = 1-\alpha_1\mu-\alpha_2\mu^2-\cdots-\alpha_p\mu^p = 0$ lie outside the unit circle. When $p = 1$, the serial correlation is called the first order serial correlation; otherwise, the higher order serial correlation.

The consequences of serial correlation include inefficient or inconsistent estima-

tion of the regression coefficients, underestimation of the error variance, underestimation of the variance of the regression coefficients estimators and inaccurate confidence intervals and so on. In addition, strong serial correlation may be evidence of omitting important explanatory variables or functional form misspecification. So it is important to test the serial correlation for the error terms.

Most of the literature on the problem of testing serial correlation is concerned with the cases in which the regression function is parametric, in particular with linear regression. The Durbin-Watson test procedure by Durbin and Watson (1950), which can be found in any standard econometrics textbooks, is the most popular method to test first-order serial correlation. To test higher-order serial correlation, commonly used tests include the Lagrange Multiplier tests of Breusch (1978) and Godfrey (1978), the Box-Pierce test of Box and Pierce (1970), and the Ljung-Box test of Ljung and Box (1978).

Recently, some approaches have been proposed to test serial correlation in semiparametric models, see Godfrey (1978), Li and Hsiao (1998), Li and Stengos (1986), Godfrey (2007), Liu *et al.* (2008) and Hu *et al.* (2009), Zhou *et al.* (2010) and Liu *et al.* (2011). As we all know, little has been discussed on how to detect serial correlation in model (1.1) and (1.3). This thesis is ready to fill this gap.

The null hypothesis to be tested is that the errors ε_i are serially uncorrelated. The alternative hypothesis of interest is (1.4) or (1.5). Follow the idea of Liu *et al.* (2008) and Hu *et al.* (2009), let $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p)^T$ be a vector of p-dimensional coefficients, then our aim is to test whether $\boldsymbol{\alpha} = \mathbf{0}$ or not. Denote $\gamma = (\gamma_1, \cdots, \gamma_p)^{\mathrm{T}}, \gamma_k = E\varepsilon_i\varepsilon_{i+k}, \quad k = 1, \cdots, p, \quad i = 1, \ldots, N$, with $N = n - p$. According to the Yule-Walker equation, it is easy to show that testing whether $\boldsymbol{\alpha} = \mathbf{0}$ or not is equivalent

to testing whether $\boldsymbol{\gamma} = \mathbf{0}$ or not. Denote $\mathbf{U}_i = (U_{i1}, \cdots, U_{ip})^{\mathrm{T}}, U_{ik} = \varepsilon_i \varepsilon_{i+k}, k = 1, 2, \cdots, p, i = 1, 2, \cdots, N, N = n - p$. Then testing the zero finite-order serial correlation in model (1.1) or (1.3) is equivalent to testing whether $E\mathbf{U}_i = \mathbf{0}$. By Owen (1990), this can be done by using the empirical likelihood method.

1.3 Empirical Likelihood

Likelihood is arguably the most significant concept for inference in parametric models and it had also been shown to be useful in nonparametric estimates of distribution functions by Kaplan and Meier (1958); Vardi (1985). Owen (1988,1990,1991), baesd on an earlier suggestion of Thomas and Grunkemeier (1975), had introduced an "empirical" likelihood ratio statistic for nonparametric problems. He also had shown that the statistics have limiting χ^2-distributions in certain situations, and shown how to get tests and confidence limits for parameters, expressed as function $\theta(F)$ of unknown distribution function F. The most attractive characteristics of the empirical likelihood method include avoiding estimation of the covariance of the estimators, developing coverage accuracy because it contains auxiliary information, and convenience of implementation.

Due to its nice qualities, the method of empirical likelihood had been widely studied in statistics and econometrics. More references and techniques can be found in the monograph by Owen (2001). Kolaczyk (1994) had made further extensions to generalized linear model. Although further investigation of this methodology is needed, especially in small to moderate size samples, it appears to provide a valuable method to tests and interval estimation in nonparametric or distribution-

free contexts.

By Owen (1991), we can construct the corresponding empirical log-likelihood ratio. However, $\varepsilon_i's$ are unknown. To solve the problem, we can replace ε_i by its estimator. Then we should estimate the parametric component $\boldsymbol{\beta}$ and the nonparametric functions $m_k(\cdot), k = 1, 2, \cdots, q$.

Several methods for estimating both components have been proposed, including the backfitting method of Ospomer and Ruppert (1999), the series approach of Li (2002), the marginal integration method of Manzan and Zerom (2005), and the polynomial splines procedure of Liu *et al.* (2011). Liang *et al.* (2008) proposed an attenuation-to-correction estimator of the parametric component when the linear covariate $\mathbf{X}$ is measured with additive error, and showed that the estimator is asymptotically normal and requires no undersmoothing. Wei and Liu (2012) proposed the restricted profile least-squares estimator for the parametric component. In addition, they proposed profile generalized likelihood ratio statistics for testing problems on the parametric component.

In this article, we assume $q = 2$ in model (1.1)and (1.2)for notational simplicity as Liang *et al.* (2008), then we can obtain a bivariate additive model. Following Opsomer and Ruppert (1997), we will use the backfitting method, which was proposed by Buja *et al.* (1989), to estimate m_1 and m_2. The related fitting procedure in S-PLUS of Chambers and Hastie (1992) have made the additive model a popular choice for multivariate nonparametric fitting. After it, we still follow Liang *et al.* (2008) and Wei and Liu (2012), to adopt the profile least-squares approach to estimate the $\boldsymbol{\beta}$.

After estimating the parameters $\boldsymbol{\beta}$ and $m_k(\cdot), k = 1, 2, \cdots, q$, we can get the estimator $\hat{e}_i = y_i - \mathbf{V}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \hat{m}_1(Z_{1i}) - \hat{m}_2(Z_{2i})$ to replace e_i. By using the Lagrange multiplier technique, we can obtain the empirical log-likelihood ratio test statistic, which is asymptotically chi-square distributed.

1.4 Outline of the Thesis

Despite the extensive research on the serial correlation under various semiparametric models, to the best of our knowledge, there is little study on the partially linear additive models. In this thesis, we investigate the serial correlation for the partially linear additive models without and with errors-in-variables by the empirical likelihood method. This thesis includes my two pieces of work. One discussed the testing serial correlation in partially linear additive models, which has been accepted for publication, see Wei and Yang (2013). The other work which also has been accepted for publication, discussed the testing serial correlation in partially linear additive errors-in-variables models, see Yang *et al.* (2013).

In Chapter 2, we propose an empirical likelihood based approach for testing serial correlation in this semiparametric model. The proposed test statistic is not only for testing zero first-order serial correlation, but also for testing higher-order serial correlation. Under the null hypothesis of no serial correlation, it is shown that the test statistic asymptotically follows a chi-square distribution. Furthermore, a simulation study is conducted to illustrate the performance of the proposed method.

In Chapter 3, based on the empirical likelihood method, a test statistic was proposed, which is asymptotical chi-square distributed under the null hypothesis of

no serial correlation. Extensive simulation studies are conducted to assess the size and power of the proposed test.

In Chapter 4, we make a brief conclusion for this thesis and point out some further research associated with the problems studied here.

Chapter 2

Testing Serial Correlation in Partially Linear Additive Models

2.1 Introduction

In this chapter, we consider the following partially linear additive models

$$Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + m_1(Z_1) + \cdots + m_q(Z_q) + \varepsilon, \tag{2.1}$$

where Y is a scalar response variable, $\mathbf{X}$ is an $l \times 1$ vector of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_l)^{\mathrm{T}}$ is a vector of l-dimensional unknown parameters, $Z_1, Z_2, \cdots, Z_q$ are univariate observed covariate, $m_1(\cdot), \cdots, m_q(\cdot)$ are unknown smooth functions, and ε is a random error with zero mean and finite variance σ^2. To ensure identifiability of the nonparametric functions, we assume that $E\{m_k(Z_k)\} = 0$ for $k = 1, 2, \cdots, q$. Without loss of generality, we also assume that both Y and $\mathbf{X}$ have been centered

about their respective means.

As mentioned in Chapter 1, there is much literature investigating this model. But the aforementioned studies mainly discuss the inference of the partially linear additive models when the errors $\varepsilon_i's$ are i.i.d. random variables. Sometimes, the observations in the sample data cannot be assumed to be independent. For example, if they are gathered sequentially in time, time series models often exhibit the phenomenon of serial correlation, where successive residuals appear to be correlated with each other. To the best of our knowledge, there is little literature on how to detect serial correlation in semiparametric partially linear additive models which are frequently used in statistical modeling. Follow the idea of Liu *et al.* (2008) and Hu *et al.* (2009), we propose an empirical-likelihood based test statistic to test finite-order serial correlation in model (1.1).

The rest of this chapter is organized as follows. In Section 2.2, we propose the empirical long-likelihood ratio test statistic and show that it is asymptotically distributed as a χ^2. Some simulations are conducted in Section 2.3 to illustrate the performance of our approach. Section 2.4 provides the proofs of the main results.

2.2 Test Statistic and Its Properties

2.2.1 Profile Least-squares Estimation

For the need of constructing the test statistic, we first develop an estimating approach for the regression coefficient in model (2.1). Following Liang *et al.* (2008) and Wei and Liu (2012), we will adopt the profile least-squares approach to estimate the

regression coefficient in model (2.1). We also assume $q = 2$ in model (2.1) for notational simplicity as Liang *et al.* (2008). Let $\{Y_i, \mathbf{X}_i, Z_{1i}, Z_{2i},\}_{i=1}^n$ be a random sample from model (2.1) with $q = 2$. Then we have

$$Y_i = \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + m_1(Z_{1i}) + m_2(Z_{2i}) + \varepsilon_i, \quad i = 1, 2, \cdots, n. \tag{2.2}$$

If the parametric component $\boldsymbol{\beta}$ is known, then model (2.2) can be rewritten as

$$Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} = m_1(Z_{1i}) + m_2(Z_{2i}) + \varepsilon_i, \quad i = 1, 2, \cdots, n. \tag{2.3}$$

Then the partially linear additive model (2.2) becomes a bivariate additive model (2.3) which has been studied by Opsomer and Ruppert (1999). For $k = 1, 2$, let

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{X}_1^{\mathrm{T}} \\ \mathbf{X}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{X}_n^{\mathrm{T}} \end{bmatrix}, \mathbf{m}_k = \begin{bmatrix} m_k(Z_{k1}) \\ m_k(Z_{k2}) \\ \vdots \\ m_k(Z_{kn}) \end{bmatrix}, \mathbf{D}_{Z_k}^k = \begin{bmatrix} 1 & (Z_{k1} - Z_k)/h_k \\ 1 & (Z_{k2} - Z_k)/h_k \\ \vdots & \vdots \\ 1 & (Z_{kn} - Z_k)/h_k \end{bmatrix},$$

where Z_k is one of $Z_{k1}, \cdots, Z_{kn}$. The smoothing matrix for local linear regression with respect to the kth covariate vector is

$$\mathbf{S}_k = \begin{bmatrix} \mathbf{e}_1^{\mathrm{T}}\{\mathbf{D}_{Z_{k1}}^{k\mathrm{T}}\mathbf{K}_{Z_{k1}}\mathbf{D}_{Z_{k1}}^k\}^{-1}\mathbf{D}_{Z_{k1}}^{\mathrm{T}}\mathbf{K}_{Z_{k1}} \\ \mathbf{e}_1^{\mathrm{T}}\{\mathbf{D}_{Z_{k2}}^{k\mathrm{T}}\mathbf{K}_{Z_{k2}}\mathbf{D}_{Z_{k2}}^k\}^{-1}\mathbf{D}_{Z_{k2}}^{\mathrm{T}}\mathbf{K}_{Z_{k2}} \\ \vdots \\ \mathbf{e}_1^{\mathrm{T}}\{\mathbf{D}_{Z_{kn}}^{k\mathrm{T}}\mathbf{K}_{Z_{kn}}\mathbf{D}_{Z_{kn}}^k\}^{-1}\mathbf{D}_{Z_{kn}}^{\mathrm{T}}\mathbf{K}_{Z_{kn}} \end{bmatrix},$$

with $\mathbf{e}_1 = (1,0)^{\mathrm{T}}$ and $\mathbf{K}_{Z_k} = \mathrm{diag}\{K_{h_k}(Z_{k1}-Z_k), K_{h_k}(Z_{k2}-Z_k), \cdots, K_{h_k}(Z_{kn}-Z_k)\}$, where $K_{h_k}(\cdot) = K(\cdot/h_k)/h_k$, $K(\cdot)$ is a kernel function and h_k is a bandwidth. By Opsomer and Ruppert (1997), we know that the unknown nonparametric functions $\mathbf{m}_k$'s can be estimated by solving the estimating equation system

