INVESTIGATION OF SIMPLE LINEAR MEASUREMENT ERROR MODELS (SLMEMS) WITH CORRELATED DATA

A Dissertation

by

MING LU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Fred Dahm
Co-Chair of Committee,	Hongwei Zhao
Committee Members,	Michael Longnecker
	Paulo Lima-Filho
Head of Department,	Valen Johnson

December 2014

Major Subject: Statistics

Copyright 2014 Ming Lu

ABSTRACT

The primary goal of this research is to develop statistical methods to determine if observed real responses are adequately modeled by (possibly stochastic) simulation models that incorporate first-order autoregressive measurement errors. We assume the measurement errors are normally distributed to allow development of likelihood-based methods of inference. Simulated true responses are modeled as a simple linear regression on the true response values. That is, we wish to detect if either additive or multiplicative biases exist in the simulation model. Efficient score and likelihood ratio tests using observed real process data are developed to test the joint null hypothesis that no significant additive or multiplicative biases exist in the stochastic simulation model. Tests for adequacy of both stochastic and deterministic simulation models are developed using, respectively, structural and functional simple linear measurement error models that allow the measurement errors to satisfy normal first-order autoregressive processes.

A byproduct of this research is developments of analogous tests of the null hypothesis that errors of measurement are independent. Such tests would be of use if the real process is not a times series and there was uncertainty whether the simulation model should allow for correlated measurement errors.

Analytic and simulation results show that all maximum likelihood estimators (MLEs) of model parameters MLEs are consistent under the structural model, but some MLEs of parameters are inconsistent under functional model. Test statistics developed under the structural model are shown to be asymptotically distributed as chi-squared random variables with two degrees of freedom when testing for additive and multiplicative biases in the simulation model having correlated measurement errors. Test statistics developed under the structural model are shown to be asymptotically distributed as chi-squared random variables with one degree of freedom when testing for independence of the measurement errors. However, for functional models, the corresponding test statistics are asymptotically distributed as random variables that are two times the chi-squared distributions. Empirical power curves are plotted under different parameter configurations. Behaviors of test statistics and power curves are found to be affected by the sample size, signal to noise ratio and strength of correlations among measurement errors.

ACKNOWLEDGEMENTS

A special debt of gratitude is owed to Dr. P. Fred Dahm for providing guidance and support throughout the course of this research.

For invaluable supports throughout my research, Dr. Hongwei Zhao is gratefully acknowledged.

Thanks also go to my committee members, Dr. Michael Longnecker and Dr. Paulo Lima-Filho.

This work was coordinated under the auspices of and supported financially by the units within the Texas A&M University System including the Department of Statistics, College of Science, and the Texas A&M Supercomputing Facility.

Finally, thanks to my daughter Sophia, my husband Xiao Long, and my parents for their encouragement, patience, and love.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Introduction and Motivation

1.1.1 Introduction

Regression models are widely used to investigate if, and how, potential predictor (or independent) variables might be related to a response (or dependent) variable of interest. Standard assumptions for regression models are a) response errors are independently and identically distributed as $N(0, \sigma^2)$ random variables, b) the regression model is correctly specified and c) predictor variables have been measured or observed without error. However, in practice independent variables may be measured with errors. If predictor variables are measured with errors, it is known that the least squares estimators of some regression model parameters are inconsistent and inferences based on these estimators are invalid (Neyman and Scott, 1948).

Regression models with measurement errors in independent variables are referred to as measurement error models. This dissertation will develop joint hypothesis test statistics under simple linear measurement error models when errors of measurement in both the response and the predictor variables follow the same first-order autoregressive scheme. Motivation for consideration of this model and the joint hypotheses of interest are given in section 1.1.2.

Recall that the classic simple linear model with only one explanatory variable is,

$$Y_t = \beta_0 + \beta_1 x_t + u_t \qquad t = 1, 2, \dots, n \qquad (1.1.1)$$

where x_t is the predictor variable, Y_t is the response variable, and u_t is an independent identically normally distributed random variable with mean 0 and variance σ_u^2. The maximum likelihood estimators for β_0 and β_1 are also the least squares estimators of these parameters. These estimators are consistent and, in fact, minimum variance unbiased estimators under the standard assumptions. Furthermore, the (uncorrected) ANOVA F test statistic for testing the hypothesis $H_0: (\beta_0, \beta_1) = (b_0, b_1)$ has an F distribution with 2 and $n-2$ numerator and denominator degrees of freedom, respectively.

But it is common that not all the standard assumptions are satisfied, and in particular, that x_t is measured with error. For example, people are interested in predicting heart rate using body mass index (BMI). The BMI is determined by height and weight, both of which are likely measured with error. Instead of the true value of x_t, we actually observe,

$$X_t = x_t + e_t \qquad t = 1, 2, \dots, n \qquad (1.1.2)$$

where e_t is an error of measurement. In such a case, x_t is often called a latent (or error-prone) predictor variable, and the model (1.1.1) and (1.1.2) is known as a simple (linear) measurement, or errors-in-variables, model. Simple (and multivariate) linear as well as certain types of nonlinear, measurement error models are thoroughly covered in Fuller (1987). Thorough coverage of nonlinear measurement error models can be found in Carroll, Ruppert, Stefanski and Crainiceanuet (2006). Both the Fuller (1987) and the Carroll, et al. (2006) references contain comprehensive reviews of the substantial literature that is concerned with measurement error models.

Unlike standard regression models, the distinction between random and fixed x_t values is crucial in measurement error models. A structural model is said to occur when latent predictor variables are considered to be random variables. A functional models is said to occur when latent predictor variables are considered to be fixed, in which case the x_t's are unknown parameters. Neyman and Scott (1948) showed that under the functional model, the maximum likelihood estimators (MLEs) of some parameters are inconsistent due to the indefinitely increasing number of parameters that must be estimated.

Although measurement error models have been extensively investigated, little attention has been given to testing the joint hypothesis of intercept and slope, $H_0: \beta_0 = b_0, \beta_1 = b_1$. In the next section, we describe how evaluation of complex stochastic simulation models using experimental data can be formulated in terms of this joint hypothesis and simple linear measurement error models. Abdul-Salam (1996) apply this formulation to cases in which measurement errors are independent of each other and both sets of measurement errors are independent, identically distributed normal random variables.

1.1.2 Motivation

One of the most important methods in the engineering and natural sciences is simulation models of real systems. People are interested in evaluating the performance of the model which is designed to represent the real system. The method of using regression theory to evaluate models arose from applications in the management sciences (Cohen and Cyert, 1961; Aigner, 1972). The adequacy of a model is evaluated by regressing of real data on simulation data and failing to reject the null hypothesis of zero intercept and unit slope. Usually, problems exist when there are stochastic components in the model.

We construct a stochastic model, for example, to mimic contaminant movement through an aquifer and want to evaluate its performance at some different initial conditions and/or input values. From the knowledge of experimenters, initial conditions including aquifer flow rates and contamination concentrations and/or input values of aquifer elevation etc. may have an effect on contaminant flow rates. For every set of baseline conditions and/or input values, we have (x_t, y_t), $t = 1,2,\ldots,n$ representing the true simulated contaminate flow rates and observations from the real system, respectively. The 'true' simulated flow rate is obtained by solving the model with means of stochastic components.

The true flow rates (x_t, y_t) , $t = 1,2,\ldots,n$ are unobservable due to random components such as the amount of rainfall in the real system and also stochastic components in the simulation model. Further, it is likely that contaminate flow rates measured over time result in auto-correlated responses. Therefore, instead of true values, we observe:

$$Z_t \equiv \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} e_t \\ u_t \end{pmatrix},$$

where u_t and e_t are additive random measurement errors and from autoregressive processes.

The bias of the simulation model is assumed to be additive and/or multiplicative, that is:

$$y_t = \beta_0 + \beta_1 x_t.$$

Another assumption is that the variability of the simulation model and the real system are equal because the randomness of the simulation model can be adjusted with sufficient accuracy. That is,

$$Var(u_t) = Var(e_t) = \sigma^2.$$

Similarly, we assume that both measurement errors follow first-order autoregressive (AR(1)) processes, and that the correlation is (at least approximately) the same for both processes.

To evaluate the accuracy of the simulation model, the hypothesis test of $H_0: \beta_0 = 0$, $\beta_1 = 1$ against $H_a: \beta_0 \neq 0$ and/or $\beta_1 \neq 1$ is performed. The null hypothesis means that there is no additive or multiplicative biases exist. And the alternative hypothesis refers to an additive and/or a multiplicative bias exists in the simulation model.

Harrison (1990), Mayer, Stuart and Swain (1994), and Mitchell (1997) discuss the inappropriateness of using the F test in an (ordinary) least squares analysis to test $H_0: \beta_0 = b_0, \beta_1 = b_1$ in the presence of measurement error. We extend the work of Abdul-Salam (1996) to cases in which measurement errors follow first-order autoregressive processes. This extension of Abdul-Salam (1996) permits a more appropriate evaluation of such complex stochastic simulation models for responses that may be temporally or spatially correlated. We derive appropriate large-sample statistical tests under this extension, and evaluate the performance of these tests in small sample settings.

We also develop methods for testing whether or not measurement errors follow identical first-order autoregressive processes. These methods could be useful in settings where it is not certain that the process being modeled has correlated responses. In such cases a modeler could test whether a correlated error structure should be incorporated into their simulation model.

1.2 Organization

The remainder of this dissertation of this dissertation is organized as follows. We define the structural and functional simple linear measurement error models that we consider, and state the hypothesis of interest in terms of parameters of these models in Section 2. Likelihood functions are derived for both models, from which we derive maximum likelihood estimators of certain model parameters, likelihood ratio test statistics and score test statistics. Section 3 reviews preliminary theoretical results to be are used in Section 4. In Section 4, we discuss the consistency of parameter estimators and derive asymptotic distributions of the derived test statistics. Simulation studies of small sample properties of estimators and test statistics are presented and discussed in sections 5. In Section 6, we summarize our conclusions and propose problems for future study.

2. DERIVATION OF LIKELIHOODS, ESTIMATORS AND TEST STATISTICS

2.1 Model Introduction

We focus on the linear relationship between the two sets of unobservable values. That is,

$$y_t = \beta_0 + \beta_1 x_t , \qquad t = 1,2,\dots,n \tag{2.1}$$

where β_0 and β_1 are unknown parameters need to estimate, and (x_t, y_t) 's are unobservable values. We assume that both y_t and x_t are unobservable due to additive random components in both the real and simulated data. That is, we observe

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} e_t \\ u_t \end{pmatrix} , \qquad t = 1,2,\dots,n \tag{2.2}$$

where (X_t, Y_t) are random variables and, e_t and u_t are unobservable additive error terms.

2.1.1 Assumptions

We develop the likelihood functions under the following assumptions. The first scenario leads to the so-called structural simple linear measurement error model and the second scenario leads to the so-called functional simple linear measurement error model (Kendall, 1951, 1952). For convenience, we simply refer to these models as structural and functional models, respectively, in what follows.

2.1.1.1 Assumptions for the structural model

In the structural model, we regard both x_t and y_t's as random variables. We assume that, $x_t, t = 1,2,\dots,n$ are distributed as i.i.d. $N(\mu_x, \sigma_x^2)$ random variables. The error terms, e_t and u_t, are mutually-independent of each other and the x_t's. Both e_t and u_t follow first order autoregressive (AR1) processes,

$$e_t = \rho e_{t-1} + v_t\,, \qquad t = 1,2,\dots,n \tag{2.3}$$

$$u_t = \rho u_{t-1} + \delta_t\,, \qquad t = 1,2,\dots,n \tag{2.4}$$

where $(v_t, \delta_t), t = \cdots, -1,0,1,2,\dots,n$ are i.i.d. bivariate normally distributed as $N_2(\mathbf{0}, \sigma^2 \boldsymbol{I})$ random vectors. The same correlation is assumed for both autoregressive processes under the null hypothesis that the simulation model is correct.

2.1.1.2 Assumptions for the functional model

In the functional model, we assume both x_t and y_t are fixed rather than random, and they must therefore be regarded as unknown parameters. Additionally we assume that $\{x_t, t = 1,2,\dots,n\}$ is a sequence of fixed numbers with a sample mean and sample variance having some constant limit, say,

$$\bar{x}_n = \frac{1}{n}\sum_{t=1}^{n} x_t \to \mu_x \tag{2.5}$$

$$s_n^2 = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \bar{x}_n)^2 \to \sigma_x^2 \tag{2.6}$$

The assumption with respect to error terms is the same as in the structural case.

2.1.2 Tests of hypotheses

We are interested in two problems. Our primary interest is testing

$$H_0\colon (\beta_0, \beta_1) = (b_0, b_1) \text{ vs. } H_a\colon (\beta_0, \beta_1) \neq (b_0, b_1) \tag{2.7}$$

under the scenario described in our motivating example in section 1.2. Of particular interest is the case in which $(b_0, b_1) = (0,1)$, i.e., that the simulation model an adequate representation of the real-system process. However, the more general formulation (2.7) also allows for possible construction of large sample joint confidence regions for (β_0, β_1) in this scenario.

In the process of investigating hypothesis (2.7), we develop tests of the hypotheses

$$H_0\colon \rho = 0 \text{ vs. } H_a\colon \rho \neq 0 \tag{2.8}$$

Such tests may be of value when simulation modelers are not sure that the real-system process produces correlated responses, but wish to test if such correlations should be incorporated into their model. For both problems, we develop likelihood ratio and score tests. Likelihood functions under both the structural and functional cases are developed in the next section.

2.2 Likelihood Functions

A bivariate normal random variable, $Z_t = (Z_1, Z_2)^T, t = 1,2,\dots,n$, with mean μ and covariance matrix Σ, has likelihood function

$$\mathcal{L} = (2\pi)^{-n}|\Sigma|^{-\frac{n}{2}} exp\left(-\frac{1}{2}\sum_{t=1}^{n}(Z_t - \mu)^T\Sigma^{-1}(Z_t - \mu)\right) \tag{2.9}$$

Therefore, the log likelihood function is,

$$\ell = log\mathcal{L} = -nlog(2\pi) - \frac{n}{2}log|\Sigma| - \frac{1}{2}\sum_{t=1}^{n}(Z_t - \mu)^T\Sigma^{-1}(Z_t - \mu) \tag{2.10}$$

This is the general form of the log likelihood function for a bivariate normal random variable. The specific forms for each case are developed below.

2.2.1 Likelihood function for the structural case

Since the observations are correlated, we first construct a conditional likelihood function. From (2.1) – (2.4), we have,

$$X_t = \rho X_{t-1} + (x_t - \rho x_{t-1}) + v_t\,, \tag{2.11}$$

$$Y_t = \rho Y_{t-1} + (1-\rho)\beta_0 + \beta_1(x_t - \rho x_{t-1}) + \delta_t\,, \qquad t = 2,3,\dots,n \tag{2.12}$$

According to the assumptions in section 2.1.1.1, x_t, u_t and e_t are independent normally distributed with mean $(\mu_x, 0, 0)^T$ and variance, a diagonal matrix $diag(\sigma_x^2, \frac{\sigma^2}{1-\rho^2}, \frac{\sigma^2}{1-\rho^2})$. Therefore, $(x_{t-1}, X_{t-1} - x_{t-1}, Y_{t-1} - y_{t-1})^T$ has a multivariate normal distribution with the above mean and variance. Furthermore, the random vector $(X_{t-1}, Y_{t-1})^T$ is normally distributed with mean and variance, respectively

$$\mu_d = (\mu_x\,, \beta_0 + \beta_1 \mu_x)^T$$

$$\Sigma_d = \begin{pmatrix} \sigma_x^2 + \frac{\sigma^2}{1-\rho^2} & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \beta_1^2 \sigma_x^2 + \frac{\sigma^2}{1-\rho^2} \end{pmatrix}.$$

After further derivation shown in Appendix A, we have

$$x_{t-1} \left| \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} \right. \sim N\left(\frac{A\mu_x + \sigma_x^2 E_{t-1}}{D}, \frac{\sigma_x^2 A}{D} \right)$$

where $A = \frac{\sigma^2}{1-\rho^2}$, $D = (1+\beta_1^2)\sigma_x^2 + A$, and $E_{t-1} = X_{t-1} + \beta_1(Y_{t-1} - \beta_0)$. Since x_t, x_{t-1}, ν_t, and δ_t are independent, and also x_t, ν_t, and δ_t are independent of $(X_{t-1}, Y_{t-1})^T$, it follows that

$$(x_t, x_{t-1}, \nu_t, \delta_t)^T | (X_{t-1}, Y_{t-1})^T \sim N\left(\left(\frac{A\mu_x + \sigma_x^2 E_{t-1}}{D}, \mu_x, 0{,}0 \right)^T, diag\left(\sigma_x^2 \frac{A}{D}, \sigma_x^2, \sigma^2, \sigma^2 \right) \right).$$

Thus, the conditional distribution for $(X_t, Y_t)^T$ given the previous observation $(X_{t-1}, Y_{t-1})^T$ is,

$$(X_t, Y_t)^T | (X_{t-1}, Y_{t-1})^T \sim N(\mu_s, \Sigma_s)$$

where $\mu_s = (\mu_{1t}, \mu_{2t})^T$, $\Sigma_s = \begin{pmatrix} \sigma_x^2 \left(1 + \rho^2 \frac{A}{D}\right) + \sigma^2 & \beta_1 \sigma_x^2 \left(1 + \rho^2 \frac{A}{D}\right) \\ \beta_1 \sigma_x^2 \left(1 + \rho^2 \frac{A}{D}\right) & \beta_1^2 \sigma_x^2 \left(1 + \rho^2 \frac{A}{D}\right) + \sigma^2 \end{pmatrix}$,

$$\mu_{1t} = \rho X_{t-1} - \rho \frac{A\mu_x + \sigma_x^2 E_{t-1}}{D} + \mu_x,$$

And

$$\mu_{2t} = \rho(Y_{t-1} - \beta_0) - \beta_1 \rho \frac{A\mu_x + \sigma_x^2 E_{t-1}}{D} + \beta_0 + \beta_1 \mu_x.$$

The determinant of the variance matrix is $|\Sigma_s| = (1+\beta_1^2)\sigma_x^2\sigma^2\left(1+\rho^2\frac{A}{D}\right)+\sigma^4$. The conditional log likelihood function then can be written as

$$\ell_s = -(n-1)log(2\pi) - \frac{(n-1)}{2} log|\Sigma_s|$$

$$-\frac{1}{2|\Sigma_s|}\sum_{t=2}^{n}\left\{\begin{array}{c} \left(\beta_1^2\sigma_x^2\left(1+\rho^2\frac{A}{D}\right)+\sigma^2\right)(X_t-\mu_{1t})^2 \\ -2\beta_1\sigma_x^2\left(1+\rho^2\frac{A}{D}\right)(X_t-\mu_{1t})(Y_t-\mu_{2t}) \\ +\left(\sigma_x^2\left(1+\rho^2\frac{A}{D}\right)+\sigma^2\right)(Y_t-\mu_{2t})^2 \end{array}\right\} \qquad (2.13)$$

2.2.2 Likelihood function for the functional case

As with the structural case, the data are transformed as,

$$X_t - \rho X_{t-1} - (x_t - \rho x_{t-1}) = \upsilon_t,$$

$$Y_t - \rho Y_{t-1} - (1-\rho)\beta_0 - \beta_1(x_t - \rho x_{t-1}) = \delta_t\,, \qquad t = 2,3,\dots,n$$

Since $(\upsilon_t, \delta_t)^T$ are i.i.d. $N_2(\mathbf{0}, \sigma^2\boldsymbol{I})$, the conditional log likelihood function is,

$$\ell_f = -(n-1)log(2\pi) - (n-1)log\sigma^2$$

$$-\frac{1}{2\sigma^2}\sum_{t=2}^{n}\{(X_t - \rho X_{t-1} - \lambda_t)^2 + (Y_t - \rho Y_{t-1} - \gamma_t)^2\} \qquad (2.14)$$

where $\lambda_t = x_t - \rho x_{t-1}$, and $\gamma_t = (1-\rho)\beta_0 + \beta_1(x_t - \rho x_{t-1})$.

2.3 Test Statistics

We now derive MLEs for parameters and develop likelihood ratio and efficient score test for each set of hypotheses under both the structural and functional cases.

2.3.1 Structural case

2.3.1.1 Efficient score test of $H_0: \rho = 0$ *versus* $H_a: \rho \neq 0$

Taking first derivatives of the structural log likelihood function with respect to the unknown parameters $(\mu_x, \sigma_x^2, \sigma^2, \beta_0, \beta_1)$ with ρ constrained to be zero, setting these derivatives equal to zero and then solving the system of equations yields the MLEs under the null hypothesis. These MLEs are

$$\tilde{\mu}_x = \bar{X}_n$$

$$\tilde{\beta}_0 = \bar{Y}_n - \tilde{\beta}_1 \bar{X}_n$$

$$\tilde{\sigma}_x^2 = \frac{\sum_{t=2}^{n}(\tilde{X}_t \tilde{Y}_t)}{(n-1)\tilde{\beta}_1}$$

$$\tilde{\sigma}^2 = \frac{\sum_{t=2}^{n}\left(\tilde{\beta}_1 \tilde{X}_t - \tilde{Y}_t\right)^2}{(n-1)\left(1+\tilde{\beta}_1^2\right)}$$

$$\tilde{\beta}_1 = \frac{-\sum_{t=2}^{n}(\tilde{X}_t^2 - \tilde{Y}_t^2) + \sqrt{\left(\sum_{t=2}^{n}(\tilde{X}_t^2 - \tilde{Y}_t^2)\right)^2 + 4\left(\sum_{t=2}^{n}\tilde{X}_t \tilde{Y}_t\right)^2}}{2\sum_{t=2}^{n}\tilde{X}_t \tilde{Y}_t}$$

where $\tilde{X}_t = X_t - \bar{X}_n$, $\tilde{Y}_t = Y_t - \bar{Y}_n$, and $\bar{X}_n = \sum_{t=2}^{n} X_t/(n-1)$.

The score vector is obtained as the first derivative with respect to each parameter under the null hypothesis, i.e.,

$$U(\tilde{\theta}) = \left(0,0,0,0,0, \left.\frac{\partial \ell_s}{\partial \rho}\right|_{\rho=0}\right)^T,$$

where

$$\left.\frac{\partial \ell_s}{\partial \rho}\right|_{\rho=0} = \frac{1}{(1+\tilde{\beta}_1^2)\tilde{\sigma}^2}\left\{\begin{array}{c} \sum_{t=2}^{n}\left[\tilde{\beta}_1 X_t - \left(Y_t - \tilde{\beta}_0\right)\right]\left[\tilde{\beta}_1 X_{t-1} - \left(Y_{t-1} - \tilde{\beta}_0\right)\right] \\ + \left(\frac{\tilde{\sigma}^2}{(1+\tilde{\beta}_1^2)\tilde{\sigma}_x^2 + \tilde{\sigma}^2}\right)^2 \sum_{t=2}^{n}\left(\tilde{X}_t + \tilde{\beta}_1 \tilde{Y}_t\right)\left(\tilde{X}_{t-1} + \tilde{\beta}_1 \tilde{Y}_{t-1}\right) \end{array}\right\}.$$

The information matrix, $I(\theta)$, is obtained by taking the expectation of the matrix of second derivatives of the log-likelihood function taken with respect to each parameter under the null hypothesis. All elements in the score vector are zero except for the last element. Therefore, when calculating the test statistic only the right bottom value in the inverse of information matrix is needed. It is shown in Appendix B that the right bottom value is $(n-1)^{-1}\left(1+\frac{\tilde{\sigma}^4}{\left((1+\tilde{\beta}_1^2)\tilde{\sigma}_x^2+\tilde{\sigma}^2\right)^2}\right)^{-1}$. Thus, in the structural case, the efficient score test statistic for testing $H_0\colon \rho = 0$ versus $H_a\colon \rho \neq 0$ is,

$$T_{ES,s,\rho} = (n-1)^{-1}\left(1+\frac{\tilde{\sigma}^4}{\left((1+\tilde{\beta}_1^2)\tilde{\sigma}_x^2+\tilde{\sigma}^2\right)^2}\right)^{-1}\left(\frac{\partial \ell_s}{\partial \rho}\Big|_{\rho=0}\right)^2 \qquad (2.15)$$

2.3.1.2 Likelihood ratio test of $H_0\colon \rho = 0$ *versus* $H_a\colon \rho \neq 0$

To perform the likelihood ratio test, we need to find the unconstrained MLEs in addition to the constrained MLEs already derived in section 2.3.1.2. Taking the first derivative of the structural log likelihood function with respect to μ_x and setting this derivative equal to zero, we get

$$\sum_{t=2}^{n}\left(E_t - \rho\frac{A}{D}E_{t-1} - (1+\beta_1^2)\left(1-\rho\frac{A}{D}\right)\mu_x\right) = 0 \qquad (2.16A)$$

where $E_t = X_t + \beta_1(Y_t - \beta_0)$, $A = \frac{\sigma^2}{1-\rho^2}$, and $D = (1+\beta_1^2)\sigma_x^2 + A$. Then

$$(1+\beta_1^2)\left(1-\rho\frac{A}{D}\right)\mu_x = \bar{E}_n - \rho\frac{A}{D}\bar{E}_{n-1},$$

and

$$\sum_{t=2}^{n}\left(E_t - \rho\frac{A}{D}E_{t-1} - (1+\beta_1^2)\left(1-\rho\frac{A}{D}\right)\mu_x\right) = \sum_{t=2}^{n}\left(\tilde{E}_t - \rho\frac{A}{D}\tilde{E}_{t-1}\right),$$

where $\bar{E}_n = \bar{X}_n + \beta_1(\bar{Y}_n - \beta_0)$, $\bar{E}_{n-1} = \bar{X}_{n-1} + \beta_1(\bar{Y}_{n-1} - \beta_0)$, $\bar{X}_n = \sum_{t=2}^{n} X_t/(n-1)$, $\bar{X}_{n-1} = \sum_{t=2}^{n} X_{t-1}/(n-1)$, $\bar{Y}_n = \sum_{t=2}^{n} Y_t/(n-1)$, $\bar{Y}_{n-1} = \sum_{t=2}^{n} Y_{t-1}/(n-1)$, and $\tilde{E}_t = E_t - \bar{E}_n$, $\tilde{E}_{t-1} = E_{t-1} - \bar{E}_{n-1}$.

Taking the first derivative of the structural log likelihood function with respect to σ_x^2, setting this derivative equal to zero and then substituting 2.16A into the resulting expression, we get

$$\left\{\begin{matrix} -(n-1)(1+\beta_1^2)\left(1+\rho^2\left(\frac{A}{D}\right)^2\right) \\ +\frac{\sigma^2}{|\Sigma|}\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\frac{A}{D}\tilde{E}_{t-1}\right)^2 \\ -2\rho\frac{A}{D^2}\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\frac{A}{D}\tilde{E}_{t-1}\right)E_{t-1} \end{matrix}\right\} = 0 \qquad (2.16B)$$

where $|\Sigma| = |\Sigma_s|$.

In similar fashion, we take the first derivatives of the structural log likelihood function with respect to σ^2, β_0, β_1 and ρ, set them equal to 0, which yields respectively

$$\left\{\begin{matrix} -(n-1)(1+\beta_1^2)\left(\frac{|\Sigma|}{\sigma^2}+\rho^2\frac{AC^2}{D^2}+\sigma^2\right)+\frac{|\Sigma|}{\sigma^4}\sum_{t=2}^{n}(F_t-\rho F_{t-1})^2 \\ +\frac{\sigma^2}{|\Sigma|}\left(\rho^2\frac{AC^2}{D^2}+\sigma^2\right)\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\frac{A}{D}\tilde{E}_{t-1}\right)^2 \\ +2\rho\frac{AC}{D^2}\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\frac{A}{D}\tilde{E}_{t-1}\right)E_{t-1} \end{matrix}\right\} = 0 \qquad (2.16C)$$

$$\sum_{t=2}^{n}(F_t-\rho F_{t-1}) = 0 \qquad (2.16D)$$

$$\left\{\begin{matrix} -(n-1)\beta_1 C\sigma^2\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)+\frac{\beta_1}{(1+\beta_1^2)}\frac{|\Sigma|}{\sigma^2}\sum_{t=2}^{n}(F_t-\rho F_{t-1})^2 \\ +\frac{\beta_1}{(1+\beta_1^2)}\left[\sigma^2+\frac{\sigma^4}{|\Sigma|}C\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\right]\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\frac{A}{D}\tilde{E}_{t-1}\right)^2 \\ -\frac{|\Sigma|}{\sigma^2}\sum_{t=2}^{n}(F_t-\rho F_{t-1})(X_t-\rho X_{t-1}) \\ -\sigma^2\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\frac{A}{D}\tilde{E}_{t-1}\right)\begin{bmatrix} (Y_t-\beta_0)-\rho\frac{A}{D}(Y_{t-1}-\beta_0) \\ +2\beta_1\rho\frac{A}{D}\frac{(A\mu_x+\sigma_x^2E_{t-1})}{D}-2\beta_1\mu_x \end{bmatrix} \end{matrix}\right\} = 0 \qquad (2.16E)$$

$$\left\{\begin{array}{c}-(n-1)(1+\beta_1^2)\rho C(C+\sigma^2)\left(\frac{A}{D}\right)^2+\frac{|\Sigma|}{\sigma^2}\sum_{t=2}^{n}(F_t-\rho F_{t-1})F_{t-1}\\ +\frac{\sigma^2}{|\Sigma|}\rho C(C+\sigma^2)\left(\frac{A}{D}\right)^2\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\frac{A}{D}\tilde{E}_{t-1}\right)^2\\ +\sigma^2\frac{A}{D}\left(1+\frac{2\rho^2}{1-\rho^2}\frac{C}{D}\right)\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\frac{A}{D}\tilde{E}_{t-1}\right)E_{t-1}\end{array}\right\}=0 \quad (2.16\text{F})$$

where $F_t=\beta_1X_t-(Y_t-\beta_0)$, and $C=(1+\beta_1^2)\sigma_x^2$.

From (2.16D),

$$\sum_{t=2}^{n}\big(\beta_1X_t-\rho\beta_1X_{t-1}-(Y_t-\beta_0)+\rho(Y_{t-1}-\beta_0)\big)=0,$$

so that

$$\beta_0=\{-(\beta_1\bar{X}_n-\bar{Y}_n)+\rho(\beta_1\bar{X}_{n-1}-\bar{Y}_{n-1})\}/(1-\rho).$$

To simplify notation, let $\alpha=\frac{A}{D}$ and $\frac{C}{D}=1-\alpha$, and then substitute these two expressions into (2.16B) to obtain

$$\left\{\begin{array}{c}-(n-1)D(1+\beta_1^2)(1+\rho^2\alpha^2)\\ +\frac{1}{1-\rho^2\alpha^2}(1+\rho^2\alpha^2)\sum_{t=2}^{n}\big(\tilde{E}_t-\rho\alpha\tilde{E}_{t-1}\big)^2\\ -2\rho\alpha\sum_{t=2}^{n}\big(\tilde{E}_t-\rho\alpha\tilde{E}_{t-1}\big)\tilde{E}_{t-1}\end{array}\right\}=0 \quad (2.16\text{B'})$$

Then, with (2.16B') and some organization we have,

$$\left\{\begin{array}{c}\frac{1-\rho^2\alpha^2}{(1-\rho^2)\alpha}\sum_{t=2}^{n}\big(\tilde{F}_t-\rho\tilde{F}_{t-1}\big)^2-\sum_{t=2}^{n}\big(\tilde{E}_t-\rho\alpha\tilde{E}_{t-1}\big)^2\\ +4\rho\alpha\frac{(1-\rho^2\alpha^2)}{1+\rho^2\alpha^2}\sum_{t=2}^{n}\big(\tilde{E}_t-\rho\alpha\tilde{E}_{t-1}\big)\tilde{E}_{t-1}\end{array}\right\}=0, \quad (2.16\text{C'})$$

$$\left\{\begin{array}{c}\beta_1(1-\rho^2\alpha^2)\sum_{t=2}^{n}\big(\tilde{F}_t-\rho\tilde{F}_{t-1}\big)^2\\ +\beta_1(1-\rho^2)\alpha\sum_{t=2}^{n}\big(\tilde{E}_t-\rho\alpha\tilde{E}_{t-1}\big)^2\\ -(1-\rho^2\alpha^2)\sum_{t=2}^{n}\big(\tilde{F}_t-\rho\tilde{F}_{t-1}\big)(X_t-\rho X_{t-1})\\ -(1+\beta_1^2)(1-\rho^2)\alpha\sum_{t=2}^{n}\big(\tilde{E}_t-\rho\alpha\tilde{E}_{t-1}\big)(Y_t-\rho\alpha Y_{t-1})\end{array}\right\}=0 \quad (2.16\text{E'})$$

and

$$\begin{Bmatrix} (1+\rho^2\alpha^2)\sum_{t=2}^{n}\left(\tilde{F}_t-\rho\tilde{F}_{t-1}\right)\tilde{F}_{t-1} \\ +\alpha^2(1+\rho^2)\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\alpha\tilde{E}_{t-1}\right)\tilde{E}_{t-1} \end{Bmatrix}=0. \tag{2.16F'}$$

There appears to be no explicit solution for system of equations (2.16C') (2.16E') and (2.16F'). In order to obtain $\hat{\rho}$, $\hat{\beta}_1$, and $\hat{\alpha}$, we have to solve these equations numerically with the constraints that ρ should be between -1 and 1 and α between 0 and 1. From equation (2.16B') we see that

$$\hat{D}=\frac{1}{(n-1)(1+\hat{\beta}_1^2)}\left\{\frac{1}{1-\hat{\rho}^2\hat{\alpha}^2}\sum_{t=2}^{n}\left(\tilde{E}_t-\hat{\rho}\hat{\alpha}\tilde{E}_{t-1}\right)^2-2\frac{\hat{\rho}\hat{\alpha}}{1+\hat{\rho}^2\hat{\alpha}^2}\sum_{t=2}^{n}\left(\tilde{E}_t-\hat{\rho}\hat{\alpha}\tilde{E}_{t-1}\right)\tilde{E}_{t-1}\right\},$$

so that,

$$\hat{\sigma}^2=(1-\hat{\rho}^2)\hat{D}\hat{\alpha},$$

$$\hat{\sigma}_x^2=\frac{1-\hat{\alpha}}{1+\hat{\beta}_1^2}\hat{D},$$

$$\hat{\beta}_0=\left\{-\left(\hat{\beta}_1\bar{X}_n-\bar{Y}_n\right)+\hat{\rho}\left(\hat{\beta}_1\bar{X}_{n-1}-\bar{Y}_{n-1}\right)\right\}/(1-\hat{\rho})$$

and

$$\hat{\mu}_x=\frac{1}{(1+\hat{\beta}_1^2)(1-\hat{\rho}\hat{\alpha})}(\bar{E}_n-\hat{\rho}\hat{\alpha}\bar{E}_{n-1}).$$

These MLEs will be denoted by $\hat{\theta}=\left(\hat{\mu}_x,\hat{\sigma}_x^2,\hat{\sigma}^2,\hat{\beta}_0,\hat{\beta}_1,\hat{\rho}\right)$, and the likelihood ratio test statistic is then represented as

$$T_{\mathrm{LRT},s,\rho}=2\left(\ell_{\mathrm{s}}(\hat{\theta})-\ell_{\mathrm{s}}(\tilde{\theta})\right)$$

First, it is noticed that the likelihood function $\ell_s(\hat{\theta})$ and $\ell_s(\tilde{\theta})$ for structural case can be rewritten as,

$$\ell_s=-(n-1)log(2\pi)-\frac{(n-1)}{2}log|\Sigma_s|-\frac{\sum_{t=2}^{n}(F_t-\rho F_{t-1})^2}{2(1+\beta_1^2)\sigma^2}-\frac{\sigma^2\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\frac{A}{D}\tilde{E}_{t-1}\right)^2}{2(1+\beta_1^2)|\Sigma_s|}.$$

From the first order derivatives for σ_x^2 and σ^2 as shown above, we know that C multiplies equation (2.16B) then plus (2.16C) will give us,

$$-\frac{\sum_{t=2}^{n}(F_t-\rho F_{t-1})^2}{2(1+\beta_1^2)\sigma^2}-\frac{\sigma^2\sum_{t=2}^{n}\left(\tilde{E}_t-\rho\frac{A}{D}\tilde{E}_{t-1}\right)^2}{2(1+\beta_1^2)|\Sigma_s|}=-(n-1).$$

Therefore,

$$\ell_s=-(n-1)log(2\pi)-\frac{(n-1)}{2}log|\Sigma_s|-(n-1).$$

Then the likelihood ratio test statistic will become

$$T_{LRT,s,\rho}=(n-1)\left(-log\left|\hat{\Sigma}_s\right|+log\left|\tilde{\Sigma}_s\right|\right) \qquad (2.16)$$

2.3.1.3 Efficient score test of $H_0\colon(\beta_0,\beta_1)=(b_0,b_1)$ *versus* $H_a\colon(\beta_0,\beta_1)\neq(b_0,b_1)$

Under the null hypothesis $H_0\colon(\beta_0,\beta_1)=(b_0,b_1)$, taking the first derivative of the structural log likelihood with respect to μ_x and setting it equal zero, yields

$$\mu_x=\frac{1}{(1+b_1^2)\left(1-\rho\frac{A}{D_0}\right)}\left(\bar{\dot{E}}_n-\rho\frac{A}{D_0}\bar{\dot{E}}_{n-1}\right) \qquad (2.17A)$$

where $D_0=(1+b_1^2)\sigma_x^2+A$, and $\dot{E}_t=X_t+b_1(Y_t-b_0)$, $\bar{\dot{E}}_n=\sum_{t=2}^{n}\dot{E}_t/(n-1)$.

The first derivatives with respect to σ_x^2, σ^2 and ρ give us,

$$\left\{\begin{array}{c}-(n-1)(1+b_1^2)\left(1+\rho^2\left(\frac{A}{D_0}\right)^2\right)\\+\frac{\sigma^2}{|\Sigma_0|}\left(1+\rho^2\left(\frac{A}{D_0}\right)^2\right)\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\frac{A}{D_0}\tilde{\dot{E}}_{t-1}\right)^2\\-2\rho\frac{A}{D_0^2}\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\frac{A}{D_0}\tilde{\dot{E}}_{t-1}\right)\dot{E}_{t-1}\end{array}\right\}=0 \qquad (2.17B)$$

$$\left\{\begin{array}{c}-(n-1)(1+b_1^2)\left(\frac{|\Sigma_0|}{\sigma^2}+\rho^2\frac{AC_0^2}{D_0^2}+\sigma^2\right)+\frac{|\Sigma_0|}{\sigma^4}\sum_{t=2}^{n}\left(\dot{F}_t-\rho\dot{F}_{t-1}\right)^2\\+\frac{\sigma^2}{|\Sigma_0|}\left(\rho^2\frac{AC_0^2}{D_0^2}+\sigma^2\right)\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\frac{A}{D_0}\tilde{\dot{E}}_{t-1}\right)^2\\+2\rho\frac{AC_0}{D_0^2}\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\frac{A}{D_0}\tilde{\dot{E}}_{t-1}\right)\dot{E}_{t-1}\end{array}\right\}=0 \qquad (2.17C)$$

and

$$\left\{\begin{array}{c}-(n-1)(1+b_1^2)\rho C_0(C_0+\sigma^2)\left(\frac{A}{D_0}\right)^2+\frac{|\Sigma_0|}{\sigma^2}\sum_{t=2}^{n}\left(\dot{F}_t-\rho\dot{F}_{t-1}\right)\dot{F}_{t-1}\\ +\frac{\sigma^2}{|\Sigma_0|}\rho C_0(C_0+\sigma^2)\left(\frac{A}{D_0}\right)^2\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\frac{A}{D_0}\tilde{\dot{E}}_{t-1}\right)^2\\ +\sigma^2\frac{A}{D_0}\left(1+\frac{2\rho^2}{1-\rho^2}\frac{C_0}{D_0}\right)\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\frac{A}{D_0}\tilde{\dot{E}}_{t-1}\right)\dot{E}_{t-1}\end{array}\right\}=0 \quad (2.17\text{D})$$

where $|\Sigma_0|=C_0\sigma^2\left(1+\rho^2\frac{A}{D_0}\right)+\sigma^4$, $C_0=(1+b_1^2)\sigma_x^2$, $\dot{F}_t=b_1X_t-(Y_t-b_0)$ and $\tilde{\dot{E}}_t=\dot{E}_t-\bar{\dot{E}}_n$. Upon substitution of $\alpha_0=\frac{A}{D_0}$, then $\frac{C_0}{D_0}=1-\alpha_0$ and $\frac{|\Sigma_0|}{\sigma^2}=D_0(1-\rho^2\alpha_0^2)$ into (2.17B), (2.17C) and (2.17D), these expressions become,

$$\left\{\begin{array}{c}-(n-1)(1+b_1^2)D_0(1+\rho^2\alpha_0^2)\\ +\frac{1}{1-\rho^2\alpha_0^2}(1+\rho^2\alpha_0^2)\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\alpha_0\tilde{\dot{E}}_{t-1}\right)^2\\ -2\rho\alpha_0\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\alpha_0\tilde{\dot{E}}_{t-1}\right)\dot{E}_{t-1}\end{array}\right\}=0, \quad (2.17\text{B'})$$

$$\left\{\begin{array}{c}-(n-1)(1+b_1^2)D_0(1+\alpha_0-3\rho^2\alpha_0^2+\rho^2\alpha_0^3)\\ +\frac{1-\rho^2\alpha_0^2}{(1-\rho^2)\alpha_0}\sum_{t=2}^{n}\left(\dot{F}_t-\rho\dot{F}_{t-1}\right)^2\\ +\frac{1}{1-\rho^2\alpha_0^2}(\alpha_0-2\rho^2\alpha_0^2+\rho^2\alpha_0^3)\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\alpha_0\tilde{\dot{E}}_{t-1}\right)^2\\ +2\rho\alpha_0(1-\alpha_0)\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\alpha_0\tilde{\dot{E}}_{t-1}\right)\dot{E}_{t-1}\end{array}\right\}=0 \quad (2.17\text{C'})$$

and

$$\left\{\begin{array}{c}-(n-1)(1+b_1^2)D_0\rho(1-\alpha_0)(1-\rho^2\alpha_0)\alpha_0^2\\ +(1-\rho^2\alpha_0^2)\sum_{t=2}^{n}\left(\dot{F}_t-\rho\dot{F}_{t-1}\right)\dot{F}_{t-1}\\ +\frac{1}{1-\rho^2\alpha_0^2}\rho(1-\alpha_0)(1-\rho^2\alpha_0)\alpha_0^2\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\alpha_0\tilde{\dot{E}}_{t-1}\right)^2\\ +(1-\rho^2)\alpha_0^2\left(1+\frac{2\rho^2}{1-\rho^2}(1-\alpha_0)\right)\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\alpha_0\tilde{\dot{E}}_{t-1}\right)\dot{E}_{t-1}\end{array}\right\}=0. \quad (2.17\text{D'})$$

Combining (2.17B') with (2.17C') and (2.17D'), we have

$$\left\{\begin{array}{c}\frac{1-\rho^2\alpha_0^2}{(1-\rho^2)\alpha_0}\sum_{t=2}^{n}\left(\dot{F}_t-\rho\dot{F}_{t-1}\right)^2-\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\alpha_0\tilde{\dot{E}}_{t-1}\right)^2\\ +4\rho\alpha_0\frac{1-\rho^2\alpha_0^2}{1+\rho^2\alpha_0^2}\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\rho\alpha_0\tilde{\dot{E}}_{t-1}\right)\dot{E}_{t-1}\end{array}\right\}=0 \quad (2.17\text{C''})$$

and

$$\sum_{t=2}^{n}\left(\dot{F}_t - \rho\dot{F}_{t-1}\right)\dot{F}_{t-1} + \alpha_0^2\frac{(1+\rho^2)}{1+\rho^2\alpha_0^2}\sum_{t=2}^{n}\left(\dot{\tilde{E}}_t - \rho\alpha_0\dot{\tilde{E}}_{t-1}\right)\dot{E}_{t-1} = 0 \quad \text{(2.17D'')}$$

Solving (2.17C'') and (2.17D'') numerically, we get MLEs for ρ and α_0, which we denote by $\dot{\rho}$ and $\dot{\alpha}_0$, respectively. Finally, from (2.17B') we have,

$$\dot{D}_0 = \frac{1}{(n-1)(1+b_1^2)}\left\{\frac{1}{1-\dot{\rho}^2\dot{\alpha}_0^2}\sum_{t=2}^{n}\left(\dot{\tilde{E}}_t - \dot{\rho}\dot{\alpha}_0\dot{\tilde{E}}_{t-1}\right)^2 - \frac{2\dot{\rho}\dot{\alpha}_0}{1+\dot{\rho}^2\dot{\alpha}_0^2}\sum_{t=2}^{n}\left(\dot{\tilde{E}}_t - \dot{\rho}\dot{\alpha}_0\dot{\tilde{E}}_{t-1}\right)\dot{E}_{t-1}\right\},$$

$$\dot{\sigma}^2 = (1-\dot{\rho}^2)\dot{D}_0\dot{\alpha}_0,$$

$$\dot{\sigma}_x^2 = \frac{1-\dot{\alpha}_0}{(1+b_1^2)}\dot{D}_0$$

and

$$\dot{\mu}_x = \frac{1}{(1+b_1^2)(1-\dot{\rho}\dot{\alpha}_0)}\left(\dot{\bar{E}}_n - \dot{\rho}\dot{\alpha}_0\dot{\bar{E}}_{n-1}\right).$$

The score vector is the vector containing the first derivatives of the structural log likelihood under the null hypothesis $\mathrm{H}_0: (\beta_0, \beta_1) = (\mathrm{b}_0, \mathrm{b}_1)$ taken with respect to $\mu_x, \sigma_x^2, \sigma^2, \rho, \beta_0, \beta_1$ and evaluated at the MLEs, $\dot{\theta} = (\dot{\mu}_x, \dot{\sigma}_x^2, \dot{\sigma}^2, \dot{\rho}, b_0, b_1)$. That is,

$$U(\dot{\theta}) = \left(0,0,0,0,\left.\frac{\partial \ell_s}{\partial \beta_0}\right|_{\beta=b},\left.\frac{\partial \ell_s}{\partial \beta_1}\right|_{\beta=b}\right)^T,$$

where

$$\left.\frac{\partial \ell_s}{\partial \beta_0}\right|_{\beta=b} = -\frac{1-\dot{\rho}}{\dot{\sigma}^2(1+b_1^2)}\sum_{t=2}^{n}\left(\dot{F}_t - \dot{\rho}\dot{F}_{t-1}\right)$$

and

$$\left.\frac{\partial \ell_s}{\partial \beta_1}\right|_{\beta=b} = b_1\frac{1}{(1+b_1^2)^2}\frac{1}{\dot{\sigma}^2}\sum_{t=2}^{n}\left(\dot{F}_t - \dot{\rho}\dot{F}_{t-1}\right)^2 + \frac{4b_1}{(1+b_1^2)^2}\frac{\dot{\sigma}^2}{|\dot{\Sigma}|}\dot{\rho}\frac{\dot{A}\dot{C}_0}{D_0^2}\sum_{t=2}^{n}\left(\dot{\tilde{E}}_t - \dot{\rho}\dot{\alpha}_0\dot{\tilde{E}}_{t-1}\right)\dot{E}_{t-1},$$

$$+\frac{b_1}{(1+b_1^2)^2}\frac{\dot{\sigma}^2}{|\dot{\Sigma}|}\sum_{t=2}^{n}\left(\dot{\tilde{E}}_t - \dot{\rho}\dot{\alpha}_0\dot{\tilde{E}}_{t-1}\right)^2 - \frac{1}{1+b_1^2}\frac{1}{\dot{\sigma}^2}\sum_{t=2}^{n}\left(\dot{F}_t - \dot{\rho}\dot{F}_{t-1}\right)(X_t - \dot{\rho}X_{t-1})$$

$$-\frac{\dot{\sigma}^2}{|\dot{\Sigma}|}\frac{1}{1+b_1^2}\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\dot{\rho}\dot{\alpha}_0\tilde{\dot{E}}_{t-1}\right)(Y_t-\dot{\rho}\dot{\alpha}_0Y_{t-1}).$$

The score test statistic under $H_0\colon(\beta_0,\beta_1)=(b_0,b_1)$ is thus represented as

$$T_{ES,s,\beta}=U(\dot{\theta})^T I(\dot{\theta})^{-1}U(\dot{\theta})$$

Since the first four elements of the score vector are zeros and only the last two are non-zero, only the right bottom two by two sub-block of the inverse of information matrix is required. The estimator of this sub-block is represented by $I(\dot{\theta})_{sub}^{-1}$. Therefore, the test statistic can be written as,

$$T_{ES,s,\beta}=\left(\frac{\partial\ell_s}{\partial\beta_0}\bigg|_{\beta=b},\frac{\partial\ell_s}{\partial\beta_1}\bigg|_{\beta=b}\right)^T I(\dot{\theta})_{sub}^{-1}\left(\frac{\partial\ell_s}{\partial\beta_0}\bigg|_{\beta=b},\frac{\partial\ell_s}{\partial\beta_1}\bigg|_{\beta=b}\right) \qquad (2.17)$$

2.3.1.4 Likelihood ratio test of $H_0\colon(\beta_0,\beta_1)=(b_0,b_1)$ *versus* $H_a\colon(\beta_0,\beta_1)\neq(b_0,b_1)$

The MLEs under the null hypothesis derived in section 2.3.1.3 and were denoted as $\dot{\theta}$. The unconstrained MLEs, $\hat{\theta}$, were derived shown in section 2.3.1.2. The likelihood ratio test statistic is therefore represented as

$$T_{ES,s,\beta}=2\left(\ell_s(\hat{\theta})-\ell_s(\dot{\theta})\right) \qquad (2.17)$$

And also because of the reason shown in 2.3.1.2, the last two summation terms in likelihood function will equal to $(n-1)$. Then the likelihood ratio test statistic is

$$T_{LRT,s,\beta}=(n-1)\left(-log|\hat{\Sigma}_s|+log|\dot{\Sigma}_s|\right) \qquad (2.18)$$

2.3.2 Functional case

2.3.2.1 Efficient score test of $H_0\colon\rho=0$ *versus* $H_a\colon\rho\neq0$

Under the null hypothesis $\mathrm{H}_0\colon\rho=0$, the log likelihood function for the functional case is

$$\ell_f\Big|_{\rho=0} = -(n-1)log(2\pi) - (n-1)log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{n}\{(X_t - x_t)^2 + (Y_t - y_t)^2\}.$$

Taking first derivatives of the log likelihood function with respect to x_1, x_2, …, x_n, β_0, β_1,σ^2, and then setting these derivatives equal zero results in the following set of equations:

$$(X_t - x_t) + \beta_1(Y_t - y_t) = 0, \qquad t = 1, 2, \dots, n$$

$$\sum_{t=1}^{n}(Y_t - y_t) = 0,$$

$$\sum_{t=1}^{n}(Y_t - y_t)x_t = 0$$

and

$$\sigma^2 = \frac{1}{2n}\sum_{t=2}^{n}\{(X_t - x_t)^2 + (Y_t - y_t)^2\}.$$

Letting $S_{XX} = \sum_{t=2}^{n}(X_t - \bar{X})^2$, $S_{YY} = \sum_{t=2}^{n}(Y_t - \bar{Y})^2$, and $S_{XY} = \sum_{t=2}^{n}(X_t - \bar{X})(Y_t - \bar{Y})$, the solution of this system of equations is the MLEs under H_0: $\rho = 0$, which is,

$$\tilde{\beta}_1 = \frac{-(S_{XX} - S_{YY}) + \sqrt{(S_{XX} - S_{YY})^2 + 4S_{XY}^2}}{2S_{XY}},$$

$$\tilde{\beta}_0 = \bar{Y} - \tilde{\beta}_1\bar{X},$$

$$\tilde{x}_t = \frac{(Y_t - \tilde{\beta}_0)\tilde{\beta}_1 + X_t}{1 + \tilde{\beta}_1^2} \qquad t = 1, 2, \dots, n,$$

and

$$\tilde{\sigma}^2 = \frac{1}{2(n-1)}\frac{1}{(1 + \tilde{\beta}_1^2)}\sum_{t=2}^{n}\left(\tilde{\beta}_1(X_t - \bar{X}) - (Y_t - \bar{Y})\right)^2.$$

Since in the functional case, $x_t, t = 1,2, \dots, n$ are fixed values that must be treated as unknown parameters, the score vector is the first derivatives of ℓ_f taken with respect to $\psi = (x_1, x_2, \dots, x_n, \beta_0, \beta_1, \sigma^2, \rho)$ under the null hypothesis. That is,

$$U_f(\tilde{\psi}) = \left(0,0,\dots,0,\frac{\partial \ell_f}{\partial \psi}\bigg|_{\rho=0}\right)^T,$$

where $\frac{\partial \ell_f}{\partial \rho}\Big|_{\rho=0} = \frac{1}{\tilde{\sigma}^2}\sum_{t=2}^{n}\{(X_t - \tilde{x}_t)(X_{t-1} - \tilde{x}_{t-1}) + (Y_t - \tilde{y}_t)(Y_{t-1} - \tilde{y}_{t-1})\}$.