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{S}_1^* \\ \mathbf{S}_2^* & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1^* \\ \mathbf{S}_2^* \end{bmatrix} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{2.4}$$

where $\mathbf{S}_k^* = (\mathbf{I}_n - \mathbf{1}\mathbf{1}^{\mathrm{T}})\mathbf{S}_k$ with $k = 1, 2$. Furthermore, we can obtain the explicit expression of backfitting estimators for $\mathbf{m}_1$ and $\mathbf{m}_2$. From (2.4), we get

$$\mathbf{I}_n\mathbf{m}_1 + \mathbf{S}_1^*\mathbf{m}_2 = \mathbf{S}_1^*(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{2.5}$$

$$\mathbf{S}_2^*\mathbf{m}_1 + \mathbf{I}_n\mathbf{m}_2 = \mathbf{S}_2^*(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \tag{2.6}$$

Then

$$\mathbf{I}_n \cdot (2.6) - \mathbf{S}_2^* \cdot (2.5),$$

we obtain

$$(\mathbf{I}_n\mathbf{S}_2^* - \mathbf{S}_2^*\mathbf{I}_n)\mathbf{m}_1 + (\mathbf{I}_n\mathbf{I}_n - \mathbf{S}_2^*\mathbf{S}_1^*)\mathbf{m}_2 = (\mathbf{I}_n\mathbf{S}_2^* - \mathbf{S}_2^*\mathbf{S}_1^*)(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Since

$$\mathbf{I}_n\mathbf{S}_2^* = \mathbf{S}_2^* = \mathbf{S}_2^*\mathbf{I}_n,$$

$$\mathbf{I}_n\mathbf{I}_n = \mathbf{I}_n,$$

it follows that

$$
\begin{aligned}
(\mathbf{I}_n - \mathbf{S}_2^*\mathbf{S}_1^*)\mathbf{m}_2 &= (\mathbf{S}_2^* - \mathbf{S}_2^*\mathbf{S}_1^*)(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= (\mathbf{I}_n - \mathbf{S}_2^*\mathbf{S}_1^* - \mathbf{I}_n + \mathbf{S}_2^*)(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= [(\mathbf{I}_n - \mathbf{S}_2^*\mathbf{S}_1^*) - (\mathbf{I}_n - \mathbf{S}_2^*)](\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{m}_2 &= (\mathbf{I}_n - \mathbf{S}_2^*\mathbf{S}_1^*)^{-1}[(\mathbf{I}_n - \mathbf{S}_2^*\mathbf{S}_1^*) - (\mathbf{I}_n - \mathbf{S}_2^*)](\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{I}_n - (\mathbf{I}_n - \mathbf{S}_2^*\mathbf{S}_1^*)^{-1}(\mathbf{I}_n - \mathbf{S}_2^*)(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).
\end{aligned}
$$

In the same way,

$$
\mathbf{I}_n \cdot (2.5) - \mathbf{S}_1^* \cdot (2.6),
$$

we can obtain

$$
(\mathbf{I}_n\mathbf{I}_n - \mathbf{S}_1^*\mathbf{S}_2^*)\mathbf{m}_1 + (\mathbf{I}_n\mathbf{S}_1^* - \mathbf{S}_1^*\mathbf{I}_n)\mathbf{m}_2 = (\mathbf{I}_n\mathbf{S}_1^* - \mathbf{S}_1^*\mathbf{S}_2^*)(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).
$$

Since

$$
\mathbf{I}_n\mathbf{S}_1^* = \mathbf{S}_1^* = \mathbf{S}_1^*\mathbf{I}_n,
$$

we also get

$$
\begin{aligned}
(\mathbf{I}_n - \mathbf{S}_1^*\mathbf{S}_2^*)\mathbf{m}_1 &= (\mathbf{S}_1^* - \mathbf{S}_1^*\mathbf{S}_2^*)(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= (\mathbf{I}_n - \mathbf{S}_1^*\mathbf{S}_2^* - \mathbf{I}_n + \mathbf{S}_1^*)(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= [(\mathbf{I}_n - \mathbf{S}_1^*\mathbf{S}_2^*) - (\mathbf{I}_n - \mathbf{S}_1^*)](\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{m}_1 &= (\mathbf{I}_n - \mathbf{S}_1^*\mathbf{S}_2^*)^{-1}[(\mathbf{I}_n - \mathbf{S}_1^*\mathbf{S}_2^*) - (\mathbf{I}_n - \mathbf{S}_1^*)](\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{I}_n - (\mathbf{I}_n - \mathbf{S}_1^*\mathbf{S}_2^*)^{-1}(\mathbf{I}_n - \mathbf{S}_1^*)(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),
\end{aligned}
$$

$$
\hat{\mathbf{m}}_1 = \mathbf{W}_1(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \quad \hat{\mathbf{m}}_2 = \mathbf{W}_2(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{2.7}
$$

with

$$\mathbf{W}_1 = \mathbf{I}_n - (\mathbf{I}_n - \mathbf{S}_1^*\mathbf{S}_2^*)^{-1}(\mathbf{I}_n - \mathbf{S}_1^*), \quad \mathbf{W}_2 = \mathbf{I}_n - (\mathbf{I}_n - \mathbf{S}_2^*\mathbf{S}_1^*)^{-1}(\mathbf{I}_n - \mathbf{S}_2^*).$$

Substituting $\hat{\mathbf{m}}_1$ and $\hat{\mathbf{m}}_2$ into (2.2), we obtain a synthetic linear regression model

$$Y_i - \bar{Y}_i = (\mathbf{X}_i - \bar{\mathbf{X}}_i)^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_i, i = 1, 2, \cdots, n, \tag{2.8}$$

where $\bar{\mathbf{Y}} = (\bar{Y}_1, \cdots, \bar{Y}_n)^{\mathrm{T}} = \mathbf{SY}$, $\bar{\mathbf{X}} = (\bar{\mathbf{X}}_1, \cdots, \bar{\mathbf{X}}_n)^{\mathrm{T}} = \mathbf{SX}$, and $\mathbf{S} = \mathbf{W}_1 + \mathbf{W}_2$.

With the linear model (2.8), we obtain the profile least-squares estimator for $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} = \left\{\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}}_i)(\mathbf{X}_i - \bar{\mathbf{X}}_i)^{\mathrm{T}}\right\}^{-1} \sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}}_i)(Y_i - \bar{Y}_i). \tag{2.9}$$

Moreover, the final estimation of $\mathbf{m}_1$ and $\mathbf{m}_2$ can be defined as

$$\hat{\mathbf{m}}_1 = \mathbf{W}_1(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \quad \hat{\mathbf{m}}_2 = \mathbf{W}_2(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{2.10}$$

Let $\hat{\mathbf{Y}} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n)^{\mathrm{T}}$ be the vector of the fitted values of $\mathbf{Y}$ and $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_n)^{\mathrm{T}}$ be the vector of residuals. Then according to the above fitting procedure and the results in (2.9) and (2.10), we have

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{m}}_1 + \hat{\mathbf{m}}_2 = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{S}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{LY} \quad \text{and} \quad \hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}}, \tag{2.11}$$

with $\mathbf{L} = \mathbf{S} + (\mathbf{I} - \mathbf{S})\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})^{\mathrm{T}}(\mathbf{I} - \mathbf{S})\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{I} - \mathbf{S})^{\mathrm{T}}(\mathbf{I} - \mathbf{S})$.

2.2.2 Test Statistic and Its Properties

For model (2.1), we are interested in testing whether the errors ε_i are serially uncorrelated. The null hypothesis to be tested is that the errors ε_i are serially uncorrelated. The alternative hypothesis of interest is a p-th order moving average, denoted by MA(p) and written as

$$\varepsilon_i = \mu_i + \alpha_1\mu_{i-1} + \cdots + \alpha_p\mu_{i-p}, \quad \mu_i \quad \text{i.i.d} \quad (0, \sigma^2) \tag{2.12}$$

or a p-th order autoregression, denoted by AR(p) and written as

$$\varepsilon_i = \mu_i + \alpha_1\varepsilon_{i-1} + \cdots + \alpha_p\varepsilon_{i-p}, \quad \mu_i \quad \text{i.i.d.} \quad (0, \sigma^2) \tag{2.13}$$

where α_i satisfies the stationary condition that the roots of equation $\alpha(\mu) = 1 - \alpha_1\mu - \alpha_2\mu^2 - \cdots - \alpha_p\mu^p = 0$ lie outside the unit circle.

Follow the idea of Liu *et al.* (2008) and Hu *et al.* (2009), let $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p)^T$ be a vector of p-dimensional coefficients, then our aim is to test whether $\boldsymbol{\alpha} = \mathbf{0}$ or not. Denote $\gamma = (\gamma_1, \cdots, \gamma_p)^{\mathrm{T}}, \gamma_k = E\varepsilon_i\varepsilon_{i+k}, \quad k = 1, \cdots, p, \quad i = 1, \ldots, N$, with $N = n - p$. According to the Yule-Walker equation, it is easy to show that testing whether $\boldsymbol{\alpha} = \mathbf{0}$ or not is equivalent to testing whether $\boldsymbol{\gamma} = \mathbf{0}$ or not.

Denote $\mathbf{U}_i = (U_{i1}, \cdots, U_{ip})^{\mathrm{T}}, U_{ik} = \varepsilon_i\varepsilon_{i+k}, k = 1, 2, \cdots, p, i = 1, 2, \cdots, N, N = n-p$. Then testing the zero finite-order serial correlation in model (2.1) is equivalent to testing whether $E\mathbf{U}_i = \mathbf{0}$. By Owen (1990), this can be done using the empirical likelihood method. Let $p_1, p_2, \cdots, p_N$ be nonnegative numbers summing to unity.

Then the corresponding empirical log-likelihood ratio can be defined as

$$\bar{l}_N = -2\max\left\{\sum_{i=1}^{N}\log(Np_i) : \sum_{i=1}^{N} p_i\mathbf{U}_i = 0, \sum_{i=1}^{N} p_i = 1\right\}.$$

However, $\varepsilon_i's$ are unknown, then $\bar{l}_n$ cannot be used directly. To solve the problem, we can replace ε_i by its estimator. Then, using $\hat{\varepsilon}_i$ to replace ε_i, the estimated empirical log-likelihood ratio is then defined by

$$l_N = -2\max\left\{\sum_{i=1}^{N}\log(Np_i) : \sum_{i=1}^{N} p_i\boldsymbol{\xi}_i = 0, p_i \geq 0, \sum_{i=1}^{N} p_i = 1\right\}, \tag{2.14}$$

where $\boldsymbol{\xi}_i = (\xi_{i1}, \cdots, \xi_{ip})^{\mathrm{T}}, \quad \xi_{ik} = \hat{\varepsilon}_i\hat{\varepsilon}_{i+k}, \quad k = 1, 2, \cdots, p, \quad i = 1, 2, \cdots, N.$

By the Lagrange multiplier technique, the empirical log-likelihood ratio can be represented as

$$l_N = 2\sum_{i=1}^{N}\log\left(1 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\xi}_i\right), \tag{2.15}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_p)^T$ is the solution of the equation

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{\xi}_i}{1 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\xi}_i} = 0. \tag{2.16}$$

The following theorem indicates that l_N is asymptotical distributed as a χ^2-distribution.

Theorem 2.1 Under the Assumptions 1 to 4 given in Section 2.4, and the null hypothesis of no serial correlation, we have

$$l_N \xrightarrow{D} \chi_p^2,$$

where χ_p^2 is a χ^2-distribution with p degrees of freedom.

2.3 Simulation Studies

In this section, we conduct extensive simulations to illustrate the finite sample properties of the proposed test procedure.

In our simulations, the data are generated from the following bivariate additive model

$$y_i = x_i\beta + m_1(z_{1i}) + m_2(z_{2i}) + \varepsilon_i, i = 1, 2, \cdots, n,$$

where $\beta = 1, m_1(z_{1i}) = 2\sin(\pi z_{1i})$, $m_2(z_{2i}) = \exp(z_{2i}) - 3.75$, and $x_i \sim N(0,1)$, $z_{1i} \sim U(0,1)$, and $z_{2i} \sim U(0,1)$. ε_i are generated from different processes

1:AR(1) error: $\varepsilon_i = \rho\varepsilon_{i-1} + \mu_i$;

2:MA(1) error: $\varepsilon_i = \rho\mu_{i-1} + \mu_i$;

3:AR(2) error: $\varepsilon_i = \alpha_1\varepsilon_{i-1} + \alpha_2\varepsilon_{i-2} + \mu_i$;

4:MA(2) error: $\varepsilon_i = \alpha_1\mu_{i-1} + \alpha_2\mu_{i-2} + \mu_i$.

To evaluate the effect of the error distributions on our results, u_i is supposed to follow the following three different distributions, (1)$\mu_i \sim N(0, 0.5^2)$, (2)$\mu_i \sim U(-\sqrt{3}/2, \sqrt{3}/2)$, (3) $\mu_i \sim \frac{1}{8}\chi_8^2 - 1$. The Epanechnikov kernel $K(x) = 0.75(1 - x^2)\mathbf{I}_{|x|\leq 1}$ and bandwidths $h_1 = h_2 = n^{-1/5}$ are used in the simulations.

Take $\rho = 0, 0.2, -0.2, 0.5, -0.5, 0.8, -0.8$, $(\alpha_1, \alpha_2) = (0,0), (0.2, 0.3), (0.2, -0.6), (0.5, 0.5), (-0.2$ and $n = 50, 100$. For each case, 1000 replications were run and the rejection rate at the significance level $\alpha = 0.05$ was computed as the estimated size and power of our proposed test procedure. The results are presented in Tables 2.1 to 2.4. For better

visualization, we also give plots and explanations for Tables 2.1 to 2.4 at the end of this chapter.

We summarize our findings as follows. When the null hypothesis is true (that is $\rho = 0$ or $(\alpha_1, \alpha_2) = (0, 0)$), the rejection frequencies (estimated sizes) of our proposed test are close to their nominal levels 0.05 under different error distributions and error process. Under the alternative hypothesis, the rejection rate seems very robust to the variation of the type of error distribution, and increases rapidly as the alternative hypothesis deviates from the null hypothesis.

2.4 Proof of the Theorem

We begin with the following assumptions required to prove the theorem. These assumptions are quite mild and can be easily satisfied.

Assumption 1. The function $K(\cdot)$ is a bounded symmetric density function with compact support.

Assumption 2. The densities $f_k(Z_k)$ of Z_k are Lipschitz continuous and bounded away from 0, and have bounded supports $\mathbf{\Omega}_k$ for $k = 1, 2, \cdots, q$.

Assumption 3. The second derivatives of $m_k(\cdot), k = 1, 2, \cdots, q$ exist and are bounded and continuous.

Assumption 4. As $n \to \infty$, $h_k \to 0$, $nh_k/\log n \to \infty$ and $nh_k^8 \to 0$ for $k = 1, 2, \cdots, q$.

In order to prove the main results, we first introduce several properties and lemmas.

Property 2.1 $O_p(1) \cdot O_p(1) = O_p(1)$

Proof: Denote $X_n = O_p(1)$, $Y_n = O_p(1)$, then for any $\varepsilon > 0$, there exist M_ε and N_ε such that

$$\mathbf{P}(|X_n| > M_\varepsilon) < \varepsilon, \mathbf{P}(|Y_n| > N_\varepsilon) < \varepsilon.$$

Then

$$\begin{aligned}
\mathbf{P}(|X_n \cdot Y_n| > M_\varepsilon \cdot N_\varepsilon) &= \mathbf{P}(|X_n \cdot Y_n| > M_\varepsilon \cdot N_\varepsilon, |X_n| > M_\varepsilon) \\
&+\mathbf{P}(|X_n \cdot Y_n| > M_\varepsilon \cdot N_\varepsilon, |X_n| \leq M_\varepsilon) \\
&\leqslant \mathbf{P}(|X_n| > M_\varepsilon) + \mathbf{P}(|Y_n| > N_\varepsilon) \\
&< 2\varepsilon
\end{aligned}$$

So

$$O_p(1) \cdot O_p(1) = O_p(1).$$

Property 2.2 $\quad O_p(1) \cdot o_p(1) = o_p(1)$

Proof: Denote $X_n = O_p(1)$, then for any $\varepsilon > 0$, there exists M_ε such that

$$\mathbf{P}(|X_n| > M_\varepsilon) < \varepsilon.$$

Denote $Y_n = o_p(1)$, which means that for any $\delta > 0$, there exists N, when $n > N$, we have

$$\mathbf{P}(|Y_n| > \delta) < \varepsilon.$$

Then

$$\begin{aligned}
\mathbf{P}(|X_n \cdot Y_n| > M_\varepsilon \cdot \delta) &= \mathbf{P}(|X_n \cdot Y_n| > M_\varepsilon \cdot \delta, |X_n| > M_\varepsilon) \\
&+\mathbf{P}(|X_n \cdot Y_n| > M_\varepsilon \cdot \delta, |X_n| \leq M_\varepsilon) \\
&\leqslant \mathbf{P}(|X_n| > M_\varepsilon) + \mathbf{P}(|Y_n| > \delta) \\
&< 2\varepsilon.
\end{aligned}$$

So

$$O_p(1) \cdot o_p(1) = o_p(1).$$

Property 2.3 $o_p(1) \cdot o_p(1) = o_p(1)$

Proof: Denote $X_n = o_p(1)$, $Y_n = o_p(1)$, then for any $\varepsilon > 0$, there exist δ_1 and δ_2 such that

$$\mathbf{P}(|X_n| > \delta_1) < \varepsilon, \mathbf{P}(|Y_n| > \delta_2) < \varepsilon.$$

Then

$$\begin{aligned}
\mathbf{P}(|X_n \cdot Y_n| > \delta_1 \cdot \delta_2) &= \mathbf{P}(|X_n \cdot Y_n| > \delta_1 \cdot \delta_2, |X_n| > \delta_1) \\
&+ \mathbf{P}(|X_n \cdot Y_n| > \delta_1 \cdot \delta_2, |X_n| \leq \delta_1) \\
&\leqslant \mathbf{P}(|X_n| > \delta_1) + \mathbf{P}(|Y_n| > \delta_2) \\
&< 2\varepsilon.
\end{aligned}$$

So

$$o_p(1) \cdot o_p(1) = o_p(1).$$

Property 2.4 $o_p(1) + o_p(1) = o_p(1)$

Proof: Denote $X_n = o_p(1)$, $Y_n = o_p(1)$, then for any $\varepsilon > 0$, there exist δ_1 and δ_2 such that

$$\mathbf{P}(|X_n| > \delta_1) < \varepsilon, \mathbf{P}(|Y_n| > \delta_2) < \varepsilon.$$

Then

$$\begin{aligned}
\mathbf{P}(|X_n + Y_n| > \delta_1 + \delta_2) &\leqslant \mathbf{P}(|X_n| > \delta_1) + \mathbf{P}(|Y_n| > \delta_2) \\
&< 2\varepsilon.
\end{aligned}$$

So

$$o_p(1) + o_p(1) = o_p(1).$$

Lemma 2.1 For the backfitting estimation of unknown functions, we have

$$\hat{m}_k(z) - m_k(z) = O_p\left(h_k^2 + \frac{1}{\sqrt{nh_k}}\right), k = 1, 2, \cdots, q.$$

This lemma can be obtained by Theorem 1 of Wand (1999).

Lemma 2.2 Let $G_i, i = 1, 2, \cdots, n$ be i.i.d random variables with $E(G_i) = 0$ and $E(G_i^2) < \infty$. Then for any permutation $(j_1, j_2, \cdots, j_n)$ of $(1, 2, \cdots, n)$,

$$\max_{1 \le k \le n} \mid \sum_{i=1}^{k} G_{j_i} \mid = O_p(n^{\frac{1}{2}} \log n).$$

The lemma comes from Gao (1995).

Lemma 2.3 (Kolmogorov Strong Law of Large Numbers) If $\{X_n\}, n \ge 1$ is an i.i.d sequence of random variables and $S_n = \sum X_n$. Then, there exists $C \in \mathbf{R}$ such that

$$\frac{S_n}{n} \xrightarrow{a.s.} C \Leftrightarrow E(\mid X_1 \mid) < \infty, C = E(X_1).$$

Lemma 2.4 Under the Assumptions 1 to 4, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\xi}_i \xrightarrow{D} N(\mathbf{0}, \sigma^4 \mathrm{I}_p).$$

Proof: Denote $e_i = \mathbf{x}_i^T(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + m_i - \hat{m}_i$ with $m_i = m_1(Z_{1i}) + m_2(Z_{2i}), \hat{m}_i = \hat{m}_1(Z_{1i}) + \hat{m}_2(Z_{2i})$. Then, by the definition of $\boldsymbol{\xi}_i$, we have

$$\begin{aligned}
\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_{ik} &= \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_i+\varepsilon_i)(e_{i+k}\varepsilon_{i+k}) \\
&= \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\varepsilon_i\varepsilon_{i+k} + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}e_ie_{i+k} + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}e_i\varepsilon_{i+k} + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{i+k}\varepsilon_i \\
&\doteq \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}U_{ik} + I_1 + I_2 + I_3.
\end{aligned}$$

For I_1, we have

$$\begin{aligned}
I_1 &= \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\mathbf{X}_i^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) + m_i - \hat{m}_i\right]\left[\mathbf{X}_{i+k}^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) + m_{i+k} - \hat{m}_{i+k}\right] \\
&= \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\mathbf{X}_i^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\mathbf{X}_{i+k}^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right] + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[(m_i-\hat{m}_i)(m_{i+k}-\hat{m}_{i+k})\right] \\
&\quad + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\mathbf{X}_i^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(m_{i+k}-\hat{m}_{i+k})\right] + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[(m_i+\hat{m}_i)\mathbf{X}_{i+k}^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right] \\
&= I_{11} + I_{12} + I_{13} + I_{14}.
\end{aligned} \tag{2.17}$$

By Lemmas 2.1 and 2.2, we prove

$$\begin{aligned}
I_{11} &= \|\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\|^2\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{X}_{i+k}^T \\
&\le \sqrt{N}\|\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\|^2\cdot\tfrac{1}{N}\sum_{i=1}^{N}\|\mathbf{X}_i\|\|\mathbf{X}_{i+k}\| \\
&\le \sqrt{N}\|\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\|^2\cdot\left(\tfrac{1}{N}\sum_{i=1}^{N}\|\mathbf{X}_i\|^2\right)^{1/2}\cdot\left(\tfrac{1}{N}\sum_{i=1}^{N}\|\mathbf{X}_{i+k}\|^2\right)^{1/2} \\
&= \sqrt{N}\cdot O_p(N^{-1})\cdot O_p(1)\cdot O_p(1) = o_p(1).
\end{aligned} \tag{2.18}$$

For $\|\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\|^2$, $\hat{\boldsymbol{\beta}}$ is the profile least-square estimator of $\boldsymbol{\beta}$, so $E\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$. Then we get

$$\sqrt{N}(\boldsymbol{\beta} - E\hat{\boldsymbol{\beta}}) \longrightarrow N(0,\Sigma),$$

which means,

$$\boldsymbol{\beta} - E\hat{\boldsymbol{\beta}} = O_p(\frac{1}{\sqrt{N}}),$$

so

$$\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|^2 = O_p(N^{-1}).$$

For $\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{X}_i\|^2\right)^{1/2}$, by Lemma 2.3, we obtain

$$\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{X}_i\|^2\right) \xrightarrow{a.s.} E(\|\mathbf{X}_i\|^2).$$

Since $E(\|\mathbf{X}_i\|^2)$ is a constant, then

$$\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{X}_i\|^2\right) = O_p(1).$$

Similarly, we also obtain

$$\left(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{X}_{i+k}\|^2\right) = O_p(1).$$

Then

$$\begin{aligned} |I_{12}| &= |\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}(m_i - \hat{m}_i)(m_{i+k} - \hat{m}_{i+k})| \\ &\leqslant \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\sup_i|m_i - \hat{m}_i|\sup_i|m_{i+k} - \hat{m}_{i+k}| \\ &= \sqrt{N}O_p(c_n^2) = o_p(1), \end{aligned} \tag{2.19}$$

with $c_n = h_1^2 + h_2^2 + \left\{\frac{\log(1/h_1)}{nh_1}\right\}^{1/2} + \left\{\frac{\log(1/h_2)}{nh_2}\right\}^{1/2}$. For I_{13}, we have

$$\begin{aligned} |I_{13}| &\leqslant \tfrac{1}{\sqrt{N}} \left\{\sum_{i=1}^{N} [\mathbf{X}_i^T(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]^2\right\}^{\frac{1}{2}} \left\{\sum_{i=1}^{N} (m_{i+k} - \hat{m}_{i+k})^2\right\}^{\frac{1}{2}} \\ &\leqslant \sqrt{N}\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\| \left(\tfrac{1}{N}\sum_{i=1}^{N} \|\mathbf{X}_i\|^2\right)^{\frac{1}{2}} \sup_i |m_i - \hat{m}_i| \\ &= o_p(1). \end{aligned} \tag{2.20}$$

By (2.18)-(2.20), we have $I_1 = o_p(1)$.