As previously discussed in the structural case, the only element needed to construct the score statistic is the lower right diagonal element of the inverse of information matrix. After some derivations (see Appendix C), this value is found to be $\frac{1}{2(n-1)}$. Therefore, the efficient score test statistic for testing $H_0: \rho = 0$ versus $H_a: \rho \neq 0$ in the functional case is

$$T_{ES,f,\rho} = \frac{1}{2(n-1)}\left(\frac{\partial \ell_f}{\partial \rho}\bigg|_{\rho=0}\right)^2 \quad (2.18)$$

2.3.2.2 Likelihood ratio test of $H_0: \rho = 0$ *versus* $H_a: \rho \neq 0$

The MLEs under the null hypothesis were derived in section 2.3.2.1. We now calculate the unconstrained MLEs. Taking first derivatives of the log likelihood function with respect to $\psi = (x_2, x_3, \dots, x_n, \beta_0, \beta_1, \sigma^2, \rho)$, and setting these derivatives equal to zero produces the following system of equations:

$$(X_t - \rho X_{t-1} - \lambda_t) + (Y_t - \rho Y_{t-1} - \gamma_t)\beta_1 + (X_{t+1} - \rho X_t - \lambda_{t+1})(-\rho)$$
$$+(Y_{t+1} - \rho Y_t - \gamma_{t+1})(-\rho\beta_1) = 0 \qquad t = 2,\dots,n-1 \quad (2.19\text{A})$$

$$(X_n - \rho X_{n-1} - \lambda_n) + (Y_n - \rho Y_{n-1} - \gamma_n)\beta_1 = 0 \quad (2.19\text{B})$$

$$\sum_{t=2}^{n}(Y_t - \rho Y_{t-1} - \gamma_t)(1-\rho) = 0 \quad (2.19\text{C})$$

$$\sum_{t=2}^{n}(Y_t - \rho Y_{t-1} - \gamma_t)(x_t - \rho x_{t-1}) = 0 \quad (2.19\text{D})$$

$$-2(n-1)\sigma^2 + \sum_{t=2}^{n}\{(X_t - \rho X_{t-1} - \lambda_t)^2 + (Y_t - \rho Y_{t-1} - \gamma_t)^2\} = 0 \quad (2.19\text{E})$$

$$\sum_{t=2}^{n}\{(X_t - \rho X_{t-1} - \lambda_t)(X_{t-1} - x_{t-1}) + (Y_t - \rho Y_{t-1} - \gamma_t)(Y_{t-1} - y_{t-1})\} = 0 \quad (2.19\text{F})$$

We regard the x_t as nuisance parameters. Therefore, we reparameterize from x_t to

$$\lambda_t = x_t - \rho x_{t-1}$$

to simplify subsequent calculations. Equations (2.19A) and (2.19B) show that,

$$\hat{\lambda}_t = \frac{(X_t - \hat{\rho} X_{t-1}) + \hat{\beta}_1 (Y_t - \hat{\rho} Y_{t-1}) - \hat{\beta}_0 \hat{\beta}_1 (1-\hat{\rho})}{1+\hat{\beta}_1^2} \qquad t = 2,3,\dots,n$$

Substituting of $\hat{\lambda}_t$ into equation (2.19C), we have

$$\hat{\beta}_0 = \frac{(\bar{Y}_n - \hat{\rho}\bar{Y}_{n-1}) - \hat{\beta}_1(\bar{X}_n - \hat{\rho}\bar{X}_{n-1})}{1-\hat{\rho}}.$$

From equation (2.19E) we see that

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{t=2}^{n} \left\{ \left(X_t - \hat{\lambda}_t - \hat{\rho} X_{t-1}\right)^2 + (Y_t - \hat{\gamma}_t - \hat{\rho} Y_{t-1})^2 \right\}.$$

Equations (2.19D) and (2.19F) reveal that

$$\hat{\rho} = \frac{\sum_{t=2}^{n} \{(\hat{\beta}_1 X_t - Y_t) - (\hat{\beta}_1 \bar{X}_n - \bar{Y}_n)\}\{(\hat{\beta}_1 X_{t-1} - Y_{t-1}) - (\hat{\beta}_1 \bar{X}_{n-1} - \bar{Y}_{n-1})\}}{\sum_{t=2}^{n} \{(\hat{\beta}_1 X_{t-1} - Y_{t-1}) - (\hat{\beta}_1 \bar{X}_{n-1} - \bar{Y}_{n-1})\}^2}$$

and

$$\hat{\beta}_1 = \frac{-(\hat{S}_{XX} - \hat{S}_{YY}) + \sqrt{(\hat{S}_{XX} - \hat{S}_{YY})^2 + 4\hat{S}_{XY}^2}}{2\hat{S}_{XY}},$$

where

$$\hat{S}_{XX} = \sum_{t=2}^{n} \{(X_t - \hat{\rho} X_{t-1}) - (\bar{X}_n - \hat{\rho}\bar{X}_{n-1})\}^2,$$

$$\hat{S}_{YY} = \sum_{t=2}^{n} \{(Y_t - \hat{\rho} Y_{t-1}) - (\bar{Y}_n - \hat{\rho}\bar{Y}_{n-1})\}^2$$

and

$$\hat{S}_{XY} = \sum_{t=2}^{n} \{(X_t - \hat{\rho} X_{t-1}) - (\bar{X}_n - \hat{\rho}\bar{X}_{n-1})\}\{(Y_t - \hat{\rho} Y_{t-1}) - (\bar{Y}_n - \hat{\rho}\bar{Y}_{n-1})\}.$$

The expressions for $\hat{\beta}_1$ and $\hat{\rho}$ are used to obtain numerical solutions for the MLEs of β_1 and ρ. Therefore, the likelihood ratio test statistic for H_0: $\rho = 0$ is

$$T_{LRT,f,\rho} = 2\left(\ell_f(\hat{\psi}) - \ell_f(\tilde{\psi})\right) = -2(n-1)(log\hat{\sigma}^2 - log\tilde{\sigma}^2) \qquad (2.19)$$

2.3.2.3 Efficient score test of $H_0: (\beta_0, \beta_1) = (b_0, b_1)$ versus $H_a: (\beta_0, \beta_1) \neq (b_0, b_1)$

Again, we take the first derivative of the functional case log likelihood, ℓ_f, with respect to $x_2, x_3, \ldots, x_n$, and set these derivatives equal to zero. Under the null hypothesis, $H_0: (\beta_0, \beta_1) = (b_0, b_1)$, we can the solve for the nuisance parameter, λ_t, to obtain

$$\ddot{\lambda}_t = \frac{(X_t - \ddot{\rho}X_{t-1}) + b_1(Y_t - \ddot{\rho}Y_{t-1}) - b_0 b_1(1-\ddot{\rho})}{1+b_1^2} \qquad t = 2,3,\ldots,n.$$

Upon taking the first derivative of ℓ_f, with respect to ρ, and substituting the nuisance parameter estimator into it, we find that

$$\ddot{\rho} = \frac{\sum_{t=2}^{n}(b_1X_t - Y_t + b_0)(b_1X_{t-1} - Y_{t-1} + b_0)}{\sum_{t=2}^{n}(b_1X_{t-1} - Y_{t-1} + b_0)^2}.$$

Similarly taking the first derivative of ℓ_f, with respect to σ^2 gives us

$$\ddot{\sigma}^2 = \frac{1}{2(n-1)}\sum_{t=2}^{n}\left\{\left(X_t - \ddot{\lambda}_t - \ddot{\rho}X_{t-1}\right)^2 + (Y_t - \ddot{\gamma}_t - \ddot{\rho}Y_{t-1})^2\right\}.$$

The score vector is formed by evaluating the first derivative of ℓ_f with respect to $\varphi = (\lambda_2, \lambda_3, \ldots, \lambda_n, \beta_0, \beta_1, \sigma^2, \rho)$ at $\ddot{\varphi} = \left(\ddot{\lambda}_2, \ddot{\lambda}_3, \ldots, \ddot{\lambda}_n, b_0, b_1, \ddot{\sigma}^2, \ddot{\rho}\right)$. The first derivative again is taken with respect to λ_t instead of x_t because the inverse of the corresponding information matrix is easier to derive for λ_t than for x_t. Thus, the score vector is

$$U(\ddot{\varphi}) = \left(0,0,\ldots,0, \frac{\partial \ell_f}{\partial \beta_0}\bigg|_{\beta=b}, \frac{\partial \ell_f}{\partial \beta_1}\bigg|_{\beta=b}, 0,0\right)^T,$$

where

$$\frac{\partial \ell_f}{\partial \beta_0}\bigg|_{\beta=b} = \ddot{\sigma}^{-2}\sum_{t=2}^{n}\{(Y_t - \ddot{\gamma}_t - \ddot{\rho}Y_{t-1})(1-\ddot{\rho})\}$$

and

$$\left.\frac{\partial \ell_f}{\partial \beta_1}\right|_{\beta=b} = \ddot{\sigma}^{-2} \sum_{t=2}^{n} \left\{ (Y_t - \ddot{\gamma}_t - \ddot{\rho} Y_{t-1}) \ddot{\lambda}_t \right\}.$$

The information matrix is of dimension $(n+3) \times (n+3)$. In order to form our test statistic, we require only the 2×2 diagonal sub-block of the inverse information matrix that corresponds to β_0 and β_1. This sub-block is (see Appendix D)

$$I(\ddot{\varphi})_{2\times 2}^{-1} = \frac{\ddot{\sigma}^2 (1+b_1^2)}{(1-\ddot{\rho})^2 \left\{ (n-1) \sum_{t=2}^{n} \ddot{\lambda}_t^2 - \left(\sum_{t=2}^{n} \ddot{\lambda}_t \right)^2 \right\}} \begin{bmatrix} \sum_{t=2}^{n} \ddot{\lambda}_t^2 & -(1-\ddot{\rho}) \sum_{t=2}^{n} \ddot{\lambda}_t \\ -(1-\ddot{\rho}) \sum_{t=2}^{n} \ddot{\lambda}_t & (n-1)(1-\ddot{\rho})^2 \end{bmatrix}.$$

Thus, the score test statistic for testing $H_0: (\beta_0, \beta_1) = (b_0, b_1)$ versus $H_{\mathrm{a}}: (\beta_0, \beta_1) \neq (b_0, b_1)$ in the functional case is

$$T_{ES,f,\beta} = U(\ddot{\varphi})^T I(\ddot{\varphi})_{2\times 2}^{-1} U(\ddot{\varphi}) \qquad (2.20)$$

2.3.2.4 Likelihood ratio test of $H_0: (\beta_0, \beta_1) = (b_0, b_1)$ *versus* $H_a: (\beta_0, \beta_1) \neq (b_0, b_1)$

The MLEs under the null hypothesis and the MLEs under the alternative hypothesis were derived in section 2.3.2.3 and 2.3.2.2, respectively. Substituting these estimators into the log likelihood function, ℓ_f, completes the likelihood ratio test statistic in this functional case:

$$T_{LRT,f,\beta} = 2\left(\ell_f(\hat{\psi}) - \ell_f(\ddot{\varphi}) \right) = -2(n-1)(log\hat{\sigma}^2 - log\ddot{\sigma}^2). \qquad (2.21)$$

3. PRELIMINARY THEORIES

We present without proof in this section known results that we use to derive properties of our estimators and test statistics. Section 3.1 reviews some large sample convergence properties. Various central limit theorems are presented in Section 3.2. These theorems will be used to find the asymptotic distributions of our test statistics. Section 3.3 contains additional useful mathematical and statistical tools, including relationships among normal and chi-squared distributions, derivatives for implicit functions, and Satterthwaite approximations.

3.1 Convergence Properties

To investigate the consistency of the maximum likelihood estimators and their large sample distributions, we need the following results.

Theorem 3.1 *Let* $X_1, X_2, \ldots$ *and* X *be random vectors*, *and* g *is a continuous function*. *Then*

$$X_n \xrightarrow{p} X \ \Rightarrow g(X_n) \xrightarrow{p} g(X),$$

and

$$X_n \xrightarrow{d} X \ \Rightarrow g(X_n) \xrightarrow{d} g(X).$$

Proof: (Serfling, 1980).

Thus, e.g., we know that if $X_n \xrightarrow{d} N(0,1)$, then $X_n^2 \xrightarrow{d} \chi_1^2$. Furthermore, if $X_n \to X$ in probability or in distribution, and A and B are two matrices with correct dimensions, we then have that

$$AX_n \to AX \text{ and } X_n^T B X_n \to X^T B X,$$

in the same way of convergence as X_n. A useful corollary of Theorem 3.1 is given below (Serfling 1980).

Corollary 3.2 If $X_n \xrightarrow{d} N(\mu, \Sigma)$, and A is a matrix, then $AX_n \xrightarrow{d} N(A\mu, A^T\Sigma A)$.

The following theorem presents a weak law of large numbers for uncorrelated, but not necessarily independent, random variable.

Theorem 3.3 (Chebyshev Theorem) *Let $X_1, X_2, \ldots$ be a sequence of random variables with means $\mu_1, \mu_2, \ldots$ and variances $\sigma_1^2, \sigma_2^2, \ldots$. Suppose $cov(X_t, X_s) = 0,\ t \neq s$. If $\lim_{n\to\infty} \frac{\sum_{t=1}^{n} \sigma_t^2}{n^2} = 0$, then*

$$\frac{\sum_{t=1}^{n} X_t}{n} - \frac{\sum_{t=1}^{n} \mu_t}{n} \xrightarrow{p} 0.$$

Proof: (Rao, 1973).

3.2 Central Limit Theorems

The classical Lindeberg central limit theorem is for a sequence of independent and identically distributed (i.i.d.) random variables with finite variance. A useful extension (Theorem 3.4 below) was proposed by Rao (1973) for random variables with different means and perhaps different covariance matrices. Although Rao's extension relaxes the assumptions of the variables having the same means and variances, independence between variables is still required.

Theorem 3.4 *Let $X_1, X_2, \ldots$ be independent random vectors with mean $\mu_1, \mu_2, \ldots$, covariances $\Sigma_1, \Sigma_2, \ldots$, and distribution functions $F_1, F_2, \ldots$. If*

$$\lim_{n\to\infty} \frac{\Sigma_1 + \Sigma_2 + \cdots + \Sigma_n}{n} = \Sigma,$$

and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \int_{\|x-\mu_t\|>\varepsilon\sqrt{n}} \|x-\mu_t\|^2 dF_t(x) = 0, \quad \forall\varepsilon > 0.$$

Then

$$\frac{\sum_{t=1}^{n} X_t}{n} \sim AN\left(\frac{\sum_{t=1}^{n} \mu_t}{n}, \frac{\Sigma}{n}\right).$$

However, in our derivations of the asymptotic distributions of the test statistics, we will encounter dependent variables. Therefore, we need a theorem for dependent variables. First, we define m-dependence following the definition and theorem by DasGupta (2008).

Definition 3.1 A stationary sequence $X_1, X_2, \dots$ is called m-dependent if $(X_1, X_2, \dots, X_i)$ and $(X_{i+j}, X_{i+j+1}, \dots)$ are independent whenever $j > m$ (DasGupta, 2008).

Theorem 3.5 Let $X_1, X_2, \dots$ be a stationary m-dependent sequence with $E(X_t) = \mu$ and $Var(X_t) = \sigma^2 < \infty$. Then

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} N(0, \tau^2),$$

where

$$\tau^2 = \sigma^2 + 2\sum_{i=1}^{m} cov(X_1, X_{1+i}).$$

Proof: (Lehmann, 1999).

Another useful result is the Cramer-Wold theorem as presented by Athreya and Lahiri (2006).

Theorem 3.6 (Cramer-Wold) Let $X_1, X_2, \dots$ be a sequence of p-dimensional random vectors and *X* be a p-dimensional random vector. $X_n \xrightarrow{d} X$ if and only if $aX_n \xrightarrow{d} aX$ for all $a \in \mathcal{R}^p$.

Proof: (Athreya and Lahiri, 2006).

It will be shown in section 4 that the test statistics are functions of various sample moments. For those without explicit solutions, we use Taylor's series expansions to find the large sample distribution of test statistics. These expansions are moment expansions, i.e., expansions taken with respect to sample moments. Consequently, we require the large sample, distributions for these sample moments. The following two theorems are central limit theorems for sample moments. As shown by Fuller (1987), Theorems 3.7 and 3.8 present limiting distributions of sample moments in structural and functional case, respectively.

Theorem 3.7 Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed k-dimensional random vectors with mean μ, covariance matrix Σ, and finite fourth moments. Let $\bar{X}$ denote the sample mean, and $M = \left(m_{ij}\right)$ the sample covariance matrix. Define the vector half of M and Σ to be

$$\text{vech}M = \left(m_{11}, m_{21}, \ldots, m_{k1}; m_{22}, m_{32}, \ldots, m_{k2}; \ldots; m_{(k-1)(k-1)}, m_{k(k-1)}; m_{kk}\right)^T,$$

$$\text{vech}\Sigma = \left(\sigma_{11}, \sigma_{21}, \ldots, \sigma_{k1}; \sigma_{22}, \sigma_{32}, \ldots, \sigma_{k2}; \ldots; \sigma_{(k-1)(k-1)}, \sigma_{k(k-1)}; \sigma_{kk}\right)^T,$$

$$y_t = (X_t - \mu, [\text{vech}\{(X_t - \mu)^T (X_t - \mu) - \Sigma\}]^T),$$

and $\Omega = E(y_t y_t^T)$. Then

$$n^{1/2}(\bar{X} - \mu, [\text{vech}\{M - \Sigma\}]^T)^T \xrightarrow{\mathcal{L}} N(0, \Omega).$$

Proof: (Fuller, 1987).

Theorem 3.8 Let $X_t = x_t + \varepsilon_t$, where ε_t are independent identically distributed k-dimensional random row vectors with zero mean vector, positive definite covariance matrix Σ, and finite fourth moments. Let $x_1, x_2, \ldots$ be a fixed sequence satisfying

$$\lim_{n\to\infty} \bar{x} = \mu_x,$$

$$\lim_{n\to\infty} M = \lim_{n\to\infty} \sum_{t=1}^{n} (x_t - \bar{x})^T (x_t - \bar{x})/(n-1) = \bar{M}.$$

Let $\hat{\theta} = (\bar{X}, (\text{vech}M)^T)^T$ and $\theta_n = (\bar{x}, (\text{vech}\{M + \Sigma\})^T)^T$, where the vech operators are defined analogously to those in Theorem 3.7. Then

$$\Omega_n^{-1/2}\left(\hat{\theta} - \theta_n\right) \xrightarrow{\mathcal{L}} N(0, I),$$

where Ω_n the covariances matrix of $\hat{\theta}$, and

$$cov\left(\bar{X}_i, \left(\bar{X}_j, M_{jk}\right)\right) = n^{-1}\left(E\left(\varepsilon_i\varepsilon_j\right), E\left(\varepsilon_i\varepsilon_j\varepsilon_k\right)\right),$$

$$cov\left(M_{ij}, M_{kl}\right) = \frac{M_{ik}E\left(\varepsilon_l\varepsilon_j\right) + M_{il}E\left(\varepsilon_k\varepsilon_j\right) + M_{jk}E\left(\varepsilon_l\varepsilon_i\right) + M_{jl}E\left(\varepsilon_k\varepsilon_i\right) + E\left[\varepsilon_i\varepsilon_j - E\left(\varepsilon_i\varepsilon_j\right)\right]\left[\varepsilon_k\varepsilon_l - E\left(\varepsilon_k\varepsilon_l\right)\right]}{n-1}$$

$$+O(n^{-2}).$$

Proof: (Fuller, 1987).

3.3 Other Useful Methods

Relationships between normal and chi-squared distributions are helpful in deriving the large sample distributions of our estimators and test statistics. We know that if a random variable X is normally distributed with mean μ and variance σ^2, then X^2/σ^2 has a noncentral chi-squared distribution with one degree of freedom and noncentrality parameter μ^2/σ^2. It is well known that the distribution of a sum of squares of k independent standard normal random variables $X_1, X_2, \ldots, X_k$, is chi-squared with k degrees of freedom. A more general theorem relating multivariate normal random variables to chi-square distribution is provided by the following theorem (Serfling 1980).

Theorem 3.9 *Let* $X = (X_1, X_2, \dots, X_k)$ *be a multivariate variable normally distributed with mean vector* μ *and covariance matrix* Σ. *Let* A *be a* $k \times k$ *symmetric matrix. Suppose for* $\theta = (\theta_1, \theta_2, \dots, \theta_k)$,

$$\theta\Sigma = 0 \Rightarrow \theta\mu^T = 0.$$

Then XAX^T *has a chi-squared distribution if and only if*

$$\Sigma A \Sigma A \Sigma A = \Sigma A \Sigma,$$

the degrees of freedom is the trace of $A\Sigma$ *and the noncentrality parameter is* $\mu A \mu^T$.

Proof: (Serfling, 1980).

Some of the maximum likelihood estimators shown in next section do not have explicit expressions. We obtain derivatives for these estimators using the implicit function theorem (Krantz and Parks, 2002). Suppose we have a differentiable function

$$F(y_1, x_1, x_2, \dots, x_k) = 0$$

and y is an implicit function of x, i.e., $y = y(x_1, x_2, \dots, x_k)$. Subject to certain conditions (Krantz and Parks, 2002), then for each $i = 1,2, \dots, k$ application of the chain rule yields

$$\frac{\partial y}{\partial x_i} = -\frac{\partial F / \partial x_i}{\partial F / \partial y}.$$

Applying the chain rule again, we can get the second order derivatives of y with respect to x by using the equation below,

$$\frac{\partial F}{\partial y} \frac{\partial^2 y}{\partial x_i \partial x_j} + \left(\frac{\partial^2 F}{\partial y^2} \frac{\partial y}{\partial x_j} + \frac{\partial^2 F}{\partial y \partial x_j} \right) \frac{\partial y}{\partial x_i} + \frac{\partial^2 F}{\partial x_i \partial y} \frac{\partial y}{\partial x_j} + \frac{\partial^2 F}{\partial x_i \partial x_j} = 0.$$

This is for the situation where function y only has one variable x. For our case, in which we have three implicit functions of several variables, it is more complicated to get first and second order partial derivatives using this technique. For example, suppose we have three differentiable functions

$$G_i(y_1, y_2, y_3, x_1, x_2, \dots, x_k) = 0,$$

and three implicit functions $y_i = y_i(x_1, x_2, \dots, x_k)$, $i = 1,2,3$. To find the first order derivatives of y with respect to x_i, we need to derive the set of equations,

$$\frac{\partial G_1}{\partial y_1}\frac{\partial y_1}{\partial x_i} + \frac{\partial G_1}{\partial y_2}\frac{\partial y_2}{\partial x_i} + \frac{\partial G_1}{\partial y_3}\frac{\partial y_3}{\partial x_i} + \frac{\partial G_1}{\partial x_i} = 0,$$

$$\frac{\partial G_2}{\partial y_1}\frac{\partial y_1}{\partial x_i} + \frac{\partial G_2}{\partial y_2}\frac{\partial y_2}{\partial x_i} + \frac{\partial G_2}{\partial y_3}\frac{\partial y_3}{\partial x_i} + \frac{\partial G_2}{\partial x_i} = 0,$$

$$\frac{\partial G_3}{\partial y_1}\frac{\partial y_1}{\partial x_i} + \frac{\partial G_3}{\partial y_2}\frac{\partial y_2}{\partial x_i} + \frac{\partial G_3}{\partial y_3}\frac{\partial y_3}{\partial x_i} + \frac{\partial G_3}{\partial x_i} = 0.$$

Applying the chain rule again to these equations, we derive the second order derivatives of y with respect to x_i as shown below, $k = 1,2,3$,

$$\begin{aligned}
0 = {} & \frac{\partial y_1}{\partial x_i}\left(\frac{\partial^2 G_k}{\partial y_1^2}\frac{\partial y_1}{\partial x_j} + \frac{\partial^2 G_k}{\partial y_1 \partial y_2}\frac{\partial y_2}{\partial x_j} + \frac{\partial^2 G_k}{\partial y_1 \partial y_3}\frac{\partial y_3}{\partial x_j} + \frac{\partial^2 G_k}{\partial y_1 \partial x_j}\right) + \frac{\partial G_k}{\partial y_1}\frac{\partial^2 y_1}{\partial x_i \partial x_j} \\
& + \frac{\partial y_2}{\partial x_i}\left(\frac{\partial^2 G_k}{\partial y_1 \partial y_2}\frac{\partial y_1}{\partial x_j} + \frac{\partial^2 G_k}{\partial y_2^2}\frac{\partial y_2}{\partial x_j} + \frac{\partial^2 G_k}{\partial y_2 \partial y_3}\frac{\partial y_3}{\partial x_j} + \frac{\partial^2 G_k}{\partial y_2 \partial x_j}\right) + \frac{\partial G_k}{\partial y_2}\frac{\partial^2 y_2}{\partial x_i \partial x_j} \\
& + \frac{\partial y_3}{\partial x_i}\left(\frac{\partial^2 G_k}{\partial y_1 \partial y_3}\frac{\partial y_1}{\partial x_j} + \frac{\partial^2 G_k}{\partial y_2 \partial y_3}\frac{\partial y_2}{\partial x_j} + \frac{\partial^2 G_k}{\partial y_3^2}\frac{\partial y_3}{\partial x_j} + \frac{\partial^2 G_k}{\partial y_3 \partial x_j}\right) + \frac{\partial G_k}{\partial y_3}\frac{\partial^2 y_3}{\partial x_i \partial x_j} \\
& + \frac{\partial^2 G_k}{\partial x_i \partial y_1}\frac{\partial y_1}{\partial x_j} + \frac{\partial^2 G_k}{\partial x_i \partial y_2}\frac{\partial y_2}{\partial x_j} + \frac{\partial^2 G_k}{\partial x_i \partial y_3}\frac{\partial y_3}{\partial x_j} + \frac{\partial^2 G_k}{\partial x_i \partial x_j}.
\end{aligned}$$

This system of equations appears to be quite complicated. However, in our application of these results, many of the terms are zero.

In our derivations we encounter sums of weighted independent chi-squared random variables. In these cases, we employ Satterthwaite approximation (Satterthwaite, 1946). Suppose we have k independent chi-square random variables X_1, X_2, …, X_k with degrees of freedom υ_1, υ_2, …, υ_k respectively. Let

$$X = c_1X_1 + c_2X_2 + \cdots + c_kX_k.$$

Then $\upsilon X/E(X)$ is asymptotically distributed as a chi-square distribution with degrees of freedom υ where

$$\upsilon = \frac{X^2}{\sum_{i=1}^{k}\{(c_iX_i)^2/\upsilon_i\}}.$$

4. DISTRIBUTIONS OF TEST STATISTICS

As we have shown in Section 2, there are two models which are structural and functional simple linear measurement error models. For each model, we are interested in testing the hypothesis of $H_0: \rho = 0$, and $H_0: \beta = b$. We derive the likelihood ratio test statistic and efficient score test statistic for every scenario. In this section, we will investigate the consistency of the maximum likelihood estimators and the asymptotic distribution for test statistics.

4.1 Consistency for Parameter Estimators

As stated in section 2.2.1, the likelihood function is shown in (2.13). After that, the maximum likelihood estimators for structural model and functional model under two different null hypothesis are derived respectively. Also, the MLE without any constraint is derived in each case. Below, we are going to show the consistency of these parameter estimators.

4.1.1 Structural case

In structural model, we have the relationship $y_t = \beta_0 + \beta_1 x_t$ and,

$$\binom{X_t}{Y_t} = \binom{x_t}{y_t} + \binom{e_t}{u_t}, \qquad t = 1,2,\ldots,n.$$

The main difference from a functional model is that x_t is a unobservable random sample from $N(\mu_x, \sigma_x^2)$. Under the null hypothesis of $H_0: \rho = 0$, e_t and u_t are two independent variables distributed as $N(0, \sigma^2)$. Therefore, under the null hypothesis, the consistency of MLE $\left(\tilde{\mu}_x, \tilde{\sigma}_x^2, \tilde{\beta}_0, \tilde{\beta}_1, \tilde{\sigma}^2\right)$ can be shown as follows.

From the set up in Section 2, we know that

$$\bar{X}_n = \sum_{t=2}^{n} X_t/(n-1)$$

and

$$X_t = x_t + e_t, \qquad t = 1,2,\dots,n.$$

Therefore,

$$plim\tilde{\mu}_x = plim\bar{X}_n = plim\sum_{t=2}^{n} x_t/(n-1) + plim\sum_{t=2}^{n} e_t/(n-1) = \mu_x.$$

Thus, the MLE for mean of the distribution where x_t comes from is consistent. Since x_t, e_t and u_t are independent with one another, the consistency for other parameter estimators is easy to show. Remember that for $t = 1,2,\dots,n$,

$$\tilde{X}_t = X_t - \bar{X}_n = x_t + e_t - \bar{x}_n - \bar{e}_n$$

and

$$\tilde{Y}_t = Y_t - \bar{Y}_n = \beta_0 + \beta_1 x_t + u_t - \beta_0 - \beta_1\bar{x}_n - \bar{u}_n = \beta_1 x_t + u_t - \beta_1\bar{x}_n - \bar{u}_n.$$

First, by independency of x_t, e_t and u_t for any $t = 1,2,\dots,n$, we have,

$$plim\sum_{t=2}^{n}(x_t - \bar{x}_n)(e_t - \bar{e}_n)/(n-1) = 0,$$

$$plim\sum_{t=2}^{n}(x_t - \bar{x}_n)(u_t - \bar{u}_n)/(n-1) = 0,$$

$$plim\sum_{t=2}^{n}(e_t - \bar{e}_n)(u_t - \bar{u}_n)/(n-1) = 0.$$

Also,

$$plim\sum_{t=2}^{n}(x_t - \bar{x}_n)^2/(n-1) = \sigma_x^2.$$

Therefore, substituting the above relationships we have,

$$plim\sum_{t=2}^{n}\tilde{X}_t\tilde{Y}_t/(n-1) = \beta_1 plim\frac{\sum_{t=2}^{n}(x_t - \bar{x}_n)^2}{(n-1)} = \beta_1\sigma_x^2. \qquad (4.1)$$

Furthermore, since

$$plim \sum_{t=2}^{n}(e_t - \bar{e}_n)^2/(n-1) = plim \sum_{t=2}^{n}(u_t - \bar{u}_n)^2/(n-1) = \sigma^2,$$

then

$$plim \sum_{t=2}^{n}\left(\tilde{X}_t^2 - \tilde{Y}_t^2\right)/(n-1) = plim \sum_{t=2}^{n} \frac{(x_t-\bar{x}_n+e_t-\bar{e}_n)^2-(\beta_1 x_t-\beta_1\bar{x}_n+u_t-\bar{u}_n)^2}{(n-1)}$$

$$= plim \sum_{t=2}^{n} \frac{\left((1-\beta_1^2)(x_t-\bar{x}_n)^2+(e_t-\bar{e}_n)^2-(u_t-\bar{u}_n)^2\right)}{(n-1)}$$

$$= (1-\beta_1^2)\sigma_x^2. \tag{4.2}$$

Substituting these two quantities (4.1) and (4.2) into the estimators for β_0 and β_1, we can see the consistency of $\tilde{\beta}_0$ and $\tilde{\beta}_1$.

$$plim\tilde{\beta}_1 = plim \frac{-\sum_{t=2}^{n}(\tilde{X}_t^2-\tilde{Y}_t^2)+\sqrt{\left(\sum_{t=2}^{n}(\tilde{X}_t^2-\tilde{Y}_t^2)\right)^2+4\left(\sum_{t=2}^{n}\tilde{X}_t\tilde{Y}_t\right)^2}}{2\sum_{t=2}^{n}\tilde{X}_t\tilde{Y}_t}$$

$$= \frac{-(1-\beta_1^2)\sigma_x^2+\sqrt{\left((1-\beta_1^2)\sigma_x^2\right)^2+4\left(\beta_1\sigma_x^2\right)^2}}{2\beta_1\sigma_x^2} = \frac{-(1-\beta_1^2)\sigma_x^2+(1+\beta_1^2)\sigma_x^2}{2\beta_1\sigma_x^2} = \beta_1$$

And thereafter,

$$plim\tilde{\beta}_0 = plim\left(\bar{Y}_n - \tilde{\beta}_1\bar{X}_n\right) = plim\left\{\frac{\sum_{t=2}^{n}(\beta_0+\beta_1 x_t+u_t)}{(n-1)} - \tilde{\beta}_1\frac{\sum_{t=2}^{n}(x_t+e_t)}{(n-1)}\right\}$$

$$= \beta_0 + \beta_1\mu_x - \beta_1\mu_x = \beta_0$$

From the above derivation, we can see that MLE for (β_0, β_1) are consistent estimators. Similarly, we can check the consistency of the variance estimators below. For the variance estimator of σ_x^2, from (4.1), we have

$$plim\tilde{\sigma}_x^2 = plim\frac{\sum_{t=2}^{n}(\tilde{X}_t\tilde{Y}_t)}{(n-1)\tilde{\beta}_1} = \sigma_x^2.$$

For the variance estimator of σ^2, we have

$$plim\tilde{\sigma}^2 = plim\frac{\sum_{t=2}^{n}\left(\tilde{\beta}_1\tilde{X}_t-\tilde{Y}_t\right)^2}{(n-1)\left(1+\tilde{\beta}_1^2\right)} = plim\frac{\sum_{t=2}^{n}\left(\tilde{\beta}_1(x_t-\bar{x}_n+e_t-\bar{e}_n)-(\beta_1 x_t-\beta_1\bar{x}_n+u_t-\bar{u}_n)\right)^2}{(n-1)\left(1+\tilde{\beta}_1^2\right)}$$

$$= plim \frac{\sum_{t=2}^{n}\left(\tilde{\beta}_1^2(x_t-\bar{x}_n)^2+\tilde{\beta}_1^2(e_t-\bar{e}_n)^2-2\tilde{\beta}_1^2(x_t-\bar{x}_n)^2+\tilde{\beta}_1^2(x_t-\bar{x}_n)^2+(u_t-\bar{u}_n)^2\right)}{(n-1)\left(1+\tilde{\beta}_1^2\right)}$$

$$= plim \frac{\sum_{t=2}^{n}\left(\tilde{\beta}_1^2(e_t-\bar{e}_n)^2+(u_t-\bar{u}_n)^2\right)}{(n-1)\left(1+\tilde{\beta}_1^2\right)} = \frac{\beta_1^2\sigma^2+\sigma^2}{\left(1+\beta_1^2\right)} = \sigma^2$$

Then, the MLE for the variance of covariate x_t and error terms are consistent estimators. From the above results, all the MLE estimators under the null hypothesis of $H_0: \rho = 0$ in a structural model are consistent. Now, we consider the MLE's consistency under the null hypothesis of $H_0: \beta_0 = b_0, \beta_1 = b_1$. To get the MLE in this scenario, we have solve the two equations (2.17C'') and (2.17D'') numerically as shown in Section 2.3.1.3. Therefore, to see the consistency of $\dot{\rho}$ and $\dot{\alpha}_0$, we have to look into those two equations simultaneously. Then (4.3) and (4.4) as shown below should be satisfied for any pair of $\dot{\rho}$ and $\dot{\alpha}_0$.

$$plim \left\{ \begin{array}{c} \frac{1-\dot{\rho}^2\dot{\alpha}_0^2}{(1-\dot{\rho}^2)\dot{\alpha}_0} \frac{\sum_{t=2}^{n}(\dot{F}_t-\dot{\rho}\dot{F}_{t-1})^2}{(n-1)} - \frac{\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\dot{\rho}\dot{\alpha}_0\tilde{\dot{E}}_{t-1}\right)^2}{(n-1)} \\ + \frac{4\dot{\rho}\dot{\alpha}_0\left(1-\dot{\rho}^2\dot{\alpha}_0^2\right)}{1+\dot{\rho}^2\dot{\alpha}_0^2} \frac{\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\dot{\rho}\dot{\alpha}_0\tilde{\dot{E}}_{t-1}\right)\dot{E}_{t-1}}{(n-1)} \end{array} \right\} = 0 \qquad (4.3)$$

$$plim \left\{ \frac{\sum_{t=2}^{n}(\dot{F}_t-\dot{\rho}\dot{F}_{t-1})\dot{F}_{t-1}}{(n-1)} + \dot{\alpha}_0^2 \frac{(1+\dot{\rho}^2)}{1+\dot{\rho}^2\dot{\alpha}_0^2} \frac{\sum_{t=2}^{n}\left(\tilde{\dot{E}}_t-\dot{\rho}\dot{\alpha}_0\tilde{\dot{E}}_{t-1}\right)\dot{E}_{t-1}}{(n-1)} \right\} = 0 \qquad (4.4)$$

First, we consider each term in the above equations separately. We know that e_t and u_t are independent and both of them are from first order autoregressive process. Then

$$plim \frac{\sum_{t=2}^{n} e_{t-1}^2}{(n-1)} = plim \frac{\sum_{t=2}^{n} u_{t-1}^2}{(n-1)} = \frac{\sigma^2}{1-\rho^2},$$

$$plim \frac{\sum_{t=2}^{n} v_t^2}{(n-1)} = plim \frac{\sum_{t=2}^{n} \delta_t^2}{(n-1)} = \sigma^2.$$

Suppose that

$$plim\dot{\rho} = \rho_1 \text{ and } plim\dot{\alpha}_0 = \alpha_{01}.$$

From the model setup, v_t and e_t are independent and δ_t and u_{t-1} are independent. Thus, for the first summation term in (2.17C''), we have

$$plim\frac{\sum_{t=2}^{n}(\dot{F}_t-\dot{\rho}\dot{F}_{t-1})^2}{(n-1)} = plim\frac{\sum_{t=2}^{n}\big(b_1X_t-(Y_t-b_0)-\dot{\rho}b_1X_{t-1}+\dot{\rho}(Y_{t-1}-b_0)\big)^2}{(n-1)}$$

$$= plim\frac{\sum_{t=2}^{n}(b_1e_t-u_t-\dot{\rho}b_1e_{t-1}+\dot{\rho}u_{t-1})^2}{(n-1)}$$

$$= plim\frac{\sum_{t=2}^{n}(b_1(\rho-\dot{\rho})e_{t-1}+b_1v_t-(\rho-\dot{\rho})u_{t-1}-\delta_t)^2}{(n-1)}$$

$$= plim\frac{\sum_{t=2}^{n}\big(b_1^2(\rho-\dot{\rho})^2e_{t-1}^2+b_1^2v_t^2+(\rho-\dot{\rho})^2u_{t-1}^2+\delta_t^2\big)}{(n-1)}$$

$$= (1+b_1^2)(\rho-\rho_1)^2\frac{\sigma^2}{1-\rho^2} + (1+b_1^2)\sigma^2$$

$$= (1+b_1^2)D_0\big((\rho-\rho_1)^2+(1-\rho^2)\big)\alpha_0$$

where $A=\frac{\sigma^2}{1-\rho^2}$, $D_0=(1+b_1^2)\sigma_x^2+A$, and $\alpha_0=\frac{A}{D_0}$.

For the second summation term in (2.17C''), recall that for $t=1,2,\ldots,n$

$$\tilde{\dot{E}}_t = \dot{E}_t - \bar{\dot{E}}_n = X_t - \bar{X}_n + b_1(Y_t-\bar{Y}_n)$$

$$= (1+b_1^2)(x_t-\bar{x}_t) + (e_t-\bar{e}_t) + b_1(u_t-\bar{u}_t)$$

$$= (1+b_1^2)\tilde{x}_t + \tilde{e}_t + b_1\tilde{u}_t.$$

Then

$$plim\frac{\sum_{t=2}^{n}\big(\tilde{\dot{E}}_t-\dot{\rho}\dot{\alpha}_0\tilde{\dot{E}}_{t-1}\big)^2}{(n-1)} = plim\frac{\sum_{t=2}^{n}\big((1+b_1^2)\tilde{x}_t+\tilde{e}_t+b_1\tilde{u}_t-\dot{\rho}\dot{\alpha}_0(1+b_1^2)\tilde{x}_{t-1}-\dot{\rho}\dot{\alpha}_0\tilde{e}_{t-1}-\dot{\rho}\dot{\alpha}_0b_1\tilde{u}_{t-1}\big)^2}{(n-1)}$$

$$= plim\frac{\sum_{t=2}^{n}\big((1+b_1^2)(\tilde{x}_t-\dot{\rho}\dot{\alpha}_0\tilde{x}_{t-1})+\tilde{v}_t+b_1\tilde{\delta}_t+(\rho-\dot{\rho}\dot{\alpha}_0)(\tilde{e}_{t-1}+b_1\tilde{u}_{t-1})\big)^2}{(n-1)}$$

$$= (1+b_1^2)\left[(1+b_1^2)\sigma_x^2(1+\rho_1^2\alpha_{01}^2)+\sigma^2+(\rho-\rho_1\alpha_{01})^2\frac{\sigma^2}{1-\rho^2}\right]$$

$$= (1+b_1^2)D_0\big((1+\rho_1^2\alpha_{01}^2)-2\rho\alpha_0\rho_1\alpha_{01}\big).$$

Similarly, for the third term in equation (2.17C''),

$$plim\frac{\sum_{t=2}^{n}(\tilde{\dot{E}}_t-\dot{\rho}\dot{\alpha}_0\tilde{\dot{E}}_{t-1})\dot{E}_{t-1}}{(n-1)} = plim\frac{\sum_{t=2}^{n}\left((1+b_1^2)^2(\tilde{x}_t-\dot{\rho}\dot{\alpha}_0\tilde{x}_{t-1})\tilde{x}_{t-1}+(\rho-\dot{\rho}\dot{\alpha}_0)(\tilde{e}_{t-1}^2+b_1^2\tilde{u}_{t-1}^2)\right)}{(n-1)}$$

$$= (1+b_1^2)D_0(-\rho_1\alpha_{01}+\rho\alpha_0).$$

The last one is the first term in (2.17D'') which is

$$plim\frac{\sum_{t=2}^{n}(\dot{F}_t-\dot{\rho}\dot{F}_{t-1})\dot{F}_{t-1}}{(n-1)} = plim\frac{\sum_{t=2}^{n}(b_1(\rho-\dot{\rho})e_{t-1}+b_1v_t-(\rho-\dot{\rho})u_{t-1}-\delta_t)(b_1e_{t-1}-u_{t-1})}{(n-1)}$$

$$= (1+b_1^2)D_0(\rho-\rho_1)\alpha_0.$$

Using the results of those four terms, equations (4.3) and (4.4) become

$$\left\{\begin{matrix}\frac{1-\rho_1^2\alpha_{01}^2}{(1-\rho_1^2)\alpha_{01}}\left((\rho-\rho_1)^2+(1-\rho^2)\right)\alpha_0-(1+\rho_1^2\alpha_{01}^2) \\ +2\rho\alpha_0\rho_1\alpha_{01}+4\rho_1\alpha_{01}\frac{1-\rho_1^2\alpha_{01}^2}{1+\rho_1^2\alpha_{01}^2}(-\rho_1\alpha_{01}+\rho\alpha_0)\end{matrix}\right\} = 0, \tag{4.5}$$

$$(\rho-\rho_1)\alpha_0+\frac{(\alpha_{01}^2+\rho_1^2\alpha_{01}^2)}{1+\rho_1^2\alpha_{01}^2}(-\rho_1\alpha_{01}+\rho\alpha_0) = 0. \tag{4.6}$$

Now we need to show that $\rho_1 = \rho$ and $\alpha_{01} = \alpha_0$. Since $-1 < \rho, \rho_1 < 1$, and $0 < \alpha_0, \alpha_{01} < 1$, then

$$0 < \frac{(\alpha_{01}^2+\rho_1^2\alpha_{01}^2)}{1+\rho_1^2\alpha_{01}^2} < 1.$$

Assume $\rho_1 \neq \rho$ and $\alpha_{01} \neq \alpha_0$, we will show there is a contradiction. From (4.6) we have

$$0 < \frac{(\alpha_{01}^2+\rho_1^2\alpha_{01}^2)}{1+\rho_1^2\alpha_{01}^2} = \frac{\rho\alpha_0-\rho_1\alpha_0}{\rho_1\alpha_{01}-\rho\alpha_0} < 1.$$

If $\rho\rho_1 < 0$, then there is no solution. If $0 \leq \rho_1 < \rho$, then $\rho\alpha_0 < \rho_1\alpha_{01}$, $\alpha_0 < \alpha_{01}$ and $2\rho\alpha_0 < \rho_1\alpha_0 + \rho_1\alpha_{01}$. If $\rho_1 < \rho < 0$, then $\rho\alpha_0 < \rho_1\alpha_{01}$, $\alpha_0 > \alpha_{01}$ and $2\rho\alpha_0 < \rho_1\alpha_0 + \rho_1\alpha_{01}$. If $0 \leq \rho < \rho_1$, then $\rho\alpha_0 > \rho_1\alpha_{01}$, $\alpha_0 > \alpha_{01}$ and $2\rho\alpha_0 > \rho_1\alpha_0 + \rho_1\alpha_{01}$.

If $\rho < \rho_1 < 0$, then $\rho\alpha_0 > \rho_1\alpha_{01}$, $\alpha_0 < \alpha_{01}$ and $2\rho\alpha_0 > \rho_1\alpha_0 + \rho_1\alpha_{01}$. Consider the first case where $0 \leq \rho_1 < \rho$, the left hand side of (4.5) becomes

$$\left\{\begin{matrix} \frac{1-\rho_1^2\alpha_{01}^2}{(1-\rho_1^2)\alpha_{01}}(1-2\rho_1\rho+\rho_1^2)\alpha_0 - (1+\rho_1^2\alpha_{01}^2) \\ +2\rho\alpha_0\rho_1\alpha_{01} + 4\rho_1\alpha_{01}\frac{1-\rho_1^2\alpha_{01}^2}{1+\rho_1^2\alpha_{01}^2}(-\rho_1\alpha_{01}+\rho\alpha_0) \end{matrix}\right\}$$

$$< \frac{1-\rho_1^2\alpha_{01}^2}{(1-\rho_1^2)\alpha_{01}}(1-2\rho_1\rho+\rho_1^2)\alpha_0 - (1+\rho_1^2\alpha_{01}^2) + 2\rho\alpha_0\rho_1\alpha_{01}$$

$$< \frac{1-\rho_1^2\alpha_{01}^2}{(1-\rho_1^2)\alpha_{01}}(1-2\rho_1^2+\rho_1^2)\alpha_{01} - (1+\rho_1^2\alpha_{01}^2) + 2\rho_1^2\alpha_{01}^2 = 0$$

Thus, the left hand side of (4.5) is less than zero which contradicts the right hand side of (4.5) which equals 0. This holds for the other three situations. Therefore, the assumption of $\rho_1 \neq \rho$ and $\alpha_{01} \neq \alpha_0$ is not correct. Thus, $\rho_1 = \rho$, or $\alpha_{01} = \alpha_0$, or $\rho_1 = \rho$ and $\alpha_{01} = \alpha_0$. From (4.6), it is obvious that if $\rho_1 = \rho$ then $\alpha_{01} = \alpha_0$ and vice versa. So we can conclude that $\rho_1 = \rho$ and $\alpha_{01} = \alpha_0$. This means that MLE $\dot{\rho}$ and $\dot{\alpha}_0$ are consistent estimators. Also,

$$\text{plim}\dot{D}_0 = \text{plim}\frac{1}{(1+b_1^2)}\left\{\frac{1}{1-\dot{\rho}^2\dot{\alpha}_0^2}\frac{\sum_{t=2}^{n}\left(\tilde{\bar{E}}_t-\dot{\rho}\dot{\alpha}_0\tilde{\bar{E}}_{t-1}\right)^2}{(n-1)} - \frac{2\dot{\rho}\dot{\alpha}_0}{1+\dot{\rho}^2\dot{\alpha}_0^2}\frac{\sum_{t=2}^{n}(\tilde{\bar{E}}_t-\dot{\rho}\dot{\alpha}_0\tilde{\bar{E}}_{t-1})\dot{E}_{t-1}}{(n-1)}\right\}$$

$$= \frac{1}{(1+b_1^2)}\left\{\frac{1}{1-\rho^2\alpha_0^2}\left((1+b_1^2)D_0(1-\rho^2\alpha_0^2)\right)\right\} = D_0.$$

Then

$$\text{plim}\dot{\sigma}^2 = \text{plim}(1-\dot{\rho}^2)\dot{D}_0\dot{\alpha}_0 = \sigma^2,$$

$$\text{plim}\dot{\sigma}_x^2 = \text{plim}\frac{1-\dot{\alpha}_0}{(1+b_1^2)}\dot{D}_0 = \sigma_x^2,$$

$$\text{plim}\dot{\mu}_x = \text{plim}\frac{(\bar{E}_n-\dot{\rho}\dot{\alpha}_0\bar{E}_{n-1})}{(1+b_1^2)(1-\dot{\rho}\dot{\alpha}_0)} = \frac{1}{(1+b_1^2)(1-\rho\alpha_0)}(1+b_1^2)\mu_x(1-\rho\alpha_0) = \mu_x.$$

Thus, under the null hypothesis of $H_0: \beta_0 = b_0, \beta_1 = b_1$, all the MLEs $(\dot{\mu}_x, \dot{\sigma}_x^2, \dot{\sigma}^2, \dot{\rho})$ are consistent estimators in structural case.

At this point, we have shown that for a structural model the maximum likelihood estimators under the two different null hypotheses are all consistent. In the same manner, the consistency of the MLEs under $H_0: \beta_0 = b_0, \beta_1 = b_1$ can be shown. Also, the consistency of the MLEs without any constraints can be shown as well.

4.1.2 Functional case

In a functional model, $\{x_t: t = 1,2, \dots, n\}$ is a sequence of fixed value with assumptions (2.5) and (2.6). Therefore, it is treated as a sequence of nuisance parameters that must be estimated. One of the consequences of fixed x_t is that the usual MLE may not be consistent. Now, let us consider the MLEs under $H_0: \rho = 0$ and $H_0: \beta_0 = b_0, \beta_1 = b_1$ respectively.

The MLEs under the null hypothesis of $H_0: \rho = 0$ were derived in section 2.3.2.1. We notice that

$$(S_{XX} - S_{YY})/(n-1) = \sum_{t=2}^{n}(X_t - \bar{X})^2/(n-1) - \sum_{t=2}^{n}(Y_t - \bar{Y})^2/(n-1)$$

$$= \frac{\sum_{t=2}^{n} X_t^2 - \bar{X}^2 - \sum_{t=2}^{n} Y_t^2 + \bar{Y}^2}{(n-1)}$$

$$= \frac{\sum_{t=2}^{n}\left((x_t+e_t)^2 - (\bar{x}_n+\bar{e}_n)^2\right) - \sum_{t=2}^{n}\left((y_t+u_t)^2 - (\bar{y}_n+\bar{u}_n)^2\right)}{(n-1)}.$$

Since e_t and u_t are from normal distribution with mean 0 and variance σ^2 under $H_0: \rho = 0$ and $y_t = \beta_0 + \beta_1 x_t$, then

$$p\text{lim}\,(S_{XX} - S_{YY})/(n-1) = \frac{\sum_{t=2}^{n} x_t^2}{(n-1)} + \sigma^2 - \bar{x}_n^2 - \frac{\sum_{t=2}^{n} y_t^2}{(n-1)} - \sigma^2 + \bar{y}_n^2$$

$$= \frac{\sum_{t=2}^{n}(x_t-\bar{x}_n)^2}{(n-1)} - \frac{\sum_{t=2}^{n}(y_t-\bar{y}_n)^2}{(n-1)}$$

$$= (1-\beta_1^2)\text{plim}\frac{\sum_{t=2}^{n}(x_t-\bar{x}_n)^2}{(n-1)}.$$

Similarly,

$$S_{XY}/(n-1) = \frac{\sum_{t=2}^{n}(X_t-\bar{X})(Y_t-\bar{Y})}{(n-1)} = \frac{\sum_{t=2}^{n}(x_t+e_t)(y_t+u_t)}{(n-1)} - (\bar{x}_n+\bar{e}_n)(\bar{y}_n+\bar{u}_n),$$

and

$$p\text{lim}\, S_{XY}/(n-1) = p\text{lim}\left\{\frac{\sum_{t=2}^{n}x_t y_t}{(n-1)} - \bar{x}_n\bar{y}_n\right\} = p\text{lim}\beta_1\frac{\sum_{t=2}^{n}(x_t-\bar{x}_n)^2}{(n-1)}.$$

Therefore, the MLE for β_1 is a consistent estimator as shown below.

$$p\text{lim}\tilde{\beta}_1 = p\text{lim}\frac{-(S_{XX}-S_{YY})+\sqrt{(S_{XX}-S_{YY})^2+4S_{XY}^2}}{2S_{XY}}$$

$$= p\text{lim}\frac{-\frac{(1-\beta_1^2)\sum_{t=2}^{n}(x_t-\bar{x}_n)^2}{(n-1)}+\sqrt{\left[\frac{(1-\beta_1^2)\sum_{t=2}^{n}(x_t-\bar{x}_n)^2}{(n-1)}\right]^2+4\left[\frac{\beta_1\sum_{t=2}^{n}(x_t-\bar{x}_n)^2}{(n-1)}\right]^2}}{2\beta_1\frac{\sum_{t=2}^{n}(x_t-\bar{x}_n)^2}{(n-1)}}$$

$$= \frac{-(1-\beta_1^2)+\sqrt{(1-\beta_1^2)^2+4\beta_1^2}}{2\beta_1} = \beta_1$$

Following the consistency of the estimator for β_1, the consistency of $\tilde{\beta}_0$ can be shown as

$$p\text{lim}\tilde{\beta}_0 = p\text{lim}(\bar{Y}-\tilde{\beta}_1\bar{X}) = p\text{lim}(\beta_0+\beta_1\bar{x}_n-\beta_1\bar{x}_n) = \beta_0.$$

For the nuisance parameters, we have

$$p\text{lim}\tilde{x}_t = p\text{lim}\frac{(Y_t-\tilde{\beta}_0)\tilde{\beta}_1+X_t}{1+\tilde{\beta}_1^2} = x_t + \frac{\beta_1 u_t+e_t}{1+\beta_1^2} \qquad t=1,2,\dots,n.$$

Thus, under the null hypothesis the MLE $(\tilde{x}_1,\tilde{x}_2,\dots,\tilde{x}_n,\tilde{\beta}_0,\tilde{\beta}_1)$ are consistent estimators.

However, the estimator for σ^2 is inconsistent because

$$p\text{lim}\tilde{\sigma}^2 = p\text{lim}\frac{1}{2(n-1)}\sum_{t=2}^{n}\{(X_t-\tilde{x}_t)^2+(Y_t-\tilde{y}_t)^2\}$$

$$= plim \frac{\sum_{t=2}^{n}\left\{\left(x_t+e_t-x_t-\frac{\beta_1 u_t+e_t}{1+\beta_1^2}\right)^2+\left(\beta_1 x_t+u_t-\beta_1 x_t-\beta_1\frac{\beta_1 u_t+e_t}{1+\beta_1^2}\right)^2\right\}}{2(n-1)}$$

$$= plim \frac{\sum_{t=2}^{n}\left\{\left(e_t-\frac{\beta_1 u_t+e_t}{1+\beta_1^2}\right)^2+\left(u_t-\beta_1\frac{\beta_1 u_t+e_t}{1+\beta_1^2}\right)^2\right\}}{2(n-1)}$$

$$= plim \frac{\sum_{t=2}^{n}\left\{\left(\frac{\beta_1^2}{1+\beta_1^2}e_t-\frac{\beta_1}{1+\beta_1^2}u_t\right)^2+\left(\frac{1}{1+\beta_1^2}u_t-\frac{\beta_1}{1+\beta_1^2}e_t\right)^2\right\}}{2(n-1)}$$

$$= plim \frac{\frac{1}{(1+\beta_1^2)}\sum_{t=2}^{n}(\beta_1 e_t-u_t)^2}{2(n-1)} = \frac{\sigma^2}{2}.$$

The MLE of σ^2 is a consistent estimator for $\sigma^2/2$. For the case $H_0: (\beta_0, \beta_1) = (b_0, b_1)$, the inconsistency problem will be shown to be the same as the above. In section 2.3.2.3, the maximum likelihood estimators are derived for all parameters including the nuisance parameters. For the MLE of ρ, we have

$$plim\ddot{\rho} = plim \frac{\sum_{t=2}^{n}(b_1 X_t-Y_t+b_0)(b_1 X_{t-1}-Y_{t-1}+b_0)}{\sum_{t=2}^{n}(b_1 X_{t-1}-Y_{t-1}+b_0)^2}$$

$$= plim \frac{\sum_{t=2}^{n}(b_1 e_t-u_t)(b_1 e_{t-1}-u_{t-1})}{\sum_{t=2}^{n}(b_1 e_{t-1}-u_{t-1})^2}$$

$$= plim \frac{\sum_{t=2}^{n}\left(b_1^2 e_t e_{t-1}+u_t u_{t-1}\right)}{\sum_{t=2}^{n}\left(b_1^2 e_{t-1}^2+u_{t-1}^2\right)^2}$$

$$= plim\left\{\rho+\frac{\sum_{t=2}^{n}\left(b_1^2 e_{t-1}v_t+u_{t-1}\delta_t\right)}{\sum_{t=2}^{n}\left(b_1^2 e_{t-1}^2+u_{t-1}^2\right)^2}\right\} = \rho.$$

The second equality is because the independence of u_t and e_t for any $t = 1,2,\dots,n$. The last equality is because e_{t-1} and v_t are independent, and also u_{t-1} is independent with δ_t. Similarly, the limit for nuisance parameters would be

$$plim\ddot{\lambda}_t = plim \frac{(X_t-\ddot{\rho}X_{t-1})+b_1(Y_t-\ddot{\rho}Y_{t-1})-b_0 b_1(1-\ddot{\rho})}{1+b_1^2}$$

$$= p\lim \frac{(x_t - \rho x_{t-1} + v_t) + b_1^2 (x_t - \rho x_{t-1}) + b_1 \delta_t}{1 + b_1^2}$$

$$= \lambda_t + \frac{v_t + b_1 \delta_t}{1 + b_1^2} \qquad t = 1,2, \ldots, n.$$

Then with these two results, the inconsistency of $\ddot{\sigma}^2$ can be shown as

$$p\lim \ddot{\sigma}^2 = p\lim \frac{1}{2(n-1)} \sum_{t=2}^{n} \left\{ \left(X_t - \ddot{\lambda}_t - \ddot{\rho} X_{t-1}\right)^2 + (Y_t - \ddot{\gamma}_t - \ddot{\rho} Y_{t-1})^2 \right\}$$

$$= p\lim \frac{1}{2(n-1)} \sum_{t=2}^{n} \left\{ \left(v_t - \frac{v_t + b_1 \delta_t}{1 + b_1^2}\right)^2 + \left(\delta_t - b_1 \frac{v_t + b_1 \delta_t}{1 + b_1^2}\right)^2 \right\}$$

$$= p\lim \frac{1}{2(n-1)\left(1 + b_1^2\right)^2} \sum_{t=2}^{n} \{ b_1^2 (\delta_t - b_1 v_t)^2 + (\delta_t - b_1 v_t)^2 \}$$

$$= \frac{1}{2\left(1 + b_1^2\right)} (\sigma^2 + b_1^2 \sigma^2) = \frac{\sigma^2}{2}.$$

Thus, we have shown that $\ddot{\sigma}^2$ is an inconsistent MLE estimator for σ^2 but the MLEs for the other parameters are consistent.

The maximum likelihood estimators for ρ and β_1 without any constraint are obtained by solving an equation system iteratively. Thus,

$$p\lim \hat{\rho} = p\lim \frac{\sum_{t=2}^{n} \{ (\hat{\beta}_1 X_t - Y_t) - (\hat{\beta}_1 \bar{X}_n - \bar{Y}_n) \} \{ (\hat{\beta}_1 X_{t-1} - Y_{t-1}) - (\hat{\beta}_1 \bar{X}_{n-1} - \bar{Y}_{n-1}) \}}{\sum_{t=2}^{n} \{ (\hat{\beta}_1 X_{t-1} - Y_{t-1}) - (\hat{\beta}_1 \bar{X}_{n-1} - \bar{Y}_{n-1}) \}^2} \stackrel{\text{def}}{=} \rho_0,$$

$$p\lim \hat{\beta}_1 = p\lim \frac{-(\ddot{S}_{XX} - \ddot{S}_{YY}) + \sqrt{(\ddot{S}_{XX} - \ddot{S}_{YY})^2 + 4 \ddot{S}_{XY}^2}}{2 \ddot{S}_{XY}} \stackrel{\text{def}}{=} \beta_{10}.$$

Because the terms in $\hat{\beta}_1$ have limits as shown below,

$$p\lim \hat{S}_{XX}/(n-1) = p\lim \frac{\sum_{t=2}^{n} \{ (X_t - \hat{\rho} X_{t-1}) - (\bar{X}_n - \hat{\rho} \bar{X}_{n-1}) \}^2}{(n-1)}$$

$$= p\lim \frac{\sum_{t=2}^{n} (\tilde{x}_t + \tilde{e}_t - \rho_0 \tilde{x}_{t-1} - \rho_0 \tilde{e}_{t-1})^2}{(n-1)}$$

$$= (1 + \rho_0^2) \sigma_x^2 + (1 + \rho_0^2) \frac{\sigma^2}{1 - \rho^2} - 2 \rho_0 \rho \frac{\sigma^2}{1 - \rho^2},$$

and

$$plim\,\hat{S}_{YY}/(n-1) = plim\frac{\sum_{t=2}^{n}\{(Y_t-\hat{\rho}Y_{t-1})-(\bar{Y}_n-\hat{\rho}\bar{Y}_{n-1})\}^2}{(n-1)}$$

$$= \beta_1^2(1+\rho_0^2)\sigma_x^2 + (1+\rho_0^2)\frac{\sigma^2}{1-\rho^2} - 2\rho_0\rho\frac{\sigma^2}{1-\rho^2},$$

and also,

$$plim\,\hat{S}_{XY}/(n-1) = plim\frac{\sum_{t=2}^{n}\{(X_t-\hat{\rho}X_{t-1})-(\bar{X}_n-\hat{\rho}\bar{X}_{n-1})\}\{(Y_t-\hat{\rho}Y_{t-1})-(\bar{Y}_n-\hat{\rho}\bar{Y}_{n-1})\}}{(n-1)}$$

$$= \beta_1(1+\rho_0^2)\sigma_x^2,$$

then $\hat{\beta}_1$ is a consistent estimator for β_1,

$$plim\hat{\beta}_1 = plim\frac{-(\ddot{S}_{XX}-\ddot{S}_{YY})+\sqrt{(\ddot{S}_{XX}-\ddot{S}_{YY})^2+4\ddot{S}_{XY}^2}}{2\ddot{S}_{XY}} = \beta_1.$$

Furthermore, the limits of terms in $\hat{\rho}$ are,

$$plim\frac{\sum_{t=2}^{n}\{(\hat{\beta}_1X_t-Y_t)-(\hat{\beta}_1\bar{X}_n-\bar{Y}_n)\}\{(\hat{\beta}_1X_{t-1}-Y_{t-1})-(\hat{\beta}_1\bar{X}_{n-1}-\bar{Y}_{n-1})\}}{(n-1)}$$

$$= plim\frac{\sum_{t=2}^{n}(\beta_1\tilde{e}_t-\tilde{u}_t)(\beta_1\tilde{e}_{t-1}-\tilde{u}_{t-1})}{(n-1)} = (1+\beta_1^2)\rho\frac{\sigma^2}{1-\rho^2},$$

and

$$plim\frac{\sum_{t=2}^{n}\{(\hat{\beta}_1X_{t-1}-Y_{t-1})-(\hat{\beta}_1\bar{X}_{n-1}-\bar{Y}_{n-1})\}^2}{(n-1)} = plim\frac{\sum_{t=2}^{n}(\beta_1\tilde{e}_{t-1}-\tilde{u}_{t-1})^2}{(n-1)} = (1+\beta_1^2)\frac{\sigma^2}{1-\rho^2}.$$

Therefore,

$$plim\hat{\rho} = plim\frac{\sum_{t=2}^{n}\{(\hat{\beta}_1X_t-Y_t)-(\hat{\beta}_1\bar{X}_n-\bar{Y}_n)\}\{(\hat{\beta}_1X_{t-1}-Y_{t-1})-(\hat{\beta}_1\bar{X}_{n-1}-\bar{Y}_{n-1})\}}{\sum_{t=2}^{n}\{(\hat{\beta}_1X_{t-1}-Y_{t-1})-(\hat{\beta}_1\bar{X}_{n-1}-\bar{Y}_{n-1})\}^2} = \rho.$$

That is, the estimator for ρ is also consistent. Because of the consistency of $\hat{\rho}$ and $\hat{\beta}_1$, we have

$$plim\hat{\beta}_0 = plim\frac{(\bar{Y}_n-\hat{\rho}\bar{Y}_{n-1})-\hat{\beta}_1(\bar{X}_n-\hat{\rho}\bar{X}_{n-1})}{1-\hat{\rho}} = \frac{(\beta_0+\beta_1\bar{x}_n-\rho\beta_0-\rho\beta_1\bar{x}_{n-1})-\beta_1(\bar{x}_n-\rho\bar{x}_{n-1})}{1-\rho} = \beta_0,$$

$$\text{plim}\hat{\lambda}_t = \text{plim}\frac{(X_t-\hat{\rho}X_{t-1})+\hat{\beta}_1(Y_t-\hat{\rho}Y_{t-1})-\hat{\beta}_0\hat{\beta}_1(1-\hat{\rho})}{1+\hat{\beta}_1^2} = x_t-\rho x_{t-1}+\frac{\upsilon_t+\beta_1\delta_t}{1+\beta_1^2} \quad t=2,3,\dots,n$$

$$\text{plim}\hat{\sigma}^2 = \text{plim}\frac{1}{2(n-1)}\sum_{t=2}^{n}\left\{\left(X_t-\hat{\lambda}_t-\hat{\rho}X_{t-1}\right)^2+(Y_t-\hat{\gamma}_t-\hat{\rho}Y_{t-1})^2\right\}$$

$$= \text{plim}\frac{1}{2(n-1)}\sum_{t=2}^{n}\left\{\left(\upsilon_t-\frac{\upsilon_t+\beta_1\delta_t}{1+\beta_1^2}\right)^2+\left(\delta_t-\beta_1\frac{\upsilon_t+\beta_1\delta_t}{1+\beta_1^2}\right)^2\right\}=\frac{\sigma^2}{2}.$$

Therefore, all MLE parameter estimators are consistent except $\hat{\sigma}^2$ which converges to $\sigma^2/2$ instead of σ^2.

4.2 Property of Test Statistics for Structural Model

4.2.1 Efficient score test statistic under H_0: $\rho = 0$

In section 4.1, it was shown that all the MLE estimators are consistent estimators. In this section, the large sample distribution for the efficient score test statistic under the null hypothesis will be derived. Because of the consistency of each estimator, we have

$$\frac{1}{\sqrt{n-1}}\text{plim}\left.\frac{\partial\ell_s}{\partial\rho}\right|_{\rho=0} = \frac{\sigma^2}{(1+\beta_1^2)\left[(1+\beta_1^2)\sigma_x^2+\sigma^2\right]^2}\frac{1}{\sqrt{n-1}}\text{plim}\sum_{t=2}^{n}\left(\tilde{X}_t+\tilde{\beta}_1\tilde{Y}_t\right)\left(\tilde{X}_{t-1}+\tilde{\beta}_1\tilde{Y}_{t-1}\right)$$

$$+\frac{1}{(1+\beta_1^2)\sigma^2}\text{plim}\frac{\sum_{t=2}^{n}[\beta_1X_t-(Y_t-\beta_0)][\beta_1X_{t-1}-(Y_{t-1}-\beta_0)]}{\sqrt{n-1}}$$

$$=\text{plim}\frac{\sigma^2\sum_{t=2}^{n}\left[(1+\beta_1^2)(x_t-\mu_x)+e_t+\beta_1u_t\right]\left[(1+\beta_1^2)(x_{t-1}-\mu_x)+e_{t-1}+\beta_1u_{t-1}\right]}{(1+\beta_1^2)\left[(1+\beta_1^2)\sigma_x^2+\sigma^2\right]^2\sqrt{n-1}}$$

$$+\text{plim}\frac{1}{(1+\beta_1^2)\sigma^2}\frac{\sum_{t=2}^{n}\left(\beta_1^2e_te_{t-1}-\beta_1e_{t-1}u_t-\beta_1e_tu_{t-1}+u_tu_{t-1}\right)}{\sqrt{n-1}}.$$

Let

$$W_{s1t} = \frac{\left[(1+\beta_1^2)(x_t-\mu_x)+e_t+\beta_1u_t\right]\left[(1+\beta_1^2)(x_{t-1}-\mu_x)+e_{t-1}+\beta_1u_{t-1}\right]}{\sigma^{-2}(1+\beta_1^2)D_1^2}+\frac{(\beta_1e_t-u_t)(\beta_1e_{t-1}-u_{t-1})}{(1+\beta_1^2)\sigma^2}.$$

where $D_1 = (1+\beta_1^2)\sigma_x^2+\sigma^2$.

Then,

$$\frac{1}{\sqrt{n-1}} \text{plim} \frac{\partial \ell_s}{\partial \rho}\bigg|_{\rho=0} = \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} W_{s1t}.$$

It is obvious that the random variable W_{s1t} is 1-dependent. Further, notice that the expectation of W_{s1t} is

$$E(W_{s1t}) = 0,$$

and the variance of W_{s1t} is

$$Var(W_{s1t}) = 1 + \frac{\sigma^4}{D_1^2}.$$

Then the covariance of W_{s12} and W_{s13} is,

$$cov(W_{s12}, W_{s13}) = 0.$$

Thus, according to Theorem 3.5 in section 3, we have

$$\frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} W_{s1t} \xrightarrow{d} N\left(0,1 + \frac{\sigma^4}{D_0^2}\right).$$

Therefore, the score element with respect to ρ is asymptotically normally distributed with mean 0 and variance $1 + \frac{\sigma^4}{D_1^2}$.

Because all the estimators are consistent, the efficient score test statistic for testing $\rho = 0$ will satisfy

$$plim T_{ES,s,\rho} = \left(1 + \frac{\sigma^4}{D_1^2}\right)^{-1} \left(\frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} W_{s1t}\right)^2,$$

which is the square of an asymptotically normally distributed random variable with mean zero and variance one. Therefore,

$$T_{ES,s,\rho} \xrightarrow{d} \chi_1^2.$$

That is, the distribution of the efficient score test statistic under the null hypothesis of $\rho = 0$ converges to a chi-squared distribution with one degree of freedom.

4.2.2 Likelihood ratio test statistic under H_0: $\rho = 0$

There is no explicit expression for the MLEs in this case. The Taylor expansion will be useful in determining the large sample distribution for the test statistic. To simplify the notation, we need to define some variables first. Let

$$\hat{\psi} = (\hat{\mu}_X, \hat{\mu}_Y, \hat{\mu}_{X^2}, \hat{\mu}_{Y^2}, \hat{\mu}_{XY}, \hat{\mu}_{XY1}, \hat{\mu}_{X1Y}, \hat{\mu}_{XX}, \hat{\mu}_{YY})$$

where

$$\hat{\mu}_X = \bar{X}, \hat{\mu}_Y = \bar{Y}, \hat{\mu}_{X^2} = \frac{1}{n-1}\sum_{t=2}^{n}(X_t - \bar{X})^2, \hat{\mu}_{Y^2} = \frac{1}{n-1}\sum_{t=2}^{n}(Y_t - \bar{Y})^2,$$

$$\hat{\mu}_{XY} = \frac{1}{n-1}\sum_{t=2}^{n}(X_t - \bar{X})(Y_t - \bar{Y}),$$

$$\hat{\mu}_{XY1} = \frac{1}{n-1}\sum_{t=2}^{n}(X_t - \bar{X})(Y_{t-1} - \bar{Y}), \hat{\mu}_{X1Y} = \frac{1}{n-1}\sum_{t=2}^{n}(X_{t-1} - \bar{X})(Y_t - \bar{Y}),$$

$$\hat{\mu}_{XX} = \frac{1}{n-1}\sum_{t=2}^{n}(X_t - \bar{X})(X_{t-1} - \bar{X}), \hat{\mu}_{YY} = \frac{1}{n-1}\sum_{t=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y}).$$

All the estimators are functions of these new variables. Thus, the test statistic is a complicated function of the new variables. Next we expand the test statistic about $\hat{\psi}$. Under the null hypothesis, $\hat{\psi}$ converges to $\psi = (\mu_x, \beta_0 + \beta_1\mu_x, \sigma_x^2 + \sigma^2, \beta_1^2\sigma_x^2 + \sigma^2, \beta_1\sigma_x^2, 0,0,0,0)$. It is obvious that the first order derivatives of the test statistic are all zero.

The second order derivatives of the test statistic can be found using the implicit function theorem and the chain rule because we do not have an explicit expression for each variable. The second order derivatives of the test statistic is generally expressed as,

$$\frac{\partial^2 T_{LRT,s,\rho}}{\partial\hat{\psi}^T\partial\hat{\psi}} = -(n-1)\left(-\frac{1}{|\hat{\Sigma}_s|^2}\frac{\partial|\hat{\Sigma}_s|}{\partial\hat{\psi}^T}\frac{\partial|\hat{\Sigma}_s|}{\partial\hat{\psi}} + \frac{1}{|\tilde{\Sigma}_s|^2}\frac{\partial|\tilde{\Sigma}_s|}{\partial\hat{\psi}^T}\frac{\partial|\tilde{\Sigma}_s|}{\partial\hat{\psi}} + \frac{1}{|\hat{\Sigma}_s|}\frac{\partial^2|\hat{\Sigma}_s|}{\partial\hat{\psi}^T\partial\hat{\psi}} - \frac{1}{|\tilde{\Sigma}_s|}\frac{\partial^2|\tilde{\Sigma}_s|}{\partial\hat{\psi}^T\partial\hat{\psi}}\right).$$

The second order derivatives of $\left|\tilde{\Sigma}_s\right|$ is easy to obtain and the non-zero part of it is given by

$$\frac{\partial^2|\tilde{\Sigma}_s|}{\partial(\hat{\mu}_{X^2},\hat{\mu}_{Y^2},\hat{\mu}_{XY})^T\partial(\hat{\mu}_{X^2},\hat{\mu}_{Y^2},\hat{\mu}_{XY})} = \begin{pmatrix} -\frac{2\beta_1^2}{(1+\beta_1^2)^2}\frac{\sigma^2}{\sigma_x^2} & 2\frac{\beta_1^2}{(1+\beta_1^2)^2}\frac{\sigma^2}{\sigma_x^2}+1 & 0 \\ 2\frac{\beta_1^2}{(1+\beta_1^2)^2}\frac{\sigma^2}{\sigma_x^2}+1 & 0 & 0 \\ 0 & 0 & c_{33} \end{pmatrix}$$

where

$$c_{33} = -\frac{(1-\beta_1^2)^2}{(1+\beta_1^2)^2}\frac{\sigma^2}{\sigma_x^2} - \frac{8\beta_1^2(1-\beta_1^2)}{(1+\beta_1^2)^2} - \frac{2(1-\beta_1^2)^2}{(1+\beta_1^2)^2}.$$

However, the second order derivatives for $\left|\hat{\Sigma}_s\right|$ is very complicated. After we obtain all the derivatives, we know the asymptotical distribution for this test statistic is chi-squared with one degree of freedom. The large sample distribution for the test statistics used to test the intercept and slope can be achieved using the same idea since the central limit theorem works for weakly dependent data.

4.3 Property of Test Statistics for Functional Model

4.3.1 Efficient score test statistic under H_0: $\rho = 0$

Now we are going to determine the large sample distribution of the efficient score test statistic under the null hypothesis. In section 2.3.2.1, the efficient score test statistic for the null hypothesis $H_0: \rho = 0$ is derived as

$$T_{ES,f,\rho} = \frac{1}{2(n-1)}\left(\left.\frac{\partial\ell_f}{\partial\rho}\right|_{\rho=0}\right)^2$$

where $\left.\frac{\partial\ell_f}{\partial\rho}\right|_{\rho=0} = \frac{1}{\tilde{\sigma}^2}\sum_{t=2}^{n}\{(X_t-\tilde{x}_t)(X_{t-1}-\tilde{x}_{t-1})+(Y_t-\tilde{y}_t)(Y_{t-1}-\tilde{y}_{t-1})\}.$

Notice that the MLE estimator for σ^2 is inconsistent and converges to $\sigma^2/2$. This has been shown in section 4.2. Also,

$$p\text{lim}\tilde{x}_t = x_t + \frac{\beta_1 u_t + e_t}{1+\beta_1^2} \qquad t = 1,2,\dots,n.$$

Using the relationships derived in section 4.2, we consider the first derivative with respect to ρ under the null hypothesis. Therefore,

$$p\text{lim}\frac{\partial \ell_f}{\partial \rho}\bigg|_{\rho=0} = p\text{lim}\frac{1}{\tilde{\sigma}^2}\sum_{t=2}^{n}\{(X_t - \tilde{x}_t)(X_{t-1} - \tilde{x}_{t-1}) + (Y_t - \tilde{y}_t)(Y_{t-1} - \tilde{y}_{t-1})\}$$

$$= \frac{2}{\sigma^2}\sum_{t=2}^{n}\left(e_t - \frac{\beta_1 u_t + e_t}{1+\beta_1^2}\right)\left(e_{t-1} - \frac{\beta_1 u_{t-1} + e_{t-1}}{1+\beta_1^2}\right)$$

$$+ \frac{2}{\sigma^2}\sum_{t=2}^{n}\left(u_t - \beta_1\frac{\beta_1 u_t + e_t}{1+\beta_1^2}\right)\left(u_{t-1} - \beta_1\frac{\beta_1 u_{t-1} + e_{t-1}}{1+\beta_1^2}\right)$$

$$= \frac{2\sum_{t=2}^{n}\{(\beta_1^2 e_t - \beta_1 u_t)(\beta_1^2 e_{t-1} - \beta_1 u_{t-1}) + (u_t - \beta_1 e_t)(u_{t-1} - \beta_1 e_{t-1})\}}{(1+\beta_1^2)^2\sigma^2}$$

$$= \frac{2\sum_{t=2}^{n}(\beta_1 e_t - u_t)(\beta_1 e_{t-1} - u_{t-1})}{(1+\beta_1^2)\sigma^2}.$$

From the assumption of the model, e_t and u_t are independent for any $t = 1,2,\dots,n$. And furthermore, under the null hypothesis of $\rho = 0$, we have

$$\binom{e_t}{u_t} \sim i.i.d.\,N(0, \sigma^2 I) \qquad t = 1,2,\dots,n.$$

Therefore,

$$\beta_1 e_t - u_t = (\beta_1, -1)\binom{e_t}{u_t}$$

$$\sim i.i.d.\,N(0, (1+\beta_1^2)\sigma^2) \qquad t = 1,2,\dots,n.$$

Thus, after standardizing,

$$Z_t \stackrel{\text{def}}{=} \frac{\beta_1 e_t - u_t}{\sqrt{(1+\beta_1^2)\sigma^2}} \sim i.i.d.\,N(0,1) \qquad t = 1,2,\dots,n.$$

And the limit of first derivative with respect to ρ becomes,

$$p\lim \frac{\partial \ell_f}{\partial \rho}\bigg|_{\rho=0} = 2\sum_{t=2}^{n} Z_t Z_{t-1}.$$

Let $W_t = Z_t Z_{t-1}$, $t = 2,3,\dots,n$. It is noticed that W_t is a 1-dependent random variable since W_t and W_{t+2} are independent. The expectation and variance of this random variable would be

$$E(W_t) = 0,$$

and

$$Var(W_t) = E(Z_t^2 Z_{t-1}^2) = E(Z_t^2)E(Z_{t-1}^2) = 1.$$

Then by the central limit theorem for m-dependent random variables discussed in Section 3, we have

$$\sqrt{n-1}\frac{\sum_{t=2}^{n} W_t}{n-1} \xrightarrow{d} N(0, \tau_W^2)$$

where $\tau_W^2 = 1 + 2cov(W_2, W_3)$.

Because

$$cov(W_2, W_3) = E(W_2 W_3) - E(W_2)E(W_3) = E(Z_1 Z_2^2 Z_3) = E(Z_1)E(Z_2^2)E(Z_3) = 0,$$

$$\tau_W^2 = 1.$$

The limit of first derivative with respect to ρ can be expressed as

$$p\lim \frac{\partial \ell_f}{\partial \rho}\bigg|_{\rho=0} = 2\sum_{t=2}^{n} W_t = 2\sqrt{n-1}\left(\sqrt{n-1}\frac{\sum_{t=2}^{n} W_t}{n-1}\right).$$

And then the limit of test statistic can be expressed as

$$p\lim T_{ES,f,\rho} = \frac{1}{2(n-1)}\left\{2\sqrt{n-1}\left(\sqrt{n-1}\frac{\sum_{t=2}^{n} W_t}{n-1}\right)\right\}^2 = 2\left\{\left(\sqrt{n-1}\frac{\sum_{t=2}^{n} W_t}{n-1}\right)\right\}^2 \xrightarrow{d} 2\chi_1^2.$$

Thus, the asymptotic distribution for the efficient score test statistic under the null hypothesis $\rho = 0$ is two times the chi-square distribution with one degree of freedom. This phenomenon must be due to the inconsistency of the variance estimator which is the denominator of the test statistic and converges to a half of the true variance.

4.3.2 Likelihood ratio test statistic under H_0: $\rho = 0$

The lack of clear expression for the MLEs causes some difficulty in finding the asymptotic distribution for the likelihood ratio test statistic under this scenario. We are going to use the Taylor expansion to solve this problem similar to the development in section 4.2.

The first step is to take the first order derivative of the test statistic with respect to $\hat{\psi}$ which is defined in section 4.2.2. and evaluate it under the null hypothesis $\rho = 0$. The first order derivatives are

$$\frac{\partial T_{LRT,f,\rho}}{\partial \hat{\psi}} = -2(n-1)\left(\frac{1}{\hat{\sigma}^2}\frac{\partial \hat{\sigma}^2}{\partial \hat{\psi}} - \frac{1}{\tilde{\sigma}^2}\frac{\partial \tilde{\sigma}^2}{\partial \hat{\psi}}\right).$$

Because

$$\tilde{\sigma}^2 = \frac{1}{2(n-1)}\frac{1}{\left(1+\tilde{\beta}_1^2\right)}\sum_{t=2}^{n}\left\{\tilde{\beta}_1^2(X_t - \bar{X})^2 - 2\tilde{\beta}_1(X_t - \bar{X})(Y_t - \bar{Y}) + (Y_t - \bar{Y})^2\right\}$$

$$= \frac{1}{2\left(1+\tilde{\beta}_1^2\right)}\left(\tilde{\beta}_1^2\hat{\mu}_{X^2} - 2\tilde{\beta}_1\hat{\mu}_{XY} + \hat{\mu}_{Y^2}\right),$$

$$\tilde{\beta}_1 = \frac{-\left(\hat{\mu}_{X^2} - \hat{\mu}_{Y^2}\right) + \sqrt{\left(\hat{\mu}_{X^2} - \hat{\mu}_{Y^2}\right)^2 + 4\hat{\mu}_{XY}^2}}{2\hat{\mu}_{XY}},$$

and

$$\frac{\partial \tilde{\beta}_1}{\partial \hat{\psi}} = \left[0,0, \frac{\frac{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})}{\sqrt{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2+4\hat{\mu}_{XY}^2}}-1}{2\hat{\mu}_{XY}}, \frac{1-\frac{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})}{\sqrt{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2+4\hat{\mu}_{XY}^2}}}{2\hat{\mu}_{XY}}, \frac{\frac{-(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2}{\sqrt{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2+4\hat{\mu}_{XY}^2}}+(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})}{2\hat{\mu}_{XY}^2}, 0,0,0,0 \right].$$

Then the second term of the first order derivative can be easily obtained as

$$\frac{\partial \tilde{\sigma}^2}{\partial \hat{\psi}} = \left(0,0, \frac{\partial \tilde{\sigma}^2}{\partial \hat{\mu}_{X^2}}, \frac{\partial \tilde{\sigma}^2}{\partial \hat{\mu}_{Y^2}}, \frac{\partial \tilde{\sigma}^2}{\partial \hat{\mu}_{XY}}, 0,0,0,0\right)$$

where

$$\frac{\partial \tilde{\sigma}^2}{\partial \hat{\mu}_{X^2}} = \frac{2[\tilde{\beta}_1\hat{\mu}_{X^2}-(1-\tilde{\beta}_1^2)\hat{\mu}_{XY}-\tilde{\beta}_1\hat{\mu}_{Y^2}]\frac{\partial \tilde{\beta}_1}{\partial \hat{\mu}_{X^2}}+\tilde{\beta}_1^2(1+\tilde{\beta}_1^2)}{2(1+\tilde{\beta}_1^2)^2},$$

$$\frac{\partial \tilde{\sigma}^2}{\partial \hat{\mu}_{Y^2}} = \frac{2[\tilde{\beta}_1\hat{\mu}_{X^2}-(1-\tilde{\beta}_1^2)\hat{\mu}_{XY}-\tilde{\beta}_1\hat{\mu}_{Y^2}]\frac{\partial \tilde{\beta}_1}{\partial \hat{\mu}_{Y^2}}+(1+\tilde{\beta}_1^2)}{2(1+\tilde{\beta}_1^2)^2},$$

$$\frac{\partial \tilde{\sigma}^2}{\partial \hat{\mu}_{XY}} = \frac{[\tilde{\beta}_1\hat{\mu}_{X^2}-(1-\tilde{\beta}_1^2)\hat{\mu}_{XY}-\tilde{\beta}_1\hat{\mu}_{Y^2}]\frac{\partial \tilde{\beta}_1}{\partial \hat{\mu}_{XY}}-\tilde{\beta}_1(1+\tilde{\beta}_1^2)}{(1+\tilde{\beta}_1^2)^2}.$$

The first order derivatives for $\hat{\sigma}^2$ are complicated to obtain because it does not have an explicit expression and involves an implicit function. Before starting, we define

$$F_f = \hat{S}_{XY}\hat{\beta}_1^2 + (\hat{S}_{XX} - \hat{S}_{YY})\hat{\beta}_1 + \hat{S}_{XY} = 0,$$

$$\hat{N}_f = \frac{\sum_{t=2}^{n}\{(\hat{\beta}_1 X_t - Y_t)-(\hat{\beta}_1\bar{X}_n-\bar{Y}_n)\}\{(\hat{\beta}_1 X_{t-1}-Y_{t-1})-(\hat{\beta}_1\bar{X}_{n-1}-\bar{Y}_{n-1})\}}{n-1},$$

and

$$\hat{D}_f = \frac{\sum_{t=2}^{n}\{(\hat{\beta}_1 X_{t-1}-Y_{t-1})-(\hat{\beta}_1\bar{X}_{n-1}-\bar{Y}_{n-1})\}^2}{n-1}.$$

Then

$$\hat{\rho} = \hat{N}_f/\hat{D}_f.$$

Applying the rules for derivatives of an implicit function given in Section 3, we can find the first order derivatives of $\hat{\beta}_1$ as

$$\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}} = -\frac{\partial F_f/\partial\hat{\psi}}{\partial F_f/\partial\hat{\beta}_1}.$$

Then we have the first order derivatives of $\hat{\sigma}^2$ as

$$\frac{\partial\hat{\sigma}^2}{\partial\hat{\psi}} = -\frac{2\hat{\beta}_1}{1+\hat{\beta}_1^2}\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}}\hat{\sigma}^2 + \frac{1}{2(1+\hat{\beta}_1^2)}\left\{(1+\hat{\rho}^2)\frac{\partial\hat{D}_f}{\partial\hat{\psi}} - 2\hat{\rho}\frac{\partial\hat{N}_f}{\partial\hat{\psi}}\right\},$$

where $\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}}$, $\frac{\partial\hat{D}_f}{\partial\hat{\psi}}$, and $\frac{\partial\hat{N}_f}{\partial\hat{\psi}}$ are vectors of the first order derivatives.

After deriving the first order derivatives, we have to find the second order derivatives. For the second order derivatives of $\tilde{\beta}_1$, we have

$$\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_X\partial\hat{\psi}} = \frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_Y\partial\hat{\psi}} = \frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{XY^1}\partial\hat{\psi}} = \frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{X^1Y}\partial\hat{\psi}} = \frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{XX}\partial\hat{\psi}} = \frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{YY}\partial\hat{\psi}} = 0,$$

$$\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{X^2}^2} = \frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{Y^2}^2} = -\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{X^2}\partial\hat{\mu}_{Y^2}} = \frac{2\hat{\mu}_{XY}}{\left[(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2+4\hat{\mu}_{XY}^2\right]^{3/2}},$$

$$\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{X^2}\partial\hat{\mu}_{XY}} = -\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{Y^2}\partial\hat{\mu}_{XY}} = -2\frac{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})}{\left[(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2+4\hat{\mu}_{XY}^2\right]^{3/2}} - \frac{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})}{2\hat{\mu}_{XY}^2\sqrt{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2+4\hat{\mu}_{XY}^2}} + \frac{1}{2\hat{\mu}_{XY}^2},$$

$$\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{XY}^2} = 2\frac{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2}{\hat{\mu}_{XY}\left[(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2+4\hat{\mu}_{XY}^2\right]^{3/2}} + \frac{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2}{\hat{\mu}_{XY}^3\sqrt{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})^2+4\hat{\mu}_{XY}^2}} - \frac{(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})}{\hat{\mu}_{XY}^3}.$$

Therefore, the second order derivatives for $\tilde{\sigma}^2$ would be

$$\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_X\partial\hat{\psi}} = \frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_Y\partial\hat{\psi}} = \frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{XY^1}\partial\hat{\psi}} = \frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{X^1Y}\partial\hat{\psi}} = \frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{XX}\partial\hat{\psi}} = \frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{YY}\partial\hat{\psi}} = 0,$$

$$\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{X^2}^2} = \frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{Y^2}^2} = \frac{[\tilde{\beta}_1(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})-(1-\tilde{\beta}_1^2)\hat{\mu}_{XY}]\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{X^2}^2}+2\tilde{\beta}_1\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{X^2}}}{(1+\tilde{\beta}_1^2)^2}$$

$$+\frac{[(1-3\tilde{\beta}_1^2)(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})+2\tilde{\beta}_1(3-\tilde{\beta}_1^2)\hat{\mu}_{XY}]\left(\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{X^2}}\right)^2}{(1+\tilde{\beta}_1^2)^3},$$

$$\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{X^2}\partial\hat{\mu}_{Y^2}} = \frac{[\tilde{\beta}_1(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})-(1-\tilde{\beta}_1^2)\hat{\mu}_{XY}]\frac{\partial^2\tilde{\beta}_1}{\partial\hat{\mu}_{X^2}\partial\hat{\mu}_{Y^2}}-\tilde{\beta}_1\left(\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{X^2}}-\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{Y^2}}\right)}{\left(1+\tilde{\beta}_1^2\right)^2}$$

$$+\frac{[(1-3\tilde{\beta}_1^2)(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})+2\tilde{\beta}_1(3-\tilde{\beta}_1^2)\hat{\mu}_{XY}]\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{X^2}}\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{Y^2}}}{\left(1+\tilde{\beta}_1^2\right)^3},$$

$$\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{X^2}\partial\hat{\mu}_{XY}} = -\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{Y^2}\partial\hat{\mu}_{XY}} = \frac{[\tilde{\beta}_1(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})-(1-\tilde{\beta}_1^2)\hat{\mu}_{XY}]\frac{\partial^2\tilde{\beta}_1}{\partial\hat{\mu}_{X^2}\partial\hat{\mu}_{XY}}+\tilde{\beta}_1\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{XY}}-(1-\tilde{\beta}_1^2)\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{X^2}}}{\left(1+\tilde{\beta}_1^2\right)^2}$$

$$+\frac{[(1-3\tilde{\beta}_1^2)(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})+2\tilde{\beta}_1(3-\tilde{\beta}_1^2)\hat{\mu}_{XY}]\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{X^2}}\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{XY}}}{\left(1+\tilde{\beta}_1^2\right)^3},$$

$$\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{XY}^2} = \frac{[\tilde{\beta}_1[\hat{\mu}_{X^2}-\hat{\mu}_{Y^2}]-[1-\tilde{\beta}_1^2]\hat{\mu}_{XY}]\frac{\partial^2\tilde{\beta}_1}{\partial\hat{\mu}_{XY}^2}-2(1-\tilde{\beta}_1^2)\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{XY}}}{\left(1+\tilde{\beta}_1^2\right)^2}+\frac{[(1-3\tilde{\beta}_1^2)(\hat{\mu}_{X^2}-\hat{\mu}_{Y^2})+2\tilde{\beta}_1(3-\tilde{\beta}_1^2)\hat{\mu}_{XY}]\left[\frac{\partial\tilde{\beta}_1}{\partial\hat{\mu}_{XY}}\right]^2}{\left(1+\tilde{\beta}_1^2\right)^3}.$$

The second order derivatives of $\hat{\sigma}^2$ can be obtained by using the method for implicit functions as well. The first thing is to find the second order derivatives for F_f, $\hat{D}_f$, and $\hat{N}_f$. Then we can figure out the second order derivatives of $\hat{\beta}_1$ after that. By the rule of derivatives for implicit functions, we have

$$\frac{\partial^2\hat{\beta}_1}{\partial\hat{\psi}^T\partial\hat{\psi}} = -\frac{1}{\partial F_f/\partial\hat{\beta}_1}\left\{\frac{\partial^2 F_f}{\partial\hat{\beta}_1^2}\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}^T}\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}}+\frac{\partial^2 F_f}{\partial\hat{\beta}_1\partial\hat{\psi}^T}\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}}+\frac{\partial^2 F_f}{\partial\hat{\beta}_1\partial\hat{\psi}}\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}^T}+\frac{\partial^2 F_f}{\partial\hat{\beta}_1^2}\right\}.$$

Finally, we can get the second order derivatives for $\hat{\sigma}^2$ by substituting these values. Therefore,

$$\frac{\partial^2\hat{\sigma}^2}{\partial\hat{\psi}^T\partial\hat{\psi}} = \frac{1}{2(1+\hat{\beta}_1^2)}\left\{(1+\hat{\rho}^2)\frac{\partial^2\hat{D}_f}{\partial\hat{\psi}^T\partial\hat{\psi}}-2\hat{\rho}\frac{\partial^2\hat{N}_f}{\partial\hat{\psi}^T\partial\hat{\psi}}-\frac{2}{\hat{D}_f}\left(\frac{\partial\hat{N}_f}{\partial\hat{\psi}^T}-\hat{\rho}\frac{\partial\hat{D}_f}{\partial\hat{\psi}^T}\right)\left(\frac{\partial\hat{N}_f}{\partial\hat{\psi}}-\hat{\rho}\frac{\partial\hat{D}_f}{\partial\hat{\psi}}\right)\right\}$$

$$+\frac{2\hat{\sigma}^2}{\left(1+\hat{\beta}_1^2\right)^2}\left\{\left(3\hat{\beta}_1^2-1\right)\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}^T}\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}}-\hat{\beta}_1\left(1+\hat{\beta}_1^2\right)\frac{\partial^2\hat{\beta}_1}{\partial\hat{\psi}^T\partial\hat{\psi}}\right\}$$

$$-\frac{\hat{\beta}_1}{\left(1+\hat{\beta}_1^2\right)^2}\left\{\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}^T}\left[(1+\hat{\rho}^2)\frac{\partial\hat{D}_f}{\partial\hat{\psi}}-2\hat{\rho}\frac{\partial\hat{N}_f}{\partial\hat{\psi}}\right]+\left[(1+\hat{\rho}^2)\frac{\partial\hat{D}_f}{\partial\hat{\psi}^T}-2\hat{\rho}\frac{\partial\hat{N}_f}{\partial\hat{\psi}^T}\right]\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}}\right\}.$$

Other second order derivatives are obtained using the same operations as these two

Under the null hypothesis of $\rho = 0$, we have the limit of $\hat{\psi}$ as

$$\hat{\psi} \to \psi = (\mu_x, \beta_0 + \beta_1\mu_x, \sigma_x^2 + \sigma^2, \beta_1^2\sigma_x^2 + \sigma^2, \beta_1\sigma_x^2, 0,0,0,0).$$

Therefore, substituting ψ into these derivatives, we have the first order derivatives of $\tilde{\beta}_1$ and $\tilde{\sigma}^2$ at ψ as

$$\left.\frac{\partial\tilde{\beta}_1}{\partial\hat{\psi}}\right|_{\psi} = \left[0,0,-\frac{\beta_1}{(1+\beta_1^2)\sigma_x^2}, \frac{\beta_1}{(1+\beta_1^2)\sigma_x^2}, \frac{1-\beta_1^2}{(1+\beta_1^2)\sigma_x^2}, 0,0,0,0\right],$$

and

$$\left.\frac{\partial\tilde{\sigma}^2}{\partial\hat{\psi}}\right|_{\psi} = \left(0,0,\frac{\beta_1^2}{2(1+\beta_1^2)}, \frac{1}{2(1+\beta_1^2)}, -\frac{\beta_1}{(1+\beta_1^2)}, 0,0,0,0\right).$$

The first order derivatives of $\hat{\beta}_1$, $\hat{D}_f$, and $\hat{N}_f$ are

$$\left.\frac{\partial\hat{\beta}_1}{\partial\hat{\psi}}\right|_{\psi} = \left(0,0,-\frac{\beta_1}{(1+\beta_1^2)\sigma_x^2}, \frac{\beta_1}{(1+\beta_1^2)\sigma_x^2}, \frac{1-\beta_1^2}{(1+\beta_1^2)\sigma_x^2}, 0,0,0,0\right),$$

$$\left.\frac{\partial\hat{D}_f}{\partial\hat{\psi}}\right|_{\psi} = \left(0,0,\beta_1^2 - \frac{2\beta_1^2\sigma^2}{(1+\beta_1^2)\sigma_x^2}, \frac{2\beta_1^2\sigma^2}{(1+\beta_1^2)\sigma_x^2} + 1, 2\beta_1\frac{(1-\beta_1^2)\sigma^2}{(1+\beta_1^2)\sigma_x^2} - 2\beta_1, 0,0,0,0\right),$$

$$\left.\frac{\partial\hat{N}_f}{\partial\hat{\psi}}\right|_{\psi} = (0,0,0,0,0,-\beta_1,-\beta_1,\beta_1^2,1).$$

Therefore, the first order derivatives of $\hat{\sigma}^2$ at ψ would be

$$\left.\frac{\partial\hat{\sigma}^2}{\partial\hat{\psi}}\right|_{\psi} = \left(0,0,\frac{\beta_1^2}{2(1+\beta_1^2)}, \frac{1}{2(1+\beta_1^2)}, -\frac{\beta_1}{(1+\beta_1^2)}, 0,0,0,0\right).$$

Then we have the first order derivatives of test statistic evaluated at ψ,

$$\left.\frac{\partial T_{LRT,f,\rho}}{\partial\hat{\psi}}\right|_{\psi} = -2(n-1)\left(\frac{1}{\hat{\sigma}^2}\frac{\partial\hat{\sigma}^2}{\partial\hat{\psi}} - \frac{1}{\tilde{\sigma}^2}\frac{\partial\tilde{\sigma}^2}{\partial\hat{\psi}}\right) = 0_{1\times 9}.$$

In order to find the second order derivatives of the test statistic evaluate at ψ, we need to find the second order derivatives of $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ respectively. For $\tilde{\sigma}^2$, because

$$\left.\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{X^2}^2}\right|_{\psi} = \left.\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{Y^2}^2}\right|_{\psi} = -\left.\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{X^2}\partial\hat{\mu}_{Y^2}}\right|_{\psi} = \frac{2\beta_1}{(1+\beta_1^2)^3\sigma_x^4},$$

$$\left.\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{X^2}\partial\hat{\mu}_{XY}}\right|_{\psi} = -\left.\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{Y^2}\partial\hat{\mu}_{XY}}\right|_{\psi} = -2\frac{(1-\beta_1^2)}{(1+\beta_1^2)^3\sigma_x^4} + \frac{1}{(1+\beta_1^2)\sigma_x^4},$$

$$\left.\frac{\partial^2\tilde{\beta}_1^2}{\partial\hat{\mu}_{XY}^2}\right|_{\psi} = 2\frac{(1-\beta_1^2)^2}{\beta_1(1+\beta_1^2)^3\sigma_x^4} - 2\frac{(1-\beta_1^2)}{\beta_1(1+\beta_1^2)\sigma_x^4},$$

we have

$$\left.\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{X^2}^2}\right|_{\psi} = \left.\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{Y^2}^2}\right|_{\psi} = -\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{X^2}\partial\hat{\mu}_{Y^2}} = -\frac{\beta_1^2}{(1+\beta_1^2)^3\sigma_x^2},$$

$$\left.\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{X^2}\partial\hat{\mu}_{XY}}\right|_{\psi} = -\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{Y^2}\partial\hat{\mu}_{XY}} = \frac{\beta_1(1-\beta_1^2)}{(1+\beta_1^2)^3\sigma_x^2},$$

$$\left.\frac{\partial^2\tilde{\sigma}^2}{\partial\hat{\mu}_{XY}^2}\right|_{\psi} = -\frac{(1-\beta_1^2)^2}{(1+\beta_1^2)^3\sigma_x^2}.$$

Other second order derivatives of $\tilde{\sigma}^2$ are all zero as shown above. The second order derivatives of $\hat{\sigma}^2$ evaluated at ψ are,

$$\left.\frac{\partial^2\hat{\sigma}^2}{\partial\hat{\psi}^T\partial\hat{\psi}}\right|_{\psi} = \begin{pmatrix} 0_{2\times2} & 0_{2\times3} & 0_{2\times4} \\ 0_{3\times2} & A_{\psi} & 0_{3\times4} \\ 0_{4\times2} & 0_{4\times3} & B_{\psi} \end{pmatrix},$$

where

$$A_{\psi} = \frac{1}{(1+\beta_1^2)^3\sigma_x^2}\begin{pmatrix} -\beta_1^2 & \beta_1^2 & \beta_1(1-\beta_1^2) \\ \beta_1^2 & -\beta_1^2 & -\beta_1(1-\beta_1^2) \\ \beta_1(1-\beta_1^2) & -\beta_1(1-\beta_1^2) & -(1-\beta_1^2)^2 \end{pmatrix},$$

$$B_{\psi} = \frac{1}{(1+\beta_1^2)^2\sigma^2}\begin{pmatrix} -\beta_1^2 & -\beta_1^2 & \beta_1^3 & \beta_1 \\ -\beta_1^2 & -\beta_1^2 & \beta_1^3 & \beta_1 \\ \beta_1^3 & \beta_1^3 & -\beta_1^4 & -\beta_1^2 \\ \beta_1 & \beta_1 & -\beta_1^2 & -1 \end{pmatrix}.$$

Notice that the second order derivative for the likelihood ratio test statistic can be expressed as following,

$$\frac{\partial^2 T_{LRT,f,\rho}}{\partial\widehat{\psi}^T\partial\widehat{\psi}} = -2(n-1)\left(-\frac{1}{\hat{\sigma}^4}\frac{\partial\hat{\sigma}^2}{\partial\widehat{\psi}^T}\frac{\partial\hat{\sigma}^2}{\partial\widehat{\psi}} + \frac{1}{\tilde{\sigma}^4}\frac{\partial\tilde{\sigma}^2}{\partial\widehat{\psi}^T}\frac{\partial\tilde{\sigma}^2}{\partial\widehat{\psi}} + \frac{1}{\hat{\sigma}^2}\frac{\partial^2\hat{\sigma}^2}{\partial\widehat{\psi}^T\partial\widehat{\psi}} - \frac{1}{\tilde{\sigma}^2}\frac{\partial^2\tilde{\sigma}^2}{\partial\widehat{\psi}^T\partial\widehat{\psi}}\right).$$

Therefore, substituting the values of the first and the second order derivatives for $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ at ψ, we have

$$\left.\frac{\partial^2 T_{LRT,f,\rho}}{\partial\widehat{\psi}^T\partial\widehat{\psi}}\right|_{\psi} = -\frac{4(n-1)}{\sigma^2}\begin{pmatrix} 0_{5\times4} & 0_{5\times4} \\ 0_{4\times5} & B_{\psi} \end{pmatrix}.$$

Let row vectors, $Z_{\rho t} = (X_t, X_{t-1}, Y_t, Y_{t-1})$, $z_{\rho t} = (x_t, x_{t-1}, y_t, y_{t-1})$, and $\varepsilon_{\rho t} = (e_t, e_{t-1}, u_t, u_{t-1})$, then

$$Z_{\rho t} = z_{\rho t} + \varepsilon_{\rho t} \qquad\qquad t = 2,3,\dots,n.$$

Under the null hypothesis, $\varepsilon_{\rho t}$ are independent identically distributed random vector with mean vector 0 and covariance matrix $\Sigma_{\varepsilon\varepsilon} = \sigma^2 I$. Define

$$\widehat{\omega} = (\bar{X}_n, \bar{X}_{n-1}, \bar{Y}_n, \bar{Y}_{n-1}, \hat{\mu}_{X^2}, \hat{\mu}_{XX}, \hat{\mu}_{XY}, \hat{\mu}_{XY1}, \hat{\mu}_{X1^2}, \hat{\mu}_{X1Y}, \hat{\mu}_{X1Y1}, \hat{\mu}_{Y^2}, \hat{\mu}_{YY}, \hat{\mu}_{Y1^2})^T.$$

Then applying Theorem 3.8 in Section 3, we have

$$\widehat{\omega} \to AN(\omega_n, G_{\omega n}),$$

where

$$\omega_n = \left(\bar{z}_\rho, (\mathrm{vech}m_{zz} + \mathrm{vech}\Sigma_{\varepsilon\varepsilon})^T\right)^T,$$

and $G_{\omega n}$ is the variance covariance matrix for $\widehat{\omega}$. From the second order derivatives of test statistic at ψ, we know that only the last four elements in $\widehat{\psi}$ have non-zero second order derivatives. These four elements $\hat{\gamma} = (\hat{\mu}_{XY1}, \hat{\mu}_{X1Y}, \hat{\mu}_{XX}, \hat{\mu}_{YY})^T$ are a subset of $\widehat{\omega}$. The corresponding variance covariance matrix would be

$$G_{\gamma n} = \frac{1}{(n-1)}\begin{pmatrix} G_{\gamma n11} & G_{\gamma n12} \\ G_{\gamma n21} & G_{\gamma n22} \end{pmatrix},$$

where

$$G_{\gamma n11} = \begin{pmatrix} (1+\beta_1^2)\sigma_x^2\sigma^2 + \sigma^4 & 0 \\ 0 & (1+\beta_1^2)\sigma_x^2\sigma^2 + \sigma^4 \end{pmatrix},$$

$$G_{\gamma n12} = G_{\gamma n21} = \beta_1\sigma_x^2\sigma^2 I_{2\times 2},$$

$$G_{\gamma n22} = \begin{pmatrix} 2\sigma_x^2\sigma^2 + \sigma^4 & 0 \\ 0 & 2\beta_1^2\sigma_x^2\sigma^2 + \sigma^4 \end{pmatrix}.$$

That is, $\hat{\gamma}$ is asymptotic normally distributed with mean vector γ_n and covariance matrix $G_{\gamma n}$.