For I_2, we have

$$I_2 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{X}_i^T(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\varepsilon_{i+k} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (m_i - \hat{m}_i)\varepsilon_{i+k} = I_{21} + I_{22}. \tag{2.21}$$

Since

$$I_{21} = \sqrt{N}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^{\mathrm{T}} \frac{1}{N} \sum_{i=1}^{N} \mathbf{X}_i \varepsilon_{i+k} = o_p(1), \tag{2.22}$$

and

$$I_{22} \leq \frac{1}{\sqrt{N}} \sup_i |m_i - \hat{m}_i| \max_{1 \leq k \leq N} |\sum_{i=1}^{k} v_{j_i}| = o_p(1), \tag{2.23}$$

we can obatin $I_2 = o_p(1)$.

Similarly, we obtain $I_3 = o_p(1)$. Then, we get

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{ik} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_{ik} + o_p(1),$$

and

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\xi}_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{U}_i + o_p(1).$$

By the same argument used in the proof of Theorem 3.1 in Hu *et al.* (2009), we get

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\xi}_i \xrightarrow{D} N(\mathbf{0}, \sigma^4 \mathrm{I}_p).$$

Lemma 2.5 Under the assumptions 1-4, we have

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}} \xrightarrow{p} \sigma^4 \mathrm{I}_p.$$

By the same argument used in the proof of Lemma 2.4, we can prove Lemma 2.5 by the law of large numbers. We omit the details here.

Proof of Theorem 2.1 Using the same strategy as the proof of Theorem 3.2 in Owen (1991), we prove that

$$\|\boldsymbol{\lambda}\| = O_p(N^{-1/2}). \tag{2.24}$$

On the other hand, based on the assumptions and Lemma 2.3, and the strong law of large numbers, we have

$$\max_{1\le i\le N}\|\boldsymbol{\xi}_i\| = o_p(N^{1/2}). \tag{2.25}$$

Based on the equation (2.16), by Lemma 2.3, (2.24) and (2.25), we have

$$\boldsymbol{\lambda} = \left(\sum_{i=1}^{N} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^{\mathrm{T}}\right)^{-1} \sum_{i=1}^{N} \boldsymbol{\xi}_i + o_p(N^{-1/2}), \tag{2.26}$$

and

$$\sum_{i=1}^{N} \boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{\xi}_i = \sum_{i=1}^{N} \left(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{\xi}_i\right)^2 + o_p(1). \tag{2.27}$$

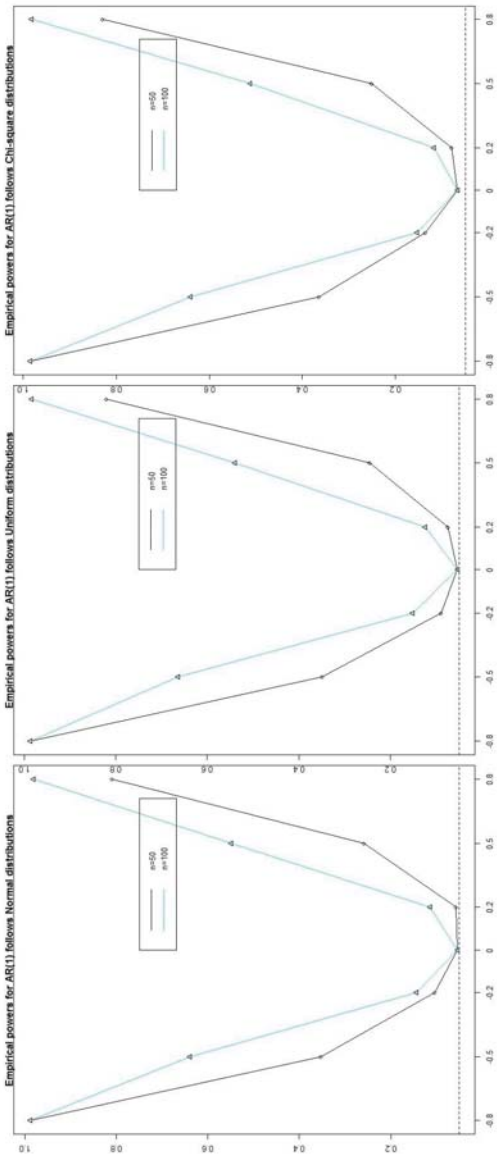
By (2.24)-(2.27), we know that

$$\begin{aligned} l_N &= \sum_{i=1}^{N} \boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^{\mathrm{T}} \boldsymbol{\lambda} + o_p(1) \\ &= \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\xi}_i\right)^{\mathrm{T}} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^{\mathrm{T}}\right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\xi}_i\right) + o_p(1). \end{aligned}$$

Finally, combining Lemmas 2.4 and 2.5, we have $l_N \xrightarrow{D} \chi_p^2$ as $N \to \infty$. The theorem is then proved.

Table 2.1. Size and power for $H_0 : \rho = 0$ when $\varepsilon_i = \rho\varepsilon_{i-1} + \mu_i$

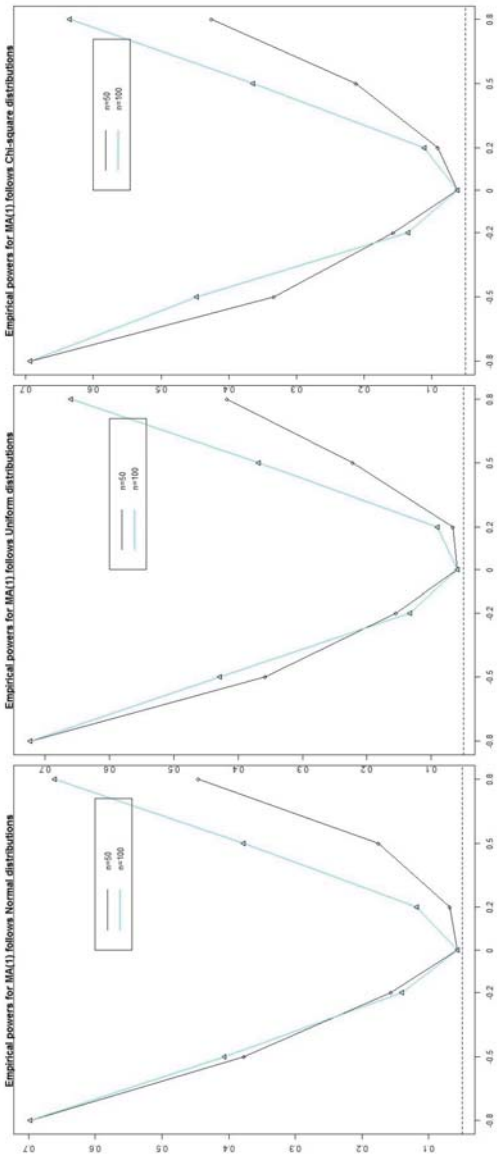
		Error		Distribution
n	ρ	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi^2_8 - 1$
n=50	0	0.067	0.059	0.063
	0.2	0.070	0.075	0.073
	-0.2	0.107	0.087	0.117
	0.5	0.230	0.210	0.208
	-0.5	0.305	0.292	0.297
	0.8	0.669	0.663	0.663
	-0.8	0.811	0.794	0.785
n=100	0	0.054	0.054	0.067
	0.2	0.114	0.124	0.117
	-0.2	0.144	0.152	0.154
	0.5	0.549	0.540	0.512
	-0.5	0.639	0.664	0.640
	0.8	0.981	0.983	0.981
	-0.8	0.988	0.986	0.984



From the figure PLAM-AR(1), we obtain that when $n = 50$ and $\rho = 0$, the rejection rate of our proposed test is close to their significance levels 0.05 under three different error distributions. When $\rho \neq 0$, the rejection rate seems very robust to the variation of the type of error distribution, and increases rapidly as the ρ's value becomes larger. When $n = 100$, we get the same conclusion and this phenomenon becomes more and more obvious.

Table 2.2. Size and power for $H_0 : \rho = 0$ when $\varepsilon_i = \rho u_{i-1} + \mu_i$.

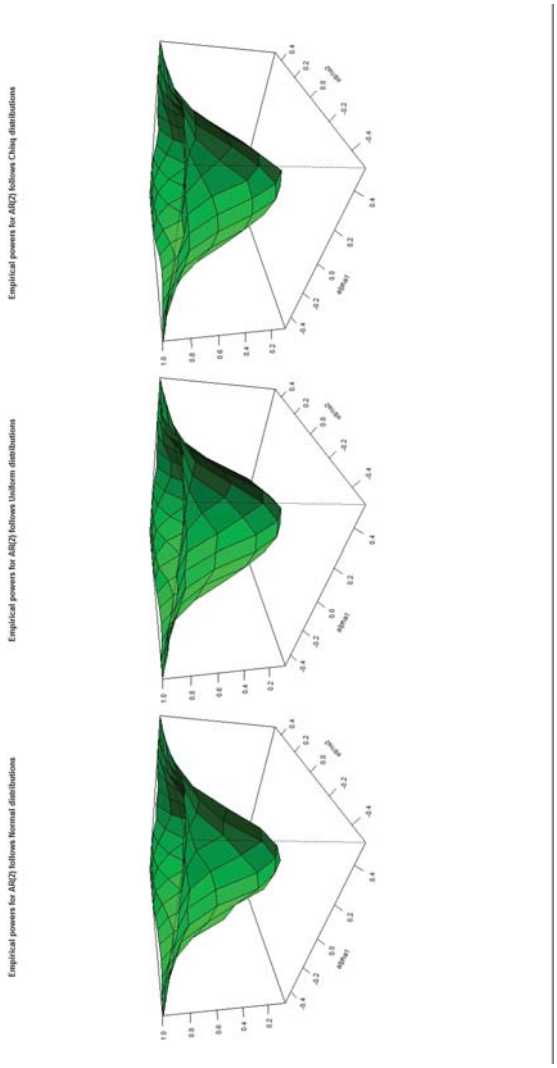
		Error		Distribution
n	ρ	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi_8^2 - 1$
n=50	0	0.073	0.070	0.068
	0.2	0.078	0.073	0.081
	-0.2	0.115	0.112	0.111
	0.5	0.123	0.142	0.136
	-0.5	0.208	0.202	0.191
	0.8	0.237	0.228	0.233
	-0.8	0.343	0.363	0.354
n=100	0	0.057	0.059	0.062
	0.2	0.118	0.09	0.111
	-0.2	0.140	0.133	0.135
	0.5	0.377	0.368	0.364
	-0.5	0.406	0.428	0.447
	0.8	0.661	0.660	0.635
	-0.8	0.697	0.723	0.693



From the figure PLAM-MA(1), we obtain that when $n = 50$ and $\rho = 0$, the rejection rate of our proposed test is close to their significance levels 0.05 under three different error distributions. When $\rho \neq 0$, the rejection rate seems very robust to the variation of the type of error distribution, and increases rapidly as the ρ's value becomes larger. When $n = 100$, we get the same conclusion and this phenomenon becomes more and more obvious.

Table 2.3. Size and power for $H_0 : \alpha_1 = \alpha_2 = 0$ when $\varepsilon_i = \alpha_1\varepsilon_{i-1} + \alpha_2\varepsilon_{i-2} + \mu_i$.

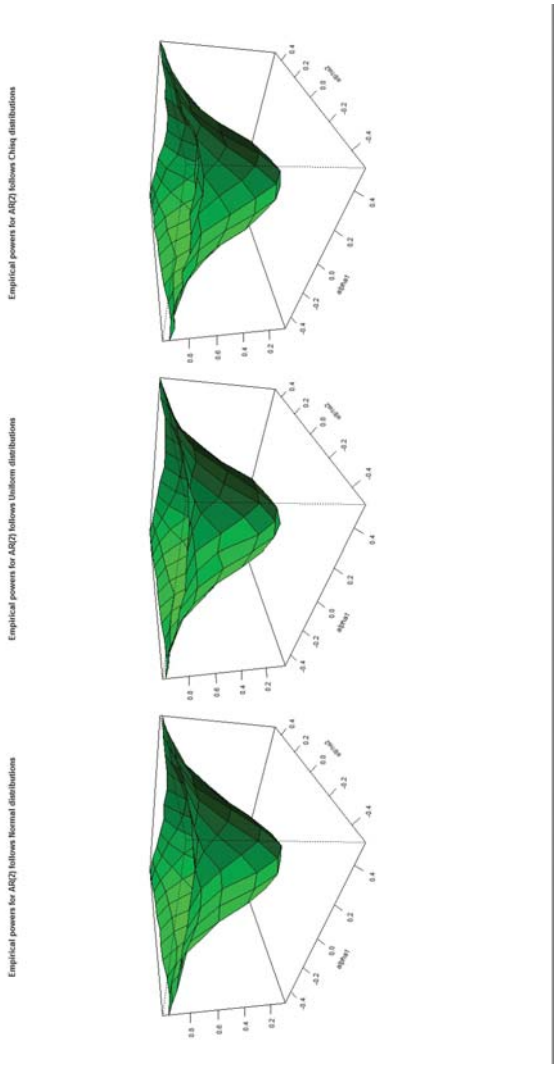
		Error		Distribution
n	(α_1, α_2)	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi_8^2 - 1$
n=50	(0,0)	0.093	0.081	0.101
	(0.2,0.3)	0.428	0.415	0.419
	(-0.2,-0.3)	0.489	0.461	0.473
	(0.2,-0.6)	0.884	0.915	0.884
	(0,0.5)	0.640	0.674	0.659
	(0.5,0.5)	0.964	0.976	0.973
n=100	(0,0)	0.073	0.061	0.070
	(0.2,0.3)	0.820	0.839	0.828
	(-0.2,-0.3)	0.772	0.807	0.780
	(0.2,-0.6)	0.999	0.999	0.999
	(0,0.5)	0.962	0.974	0.970
	(0.5,0.5)	1.000	1.000	1.000



From the figure PLAM-AR(2), we can obtain that when $n = 50$ and $(\alpha_1, \alpha_2) = (0, 0)$, the rejection rate of our proposed test are close to their significance levels 0.05 under three different error distributions. When $\rho \neq 0$, the rejection rate seems very robust to the variation of the type of error distribution, and increase rapidly as the (α_1, α_2)'s value becomes larger. When $n = 100$, we can get the same conclusion and this phenomenon becomes more and more obvious.