Because the first order derivatives of the test statistic at ψ are all zero, and the second order derivatives of the test statistic at ψ are also zero, except those corresponding to $\hat{\mu}_{XY1}$, $\hat{\mu}_{X1Y}$, $\hat{\mu}_{XX}$, and $\hat{\mu}_{YY}$, the Taylor expansion of the test statistic would be,

$$T_{LRT,f,\rho} \doteq T_{LRT,f,\rho}\Big|_{\psi} + \frac{\partial T_{LRT,f,\rho}}{\partial \hat{\psi}}\Big|_{\psi}\left(\hat{\psi}-\psi\right)^T + \frac{1}{2}\left(\hat{\psi}-\psi\right)\frac{\partial^2 T_{LRT,f,\rho}}{\partial\hat{\psi}\partial\hat{\psi}^T}\Big|_{\psi}\left(\hat{\psi}-\psi\right)^T$$

$$= \frac{1}{2}(\hat{\gamma}-\gamma)\frac{\partial^2 T_{LRT,f,\rho}}{\partial\hat{\gamma}\partial\hat{\gamma}^T}\Big|_{\gamma}(\hat{\gamma}-\gamma)^T = 2(\hat{\gamma}-\gamma)C_\rho(\hat{\gamma}-\gamma)^T$$

where

$$C_\rho = \frac{1}{\left(1+\beta_1^2\right)^2}\frac{(n-1)}{\sigma^4}\begin{pmatrix} \beta_1^2 & \beta_1^2 & -\beta_1^3 & -\beta_1 \\ \beta_1^2 & \beta_1^2 & -\beta_1^3 & -\beta_1 \\ -\beta_1^3 & -\beta_1^3 & \beta_1^4 & \beta_1^2 \\ -\beta_1 & -\beta_1 & \beta_1^2 & 1 \end{pmatrix}.$$

Then

$$G_{\gamma n}C_\rho G_{\gamma n} = G_{\gamma n}C_\rho G_{\gamma n}C_\rho G_{\gamma n} = \frac{1}{(1+\beta_1^2)^2}\frac{\sigma^4}{(n-1)}\begin{pmatrix} \beta_1^2 & \beta_1^2 & -\beta_1^3 & -\beta_1 \\ \beta_1^2 & \beta_1^2 & -\beta_1^3 & -\beta_1 \\ -\beta_1^3 & -\beta_1^3 & \beta_1^4 & \beta_1^2 \\ -\beta_1 & -\beta_1 & \beta_1^2 & 1 \end{pmatrix},$$

and

$$C_\rho G_{\gamma n} = \frac{1}{(1+\beta_1^2)^2}\begin{pmatrix} \beta_1^2 & \beta_1^2 & -\beta_1^3 & -\beta_1 \\ \beta_1^2 & \beta_1^2 & -\beta_1^3 & -\beta_1 \\ -\beta_1^3 & -\beta_1^3 & \beta_1^4 & \beta_1^2 \\ -\beta_1 & -\beta_1 & \beta_1^2 & 1 \end{pmatrix}.$$

Notice that the trace of $C_\rho G_{\gamma n}$ is 1. Therefore, according to Theorem 3.9 in Section 3, the distribution of $(\hat{\gamma} - \gamma)C_\rho(\hat{\gamma} - \gamma)^T$ converges to a chi-squared distribution with one degree of freedom. Thus, under the null hypothesis of $\rho = 0$,

$$T_{LRT,f,\rho} \xrightarrow{d} 2\chi_1^2.$$

This result is the same as the efficient score test statistic for testing $\rho = 0$ under the null hypothesis as shown in section 4.3.1. The coefficient 2 before chi-squared distribution must be due to the inconsistent estimator for σ^2.

4.3.3 Efficient score test statistic under $\boldsymbol{H_0}: (\boldsymbol{\beta_0}, \boldsymbol{\beta_1}) = (\boldsymbol{b_0}, \boldsymbol{b_1})$

In section 2.3, the efficient score test statistic for the test hypothesis $H_0: (\beta_0, \beta_1) = (b_0, b_1)$ has been derived. The large sample distribution for the test statistic will be derived in this section. As shown in (2.20), the test statistic is

$$T_{ES,f,\beta} = U_\beta(\ddot{\varphi})^T I(\ddot{\varphi})^{-1}_{2\times 2} U_\beta(\ddot{\varphi})$$

where

$$U_\beta(\ddot{\varphi}) = \left(\frac{\sum_{t=2}^n\{(Y_t - \ddot{\gamma}_t - \ddot{\rho}Y_{t-1})(1-\ddot{\rho})\}}{\ddot{\sigma}^2}, \frac{\sum_{t=2}^n\{(Y_t - \ddot{\gamma}_t - \ddot{\rho}Y_{t-1})\ddot{\lambda}_t\}}{\ddot{\sigma}^2}\right)^T,$$

$$I(\ddot{\varphi})_{2\times 2} = \frac{1}{\ddot{\sigma}^2(1+b_1^2)}\begin{bmatrix}(n-1)(1-\ddot{\rho})^2 & (1-\ddot{\rho})\sum_{t=2}^{n}\ddot{\lambda}_t \\ (1-\ddot{\rho})\sum_{t=2}^{n}\ddot{\lambda}_t & \sum_{t=2}^{n}\ddot{\lambda}_t^2\end{bmatrix},$$

and $I(\ddot{\varphi})_{2\times 2}^{-1}$ is the inverse of the information matrix.

Under the null hypothesis, we have $\beta_0 = b_0$ and $\beta_1 = b_1$. In section 4.1, we have investigated the consistency for each parameter estimator and we know that $\ddot{\rho}$ is a consistent estimator and $\ddot{\sigma}^2$ converges to $\sigma^2/2$ though inconsistent. By using those results, the first term of $U_\beta(\ddot{\varphi})$ would have a limit as

$$p\lim\frac{\sum_{t=2}^{n}(Y_t-\ddot{\gamma}_t-\ddot{\rho}Y_{t-1})(1-\ddot{\rho})}{\ddot{\sigma}^2\sqrt{n-1}} = p\lim\frac{2(1-\rho)\sum_{t=2}^{n}\left\{b_1x_t+u_t-b_1\left(x_t-\rho x_{t-1}+\frac{v_t+b_1\delta_t}{1+b_1^2}\right)-\rho(b_1x_{t-1}+u_{t-1})\right\}}{\sigma^2\sqrt{n-1}}$$

$$= 2(1-\rho)\frac{1}{\sigma^2(1+b_1^2)}p\lim\frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}(\delta_t - b_1v_t)$$

$$= 2(1-\rho)\sqrt{\frac{1}{\sigma^2(1+b_1^2)}}p\lim\frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}\frac{(\delta_t-b_1v_t)}{\sqrt{\sigma^2(1+b_1^2)}}$$

Let

$$Z_{1t} = \frac{(\delta_t-b_1v_t)}{\sqrt{\sigma^2(1+b_1^2)}}$$

and

$$Z_1 = \frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}Z_{1t} \qquad \text{for } t = 2,3,\dots,n.$$

Thus,

$$p\lim\frac{\sum_{t=2}^{n}(Y_t-\ddot{\gamma}_t-\ddot{\rho}Y_{t-1})}{\sqrt{n-1}} = 2\sqrt{\frac{1}{\sigma^2(1+b_1^2)}}p\lim Z_1.$$

Considering the second non-zero element in the score vector, its probability limit would be

$$plim\frac{\sum_{t=2}^{n}\{(Y_t-\ddot{\gamma}_t-\ddot{\rho}Y_{t-1})\ddot{\lambda}_t\}}{\ddot{\sigma}^2\sqrt{n-1}} = 2\frac{1}{\sigma^2(1+b_1^2)}plim\frac{\sum_{t=2}^{n}\left\{(\delta_t-b_1v_t)\left(x_t-\rho x_{t-1}+\frac{v_t+b_1\delta_t}{1+b_1^2}\right)\right\}}{\sqrt{n-1}}$$

$$= 2\sqrt{\frac{1}{\sigma^2(1+b_1^2)}}plim\frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}\frac{(\delta_t-b_1v_t)}{\sqrt{\sigma^2(1+b_1^2)}}(x_t-\rho x_{t-1})$$

$$+2\frac{1}{1+b_1^2}plim\frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}\frac{(\delta_t-b_1v_t)}{\sqrt{\sigma^2(1+b_1^2)}}\frac{(v_t+b_1\delta_t)}{\sqrt{\sigma^2(1+b_1^2)}}.$$

Let

$$Z_{2t} = \frac{(\delta_t-b_1v_t)}{\sqrt{\sigma^2(1+b_1^2)}}(x_t-\rho x_{t-1}),$$

$$Z_2 = \frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}Z_{2t},$$

and

$$Z_{3t} = \frac{(\delta_t-b_1v_t)}{\sqrt{\sigma^2(1+b_1^2)}}\frac{(v_t+b_1\delta_t)}{\sqrt{\sigma^2(1+b_1^2)}}$$

$$Z_3 = \frac{1}{\sqrt{n-1}}\sum_{t=2}^{n}Z_{3t} \qquad \text{for } t = 2,3,\ldots,n.$$

Then we have

$$plim\frac{\sum_{t=2}^{n}\{(Y_t-\ddot{\gamma}_t-\ddot{\rho}Y_{t-1})\ddot{\lambda}_t\}}{\ddot{\sigma}^2\sqrt{n-1}} = 2\sqrt{\frac{1}{\sigma^2(1+b_1^2)}}\text{plim}Z_2 + 2\frac{1}{1+b_1^2}plimZ_3.$$

Since (v_t, δ_t) is an i.i.d. normally distributed random vector according to assumptions of the model, then Z_{1t} is i.i.d. normally distributed with mean 0 and variance 1. Therefore, Z_1 is distributed as a normal random variable with mean 0 and variance 1. Next, we will determine the expectations and covariance for Z_{1t}, Z_{2t} and Z_{3t}. For Z_{2t}, we have

$$E(Z_{2t}) = 0,$$

and

$$Var(Z_{2t}) = (x_t - \rho x_{t-1})^2 Var(Z_{1t}) = (x_t - \rho x_{t-1})^2 \quad t = 2,3,\dots,n.$$

For Z_{3t}, we have

$$E(Z_{3t}) = E\,\frac{(\delta_t - b_1 v_t)(v_t + b_1\delta_t)}{\sigma^2(1+b_1^2)} = 0,$$

and

$$Var(Z_{3t}) = Var\,\frac{(\delta_t - b_1 v_t)(v_t + b_1\delta_t)}{\sigma^2(1+b_1^2)} = \frac{E\left((1-b_1^2)\delta_t v_t - b_1 v_t^2 + b_1\delta_t^2\right)^2}{\sigma^4(1+b_1^2)^2}$$

$$= \frac{E\left((1-b_1^2)^2\delta_t^2 v_t^2 + b_1^2 v_t^4 + b_1^2\delta_t^4 - 2b_1(1-b_1^2)\delta_t v_t^3 + 2b_1(1-b_1^2)\delta_t^3 v_t - 2b_1^2\delta_t^2 v_t^2\right)}{\sigma^4(1+b_1^2)^2}$$

$$= \frac{(1-b_1^2)^2\sigma^4 + 3b_1^2\sigma^4 + 3b_1^2\sigma^4 - 2b_1^2\sigma^4}{\sigma^4(1+b_1^2)^2}$$

$$= 1.$$

The covariance among Z_{1t}, Z_{2t}, and Z_{3t} would be

$$cov(Z_{1t}, Z_{2t}) = E(Z_{1t}Z_{2t}) - E(Z_{1t})E(Z_{2t})$$

$$= E\,\frac{(\delta_t - b_1 v_t)}{\sqrt{\sigma^2(1+b_1^2)}}\frac{(\delta_t - b_1 v_t)}{\sqrt{\sigma^2(1+b_1^2)}}(x_t - \rho x_{t-1})$$

$$= x_t - \rho x_{t-1},$$

$$cov(Z_{1t}, Z_{3t}) = E(Z_{1t}Z_{3t}) - E(Z_{1t})E(Z_{3t}) = E(Z_{1t}Z_{1t}Z_{2t})$$

$$= E\,\frac{(\delta_t^2 v_t - 2b_1\delta_t v_t^2 + b_1^2 v_t^3 + b_1\delta_t^3 - 2b_1^2\delta_t^2 v_t + b_1^3\delta_t v_t^2)}{\sigma^2(1+b_1^2)\sqrt{\sigma^2(1+b_1^2)}} = 0,$$

and

$$cov(Z_{2t}, Z_{3t}) = E(Z_{2t}Z_{3t}) - E(Z_{2t})E(Z_{3t}) = E(Z_{2t}Z_{1t}Z_{2t})$$

$$= E\,\frac{(v_t^2\delta_t + 2b_1\delta_t^2 v_t + b_1^2\delta_t^3 - b_1 v_t^3 - 2b_1^2\delta_t v_t^2 - b_1^3\delta_t^2 v_t)(x_t - \rho x_{t-1})^2}{\sigma^2(1+b_1^2)\sqrt{\sigma^2(1+b_1^2)}} = 0.$$

Therefore, we have the mean and covariance structure of this random vector for any value $t = 2,3, \ldots, n$ as

$$Z_t = \begin{pmatrix} (1-\rho)\sqrt{\frac{1}{\sigma^2(1+b_1^2)}} Z_{1t} \\ \sqrt{\frac{1}{\sigma^2(1+b_1^2)}} Z_{2t} + \frac{1}{1+b_1^2} Z_{3t} \end{pmatrix} \sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{U11t} & \sigma_{U12t} \\ \sigma_{U21t} & \sigma_{U22t} \end{pmatrix} \right]$$

where

$$\sigma_{U11t} = (1-\rho)^2 \frac{1}{\sigma^2(1+b_1^2)},$$

$$\sigma_{U12t} = \sigma_{U21t} = cov\left((1-\rho)\sqrt{\frac{1}{\sigma^2(1+b_1^2)}} Z_{1t}, \sqrt{\frac{1}{\sigma^2(1+b_1^2)}} Z_{2t} + \frac{1}{1+b_1^2} Z_{3t} \right)$$

$$= (1-\rho) \frac{1}{\sigma^2(1+b_1^2)} (x_t - \rho x_{t-1}),$$

and

$$\sigma_{U22t} = \frac{1}{\sigma^2(1+b_1^2)} (x_t - \rho x_{t-1})^2 + \frac{1}{(1+b_1^2)^2}.$$

It is noticed that σ_{U11t} does not depend on t, and

$$\frac{\sum_{t=2}^{n} \sigma_{U12t}}{n-1} = \frac{\sum_{t=2}^{n} \sigma_{U21t}}{n-1} = (1-\rho) \frac{1}{\sigma^2(1+b_1^2)} \frac{\sum_{t=2}^{n}(x_t - \rho x_{t-1})}{n-1} \to \frac{\sigma^2}{1+b_1^2} (1-\rho)^2 \mu_x,$$

$$\frac{\sum_{t=2}^{n} \sigma_{U22t}}{n-1} = \frac{1}{\sigma^2(1+b_1^2)} \frac{\sum_{t=2}^{n}(x_t - \rho x_{t-1})^2}{n-1} + \frac{1}{(1+b_1^2)^2}$$

$$\to \frac{1}{\sigma^2(1+b_1^2)} \left((1+\rho^2)\sigma_x^2 + (1-\rho)^2 \mu_x^2 \right) + \frac{1}{(1+b_1^2)^2}.$$

For any $a = (a_1, a_2) \in \mathcal{R}^2$, we have

$$aZ_t = a_1(1-\rho)\sqrt{\frac{1}{\sigma^2(1+b_1^2)}} Z_{1t} + a_2\sqrt{\frac{1}{\sigma^2(1+b_1^2)}} Z_{2t} + a_2 \frac{1}{1+b_1^2} Z_{3t}.$$

Its expectation and variance are

$$EaZ_t = 0,$$

and

$$Var(aZ_t) = \frac{a_1^2(1-\rho)^2}{\sigma^2\left(1+b_1^2\right)} + a_2^2\frac{(x_t-\rho x_{t-1})^2}{\sigma^2\left(1+b_1^2\right)} + a_2^2\frac{1}{\left(1+b_1^2\right)^2} + 2a_1a_2\frac{(1-\rho)(x_t-\rho x_{t-1})}{\sigma^2\left(1+b_1^2\right)}.$$

Thus,

$$\lim_{n\to\infty}\frac{\sum_{t=2}^{n} Var(aZ_t)}{n-1} = (a_1+a_2\mu_x)^2\frac{(1-\rho)^2}{\sigma^2\left(1+b_1^2\right)} + a_2^2\frac{(1+\rho^2)\sigma_x^2}{\sigma^2\left(1+b_1^2\right)} + a_2^2\frac{1}{\left(1+b_1^2\right)^2}.$$

For any $\varepsilon > 0$,

$$\lim_{n\to\infty}\frac{1}{n-1}\sum_{t=2}^{n}\int_{|aZ_t|>\varepsilon\sqrt{n-1}}(aZ_t)^2 dF_t(aZ_t)$$

$$= \lim_{n\to\infty}\frac{1}{n-1}\sum_{t=2}^{n} E\left\{(aZ_t)^2 I\left(|aZ_t| > \varepsilon\sqrt{n-1}\right)\right\}$$

$$\le \lim_{n\to\infty} max\left\{(aZ_t)^2 I\left(|aZ_t| > \varepsilon\sqrt{n-1}\right)\right\} = 0,$$

conditions (2.5) and (2.6) are sufficient to ensure the inequality holds and then the result is followed by dominated convergence theorem since $E(aZ_t)^2 < \infty$.

Therefore, by theorem 3.4 in section 3, we have

$$\frac{\sum_{t=2}^{n} aZ_t}{n-1} \xrightarrow{d} N\left(0, \frac{(a_1+a_2\mu_x)^2\frac{(1-\rho)^2}{\sigma^2\left(1+b_1^2\right)} + a_2^2\frac{(1+\rho^2)\sigma_x^2}{\sigma^2\left(1+b_1^2\right)} + a_2^2\frac{1}{\left(1+b_1^2\right)^2}}{n-1}\right) = aZ,$$

where

$$Z = N\left[\begin{pmatrix}0\\0\end{pmatrix}, \frac{1}{\sigma^2\left(1+b_1^2\right)}\begin{pmatrix}(1-\rho)^2 & (1-\rho)^2\mu_x \\ (1-\rho)^2\mu_x & (1+\rho^2)\sigma_x^2 + (1-\rho)^2\mu_x^2 + \frac{\sigma^2}{1+b_1^2}\end{pmatrix}\right].$$

Then, by Cramer-Wold theorem,

$$\frac{\sum_{t=2}^{n} Z_t}{n-1} \xrightarrow{d} Z.$$

Therefore,

$$plim\frac{U_\beta(\ddot{\varphi})}{\sqrt{n-1}} = plim\left(\frac{(1-\ddot{\rho})}{\ddot{\sigma}^2\sqrt{n-1}}\sum_{t=2}^{n}(Y_t - \ddot{\gamma}_t - \ddot{\rho}Y_{t-1}), \frac{1}{\ddot{\sigma}^2\sqrt{n-1}}\sum_{t=2}^{n}\{(Y_t - \ddot{\gamma}_t - \ddot{\rho}Y_{t-1})\ddot{\lambda}_t\}\right)^T$$

$$= 2plim\left((1-\rho)\sqrt{\frac{1}{\sigma^2(1+b_1^2)}}Z_1, \sqrt{\frac{1}{\sigma^2(1+b_1^2)}}Z_2 + \frac{1}{1+b_1^2}Z_3\right)^T$$

$$\xrightarrow{d} N\left[\begin{pmatrix}0\\0\end{pmatrix}, \frac{4}{\sigma^2(1+b_1^2)}\begin{pmatrix}(1-\rho)^2 & (1-\rho)^2\mu_x \\ (1-\rho)^2\mu_x & (1+\rho^2)\sigma_x^2 + (1-\rho)^2\mu_x^2 + \frac{\sigma^2}{1+b_1^2}\end{pmatrix}\right].$$

We have looked into the property of the score vector in the test statistic. Now we will consider the terms in the information matrix. There are two summation terms in that matrix. For the average of the nuisance variable, we have

$$plim\frac{\sum_{t=2}^{n}\ddot{\lambda}_t}{n-1} = plim\frac{\sum_{t=2}^{n}\left(x_t - \rho x_{t-1} + \frac{v_t + b_1\delta_t}{1+b_1^2}\right)}{n-1} = (1-\rho)\mu_x.$$

For the second sample moments of the nuisance variable, we have

$$plim\frac{\sum_{t=2}^{n}\ddot{\lambda}_t^2}{n-1} = plim\frac{\sum_{t=2}^{n}\left(x_t - \rho x_{t-1} + \frac{v_t + b_1\delta_t}{1+b_1^2}\right)^2}{n-1}$$

$$= plim\frac{\sum_{t=2}^{n}\left((x_t - \rho x_{t-1})^2 + 2(x_t - \rho x_{t-1})\frac{v_t + b_1\delta_t}{1+b_1^2} + \left(\frac{v_t + b_1\delta_t}{1+b_1^2}\right)^2\right)}{n-1}$$

$$= (1+\rho^2)\sigma_x^2 + (1-\rho)^2\mu_x^2 + \frac{\sigma^2}{1+b_1^2}.$$

Therefore,

$$plim\frac{I(\ddot{\varphi})_{2\times2}}{(n-1)} = plim\frac{1}{\ddot{\sigma}^2(1+b_1^2)}\begin{bmatrix}(1-\ddot{\rho})^2 & (1-\ddot{\rho})\frac{\sum_{t=2}^{n}\ddot{\lambda}_t}{(n-1)} \\ (1-\ddot{\rho})\frac{\sum_{t=2}^{n}\ddot{\lambda}_t}{(n-1)} & \frac{\sum_{t=2}^{n}\ddot{\lambda}_t^2}{(n-1)}\end{bmatrix}$$

$$= \frac{2}{\sigma^2(1+b_1^2)}\begin{bmatrix}(1-\rho)^2 & (1-\rho)^2\mu_x \\ (1-\rho)^2\mu_x & (1+\rho^2)\sigma_x^2 + (1-\rho)^2\mu_x^2 + \frac{\sigma^2}{1+b_1^2}\end{bmatrix}.$$

Thus, for the test statistic,

$$\text{plim}T_{ES,f,\beta} = \text{plim}\, U_\beta(\ddot{\varphi})^T/(n-1)\,(I(\ddot{\varphi})_{2\times 2}/(n-1))^{-1}\, U_\beta(\ddot{\varphi})/(n-1)$$

$$= 2\text{plim}\begin{pmatrix} \sqrt{\frac{\sigma^2}{1+b_1^2}}Z_1 \\ \sqrt{\frac{\sigma^2}{1+b_1^2}}Z_2 + \frac{\sigma^2}{1+b_1^2}Z_3 \end{pmatrix}^T \Sigma_\beta^{-1} \begin{pmatrix} \sqrt{\frac{\sigma^2}{1+b_1^2}}Z_1 \\ \sqrt{\frac{\sigma^2}{1+b_1^2}}Z_2 + \frac{\sigma^2}{1+b_1^2}Z_3 \end{pmatrix},$$

where

$$\Sigma_\beta^{-1} = \frac{4}{(1+b_1^2)\sigma^2}\begin{pmatrix} (1-\rho)^2 & (1-\rho)^2\mu_x \\ (1-\rho)^2\mu_x & (1+\rho^2)\sigma_x^2 + (1-\rho)^2\mu_x^2 + \frac{\sigma^2}{1+b_1^2} \end{pmatrix}$$

which is the same with the covariance matrix for $\text{plim}\frac{U_\beta(\ddot{\varphi})}{\sqrt{n-1}}$.

Therefore, by Theorem 3.9 in Section 3,

$$T_{ES,f,\beta} \xrightarrow{d} 2\chi_2^2.$$

After the detailed derivation, it is found that the efficient score test statistic for testing $H_0: (\beta_0, \beta_1) = (b_0, b_1)$ under the null hypothesis has asymptotic distribution as two times a chi-square distribution with two degrees of freedom. The coefficient of two before the chi-square distribution is probably due to the inconsistent estimate of σ^2 that converges to one half of the true value.

4.3.4 Likelihood ratio test statistic under $H_0: (\beta_0, \beta_1) = (b_0, b_1)$

This is the most complicated situation since there is no explicit expression for the maximum likelihood estimator with no constraint. Furthermore, even under the null hypothesis, we still face the fact that the correlation coefficient ρ is non-zero and the data are dependent. We will investigate the large sample distribution using the technique of Taylor expansion in the same way with that for test statistic of testing $\rho = 0$ as shown in section 4.3.2.

All the maximum likelihood estimators and the test statistic can be implicitly or explicitly expressed as a function of $\hat{\psi}$ defined in section 4.3.2. Therefore, firstly, we consider the first order derivatives,

$$\frac{\partial T_{LRT,f,\beta}}{\partial \hat{\psi}} = -2(n-1)\left(\frac{1}{\hat{\sigma}^2}\frac{\partial \hat{\sigma}^2}{\partial \hat{\psi}} - \frac{1}{\ddot{\sigma}^2}\frac{\partial \ddot{\sigma}^2}{\partial \hat{\psi}}\right).$$

The first order derivatives of $\ddot{\sigma}^2$ are

$$\frac{\partial \ddot{\sigma}^2}{\partial \hat{\psi}} = \left(\frac{b_1\hat{\mu}_Z(1-\ddot{\rho})^2}{1+b_1^2}, -\frac{\hat{\mu}_Z(1-\ddot{\rho})^2}{1+b_1^2}, \frac{b_1^2(1+\ddot{\rho}^2)}{2(1+b_1^2)}, \frac{(1+\ddot{\rho}^2)}{2(1+b_1^2)}, -\frac{b_1(1+\ddot{\rho}^2)}{1+b_1^2}, \frac{b_1\ddot{\rho}}{1+b_1^2}, \frac{b_1\ddot{\rho}}{1+b_1^2}, -\frac{b_1^2\ddot{\rho}}{(1+b_1^2)}, -\frac{\ddot{\rho}}{1+b_1^2}\right)$$

where

$$\hat{\mu}_Z = b_0 + b_1\hat{\mu}_X - \hat{\mu}_Y.$$

However, the first order derivative of $\hat{\sigma}^2$ can only be obtained using the method for implicit function. It can be expressed as

$$\frac{\partial \hat{\sigma}^2}{\partial \hat{\psi}} = -\frac{2\hat{\beta}_1}{1+\hat{\beta}_1^2}\frac{\partial \hat{\beta}_1}{\partial \hat{\psi}}\hat{\sigma}^2 + \frac{1}{2(1+\hat{\beta}_1^2)}\left\{(1+\hat{\rho}^2)\frac{\partial \hat{D}_f}{\partial \hat{\psi}} - 2\hat{\rho}\frac{\partial \hat{N}_f}{\partial \hat{\psi}}\right\}$$

where $\hat{D}_f$ and $\hat{N}_f$ are defined in section 4.3.2. The way to calculate the first order derivative of $\hat{\beta}_1$ is also described in section 4.3.2. The first order derivatives for $\hat{D}_f$ and $\hat{N}_f$ are

$$\frac{\partial \hat{D}_f}{\partial \hat{\psi}} = 2\left(\hat{\beta}_1\hat{\mu}_{X^2} - \hat{\mu}_{XY}\right)\frac{\partial \hat{\beta}_1}{\partial \hat{\psi}} + \left(0,0,\hat{\beta}_1^2,1,-2\hat{\beta}_1,0,0,0,0\right),$$

$$\frac{\partial \hat{N}_f}{\partial \hat{\psi}} = \left\{2\hat{\beta}_1\hat{\mu}_{XX} - (\hat{\mu}_{XY1} + \hat{\mu}_{X1Y})\right\}\frac{\partial \hat{\beta}_1}{\partial \hat{\psi}} + \left(0,0,0,0,0,-\hat{\beta}_1,-\hat{\beta}_1,\hat{\beta}_1^2,1\right).$$

The second order derivative of the test statistic is,

$$\frac{\partial^2 T_{LRT,f,\beta}}{\partial \hat{\psi}^T\partial \hat{\psi}} = -2(n-1)\left(-\frac{1}{\hat{\sigma}^4}\frac{\partial \hat{\sigma}^2}{\partial \hat{\psi}^T}\frac{\partial \hat{\sigma}^2}{\partial \hat{\psi}} + \frac{1}{\ddot{\sigma}^4}\frac{\partial \ddot{\sigma}^2}{\partial \hat{\psi}^T}\frac{\partial \ddot{\sigma}^2}{\partial \hat{\psi}} + \frac{1}{\hat{\sigma}^2}\frac{\partial^2 \hat{\sigma}^2}{\partial \hat{\psi}^T\partial \hat{\psi}} - \frac{1}{\ddot{\sigma}^2}\frac{\partial^2 \ddot{\sigma}^2}{\partial \hat{\psi}^T\partial \hat{\psi}}\right).$$

To calculate this, we have to find the second order derivatives for $\ddot{\sigma}^2$ and $\hat{\sigma}^2$ first. In section 4.3.2, we have shown how to obtain the second order derivatives of $\hat{\sigma}^2$. Now we focus on the second order derivatives of $\ddot{\sigma}^2$. Let

$$\ddot{N}_f = b_1^2\hat{\mu}_{XX} - b_1(\hat{\mu}_{XY1} + \hat{\mu}_{X1Y}) + \hat{\mu}_{YY} + \hat{\mu}_Z^2,$$

$$\ddot{D}_f = b_1^2\hat{\mu}_{X^2} + \hat{\mu}_{Y^2} - 2b_1\hat{\mu}_{XY} + \hat{\mu}_Z^2.$$

Then $\ddot{\rho}$ can be expressed as

$$\ddot{\rho} = \ddot{N}_f/\ddot{D}_f.$$

After some derivations, the second order derivatives of $\ddot{\sigma}^2$ would be

$$\frac{\partial^2\ddot{\sigma}^2}{\partial\hat{\psi}^T\partial\hat{\psi}} = \begin{pmatrix} \ddot{A} & \ddot{b}^T\ddot{a} \\ \ddot{a}^T\ddot{b} & -\frac{1}{1+b_1^2}\frac{1}{\ddot{D}_f}\ddot{a}^T\ddot{a} \end{pmatrix}$$

where

$$\ddot{A}_f = \left(-\frac{4(1-\ddot{\rho})^2}{1+b_1^2}\frac{\hat{\mu}_Z^2}{\ddot{D}} + \frac{(1-\ddot{\rho})^2}{1+b_1^2}\right)\begin{pmatrix} b_1^2 & -b_1 \\ -b_1 & 1 \end{pmatrix},$$

$$\ddot{a} = (-b_1^2\ddot{\rho}, -\ddot{\rho}, 2b_1\ddot{\rho}, -b_1, -b_1, b_1^2, 1),$$

$$\ddot{b} = \frac{2(1-\ddot{\rho})}{1+b_1^2}\frac{\hat{\mu}_Z}{\ddot{D}_f}(-b_1, 1).$$

Under the null hypothesis, we know that $(\beta_0, \beta_1) = (b_0, b_1)$, and

$$\hat{\psi} \to \psi_\beta = \left(\mu_x, b_0 + b_1\mu_x, \sigma_x^2 + \frac{\sigma^2}{1-\rho^2}, b_1^2\sigma_x^2 + \frac{\sigma^2}{1-\rho^2}, b_1\sigma_x^2, 0, 0, \rho\frac{\sigma^2}{1-\rho^2}, \rho\frac{\sigma^2}{1-\rho^2}\right).$$

Then the first order derivatives of $\ddot{\sigma}^2$ under the null hypothesis at ψ_β would be

$$\left.\frac{\partial\ddot{\sigma}^2}{\partial\hat{\psi}}\right|_{\psi_\beta} = \left(0, 0, \frac{b_1^2(1+\rho^2)}{2(1+b_1^2)}, \frac{(1+\rho^2)}{2(1+b_1^2)}, -\frac{b_1(1+\rho^2)}{1+b_1^2}, \frac{b_1\rho}{1+b_1^2}, \frac{b_1\rho}{1+b_1^2}, -\frac{b_1^2\rho}{(1+b_1^2)}, -\frac{\rho}{1+b_1^2}\right).$$

The first order derivatives of $\hat{\sigma}^2$ at ψ_β become

$$\left.\frac{\partial\hat{\sigma}^2}{\partial\hat{\psi}}\right|_{\psi_\beta} = \left(0,0,\frac{b_1^2(1+\rho^2)}{2(1+b_1^2)},\frac{(1+\rho^2)}{2(1+b_1^2)},-\frac{b_1(1+\rho^2)}{1+b_1^2},\frac{b_1\rho}{1+b_1^2},\frac{b_1\rho}{1+b_1^2},-\frac{b_1^2\rho}{(1+b_1^2)},-\frac{\rho}{1+b_1^2}\right).$$

It is obvious that $\left.\frac{\partial\ddot{\sigma}^2}{\partial\hat{\psi}}\right|_{\psi_\beta}$ and $\left.\frac{\partial\hat{\sigma}^2}{\partial\hat{\psi}}\right|_{\psi_\beta}$ are equal. Therefore, the first order derivatives for the test statistic are all zero. That is,

$$\left.\frac{\partial T_{LRT,f,\beta}}{\partial\hat{\psi}}\right|_{\psi_\beta} = 0_{1\times 9}.$$

Furthermore, at ψ_β, we have

$$\left.\ddot{D}_f\right|_{\psi_\beta} = (1+b_1^2)\frac{\sigma^2}{1-\rho^2}.$$

Then the second order derivatives of $\ddot{\sigma}^2$ under the null hypothesis at ψ_β would be

$$\left.\frac{\partial^2\ddot{\sigma}^2}{\partial\hat{\psi}^T\partial\hat{\psi}}\right|_{\psi_\beta} = \begin{pmatrix} A_{\psi_\beta} & 0_{2\times 7} \\ 0_{7\times 2} & a_\beta^T b_\beta \end{pmatrix},$$

where

$$A_{\psi_\beta} = \frac{b_1}{1+b_1^2}(1-\rho)^2\begin{pmatrix} b_1 & -1 \\ -1 & 1 \end{pmatrix},$$

$$a_\beta = -\frac{1}{\left(1+b_1^2\right)^2}\frac{1-\rho^2}{\sigma^2}b_\beta,$$

$$b_\beta = (-b_1^2\rho, -\rho, 2b_1\rho, -b_1, -b_1, b_1^2, 1).$$

Though complicated, we can derive the second order derivatives of F_f, $\widehat{D}_f$, $\widehat{N}_f$, and $\hat{\beta}_1^2$. And then, the second order derivatives of $\hat{\sigma}^2$ at ψ_β can be obtained as

$$\left.\frac{\partial^2\hat{\sigma}^2}{\partial\hat{\psi}^T\partial\hat{\psi}}\right|_{\psi_\beta} = \begin{pmatrix} 0_{2\times 2} & 0_{2\times 7} \\ 0_{7\times 2} & B_{\psi_\beta} \end{pmatrix},$$

where

$$B_{\psi_\beta} = c_\beta^T d_\beta + a_\beta^T b_\beta.$$

In the definition of B_{ψ_β}, a_β and b_β have been defined above. And

$$d_\beta = \left(-b_1, b_1, (1-b_1^2), -\frac{(1-b_1^2)\rho}{(1+\rho^2)}, -\frac{(1-b_1^2)\rho}{(1+\rho^2)}, \frac{2b_1\rho}{(1+\rho^2)}, -\frac{2b_1\rho}{(1+\rho^2)}\right),$$

$$c_\beta = -\frac{(1+\rho^2)}{\left(1+b_1^2\right)^3 \sigma_x^2} d_\beta.$$

Then substituting the second order derivatives of $\ddot{\sigma}^2$ and $\hat{\sigma}^2$ at ψ_β, the second order derivatives of the test statistic at ψ_β would be

$$\left.\frac{\partial^2 T_{LRT,f,\beta}}{\partial \hat{\psi}^T \partial \hat{\psi}}\right|_{\psi_\beta} = 4(n-1)\frac{1}{\sigma^2}\begin{pmatrix} A_{\psi_\beta} & 0_{2\times 7} \\ 0_{7\times 2} & -c_\beta^T d_\beta \end{pmatrix}.$$

Therefore, the second order Taylor expansion of the test statistic would be

$$T_{LRT,f,\beta} \doteq T_{LRT,f,\beta}\big|_{\psi_\beta} + \left.\frac{\partial T_{LRT,f,\beta}}{\partial \hat{\psi}}\right|_{\psi_\beta} \left(\hat{\psi} - \psi_\beta\right)^T + \frac{1}{2}\left(\hat{\psi} - \psi_\beta\right) \left.\frac{\partial^2 T_{LRT,f,\beta}}{\partial \hat{\psi}^T \partial \hat{\psi}}\right|_{\psi_\beta} \left(\hat{\psi} - \psi_\beta\right)^T$$

$$= 2(n-1)\frac{1}{\sigma^2}\left(\hat{\psi} - \psi_\beta\right)\begin{pmatrix} A_{\psi_\beta} & 0_{2\times 7} \\ 0_{7\times 2} & -c_\beta^T d_\beta \end{pmatrix}\left(\hat{\psi} - \psi_\beta\right)^T.$$

To apply Theorem 3.8 in Section 3, we need to make some variable transformations first. Let

$$U_1 = \sqrt{1-\rho^2} X_1,$$

$$U_t = X_t - \rho X_{t-1} = x_t - \rho x_{t-1} + \upsilon_t \qquad t = 2,3,\dots,n,$$

and

$$V_1 = \sqrt{1-\rho^2} V_1,$$

$$V_t = Y_t - \rho Y_{t-1} = y_t - \rho y_{t-1} + \delta_t \qquad t = 2,3,\dots,n.$$

Furthermore, let

$$Z_{\beta t} = (U_t, V_t),$$

$$z_{\beta t} = (x_t - \rho x_{t-1}, y_t - \rho y_{t-1}),$$

$$\varepsilon_{\beta t} = (\upsilon_t, \delta_t).$$

Then

$$Z_{\beta t} = z_{\beta t} + \varepsilon_{\beta t},$$

and $Z_{\beta t}$ is an independent random vector. The sample covariance matrix of $Z_{\beta t}$ becomes

$$m_{\beta zz} = \frac{1}{n-1}\begin{pmatrix} \sum_{t=2}^{n}(U_t - \bar{U})^2 & \sum_{t=2}^{n}(U_t - \bar{U})(V_t - \bar{V}) \\ \sum_{t=2}^{n}(U_t - \bar{U})(V_t - \bar{V}) & \sum_{t=2}^{n}(V_t - \bar{V})^2 \end{pmatrix}.$$

Let the sample mean and sample covariance be given as follows,

$$\hat{\theta}_\beta = \left(\hat{\theta}_{\beta 1}^T, \hat{\theta}_{\beta 2}^T\right)^T,$$

where

$$\hat{\theta}_{\beta 1} = (\bar{U}, \bar{V})^T,$$

$$\hat{\theta}_{\beta 2} = \left(\frac{1}{n-1}\sum_{t=2}^{n}(U_t - \bar{U})^2, \frac{1}{n-1}\sum_{t=2}^{n}(U_t - \bar{U})(V_t - \bar{V}), \frac{1}{n-1}\sum_{t=2}^{n}(V_t - \bar{V})^2\right)^T.$$

Following Theorem 3.8 in Section 3, we know that the new vector $\hat{\theta}_\beta$ is asymptotically distributed as a normal distribution with mean

$$\theta_{\beta n} = \left(\theta_{\beta n1}^T, \theta_{\beta n2}^T\right)^T,$$

where

$$\theta_{\beta n1} = (\bar{x}_n - \rho\bar{x}_{n-1}, \bar{y}_n - \rho\bar{y}_{n-1})^T,$$

$$\theta_{\beta n2} = \begin{pmatrix} \frac{\sum_{t=2}^{n}(x_t - \rho x_{t-1} - \bar{x}_n + \rho\bar{x}_{n-1})^2}{n-1} + \sigma^2 \\ \frac{\sum_{t=2}^{n}(x_t - \rho x_{t-1} - \bar{x}_n + \rho\bar{x}_{n-1})(y_t - \rho y_{t-1} - \bar{y}_n + \rho\bar{y}_{n-1})}{n-1} \\ \frac{\sum_{t=2}^{n}(y_t - \rho y_{t-1} - \bar{y}_n + \rho\bar{y}_{n-1})^2}{n-1} + \sigma^2 \end{pmatrix},$$

and covariance matrix

$$G_{\beta n} = \begin{pmatrix} G_{\beta n1} & 0_{2\times 3} \\ 0_{3\times 2} & G_{\beta n2} \end{pmatrix},$$

where

$$G_{\beta n1} = \frac{1}{n-1}\sigma^2 I_{2\times 2},$$

$$G_{\beta n2} = \frac{4\sigma^2}{n-1}\begin{pmatrix} \frac{\sum_{t=2}^{n}(x_t-\bar{x}_n)^2}{n-1}+\sigma^2 & \frac{\sum_{t=2}^{n}(x_t-\bar{x}_n)(y_t-\bar{y}_n)}{2(n-1)} & 0 \\ \frac{\sum_{t=2}^{n}(x_t-\bar{x}_n)(y_t-\bar{y}_n)}{2(n-1)} & \frac{\sum_{t=2}^{n}\left[(x_t-\bar{x}_n)^2+(y_t-\bar{y}_n)^2\right]}{4(n-1)}+\frac{\sigma^2}{4} & \frac{\sum_{t=2}^{n}(x_t-\bar{x}_n)(y_t-\bar{y}_n)}{2(n-1)} \\ 0 & \frac{\sum_{t=2}^{n}(x_t-\bar{x}_n)(y_t-\bar{y}_n)}{2(n-1)} & \frac{\sum_{t=2}^{n}(y_t-\bar{y}_n)^2}{n-1}+\sigma^2 \end{pmatrix}.$$

It is obvious that there is some relationships between $\hat{\theta}_\beta$ and $\hat{\psi}$ and we can make a transformation to obtain $\hat{\theta}_\beta$ from $\hat{\psi}$. That is,

$$\hat{\theta}_\beta = D_\beta \hat{\psi},$$

where

$$D_\beta = \begin{pmatrix} 1-\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1+\rho^2 & 0 & 0 & 0 & 0 & -2\rho & 0 \\ 0 & 0 & 0 & 0 & 1+\rho^2 & -\rho & -\rho & 0 & 0 \\ 0 & 0 & 0 & 1+\rho^2 & 0 & 0 & 0 & 0 & -2\rho \end{pmatrix}.$$

One of the generalized inverse matrices of D that satisfies the equality of $D_\beta D_{\beta I} D_\beta = D_\beta$ is

$$D_{\beta I} = \begin{pmatrix} D_{\beta I1} & 0_{3\times 2} \\ 0_{2\times 3} & D_{\beta I2} \\ 0_{4\times 3} & 0_{4\times 2} \end{pmatrix},$$

where

$$D_{\beta I1} = \begin{pmatrix} (1-\rho)^{-1} & 0 & 0 \\ 0 & (1-\rho)^{-1} & 0 \\ 0 & 0 & (1+\rho^2)^{-1} \end{pmatrix},$$

$$D_{\beta I2} = \begin{pmatrix} 0 & (1+\rho^2)^{-1} \\ (1+\rho^2)^{-1} & 0 \end{pmatrix}.$$

Therefore,

$$B_\beta = D_{\beta I}^T \frac{\partial^2 T_{LRT,f,\beta}}{\partial \hat{\psi}^T \partial \hat{\psi}}\bigg|_{\psi_\beta} D_{\beta I} = \begin{pmatrix} B_{\beta 11} & 0_{2\times 3} \\ 0_{3\times 2} & B_{\beta 22} \end{pmatrix},$$

where

$$B_{\beta 11} = 4(n-1)\frac{1}{(1-\rho)^2\sigma^2} A_{\psi_\beta},$$

$$B_{\beta 22} = 4(n-1)\frac{1}{\sigma^2}\frac{1}{\left(1+b_1^2\right)^3 \sigma_x^2 (1+\rho^2)} \begin{pmatrix} b_1^2 & -b_1(1-b_1^2) & -b_1^2 \\ -b_1(1-b_1^2) & (1-b_1^2)^2 & b_1(1-b_1^2) \\ -b_1^2 & b_1(1-b_1^2) & b_1^2 \end{pmatrix}.$$

After some calculations, we find that

$$\frac{\partial^2 T_{LRT,f,\beta}}{\partial \hat{\psi}^T \partial \hat{\psi}}\bigg|_{\psi_\beta} = D_\beta^T B_\beta D_\beta.$$

Substituting these equations into the Taylor expansion of the test statistic, we have the expansion after variable transformation,

$$T_{LRT,f,\beta} \doteq \frac{1}{2}\left(\hat{\psi} - \psi_\beta\right) \frac{\partial^2 T_{LRT,f,\beta}}{\partial \hat{\psi}^T \partial \hat{\psi}}\bigg|_{\psi_\beta} \left(\hat{\psi} - \psi_\beta\right)^T = \frac{1}{2}\left(\hat{\psi} - \psi_\beta\right) D_\beta^T B_\beta D_\beta \left(\hat{\psi} - \psi_\beta\right)^T$$

$$= \frac{1}{2}\left(\hat{\theta}_\beta - \theta_\beta\right) B_\beta \left(\hat{\theta}_\beta - \theta_\beta\right)^T$$

$$= \frac{1}{2}\left(\hat{\theta}_{\beta 1} - \theta_{\beta 1}\right) B_{\beta 11} \left(\hat{\theta}_{\beta 1} - \theta_{\beta 1}\right)^T + \frac{1}{2}\left(\hat{\theta}_{\beta 2} - \theta_{\beta 2}\right) B_{\beta 22} \left(\hat{\theta}_{\beta 2} - \theta_{\beta 2}\right)^T.$$

Then the Taylor expansion of the test statistic becomes a summation of two parts. We will consider the two parts separately. First of all, from the structure of the covariance matrix, we can see that the mean of $Z_{\beta t}$, $\hat{\theta}_{\beta 1}$, and the covariance elements, $\hat{\theta}_{\beta 2}$, are independent. Furthermore, since $\hat{\theta}_\beta - \theta_\beta$ is asymptotically normally distributed with mean zero and covariance matrix $G_{\beta n}$, then $\hat{\theta}_{\beta 1} - \theta_{\beta 1}$ and $\hat{\theta}_{\beta 2} - \theta_{\beta 2}$ are asymptotically independently normally distributed with mean zero,

$$\hat{\theta}_{\beta 1} - \theta_{\beta 1} \to AN(0, G_{\beta n1}),$$

$$\hat{\theta}_{\beta 2} - \theta_{\beta 2} \to AN(0, G_{\beta n2}).$$

For the first term in the expansion, we notice that,

$$\frac{B_{\beta 11}}{4} G_{\beta n1} \frac{B_{\beta 11}}{4} = \frac{B_{\beta 11}}{4}.$$

Therefore, according to Theorem 3.9 in Section 3, $\frac{1}{4}(\hat{\theta}_{\beta 1} - \theta_{\beta 1}) B_{\beta 11} (\hat{\theta}_{\beta 1} - \theta_{\beta 1})^T$ converges to a chi-squared distribution with one degree of freedom. The reason why the degree of freedom is one is that the rank of $B_{\beta 11} G_{\beta n1}$ is one. For the second term in the expansion, we have

$$\frac{B_{\beta 22}}{4k} G_{\beta n2} \frac{B_{\beta 22}}{4k} = \frac{B_{\beta 22}}{4k},$$

where

$$k = \frac{1}{(1+b_1^2)(1+\rho^2)\sigma_x^2} \left\{ \frac{\sum_{t=2}^{n} [(x_t - \bar{x}_n)^2 + (y_t - \bar{y}_n)^2]}{(n-1)} + \sigma^2 + \frac{4b_1^2}{(1+b_1^2)} \sigma^2 \right\}.$$

Therefore, following Theorem 3.9, we have $\frac{1}{4k}(\hat{\theta}_{\beta 2} - \theta_{\beta 2}) B_{\beta 22} (\hat{\theta}_{\beta 2} - \theta_{\beta 2})^T$ asymptotically distributed as a chi-squared random variable with one degree of freedom. It is one degree of freedom because the rank of $B_{\beta 22} G_{\beta n2}$ is one. Then the test statistic is a summation of two independent chi-squared random variables with both having one degree of freedom. Let

$$T_{1f} = \frac{1}{4}(\hat{\theta}_{\beta 1} - \theta_{\beta 1}) B_{\beta 11} (\hat{\theta}_{\beta 1} - \theta_{\beta 1})^T,$$

$$T_{2f} = \frac{1}{4k}(\hat{\theta}_{\beta 2} - \theta_{\beta 2}) B_{\beta 22} (\hat{\theta}_{\beta 2} - \theta_{\beta 2})^T.$$

Then,

$$T_{LRT,f,\beta} \doteq 2(\hat{\theta}_{\beta 1} - \theta_{\beta 1})\frac{B_{\beta 11}}{4}(\hat{\theta}_{\beta 1} - \theta_{\beta 1})^T + 2k(\hat{\theta}_{\beta 2} - \theta_{\beta 2})\frac{B_{\beta 22}}{4k}(\hat{\theta}_{\beta 2} - \theta_{\beta 2})^T$$

$$= 2T_{1f} + 2kT_{2f}.$$

Notice that the two chi-squared distributed random variables, T_{1f} and T_{2f} are multiplied by constants which are not equal. Thus, using the Satterthwaite approximation, we know that $\frac{\nu T_{LRT,f,\beta}}{E(T_{LRT,f,\beta})}$ is asymptotically distributed as a chi-squared random variable with degree of freedom equal to

$$\nu = \frac{(2T_{1f}+2kT_{2f})^2}{(2T_{1f})^2+(2kT_{2f})^2}.$$

Since the value of degree of freedom depends on the variance estimators, it makes this scenario complicated to investigate and the simulation result is not ideal especially when the absolute value of the correlation coefficient is large.

5. SIMULATION STUDIES

Simulation studies are used to examine the small sample behavior of our estimators and test statistics. The results presented are a representative subset of results taken from a much larger simulation study. Section 5.1 investigates the empirical biases of parameter estimators for both structural and function models as the ratio σ_x^2/σ^2 varies. Abdul-Salam (1996) showed that when $\rho = 0$ this ratio played a crucial role in the small sample behavior of parameter estimators and test statistics in the structural and functional models. Assessment of sample sizes needed to achieve adequate approximations to the distributions of our test statistics by the limiting null distributions is presented in Section 5.2. Empirical type I error rates are studied in Section 5.3. Power studies conclude the empirical investigation of our test statistics in Section 5.4.

5.1 Empirical Bias and Consistency of Maximum Likelihood Estimates

5.1.1 Empirical bias and consistency under the structural model

Simulated trivariate normal data $(x_t, \delta_t, \upsilon_t)^T$ with mean $(\mu_x, 0,0)^T$ and diagonal covariance matrix $diag(\sigma_x^2, \sigma^2, \sigma^2)$ are generated using the R package MASS. Values of $e_t = \rho e_{t-1} + \delta_t$ and $u_t = \rho u_{t-1} + \upsilon_t$ are calculated and used to compute the simulated (X_t, Y_t) values, with $X_t = x_t + e_t$ and $Y_t = \beta_0 + \beta_1 x_t + u_t = y_t + u_t$. Empirical biases of parameter estimators are investigated for varying sample sizes given in tables 5.1-5.7. Three sets of parameter conditions are considered: 1) $\rho = 0$; 2) $(\beta_0, \beta_1) = (b_0, b_1)$, where (b_0, b_1) is known; and, 3) no parameter constraints. The number of simulated data sets is $N = 2000$ for each parameter configuration listed below. Empirical bias is calculated as the true parameter value subtracted from the sample mean of the $N = 2000$ parameter estimates.