Table 2.4. Size and power for $H_0 : \alpha_1 = \alpha_2 = 0$ when $\varepsilon_i = \alpha_1 u_{i-1} + \alpha_2 u_{i-2} + \mu_i$.

		Error		Distribution
n	(α_1, α_2)	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi_8^2 - 1$
n=50	(0,0)	0.079	0.095	0.097
	(0.2,0.3)	0.304	0.284	0.306
	(-0.2,-0.3)	0.423	0.419	0.386
	(0.2,-0.6)	0.711	0.701	0.660
	(0,0.5)	0.482	0.548	0.479
	(0.5,0.5)	0.719	0.733	0.704
n=100	(0,0)	0.075	0.056	0.082
	(0.2,0.3)	0.627	0.667	0.648
	(-0.2,-0.3)	0.693	0.732	0.718
	(0.2,-0.6)	0.958	0.978	0.952
	(0,0.5)	0.889	0.902	0.866
	(0.5,0.5)	0.984	0.982	0.978



From the figure PLAM-MA(2), we can obtain that when $n = 50$ and $(\alpha_1, \alpha_2) = (0, 0)$, the rejection rate of our proposed test are close to their significance levels 0.05 under three different error distributions. When $\rho \neq 0$, the rejection rate seems very robust to the variation of the type of error distribution, and increase rapidly as the (α_1, α_2)'s value becomes larger. When $n = 100$, we can get the same conclusion and this phenomenon becomes more and more obvious.

Chapter 3

Testing Serial Correlation in Partially Linear Additive Models with Errors in Variables

3.1 Introduction

In this chapter, we consider the following partially linear additive models with errors in variable or partially linear additive errors-in-variables models

$$\begin{cases} Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + m_1(Z_1) + \cdots + m_q(Z_q) + \varepsilon, \\ \mathbf{V} = \mathbf{X} + \boldsymbol{\eta}, \end{cases} \tag{3.1}$$

where Y is a scalar response variable, $\mathbf{X}$ is an $l \times 1$ vector of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_l)^{\mathrm{T}}$ is a vector of l-dimensional unknown parameters, Z_k's are univariate continuous variables, $m_1(\cdot), \cdots, m_q(\cdot)$ are unknown smooth functions, and

ε is a random error with zero mean and finite variance σ^2. To ensure identifiability of the nonparametric functions, we assume that $E\{m_k(Z_k)\} = 0$ for $k = 1, 2, \cdots, q$. Measurement errors $\boldsymbol{\eta}$ are independent and identically distributed, independent of $(Y, \mathbf{X}, Z_1, \cdots, Z_q)$, with mean zero and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$. Here, we assume that $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ is known. If $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ is unknown, we can estimate it by repeatedly measuring $\mathbf{V}$. Model (3.1) includes a variety of existing models. For example, when $m_1(Z_1) = \cdots = m_q(Z_q) = 0$, model (3.1) becomes linear EV models. When $q = 1$, model (3.1) reduces to the partially linear EV model which has been discussed by Cui and Li (1998), Liang *et al.* (2007), and Wang (1999).

There is much literature studying model (3.1). Relevant discussion can be found in Chapter 1. As we know, little has been done on how to detect serial correlation in model (3.1). The present chapter is ready to fill this gap. Following the idea of Liu *et al.* (2008) and Hu *et al.* (2009), we propose an empirical-likelihood based test statistic to test finite-order serial correlation in model (3.1).

3.2 Test Statistic and Its Properties

We assume $q = 2$ in model (3.1) for notational simplicity as Liang *et al.*(2008). Suppose that $\{Y_i, \mathbf{V}_i, Z_{1i}, Z_{2i}, \}_{i=1}^n$ is a random sample of incomplete data from model (3.1) with $q = 2$. Then we have

$$\begin{cases} Y_i = \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + m_1(Z_{1i}) + m_2(Z_{2i}) + \varepsilon_i, \\ \mathbf{V}_i = \mathbf{X}_i + \boldsymbol{\eta}_i, \end{cases} \tag{3.2}$$

For model (3.2), we assume that ε_i follows p-th order moving average, denoted by MA(p) and written as

$$\varepsilon_i = \mu_i + \alpha_1 \mu_{i-1} + \cdots + \alpha_p \mu_{i-p}, \quad \mu_i \quad \text{i.i.d} \quad (0, \sigma^2),$$

or a p-th order autoregression, denoted by AR(p) and written as

$$\varepsilon_i = \mu_i + \alpha_1 \varepsilon_{i-1} + \cdots + \alpha_p \varepsilon_{i-p}, \quad \mu_i \quad \text{i.i.d.} \quad (0, \sigma^2),$$

where α_i satisfies the stationary condition that the roots of equation $\alpha(\mu) = 1 - \alpha_1 \mu - \alpha_2 \mu^2 - \cdots - \alpha_p \mu^p = 0$ lie outside the unit circle.

We are interested in testing whether the model error ε_i is serially uncorrelated. Denote $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_p)^{\mathrm{T}}, \gamma_k = E\varepsilon_i \varepsilon_{i+k}, \quad k = 1, \cdots, p, \quad i = 1, \ldots, N$, with $N = n - p$. It is easy to show that testing whether ε_i is serially uncorrelated or not is equivalent to testing whether $\boldsymbol{\gamma} = \mathbf{0}$ or not. Follow the idea of Liu *et al.* (2008) and Hu *et al.* (2009), let $e_i = \varepsilon_i - \boldsymbol{\eta}_i^{\mathrm{T}} \boldsymbol{\beta}$, $\bar{\boldsymbol{\gamma}} = (\bar{\gamma}_1, \cdots, \bar{\gamma}_p)^{\mathrm{T}}$, with $\bar{\gamma}_k = E e_i e_{i+k}$. Noting that ε_i is independent of $\boldsymbol{\eta}_i$, then under the null hypothesis of no serial correlation, we have

$$\bar{\gamma}_k = E e_i e_{i+k} = E(\varepsilon_i - \boldsymbol{\eta}_i^{\mathrm{T}} \boldsymbol{\beta})(\varepsilon_{i+k} - \boldsymbol{\eta}_{i+k}^{\mathrm{T}} \boldsymbol{\beta}) = E\varepsilon_i \varepsilon_{i+k} = \gamma_k.$$

Denote $\mathbf{U}_i = (U_{i1}, \cdots, U_{ip})^{\mathrm{T}}, U_{ik} = e_i e_{i+k}, k = 1, 2, \cdots, p, i = 1, 2, \cdots, N$, then testing the zero finite-order serial correlation in model (3.2) is equivalent to testing whether $E\mathbf{U}_i = \mathbf{0}$. By Owen (2001), this can be done by using the empirical likelihood method.

Empirical likelihood is a nonparametric method of statistical inference. It allows the data analyst to use likelihood method, without having to assume that the data comes from a known family of distribution.

Let $x_1, x_2, \cdots, x_n$ be independent random vectors in $\mathbf{R}^P$, for $P \geq 1$, with common distribution function F_0. The empirical distribution

$$F_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$$

is well known to be the nonparametric maximum likelihood estimate of F_0 based on $x_1, x_2, \cdots, x_n$. Here δ_x denotes a point mass at x. Then the likelihood function that F_n maximizes is

$$L(F) = \prod_{i=1}^{n} F\{x_i\},$$

where $F\{x_i\}$ is the probability of $\{x_i\}$ under F, x_i is the observed value, and F is any probability measure on $\mathbf{R}^P$. This motivates the term nonparametric *mle* for the estimate $T(F_n)$ of the parameter $T(F_0)$, where T is a statistical functional. In some cases the empirical likelihood ratio function

$$R(F) = L(F)/L(F_n)$$

can be used to construct nonparametric confidence regions and test for $T(F_0)$. Consider sets of the form

$$C = \{T(F)|R(F) \geq r\},$$

where C maybe used as confidence regions for $T(F_0)$. Under such conditions, a test of $T(F_0) = t$ rejects when $t \notin C$. That means, when no distribution F with $T(F) = t$

has likelihood $L(F) \geq rL(F_n)$.

Theorem 1 (Owen (1990)) Let $x, x_1, x_2, \cdots, x_n$ be $i.i.d.$ random vectors in $\mathbf{R}^P$, with $E(X) = \mu_0$ and $Var(X) = \Sigma$ of rank $q > 0$. For positive $r < 1$, let

$$C_{r,n} = \{\int xdF | F \leq F_n, R(F) \geq r\},$$

then $C_{r,n}$ is a convex set and

$$\lim_{x \to \infty} P(\mu_0 \in C_{r,n}) = P(\chi^2_{(q)} \leq -2log(r)).$$

Assume there are no ties among the x_i, let $\omega_i = F(\{x_i\}), \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1$, then

$$L(F) = \prod_{i=1}^{n} \omega_i.$$

Since mle of F is $\hat{F} = \frac{1}{n}$, we get

$$L(\hat{F}) = \prod_{i=1}^{n} \frac{1}{n}.$$

Thus

$$R(F) = L(F)/L(\hat{F}) = \prod_{i=1}^{n} n\omega_i.$$

For

$$T(F) = \int xdF = \sum_{i=1}^{n} \omega_i x_i,$$

Therefore

$$C_{r,n} = \{\sum_{i=1}^{n} \omega_i x_i | \sum_{i=1}^{n} n\omega_i \geq r, \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1\}.$$

The corresponding profile likelihood is

$$R(\mu) = \max\{\sum_{i=1}^{n} n\omega_i | \sum_{i=1}^{n} \omega_i x_i = \mu, \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1\}.$$

In our thesis, let $p_1, p_2, \cdots, p_N$ be nonnegative numbers summing to unity. Then the corresponding empirical log-likelihood ratio can be defined as

$$\bar{l}_N = -2 \max \left\{ \sum_{i=1}^{N} \log(Np_i) | \sum_{i=1}^{N} p_i \mathbf{U}_i = 0, p_i \geq 0, \sum_{i=1}^{N} p_i = 1 \right\}. \tag{3.3}$$

However, $e_i's$ are unknown, then $\bar{l}_n$ cannot be used directly. To solve the problem, we can replace e_i by its estimator. In the following, we will apply the corrected profile least-squares approach of Liang *et al.* (2008) to estimate model (3.2).