Table 5.1 Consistency for MLE under $\rho = 0$ with $(\beta_0, \beta_1, \sigma^2, \mu_x) = (0, 1, 4, 1)$ in structural model

Sample Size	Parameter Value	$\tilde{\mu}_x$	$\tilde{\sigma}_x^2$	$\tilde{\sigma}^2$	$\tilde{\beta}_0$	$\tilde{\beta}_1$
n=10	$\sigma_x^2 = 1$	-0.0123 (0.0171)	0.9903 (0.0449)	-1.5773 (0.0279)	-0.9419 (1.4852)	1.8207 (1.7709)
	$\sigma_x^2 = 4$	0.0018 (0.0209)	0.3345 (0.0772)	-1.1871 (0.0322)	-0.8198 (0.4467)	0.4886 (0.3397)
	$\sigma_x^2 = 16$	-0.0326 (0.0332)	-1.2217 (0.1934)	-0.9099 (0.0375)	-0.0735 (0.0386)	0.0472 (0.0181)
n=20	$\sigma_x^2 = 1$	-0.0061 (0.0113)	0.5832 (0.0309)	-0.8542 (0.0226)	1.0764 (2.2048)	-0.5804 (1.8951)
	$\sigma_x^2 = 4$	-0.0209 (0.0141)	0.1096 (0.0544)	-0.5511 (0.0272)	-0.1157 (0.0455)	0.1008 (0.0424)
	$\sigma_x^2 = 16$	-0.0006 (0.0227)	-0.5108 (0.1414)	-0.4549 (0.0270)	-0.0187 (0.0154)	0.0165 (0.0045)
n=30	$\sigma_x^2 = 1$	0.0031 (0.0093)	0.3854 (0.0254)	-0.6060 (0.0191)	0.3325 (0.4370)	-0.3735 (0.4057)
	$\sigma_x^2 = 4$	-0.0180 (0.0116)	0.0744 (0.0448)	-0.3407 (0.0223)	-0.0755 (0.0206)	0.0943 (0.0153)
	$\sigma_x^2 = 16$	-0.0160 (0.0187)	-0.3431 (0.1158)	-0.2804 (0.0232)	0.0059 (0.0127)	0.0063 (0.0035)
n=50	$\sigma_x^2 = 1$	-0.0051 (0.0073)	0.2591 (0.0206)	-0.3471 (0.0160)	0.1287 (0.4676)	-0.1881 (0.4731)
	$\sigma_x^2 = 4$	0.0128 (0.0090)	0.0593 (0.0349)	-0.2288 (0.0178)	-0.0428 (0.0117)	0.0370 (0.0065)
	$\sigma_x^2 = 16$	0.0094 (0.0137)	-0.2612 (0.0892)	-0.1630 (0.0178)	-0.0061 (0.0094)	0.0034 (0.0025)
n=100	$\sigma_x^2 = 1$	-0.0004 (0.0051)	0.1210 (0.0154)	-0.1603 (0.0118)	-0.0408 (0.4253)	0.0864 (0.3397)
	$\sigma_x^2 = 4$	-0.0028 (0.0066)	-0.0104 (0.0256)	-0.1058 (0.0130)	-0.0246 (0.0079)	0.0218 (0.0043)
	$\sigma_x^2 = 16$	0.0074 (0.0100)	-0.0780 (0.0631)	-0.0833 (0.0126)	-0.0054 (0.0067)	0.0008 (0.0017)
n=500	$\sigma_x^2 = 1$	0.0003 (0.0022)	0.0331 (0.0070)	-0.0435 (0.0057)	-0.0224 (0.0062)	0.0221 (0.0056)
	$\sigma_x^2 = 4$	0.0017 (0.0028)	-0.0187 (0.0114)	-0.0041 (0.0056)	-0.0084 (0.0033)	0.0038 (0.0017)
	$\sigma_x^2 = 16$	-0.0092 (0.0045)	-0.0154 (0.0295)	-0.0179 (0.0057)	0.0018 (0.0029)	-0.0008 (0.0007)
n=1000	$\sigma_x^2 = 1$	0.0009 (0.0016)	0.0133 (0.0049)	-0.0199 (0.0039)	-0.0165 (0.0042)	0.1535 (0.0036)
	$\sigma_x^2 = 4$	0.0033 (0.0020)	-0.0043 (0.0079)	-0.0124 (0.0040)	-0.0053 (0.0023)	0.0023 (0.0013)
	$\sigma_x^2 = 16$	-0.0020 (0.0032)	-0.0089 (0.0202)	-0.0021 (0.0040)	-0.0017 (0.0020)	0.0000 (0.0005)

* In each cell, the numbers are in the form of bias (standard error).

Table 5.2 Consistency for MLE under $\rho = 0$ with $(\beta_0, \beta_1, \sigma^2, \mu_x) = (1, 3, 4, 1)$ in structural model

Sample Size	Parameter Value	$\tilde{\mu}_x$	$\tilde{\sigma}_x^2$	$\tilde{\sigma}^2$	$\tilde{\beta}_0$	$\tilde{\beta}_1$
n=10	$\sigma_x^2 = 1$	0.0254 (0.0167)	0.5854 (0.0384)	-1.1309 (0.0333)	3.0140 (2.0899)	-1.7296 (1.1940)
	$\sigma_x^2 = 4$	-0.0069 (0.0216)	-0.0099 (0.0708)	-0.9002 (0.0362)	1.3810 (1.7351)	-0.8165 (1.1580)
	$\sigma_x^2 = 16$	-0.0435 (0.0337)	-1.2335 (0.1939)	-0.8519 (0.0375)	-0.1481 (0.0681)	0.1696 (0.0224)
n=20	$\sigma_x^2 = 1$	0.0047 (0.0114)	0.2677 (0.0262)	-0.5228 (0.0260)	-0.5260 (1.1511)	0.9116 (0.8526)
	$\sigma_x^2 = 4$	-0.0066(0.0145)	0.0568 (0.0501)	-0.4503 (0.0262)	-0.3309 (0.0823)	0.2802 (0.0523)
	$\sigma_x^2 = 16$	-0.0053 (0.0234)	-0.5174 (0.1425)	-0.4045 (0.0270)	-0.0579 (0.0361)	0.0473 (0.0096)
n=30	$\sigma_x^2 = 1$	0.0084 (0.0095)	0.1182 (0.0204)	-0.3310 (0.0227)	0.5735 (1.2321)	-0.2704 (1.0006)
	$\sigma_x^2 = 4$	0.0057 (0.0117)	0.0297 (0.0397)	-0.2801 (0.0230)	0.0052 (0.1254)	0.0384 (0.1102)
	$\sigma_x^2 = 16$	-0.0162 (0.0188)	-0.4580 (0.1110)	-0.2822 (0.0224)	0.0058 (0.0280)	0.0335 (0.0072)
n=50	$\sigma_x^2 = 1$	0.0099 (0.0070)	0.1003 (0.0159)	-0.2012 (0.0177)	-0.1664 (0.2482)	0.0738 (0.2373)
	$\sigma_x^2 = 4$	0.0155 (0.0090)	-0.0258 (0.0317)	-0.1971 (0.0176)	-0.1099 (0.0244)	0.0892 (0.0125)
	$\sigma_x^2 = 16$	-0.0223 (0.0146)	-0.1960 (0.0885)	-0.1744 (0.0177)	-0.0229 (0.0216)	0.0171 (0.0054)
n=100	$\sigma_x^2 = 1$	-0.0006 (0.0050)	0.0585 (0.0116)	-0.1067 (0.0126)	-0.1962 (0.0298)	0.2113 (0.0262)
	$\sigma_x^2 = 4$	-0.0042 (0.0064)	-0.0092 (0.0225)	-0.0884 (0.0123)	-0.0174 (0.0167)	0.0342 (0.0079)
	$\sigma_x^2 = 16$	0.0147 (0.0103)	-0.0650 (0.0625)	-0.0927 (0.0125)	-0.0137 (0.0150)	0.0041 (0.0037)
n=500	$\sigma_x^2 = 1$	0.0028 (0.0022)	0.0153 (0.0052)	-0.0208 (0.0058)	-0.0416 (0.0102)	0.0323 (0.0081)
	$\sigma_x^2 = 4$	-0.0052(0.0029)	0.0114 (0.0102)	-0.0207 (0.0056)	0.0079 (0.0070)	0.0016 (0.0034)
	$\sigma_x^2 = 16$	0.0042 (0.0045)	-0.0533 (0.0285)	-0.0167 (0.0055)	0.0026 (0.0064)	0.0024 (0.0016)
n=1000	$\sigma_x^2 = 1$	-0.0019 (0.0016)	0.0040 (0.0035)	-0.0130 (0.0040)	-0.0161 (0.0072)	0.0190 (0.0055)
	$\sigma_x^2 = 4$	0.0003 (0.0020)	-0.0033 (0.0072)	-0.0096 (0.0040)	-0.0018 (0.0051)	0.0051 (0.0024)
	$\sigma_x^2 = 16$	0.0023 (0.0031)	-0.0288 (0.0196)	-0.0047 (0.0040)	-0.0044 (0.0046)	0.0010 (0.0011)

* In each cell, the numbers are in the form of empirical bias (standard error).

In tables 5.1 and Table 5.2, both the empirical bias and the standard error of the empirical bias for each estimator decrease as sample size increases. As the ratio of σ_x^2/σ^2 increases, empirical biases of the estimators for μ_x, σ_x^2, and σ^2 tend to increase somewhat, whereas the empirical biases and standard errors for intercept and slope estimators tend to decrease. Except in cases of substantial measurement error, viz., $\sigma_x^2/\sigma^2 = 0.25$ or $\sigma_x^2/\sigma^2 = 1$, empirical biases fall well within two standard errors of zero for sample sizes greater than $n = 30$. In general, the values of intercept and slope appear to have little effect on the empirical biases of the estimators.

Table 5.3 Consistency for MLE under $(\beta_0, \beta_1) = (0, 1)$ with $(\sigma^2, \mu_x, \rho) = (4, 1, 0)$ in structural model

Sample Size	Parameter Value	$\dot{\mu}_x$	$\dot{\sigma}_x^2$	$\dot{\sigma}^2$	$\dot{\rho}$
n=10	$\sigma_x^2 = 1$	0.0084 (0.0129)	0.5282 (0.0273)	-0.9410 (0.0307)	-0.0356 (0.0062)
	$\sigma_x^2 = 4$	0.0031 (0.0181)	-0.3052 (0.0584)	-0.4501 (0.0370)	-0.0436 (0.0066)
	$\sigma_x^2 = 16$	-0.0515 (0.0318)	-2.0202 (0.1850)	-0.2464 (0.0417)	-0.0133 (0.0068)
n=20	$\sigma_x^2 = 1$	-0.0064 (0.0087)	0.2629 (0.0200)	-0.4823 (0.0237)	-0.0185 (0.0046)
	$\sigma_x^2 = 4$	0.0034 (0.0126)	-0.2422 (0.0434)	-0.1581 (0.0289)	-0.0166 (0.0048)
	$\sigma_x^2 = 16$	0.0170 (0.0217)	-1.0942 (0.1255)	-0.1705 (0.0290)	0.0006 (0.0050)
n=30	$\sigma_x^2 = 1$	-0.0046 (0.0073)	0.1938 (0.0175)	-0.3629 (0.0202)	-0.0107 (0.0036)
	$\sigma_x^2 = 4$	-0.0172 (0.0103)	-0.2685 (0.0365)	-0.1119 (0.0232)	-0.0130 (0.0040)
	$\sigma_x^2 = 16$	-0.0002 (0.01772)	-0.5649 (0.1063)	-0.0855 (0.0231)	-0.0010(0.0040)
n=50	$\sigma_x^2 = 1$	-0.0135 (0.0055)	0.0770 (0.0144)	-0.1545 (0.0166)	-0.0073 (0.0029)
	$\sigma_x^2 = 4$	-0.0125 (0.0079)	-0.1118 (0.0286)	-0.0508 (0.0179)	-0.0076 (0.0030)
	$\sigma_x^2 = 16$	-0.0309 (0.0138)	-0.4708 (0.0797)	-0.0438 (0.0178)	-0.0035 (0.0031)
n=100	$\sigma_x^2 = 1$	-0.0025 (0.0055)	-0.0296 (0.0160)	-0.0171 (0.0180)	-0.0025 (0.0027)
	$\sigma_x^2 = 4$	-0.0035 (0.0075)	-0.1329 (0.0290)	-0.0011 (0.0183)	-0.0002 (0.0029)
	$\sigma_x^2 = 16$	0.0130 (0.0134)	-0.3078 (0.0879)	-0.1045 (0.0181)	-0.0009 (0.0034)
n=500	$\sigma_x^2 = 1$	-0.0004 (0.0025)	-0.0034 (0.0071)	-0.0095 (0.0080)	0.0001 (0.0012)
	$\sigma_x^2 = 4$	0.0034 (0.0035)	-0.0167 (0.0123)	-0.0054 (0.0081)	0.0009 (0.0013)
	$\sigma_x^2 = 16$	-0.0080 (0.0061)	-0.0476 (0.0363)	0.0017 (0.0078)	-0.0021 (0.0014)
n=1000	$\sigma_x^2 = 1$	0.0010 (0.0017)	-0.0067 (0.0051)	0.0103 (0.0056)	0.0000 (0.0008)
	$\sigma_x^2 = 4$	0.0011 (0.0024)	-0.0023 (0.0090)	-0.0060 (0.0057)	-0.0006 (0.0009)
	$\sigma_x^2 = 16$	-0.0073 (0.0044)	-0.0259 (0.0256)	-0.0088 (0.0058)	-0.0004 (0.0009)

* In each cell, the numbers are in the form of empirical bias (standard error).

Table 5.3 - 5.5 summarize behaviors of empirical biases of estimators when $(\beta_0, \beta_1) = (b_0, b_1)$ is known, but values of ρ vary.

Table 5.4 Consistency for MLE under $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}) = (\mathbf{0}, \mathbf{1})$ with $(\boldsymbol{\sigma^2}, \boldsymbol{\mu_x}, \boldsymbol{\rho}) = (\mathbf{4}, \mathbf{1}, \mathbf{0.5})$ in structural model

Sample Size	Parameter Value	$\dot{\mu}_x$	$\dot{\sigma}_x^2$	$\dot{\sigma}^2$	$\dot{\rho}$
n=10	$\sigma_x^2 = 1$	-0.0338 (0.0214)	0.5009 (0.0267)	-0.9239 (0.0312)	-0.1594 (0.0060)
	$\sigma_x^2 = 4$	0.0254 (0.0263)	-0.1671 (0.0594)	-0.5017 (0.0382)	-0.1862 (0.0062)
	$\sigma_x^2 = 16$	-0.0189 (0.0364)	-1.7710 (0.1784)	-0.4048 (0.0431)	-0.1532 (0.0065)
n=20	$\sigma_x^2 = 1$	-0.0170 (0.0149)	0.2386 (0.0190)	-0.4448 (0.0245)	-0.0741 (0.0038)
	$\sigma_x^2 = 4$	0.0098 (0.0170)	-0.2443 (0.0431)	-0.2148 (0.0277)	-0.0935 (0.0043)
	$\sigma_x^2 = 16$	0.0406 (0.0250)	-0.7067 (0.1301)	-0.1730 (0.0290)	-0.0650 (0.0043)
n=30	$\sigma_x^2 = 1$	-0.0011 (0.0122)	0.1457 (0.0154)	-0.2858 (0.0205)	-0.0518 (0.0030)
	$\sigma_x^2 = 4$	0.0140 (0.0141)	-0.1390 (0.0369)	-0.1346 (0.0231)	-0.0441 (0.0034)
	$\sigma_x^2 = 16$	-0.0235 (0.0205)	-0.6571 (0.1061)	-0.1429 (0.0226)	-0.0479 (0.0037)
n=50	$\sigma_x^2 = 1$	0.0035 (0.0094)	0.0617 (0.0127)	-0.1353 (0.0162)	-0.0227 (0.0023)
	$\sigma_x^2 = 4$	0.0094 (0.0110)	-0.0429 (0.0273)	-0.0809 (0.0175)	-0.0252 (0.0025)
	$\sigma_x^2 = 16$	-0.0166 (0.0158)	-0.4303 (0.0837)	-0.0879 (0.0174)	-0.0243 (0.0029)
n=100	$\sigma_x^2 = 1$	0.0090 (0.0067)	0.0074 (0.0092)	-0.0610 (0.0118)	-0.0123 (0.0016)
	$\sigma_x^2 = 4$	0.0006 (0.0078)	-0.0196 (0.0194)	-0.0731 (0.0122)	-0.0115 (0.0018)
	$\sigma_x^2 = 16$	-0.0068 (0.0108)	-0.0285 (0.0583)	-0.0349 (0.0123)	-0.0139 (0.0019)
n=500	$\sigma_x^2 = 1$	0.0008 (0.0030)	0.0004 (0.0042)	-0.0092 (0.0054)	-0.0026 (0.0007)
	$\sigma_x^2 = 4$	-0.0015 (0.0035)	-0.0239 (0.0088)	-0.0043 (0.0055)	-0.0016 (0.0008)
	$\sigma_x^2 = 16$	0.0062 (0.0048)	-0.0298 (0.0260)	-0.0037 (0.0056)	-0.0017 (0.0009)
n=1000	$\sigma_x^2 = 1$	0.0035 (0.0021)	-0.0003 (0.0030)	-0.0063 (0.0039)	-0.0013 (0.0005)
	$\sigma_x^2 = 4$	-0.0025 (0.0025)	-0.0011 (0.0063)	-0.0053 (0.0040)	0.0010 (0.0006)
	$\sigma_x^2 = 16$	0.0019 (0.0035)	-0.0025 (0.0187)	-0.0042 (0.0040)	-0.0008 (0.0006)

* In each cell, the numbers are in the form of empirical bias (standard error).

Table 5.5 Consistency for MLE under $(\beta_0, \beta_1) = (0, 1)$ with $(\sigma^2, \mu_x, \rho) = (4, 1, 0.9)$ in structural model

Sample Size	Parameter Value	$\dot{\mu}_x$	$\dot{\sigma}_x^2$	$\dot{\sigma}^2$	$\dot{\rho}$
n=10	$\sigma_x^2 = 1$	0.0466 (0.0911)	0.3251 (0.0314)	-1.0845 (0.0415)	-0.3317 (0.0076)
	$\sigma_x^2 = 4$	-0.0470 (0.0712)	-0.0141 (0.0689)	-0.7137 (0.0436)	-0.3509 (0.0072)
	$\sigma_x^2 = 16$	-0.1012 (0.0631)	-1.2225 (0.1999)	-0.3973 (0.0582)	-0.3129 (0.0064)
n=20	$\sigma_x^2 = 1$	-0.0078 (0.0731)	0.2213 (0.0253)	-0.5238 (0.0355)	-0.1869 (0.0048)
	$\sigma_x^2 = 4$	0.0912 (0.0600)	0.0892 (0.0571)	-0.3467 (0.0310)	-0.1976 (0.0039)
	$\sigma_x^2 = 16$	-0.0089 (0.0562)	-0.2220 (0.1442)	-0.3694 (0.0304)	-0.1871 (0.0037)
n=30	$\sigma_x^2 = 1$	0.0055 (0.0696)	0.3714 (0.1586)	-0.3238 (0.0326)	-0.1394 (0.0038)
	$\sigma_x^2 = 4$	-0.0213 (0.0613)	0.0934 (0.0488)	-0.2726 0.0289)	-0.1450 (0.0032)
	$\sigma_x^2 = 16$	-0.1578 (0.0529)	0.2621 (0.1276)	-0.3677 (0.0240)	-0.1410 (0.0028)
n=50	$\sigma_x^2 = 1$	0.0405 (0.0776)	0.3683 (0.3116)	-0.1482 (0.0327)	-0.0867 (0.0038)
	$\sigma_x^2 = 4$	0.0676 (0.0617)	0.2241 (0.0729)	-0.2440 (0.0246)	-0.0937 (0.0026)
	$\sigma_x^2 = 16$	-0.0442 (0.0478)	0.6130 (0.1050)	-0.2107 (0.0199)	-0.1018 (0.0020)
n=100	$\sigma_x^2 = 1$	0.0070 (0.0327)	0.0225 (0.0081)	-0.0804 (0.0127)	-0.0275 (0.0009)
	$\sigma_x^2 = 4$	-0.0325 (0.0302)	0.0247 (0.0188)	-0.0949 (0.0125)	-0.0232(0.0009)
	$\sigma_x^2 = 16$	0.0245 (0.0301)	-0.0401(0.0611)	-0.0780 (0.0128)	-0.0239 (0.0009)
n=500	$\sigma_x^2 = 1$	0.0100 (0.0151)	0.0111 (0.0036)	-0.0172 (0.0058)	-0.0063 (0.0004)
	$\sigma_x^2 = 4$	-0.0173 (0.0136)	0.0146 (0.0087)	-0.0139(0.0054)	-0.0051 (0.0004)
	$\sigma_x^2 = 16$	0.0109 (0.0143)	-0.0623 (0.0285)	-0.0244 (0.0056)	-0.0048 (0.0004)
n=1000	$\sigma_x^2 = 1$	0.0061 (0.0106)	0.0073 (0.0026)	-0.0089 (0.0042)	-0.0033 (0.0002)
	$\sigma_x^2 = 4$	0.0025 (0.0103)	0.0089 (0.0059)	-0.0144 (0.0039)	-0.0024 (0.0002)
	$\sigma_x^2 = 16$	-0.0101 (0.0101)	-0.0037 (0.0208)	-0.0028 (0.0038)	-0.0021 (0.0003)

* In each cell, the numbers are in the form of empirical bias (standard error).

As in tables 5.1 and 5.2, the empirical biases of the estimators decrease when either the sample size or the signal to noise ratio increases shown in Table 5.3-5.5. In contrast, the empirical biases increase as ρ increases and other parameters remain constant. Substantially larger sample sizes or signal to noise ratios are required to offset the empirical biases in the presence of significant correlation. Also, empirical biases for estimators of μ_x and σ_x^2 increase as the signal to noise ratio increases.

Tables 5.6 and Table 5.7 show empirical biases and standard errors of the empirical biases when there are no constraints on either (β_0, β_1) or ρ.

Table 5.6 Consistency of MLE without constraints with $\rho = 0$ in structural model

Sample Size	σ_x^2	$\hat{\mu}_x$ ($\mu_x = 1$)	$\hat{\sigma}_x^2$	$\hat{\sigma}^2$ ($\sigma^2 = 4$)	$\hat{\beta}_0$ ($\beta_0 = 0$)	$\hat{\beta}_1$ ($\beta_1 = 1$)	$\hat{\rho}$ ($\rho = 0$)
n=10	1	-0.0012(0.0166)	-0.4431(0.0234)	-0.2043(0.0383)	0.4401(0.0279)	-0.4762(0.0188)	-0.1082(0.0058)
	4	-0.0189(0.0215)	-0.8868(0.0652)	0.1202(0.0526)	0.1806(0.0332)	-0.1662(0.0201)	-0.1548(0.0072)
	16	0.1045(0.0329)	-1.7229(0.1789)	0.3244(0.0832)	-0.0079(0.0288)	-0.0247(0.0113)	-0.1233(0.0072)
n=20	1	0.0035(0.0114)	-0.2248(0.0228)	-0.1548(0.0292)	0.3946(0.0374)	-0.4045(0.0419)	-0.0725(0.0044)
	4	-0.0052(0.0147)	-0.1709(0.0500)	-0.1457(0.0334)	0.0172(0.0226)	-0.0181(0.0141)	-0.0782(0.0051)
	16	0.0190(0.0224)	-0.1565(0.1299)	-0.2500(0.0446)	0.0175(0.0165)	0.0101(0.0065)	-0.0535(0.0053)
n=30	1	0.0038(0.0094)	-0.0951(0.0219)	-0.1096(0.0240)	0.2296(0.0227)	-0.2415(0.0181)	-0.0488(0.0036)
	4	0.0104(0.0120)	0.0340(0.0449)	-0.1947(0.0274)	0.0097(0.0157)	-0.0192(0.0098)	-0.0511(0.0041)
	16	0.0217(0.0183)	-0.2262(0.1127)	-0.1894(0.0366)	0.0155(0.0132)	-0.0062(0.0043)	-0.0355(0.0041)
n=50	1	-0.0039(0.0075)	0.0245(0.0194)	-0.1085(0.0193)	0.1823(0.0184)	-0.1748(0.0154)	-0.0280(0.0027)
	4	0.0070(0.0091)	-0.0980(0.0369)	-0.0939(0.0227)	0.0097(0.0124)	-0.0138(0.0081)	-0.0283(0.0032)
	16	0.0163(0.0142)	-0.1157(0.0885)	-0.1903(0.0219)	-0.0036(0.0096)	0.0016(0.0029)	-0.0209(0.0032)
n=100	1	0.0108(0.0061)	-0.0092(0.0187)	-0.0297(0.0181)	0.1203(0.0188)	-0.1317(0.0165)	-0.0106(0.0023)
	4	0.0041(0.0065)	0.0175(0.0277)	-0.0743(0.0157)	-0.0011(0.0083)	-0.0008(0.0052)	-0.0127(0.0022)
	16	0.0054(0.0102)	-0.1261(0.0632)	-0.1044(0.0139)	-0.0048(0.0066)	0.0011(0.0018)	-0.0089(0.0022)
n=500	1	-0.0005(0.0023)	-0.0032(0.0084)	-0.0010(0.0070)	0.0283(0.0077)	-0.0281(0.0071)	-0.0017(0.0009)
	4	-0.0027(0.0030)	-0.0143(0.0113)	-0.0274(0.0056)	-0.0092(0.0034)	0.0061(0.0018)	-0.0023(0.0010)
	16	-0.0058(0.0045)	-0.0386(0.0282)	-0.0273(0.0056)	-0.0032(0.0030)	0.0018(0.0007)	-0.0042(0.0010)
n=1000	1	-0.0009(0.0016)	0.0080(0.0055)	-0.0137(0.0045)	0.0046(0.0049)	-0.0039(0.0044)	-0.0006(0.0006)
	4	-0.0014(0.0020)	0.0113(0.0082)	-0.0209(0.0040)	-0.0004(0.0024)	0.0012(0.0013)	-0.0015(0.0007)
	16	-0.0017(0.0032)	-0.0126(0.0205)	-0.0122(0.0039)	-0.0001(0.0021)	0.0010(0.0005)	-0.0002(0.0007)

* In each cell, the numbers are in the form of empirical bias (standard error).

Table 5.7 Consistency of MLE without constraints with $\rho = 0.5$ in structural model

Sample Size	σ_x^2	$\hat{\mu}_x$ ($\mu_x = 1$)	$\hat{\sigma}_x^2$	$\hat{\sigma}^2$ ($\sigma^2 = 4$)	$\hat{\beta}_0$ ($\beta_0 = 0$)	$\hat{\beta}_1$ ($\beta_1 = 1$)	$\hat{\rho}$ ($\rho = 0.5$)
n=10	1	-0.0216(0.0312)	-0.2540(0.0272)	-0.3475(0.0359)	0.3517(0.0504)	-0.3108(0.0209)	-0.2857(0.0064)
	4	0.0128(0.0333)	-0.2014(0.0658)	-0.4607(0.0439)	-0.0136(0.0540)	-0.0018(0.0181)	-0.3154(0.0072)
	16	-0.0266(0.0413)	-0.9191(0.1776)	-0.1272(0.0737)	-0.0121(0.0500)	-0.0130(0.0124)	-0.2970(0.0076)
n=20	1	-0.0318(0.0208)	0.0518(0.0255)	-0.2726(0.0272)	0.1595(0.0368)	-0.1372(0.0204)	-0.1360(0.0043)
	4	0.0031(0.0229)	0.1657(0.0501)	-0.3967(0.0307)	-0.0513(0.0341)	0.0184(0.0113)	-0.1659(0.0050)
	16	-0.0215(0.0292)	0.0161(0.1310)	-0.2881(0.0462)	-0.0063(0.0324)	-0.0047(0.0076)	-0.1307(0.0050)
n=30	1	0.0027(0.0171)	0.1475(0.0222)	-0.2478(0.0220)	-0.0010(0.0325)	-0.0179(0.0172)	-0.0885(0.0035)
	4	-0.0142(0.0181)	0.2352(0.0435)	-0.2775(0.0268)	-0.0185(0.0255)	0.0101(0.0087)	-0.0982(0.0039)
	16	0.0180(0.0229)	0.1403(0.1108)	-0.2813(0.0352)	-0.0136(0.0241)	-0.0074(0.0036)	-0.0831(0.0041)
n=50	1	-0.0047(0.0133)	0.0985(0.0180)	-0.1409(0.0181)	-0.0090(0.0254)	-0.0234(0.0154)	-0.0493(0.0025)
	4	0.0098(0.0137)	0.1358(0.0337)	-0.2059(0.0202)	0.0082(0.0197)	0.0045(0.0079)	-0.0503(0.0028)
	16	-0.0136(0.0189)	-0.1456(0.0875)	-0.2083(0.0228)	0.0153(0.0188)	0.0038(0.0027)	-0.0481(0.0029)
n=100	1	0.0038(0.0101)	0.0207(0.0140)	-0.0845(0.0155)	0.0054(0.0226)	-0.0201(0.0150)	-0.0242(0.0019)
	4	0.0245(0.0100)	0.0950(0.0243)	-0.1201(0.0139)	-0.0159(0.0134)	0.0077(0.0039)	-0.0213(0.0019)
	16	0.0026(0.0125)	-0.0961(0.0652)	-0.1314(0.0141)	0.0045(0.0125)	-0.0027(0.0017)	-0.0248(0.0020)
n=500	1	-0.0026(0.0042)	0.0100(0.0063)	-0.0216(0.0061)	-0.0012(0.0077)	0.0022(0.0050)	-0.0035(0.0008)
	4	0.0042(0.0045)	-0.0021(0.0105)	-0.0287(0.0056)	-0.0085(0.0058)	0.0037(0.0015)	-0.0043(0.0008)
	16	0.0095(0.0057)	0.0054(0.0282)	-0.0278(0.0056)	-0.0069(0.0057)	0.0005(0.0007)	-0.0039(0.0009)
n=1000	1	-0.0013(0.0030)	0.0071(0.0041)	-0.0131(0.0039)	-0.0033(0.0050)	0.0083(0.0029)	-0.0009(0.0005)
	4	-0.0045(0.0032)	0.0008(0.0075)	-0.0115(0.0039)	0.0006(0.0043)	0.0017(0.0010)	-0.0022(0.0006)
	16	0.0002(0.0040)	0.0272(0.0201)	-0.0181(0.0040)	-0.0024(0.0040)	0.0000(0.0005)	-0.0019(0.0006)

* In each cell, the numbers are in the form of empirical bias (standard error).

In general, the empirical biases of the unconstrained maximum likelihood estimators decrease as sample size increases. Also, the empirical bias of these estimators increases as the correlation gets stronger. The empirical bias of estimator of σ^2 is generally larger than those of other estimators.

5.1.2 Empirical biases and consistency under the functional model

Bivariate normal data $(\delta_t, \upsilon_t)^T$ with mean vector $(0,0)^T$ and covariance matrix $\sigma^2 I$ are generated using the MASS package in R and $e_t = \rho e_{t-1} + \delta_t$ and $u_t = \rho u_{t-1} + \upsilon_t$ are computed. A set of x_t values is independently generated from a normal distribution with mean μ_x and variance σ_x^2. This set of x_t values is held fixed for each of the 2000 simulated data sets for a given parameter configuration. Simulated data for the functional model are computed as $X_t = x_t + e_t$, y_t, $Y_t = \beta_0 + \beta_1 x_t + u_t = y_t + u_t$. Remaining parameter settings are identical to those for the structural model simulations used to generate results reported tables 5.1-5.7. The empirical biases and standard errors of the empirical biases for all parameter estimates (except the x_t are displayed in tables 5.8-5.14. As shown in Section 4, the estimator of σ^2 converges to a half of its true value. Therefore, these empirical biases are calculated by subtracting the value of $\sigma^2/2$ from the average of the 2000 simulated variance estimates for each set of parameters.

Empirical biases and standard errors of the empirical biases decrease as sample size increase for each estimator in tables 5.8 and 5.9. As the ratio of σ_x^2 to σ^2 increases, the empirical biases of the estimates of μ_x, σ_x^2, and σ^2 increase slightly, whereas the biases and standard errors for the intercept and slope estimates decrease. As the signal to noise ratio increases, empirical biases of the regression coefficients decrease. We note that the values of the intercept and slope appeared to have little effect on the empirical biases of these estimators.

Table 5.8 Consistency for MLE under $\rho = 0$ with $(\beta_0, \beta_1, \sigma^2, \mu_x) = (0, 1, 4, 1)$ in functional model

Sample Size	Parameter Value	$\tilde{\sigma}^2$	$\tilde{\beta}_0$	$\tilde{\beta}_1$
n=10	$\sigma_x^2 = 1$	-0.8406 (0.0128)	2.9287 (1.7575)	-4.4115 (2.9132)
	$\sigma_x^2 = 4$	-0.5564 (0.0168)	0.8522 (0.8090)	-1.3816 (1.3689)
	$\sigma_x^2 = 16$	-0.4608 (0.0192)	-0.0136 (0.0232)	0.0527 (0.0071)
n=20	$\sigma_x^2 = 1$	-0.4181 (0.0114)	-0.1121 (0.2782)	0.0291 (0.3838)
	$\sigma_x^2 = 4$	-0.2798 (0.0129)	-0.0799 (0.0231)	0.1356 (0.0239)
	$\sigma_x^2 = 16$	-0.2029 (0.0135)	0.0235 (0.0152)	0.0229 (0.0041)
n=30	$\sigma_x^2 = 1$	-0.3231 (0.0095)	1.6682 (0.7671)	-1.3280 (0.5660)
	$\sigma_x^2 = 4$	-0.1656 (0.0112)	-0.1146 (0.0195)	0.0738 (0.0107)
	$\sigma_x^2 = 16$	-0.1573 (0.0112)	-0.0214 (0.0127)	0.0058 (0.0027)
n=50	$\sigma_x^2 = 1$	-0.1468 (0.0084)	-0.3370 (0.0892)	0.3805 (0.1026)
	$\sigma_x^2 = 4$	-0.1007 (0.0087)	-0.0297 (0.0123)	0.0345 (0.0063)
	$\sigma_x^2 = 16$	-0.0833 (0.0089)	-0.0015 (0.0093)	0.0051 (0.0022)
n=100	$\sigma_x^2 = 1$	-0.0913 (0.0059)	-0.1681 (0.0460)	0.1954 (0.0537)
	$\sigma_x^2 = 4$	-0.0559 (0.0062)	-0.0244 (0.0084)	0.0172 (0.0042)
	$\sigma_x^2 = 16$	-0.0433 (0.0063)	0.0038 (0.0073)	0.0032 (0.0018)
n=500	$\sigma_x^2 = 1$	-0.0183 (0.0029)	-0.0339 (0.0061)	0.0324 (0.0056)
	$\sigma_x^2 = 4$	-0.0175 (0.0028)	-0.0037 (0.0035)	0.0017 (0.0017)
	$\sigma_x^2 = 16$	-0.0043 (0.0028)	-0.0027 (0.0029)	0.0010 (0.0008)
n=1000	$\sigma_x^2 = 1$	-0.0092 (0.0020)	-0.0115 (0.0039)	0.0125 (0.0037)
	$\sigma_x^2 = 4$	-0.0019 (0.0020)	-0.0038 (0.0024)	0.0016 (0.0013)
	$\sigma_x^2 = 16$	-0.0053 (0.0020)	0.0011 (0.0021)	0.0004 (0.0005)

* In each cell, the numbers are in the form of empirical bias (standard error).

Table 5.9 Consistency for MLE under $\rho = 0$ with $(\beta_0, \beta_1, \sigma^2, \mu_x) = (1, 3, 4, 1)$ in functional model

Sample Size	Parameter Value	$\tilde{\sigma}^2$	$\tilde{\beta}_0$	$\tilde{\beta}_1$
n=10	$\sigma_x^2 = 1$	-0.5095 (0.0176)	-5.8508 (3.9770)	6.0422 (7.3714)
	$\sigma_x^2 = 4$	-0.4482 (0.0188)	3.2476 (3.3543)	-0.9426 (1.4153)
	$\sigma_x^2 = 16$	-0.4493 (0.0186)	-0.0590 (0.0519)	0.1094 (0.0132)
n=20	$\sigma_x^2 = 1$	-0.2018 (0.0140)	-4.1158 (3.3853)	4.5791 (3.9877)
	$\sigma_x^2 = 4$	-0.2090 (0.0136)	-0.1342 (0.0400)	0.1096 (0.0166)
	$\sigma_x^2 = 16$	-0.2191 (0.0137)	-0.0861 (0.0332)	0.0337 (0.0076)
n=30	$\sigma_x^2 = 1$	-0.1823 (0.0112)	-0.1134 (2.0318)	0.9668 (2.3900)
	$\sigma_x^2 = 4$	-0.1387 (0.0112)	-0.1208 (0.0318)	0.1017 (0.0136)
	$\sigma_x^2 = 16$	-0.1358 (0.0114)	-0.0054 (0.0273)	0.0223 (0.0067)
n=50	$\sigma_x^2 = 1$	-0.0923 (0.0088)	-0.6026 (0.0896)	0.6257 (0.0891)
	$\sigma_x^2 = 4$	-0.0815 (0.0087)	-0.0741 (0.0237)	0.0894 (0.0141)
	$\sigma_x^2 = 16$	-0.0545 (0.0089)	-0.0421 (0.0206)	0.0207 (0.0056)
n=100	$\sigma_x^2 = 1$	-0.0489 (0.0063)	-0.1641 (0.0238)	0.1884 (0.0199)
	$\sigma_x^2 = 4$	-0.0443 0.0063)	-0.0457 (0.0166)	0.0409 (0.0081)
	$\sigma_x^2 = 16$	-0.0454 (0.0063)	-0.0185 (0.0149)	0.0104 (0.0037)
n=500	$\sigma_x^2 = 1$	-0.0039 (0.0028)	-0.0356 (0.0098)	0.0317 (0.0071)
	$\sigma_x^2 = 4$	-0.0099 (0.0028)	-0.0093 (0.0071)	0.0048 (0.0033)
	$\sigma_x^2 = 16$	-0.0113 (0.0028)	-0.0023 (0.0065)	0.0047 (0.0016)
n=1000	$\sigma_x^2 = 1$	-0.0022 (0.0020)	-0.0182 (0.0073)	0.0164 (0.0054)
	$\sigma_x^2 = 4$	-0.0052 (0.0020)	-0.0138 (0.0052)	0.0053 (0.0025)
	$\sigma_x^2 = 16$	-0.0037 (0.0020)	-0.0079 (0.0047)	0.0004 (0.0011)

* In each cell, the numbers are in the form of empirical bias (standard error).

Table 5.10 Consistency for MLE under $(\beta_0, \beta_1) = (0, 1)$ with $(\sigma^2, \mu_x) = (4, 1)$ in functional model

Sample Size	σ_x^2	$\ddot{\sigma}^2$ ($\rho = 0$)	$\ddot{\rho}$ ($\rho = 0$)	$\ddot{\sigma}^2$ ($\rho = 0.5$)	$\ddot{\rho}$ ($\rho = 0.5$)	$\ddot{\sigma}^2$ ($\rho = 0.9$)	$\ddot{\rho}$ ($\rho = 0.9$)
n=10	1	-0.1713(0.0205)	0.0024(0.0071)	-0.2069(0.0193)	-0.0908(0.0068)	-0.2219(0.0206)	-0.1464(0.0063)
	4	-0.1882(0.0196)	-0.0038(0.0069)	-0.1944(0.0192)	-0.0821(0.0067)	-0.2203(0.0199)	-0.1393(0.0061)
	16	-0.1768(0.0202)	0.0107(0.0070)	-0.2046(0.0192)	-0.0777(0.0067)	-0.1571(0.0208)	-0.1381(0.0060)
n=20	1	-0.1165(0.0140)	0.0059(0.0050)	-0.0829(0.0142)	-0.0401(0.0045)	-0.0968(0.0143)	-0.0791(0.0034)
	4	-0.0996(0.0136)	0.0012(0.0049)	-0.0773(0.0144)	-0.0415(0.0045)	-0.0983(0.0140)	-0.0759(0.0035)
	16	-0.0834(0.0139)	-0.0006(0.0047)	-0.0961(0.0140)	-0.0372(0.0044)	-0.1157(0.0142)	-0.0867(0.0036)
n=30	1	-0.0611(0.0115)	-0.0022(0.0040)	-0.0592(0.0117)	-0.0346(0.0036)	-0.0718(0.0113)	-0.0538(0.0025)
	4	-0.0558(0.0115)	0.0053(0.0040)	-0.0526(0.0117)	-0.0310(0.0035)	-0.0613(0.0119)	-0.0563(0.0026)
	16	-0.0664(0.0116)	0.0056(0.0040)	-0.0710(0.0113)	-0.0380(0.0037)	-0.0549(0.0113)	-0.0510(0.0026)
n=50	1	-0.0423(0.0089)	0.0025(0.0031)	-0.0295(0.0090)	-0.0206(0.0029)	-0.0476(0.0091)	-0.0331(0.0018)
	4	-0.0313(0.0090)	-0.0006(0.0031)	-0.0368(0.0091)	-0.0204(0.0028)	-0.0272(0.0090)	-0.0332(0.0018)
	16	-0.0462(0.0087)	0.0043(0.0032)	-0.0286(0.0090)	-0.0191(0.0028)	-0.0487(0.0088)	-0.0347(0.0018)
n=100	1	-0.0135(0.0062)	0.0038(0.0022)	-0.0173(0.0063)	-0.0145(0.0020)	-0.0239(0.0065)	-0.0182(0.0012)
	4	-0.0256(0.0065)	0.0032(0.0022)	-0.0155(0.0064)	-0.0117(0.0019)	-0.0179(0.0063)	-0.0173(0.0011)
	16	-0.0276(0.0063)	-0.0011(0.0022)	-0.0177(0.0065)	-0.0094(0.0019)	-0.0310(0.0063)	-0.0187(0.0011)
n=500	1	-0.0044(0.0029)	0.0007(0.0010)	-0.0048(0.0027)	-0.0023(0.0009)	-0.0069(0.0028)	-0.0041(0.0005)
	4	-0.0008(0.0029)	0.0006(0.0010)	-0.0067(0.0028)	-0.0014(0.0009)	-0.0050(0.0029)	-0.0032(0.0005)
	16	-0.0016(0.0028)	-0.0019(0.0010)	0.0003(0.0028)	-0.0026(0.0009)	-0.0022(0.0029)	-0.0036(0.0005)
n=1000	1	-0.0008(0.0020)	0.0014(0.0007)	-0.0014(0.0020)	-0.0009(0.0006)	-0.0023(0.0020)	-0.0020(0.0003)
	4	-0.0026(0.0021)	-0.0008(0.0007)	-0.0019(0.0020)	-0.0012(0.0006)	-0.0051(0.0020)	-0.0020(0.0003)
	16	-0.0048(0.0020)	0.0009(0.0007)	0.0023(0.0020)	-0.0007(0.0006)	-0.0036(0.0020)	-0.0024(0.0003)

* In each cell, the numbers are in the form of empirical bias (standard error).

Table 5.10 summarizes behavior of the estimates derived under the null hypothesis that $(\beta_0, \beta_1) = (b_0, b_1)$. The first pair of columns, labeled $\ddot{\sigma}^2$ and $\ddot{\rho}$, are estimates of σ^2 and ρ, respectively, when the correlation is $\rho = 0$. The second and third pair of columns correspond to cases in which $\rho = 0.5$ and $\rho = 0.9$, respectively. As expected, empirical biases are reduced for larger sample sizes. Unlike the signal to noise ratio, the value of ρ seems to affect empirical biases of the estimators. Larger correlations require larger sample sizes to reduce empirical biases to levels observed for smaller (or no) correlation.

Table 5.11 Consistency for MLE without constraints with $(\sigma^2, \beta_0, \beta_1, \rho) = (4, 0, 1, 0)$ in functional model

Sample Size	Parameter Value	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\rho}$
n=10	$\sigma_x^2 = 1$	-0.8301 (0.0152)	0.1194 (1.8599)	0.1346 (1.0725)	-0.1265 (0.0086)
	$\sigma_x^2 = 4$	-0.7632 (0.0158)	-0.1230 (0.1902)	0.0552 (0.1731)	-0.1223 (0.0086)
	$\sigma_x^2 = 16$	-0.7145 (0.0165)	0.1316 (0.1038)	0.0605 (0.0116)	-0.1706 (0.0078)
n=20	$\sigma_x^2 = 1$	-0.5784 (0.0107)	0.4309 (0.7813)	-0.9904 (0.9473)	-0.0697 (0.0058)
	$\sigma_x^2 = 4$	-0.3706 (0.0129)	-0.1369 (0.0287)	0.1468 (0.0228)	-0.0591 (0.0056)
	$\sigma_x^2 = 16$	-0.3244 (0.0132)	-0.0282 (0.0151)	0.0307 (0.0061)	-0.0 538 (0.0054)
n=30	$\sigma_x^2 = 1$	-0.4128 (0.0094)	0.4262 (0.3435)	0.6258 (0.7988)	-0.0604 (0.0048)
	$\sigma_x^2 = 4$	-0.2392 (0.0109)	-0.0285 (0.0141)	0.0384 (0.0065)	-0.0385 (0.0043)
	$\sigma_x^2 = 16$	-0.2253 (0.0112)	-0.0709 (0.0158)	0.0178 (0.0034)	-0.0501 (0.0043)
n=50	$\sigma_x^2 = 1$	-0.2236 (0.0080)	-0.1738 (0.1964)	0.2906 (0.2555)	-0.0144 (0.0035)
	$\sigma_x^2 = 4$	-0.1574 (0.0084)	-0.0399 (0.0122)	0.0519 (0.0079)	-0.0202 (0.0033)
	$\sigma_x^2 = 16$	-0.1370 (0.0086)	0.0102 (0.0089)	0.0071 (0.0022)	-0.0229 (0.0033)
n=100	$\sigma_x^2 = 1$	-0.1095 (0.0060)	-0.1894 (0.0710)	0.1891 (0.0726)	-0.0104 (0.0024)
	$\sigma_x^2 = 4$	-0.0707 (0.0061)	-0.0116 (0.0070)	0.0156 (0.0040)	-0.0078 (0.0023)
	$\sigma_x^2 = 16$	-0.0706 (0.0061)	0.0046 (0.0067)	0.0043 (0.0017)	-0.0137 (0.0023)
n=500	$\sigma_x^2 = 1$	-0.0238 (0.0028)	-0.0428 (0.0066)	0.0434 (0.0059)	-0.0013 (0.0010)
	$\sigma_x^2 = 4$	-0.0115 (0.0028)	-0.0045 (0.0034)	0.0033 (0.0018)	-0.0026 (0.0010)
	$\sigma_x^2 = 16$	-0.0155 (0.0029)	0.0005 (0.0029)	0.0015 (0.0008)	-0.0027 (0.0010)
n=1000	$\sigma_x^2 = 1$	-0.0099 (0.0020)	-0.0074 (0.0044)	0.0089 (0.0037)	-0.0009 (0.0007)
	$\sigma_x^2 = 4$	-0.0069 (0.0020)	-0.0002 (0.0024)	0.0012 (0.0012)	-0.0011 (0.0007)
	$\sigma_x^2 = 16$	-0.0108 (0.0019)	0.0007 (0.0020)	-0.0004 (0.0005)	-0.0005 (0.0007)

* In each cell, the numbers are in the form of empirical bias (standard error).

Tables 5.11-5.13 show empirical biases of the unconstrained MLEs when ρ =0, 0.5, and 0.9.

Table 5.12 Consistency for MLE without constraints with $(\boldsymbol{\sigma}^2, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\rho}) = (4, 0, 1, 0.5)$ in functional model

Sample Size	Parameter Value	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\rho}$
n=10	$\sigma_x^2 = 1$	-0.9873 (0.0131)	1.1360 (0.6339)	-0.6777 (0.4593)	-0.3741 (0.0087)
	$\sigma_x^2 = 4$	-0.6508 (0.0168)	0.1818 (0.0779)	0.0674 (0.0083)	-0.2430 (0.0076)
	$\sigma_x^2 = 16$	-0.6984 (0.0165)	-0.1435 (0.1278)	0.0939 (0.0135)	-0.2952 (0.0085)
n=20	$\sigma_x^2 = 1$	-0.3921 (0.0125)	0.1461 (0.1728)	-0.0253 (0.1119)	-0.1574 (0.0059)
	$\sigma_x^2 = 4$	-0.4020 (0.0122)	-0.0568 (0.1399)	0.3982 (0.2312)	-0.1652 (0.0059)
	$\sigma_x^2 = 16$	-0.3368 (0.0131)	-0.0118 (0.0293)	0.0286 (0.0051)	-0.1364 (0.0052)
n=30	$\sigma_x^2 = 1$	-0.3254 (0.0103)	-0.4438 (0.2880)	0.8320 (0.3082)	-0.1320 (0.0047)
	$\sigma_x^2 = 4$	-0.2472 (0.0109)	-0.0204 (0.0410)	0.0405 (0.0272)	-0.0912 (0.0042)
	$\sigma_x^2 = 16$	-0.2193 (0.0112)	-0.0282 (0.0243)	0.0126 (0.0035)	-0.0887 (0.0041)
n=50	$\sigma_x^2 = 1$	-0.2069 (0.0081)	0.6064 (0.5529)	-0.8502 (0.7458)	-0.0691 (0.0033)
	$\sigma_x^2 = 4$	-0.1368 (0.0090)	-0.0443 (0.0223)	0.0350 (0.0070)	-0.0500 (0.0030)
	$\sigma_x^2 = 16$	-0.1310 (0.0089)	0.0043 (0.0183)	0.0014 (0.0017)	-0.0422 (0.0029)
n=100	$\sigma_x^2 = 1$	-0.0911 (0.0059)	-0.1854 (0.0468)	0.1673 (0.0382)	-0.0298 (0.0021)
	$\sigma_x^2 = 4$	-0.0676 (0.0062)	0.0210 (0.0129)	0.0116 (0.0036)	-0.0242 (0.0020)
	$\sigma_x^2 = 16$	-0.0649 (0.0063)	0.0095 (0.0127)	0.0029 (0.0016)	-0.0243 (0.0020)
n=500	$\sigma_x^2 = 1$	-0.0189 (0.0028)	-0.0154 (0.0074)	0.0145 (0.0043)	-0.0052 (0.0009)
	$\sigma_x^2 = 4$	-0.0146 (0.0028)	-0.0108 (0.0061)	0.0039 (0.0016)	-0.0049 (0.0009)
	$\sigma_x^2 = 16$	-0.0168 (0.0028)	-0.0085 (0.0057)	0.0015 (0.0006)	-0.0048 (0.0009)
n=1000	$\sigma_x^2 = 1$	-0.0089 (0.0020)	-0.0176 (0.0051)	0.0146 (0.0030)	-0.0026 (0.0006)
	$\sigma_x^2 = 4$	-0.0073 (0.0020)	0.0059 (0.0042)	-0.0010 (0.0011)	-0.0031 (0.0006)
	$\sigma_x^2 = 16$	-0.0060 (0.0021)	0.0020 (0.0040)	-0.0003 (0.0005)	-0.0027 (0.0006)

* In each cell, the numbers are in the form of empirical bias (standard error).

Table 5.13 Consistency for MLE without constraints with $(\boldsymbol{\sigma^2}, \boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \boldsymbol{\rho}) = (\mathbf{4}, \mathbf{0}, \mathbf{1}, \mathbf{0.9})$ in functional model

Sample Size	Parameter Value	$\hat{\sigma}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\rho}$
n=10	$\sigma_x^2 = 1$	-0.9943 (0.0132)	-0.3859 (0.6506)	-0.2462 (0.2714)	-0.5546 (0.0090)
	$\sigma_x^2 = 4$	-0.8267 (0.0159)	0.0217 (0.9975)	0.4389 (0.2504)	-0.5149 (0.0092)
	$\sigma_x^2 = 16$	-0.8147 (0.0160)	0.7169 (0.8859)	0.0210 (0.0046)	-0.4834 (0.0089)
n=20	$\sigma_x^2 = 1$	-0.5031 (0.0115)	-0.9249 (6.1703)	-0.2449 (0.1964)	-0.2943 (0.0060)
	$\sigma_x^2 = 4$	-0.4137 (0.0131)	0.5156 (0.3488)	0.0607 (0.0118)	-0.2439 (0.0049)
	$\sigma_x^2 = 16$	-0.3804 (0.0128)	-0.0752 (0.1644)	0.0020 (0.0031)	-0.2297 (0.0044)
n=30	$\sigma_x^2 = 1$	-0.3313 (0.0106)	-0.4371 (0.3813)	-0.2340 (0.4471)	-0.1809 (0.0041)
	$\sigma_x^2 = 4$	-0.2767 (0.0110)	0.2810 (0.4263)	0.0456 (0.0070)	-0.1638 (0.0035)
	$\sigma_x^2 = 16$	-0.2670 (0.0110)	0.3687 (0.3803)	0.0042 (0.0021)	-0.1538 (0.0033)
n=50	$\sigma_x^2 = 1$	-0.1878 (0.0081)	-0.0490 (1.3443)	0.3391 (0.2514)	-0.1032 (0.0026)
	$\sigma_x^2 = 4$	-0.1536 (0.0087)	0.0897 (0.1085)	0.0130 (0.0043)	-0.0904 (0.0022)
	$\sigma_x^2 = 16$	-0.1488 (0.0088)	-0.0137 (0.0995)	0.0020 (0.0015)	-0.0846 (0.0022)
n=100	$\sigma_x^2 = 1$	-0.0988 (0.0060)	0.0969 (0.1607)	0.0027 (0.0475)	-0.0436 (0.0014)
	$\sigma_x^2 = 4$	-0.0781 (0.0062)	-0.0646 (0.0652)	0.0093 (0.0028)	-0.0424 (0.0013)
	$\sigma_x^2 = 16$	-0.0701 (0.0062)	-0.0576 (0.0622)	0.0025 (0.0014)	-0.0403 (0.0013)
n=500	$\sigma_x^2 = 1$	-0.0179 (0.0028)	0.0042 (0.0286)	-0.0018 (0.0033)	-0.0071 (0.0005)
	$\sigma_x^2 = 4$	-0.0157 (0.0028)	0.0077 (0.0293)	0.0006 (0.0012)	-0.0083 (0.0005)
	$\sigma_x^2 = 16$	-0.0135 (0.0028)	-0.0273 (0.0281)	0.0005 (0.0005)	-0.0082 (0.0005)
n=1000	$\sigma_x^2 = 1$	-0.0099 (0.0019)	-0.0129 (0.0197)	0.0028 (0.0023)	-0.0042 (0.0003)
	$\sigma_x^2 = 4$	-0.0078 (0.0019)	-0.0156 (0.0202)	0.0003 (0.0009)	-0.0036 (0.0003)
	$\sigma_x^2 = 16$	-0.0065 (0.0020)	0.0216 (0.0196)	-0.0004 (0.0004)	-0.0034 (0.0003)
$n=10^4$	$\sigma_x^2 = 1$	-0.0007 (0.0006)	0.0033 (0.0063)	0.0005 (0.0007)	-0.0004 (0.0001)
	$\sigma_x^2 = 4$	-0.0012 (0.0006)	-0.0038 (0.0062)	0.0003 (0.0003)	-0.0004 (0.0001)
	$\sigma_x^2 = 16$	-0.0011 (0.0006)	-0.0037 (0.0063)	0.0001 (0.0001)	-0.0002 (0.0001)

* In each cell, the numbers are in the form of empirical bias (standard error).