3.2.1 Corrected Profile Least-squares Estimation

For convenience, we first suppose that $\mathbf{X}$ can be observed without measurement error. If the parametric component $\boldsymbol{\beta}$ is known, then the first part of model (3.2) can be rewritten as

$$Y_i - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta} = m_1(Z_{1i}) + m_2(Z_{2i}) + \varepsilon_i, \quad i = 1, 2, \cdots, n. \tag{3.4}$$

Obviously, model (3.4) is a bivariate additive model that has been studied by Opsomer and Ruppert (1997). Let $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_n)^{\mathrm{T}}$,

$\mathbf{m}_k = (m_k(Z_{k1}), m_k(Z_{k2}), \cdots, m_k(Z_{kn}))$,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1^{\mathrm{T}} \\ \mathbf{X}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{X}_n^{\mathrm{T}} \end{bmatrix}, \mathbf{S}_k = \begin{bmatrix} \mathbf{e}_1^{\mathrm{T}}\{\mathbf{D}_{Z_{k1}}^{k\mathrm{T}}\mathbf{K}_{Z_{k1}}\mathbf{D}_{Z_{k1}}^{k}\}^{-1}\mathbf{D}_{Z_{k1}}^{\mathrm{T}}\mathbf{K}_{Z_{k1}} \\ \mathbf{e}_1^{\mathrm{T}}\{\mathbf{D}_{Z_{k2}}^{k\mathrm{T}}\mathbf{K}_{Z_{k2}}\mathbf{D}_{Z_{k2}}^{k}\}^{-1}\mathbf{D}_{Z_{k2}}^{\mathrm{T}}\mathbf{K}_{Z_{k2}} \\ \vdots \\ \mathbf{e}_1^{\mathrm{T}}\{\mathbf{D}_{Z_{kn}}^{k\mathrm{T}}\mathbf{K}_{Z_{kn}}\mathbf{D}_{Z_{kn}}^{k}\}^{-1}\mathbf{D}_{Z_{kn}}^{\mathrm{T}}\mathbf{K}_{Z_{kn}} \end{bmatrix},$$

$$\mathbf{D}_{Z_k}^{k} = \begin{bmatrix} 1 & (Z_{k1} - Z_k)/h_k \\ 1 & (Z_{k2} - Z_k)/h_k \\ \vdots & \vdots \\ 1 & (Z_{kn} - Z_k)/h_k \end{bmatrix}.$$

$Z_k = [Z_{k1}, Z_{k2}, \cdots, Z_{kn}]^{\mathrm{T}}$, $\mathbf{e}_1 = (1, 0)^{\mathrm{T}}$ and $\mathbf{K}_{Z_k} = \mathrm{diag}\{K_{h_k}(Z_{k1} - Z_k), K_{h_k}(Z_{k2} - Z_k), \cdots, K_{h_k}(Z_{kn} - Z_k)\}$, where $K_{h_k}(\cdot) = K(\cdot/h_k)/h_k$, $K(\cdot)$ is a kernel function and h_k is a bandwidth, $k = 1, 2$.

By Opsomer and Ruppert (1997), the backfitting estimators of $\mathbf{m}_1$ and $\mathbf{m}_2$ can be obtained as

$$\hat{\mathbf{m}}_1 = \mathbf{W}_1(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \quad \hat{\mathbf{m}}_2 = \mathbf{W}_2(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{3.5}$$

with

$$\mathbf{W}_1 = \mathbf{I}_n - (\mathbf{I}_n - \mathbf{S}_1^*\mathbf{S}_2^*)^{-1}(\mathbf{I}_n - \mathbf{S}_1^*), \quad \mathbf{W}_2 = \mathbf{I}_n - (\mathbf{I}_n - \mathbf{S}_2^*\mathbf{S}_1^*)^{-1}(\mathbf{I}_n - \mathbf{S}_2^*).$$

Substituting $\hat{\mathbf{m}}_1$ and $\hat{\mathbf{m}}_2$ into model (3.4), we obtain a synthetic linear regression model

$$Y_i - \hat{Y}_i = (\mathbf{X}_i - \hat{\mathbf{X}}_i)^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_i - \hat{\varepsilon}_i, i = 1, 2, \cdots, n, \tag{3.6}$$

where $\hat{\mathbf{Y}} = (\hat{Y}_1, \cdots, \hat{Y}_n)^{\mathrm{T}} = \mathbf{SY}$, $\hat{\mathbf{X}} = (\hat{\mathbf{X}}_1, \cdots, \hat{\mathbf{X}}_n)^{\mathrm{T}} = \mathbf{SX}$,

$\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_n)^{\mathrm{T}} = \mathbf{S}\boldsymbol{\varepsilon}$, and $\mathbf{S} = \mathbf{W}_1 + \mathbf{W}_2$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)$.

Based on the linear model (3.6), when $\mathbf{X}_i$ can be observed, we get the profile least squares estimator for $\boldsymbol{\beta}$. However, in our case, $\mathbf{X}_i$ cannot be observed. By correction for attenuation technique, Liang *et al.* (2008) defined the corrected profile least-squares estimator of $\boldsymbol{\beta}$ as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in R^l} \left[(\bar{\mathbf{Y}} - \bar{\mathbf{V}}\boldsymbol{\beta})^{\mathrm{T}} (\bar{\mathbf{Y}} - \bar{\mathbf{V}}\boldsymbol{\beta}) - n\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{\eta}} \boldsymbol{\beta} \right] = (\bar{\mathbf{V}}^{\mathrm{T}} \bar{\mathbf{V}} - n\boldsymbol{\Sigma}_{\boldsymbol{\eta}})^{-1} \bar{\mathbf{V}}^{\mathrm{T}} \bar{\mathbf{Y}}, \quad (3.7)$$

where $\bar{\mathbf{Y}} = \mathbf{Y} - \hat{\mathbf{Y}}, \bar{\mathbf{V}} = \mathbf{V} - \hat{\mathbf{V}}$, $\hat{\mathbf{V}} = (\hat{\mathbf{V}}_1, \cdots, \hat{\mathbf{V}}_n)^{\mathrm{T}} = \mathbf{SV}$ and $\mathbf{V} = (\mathbf{V}_1, \cdots, \mathbf{V}_n)^{\mathrm{T}}$. Moreover, the final estimator of $\mathbf{m}_1$ and $\mathbf{m}_2$ can be defined as

$$\hat{\mathbf{m}}_1 = \mathbf{W}_1(\mathbf{Y} - \mathbf{V}\hat{\boldsymbol{\beta}}), \quad \hat{\mathbf{m}}_2 = \mathbf{W}_2(\mathbf{Y} - \mathbf{V}\hat{\boldsymbol{\beta}}). \quad (3.8)$$

3.2.2 The Empirical Log-likelihood Ratio Statistic

Using $\hat{e}_i = y_i - \mathbf{V}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}} - \hat{m}_1(Z_{1i}) - \hat{m}_2(Z_{2i})$ to replace e_i in (3.3), the estimated empirical log-likelihood ratio is then defined by

$$l_N = -2 \max \left\{ \sum_{i=1}^{N} \log(Np_i) : \sum_{i=1}^{N} p_i \boldsymbol{\xi}_i = 0, p_i \geq 0, \sum_{i=1}^{N} p_i = 1 \right\}, \quad (3.9)$$

where $\boldsymbol{\xi}_i = (\xi_{i1}, \cdots, \xi_{ip})^{\mathrm{T}}, \quad \xi_{ik} = \hat{e}_i \hat{e}_{i+k}, \quad k = 1, 2, \cdots, p, \quad i = 1, 2, \cdots, N.$

By the Lagrange multiplier technique, the empirical log-likelihood ratio can be

represented as

$$\hat{l}_N = 2\sum_{i=1}^{N} \log(1 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\xi}_i), \tag{3.10}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_p)^T$ is the solution of the equation

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\boldsymbol{\xi}_i}{1 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\xi}_i} = 0. \tag{3.11}$$

The following theorem indicates that l_N is asymptotically distributed as a χ^2-distribution.

Theorem 3.1 Under the Assumptions 1 to 4 given in Section 3.4, and the null hypothesis of no serial correlation, we have

$$l_N \xrightarrow{D} \chi_p^2,$$

where χ_p^2 is a χ^2-distribution with p degrees of freedom.

3.3 Simulation Studies

In this section, we conducted extensive simulations to illustrate the finite sample properties of the proposed test procedure. In our simulations, the data are generated from the following models

$$y_i = x_i\beta + m_1(z_{1i}) + m_2(z_{2i}) + \varepsilon_i, \quad v_i = x_i + \eta_i, i = 1, 2, \cdots, n,$$

where $\beta = 1, m_1(z_{1i}) = 2\cos(2\pi z_{1i})$, $m_2(z_{2i}) = z_{2i}^4 + 2z_{2i}^3 + 3z_{2i}^2 - 2z_{2i} - 2$, and $x_i \sim N(0,1)$, $z_{1i} \sim U(0,1)$, $z_{2i} \sim U(-1,1)$, $\eta_i \sim N(0, 0.25)$. ε_i are generated from different processes

1:AR(1) error: $\varepsilon_i = \rho\varepsilon_{i-1} + \mu_i$;

2:MA(1) error: $\varepsilon_i = \rho\mu_{i-1} + \mu_i$;

3:AR(2) error: $\varepsilon_i = \alpha_1\varepsilon_{i-1} + \alpha_2\varepsilon_{i-2} + \mu_i$;

4:MA(2) error: $\varepsilon_i = \alpha_1\mu_{i-1} + \alpha_2\mu_{i-2} + \mu_i$.

To evaluate the effect of the error distributions on our results, μ_i is supposed to follow the following three different distributions, (1)$\mu_i \sim N(0, 0.5^2)$, (2)$\mu_i \sim U(-\sqrt{3}/2, \sqrt{3}/2)$, (3) $\mu_i \sim \frac{1}{8}\chi_8^2 - 1$. The Epanechnikov kernel $K(x) = 0.75(1 - x^2)\mathbf{I}_{|x|\leq 1}$ and bandwidths $h_1 = h_2 = n^{-1/5}$ are used in the simulations.

Take $\rho = 0, \pm 0.2, \pm 0.5, \pm 0.8$ and $(\alpha_1, \alpha_2) = (0,0), (0.3, 0.4), (0.2, -0.6), (-0.4, 0.5)$, $(0, 0.5)$ and $n = 100, 200$. For each case, 1000 replications were run and the rejection rate at the significance level $\alpha = 0.05$ was computed as the estimated size and power of our proposed test procedure. The results are presented in Tables 3.1 to 3.4. For better visualization, we also give plots and explanations for Tables 3.1 to 3.4 at the end of this chapter.

We summarize our findings as follows. When the null hypothesis is true (that is $\rho = 0$ or $(\alpha_1, \alpha_2) = (0,0)$), the rejection frequencies (estimated sizes) of our proposed test are close to their nominal levels 0.05 under different error distributions and error process. Under the alternative hypothesis, the rejection rate seems very robust to the variation of the type of error distribution, and increases rapidly as the alternative hypothesis deviates from the null hypothesis.

3.4 Proof of the Theorem

We begin with the following assumptions required to derive the main results. These assumptions are quite mild and can be easily satisfied.

Assumption 1. The function $K(\cdot)$ is a bounded symmetric density function with compact support.

Assumption 2. The densities $f_k(Z_k)$ of Z_k are Lipschitz continuous and bounded away from 0, and have bounded supports $\mathbf{\Omega}_k$ for $k = 1, 2$.

Assumption 3. The second derivatives of $m_k(\cdot), k = 1, 2$ exist and are bounded and continuous.

Assumption 4. As $n \to \infty$, $h_k \to 0$, $nh_k/\log n \to \infty$ and $nh_k^8 \to 0$ for $k = 1, 2$.

In order to prove that main results, we first introduce several lemmas.

Lemma 3.1 For the backfitting estimation of unknown functions, we have

$$\hat{m}_k(z) - m_k(z) = O_p\left(h_k^2 + \frac{1}{\sqrt{nh_k}}\right), k = 1, 2.$$

This lemma can be obtained by Theorem 1 of Wand (1999).

Lemma 3.2 Let $G_i, i = 1, 2, \cdots, n$ be i.i.d. random variables with $E(G_i) = 0$ and $E(G_i^2) < \infty$. Then for any permutation $(j_1, j_2, \cdots, j_n)$ of $(1, 2, \cdots, n)$,

$$\max_{1 \le k \le n} \mid \sum_{i=1}^{k} G_{j_i} \mid = O_p(n^{\frac{1}{2}} \log n).$$

The lemma comes from Gao (1995).