Comparing tables 5.11, 5.12 and 5.13, we again see that stronger correlations require larger samples to achieve comparable levels of empirical biases, while larger signal to noise ratios have the effect of reducing empirical biases.

In summary, increasing sample size or the signal to noise ratio reduces the empirical bias. Consistency of the estimators is apparent, with empirical biases of most estimators less than 0.01 for sample sizes of 1000 or greater. Correlated data generally appear to have slower convergence and require larger sample sizes to reduce empirical bias.

5.2 Large Sample Distributions of Test Statistics

Simulations in this section mimic those of Section 5.1, but focus on the behavior of the test statistics. Sample sizes and parameter values vary to examine the effects on the test statistics developed in Section 2 under various scenarios.

5.2.1 Small sample and asymptotic behavior of test statistics under the structural model

Anderson-Darling (A-D) goodness of fit (GOF) tests and quantile vs. quantile (Q-Q) plots are used to exam the large sample distribution of the test statistics under the null hypotheses. P-value is obtained by comparing 1000 values of test statistics with chi-square distributions. Repeat this procedure 1000 times. Recall from section 4 for the structural model, the score test statistic is asymptotically distributed as chi-square with degree of freedom 1 under the null hypothesis $\rho = 0$. Table 5.14 shows empirical p-values for A-D GOF tests of the score test statistic for different sets of parameter values. Figure 5.1 displays sets of Q-Q plots for empirical score test statistic quantiles plotted against quantiles from the chi-square distribution with one degree of freedom.

Table 5.14 illustrates the general trend that varying values of (β_0, β_1) has little impact on the behavior of the score test statistic. Consequently, subsequent tables in this section are we only show the Q-Q for the primary interest, $(\beta_0, \beta_1) = (0,1)$. Sample sizes of 20 or greater, combined with relatively large signal to noise ratios, appear to be sufficient for the distribution of the score test statistic to be well approximated by the chi-square distributed with one degree of freedom. As sample sizes reach about 50, even with low signal to noise ratio, the score test statistic behaves much like a chi-square random variable with one degree of freedom.

Table 5.14 A-D test p-values* for score test statistic under $\rho = 0$ in structural model

Sample Size	Parameter $(\beta_0, \beta_1, \sigma^2, \mu_x)$	$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	(0,1,4,1)	0.0444 (0.2820,0.5210,0.6440)	0.0668 (0.2470,0.4510,0.5830)	0.0994 (0.1790,0.3700,0.5010)
	(1,1,4,1)	0.0419 (0.2910,0.5290,0.6400)	0.0584 (0.2550,0.4720,0.6000)	0.1046 (0.1790,0.3670,0.4920)
	(1,3,4,1)	0.0704 (0.2410,0.4430,0.5620)	0.1205 (0.1780,0.3530,0.46450	0.1348 (0.1640,0.3400,0.4440)
n=20	(0,1,4,1)	0.4319 (0.0150,0.0700,0.1180)	0.4570 (0.0110,0.0480,0.1100)	0.4641 (0.0150,0.0510,0.1050)
	(1,1,4,1)	0.4199 (0.0110,0.0680,0.1330)	0.4524 (0.0140,0.0570,0.1125)	0.4658 (0.0110,0.0510,0.1040)
	(1,3,4,1)	0.4223 (0.0170,0.0620,0.1200)	0.4255 (0.0120,0.0650,0.1290)	0.4414 (0.0110,0.0610,0.1240)
n=50	(0,1,4,1)	0.4776 (0.0090,0.0580,0.1050)	0.5104 (0.0040,0.0520,0.0930)	0.5026 (0.0100,0.0490,0.1020)
	(1,1,4,1)	0.4913 (0.0130,0.0560,0.1030)	0.4705 (0.0110,0.0510,0.1040)	0.4822 (0.0090,0.0510,0.1020)
	(1,3,4,1)	0.4933 (0.0180,0.0670,0.1110)	0.4513 (0.0120,0.0570,0.1140)	0.4708 (0.0110,0.0550,0.1100)
n=100	(0,1,4,1)	0.5169 (0.0090,0.0480,0.1000)	0.4824 (0.0080,0.0480,0.0980)	0.5122 (0.0090,0.0490,0.0970)
	(1,1,4,1)	0.4764 (0.0080,0.0520,0.1000)	0.5040 (0.0110,0.0500,0.1050)	0.4913 (0.0090,0.0480,0.1040)
	(1,3,4,1)	0.5131 (0.0080,0.0430,01020)	0.5001 (0.0110,0.0510,0.1090)	0.5166 (0.0090,0.0490,0.0920)

*p-values are calculated by comparing one half of test statistic with χ_1^2. Each cell has the form of median of 2000 p-values, rejection probabilities with 3 different significance level ($\alpha = 0.01,0.05,0.1$).

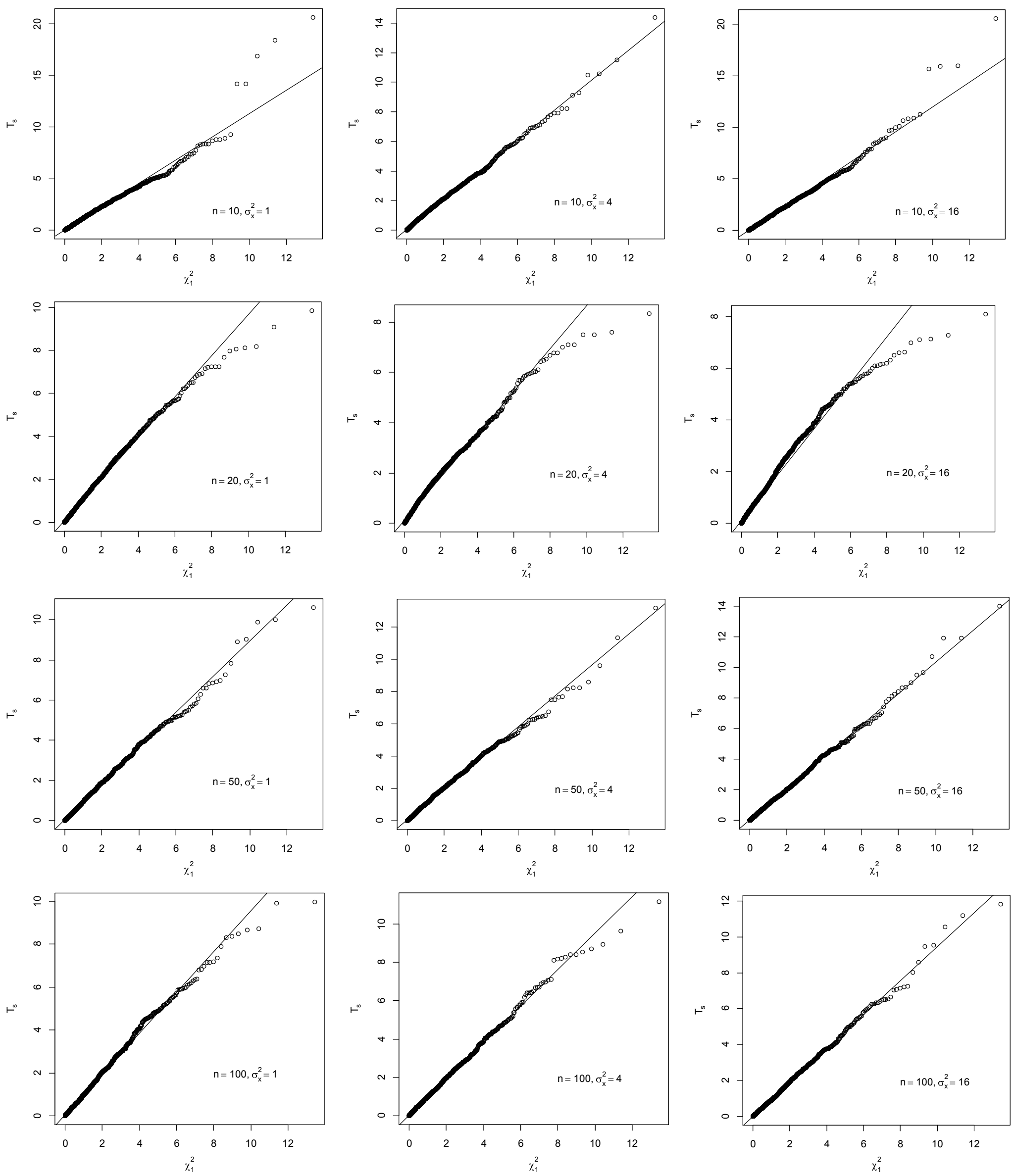

Figure 5.1 Q-Q plot of score test statistic under $\rho = 0$ versus χ_1^2. $(\beta_0, \beta_1, \sigma^2, \mu_x) = (0, 1, 4, 1)$

From the Q-Q plots in Figure 5.1, small sample sizes are tend to produce longer-tailed distributions than the chi-square distribution having one degree of freedom. As sample size increases, the Q-Q plots suggest convergence of the score test statistic null distribution to a chi-square distribution with one degree of freedom.

Table 5.15 A-D test p-values* for LRT test statistic under $\rho = 0$ in structural model

Sample Size	Parameter $(\beta_0, \beta_1, \sigma^2, \mu_x)$	$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=50	(0,1,4,1)	0.0625 (0.260,0.43,0.590)	0.2593 (0.110,0.180,0.300)	0.2535 (0.070,0.210,0.310)
	(1,1,4,1)	0.0818 (0.140,0.440,0.550)	0.2557 (0.070,0.200,0.280)	0.2618 (0.050,0.170,0.270)
	(1,3,4,1)	0.2737 (0.060,0.170,0.300)	0.3355 (0.020,0.110,0.220)	0.3870 (0.030,0.070,0.200)
n=100	(0,1,4,1)	0.3118 (0.100,0.230,0.320)	0.4302 (0.016,0.080,0.130)	0.4686 (0.030,0.050,0.112)
	(1,1,4,1)	0.2729 (0.070,0.200,0.280)	0.4496 (0.014,0.070,0.120)	0.4483 (0.014,0.060,0.124)
	(1,3,4,1)	0.2930 (0.080,0.220,0.300)	0.4174 (0.012,0.060,0.092)	0.4719 (0.012,0.0540,0.094)
n=500	(0,1,4,1)	0.4423 (0.030,0.084,0.130)	0.5831 (0.014,0.040,0.090)	0.4853 (0.010,0.052,0.094)
	(1,1,4,1)	0.4244 (0.022,0.082,0.100)	0.4551 (0.020,0.058,0.080)	0.5540 (0.012,0.048,0.092)
	(1,3,4,1)	0.4446 (0.034,0.084,0.150)	0.5048 (0.016,0.060,0.110)	0.5226 (0.010,0.090,0.100)

*p-values are calculated by comparing one half of test statistic with χ_1^2. Each cell has the form of median of 500 p-values, reject probabilities with 3 different significance level ($\alpha = 0.01, 0.05, 0.1$).

Table 5.15 lists empirical p-values of the A-D GOF test statistics for the LRT test statistic under the null hypothesis of $\rho = 0$. The corresponding Q-Q plots are shown in Figure 5.2 for the parameter setting $(\beta_0, \beta_1, \sigma^2, \mu_x) = (0,1,4,1)$. Again, larger ratios of σ_x^2 to σ^2 require smaller sample sizes for the distribution of the LRT test statistics to be well-approximated by the chi-square distribution with one degree of freedom. Sample sizes of 100 or greater appear to be sufficient for the distribution of the LRT test statistic to be well-approximated by the chi-square with one degree of freedom, even when smaller signal to noise ratios are used.

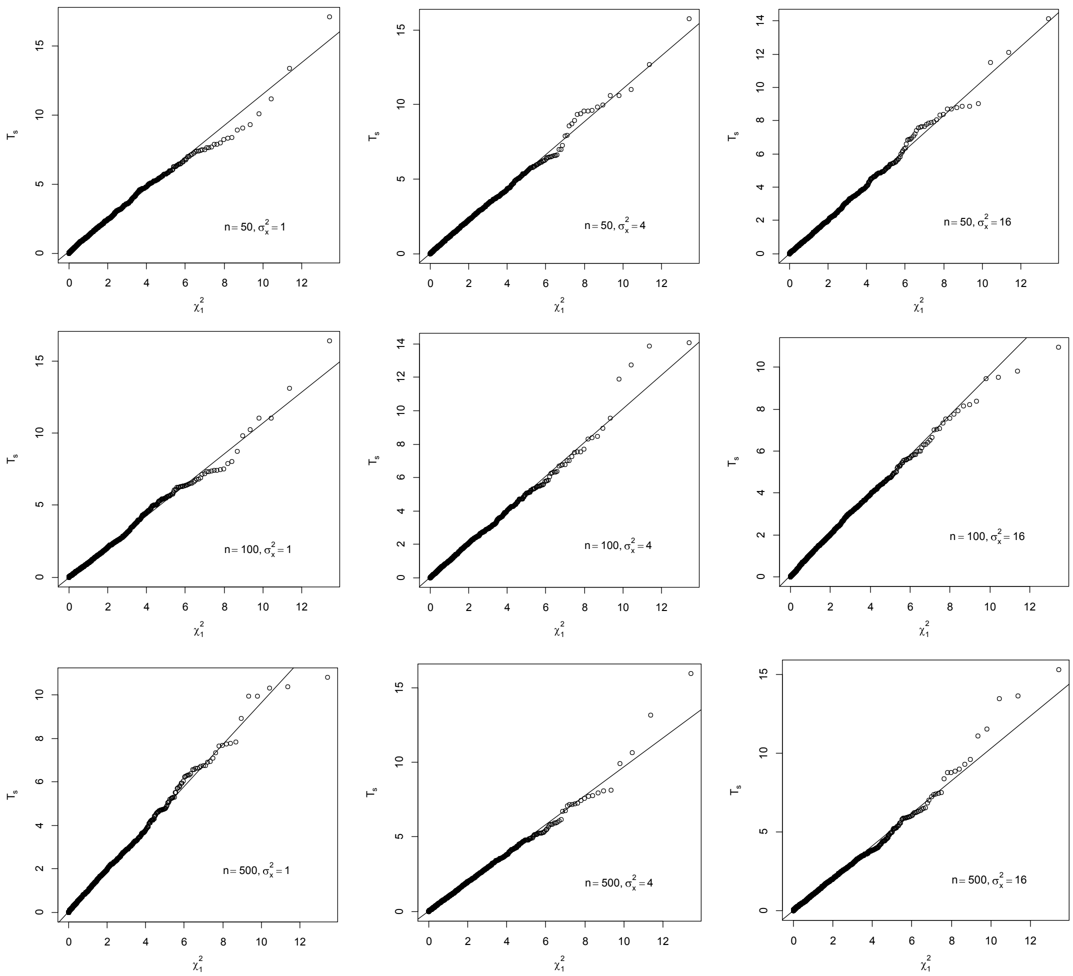

Figure 5.2 Q-Q plot of LRT test statistic under $\rho = 0$ versus χ_1^2. $(\beta_0, \beta_1, \sigma^2, \mu_x) = (0, 1, 4, 1)$

The score test statistic for testing the null hypothesis of $(\beta_0, \beta_1) = (0,1)$ is asymptotically distributed as a chi-square distribution with two degrees of freedom. The A-D GOF p-values in Table 5.16 indicate reasonable goodness of fit for sample sizes of 50 or greater, except when $\rho = 0.9$. The Q-Q plots of the score test statistic empirical quantiles versus quantiles of the chi-square distribution with two degrees of freedom are shown in Figure 5.3 for the case $\rho = 0.5$.

Table 5.16 A-D test p-values* for score test statistic under $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}) = (\mathbf{0}, \mathbf{1})$ in structural model

Sample Size	Parameter Value (μ_x, σ^2, ρ)	$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=50	(1,4,0)	0.3718 (0.026,0.086,0.160)	0.3592 (0.014,0.076,0.152)	0.3591 (0.012,0.062,0.150)
	(1,4,0.5)	0.2489 (0.064,0.208,0.304)	0.1918 (0.056,0.228,0.340)	0.2139 (0.052,0.160,0.274)
	(1,4,0.9)	0.0000 (1.000,1.00,1.000)	0.0000 (1.000,1.000,1.000)	0.0000 (1.000,1.000,1.000)
n=100	(1,4,0)	0.4660 (0.022,0.062,0.122)	0.4665 (0.012,0.068,0.120)	0.4625 (0.012,0.058,0.116)
	(1,4,0.5)	0.3973 (0.022,0.088,0.146)	0.4108 (0.018,0.076,0.156)	0.4255 (0.022,0.080,0.150)
	(1,4,0.9)	0.0003 (0.866,0.954,0.972)	0.0000 (0.980,0.996,0.998)	0.0000 (0.926,0.984,0.990)
n=500	(1,4,0)	0.5245 (0.010,0.046,0.088)	0.4791 (0.008,0.058,0.104)	0.5011 (0.010,0.058,0.090)
	(1,4,0.5)	0.4802 (0.002,0.056,0.116)	0.4869 (0.008,0.048,0.094)	0.5001 (0.012,0.050,0.110)
	(1,4,0.9)	0.2091 (0.070,0.180,0.356)	0.2951 (0.084,0.142,0.340)	0.2965 (0.070,0.122,0.280)

*p-values are calculated by comparing one half of test statistic with χ_2^2. Each cell has the form of median of 1000 p-values, reject probabilities with 3 different significance level ($\alpha = 0.01,0.05,0.1$).

From Table 5.16 and Figure 5.3, when ratio of σ_x^2 over σ^2 is large and there is no correlation, the distribution of score test statistic behaves like a chi-square distribution with sample size is only 50. As sample size increases, the distribution is more and more like chi-square distribution. High correlation among data requires more samples to achieve asymptotic distribution. And so does low ratio of σ_x^2 over σ^2. When there is moderate correlation, the sample size should be above 500 to make the test statistic like chi-square distribution for low signal to noise ratio. When high correlation exists, it needs more samples.

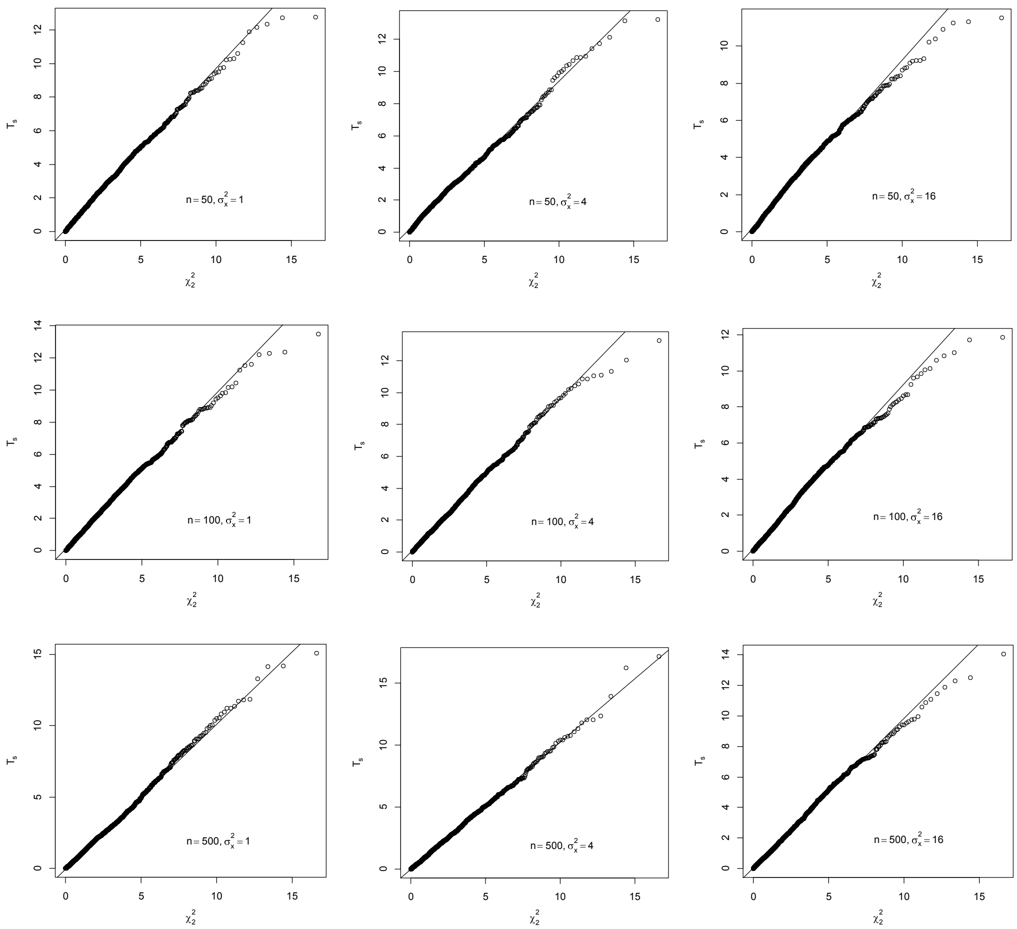

Figure 5.3 Q-Q plot of score test statistic under $(\beta_0, \beta_1) = (0, 1)$ versus χ_2^2. $(\rho, \sigma^2, \mu_x) = (0.5, 4, 1)$

Table 5.17 A-D test p-values* for LRT test statistic under $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}) = (\mathbf{0}, \mathbf{1})$ in structural model

Sample Size	Parameter Value (μ_x, σ^2, ρ)	$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=50	(1,4,0)	0.0000 (0.980,1.000,1.000)	0.0874 (0.230,0.410,0.530)	0.4788 (0.020,0.090,0.140)
	(1,4,0.5)	0.0013 (0.750,0.930,0.970)	0.0950 (0.150,0.340,0.510)	0.1678 (0.070,0.270,0.380)
	(1,4,0.9)	0.0000 (1.000,1.000,1.000)	0.0000 (1.000,1.000,1.000)	0.0001 (0.090,0.980,1.000)
n=100	(1,4,0)	0.0000 (0.980,0.990,1.000)	0.2169 (0.130,0.290,0.350)	0.4267 (0.020,0.140,0.200)
	(1,4,0.5)	0.0005 (0.780,0.950,0.960)	0.3505 (0.090,0.230,0.330)	0.3713 (0.000,0.030,0.120)
	(1,4,0.9)	0.0000 (1.000,1.000,1.000)	0.0000 (1.000,1.000,1.000)	0.0001 (0.090,0.980,1.000)
n=500	(1,4,0)	0.4036 (0.000,0.070,0.180)	0.4842 (0.012,0.058,0.120)	0.5633 (0.010,0.050,0.090)
	(1,4,0.5)	0.0079 (0.510,0.750,0.850)	0.4522 (0.014,0.060,0.140)	0.4551 (0.010,0.040,0.100)
	(1,4,0.9)	0.0001 (0.930,0.980,1.000)	0.0887 (0.160,0.370,0.530)	0.1343 (0.090,0.290,0.400)
n=1000	(1,4,0)	0.4432 (0.000,0..080,0.130)	0.4849 (0.008,0.048,0.090)	0.5819 (0.010,0.050,0.102)
	(1,4,0.5)	0.0673 (0.190,0.470,0.570)	0.4899 (0.010,0.046,0.080)	0.5235 (0.010,0.052,0.094)
	(1,4,0.9)	0. 0012 (0.800,0.890,0.960)	0.2609 (0.060,0.220,0.260)	0.2535 (0.020,0.170,0.300)

*p-values are calculated by comparing one half of test statistic with χ_2^2. Each cell has the form of median of 500 p-values, reject probabilities with 3 different significance level ($\alpha = 0.01, 0.05, 0.1$).

Table 5.17 shows the results of goodness of fit tests for LRT test statistic under $(\beta_0, \beta_1) = (0,1)$ and Figure 5.4 shows the corresponding Q-Q plots with $\rho = 0.5$. It is noticed that when signal to noise ratio is ¼ which is low, even 1000 samples cannot make the test statistic chi-square distributed.

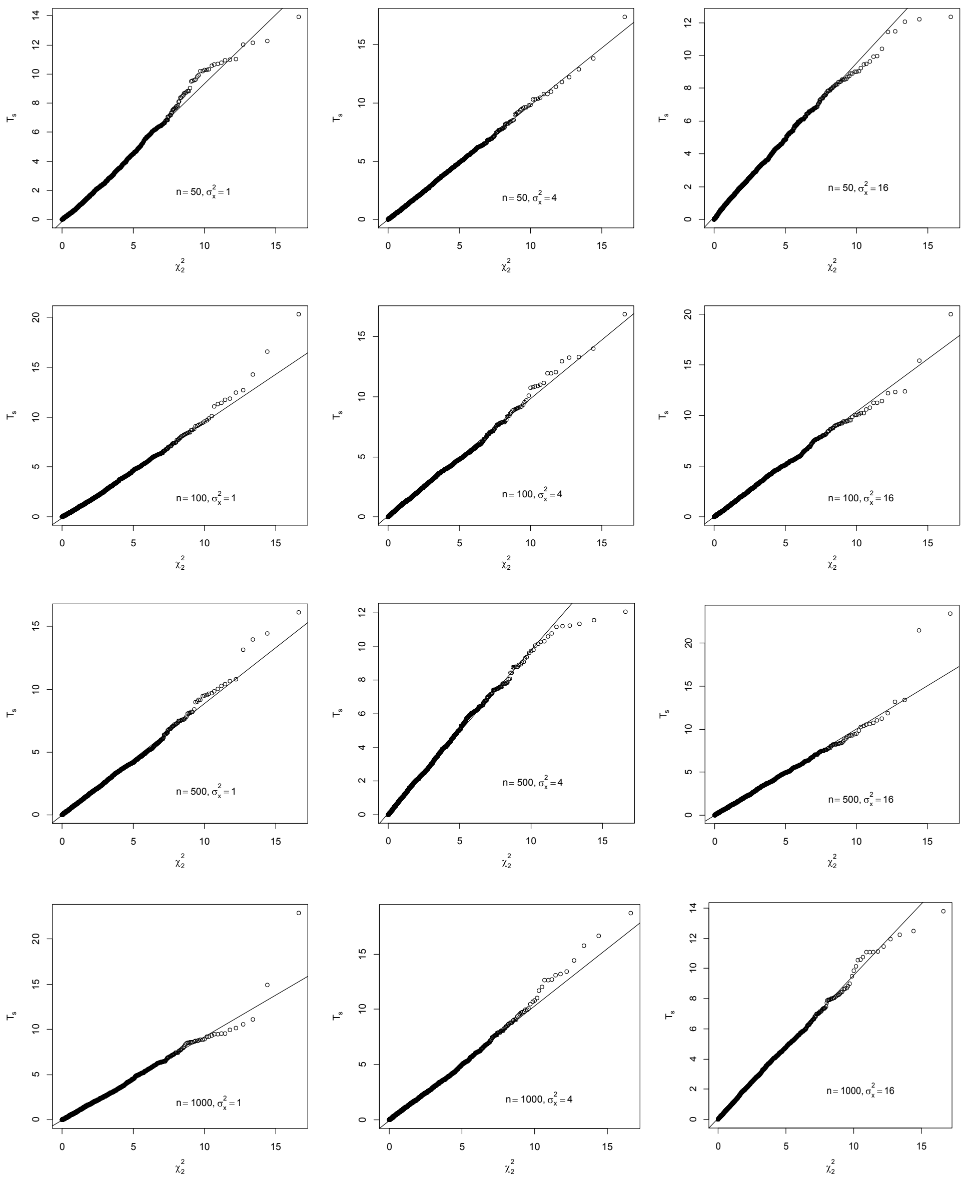

Figure 5.4 Q-Q plot of LRT test statistic under $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}) = (\mathbf{0}, \mathbf{1})$ versus $\boldsymbol{\chi^2_2}$. $(\boldsymbol{\rho}, \boldsymbol{\sigma^2}, \boldsymbol{\mu_x}) = (\mathbf{0.5}, \mathbf{4}, \mathbf{1})$

Generally speaking, when sample size is large enough, all these four test statistics for structural model are approximately chi-square distributed with either one or two degrees of freedom, depending on which hypotheses are tested. Likelihood ratio test statistics need larger sample sizes than the corresponding score test statistics to achieve the same accuracy of approximation. Lower correlation and higher ratio of σ_x^2 over σ^2 improve the performance of all four test statistics.

5.2.2 Small sample and asymptotic behavior of test statistics under the functional model

Simulation parameters in this section replicate those of section 5.2.1, except now data are generated using the functional model as described at the beginning of section 5.1.2. Table 5.18 lists p-values of the A-D GOF test for the score test statistic under the constraint $\rho = 0$. As in the structural case, varying the value of β does not seem to affect the results. Consequently, Q-Q plots in Figure 5.5 are shown only for the case of primary interest, $(\beta_0, \beta_1) = (0,1)$.

Table 5.18 A-D test p-values* for score test statistic under $\rho = 0$ in functional model

Sample Size	Parameter $(\beta_0, \beta_1, \sigma^2, \mu_x)$	$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	(0,1,4,1)	0.0342 (0.3550,0.5450,0.6390)	0.0602 (0.3360,0.4760,0.5740)	0.0547 (0.3540,0.4920,0.5780)
	(1,1,4,1)	0.0332 (0.3820,0.5650,0.6670)	0.0702 (0.3300,0.4680,0.5320)	0.0514 (0.3800,0.4990,0.5710)
	(1,3,4,1)	0.0784 (0.3220,0.4490,0.5340)	0.0673 (0.3390,0.4740,0.5540)	0.0585 (0.3730,0.4270,0.6050)
n=20	(0,1,4,1)	0.2709 (0.0410,0.1100,0.1780)	0.3692 (0.0280,0.1010,0.1700)	0.3982 (0.0160,0.0880,0.1700)
	(1,1,4,1)	0.3559 (0.0370,0.1220,0.1950)	0.3680 (0.0320,0.1060,0.1840)	0.3695 (0.0200,0.0690,0.1500)
	(1,3,4,1)	0.3534 (0.0300,0.1210,0.2110)	0.3607 (0.0310,0.1120,0.2040)	0.3555 (0.0190,0.0980,0.1720)
n=50	(0,1,4,1)	0.4640 (0.0100,0.0590,0.1260)	0.4799 (0.0100,0.0630,0.1230)	0.4937 (0.0110,0.0580,0.1110)
	(1,1,4,1)	0.4756 (0.0190,0.0660,0.1210)	0.4948 (0.0090,0.0480,0.1170)	0.5188 (0.0070,0.0480,0.0880)
	(1,3,4,1)	0.4901 (0.0140,0.0530,0.1030)	0.4943 (0.0050,0.0520,0.0940)	0.5093 (0.0070,0.0430,0.1080)
n=100	(0,1,4,1)	0.4986 (0.0080,0.0520,0.1150)	0.4907 (0.0090,0.0470,0.0910)	0.5191 (0.0070,0.0480,0.0980)
	(1,1,4,1)	0.4734 (0.0120,0.0550,0.1100)	0.4960 (0.0080,0.0480,0.1060)	0.4838 (0.0100,0.0510,0.1000)
	(1,3,4,1)	0.4867 (0.0100,0.0570,0.1130)	0.4944 (0.0110,0.0470,0.1110)	0.5045 (0.0100,0.0490,0.0910)

*p-values are calculated by comparing one half of test statistic with χ_1^2. Each cell has the form of median of 1000 p-values, reject probabilities with 3 different significance level ($\alpha = 0.01,0.05,0.1$).

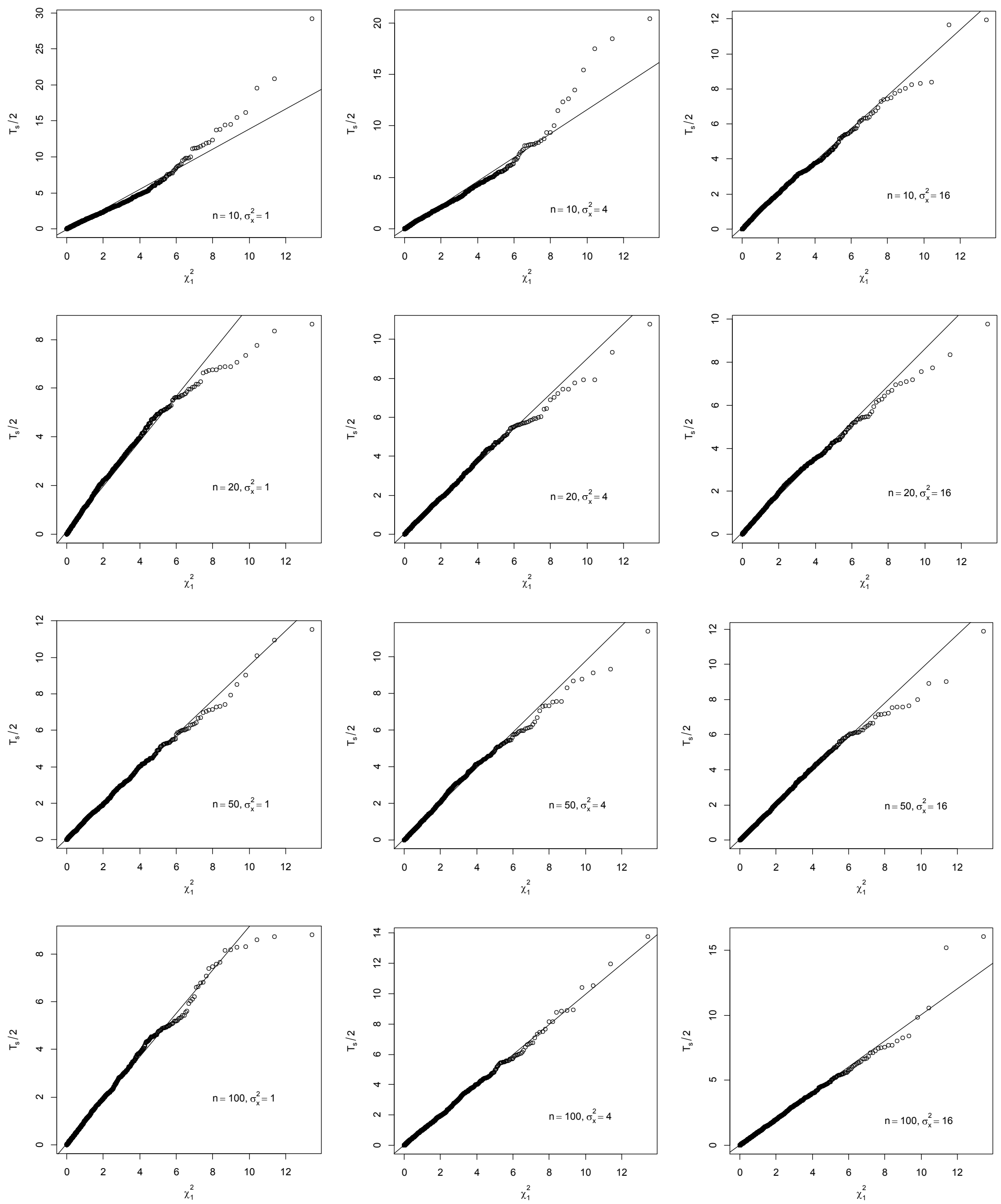

Figure 5.5 Q-Q plot of a half of score test statistic under $\rho = 0$ versus χ_1^2. $(\beta_0, \beta_1, \sigma^2, \mu_x) = (0, 1, 4, 1)$

Sample sizes of 50 or greater appear to be sufficient for the score test statistics to be reasonably well-approximated by a chi-square distribution with one degree of freedom. In particular, sample sizes as small as 20 seem to be sufficient for large sample approximations to be useful when no correlation exist in the data as shown in Figure 5.5 and Table 5.18.

The A-D GOF test results for LRT test statistic when $\rho = 0$ are shown in Table 5.19. In contrast to other cases discussed, values of (β_0, β_1) appear to affect the behaviors of the LRT test statistics. Larger values of β_1 result in somewhat better performance of the LRT test statistic. Sample sizes 100 or greater result in A-D test statistic p-values greater than 0.40 when $\beta_1 = 3$. Again, larger signal to noise ratios also improve the simulated performance of the LRT test statistic.

Table 5.19 A-D test p-values* for LRT test statistic under $\rho = 0$ in functional model

Sample Size	Parameter Value $(\beta_0, \beta_1, \sigma^2, \mu_x)$	$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=50	(0,1,4,1)	0.0661 (0.236,0.457,0.566)	0.1413 (0.147,0.312,0.427)	0.3890 (0.031,0.118,0.203)
	(1,1,4,1)	0.0534 (0.259,0.490,0.614)	0.1086 (0.154,0.358,0.476)	0.2911 (0.064,0.184,0.274)
	(1,3,4,1)	0.1919 (0.078,0.208,0.338)	0.2768 (0.060,0.164,0.266)	0.2429 (0.086,0.222,0.316)
n=100	(0,1,4,1)	0.2858 (0.084,0.188,0.280)	0.3945 (0.046,0.136,0.198)	0.4695 (0.010,0.069,0.140)
	(1,1,4,1)	0.2909 (0.075,0.199,0.305)	0.4388 (0.020,0.081,0.137)	0.4343 (0.016,0.072,0.148)
	(1,3,4,1)	0.4268 (0.024,0.092,0.164)	0.4438 (0.016,0.070,0.117)	0.4850 (0.014,0.061,0.105)
n=500	(0,1,4,1)	0.4257 (0.014,0.080,0.132)	0.5048 (0.014,0.054,0.096)	0.5044 (0.012,0.064,0.102)
	(1,1,4,1)	0.4796 (0.016,0.084,0.126)	0.4905 (0.008,0.052,0.110)	0.4998 (0.012,0.050,0.102)
	(1,3,4,1)	0.4424 (0.014,0.066,0.122)	0.5040 (0.014,0.060,0.092)	0.4809 (0.010,0.046,0.100)

*p-values are calculated by comparing one half of test statistic with χ_1^2. Each cell has the form of median of 1000 p-values, reject probabilities with 3 different significance level ($\alpha = 0.01, 0.05, 0.1$).

Table 5.20 shows A-D GOF test results for the score test statistic for testing the null hypothesis $(\beta_0, \beta_1) = (0,1)$ in the functional case. These results are similar to those observed in the structural case. Increasing either the sample size or the signal to noise ratio results in a better approximation of the small sample distribution of the score test statistic by a chi-square distribution with two degrees of freedom. Furthermore, little or no correlation also improves the chi-square distribution approximation.

Table 5.20 A-D test p-values* for score test statistic under $(\beta_0, \beta_1) = (0, 1)$ in functional model

Sample Size	Parameter (μ_x, σ^2, ρ)	$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=30	(1,4,0)	0.1240 (0.0770,0.2720,0.4410)	0.1288 (0.0690,0.2520,0.4230)	0.1196 (0.0700,0.2810,0.4400)
	(1,4,0.5)	0.0183 (0.3500,0.7660,0.9100)	0.0230 (0.3050,0.7020,0.8710)	0.0229 (0.2950,0.7120,0.8780)
	(1,4,0.9)	0.0000 (1.0000,1.0000,1.0000)	0.0000 (1.0000,1.0000,1.0000)	0.0000 (1.0000,1.0000,1.0000)
n=50	(1,4,0)	0.3006 (0.0240,0.1130,0.1960)	0.3105 (0.0240,0.0990,0.1910)	0.3222 (0.0170,0.0840,0.1650)
	(1,4,0.5)	0.1206 (0.0580,0.2650,0.4510)	0.1370 (0.0480,0.2230,0.4040)	0.1487 (0.0610,0.2080,0.3720)
	(1,4,0.9)	0.0000 (1.0000,1.0000,1.0000)	0.0000 (1.0000,1.0000,1.0000)	0.0000 (1.0000,1.0000,1.0000)
n=100	(1,4,0)	0.4604 (0.0050,0.0480,0.0960)	0.4494 (0.0110,0.0570,0.1000)	0.4435 (0.0110,0.04200,0.1070)
	(1,4,0.5)	0.3550 (0.0200,0.0850,0.1580)	0.3487 (0.0130,0.0870,0.1630)	0.3302 (0.0140,0.0790,0.1440)
	(1,4,0.9)	0.0041 (0.7070,0.9620,0.9930)	0.0040 (0.7320,0.9630,0.9900)	0.0051 (0.6630,0.9480,0.9830)
n=500	(1,4,0)	0.4963 (0.0110,0.0510,0.0980)	0.5273 (0.0110,0.0560,0.1090)	0.4946 (0.0100,0.0460,0.0960)
	(1,4,0.5)	0.5140 (0.0070,0.0550,0.1010)	0.4931 (0.0110,0.0550,0.0780)	0.5283 (0.0110,0.0475,0.0910)
	(1,4,0.9)	0.4046 (0.0170,0.0663,0.1300)	0.3768 (0.0075,0.0650,0.1250)	0.3926 (0.0130,0.0763,0.1510)
n=1000	(1,4,0)	0.5030 (0.0110,0.0520,0.1000)	0.5134 (0.0080,0.0460,0.0880)	0.4736 (0.0100,0.0480,0.1000)
	(1,4,0.5)	0.5080 (0.0120,0.0540,0.0960)	0.5206 (0.0090,0.0480,0.0980)	0.4674 (0.0090,0.0520,0.1060)
	(1,4,0.9)	0.4476 (0.0140,0.0560,0.0880)	0.4740 (0.0080,0.0520,0.1100)	0.5042 (0.0140,0.0480,0.0940)

*p-values are calculated by comparing one half of test statistic with χ_2^2. Each cell has the form of median of 1000 p-values, reject probabilities with 3 different significance level ($\alpha = 0.01, 0.05, 0.1$).

In general, larger sample sizes are required to achieve adequate approximations when there is correlation among data. And also, more samples are needed if signal to noise ratio is low. When the correlation is strong and/or the signal to noise ratio is low, sample size of 500, even 1000 is required.

5.3 Empirical Type I Error Rate for Test Statistics

Empirical Type I error rates of each test statistic are studied in this section. For each estimated Type I error rate, three types of standard errors are calculated and recorded in tables below. The first standard error is a conservative standard error calculated as $\sqrt{\alpha(1-\alpha)/N}$ with $\alpha = 0.5$. The second standard error is calculated using the nominal type I error rate. The third standard error is calculated using the estimate of α from the simulated data. The number of data sets simulated to estimate each type I error rate is $N = 5{,}000$.

5.3.1 Empirical type I error rates for the structural model test statistics

Tables 5.21-5.23 show estimated type I error rates and their three types of estimated standard errors for score test statistics of $\rho = 0$ with various values of parameters.

Tables 5.21 and 5.22 show that for a nominal type I error rates of 0.01 and 0.05, the estimated type I error rates generally do not differ significantly from the nominal rates when the signal to noise ratio is one or greater and sample size is greater than 50 using estimated standard error. But when signal to noise ratio is 0.25, hundreds of samples are needed to reach the nominal type I error rate. If we use conservative standard error, sample size of 10 seems enough.

Table 5.21 Empirical type I error rates for the structural model score test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.01$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.01$ (conservative S.E.=0.0071, nominal S.E.=0.0014)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.0134 (0.0016)	0.0104 (0.0014)	0.0119 (0.0016)
	$(\beta_0, \beta_1) = (1,1)$	0.0138 (0.0016)	0.0114 (0.0015)	0.0112 (0.0015)
	$(\beta_0, \beta_1) = (1,3)$	0.0100 (0.0014)	0.0102 (0.0014)	0.0110 (0.0015)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.0077 (0.0013)	0.0061 (0.0011)	0.0064 (0.0011)
	$(\beta_0, \beta_1) = (1,1)$	0.0060 (0.0011)	0.0082 (0.0013)	0.0058 (0.0011)
	$(\beta_0, \beta_1) = (1,3)$	0.0076 (0.0012)	0.0084 (0.0013)	0.0078 (0.0012)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.0070 (0.0011)	0.0067 (0.0011)	0.0098 (0.0014)
	$(\beta_0, \beta_1) = (1,1)$	0.0074 (0.0012)	0.0072 (0.0012)	0.0090 (0.0013)
	$(\beta_0, \beta_1) = (1,3)$	0.0070 (0.0012)	0.0076 (0.0012)	0.0090 (0.0013)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.0082 (0.0013)	0.0091 (0.0013)	0.0095 (0.0014)
	$(\beta_0, \beta_1) = (1,1)$	0.0086 (0.0013)	0.0100 (0.0014)	0.0084 (0.0013)
	$(\beta_0, \beta_1) = (1,3)$	0.0094 (0.0014)	0.0082 (0.0013)	0.0080 (0.0013)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.0117 (0.0016)	0.0089 (0.0013)	0.0094 (0.0014)
	$(\beta_0, \beta_1) = (1,1)$	0.0112 (0.0015)	0.0098 (0.0014)	0.0096 (0.0014)
	$(\beta_0, \beta_1) = (1,3)$	0.0092 (0.0014)	0.0084 (0.0013)	0.0114 (0.0015)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.0095 (0.0014)	0.0109 (0.0015)	0.0095 (0.0014)
	$(\beta_0, \beta_1) = (1,1)$	0.0100 (0.0014)	0.0110 (0.0015)	0.0124 (0.0016)
	$(\beta_0, \beta_1) = (1,3)$	0.0096 (0.0014)	0.0132 (0.0016)	0.0126 (0.0016)

*Test statistic is compared with χ_1^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from nominal rate using estimated S.E. and blue for using conservative S.E.

Table 5.22 Empirical type I error rates for the structural model score test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.05$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.05$ (conservative S.E.=0.0071, nominal S.E.=0.0031)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.0601 (0.0034)	0.0532 (0.0031)	0.0516 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0590 (0.0033)	0.0600 (0.0034)	0.0548 (0.0032)
	$(\beta_0, \beta_1) = (1,3)$	0.0566 (0.0033)	0.0486 (0.0030)	0.0524 (0.0032)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.0497 (0.0031)	0.0434 (0.0028)	0.0455 (0.0030)
	$(\beta_0, \beta_1) = (1,1)$	0.0426 (0.0029)	0.0484 (0.0030)	0.0462 (0.0030)
	$(\beta_0, \beta_1) = (1,3)$	0.0460 (0.0030)	0.0490 (0.0031)	0.0402 (0.0028)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.0459 (0.0030)	0.0451 (0.0030)	0.0490 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0444 (0.0029)	0.0456 (0.0030)	0.0484 (0.0030)
	$(\beta_0, \beta_1) = (1,3)$	0.0478 (0.0030)	0.0474 (0.0030)	0.0470 (0.0030)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.0511 (0.0031)	0.0481 (0.0030)	0.0502 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0552 (0.0032)	0.0490 (0.0031)	0.0476 (0.0030)
	$(\beta_0, \beta_1) = (1,3)$	0.0528 (0.0032)	0.0476 (0.0030)	0.0466 (0.0030)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.0503 (0.0031)	0.0508 (0.0031)	0.0473 (0.0030)
	$(\beta_0, \beta_1) = (1,1)$	0.0510 (0.0031)	0.0492 (0.0031)	0.0464 (0.0030)
	$(\beta_0, \beta_1) = (1,3)$	0.0470 (0.0030)	0.0464 (0.0030)	0.0504 (0.0031)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.0490 (0.0031)	0.0479 (0.0030)	0.0501 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0514 (0.0031)	0.0506 (0.0031)	0.0526 (0.0032)
	$(\beta_0, \beta_1) = (1,3)$	0.0462 (0.0030)	0.0554 (0.0032)	0.0550 (0.0032)

*Test statistic is compared with χ_1^2. Each cell has value of estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for conservative S.E.

Table 5.23 shows that for type I error rates 0.1, the estimated type I error rate is close to the nominal even when sample size is as small as 20 using either estimated standard error or conservative standard error.

Table 5.23 Empirical type I error rates for the structural model score test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.1$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.1$ (conservative S.E.=0.0071, nominal S.E.=0.0042)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.1124 (0.0045)	0.1119 (0.0045)	0.1100 (0.0044)
	$(\beta_0, \beta_1) = (1,1)$	0.1132 (0.0045)	0.1168 (0.0045)	0.1120 (0.0045)
	$(\beta_0, \beta_1) = (1,3)$	0.1124 (0.0045)	0.1100 (0.0044)	0.1054 (0.0043)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.0961 (0.0041)	0.0949 (0.0041)	0.0953 (0.0041)
	$(\beta_0, \beta_1) = (1,1)$	0.0970 (0.0042)	0.1040 (0.0043)	0.0984 (0.0042)
	$(\beta_0, \beta_1) = (1,3)$	0.0968 (0.0042)	0.1048 (0.0043)	0.0976 (0.0042)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.0970 (0.0042)	0.0929 (0.0041)	0.0996 (0.0042)
	$(\beta_0, \beta_1) = (1,1)$	0.0928 (0.0041)	0.0940 (0.0042)	0.0990 (0.0042)
	$(\beta_0, \beta_1) = (1,3)$	0.0960 (0.0042)	0.1002 (0.0042)	0.1016 (0.0043)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.0997 (0.0042)	0.0988 (0.0042)	0.0965 (0.0042)
	$(\beta_0, \beta_1) = (1,1)$	0.1018 (0.0043)	0.0928 (0.0041)	0.0962 (0.0042)
	$(\beta_0, \beta_1) = (1,3)$	0.1052 (0.0043)	0.0978 (0.0042)	0.0958 (0.0042)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.0968 (0.0042)	0.1009 (0.0042)	0.0978 (0.0042)
	$(\beta_0, \beta_1) = (1,1)$	0.0980 (0.0042)	0.0956 (0.0042)	0.0928 (0.0041)
	$(\beta_0, \beta_1) = (1,3)$	0.0926 (0.0041)	0.0992 (0.0042)	0.1044 (0.0043)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.0956 (0.0041)	0.0951 (0.0041)	0.1034 (0.0043)
	$(\beta_0, \beta_1) = (1,1)$	0.0998 (0.0042)	0.0990 (0.0042)	0.1038 (0.0043)
	$(\beta_0, \beta_1) = (1,3)$	0.0986 (0.0042)	0.1022 (0.0043)	0.1050 (0.0043)

*Test statistic is compared with χ_1^2. Each cell has value of estimated type I error (estimated S.E.). Red indicates significant difference from nominal rate using estimated S.E. and blue for conservative S.E.

Tables 5.24-5.26 show estimated type I error rates and their three types of estimated standard errors for LRT tests of the null hypothesis $\rho = 0$ using various values of parameters. In general, increasing the sample size will increase the accuracy of the empirical type I error rate.

Table 5.24 Empirical type I error rates for the structural model likelihood ratio test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.01$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.01$ (conservative S.E.=0.0071, nominal S.E.=0.0014)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.0226 (0.0021)	0.0174 (0.0018)	0.0097 (0.0014)
	$(\beta_0, \beta_1) = (1,1)$	0.0169 (0.0018)	0.0197 (0.0020)	0.0122 (0.0016)
	$(\beta_0, \beta_1) = (1,3)$	0.0181 (0.0019)	0.0074 (0.0012)	0.0057 (0.0011)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.0235 (0.0021)	0.0147 (0.0017)	0.0125 (0.0016)
	$(\beta_0, \beta_1) = (1,1)$	0.0194 (0.0019)	0.0154 (0.0017)	0.0118 (0.0015)
	$(\beta_0, \beta_1) = (1,3)$	0.0149 (0.0017)	0.0099 (0.0014)	0.0101 (0.0014)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.0119 (0.0015)	0.0131 (0.0016)	0.0128 (0.0016)
	$(\beta_0, \beta_1) = (1,1)$	0.0129 (0.0016)	0.0128 (0.0016)	0.0112 (0.0015)
	$(\beta_0, \beta_1) = (1,3)$	0.0103 (0.0014)	0.0122 (0.0016)	0.0120 (0.0015)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.0079 (0.0013)	0.0116 (0.0015)	0.0116 (0.0015)
	$(\beta_0, \beta_1) = (1,1)$	0.0106 (0.0015)	0.0094 (0.0014)	0.0130 (0.0016)
	$(\beta_0, \beta_1) = (1,3)$	0.0084 (0.0013)	0.0108 (0.0015)	0.0114 (0.0015)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.0119 (0.0015)	0.0106 (0.0014)	0.0114 (0.0015)
	$(\beta_0, \beta_1) = (1,1)$	0.0102 (0.0014)	0.0096 (0.0014)	0.0104 (0.0014)
	$(\beta_0, \beta_1) = (1,3)$	0.0100 (0.0014)	0.0130 (0.0016)	0.0092 (0.0014)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.0106 (0.0014)	0.0082 (0.0013)	0.0116 (0.0015)
	$(\beta_0, \beta_1) = (1,1)$	0.0102 (0.0014)	0.0122 (0.0016)	0.0094 (0.0014)
	$(\beta_0, \beta_1) = (1,3)$	0.0130 (0.0016)	0.0094 (0.0014)	0.0120 (0.0015)

*Test statistic is compared with χ_1^2. Each cell has value of estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for conservative S.E.

Table 5.25 Empirical type I error rates for the structural model likelihood ratio test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.05$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.05$ (conservative S.E.=0.0071, nominal S.E.=0.0031)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.0797 (0.0038)	0.0847 (0.0039)	0.0438 (0.0029)
	$(\beta_0, \beta_1) = (1,1)$	0.0889 (0.0040)	0.0772 (0.0038)	0.0527 (0.0032)
	$(\beta_0, \beta_1) = (1,3)$	0.0841 (0.0039)	0.0475 (0.0030)	0.0343 (0.0026)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.0798 (0.0038)	0.0698 (0.0036)	0.0594 (0.0033)
	$(\beta_0, \beta_1) = (1,1)$	0.0700 (0.0036)	0.0727 (0.0037)	0.0588 (0.0033)
	$(\beta_0, \beta_1) = (1,3)$	0.0609 (0.0034)	0.0555 (0.0032)	0.0483 (0.0030)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.0596 (0.0033)	0.0602 (0.0034)	0.0558 (0.0032)
	$(\beta_0, \beta_1) = (1,1)$	0.0576 (0.0033)	0.0538 (0.0032)	0.0548 (0.0032)
	$(\beta_0, \beta_1) = (1,3)$	0.0531 (0.0032)	0.0538 (0.0032)	0.0528 (0.0032)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.0543 (0.0032)	0.0608 (0.0034)	0.0512 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0577 (0.0033)	0.0526 (0.0032)	0.0522 (0.0032)
	$(\beta_0, \beta_1) = (1,3)$	0.0472 (0.0030)	0.0556 (0.0032)	0.0526 (0.0032)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.0552 (0.0032)	0.0522 (0.0031)	0.0486 (0.0030)
	$(\beta_0, \beta_1) = (1,1)$	0.0458 (0.0030)	0.0504 (0.0031)	0.0468 (0.0030)
	$(\beta_0, \beta_1) = (1,3)$	0.0472 (0.0030)	0.0536 (0.0032)	0.0510 (0.0031)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.0538 (0.0032)	0.0472 (0.0030)	0.0494 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0472 (0.0030)	0.0550 (0.0032)	0.0456 (0.0030)
	$(\beta_0, \beta_1) = (1,3)$	0.0494 (0.0031)	0.0516 (0.0031)	0.0502 (0.0031)

*Test statistic is compared with χ_1^2. Each cell has value of estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for conservative S.E.

Table 5.24 shows that sample sizes of 50 or greater make the empirical type I error rates generally not differ significantly from the nominal rate 0.01 using estimated standard error. But a small sample size of 10 is needed using conservative standard error. Table 5.25 shows that big sample size of 500 is required to make the empirical type I error rates not differ significantly from the nominal rate 0.05 using estimated standard error, however, only size of 50 is needed using conservative standard error.

Table 5.26 Empirical type I error rates for the structural model likelihood ratio test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.1$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.1$ (conservative S.E.=0.0071, nominal S.E.=0.0042)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.1531 (0.0051)	0.1421 (0.0049)	0.0944 (0.0041)
	$(\beta_0, \beta_1) = (1,1)$	0.1523 (0.0051)	0.1314 (0.0048)	0.1077 (0.0044)
	$(\beta_0, \beta_1) = (1,3)$	0.1488 (0.0050)	0.1038 (0.0043)	0.0868 (0.0040)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.1438 (0.0050)	0.1309 (0.0048)	0.1145 (0.0045)
	$(\beta_0, \beta_1) = (1,1)$	0.1213 (0.0046)	0.1274 (0.0047)	0.1142 (0.0045)
	$(\beta_0, \beta_1) = (1,3)$	0.1281 (0.0047)	0.1079 (0.0044)	0.0990 (0.0042)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.1192 (0.0046)	0.1240 (0.0047)	0.1114 (0.0044)
	$(\beta_0, \beta_1) = (1,1)$	0.1114 (0.0044)	0.1224 (0.0046)	0.1152 (0.0045)
	$(\beta_0, \beta_1) = (1,3)$	0.1020 (0.0043)	0.1080 (0.0044)	0.1088 (0.0044)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.1084 (0.0044)	0.1094 (0.0044)	0.1038 (0.0043)
	$(\beta_0, \beta_1) = (1,1)$	0.1133 (0.0045)	0.1058 (0.0043)	0.1046 (0.0045)
	$(\beta_0, \beta_1) = (1,3)$	0.1021 (0.0043)	0.1098 (0.0044)	0.1070 (0.0044)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.1059 (0.0044)	0.0968 (0.0042)	0.0992 (0.0042)
	$(\beta_0, \beta_1) = (1,1)$	0.0912 (0.0041)	0.0962 (0.0042)	0.0962 (0.0042)
	$(\beta_0, \beta_1) = (1,3)$	0.0986 (0.0042)	0.1054 (0.0043)	0.1020 (0.0043)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.0999 (0.0042)	0.0952 (0.0042)	0.1058 (0.0043)
	$(\beta_0, \beta_1) = (1,1)$	0.0918 (0.0041)	0.0994 (0.0042)	0.0980 (0.0042)
	$(\beta_0, \beta_1) = (1,3)$	0.1014 (0.0043)	0.1004 (0.0043)	0.1006 (0.0043)

*Half of test statistic is compared with χ_1^2. Each cell has value of estimated type I error (estimated S.E.). Red indicates significant difference from nominal rate using estimated S.E. and blue for conservative S.E.

Table 5.26 shows that sample size greater than 500 is needed generally for the empirical type I error rates not significantly different from the nominal rate 0.1 using either estimated standard error or conservative standard error. When signal to noise ratio is large, sample size of 100 is enough.

Table 5.27 Empirical type I error rates for the structural model score test of the null hypothesis $(\beta_0, \beta_1) = (0, 1)$ using a nominal type I error rate of $\alpha = 0.01$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.01$ (conservative S.E.=0.0071, nominal S.E.=0.0014)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$\rho = 0$	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
	$\rho = 0.5$	0.0008 (0.0004)	0.0004 (0.0003)	0.0014 (0.0005)
	$\rho = 0.9$	0.0036 (0.0008)	0.0022 (0.0007)	0.0026 (0.0007)
n=20	$\rho = 0$	0.0028 (0.0007)	0.0012 (0.0005)	0.0020 (0.0006)
	$\rho = 0.5$	0.0024 (0.0007)	0.0022 (0.0007)	0.0026 (0.0007)
	$\rho = 0.9$	0.0030 (0.0008)	0.0074 (0.0012)	0.0054 (0.0010)
n=50	$\rho = 0$	0.0054 (0.0010)	0.0068 (0.0012)	0.0058 (0.0011)
	$\rho = 0.5$	0.0060 (0.0011)	0.0066 (0.0011)	0.0046 (0.0010)
	$\rho = 0.9$	0.0078 (0.0012)	0.0080 (0.0013)	0.0092 (0.0014)
n=100	$\rho = 0$	0.0060 (0.0011)	0.0092 (0.0014)	0.0088 (0.0013)
	$\rho = 0.5$	0.0084 (0.0013)	0.0090 (0.0013)	0.0088 (0.0013)
	$\rho = 0.9$	0.0088 (0.0013)	0.0134 (0.0016)	0.0092 (0.0014)
n=500	$\rho = 0$	0.0094 (0.0014)	0.0104 (0.0014)	0.0114 (0.0015)
	$\rho = 0.5$	0.0098 (0.0014)	0.0088 (0.0013)	0.0100 (0.0014)
	$\rho = 0.9$	0.0104 (0.0014)	0.0114 (0.0015)	0.0118 (0.0015)
n=1000	$\rho = 0$	0.0084 (0.0013)	0.0108 (0.0015)	0.0078 (0.0012)
	$\rho = 0.5$	0.0092 (0.0014)	0.0110 (0.0015)	0.0092 (0.0014)
	$\rho = 0.9$	0.0102 (0.0014)	0.0090 (0.0013)	0.0128 (0.0016)

*Test statistic is compared with χ_2^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Tables 5.27-5.29 show estimated type I error rates and their three types of estimated standard errors for score tests of the null hypothesis $(\beta_0, \beta_1) = (0,1)$ using various values of parameters. In general, increasing the sample size will increase the accuracy of the empirical type I error rate.

Table 5.28 Empirical type I error rates for the structural model score test of the null hypothesis $(\beta_0, \beta_1) = (0, 1)$ using a nominal type I error rate of $\alpha = 0.05$

Sample Size	Parameter Value	Type I Error $\alpha = 0.05$ (conservative S.E.=0.0071, nominal S.E.=0.0031)		
	$(\sigma^2, \mu_x) = (4,1)$	$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$\rho = 0$	0.0138 (0.0016)	0.0126 (0.0016)	0.0124 (0.0016)
	$\rho = 0.5$	0.0232 (0.0021)	0.0216 (0.0021)	0.0244 (0.0022)
	$\rho = 0.9$	0.0438 (0.0029)	0.0458 (0.0030)	0.0530 (0.0032)
n=20	$\rho = 0$	0.0322 (0.0025)	0.0284 (0.0023)	0.0256 (0.0022)
	$\rho = 0.5$	0.0366 (0.0027)	0.0340 (0.0026)	0.0288 (0.0024)
	$\rho = 0.9$	0.0472 (0.0030)	0.0652 (0.0035)	0.0592 (0.0033)
n=50	$\rho = 0$	0.0432 (0.0029)	0.0410 (0.0028)	0.0424 (0.0028)
	$\rho = 0.5$	0.0416 (0.0028)	0.0438 (0.0029)	0.0390 (0.0027)
	$\rho = 0.9$	0.0488 (0.0030)	0.0702 (0.0036)	0.0612 (0.0034)
n=100	$\rho = 0$	0.0448 (0.0029)	0.0466 (0.0030)	0.0456 (0.0030)
	$\rho = 0.5$	0.0476 (0.0030)	0.0440 (0.0029)	0.0446 (0.0029)
	$\rho = 0.9$	0.0582 (0.0033)	0.0688 (0.0036)	0.0560 (0.0033)
n=500	$\rho = 0$	0.0502 (0.0031)	0.0488 (0.0030)	0.0540 (0.0032)
	$\rho = 0.5$	0.0484 (0.0030)	0.0496 (0.0031)	0.0518 (0.0031)
	$\rho = 0.9$	0.0586 (0.0033)	0.0570 (0.0033)	0.0566 (0.0043)
n=1000	$\rho = 0$	0.0470 (0.0030)	0.0480 (0.0030)	0.0464 (0.0030)
	$\rho = 0.5$	0.0488 (0.0030)	0.0474 (0.0030)	0.0518 (0.0031)
	$\rho = 0.9$	0.0506 (0.0031)	0.0500 (0.0031)	0.0534 (0.0032)

*Test statistic is compared with χ_2^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

From Table 5.27, we observe that the estimated type I error is close to nominal when sample size is greater than 100 and use estimated standard error. Sample size of 10 seems enough when using conservative standard error. Table 5.28 shows that when sample size is 100 or greater, the estimated type I error rate is close to the nominal error rate 0.05 with large signal to noise ratio. Using estimated standard error, large sample size greater than 500 is needed for low signal to noise ratio and high correlation. Using conservative standard error, sample size greater than 100 is enough.

Table 5.29 Empirical type I error rates for the structural model score test of the null hypothesis $(\beta_0, \beta_1) = (0, 1)$ using a nominal type I error rate of $\alpha = 0.1$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.1$ (conservative S.E.=0.0071, nominal S.E.=0.0042)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$\rho = 0$	0.0624 (0.0034)	0.0570 (0.0033)	0.0534 (0.0032)
	$\rho = 0.5$	0.0804 (0.0038)	0.0776 (0.0038)	0.0742 (0.0037)
	$\rho = 0.9$	0.1222 (0.0046)	0.1320 (0.0048)	0.1320 (0.0048)
n=20	$\rho = 0$	0.0781 (0.0042)	0.0768 (0.0038)	0.0744 (0.0037)
	$\rho = 0.5$	0.0918 (0.0041)	0.0908 (0.0041)	0.0728 (0.0037)
	$\rho = 0.9$	0.1224 (0.0046)	0.1542 (0.0051)	0.1384 (0.0049)
n=50	$\rho = 0$	0.0966 (0.0042)	0.0944 (0.0041)	0.0918 (0.0041)
	$\rho = 0.5$	0.0990 (0.0042)	0.0956 (0.0042)	0.0927 (0.0041)
	$\rho = 0.9$	0.1232 (0.0046)	0.1526 (0.0051)	0.1378 (0.0049)
n=100	$\rho = 0$	0.0982 (0.0042)	0.0994 (0.0042)	0.0918 (0.0041)
	$\rho = 0.5$	0.0970 (0.0042)	0.0918 (0.0041)	0.0919 (0.0041)
	$\rho = 0.9$	0.1182 (0.0046)	0.1400 (0.0049)	0.1206 (0.0046)
n=500	$\rho = 0$	0.1016 (0.0043)	0.1014 (0.0043)	0.1014 (0.0043)
	$\rho = 0.5$	0.0970 (0.0042)	0.1004 (0.0043)	0.1050 (0.0043)
	$\rho = 0.9$	0.1152 (0.0045)	0.1124 (0.0045)	0.1030 (0.0043)
n=1000	$\rho = 0$	0.0982 (0.0042)	0.0919 (0.0041)	0.0918 (0.0041)
	$\rho = 0.5$	0.1000 (0.0042)	0.0996 (0.0042)	0.1042 (0.0043)
	$\rho = 0.9$	0.0964 (0.0042)	0.1000 (0.0042)	0.1088 (0.0044)

*Test statistic is compared with χ_2^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Table 5.29 shows that generally sample size of 1000 or greater is needed to make estimated type I error rates not differ from the nominal error rate 0.1 using either estimated standard error or conservative standard error. Notice that for moderate or no correlation, sample size of 50 is enough.

Table 5.30 Empirical type I error rates for the structural model likelihood ratio test of the null hypothesis $(\beta_0, \beta_1) = (0, 1)$ using a nominal type I error rate of $\alpha = 0.01$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.01$ (conservative S.E.=0.0071, nominal S.E.=0.0014)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$\rho = 0$	0.0113 (0.0015)	0.0087 (0.0013)	0.0102 (0.0014)
	$\rho = 0.5$	0.0215 (0.0021)	0.0252 (0.0022)	0.0236 (0.0021)
	$\rho = 0.9$	0.0523 (0.0031)	0.0850 (0.0039)	0.0791 (0.0038)
n=20	$\rho = 0$	0.0065 (0.0011)	0.0067 (0.0012)	0.0148 (0.0017)
	$\rho = 0.5$	0.0135 (0.0016)	0.0134 (0.0016)	0.0108 (0.0015)
	$\rho = 0.9$	0.0545 (0.0032)	0.0014 (0.0038)	0.0698 (0.0036)
n=50	$\rho = 0$	0.0045 (0.0009)	0.0112 (0.0015)	0.0086 (0.0013)
	$\rho = 0.5$	0.0067 (0.0012)	0.0140 (0.0017)	0.0124 (0.0016)
	$\rho = 0.9$	0.0307 (0.0024)	0.0423 (0.0028)	0.0416 (0.0028)
n=100	$\rho = 0$	0.0077 (0.0012)	0.0078 (0.0012)	0.0102 (0.0014)
	$\rho = 0.5$	0.0088 (0.0013)	0.0082 (0.0013)	0.0092 (0.0014)
	$\rho = 0.9$	0.0198 (0.0020)	0.0308 (0.0024)	0.0282 (0.0023)
n=500	$\rho = 0$	0.0076 (0.0012)	0.0090 (0.0013)	0.0128 (0.0016)
	$\rho = 0.5$	0.0048 (0.0010)	0.0112 (0.0015)	0.0116 (0.0015)
	$\rho = 0.9$	0.0100 (0.0014)	0.0172 (0.0018)	0.0154 (0.0017)
n=1000	$\rho = 0$	0.0090 (0.0013)	0.0098 (0.0014)	0.0110 (0.0015)
	$\rho = 0.5$	0.0080 (0.0013)	0.0122 (0.0016)	0.0092 (0.0014)
	$\rho = 0.9$	0.0090 (0.0013)	0.0130 (0.0016)	0.0129 (0.0016)

*Test statistic is compared with χ_2^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Tables 5.30-5.32 show estimated type I error rates and their three types of estimated standard errors for score tests of the null hypothesis $(\beta_0, \beta_1) = (0,1)$ using various values of parameters. In general, increasing the sample size will increase the accuracy of the empirical type I error rate.

Table 5.31 Empirical type I error rates for the structural model likelihood ratio test of the null hypothesis $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}) = (\mathbf{0}, \mathbf{1})$ using a nominal type I error rate of $\boldsymbol{\alpha} = \mathbf{0.05}$

Sample Size	Parameter Value	Type I Error $\alpha = 0.05$ (conservative S.E.=0.0071, nominal S.E.=0.0031)		
	$(\sigma^2, \mu_x) = (4,1)$	$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$\rho = 0$	0.0374 (0.0027)	0.0414 (0.0028)	0.0488 (0.0030)
	$\rho = 0.5$	0.0785 (0.0038)	0.0951 (0.0041)	0.0973 (0.0042)
	$\rho = 0.9$	0.1731 (0.0053)	0.2252 (0.0059)	0.2366 (0.0060)
n=20	$\rho = 0$	0.0256 (0.0022)	0.0402 (0.0028)	0.0528 (0.0032)
	$\rho = 0.5$	0.0541 (0.0032)	0.0624 (0.0034)	0.0602 (0.0034)
	$\rho = 0.9$	0.1581 (0.0052)	0.2073 (0.0057)	0.1998 (0.0057)
n=50	$\rho = 0$	0.0353 (0.0026)	0.0510 (0.0031)	0.0546 (0.0032)
	$\rho = 0.5$	0.0398 (0.0028)	0.0568 (0.0033)	0.0566 (0.0033)
	$\rho = 0.9$	0.0992 (0.0042)	0.1324 (0.0048)	0.1308 (0.0048)
n=100	$\rho = 0$	0.0352 (0.0026)	0.0412 (0.0028)	0.0576 (0.0033)
	$\rho = 0.5$	0.0395 (0.0028)	0.0440 (0.0029)	0.0562 (0.0033)
	$\rho = 0.9$	0.0750 (0.0037)	0.0924 (0.0041)	0.1000 (0.0042)
n=500	$\rho = 0$	0.0384 (0.0027)	0.0492 (0.0031)	0.0556 (0.0032)
	$\rho = 0.5$	0.0356 (0.0026)	0.0496 (0.0031)	0.0500 (0.0031)
	$\rho = 0.9$	0.0486 (0.0030)	0.0684 (0.0036)	0.0652 (0.0035)
n=1000	$\rho = 0$	0.0446 (0.0029)	0.0492 (0.0031)	0.0484 (0.0030)
	$\rho = 0.5$	0.0442 (0.0029)	0.0530 (0.0032)	0.0488 (0.0030)
	$\rho = 0.9$	0.0460 (0.0030)	0.0538 (0.0032)	0.0560 (0.0033)

*Test statistic is compared with χ_2^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Table 5.30 shows that sample size of 100 or greater makes estimated type I error rates not differ significantly from the nominal error rate 0.01 when there is a moderate or no correlation using either estimated standard error or conservative standard error. However, in the situation of high correlation, sample size of 1000 is needed using estimated standard error and 500 is needed using conservative standard error.

Table 5.32 Empirical type I error rates for the structural model likelihood ratio test of the null hypothesis $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}) = (\mathbf{0}, \mathbf{1})$ using a nominal type I error rate of $\boldsymbol{\alpha} = \mathbf{0.1}$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.1$ (conservative S.E.=0.0071, nominal S.E.=0.0042)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$\rho = 0$	0.0695 (0.0036)	0.0895 (0.0040)	0.1030 (0.0043)
	$\rho = 0.5$	0.1349 (0.0048)	0.1594 (0.0052)	0.1704 (0.0053)
	$\rho = 0.9$	0.2802 (0.0064)	0.3371 (0.0067)	0.3560 (0.0068)
n=20	$\rho = 0$	0.0535 (0.0032)	0.0892 (0.0040)	0.0974 (0.0042)
	$\rho = 0.5$	0.0919 (0.0041)	0.1164 (0.0045)	0.1202 (0.0046)
	$\rho = 0.9$	0.2363 (0.0060)	0.3027 (0.0065)	0.2998 (0.0065)
n=50	$\rho = 0$	0.0670 (0.0035)	0.0978 (0.0042)	0.1056 (0.0043)
	$\rho = 0.5$	0.0809 (0.0039)	0.1006 (0.0043)	0.1032 (0.0043)
	$\rho = 0.9$	0.1652 (0.0053)	0.2056 (0.0057)	0.2136 (0.0058)
n=100	$\rho = 0$	0.0711 (0.0036)	0.0882 (0.0040	0.1060 (0.0044)
	$\rho = 0.5$	0.0811 (0.0039)	0.0824 (0.0039)	0.1080 (0.0044)
	$\rho = 0.9$	0.1286 (0.0047)	0.1510 (0.0051)	0.1690 (0.0053)
n=500	$\rho = 0$	0.0816 (0.0039)	0.0982 (0.0042)	0.1056 (0.0043)
	$\rho = 0.5$	0.0740 (0.0037)	0.1026 (0.0043)	0.1002 (0.0042)
	$\rho = 0.9$	0.1024 (0.0043)	0.1294 (0.0047)	0.1190 (0.0046)
n=1000	$\rho = 0$	0.0922 (0.0041)	0.1020 (0.0043)	0.0984 (0.0042)
	$\rho = 0.5$	0.0936 (0.0042)	0.1022 (0.0043)	0.1000 (0.0042)
	$\rho = 0.9$	0.0976 (0.0042)	0.1076 (0.0044)	0.1080 (0.0044)

*Test statistic is compared with χ_2^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Tables 5.31 and 5.32 show that sample size of 1000 or greater is needed to make estimated error rate not differ significantly from the nominal error rate 0.05 using either estimated standard error or conservative standard error.

Generally speaking, the signal to noise ratio and correlation has impact on type I error besides the sample size. Large signal to noise ratio makes type I error close to nominal even when sample size is small. Strong correlation results in the estimated type I error to deviate from the nominal value. A sample size of 100 is large enough to produce estimated error rates close to nominal for most of the cases when testing $\rho = 0$, however, a large sample size of 1000 or greater is required when testing $(\beta_0, \beta_1) = (b_0, b_1)$.

5.3.2 Empirical type I error rates for the functional model test statistics

Tables 5.33-5.35 show estimated type I error rates and their three types of estimated standard errors for functional model score tests of the null hypothesis $H_0: \rho = 0$ using various values of parameters.

Table 5.33 shows that when using conservative standard error, sample size of 10 is enough for no significant difference between estimated type I error rates and nominal error rate of 0.01. However, sample size of 500 is needed when using estimated standard error. Table 5.34 shows that sample size of 20 is enough for the estimated error rates not different from the nominal 0.05 when using conservative standard error and sample size of 100 is needed when using estimated standard error.

Table 5.33 Empirical type I error rates for the functional model score test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.01$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.01$ (conservative S.E.=0.0071, nominal S.E.=0.0014)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.0122 (0.0016)	0.0110 (0.0015)	0.0094 (0.0014)
	$(\beta_0, \beta_1) = (1,1)$	0.0132 (0.0016)	0.0128 (0.0016)	0.0082 (0.0013)
	$(\beta_0, \beta_1) = (1,3)$	0.0088 (0.0013)	0.0114 (0.0015)	0.0130 (0.0016)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.0084 (0.0013)	0.0052 (0.0010)	0.0096 (0.0014)
	$(\beta_0, \beta_1) = (1,1)$	0.0074 (0.0012)	0.0064 (0.0011)	0.0104 (0.0014)
	$(\beta_0, \beta_1) = (1,3)$	0.0078 (0.0012)	0.0058 (0.0011)	0.0076 (0.0012)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.0076 (0.0012)	0.0108 (0.0015)	0.0108 (0.0015)
	$(\beta_0, \beta_1) = (1,1)$	0.0084 (0.0013)	0.0076 (0.0012)	0.0088 (0.0013)
	$(\beta_0, \beta_1) = (1,3)$	0.0072 (0.0012)	0.0086 (0.0013)	0.0084 (0.0013)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.0074 (0.0012)	0.0074 (0.0012)	0.0084 (0.0013)
	$(\beta_0, \beta_1) = (1,1)$	0.0078 (0.0012)	0.0108 (0.0015)	0.0084 (0.0013)
	$(\beta_0, \beta_1) = (1,3)$	0.0136 (0.0016)	0.0070 (0.0012)	0.0112 (0.0015)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.0100 (0.0014)	0.0110 (0.0015)	0.0124 (0.0016)
	$(\beta_0, \beta_1) = (1,1)$	0.0076 (0.0012)	0.0080 (0.0013)	0.0106 (0.0014)
	$(\beta_0, \beta_1) = (1,3)$	0.0104 (0.0014)	0.0092 (0.0014)	0.0112 (0.0015)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.0102 (0.0014)	0.0090 (0.0013)	0.0078 (0.0012)
	$(\beta_0, \beta_1) = (1,1)$	0.0092 (0.0014)	0.0102 (0.0014)	0.0088 (0.0013)
	$(\beta_0, \beta_1) = (1,3)$	0.0132 (0.0016)	0.0088 (0.0013)	0.0104 (0.0014)

*One half of test statistic is compared with χ_1^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Table 5.34 Empirical type I error rates for the functional model score test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.05$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.05$ (conservative S.E.=0.0071, nominal S.E.=0.0031)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.0534 (0.0032)	0.0524 (0.0032)	0.0508 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0588 (0.0033)	0.0550 (0.0032)	0.0456 (0.0030)
	$(\beta_0, \beta_1) = (1,3)$	0.0512 (0.0031)	0.0524 (0.0032)	0.0646 (0.0035)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.0464 (0.0030)	0.0440 (0.0029)	0.0518 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0424 (0.0028)	0.0430 (0.0029)	0.0526 (0.0032)
	$(\beta_0, \beta_1) = (1,3)$	0.0466 (0.0030)	0.0446 (0.0029)	0.0474 (0.0030)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.0512 (0.0031)	0.0496 (0.0031)	0.0442 (0.0029)
	$(\beta_0, \beta_1) = (1,1)$	0.0484 (0.0030)	0.0430 (0.0029)	0.0406 (0.0028)
	$(\beta_0, \beta_1) = (1,3)$	0.0446 (0.0029)	0.0466 (0.0030)	0.0420 (0.0028)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.0444 (0.0029)	0.0442 (0.0029)	0.0484 (0.0030)
	$(\beta_0, \beta_1) = (1,1)$	0.0482 (0.0030)	0.0492 (0.0031)	0.0494 (0.0031)
	$(\beta_0, \beta_1) = (1,3)$	0.0548 (0.0032)	0.0480 (0.0030)	0.0494 (0.0031)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.0478 (0.0030)	0.0536 (0.0032)	0.0518 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0443 (0.0029)	0.0474 (0.0030)	0.0486 (0.0030)
	$(\beta_0, \beta_1) = (1,3)$	0.0522 (0.0031)	0.0524 (0.0032)	0.0514 (0.0031)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.0482 (0.0030)	0.0564 (0.0033)	0.0524 (0.0032)
	$(\beta_0, \beta_1) = (1,1)$	0.0482 (0.0030)	0.0556 (0.0032)	0.0452 (0.0029)
	$(\beta_0, \beta_1) = (1,3)$	0.0526 (0.0032)	0.0474 (0.0030)	0.0550 (0.0032)

*One half of test statistic is compared with χ_1^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Table 5.35 shows that sample size of 20 is enough for the estimated error rates close to the nominal error rate 0.10 using conservative error rate and sample size of 100 is required using estimated error rate.

Table 5.35 Empirical type I error rates for the functional model score test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.1$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.1$ (conservative S.E.=0.0071, nominal S.E.=0.0042)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.1132 (0.0045)	0.1092 (0.0044)	0.1116 (0.0045)
	$(\beta_0, \beta_1) = (1,1)$	0.1112 (0.0044)	0.1082 (0.0044)	0.0934 (0.0041)
	$(\beta_0, \beta_1) = (1,3)$	0.1046 (0.0043)	0.1088 (0.0044)	0.1294 (0.0047)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.0984 (0.0042)	0.0922 (0.0041)	0.1086 (0.0044)
	$(\beta_0, \beta_1) = (1,1)$	0.0940 (0.0041)	0.0966 (0.0042)	0.1096 (0.0044)
	$(\beta_0, \beta_1) = (1,3)$	0.1022 (0.0043)	0.0948 (0.0041)	0.1022 (0.0043)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.0910 (0.0041)	0.0984 (0.0042)	0.0986 (0.0042)
	$(\beta_0, \beta_1) = (1,1)$	0.0906 (0.0041)	0.0982 (0.0042)	0.0946 (0.0041)
	$(\beta_0, \beta_1) = (1,3)$	0.0958 (0.0042)	0.0970 (0.0042)	0.0982 (0.0042)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.0924 (0.0041)	0.0924 (0.0041)	0.0984 (0.0042)
	$(\beta_0, \beta_1) = (1,1)$	0.0914 (0.0041)	0.1036 (0.0043)	0.0984 (0.0042)
	$(\beta_0, \beta_1) = (1,3)$	0.1026 (0.0043)	0.0990 (0.0042)	0.1020 (0.0043)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.0992 (0.0042)	0.1002 (0.0042)	0.1012 (0.0043)
	$(\beta_0, \beta_1) = (1,1)$	0.0910 (0.0041)	0.0960 (0.0042)	0.1000 (0.0042)
	$(\beta_0, \beta_1) = (1,3)$	0.1022 (0.0043)	0.1014 (0.0043)	0.1026 (0.0043)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.1026 (0.0043)	0.1062 (0.0044)	0.1016 (0.0043)
	$(\beta_0, \beta_1) = (1,1)$	0.1002 (0.0042)	0.1082 (0.0044)	0.0972 (0.0042)
	$(\beta_0, \beta_1) = (1,3)$	0.1072 (0.0044)	0.1002 (0.0042)	0.1038 (0.0043)

*One half of test statistic is compared with χ_1^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Tables 5.36-5.38 show estimated type I error rates and their three types of estimated standard errors for functional model likelihood ratio tests of the null hypothesis $H_0: \rho = 0$ using various values of parameters.

Table 5.36 Empirical type I error rates for the functional model likelihood ratio test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.01$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.01$ (conservative S.E.=0.0071, nominal S.E.=0.0014)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.0281 (0.0023)	0.0264 (0.0023)	0.0214 (0.0020)
	$(\beta_0, \beta_1) = (1,1)$	0.0246 (0.0022)	0.0372 (0.0027)	0.0230 (0.0021)
	$(\beta_0, \beta_1) = (1,3)$	0.0208 (0.0020)	0.0260 (0.0023)	0.0262 (0.0023)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.0182 (0.0019)	0.0156 (0.0018)	0.0160 (0.0018)
	$(\beta_0, \beta_1) = (1,1)$	0.0180 (0.0019)	0.0144 (0.0017)	0.0158 (0.0018)
	$(\beta_0, \beta_1) = (1,3)$	0.0135 (0.0016)	0.0150 (0.0017)	0.0166 (0.0018)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.0138 (0.0016)	0.0124 (0.0016)	0.0121 (0.0015)
	$(\beta_0, \beta_1) = (1,1)$	0.0136 (0.0016)	0.0132 (0.0016)	0.0119 (0.0015)
	$(\beta_0, \beta_1) = (1,3)$	0.0134 (0.0016)	0.0122 (0.0016)	0.0108 (0.0015)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.0132 (0.0016)	0.0134 (0.0016)	0.0124 (0.0016)
	$(\beta_0, \beta_1) = (1,1)$	0.0106 (0.0014)	0.0120 (0.0015)	0.0132 (0.0016)
	$(\beta_0, \beta_1) = (1,3)$	0.0108 (0.0015)	0.0098 (0.0014)	0.0110 (0.0015)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.0118 (0.0015)	0.0110 (0.0015)	0.0098 (0.0014)
	$(\beta_0, \beta_1) = (1,1)$	0.0078 (0.0012)	0.0092 (0.0014)	0.0128 (0.0016)
	$(\beta_0, \beta_1) = (1,3)$	0.0110 (0.0015)	0.0106 (0.0014)	0.0110 (0.0015)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.0102 (0.0014)	0.0114 (0.0015)	0.0096 (0.0014)
	$(\beta_0, \beta_1) = (1,1)$	0.0102 (0.0014)	0.0102 (0.0014)	0.0094 (0.0014)
	$(\beta_0, \beta_1) = (1,3)$	0.0118 (0.0015)	0.0102 (0.0014)	0.0098 (0.0014)

*One half of test statistic is compared with χ_1^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Table 5.36 shows that sample size of 20 is enough for estimated type I error rates close to the nominal error rate 0.01 when using conservative standard error and sample size of 100 is needed using estimated standard error. Table 5.37 shows that sample size of 50 is enough for estimated type I error rates close to the nominal error rate 0.05 using conservative standard error and sample size of 500 is needed using estimated standard error. Table 5.38 shows that sample size greater than 100 is needed using either estimated standard error or conservative standard error.

Table 5.37 Empirical type I error rates for the functional model likelihood ratio test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.05$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.05$ (conservative S.E.=0.0071, nominal S.E.=0.0031)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.1010 (0.0043)	0.0884 (0.0040)	0.0882 (0.0040)
	$(\beta_0, \beta_1) = (1,1)$	0.0991 (0.0042)	0.1178 (0.0046)	0.0862 (0.0040)
	$(\beta_0, \beta_1) = (1,3)$	0.0865 (0.0040)	0.0900 (0.0040)	0.1002 (0.0042)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.0802 (0.0038)	0.0644 (0.0035)	0.0664 (0.0035)
	$(\beta_0, \beta_1) = (1,1)$	0.0722 (0.0037)	0.0666 (0.0035)	0.0730 (0.0037)
	$(\beta_0, \beta_1) = (1,3)$	0.0658 (0.0035)	0.0618 (0.0034)	0.0654 (0.0035)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.0586 (0.0033)	0.0590 (0.0033)	0.0604 (0.0034)
	$(\beta_0, \beta_1) = (1,1)$	0.0632 (0.0034)	0.0606 (0.0034)	0.0632 (0.0034)
	$(\beta_0, \beta_1) = (1,3)$	0.0600 (0.0034)	0.0566 (0.0033)	0.0578 (0.0033)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.0588 (0.0033)	0.0616 (0.0034)	0.0502 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0610 (0.0034)	0.0494 (0.0031)	0.0546 (0.0032)
	$(\beta_0, \beta_1) = (1,3)$	0.0498 (0.0031)	0.0532 (0.0032)	0.0532 (0.0032)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.0562 (0.0033)	0.0486 (0.0030)	0.0510 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0508 (0.0031)	0.0502 (0.0031)	0.0564 (0.0033)
	$(\beta_0, \beta_1) = (1,3)$	0.0486 (0.0030)	0.0512 (0.0031)	0.0496 (0.0031)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.0476 (0.0030)	0.0512 (0.0031)	0.0514 (0.0031)
	$(\beta_0, \beta_1) = (1,1)$	0.0508 (0.0031)	0.0530 (0.0032)	0.0490 (0.0031)
	$(\beta_0, \beta_1) = (1,3)$	0.0538 (0.0032)	0.0524 (0.0032)	0.0482 (0.0030)

*One half of test statistic is compared with χ_1^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Table 5.38 Empirical type I error rates for the functional model likelihood ratio test of the null hypothesis $\rho = 0$ using a nominal type I error rate of $\alpha = 0.10$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.10$ (conservative S.E.=0.0071, nominal S.E.=0.0042)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$(\beta_0, \beta_1) = (0,1)$	0.1682 (0.0053)	0.1564 (0.0051)	0.1486 (0.0050)
	$(\beta_0, \beta_1) = (1,1)$	0.1697 (0.0053)	0.1917 (0.0056)	0.1590 (0.0052)
	$(\beta_0, \beta_1) = (1,3)$	0.1535 (0.0051)	0.1596 (0.0052)	0.1700 (0.0053)
n=20	$(\beta_0, \beta_1) = (0,1)$	0.1419 (0.0049)	0.1246 (0.0047)	0.1252 (0.0047)
	$(\beta_0, \beta_1) = (1,1)$	0.1312 (0.0048)	0.1274 (0.0047)	0.1300 (0.0048)
	$(\beta_0, \beta_1) = (1,3)$	0.1256 (0.0047)	0.1178 (0.0046)	0.1192 (0.0046)
n=50	$(\beta_0, \beta_1) = (0,1)$	0.1204 (0.0046)	0.1126 (0.0045)	0.1148 (0.0045)
	$(\beta_0, \beta_1) = (1,1)$	0.1139 (0.0045)	0.1158 (0.0045)	0.1160 (0.0045)
	$(\beta_0, \beta_1) = (1,3)$	0.1162 (0.0045)	0.1126 (0.0045)	0.1100 (0.0044)
n=100	$(\beta_0, \beta_1) = (0,1)$	0.1110 (0.0044)	0.1104 (0.0044)	0.0968 (0.0042)
	$(\beta_0, \beta_1) = (1,1)$	0.1188 (0.0046)	0.1024 (0.0043)	0.1046 (0.0043)
	$(\beta_0, \beta_1) = (1,3)$	0.1032 (0.0043)	0.1022 (0.0043)	0.1026 (0.0043)
n=500	$(\beta_0, \beta_1) = (0,1)$	0.1080 (0.0044)	0.0998 (0.0042)	0.1002 (0.0042)
	$(\beta_0, \beta_1) = (1,1)$	0.1018 (0.0043)	0.1030 (0.0043)	0.1034 (0.0043)
	$(\beta_0, \beta_1) = (1,3)$	0.1010 (0.0043)	0.0992 (0.0042)	0.1014 (0.0043)
n=1000	$(\beta_0, \beta_1) = (0,1)$	0.1018 (0.0043)	0.0994 (0.0042)	0.1054 (0.0043)
	$(\beta_0, \beta_1) = (1,1)$	0.0998 (0.0042)	0.1022 (0.0043)	0.0968 (0.0042)
	$(\beta_0, \beta_1) = (1,3)$	0.0992 (0.0042)	0.1078 (0.0044)	0.0960 (0.0042)

*One half of test statistic is compared with χ_1^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Tables 5.39-5.41 show estimated type I error rates and their three types of estimated standard errors for functional model score tests of the null hypothesis $H_0: (\beta_0, \beta_1) = (0,1)$ using various values of parameters.

Table 5.39 Empirical type I error rates for the functional model score test of the null hypothesis $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}) = (\mathbf{0}, \mathbf{1})$ using a nominal type I error rate of $\boldsymbol{\alpha} = \mathbf{0.01}$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.01$ (conservative S.E.=0.0071, nominal S.E.=0.0014)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$\rho = 0$	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
	$\rho = 0.5$	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
	$\rho = 0.9$	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
n=20	$\rho = 0$	0.0016 (0.0006)	0.0018 (0.0006)	0.0028 (0.0007)
	$\rho = 0.5$	0.0016 (0.0006)	0.0014 (0.0005)	0.0014 (0.0005)
	$\rho = 0.9$	0.0020 (0.0006)	0.0012 (0.0005)	0.0010 (0.0004)
n=50	$\rho = 0$	0.0048 (0.0010)	0.0064 (0.0011)	0.0068 (0.0012)
	$\rho = 0.5$	0.0034 (0.0008)	0.0036 (0.0008)	0.0046 (0.0010)
	$\rho = 0.9$	0.0032 (0.0008)	0.0022 (0.0007)	0.0042 (0.0010)
n=100	$\rho = 0$	0.0062 (0.0011)	0.0108 (0.0015)	0.0084 (0.0013)
	$\rho = 0.5$	0.0068 (0.0012)	0.0056 (0.0011)	0.0066 (0.0011)
	$\rho = 0.9$	0.0032 (0.0008)	0.0048 (0.0010)	0.0060 (0.0011)
n=500	$\rho = 0$	0.0116 (0.0015)	0.0094 (0.0014)	0.0086 (0.0013)
	$\rho = 0.5$	0.0098 (0.0014)	0.0094 (0.0014)	0.0074 (0.0012)
	$\rho = 0.9$	0.0064 (0.0011)	0.0078 (0.0012)	0.0050 (0.0010)
n=1000	$\rho = 0$	0.0088 (0.0013)	0.0090 (0.0013)	0.0108 (0.0015)
	$\rho = 0.5$	0.0078 (0.0012)	0.0098 (0.0014)	0.0098 (0.0014)
	$\rho = 0.9$	0.0086 (0.0013)	0.0080 (0.0013)	0.0088 (0.0013)

*One half of test statistic is compared with χ_2^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Table 5.39 shows that sample size of 1000 is needed to make estimated type I error rates significantly different from the nominal error rate 0.01 when using estimated standard error and it seems that sample size of 10 is enough using conservative standard error. Tables 5.40 and 5.41 show that sample size of 1000 is required for estimated type I error rates to be close to the nominal error rate 0.05 when using estimated standard error and sample size of 500 is needed using conservative standard error. Table 5.41 shows that

Table 5.40 Empirical type I error rates for the functional model score test of the null hypothesis $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}) = (\mathbf{0}, \mathbf{1})$ using a nominal type I error rate of $\boldsymbol{\alpha = 0.05}$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.05$ (conservative S.E.=0.0071, nominal S.E.=0.0031)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$\rho = 0$	0.0104 (0.0014)	0.0124 (0.0016)	0.0144 (0.0017)
	$\rho = 0.5$	0.0120 (0.0015)	0.0128 (0.0016)	0.0126 (0.0016)
	$\rho = 0.9$	0.0126 (0.0016)	0.0104 (0.0014)	0.0134 (0.0016)
n=20	$\rho = 0$	0.0274 (0.0023)	0.0262 (0.0023)	0.0330 (0.0025)
	$\rho = 0.5$	0.0228 (0.0021)	0.0214 (0.0020)	0.0230 (0.0021)
	$\rho = 0.9$	0.0218 (0.0021)	0.0202 (0.0020)	0.0180 (0.0019)
n=50	$\rho = 0$	0.0402 (0.0028)	0.0408 (0.0028)	0.0396 (0.0028)
	$\rho = 0.5$	0.0372 (0.0027)	0.0328 (0.0025)	0.0340 (0.0026)
	$\rho = 0.9$	0.0210 (0.0020)	0.0230 (0.0021)	0.0268 (0.0023)
n=100	$\rho = 0$	0.0440 (0.0029)	0.0480 (0.0030)	0.0472 (0.0030)
	$\rho = 0.5$	0.0400 (0.0028)	0.0472 (0.0030)	0.0406 (0.0028)
	$\rho = 0.9$	0.0254 (0.0022)	0.0288 (0.0024)	0.0364 (0.0026)
n=500	$\rho = 0$	0.0542 (0.0032)	0.0518 (0.0031)	0.0472 (0.0030)
	$\rho = 0.5$	0.0496 (0.0031)	0.0504 (0.0031)	0.0480 (0.0030)
	$\rho = 0.9$	0.0400 (0.0028)	0.0432 (0.0029)	0.0420 (0.0028)
n=1000	$\rho = 0$	0.0510 (0.0031)	0.0536 (0.0032)	0.0496 (0.0031)
	$\rho = 0.5$	0.0506 (0.0031)	0.0524 (0.0032)	0.0540 (0.0032)
	$\rho = 0.9$	0.0464 (0.0030)	0.0470 (0.0030)	0.0446 (0.0029)

*One half of test statistic is compared with χ_2^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

Table 5.41 Empirical type I error rates for the functional model score test of the null hypothesis $(\beta_0, \beta_1) = (0, 1)$ using a nominal type I error rate of $\alpha = 0.10$

Sample Size	Parameter Value $(\sigma^2, \mu_x) = (4,1)$	Type I Error $\alpha = 0.10$ (conservative S.E.=0.0071, nominal S.E.=0.0042)		
		$\sigma_x^2 = 1$	$\sigma_x^2 = 4$	$\sigma_x^2 = 16$
n=10	$\rho = 0$	0.0608 (0.0034)	0.0578 (0.0033)	0.0510 (0.0031)
	$\rho = 0.5$	0.0542 (0.0032)	0.0516 (0.0031)	0.0506 (0.0031)
	$\rho = 0.9$	0.0556 (0.0032)	0.0502 (0.0031)	0.0472 (0.0030)
n=20	$\rho = 0$	0.0814 (0.0039)	0.0780 (0.0038)	0.0890 (0.0040)
	$\rho = 0.5$	0.0632 (0.0034)	0.0618 (0.0034)	0.0660 (0.0035)
	$\rho = 0.9$	0.0562 (0.0033)	0.0510 (0.0031)	0.0508 (0.0031)
n=50	$\rho = 0$	0.0904 (0.0041)	0.0934 (0.0041)	0.0914 (0.0041)
	$\rho = 0.5$	0.0820 (0.0039)	0.0800 (0.0038)	0.0802 (0.0038)
	$\rho = 0.9$	0.0580 (0.0033)	0.0550 (0.0032)	0.0644 (0.0035)
n=100	$\rho = 0$	0.0910 (0.0041)	0.1074 (0.0044)	0.0950 (0.0041)
	$\rho = 0.5$	0.0910 (0.0041)	0.1004 (0.0043)	0.0934 (0.0041)
	$\rho = 0.9$	0.0636 (0.0035)	0.0680 (0.0036)	0.0774 (0.0038)
n=500	$\rho = 0$	0.1036 (0.0043)	0.1048 (0.0043)	0.0978 (0.0042)
	$\rho = 0.5$	0.0976 (0.0042)	0.1030 (0.0043)	0.1016 (0.0043)
	$\rho = 0.9$	0.0954 (0.0042)	0.0910 (0.0041)	0.0886 (0.0040)
n=1000	$\rho = 0$	0.1008 (0.0043)	0.1014 (0.0043)	0.0974 (0.0042)
	$\rho = 0.5$	0.1004 (0.0043)	0.1008 (0.0043)	0.0990 (0.0042)
	$\rho = 0.9$	0.0964 (0.0042)	0.0922 (0.0041)	0.0922 (0.0041)

*One half of test statistic is compared with χ_2^2. Each cell has value of the form estimated type I error (estimated S.E.). Red indicates significant difference from the nominal rate using estimated S.E. and blue for using conservative S.E.

From all the above tables in this section, we observe that the estimated type I error rate for our test statistics is affected by sample size, correlation and signal to noise ratio. The bias of the estimated type I error rate from the nominal error rate decreases as sample size or signal to noise ratio increases and increases as correlation increases. In general, the sample size needed to make the estimated type I error rates close to the nominal is larger using estimated standard error than that using conservative standard error.

5.4 Power of the Test Statistics

We examine the empirical power of each test statistic for both structural and functional cases and gauge the effects of varying parameter values on of the empirical power. For testing the null hypothesis $\rho = 0$, relationships between empirical power and the value of ρ are studied for varying parameter values. However, when testing the null hypothesis $(\beta_0, \beta_1) = (b_0, b_1)$, relationships between power and values of β_0 and β_0 are investigated by varying parameters using the following scheme. Fayes (1996) showed analytically for the structural model and demonstrated via simulation studies for the functional model, that power of his likelihood-based test statistics is minimized when $\beta_0 = b_0 - \mu_x(\beta_1 - b_1)$. Simulation studies in this section are therefore performed under this "worst case" scenario, i.e., by setting the intercept parameters equal to $b_0 - \mu_x(\beta_1 - b_1)$.

Empirical power of score test statistic for testing $\rho = 0$ is shown in Figure 5.6. In each simulation, $\sigma^2 = 4$ and the null hypothesis is $\rho = 0$. Three pairs of (β_0, β_1) are used: $(0,1)$, $(1,1)$, and $(1,3)$. For each (β_0, β_1), values of σ_x^2 are 1, 4, and 16.

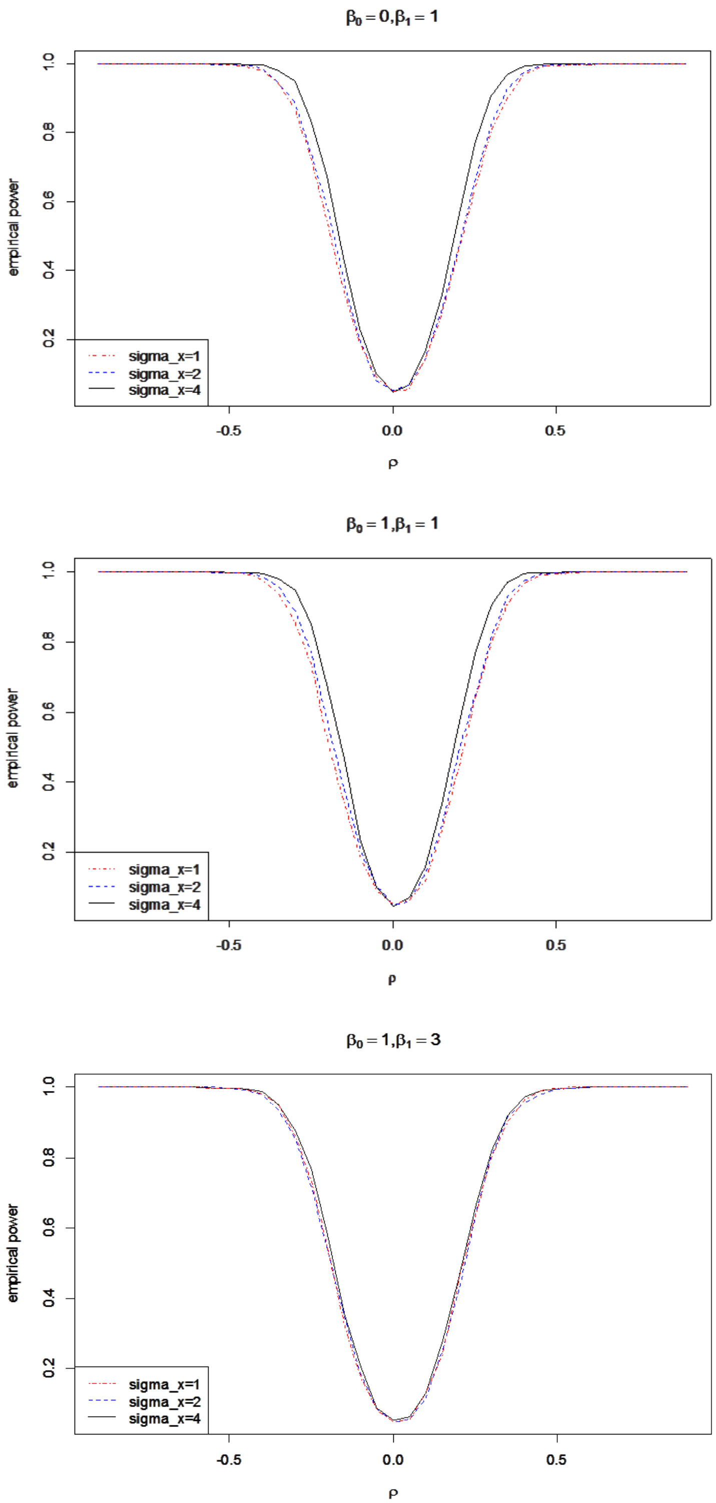

Figure 5.6 Empirical power for score test statistic $T_{s,ES,\rho=0}$

From Figure 5.6 graphs the empirical power curve for the score test statistic for testing we observe that change of β_0 does not seem to change the power of the test statistic. Larger value of σ_x^2 may have more power. As β_1 increases, the difference between different values of σ_x^2 becomes ignorable. Next, the power of LRT test statistic for testing $\rho = 0$ is similarly shown in Figure 5.7.

Comparing Figure 5.7 with Figure 5.6, we notice that they are similar with each other. The empirical power curve for LRT test statistic is very like that for score test statistic. Therefore, the properties are more or less the same. If β_1 is increased, there is almost no difference among power curves with different values of σ_x^2. From the first two plots, the empirical power curves are close to each other when signal to noise ratio is moderate to high, but very different from that with low signal to noise ratio.