Lemma 3.3 Under the Assumptions 1 to 4, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\boldsymbol{\xi}_i \xrightarrow{D} N(\mathbf{0},\sigma_0^2), \sigma_0=\sigma^2+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_\eta\boldsymbol{\beta}.$$

Proof: Denote $\varphi_i = \mathbf{V}_i^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) + m_i - \hat{m}_i$ with $m_i = m_1(Z_{1i}) + m_2(Z_{2i}), \hat{m}_i = \hat{m}_1(Z_{1i}) + \hat{m}_2(Z_{2i})$. Then, by the definition of $\boldsymbol{\xi}_i$, we have

$$\begin{aligned}
\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_{ik} &= \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_i+\varphi_i)(e_{i+k}\varphi_{i+k}) \\
&= \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}e_ie_{i+k} + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_i\varphi_{i+k} + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{i+k}\varphi_i + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}e_i\varphi_{i+k} \\
&\doteq \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}U_{ik} + I_1 + I_2 + I_3.
\end{aligned}$$

For I_1, we have

$$\begin{aligned}
I_1 &= \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\mathbf{V}_i^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) + m_i - \hat{m}_i\right]\left[\mathbf{V}_{i+k}^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) + m_{i+k} - \hat{m}_{i+k}\right] \\
&= \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\mathbf{V}_i^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\mathbf{V}_{i+k}^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right] + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[(m_i-\hat{m}_i)(m_{i+k}-\hat{m}_{i+k})\right] \\
&\quad + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\mathbf{V}_i^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(m_{i+k}-\hat{m}_{i+k})\right] + \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[(m_i+\hat{m}_i)\mathbf{V}_{i+k}^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right] \\
&= I_{11} + I_{12} + I_{13} + I_{14}.
\end{aligned}$$

Obviously, $I_{11} = \sqrt{N}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^{\mathrm{T}}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{V}_i\mathbf{V}_{i+k}^T\right](\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) = O_p(\frac{1}{\sqrt{N}}) = o_p(1)$. By Lemmas 3.1 and 3.2, we prove

$$\begin{aligned}
|I_{12}| &= \left|\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}(m_i-\hat{m}_i)(m_{i+k}-\hat{m}_{i+k})\right| \\
&\leqslant \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\sup_i|m_i-\hat{m}_i|\sup_i|m_{i+k}-\hat{m}_{i+k}| \qquad (3.12)\\
&= \sqrt{N}O_p(c_n^2) = o_p(1),
\end{aligned}$$

with $c_n = h_1^2 + h_2^2 + \left\{\frac{\log(1/h_1)}{nh_1}\right\}^{1/2} + \left\{\frac{\log(1/h_2)}{nh_2}\right\}^{1/2}$. For I_{13}, we have

$$\begin{aligned} |I_{13}| &\leqslant \tfrac{1}{\sqrt{N}} \left\{\sum_{i=1}^{N} [\mathbf{X}_i^T(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]^2\right\}^{\frac{1}{2}} \left\{\sum_{i=1}^{N} (m_{i+k} - \hat{m}_{i+k})^2\right\}^{\frac{1}{2}} \\ &\leqslant \sqrt{N}\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\| \left(\tfrac{1}{N}\sum_{i=1}^{N} \|\mathbf{X}_i\|^2\right)^{\frac{1}{2}} \sup_i |m_i - \hat{m}_i| \\ &= o_p(1). \end{aligned} \tag{3.13}$$

Similarly to I_{13}, we have $I_{14} = o_p(1)$. By the above results, we have $I_1 = o_p(1)$.

For I_2, we have

$$I_2 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{V}_i^T(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})e_{i+k} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (m_i - \hat{m}_i)e_{i+k} = I_{21} + I_{22}. \tag{3.14}$$

We have

$$I_{21} = \sqrt{N}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^{\mathrm{T}} \frac{1}{N} \sum_{i=1}^{N} \mathbf{V}_i e_{i+k} = o_p(1), \tag{3.15}$$

and

$$I_{22} \leq \frac{1}{\sqrt{N}} \sup_i |m_i - \hat{m}_i| \max_{1\leq k\leq n} |\sum_{i=1}^{k} v_{j_i}| = o_p(1). \tag{3.16}$$

By (3.15) and (3.16), we have $I_2 = o_p(1)$. Similarly, we prove $I_3 = o_p(1)$. Then, we get

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{ik} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U_{ik} + o_p(1),$$

and

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\xi}_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{U}_i + o_p(1).$$

By the same argument used in the proof of Theorem 3.1 in Hu *et al.* (2009), we get

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{\xi}_i \xrightarrow{D} N(\mathbf{0}, \sigma_0^2).$$

Lemma 3.4 Under the Assumptions 1 to 4, we have

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}} \xrightarrow{p} \sigma^4 \mathrm{I}_p.$$

By the same argument as in the proof of Lemma 3.3, we prove Lemma 3.4 by the law of large numbers. We omit the details here.

Proof of Theorem 3.1 Using the same strategy as in the proof of Theorem 3.2 in Owen (1991), we prove that

$$\boldsymbol{\lambda} = \left(\sum_{i=1}^{N}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}}\right)^{-1}\sum_{i=1}^{N}\boldsymbol{\xi}_i + o_p(N^{-1/2}), \tag{3.17}$$

and

$$\sum_{i=1}^{N}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\xi}_i = \sum_{i=1}^{N}\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\xi}_i\right)^2 + o_p(1). \tag{3.18}$$

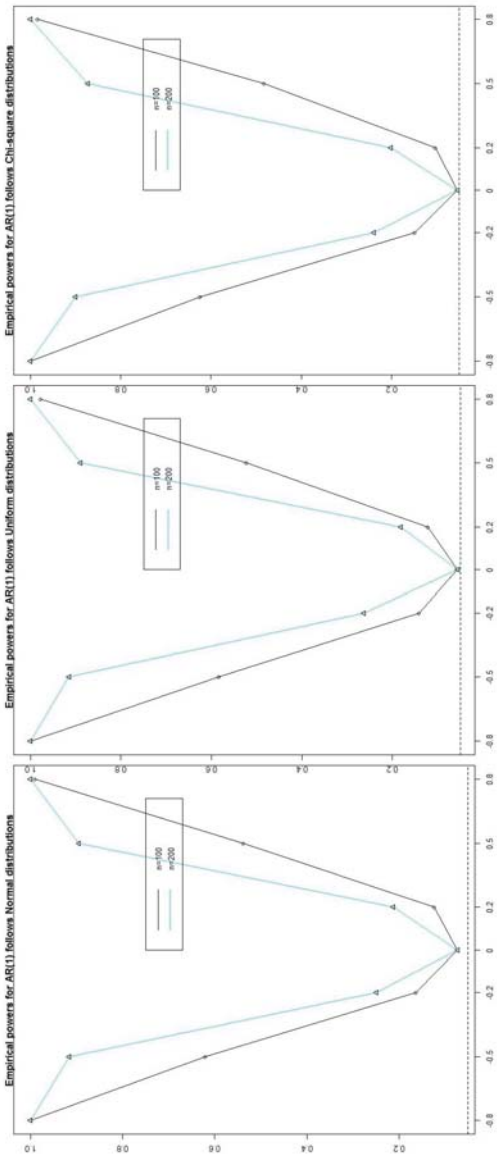
Using Taylor's expansion, and (3.17), (3.18), we obtain

$$\begin{aligned} l_N &= \sum_{i=1}^{N}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{\lambda} + o_p(1) \\ &= \left(\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\xi}_i\right)^{\mathrm{T}}\left(\tfrac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}}\right)^{-1}\left(\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\xi}_i\right) + o_p(1). \end{aligned}$$

Finally, combining Lemmas 3.3 and 3.4, we have $l_N \xrightarrow{D} \chi_p^2$ as $N \to \infty$. The theorem is then proved.

Table 3.1. Size and power for $H_0 : \rho = 0$ when $\varepsilon_i = \rho\varepsilon_{i-1} + \mu_i$

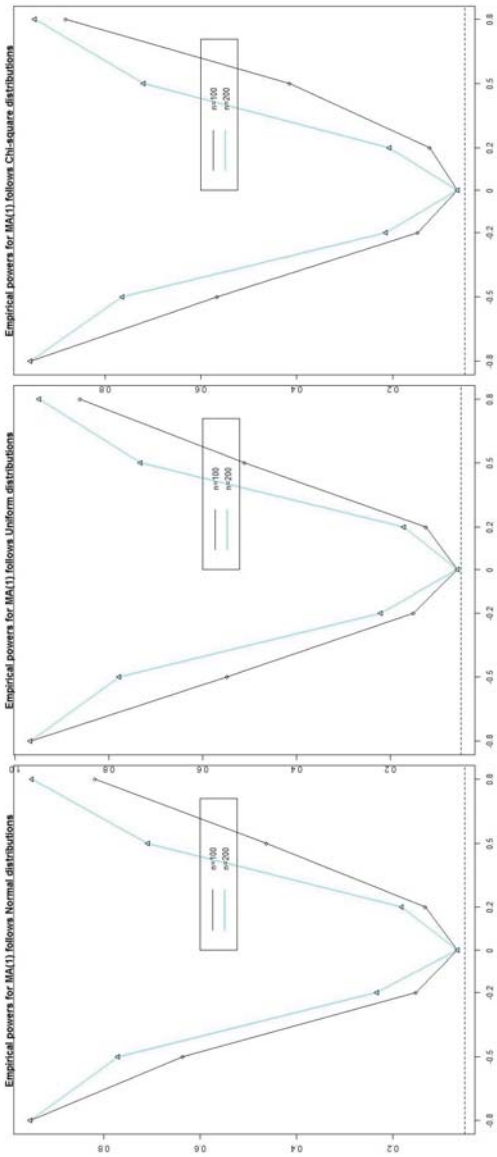
		Error		Distribution
n	ρ	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi^2_8 - 1$
n=50	0	0.055	0.060	0.060
	0.2	0.105	0.123	0.107
	-0.2	0.145	0.142	0.151
	0.5	0.517	0.512	0.474
	-0.5	0.599	0.571	0.609
	0.8	0.966	0.953	0.957
	-0.8	0.976	0.974	0.972
n=100	0	0.073	0.056	0.054
	0.2	0.212	0.181	0.202
	-0.2	0.249	0.263	0.239
	0.5	0.895	0.889	0.873
	-0.5	0.916	0.915	0.900
	0.8	1.000	1.000	1.000
	-0.8	1.000	1.000	1.000



From the figure PLAM(EV)-AR(1), we obtain that when $n = 100$ and $\rho = 0$, the rejection rate of our proposed test is close to their significance levels 0.05 under three different error distributions. When $\rho \neq 0$, the rejection rate seems very robust to the variation of the type of error distribution, and increases rapidly as the ρ's value becomes larger. When $n = 200$, we get the same conclusion and this phenomenon becomes more and more obvious.

Table 3.2. Size and power for $H_0 : \rho = 0$ when $\varepsilon_i = \rho u_{i-1} + \mu_i$.

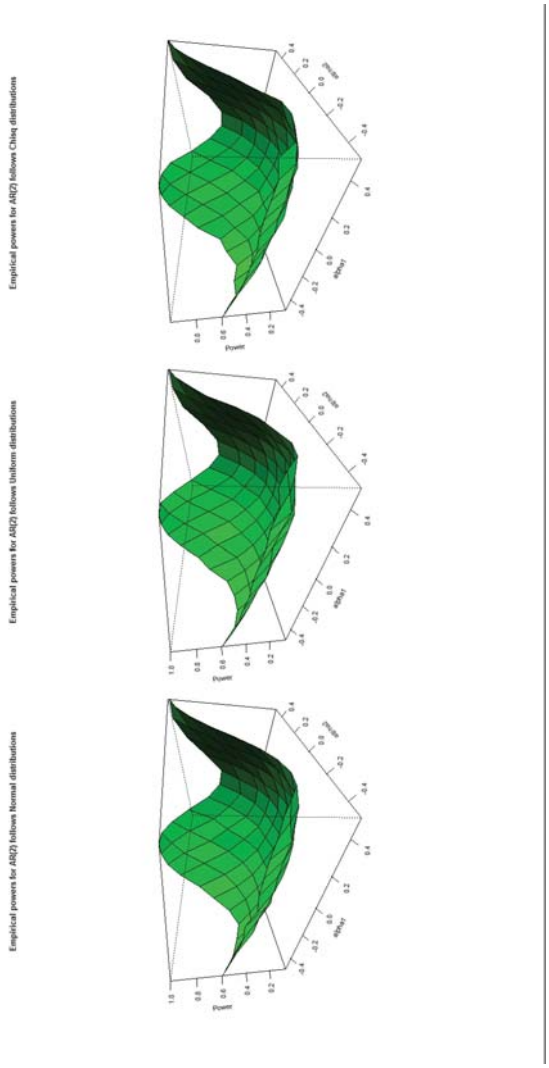
		Error		Distribution
n	ρ	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi_8^2 - 1$
n=100	0	0.062	0.071	0.063
	0.2	0.106	0.117	0.100
	-0.2	0.119	0.135	0.116
	0.5	0.323	0.382	0.288
	-0.5	0.437	0.408	0.386
	0.8	0.557	0.623	0.589
	-0.8	0.645	0.695	0.636
n=200	0	0.065	0.057	0.065
	0.2	0.181	0.171	0.207
	-0.2	0.233	0.220	0.214
	0.5	0.709	0.732	0.719
	-0.5	0.771	0.778	0.764
	0.8	0.950	0.948	0.946
	-0.8	0.953	0.967	0.955



From the figure PLAM(EV)-MA(1), we can obtain that when $n = 100$ and $\rho = 0$, the rejection rate of our proposed test are close to their significance levels 0.05 under three different error distributions. When $\rho \neq 0$, the rejection rate seems very robust to the variation of the type of error distribution, and increase rapidly as the ρ's value becomes larger. When $n = 200$, we can get the same conclusion and this phenomenon becomes more and more obvious.

Table 3.3. Size and power for $H_0 : \alpha_1 = \alpha_2 = 0$ when $\varepsilon_i = \alpha_1\varepsilon_{i-1} + \alpha_2\varepsilon_{i-2} + \mu_i$.

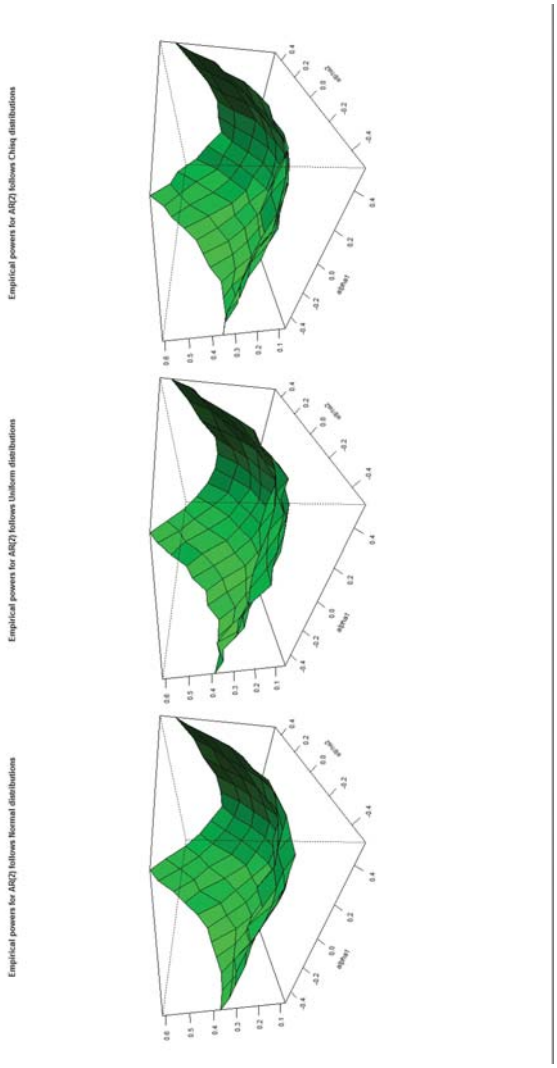
		Error		Distribution
n	(α_1, α_2)	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi^2_8 - 1$
n=100	(0,0)	0.083	0.072	0.063
	(0,0.5)	0.442	0.454	0.421
	(0.3,0.4)	0.620	0.667	0.637
	(0.2,-0.6)	0.689	0.700	0.706
	(-0.4,0.5)	0.985	0.983	0.984
n=200	(0,0)	0.065	0.054	0.050
	(0,0.5)	0.827	0.808	0.801
	(0.3,0.4)	0.970	0.962	0.962
	(0.2,-0.6)	0.968	0.963	0.955
	(-0.4,0.5)	1.000	1.000	1.000



From the figure PLAM-AR(2), we can obtain that when $n = 100$ and $(\alpha_1, \alpha_2) = (0, 0)$, the rejection rate of our proposed test are close to their significance levels 0.05 under three different error distributions. When $\rho \neq 0$, the rejection rate seems very robust to the variation of the type of error distribution, and increase rapidly as the ρ's value becomes larger. When $n = 200$, we can get the same conclusion and this phenomenon becomes more and more obvious.

Table 3.4. Size and power for $H_0 : \alpha_1 = \alpha_2 = 0$ when $\varepsilon_i = \alpha_1 u_{i-1} + \alpha_2 u_{i-2} + \mu_i$.

		Error		Distribution
n	(α_1, α_2)	$N(0, 0.5^2)$	$U(-\sqrt{3}/2, \sqrt{3}/2)$	$\frac{1}{8}\chi^2_8 - 1$
n=100	(0,0)	0.070	0.055	0.077
	(0,0.5)	0.268	0.293	0.294
	(0.3,0.4)	0.317	0.303	0.306
	(0.2,-0.6)	0.448	0.418	0.426
	(-0.4,0.5)	0.518	0.514	0.500
n=200	(0,0)	0.069	0.055	0.061
	(0,0.5)	0.647	0.624	0.614
	(0.3,0.4)	0.683	0.663	0.665
	(0.2,-0.6)	0.774	0.822	0.789
	(-0.4,0.5)	0.871	0.889	0.869



From the figure PLAM-MA(2), we can obtain that when $n = 100$ and $(\alpha_1, \alpha_2) = (0, 0))$, the rejection rate of our proposed test are close to their significance levels 0.05 under three different error distributions. When $\rho \neq 0$, the rejection rate seems very robust to the variation of the type of error distribution, and increase rapidly as the (α_1, α_2)'s value becomes larger. When $n = 200$, we can get the same conclusion and this phenomenon becomes more and more obvious.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

In this thesis, we proposed inference procedures for testing the serial correlation in the partially linear additive models based on the empirical likelihood methods. For the partially linear additive models without errors in variables, the empirical likelihood method based on the profile least-square estimation was developed. The method can be used not only for testing zero first-order serial correlation, but also for testing higher-order serial correlation. Under mild conditions, we showed that the estimated empirical log-likelihood ratio is an asymptotical χ^2-distribution under the null hypothesis of no serial correlation. Then the rejection region can be constructed easily. Our simulation results illustrate the performance of the proposed procedure. Furthermore, we considered the same problem for the partially linear additive model with errors in variable. Because of the existence of the errors, the profile-least-square-based empirical likelihood is no longer valid. The empirical likelihood based

on a corrected profile least-squares method was investigated. We showed in Chapter 3 that the estimated empirical log-likelihood ratio is asymptotical χ^2-distribution under the null hypothesis of no serial correlation. Then the rejection region can be constructed easily. Our simulation results illustrate the performance of the proposed procedure.

4.2 Future Work

In this work, we focus on the partially linear additive model for ordinary data. There remain many topics for future work. Firstly, we can extend the procedure established here to the generalized partially linear additive model which includes the partially linear additive model as a special case. Secondly, this thesis only deals with the ordinary data, but in practice, people often encounter data with incomplete observations, such as missing data, censoring data and so on. How to extend the methods established here to such types of data is worthy of further investigation.

References

[1] Bekker, P. A. (1986). Comment on identification in the Linear Errors in Variables Model. *Econometrica*, 54:215-217.

[2] Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual correlations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65:1509-1526.

[3] Breusch, T. S. and Pagan, A. R. (1980). The Lagrange multiplier test and its application to model specification in econometrics. *Review of Economic Studies*, 47:239-254.

[4] Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, 17:453-510.

[5] Carroll, R. J., Fan, J. Q., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92:477-489.

[6] Chambers, J. M. and Hastie, T. J. (1992). Statistical models in S. *London: Chapman and Hall*.

[7] Chen, S. X. and Keilegom, I. V. (2009). A review on empirical likelihood methods for regression. *Test*, 18:415-447.

[8] Cui, H. J. and Li, R. C. (1998). On parameter estimation for semi-linear errors-in-variables models. *Journal of Multivariate Analysis*, 64:1-24.

[9] Durbin, J. and Watson, G. S. (1950). Testing serial correlation in least squares regression. *Biometrika*, 37:409-428.

[10] Durbin, J. (1970). Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables. *Econometrica*, 38:410-421.

[11] Fan, J. and Jiang, J. (2005). Nonparametric inferences for additive models. *Journal of the American Statistical Association*, 100:890-907.

[12] Fan, J. Q. and Li, R. Z. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96:1348-1360.

[13] Fan, J. Q. and Lv, J. C. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B.*, 70:849-911.

[14] Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association*, 76:817-823.

[15] Fuller, W. A. (1987). Measurement Error Models. *New York: John Wiley.*

[16] Gao, J. (1995). Asymptotic theory for partly linear models. *Communications in Statistics - Theory and Methods*, 24:1985-2010.

[17] Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, 46:1303-1310.

[18] Godfrey, L. G. (2007). Alternative approaches to implementing Lagrange multiplier tests for serial correlation in dynamic regression models. *Computational Statistics & Data Analysis*, 51:3282-3295.

[19] Hardle, W. and Hall, P. and Ichimura, H. (1993). *The Annals of Statistics*, 21:157-178.

[20] Hu, X., Wang, Z. and Liu, F. (2009). Testing serial correlation in semiparametric varying coefficient partially linear models. *Communications in Statistics - Theory and Methods*, 38(13):2145-2163.

[21] Jiang, J., Zhou, H., Jiang, X. and Peng, J. (2007). Generalized likelihood ratio tests for the structures of semiparametric additive models. *The Canadian Journal of Statistics*, 35:381-398.

[22] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457-481.

[23] Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statistics Sinica*, 4:199-218.

[24] Li, Q. (2002). Efficient estimation of additive partially linear models. *International Economic Review*, 41:1073-1092.

[25] Li, D. and Stengos, T. (2003). Testing serial correlation in semi-parametric time series models. *Journal of Time Series Analysis*, 24:311-335.

[26] Li, Q. and Hsiao, C. (1998). Testing serial correlation in semiparametric panel data models.*Journal of Econometrics*, 87:207-237.

[27] Liang, H., Hardle, W. and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *Annals of Statistics*, 27:1519-1935.

[28] Liang, H. and Qin, Y. S. (2008). Empirical likelihood based inferences for partially linear models with missing covariates. *Australian and New Zealand Journal of Statistics*, 50:347-359.

[29] Liang, H., Su, H., Thurston, S., Meeker, J. D. and Hauser, R. (2009). Empirical likelihood based inference for additive partial linear measurement error models. *Statistics and Its Interface*, 2:83-90.

[30] Liang, H., Thurston, H., Ruppert, D. and Apanasovich, T. (2008). Additive partial linear models with measurement errors. *Biometrika*, 95:667-678.

[31] Liang, H., Wang, S. and Carroll, R. J. (2007). Partially linear models with missing response variables and error-prone covariates. *Biometrika*, 94:185-198.

[32] Liu, F., Chen, G. and Chen, M. (2008). Testing serial correlation in partial linear errors-in-variables models based on empirical likelihood. *Communications in Statistics-Theory and Methods*, 37(12):1905-1918.

[33] Liu, X., Wang, L. and Liang, H. (2011). Estimation and variable selection for semiparametric additive partial linear model. *Statistica Sinica*, 21(3):1225-1248.

[34] Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65:297-303.

[35] Manzan, S. and Zerom, D. (2005). Kernel estimation of a partially linear additive model. *Statistics and Probability Letters*, 72:313-322.

[36] Nelder, J. A. and Baker, R. J. (1972). Generalized linear models. *Wiely.*

[37] Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, 25:186-211.

[38] Opsomer, J. D. and Ruppert, D. (1999). A root-n consistent backfitting estimator for semiparametric additive modelling. *Journal of Computational and Graphical Statistics*, 8:715-732.

[39] Owen, A. (1988). Empirical likelihood ratio confidence intervals for single functional. *Biometrika*, 75:237-249.

[40] Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18:90-120.

[41] Owen, A. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, 19:1725-1747.

[42] Owen, A. (2001). Empirical likelihood. *New York: Chapman and Hall.*

[43] Rao, C. R. (1973). Linear statistical inference and its applications. *Wiely.*

[44] Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B.*, 50:413-436.

[45] Tang, C. Y. and Leng, C. L. (2010). Penalized high-dimensional empirical likelihood. *Biometrika*, 97:905-920.

[46] Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, 70:865-871.

[47] Vardi, Y. (1985). Empirical distribution in selection bias models. *The Annals of Statistics*, 13:178-203.

[48] Wand, M. P. (1999). A central limit theorem for local polynomial backfitting estimators. *Journal of Multivariate Analysis*, 70:57-65.

[49] Wang, Q. (1999). Estimation of partial linear error-in-variables models with validation data. *Journal of Multivariate Analysis*, 69:30-64.

[50] Wang, X., Chen, F., Lin, L. (2010). Empirical likelihood inference for the parameter in additive partially linear EV models. *Communications in Statistics-Theory and Methods*, 39(19):3513-3524.

[51] Wei, C. and Liu, C. (2012). Statistical inference on semiparametric partially linear additive models. *Journal of Nonparametric Statistics*, 24:809-823.

[52] Wei, C., Luo, Y. and Wu, X. (2012). Empirical likelihood for partially linear additive errors-in-variables models. *Statistical Papers*, 53(2):485-496.

[53] Wei, C. and Wang, Q. (2012). Statistical inference on restricted partially linear additive errors-in-variables models. *Test*, 21:757-774.

[54] Wei, C. and Yang, J. (2014). Testing serial correlation in partially linear additive models. *Acta Mathematicae Applicatae Sinica*, To appear.

[55] Yang, J., Wei, C. and Guo, S. (2014). Testing serial correlation in partially linear additive errors-in-variables models. *Communication in Statistics-Simulation and Computation*, To appear.

[56] Zhao, P.X. and Xue, L.G. (2013). Empirical likelihood for nonparametric components in additive partially linear models *Communications in Statistics - Simulation and Computation*, 42(9):1935-1947.

[57] Zhou, Z., Qian, W. and He, C. (2010). Testing serial correlation for partially nonlinear models. *Journal of the Korean Statistical Society*, 39(4):501-509.

[58] Zou, H. (2006). The adaptive Lasso and its oracle properties *Journal of the American Statistical Association*, 101(476):1418-1429.