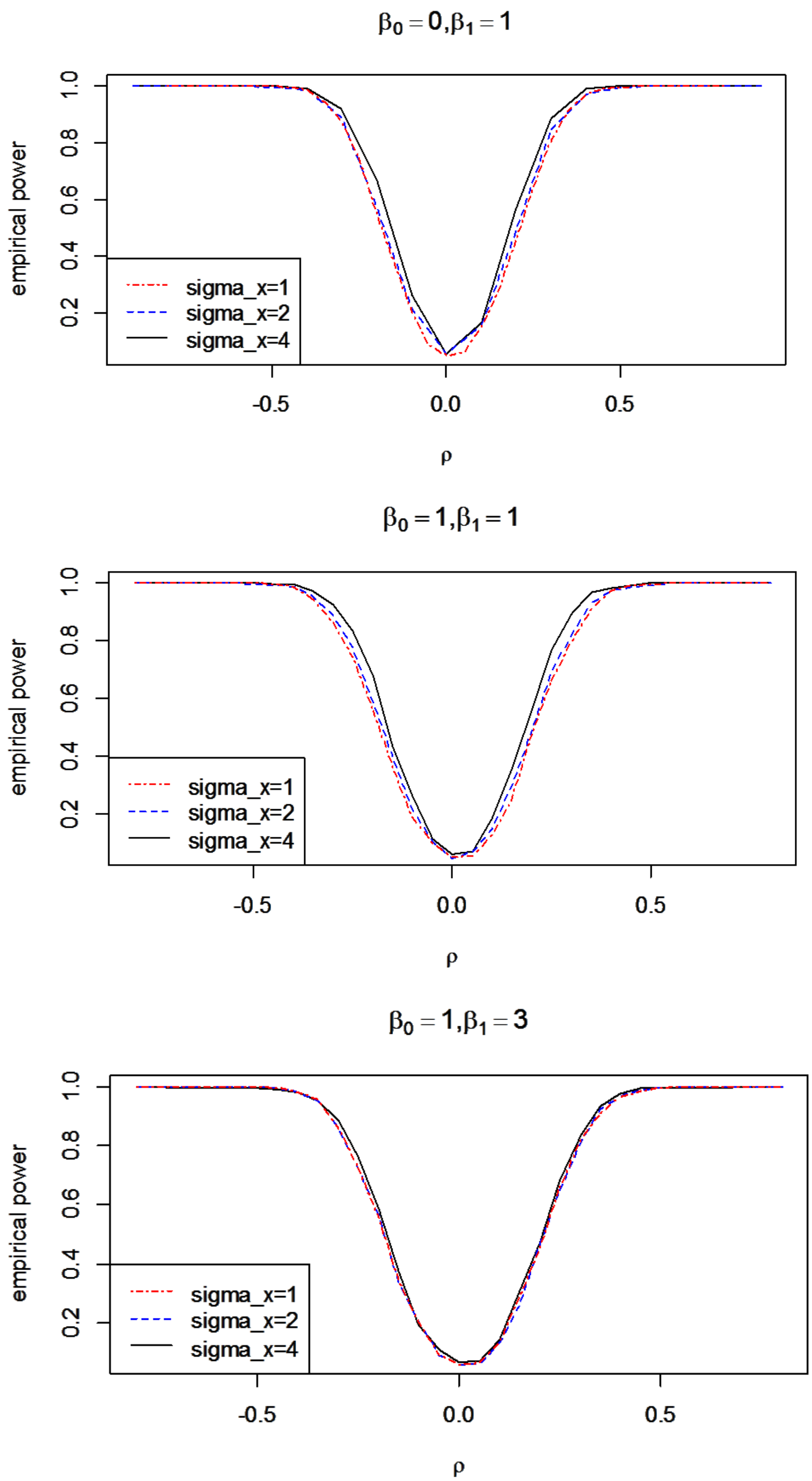

Figure 5.7 Empirical power for LRT test statistic $T_{s,LRT,\rho=0}$

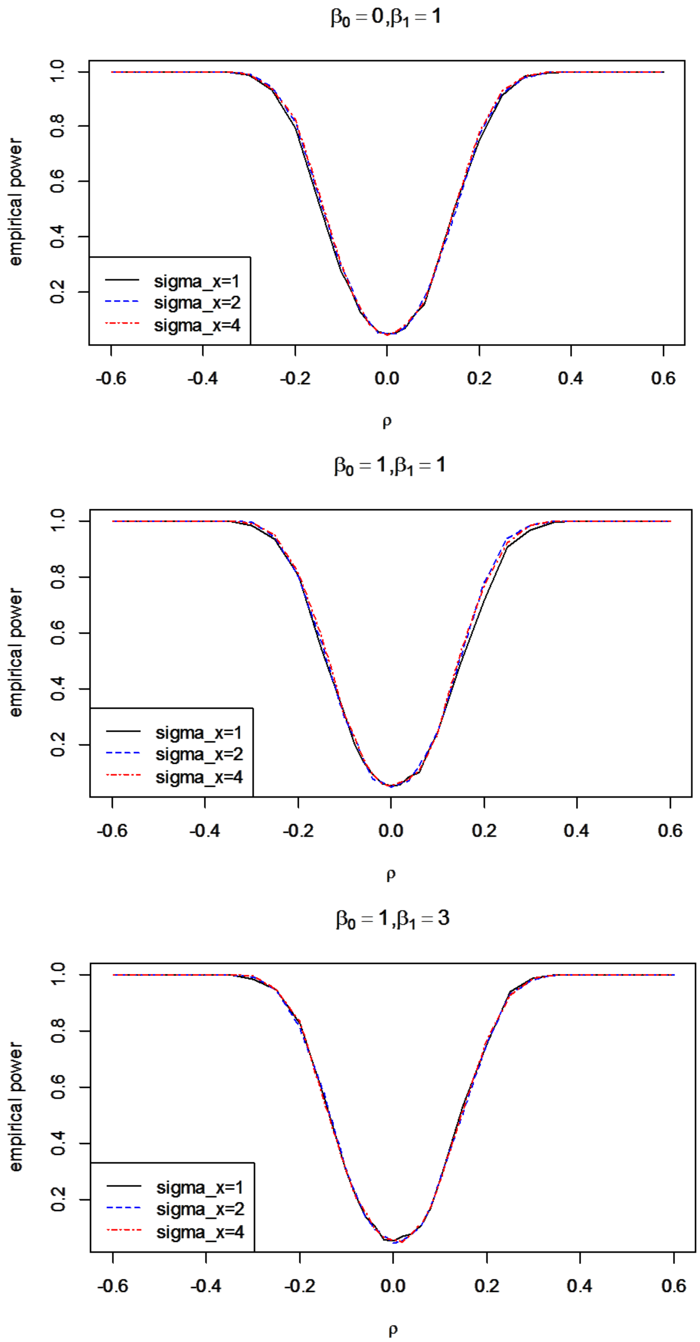

Figure 5.8 Empirical power for score test statistic $T_{f,ES,\rho=0}$, n=200

Figures 5.8 and 5.9 are empirical power curves for functional score test and LRT test statistics, respectively. As with the structural cases, these two sets of power curves look similar to each other. Both seem to not vary greatly with the changes in parameter values.

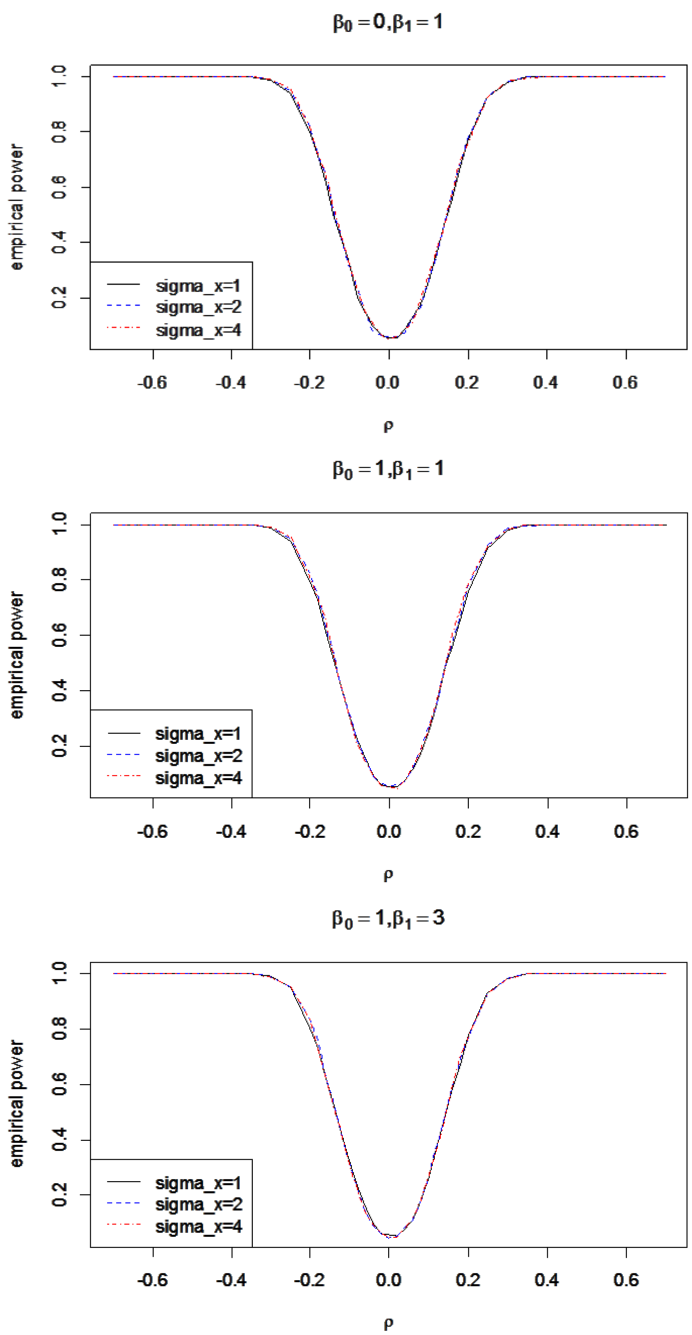

Figure 5.9 Empirical power for LRT test statistic $T_{f,LRT,\rho=0}$, n=200

Empirical power curves the score test statistic in structural model $T_{s,ES,\beta=b}$, are plotted in Figure 5.10. The null hypothesis tested is $(\beta_0, \beta_1) = (b_0, b_1) = (0,1)$, and $\sigma^2 = 4$ and $\mu_x = 1$ are held constant in these simulated power curves. Within each column of graphs, values of β_1 are 0.9, 1.0 and 1.1 from top to bottom, respectively. Recall that values of β_0 equal $b_0 - \mu_x(\beta_1 - b_1) = -(\beta_1 - 1)$ in these simulations present the "worst-case" scenario for power. Values of correlation coefficients are 0, 0.5 and 0.9 going left to right in each row of graphs. Empirical power of the LRT test statistic is poor when the correlation equals 0.9. The first column of graphs, wherein correlation equals zero, corroborates the finding of Abdul-Salam (1996) that power is minimized when β_0 equals $-(\beta_1 - 1)$, or 0.1, 0, and -0.1, respectively, as you descend from the top graph. When β_1 equals b_1, power seems not to depend on the variance ratio, except in the case of the strongest correlation. Otherwise, power is adversely affected by increases in the variance ratio.

Empirical power curves in Figure 5.11 use the same parameter combinations as those in Figure 5.9 for the functional case score test statistic $T_{f,ES,\beta=b}$.

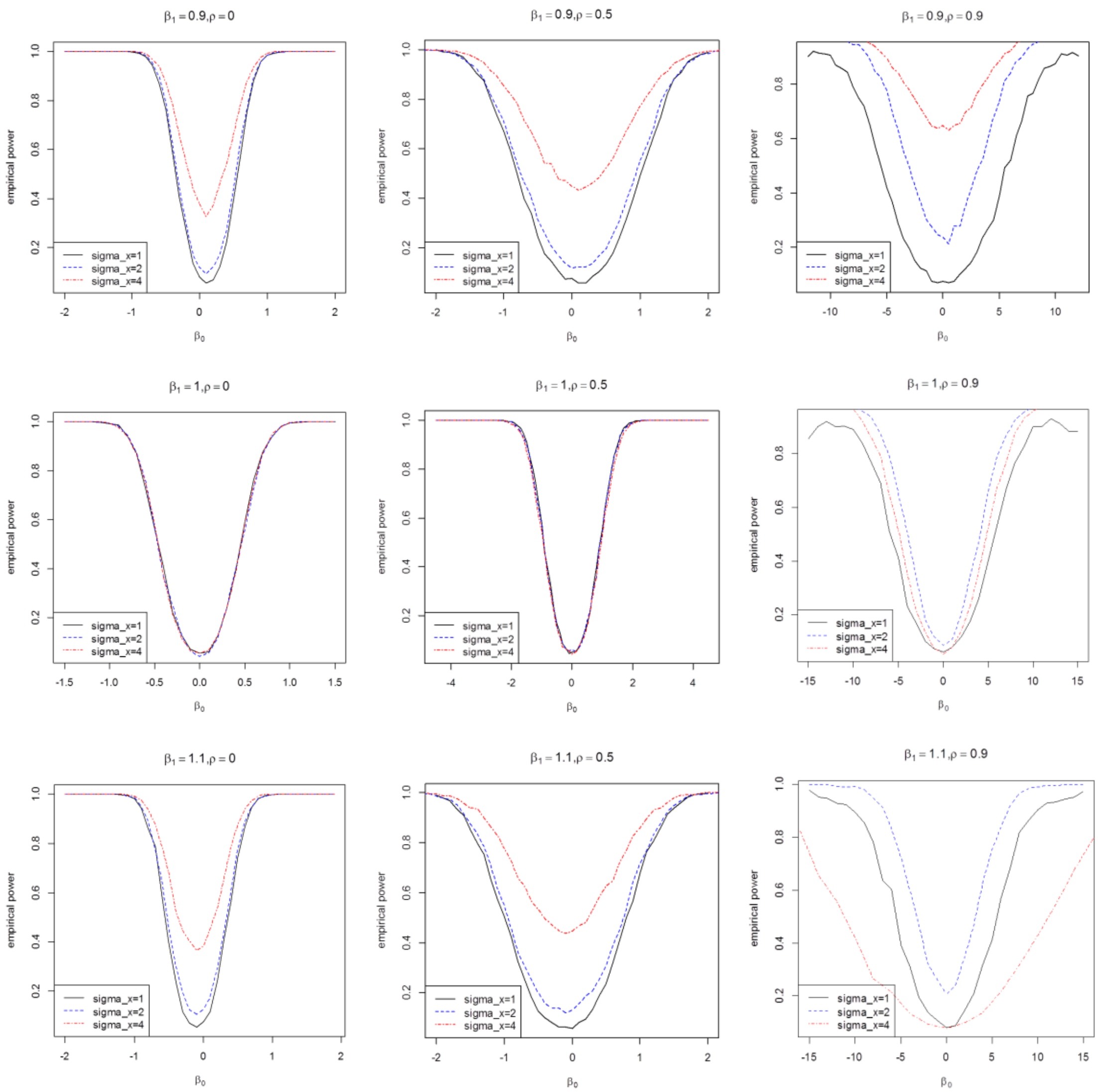

Figure 5.10 Empirical power for score test statistic $T_{s,ES,\beta=b}$, n=200

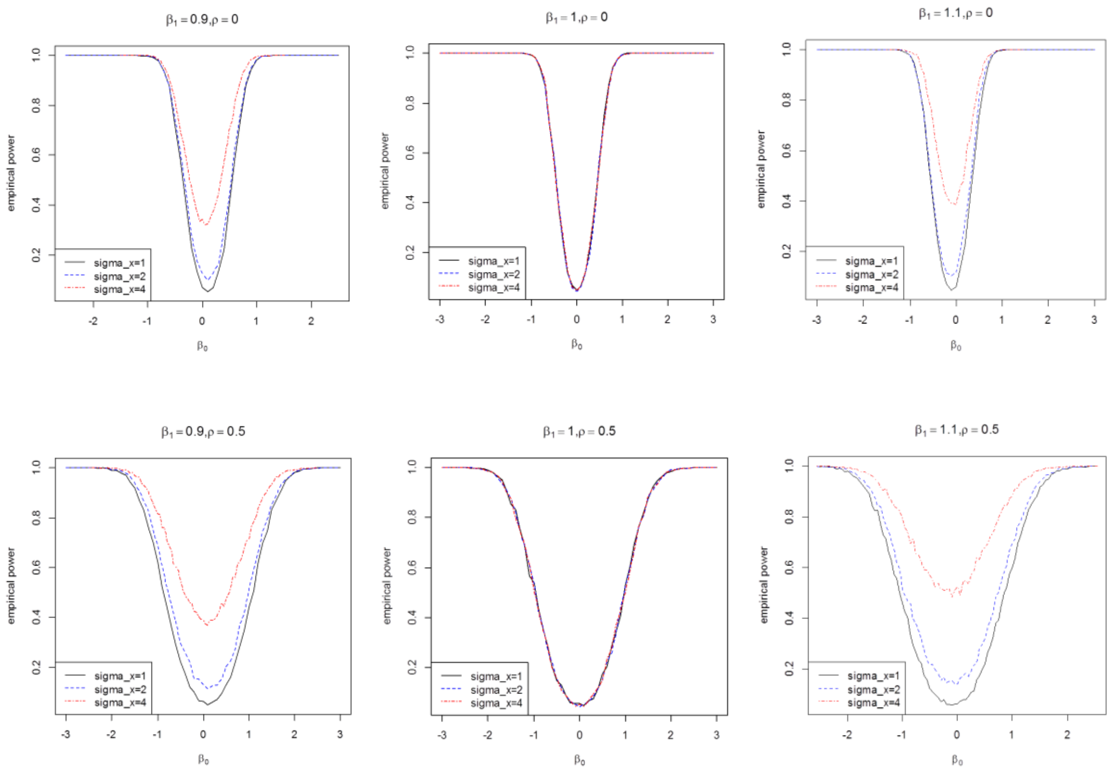

Figure 5.11 Empirical power for score test statistic $T_{f,ES,\beta=b}$, n=200

Figure 5.11 illustrates how empirical power decreases as correlation increases for the functional score test statistic. Again, when correlation equals zero, power is minimized when β_0 equals $-(\beta_1 - 1)$, or 0.1, 0 and -0.1. Also, when the true value of β_1 differs from b_1, the power of the test statistic depends the variance ratio σ_x^2/σ^2.

We investigate the relationship between power and β_1 also for functional score test statistic next. Typical results are shown in Figure 5.12.

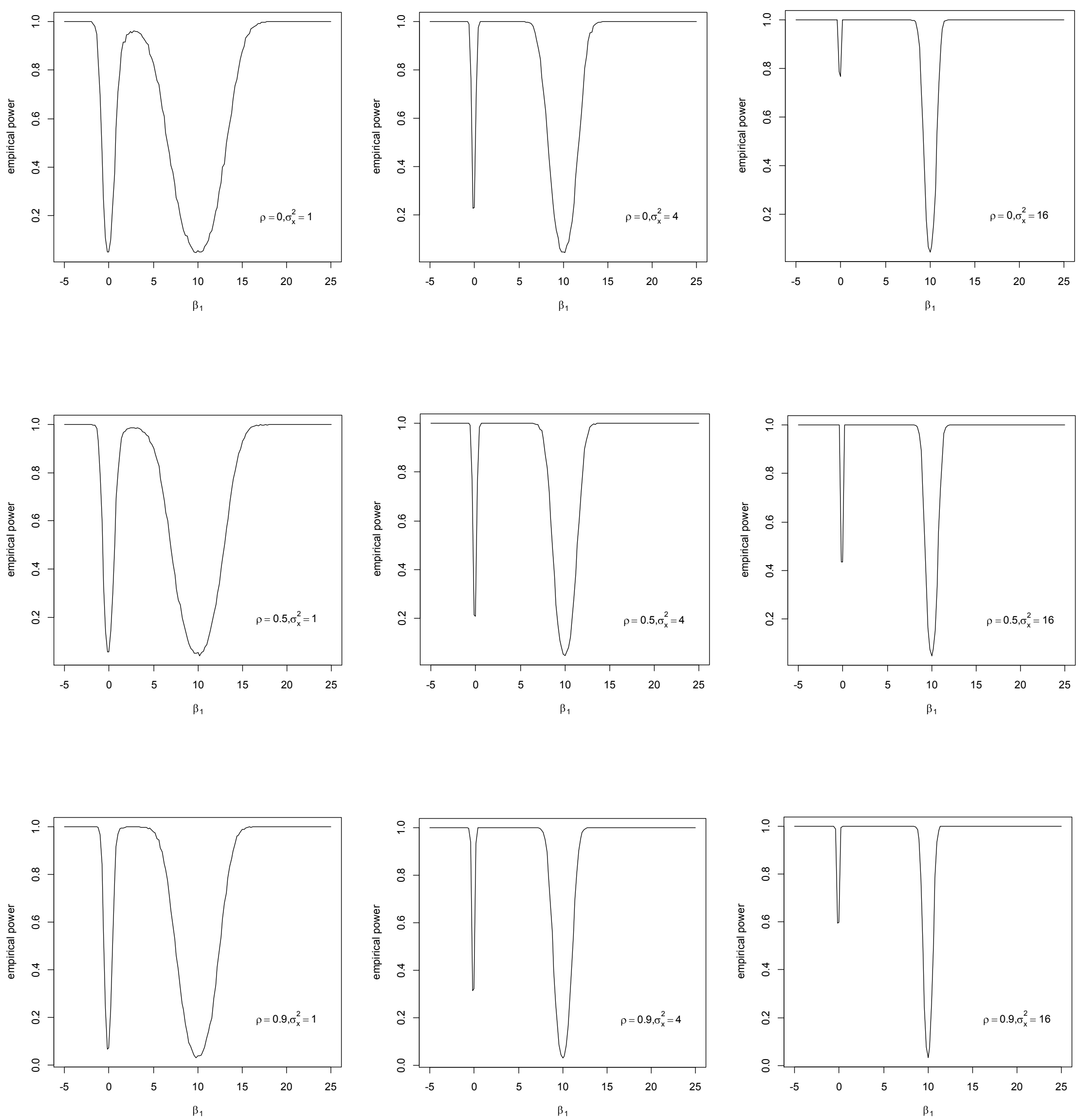

Figure 5.12 Empirical power for score test statistic $T_{f,ES,\beta=b}$, n=200

The null hypothesis is $(b_0, b_1) = (0,10)$. The value $\beta_1 = 10$ is chosen simply to provide a clearer picture of the anomalous loss of power in regions of the alternative hypothesis parameter space. The loss of power in these regions is exacerbated by decreasing the signal to noise ratio. Although not shown in Figure 5.12, as values of β_1 tend toward zero, the anomalous region becomes closer to the hypothesized null value of β_1. In particular, when the null hypothesis of interest is $(\beta_0, \beta_1) = (0,1)$, the loss of power could become problematic. Abdul-Salam (1996) showed that the point in the alternative parameter space for which power of the LRT test statistic is minimized when no correlation is present occurs when $\beta_0 = b_0 - \mu_x(\beta_1 - b_1)$.

In general, for all of the test statistics, increasing sample size or signal to noise ratio increases the empirical power. Stronger correlations among the data reduce empirical power and require larger sample sizes to achieve the results similar to those for more weakly correlated data.

6. CONCLUSION

6.1 Conclusion

In this work, we derived the likelihood based test statistics for several different scenarios. According to the properties of covariate x_t, we define two models. One model is called structural model with x_t a random variable, and the other model is called functional model with x_t fixed constants. For each of the two models, we set up two kinds of tests of hypotheses. The null hypotheses are $\rho = 0$ and $(\beta_0, \beta_1) = (b_0, b_1)$ respectively. And for each set of hypothesis, we calculate the maximum likelihood estimator and derive the likelihood ratio test statistic and score test statistic.

After careful derivation, we have eight different test statistics and six sets of parameter estimators. For each set of estimators, the consistency problem is discussed. Both theoretical and simulation results show that these estimators are consistent. The bias of these estimators generally decreases as sample size increases. Correlation among data and low ratio of σ_x^2 over σ^2 needs more samples to get the same level of bias. For the test statistics, the asymptotic distributions are derived and the corresponding simulation results are shown. The test statistics for testing null hypothesis $\rho = 0$ are approximately distributed as a chi-square distribution with one degree of freedom. Those for testing null hypothesis $(\beta_0, \beta_1) = (b_0, b_1)$ are distributed as a chi-square distribution with two degrees of freedom approximately. A-D GOF test and Q-Q plot is used to compare the distribution of test statistic with chi-square distribution. Small sample size is enough if correlation is weak and signal to noise ratio is large.

The power of these test statistics is also investigated. Simulation results show how the power curves change as one or several parameters change. For the four test statistics of testing $\rho = 0$, power seems not heavily depend on parameter (β_0, β_1) and the variance ratio especially for functional cases. The score test statistic and likelihood ratio test statistic behaves similar when both are from the same model and test the same null hypothesis. For the test statistics of testing the null hypothesis $(\beta_0, \beta_1) = (b_0, b_1)$, the minimum value of power occurs when $(\beta_0 - b_0) + \mu_x(\beta_1 - b_1) = 0$ is satisfied. And if the true value of β_1 is the same with b_1, then power is stable with change of the variance ratio. If β_1 is different from b_1, the variance ratio has a great impact on power. Also, power is usually decreased as the correlation among data increases. When plotting power curve against β_1, we find that there are two valleys in the power curve for test statistics of testing $(\beta_0, \beta_1) = (b_0, b_1)$. This may cause some problem if the value of β_1 is not very positive or very negative. However, for those with values much larger or smaller than zero, we still can get good power using the test statistics. Behaviors of the functional score test statistic and functional LRT test statistic are usually similar.

According to all the work done, increasing sample size can help get better estimators and increase power. Once we have better estimators, the asymptotic behavior of the test statistics would get better. So is the power of the test statistics. Notice that, the signal to noise ratio is an important variable that affects large sample properties and power of the test statistics. The correlation among data is also an important variable that matters.

6.2 Future Work

One of the extensions of this research would be to find a better adjustment for the LRT test statistic of testing $(\beta_0, \beta_1) = (b_0, b_1)$ in functional case. Another one is to derive the results for multivariate case. Further, only AR(1) process is considered in this work. Other popular correlation structures can also be learned and the corresponding test statistics may be derived.

REFERENCES

Abdul-Salam, F. A. Y (1996). Small sample inference for the structural and functional simple linear measurement error models. (Unpublished doctoral dissertation). Texas A&M University, College Station, Texas.

Aigner, D. J. (1972). A note on verification of computer simulation models. *Management Science*, 18(11), 615-619.

Athreya, K. B., & Lahiri, S. N. (2006). *Measure theory and probability theory*. New York: Springer.

Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models, A modern Perspective*, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC, Taylor & Francis Group.

Cohen, K. J., & Cyert, R. M. (1961). Computer models in dynamic economics. *The Quarterly Journal of Economics*, 112-127.

DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. New York: Springer.

Fuller, W. A. (1987). *Measurement error models*. New York: Wiley.

Harrison, S. R. (1990). Regression of a model on real-system output: an invalid test of model validity. *Agricultural Systems*, 34(3), 183-190.

Kendall, M. G. (1951). Regression, structure and functional relationship. Part I. *Biometrika*, 11-25.

Kendall, M. G. (1952). Regression, structure and functional relationship. Part II. *Biometrika*, 96-108.

Krantz, S. G., & Parks, H. R. (2002). *The implicit function theorem: History, theory, and applications*. Boston: Birkhäuser.

Lehmann, E. L. (1999). *Elements of large-sample theory*. New York: Springer.

Mayer, D. G., Stuart, M. A., & Swain, A. J. (1994). Regression of real-world data on model output: an appropriate overall test of validity. *Agricultural Systems*, 45(1), 93-104.

Mitchell, P. L. (1997). Misuse of regression for empirical validation of models. *Agricultural Systems*, 54(3), 313-326.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, 1-32.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. New York: Wiley.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, 110-114.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

APPENDIX A

DERIVATION OF STRUCTURAL LIKELIHOOD FUNCTION

From the assumption when introducing the structural model, we know that for any $t = 2, \dots, n$

$$Z \overset{\text{def}}{=} (x_{t-1}, u_{t-1}, e_{t-1})^T \sim N(\mu, \Sigma),$$

where

$$\mu = (\mu_x, 0,0)^T,$$

and

$$\Sigma = \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \frac{\sigma^2}{1-\rho^2} & 0 \\ 0 & 0 & \frac{\sigma^2}{1-\rho^2} \end{pmatrix}.$$

Then after variable transformation, we have

$$\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \beta_0 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 0 \\ \beta_1 & 0 & 1 \end{pmatrix} Z \sim N(\mu_1, \Sigma_1),$$

where

$$\mu_1 = \begin{pmatrix} \mu_x \\ \beta_0 + \beta_1 \mu_x \end{pmatrix},$$

and

$$\Sigma_1 = \begin{pmatrix} \sigma_x^2 + \frac{\sigma^2}{1-\rho^2} & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \beta_1^2 \sigma_x^2 + \frac{\sigma^2}{1-\rho^2} \end{pmatrix}.$$

Since the corresponding Jacobian coefficient is 1, then

$$f(x_{t-1}, X_{t-1}, Y_{t-1}) = 1 \times f(x_{t-1}, X_{t-1} - x_{t-1}, Y_{t-1} - y_{t-1}).$$

Let $Z_2 = (x_{t-1}, X_{t-1} - x_{t-1}, Y_{t-1} - y_{t-1})^T$, we know that

$$g(x_{t-1}|X_{t-1}, Y_{t-1}) = \frac{f(x_{t-1}, X_{t-1}, Y_{t-1})}{f(X_{t-1}, Y_{t-1})} = \frac{\frac{1}{\sqrt{(2\pi)^3|\Sigma|}} exp\left(-\frac{1}{2}(Z_2-\mu)^T\Sigma^{-1}(Z_2-\mu)\right)}{\frac{1}{\sqrt{(2\pi)^2|\Sigma_1|}} exp\left(-\frac{1}{2}(Z_1-\mu_1)^T\Sigma_1^{-1}(Z_1-\mu_1)\right)}$$

$$= \frac{exp\left(-\frac{1}{2}\{(Z_2-\mu)^T\Sigma^{-1}(Z_2-\mu)-(Z_1-\mu_1)^T\Sigma_1^{-1}(Z_1-\mu_1)\}\right)}{\sqrt{2\pi\sigma_x^2\left(\frac{\sigma^2}{1-\rho^2}\right)/\left(\left(1+\beta_1^2\right)\sigma_x^2+\frac{\sigma^2}{1-\rho^2}\right)}}.$$

Since

$$(Z_2-\mu)^T\Sigma^{-1}(Z_2-\mu) = \frac{(x_{t-1}-\mu_x)^2}{\sigma_x^2} + (1-\rho^2)\frac{(X_{t-1}-x_{t-1})^2}{\sigma^2} + (1-\rho^2)\frac{(Y_{t-1}-\beta_0-\beta_1 x_{t-1})^2}{\sigma^2}$$

$$= \frac{A(x_{t-1}-\mu_x)^2+\sigma_x^2(X_{t-1}-x_{t-1})^2+\sigma_x^2(Y_{t-1}-\beta_0-\beta_1 x_{t-1})^2}{\sigma_x^2 A}$$

$$= \frac{Dx_{t-1}^2-2\{A\mu_x+\sigma_x^2[X_{t-1}+\beta_1(Y_{t-1}-\beta_0)]\}x_{t-1}+\{A\mu_x^2+\sigma_x^2[X_{t-1}^2+(Y_{t-1}-\beta_0)^2]\}}{\sigma_x^2 A}$$

$$= \frac{D}{\sigma_x^2 A}\left\{x_{t-1} - \frac{\left(A\mu_x+\sigma_x^2\left(X_{t-1}+\beta_1(Y_{t-1}-\beta_0)\right)\right)}{D}\right\}^2$$

$$-\frac{\left\{A\mu_x+\sigma_x^2\left(X_{t-1}+\beta_1(Y_{t-1}-\beta_0)\right)\right\}^2}{\sigma_x^2 AD} + \frac{A\mu_x^2+\sigma_x^2\left(X_{t-1}^2+(Y_{t-1}-\beta_0)^2\right)}{\sigma_x^2 A}$$

and

$$(Z_1-\mu_1)^T\Sigma_1^{-1}(Z_1-\mu_1) = \frac{(\beta_1^2\sigma_x^2+A)}{AD}(X_{t-1}-\mu_x)^2 + \frac{(\sigma_x^2+A)}{AD}[Y_{t-1}-(\beta_0+\beta_1\mu_x)]^2$$

$$-\frac{2\beta_1\sigma_x^2}{AD}(X_{t-1}-\mu_x)[Y_{t-1}-(\beta_0+\beta_1\mu_x)]$$

$$= \frac{1}{AD}\sigma_x^2\left(\beta_1 X_{t-1}-(Y_{t-1}-\beta_0)\right)^2 + \frac{1}{D}(X_{t-1}-\mu_x)^2$$

$$+\frac{1}{D}\left(Y_{t-1}-(\beta_0+\beta_1\mu_x)\right)^2,$$

then

$(Z_2-\mu)^T\Sigma^{-1}(Z_2-\mu) - (Z_1-\mu_1)^T\Sigma_1^{-1}(Z_1-\mu_1)$

$$= \frac{D\left[x_{t-1} - \frac{A\mu_x + \sigma_x^2\left(X_{t-1} + \beta_1(Y_{t-1} - \beta_0)\right)}{D}\right]^2}{\sigma_x^2 A} - \frac{\left[A\mu_x + \sigma_x^2\left(X_{t-1} + \beta_1(Y_{t-1} - \beta_0)\right)\right]^2}{\sigma_x^2 AD} + \frac{A\mu_x^2 + \sigma_x^2\left(X_{t-1}^2 + (Y_{t-1} - \beta_0)^2\right)}{\sigma_x^2 A}$$

$$-\frac{1}{A}\left\{\frac{\sigma_x^2\left(\beta_1 X_{t-1} - (Y_{t-1} - \beta_0)\right)^2}{D} + A\frac{\left((X_{t-1} - \mu_x)^2 + \left(Y_{t-1} - (\beta_0 + \beta_1\mu_x)\right)^2\right)}{D}\right\}$$

$$= \frac{D}{\sigma_x^2 A}\left\{x_{t-1} - \frac{A\mu_x + \sigma_x^2\left(X_{t-1} + \beta_1(Y_{t-1} - \beta_0)\right)}{D}\right\}^2,$$

where $A = \frac{\sigma^2}{1-\rho^2}$, $D = (1 + \beta_1^2)\sigma_x^2 + A$.

By substituting the above equation into $g(x_{t-1}|X_{t-1}, Y_{t-1})$, we have

$$g(x_{t-1}|X_{t-1}, Y_{t-1}) = \frac{exp\left\{-\frac{1}{2\sigma_x^2 A/D}\left\{x_{t-1} - \frac{A\mu_x + \sigma_x^2\left(X_{t-1} + \beta_1(Y_{t-1} - \beta_0)\right)}{D}\right\}^2\right\}}{\sqrt{2\pi\sigma_x^2 A/D}}$$

which follows a normal density function.

Therefore, the conditional distribution of x_{t-1} given (X_{t-1}, Y_{t-1}) is

$$x_{t-1}|(X_{t-1}, Y_{t-1})^T \sim N\left(\frac{A\mu_x + \sigma_x^2\left(X_{t-1} + \beta_1(Y_{t-1} - \beta_0)\right)}{D}, \frac{\sigma_x^2 A}{D}\right).$$

APPENDIX B

DERIVATIVES AND INFORMATION MATRIX IN STRUCTURAL CASE

The first derivatives of the structural case log-likelihood function taken with respect to $(\mu_x, \sigma_x^2, \sigma^2, \beta_0, \beta_1, \rho)$ are

$$\frac{\partial \ell}{\partial \mu_x} = \frac{1}{2|\Sigma|}\left\{2\sigma^2\left(1-\rho\frac{A}{D}\right)\sum_{t=2}^{n}[(X_t-\mu_1)+\beta_1(Y_t-\mu_2)]\right\},$$

$$\frac{\partial \ell}{\partial \sigma_x^2} = -(n-1)\frac{(1+\beta_1^2)\sigma^2}{2|\Sigma|}\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)-\rho\frac{\sigma^2 A}{|\Sigma|D^2}\sum_{t=2}^{n}\left\{\left((X_t-\mu_1)+\beta_1(Y_t-\mu_2)\right)G_{t-1}\right\}$$

$$+\frac{1}{2|\Sigma|^2}\sigma^4\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\sum_{t=2}^{n}\left\{\left((X_t-\mu_1)+\beta_1(Y_t-\mu_2)\right)^2\right\},$$

$$\frac{\partial \ell}{\partial \sigma^2} = -(n-1)\frac{C\left(1+\frac{\rho^2 A}{D}+\frac{\rho^2 AC}{D^2}\right)+2\sigma^2}{2|\Sigma|}+\frac{\sigma^2\left(\rho^2\frac{AC^2}{D^2}+\sigma^2\right)}{2|\Sigma|^2(1+\beta_1^2)}\sum_{t=2}^{n}\left\{\left((X_t-\mu_1)+\beta_1(Y_t-\mu_2)\right)^2\right\}$$

$$+\frac{1}{2(1+\beta_1^2)\sigma^4}\sum_{t=2}^{n}\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\}^2+\frac{\rho A\sigma_x^2\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}G_{t-1}}{|\Sigma|D^2},$$

$$\frac{\partial \ell}{\partial \beta_0} = \frac{\beta_1\sigma^2}{(1+\beta_1^2)|\Sigma|}\left(1-\rho\frac{A}{D}\right)\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}$$

$$-\frac{(1-\rho)}{(1+\beta_1^2)\sigma^2}\sum_{t=2}^{n}\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\},$$

$$\frac{\partial \ell}{\partial \beta_1} = -(n-1)\frac{1}{|\Sigma|}\beta_1\sigma_x^2\sigma^2\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)+\frac{\beta_1}{(1+\beta_1^2)^2\sigma^2}\sum_{t=2}^{n}\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\}^2$$

$$+\frac{\beta_1\sigma^2}{(1+\beta_1^2)^2|\Sigma|^2}\left\{|\Sigma|+C\sigma^2\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\right\}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}^2$$

$$-\frac{1}{(1+\beta_1^2)\sigma^2}\sum_{t=2}^{n}\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\}(X_t-\rho X_{t-1})$$

$$-\frac{\sigma^2\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}\left\{(Y_t-\beta_0)-\rho\frac{A}{D}(Y_{t-1}-\beta_0)+2\beta_1\rho\frac{A}{D}\frac{(A\mu_x+\sigma_x^2E_{t-1})}{D}-2\beta_1\mu_x\right\}}{(1+\beta_1^2)|\Sigma|},$$

$$\frac{\partial \ell}{\partial \rho} = -(n-1)\frac{1}{|\Sigma|}\rho\frac{A^2C(C+\sigma^2)}{D^2} + \frac{1}{1+\beta_1^2}\sigma^2\rho\frac{A^2C(C+\sigma^2)}{D^2|\Sigma|^2}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}^2$$

$$+\frac{1}{1+\beta_1^2}\sigma^2\left(\frac{A}{D}+\frac{2\rho^2}{(1-\rho^2)}\frac{AC}{D^2}\right)\frac{1}{|\Sigma|}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}G_{t-1}$$

$$+\frac{1}{(1+\beta_1^2)\sigma^2}\sum_{t=2}^{n}\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\}F_{t-1}.$$

Therefore, taking the derivatives again of $\frac{\partial \ell}{\partial \mu_x}$, we have second order partial derivatives shown below

$$\frac{\partial^2 \ell}{\partial \mu_x^2} = -(n-1)\frac{\sigma^2}{|\Sigma|}(1+\beta_1^2)\left(1-\rho\frac{A}{D}\right)^2,$$

$$\frac{\partial^2 \ell}{\partial \mu_x \partial \sigma_x^2} = \frac{(1+\beta_1^2)\sigma^2\left\{\rho\frac{A}{D^2}-\frac{\sigma^2}{|\Sigma|}\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\left(1-\rho\frac{A}{D}\right)\right\}}{|\Sigma|}\sum_{t=2}^{n}\left\{E_t-\rho\frac{A}{D}E_{t-1}-(1+\beta_1^2)\left(1-\frac{\rho A}{D}\right)\mu_x\right\}$$

$$+(1+\beta_1^2)\frac{\sigma^2}{|\Sigma|}\left(1-\rho\frac{A}{D}\right)\rho\frac{A}{D^2}\sum_{t=2}^{n}(E_{t-1}-(1+\beta_1^2)\mu_x),$$

$$\frac{\partial^2 \ell}{\partial \mu_x \partial \sigma^2} = -\frac{\left\{\frac{\sigma^2}{|\Sigma|}\left(\rho^2\frac{AC^2}{D^2}+\sigma^2\right)\left(1-\rho\frac{A}{D}\right)+\rho\frac{AC}{D^2}\right\}}{|\Sigma|}\sum_{t=2}^{n}\left\{E_t-\rho\frac{A}{D}E_{t-1}-(1+\beta_1^2)\left(1-\rho\frac{A}{D}\right)\mu_x\right\}$$

$$-\frac{1}{|\Sigma|}\rho\frac{AC}{D^2}\left(1-\rho\frac{A}{D}\right)\sum_{t=2}^{n}\{E_{t-1}-(1+\beta_1^2)\mu_x\},$$

$$\frac{\partial^2 \ell}{\partial \mu_x \partial \beta_0} = -(n-1)\beta_1\frac{\sigma^2}{|\Sigma|}\left(1-\rho\frac{A}{D}\right)^2,$$

$$\frac{\partial^2 \ell}{\partial \mu_x \partial \beta_1} = 2\frac{\beta_1\sigma_x^2\sigma^2\left\{\rho\frac{A}{D^2}-\frac{\sigma^2}{|\Sigma|}\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\left(1-\rho\frac{A}{D}\right)\right\}}{|\Sigma|}\sum_{t=2}^{n}\left\{E_t-\rho\frac{A}{D}E_{t-1}-(1+\beta_1^2)\left(1-\rho\frac{A}{D}\right)\mu_x\right\}$$

$$+\frac{\sigma^2\left(1-\rho\frac{A}{D}\right)\sum_{t=2}^{n}\left\{(Y_t-\beta_0)-\rho\frac{A}{D}(Y_{t-1}-\beta_0)+2\frac{\beta_1\sigma_x^2\rho A}{D^2}E_{t-1}-2\beta_1\left[1-\rho\left(\frac{A}{D}\right)^2\right]\mu_x\right\}}{|\Sigma|},$$

$$\frac{\partial^2 \ell}{\partial \mu_x \partial \rho} = \frac{\sigma^2}{|\Sigma|}\frac{A\left\{-\frac{2\rho C}{|\Sigma|}(C+\sigma^2)\frac{A}{D}\left(1-\frac{\rho A}{D}\right)-\left(1+\frac{2\rho^2}{1-\rho^2}\frac{C}{D}\right)\right\}}{D}\sum_{t=2}^{n}\left\{E_t-\frac{\rho A}{D}E_{t-1}-(1+\beta_1^2)\left(1-\frac{\rho A}{D}\right)\mu_x\right\}$$

$$-\frac{\sigma^2}{|\Sigma|}\frac{A}{D}\left(1-\rho\frac{A}{D}\right)\left(1+\frac{2\rho^2}{1-\rho^2}\frac{C}{D}\right)\sum_{t=2}^{n}(E_{t-1}-(1+\beta_1^2)\mu_x).$$

Taking derivatives of $\frac{\partial \ell}{\partial \sigma_x^2}$, we have

$$\frac{\partial^2 \ell}{\partial \left(\sigma_x^2\right)^2} = (n-1)\frac{1}{2|\Sigma|^2}(1+\beta_1^2)^2\sigma^2\left\{\sigma^2\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)^2 + 2|\Sigma|\rho^2\frac{A^2}{D^3}\right\}$$

$$-\frac{(1+\beta_1^2)\sigma^4\left\{\sigma^2\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)^2+|\Sigma|\rho^2\frac{A^2}{D^3}\right\}}{|\Sigma|^3}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}^2$$

$$+2\frac{(1+\beta_1^2)\sigma^4\rho\frac{A}{D^2}\left\{1+\rho^2\left(\frac{A}{D}\right)^2+\frac{1}{D}\frac{|\Sigma|}{\sigma^2}\right\}}{|\Sigma|^2}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}G_{t-1}$$

$$-(1+\beta_1^2)\sigma^2\rho^2\frac{A^2}{D^4}\frac{1}{|\Sigma|}\sum_{t=2}^{n}G_{t-1}^2,$$

$$\frac{\partial^2 \ell}{\partial \sigma_x^2 \partial \sigma^2} = (n-1)(1+\beta_1^2)\frac{1}{2|\Sigma|}\left\{\frac{\sigma^2}{|\Sigma|}\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\left(\rho^2\frac{AC^2}{D^2}+\sigma^2\right)-2\rho^2\left(\frac{A}{D}\right)^2\frac{C}{D}\right\}$$

$$-\frac{\sigma^4\left\{C\left(1+\rho^2\frac{AC^2}{D^3}\right)+\sigma^2\left(1+\rho^2\left(\frac{A}{D}\right)^3\right)\right\}}{|\Sigma|^3}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}^2$$

$$+2\frac{\rho\sigma^2\frac{A}{D^2}\left(\frac{A}{D}\sigma^2-\frac{C^2}{D}\right)}{|\Sigma|^2}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}G_{t-1}+\frac{\rho^2A^2C}{|\Sigma|D^4}\sum_{t=2}^{n}G_{t-1}^2,$$

$$\frac{\partial^2 \ell}{\partial \sigma_x^2 \partial \beta_0} = \frac{\sigma^2}{|\Sigma|}\beta_1\left[\rho\frac{A}{D^2}-\frac{\sigma^2}{|\Sigma|}\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\left(1-\rho\frac{A}{D}\right)\right]\sum_{t=2}^{n}\left((X_t-\mu_1)+\beta_1(Y_t-\mu_2)\right)$$

$$+\frac{\sigma^2}{|\Sigma|}\beta_1\rho\frac{A}{D^2}\left(1-\rho\frac{A}{D}\right)\sum_{t=2}^{n}G_{t-1},$$

$$\frac{\partial^2 \ell}{\partial \sigma_x^2 \partial \beta_1} = (n-1)\beta_1\frac{\sigma^4}{|\Sigma|^2}\left\{\rho^2\frac{AC^2}{D^2}\left(2\frac{A}{D}-1+\rho^2\left(\frac{A}{D}\right)^2\right)-\sigma^2\left(1+\rho^2\left(\frac{A}{D}\right)^2-2\rho^2\left(\frac{A}{D}\right)^2\frac{C}{D}\right)\right\}$$

$$-2\beta_1\sigma_x^2\frac{\sigma^6}{|\Sigma|^3}\left\{\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)^2+\rho^2\frac{A^2}{D^3}\frac{|\Sigma|}{\sigma^2}\right\}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}^2$$

$$+2\beta_1\sigma_x^2\rho\frac{A}{D^2}\frac{\sigma^4}{|\Sigma|^2}\left\{1+\rho^2\left(\frac{A}{D}\right)^2+2\frac{|\Sigma|}{\sigma^2}\frac{1}{D}\right\}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}G_{t-1}$$

$$+\frac{\sigma^4\left\{1+\rho^2\left(\frac{A}{D}\right)^2\right\}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}\left\{\rho A\frac{(Y_{t-1}-\beta_0)D-2\beta_1(A\mu_x+\sigma_x^2E_{t-1})}{D^2}+(Y_t-\beta_0)-2\beta_1\mu_x\right\}}{|\Sigma|^2}$$

$$-\frac{1}{|\Sigma|}\sigma^2\rho\frac{A}{D^2}\sum_{t=2}^{n}\left\{\rho A\frac{(Y_{t-1}-\beta_0)D-2\beta_1\left(A\mu_x+\sigma_x^2E_{t-1}\right)}{D^2}+(Y_t-\beta_0)-2\beta_1\mu_x\right\}G_{t-1}$$

$$-\frac{1}{|\Sigma|}\sigma^2\rho\frac{A}{D^2}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}\{(Y_{t-1}-\beta_0)-2\beta_1\mu_x\},$$

$$\frac{\partial^2\ell}{\partial\sigma_x^2\partial\rho}=(n-1)(1+\beta_1^2)\rho\frac{A^2}{D^2}\frac{\sigma^2}{|\Sigma|^2}\left\{C(C+\sigma^2)\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)-|\Sigma|\left(1+\frac{2\rho^2}{1-\rho^2}\frac{C}{D}\right)\right\}$$

$$+\frac{\rho\frac{A^2}{D^2}\sigma^4\left\{-2C\left(C+\sigma^2\right)\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)+|\Sigma|\left(1+\frac{2\rho^2}{1-\rho^2}\frac{C}{D}\right)\right\}}{|\Sigma|^3}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}^2$$

$$+\frac{\frac{A\sigma^2}{D}\left\{\frac{\rho^2A^2\sigma^2}{D^2}-\left\{\frac{2\rho^2AC^2}{D^2}+\left(\frac{\rho^2AC}{D^2}+1\right)\sigma^2\right\}\left(\frac{C}{D}+1\right)-\frac{\sigma^4}{D}-\frac{4\rho^4A^3C^2}{D^4}\right\}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}G_{t-1}}{|\Sigma|^2}$$

$$+\frac{\sigma^2}{|\Sigma|}\rho\frac{A}{D^2}\left(\frac{A}{D}+\frac{2\rho^2}{(1-\rho^2)}\frac{AC}{D^2}\right)\sum_{t=2}^{n}G_{t-1}^2.$$

Similarly, we have other second order partial derivatives with respect to σ^2 as

$$\frac{\partial^2\ell}{\partial(\sigma^2)^2}=(n-1)\frac{1}{2|\Sigma|^2}\left\{C\left(1+\rho^2\frac{A}{D}+\rho^2\frac{AC}{D^2}\right)+2\sigma^2\right\}^2-\frac{1}{|\Sigma|}(n-1)\left(\frac{\rho^2}{1-\rho^2}\frac{C^3}{D^3}+1\right)$$

$$-\frac{1}{1+\beta_1^2}\frac{1}{\sigma^6}\sum_{t=2}^{n}\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\}^2-\frac{1}{|\Sigma|}\sigma_x^2\frac{\rho^2}{1-\rho^2}\frac{AC}{D^4}\sum_{t=2}^{n}G_{t-1}^2$$

$$-\frac{1}{|\Sigma|^2}\frac{1}{1+\beta_1^2}\left\{\frac{\sigma^2}{|\Sigma|}\left(\rho^2\frac{AC^2}{D^2}+\sigma^2\right)^2+\rho^2\frac{A^2C^2}{D^3}\right\}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}^2$$

$$-\frac{1}{|\Sigma|}\frac{2\rho\sigma_x^2A}{D^2}\left\{\frac{1}{1-\rho^2}\frac{1}{D}+\frac{1}{|\Sigma|}\left(\rho^2\frac{AC^2}{D^2}+\sigma^2\right)\right\}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}G_{t-1},$$

$$\frac{\partial^2\ell}{\partial\sigma^2\partial\beta_0}=\frac{1}{|\Sigma|}(1-\rho)\left[\sigma_x^2\left(\frac{1}{\sigma^2}+\frac{\rho^2}{1-\rho^2}\frac{1}{D}\right)+\frac{1}{1+\beta_1^2}\right]\sum_{t=2}^{n}\left(\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\right)$$

$$-\frac{\beta_1}{1+\beta_1^2}\frac{\sigma^2}{|\Sigma|^2}\left\{\left(\rho^2\frac{AC^2}{D^2}+\sigma^2\right)\left(1-\rho\frac{A}{D}\right)+\frac{|\Sigma|}{\sigma^2}\frac{\rho AC}{D^2}\right\}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}$$

$$-\frac{1}{|\Sigma|}\rho\beta_1\sigma_x^2\frac{A}{D^2}\left(1-\rho\frac{A}{D}\right)\sum_{t=2}^{n}G_{t-1}$$

$$\frac{\partial^2\ell}{\partial\sigma^2\partial\beta_1}=(n-1)\beta_1\sigma_x^2\frac{\sigma^2}{|\Sigma|^2}\left\{\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\left[\frac{|\Sigma|}{\sigma^2}+\left(\rho^2\frac{AC^2}{D^2}+\sigma^2\right)\right]-\frac{|\Sigma|}{\sigma^2}\rho^2\frac{A^2}{D^2}\left(1+2\frac{C}{D}\right)\right\}$$

$$-\frac{\beta_1}{\left(1+\beta_1^2\right)^2}\frac{1}{\sigma^4}\sum_{t=2}^{n}\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\}^2$$

$$+\frac{\beta_1\left\{\frac{2\rho^2A^2C^2|\Sigma|}{D^3\quad\sigma^2}-\left(\frac{\rho^2AC^2}{D^2}+\sigma^2\right)\left\{\frac{|\Sigma|}{\sigma^2}-2C\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\right\}\right\}}{\left(1+\beta_1^2\right)^2}\frac{\sigma^4}{|\Sigma|^3}\sum_{t=2}^n\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}^2$$

$$-2\frac{A}{D^2}\frac{\rho\beta_1\sigma_x^4}{|\Sigma|^2}\left\{2\frac{|\Sigma|}{D}+\sigma^2\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\right\}\sum_{t=2}^n\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}G_{t-1}$$

$$+\frac{1}{1+\beta_1^2}\frac{1}{\sigma^4}\sum_{t=2}^n\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\}(X_t-\rho X_{t-1})$$

$$+\frac{\sigma^2}{|\Sigma|^2}\frac{\left(\frac{\rho^2AC^2}{D^2}+\sigma^2\right)\sum_{t=2}^n\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}\left\{\frac{\rho A[(Y_{t-1}-\beta_0)D-2\beta_1(A\mu_x+\sigma_x^2E_{t-1})]}{D^2}+Y_t-\beta_0-2\beta_1\mu_x\right\}}{1+\beta_1^2}$$

$$+\frac{1}{|\Sigma|}\rho\sigma_x^2\frac{A}{D^2}\sum_{t=2}^n\left\{\rho A\frac{(Y_{t-1}-\beta_0)D-2\beta_1\left(A\mu_x+\sigma_x^2E_{t-1}\right)}{D^2}+(Y_t-\beta_0)-2\beta_1\mu_x\right\}G_{t-1}$$

$$+\frac{1}{|\Sigma|}\rho\sigma_x^2\frac{A}{D^2}\sum_{t=2}^n\big((X_t-\mu_1)+\beta_1(Y_t-\mu_2)\big)\big((Y_{t-1}-\beta_0)-2\beta_1\mu_x\big),$$

$$\frac{\partial^2\ell}{\partial\sigma^2\partial\rho}=(n-1)\rho\frac{AC}{D}\frac{1}{|\Sigma|}\left\{\frac{1}{1-\rho^2}\frac{\left(C+\sigma^2\right)}{D}+2\frac{\rho^2}{1-\rho^2}\frac{AC}{D^2}-\left(1+\frac{C}{D}\right)+\left(\rho^2\frac{AC^2}{D^2}+\sigma^2\right)\frac{1}{|\Sigma|}\frac{A\left(C+\sigma^2\right)}{D}\right\}$$

$$+\rho\frac{AC}{D^2}\frac{\sigma^2}{|\Sigma|^2}\frac{\left\{C\left(1+\frac{\rho^2}{1-\rho^2}\frac{D-2A}{D}\right)-2\frac{1}{|\Sigma|}(C+\sigma^2)A\left(\rho^2\frac{AC^2}{D^2}+\sigma^2\right)\right\}}{1+\beta_1^2}\sum_{t=2}^n\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}^2$$

$$+\frac{\frac{A}{D}\left\{\frac{C^2}{D}\sigma^2-\frac{2\rho^2A^2C(D+C)}{D^3}\sigma^2-\frac{2\rho^2A^2C^2|\Sigma|}{D^2\quad\sigma^2}-\frac{A}{D}\sigma^4\right\}}{\left(1+\beta_1^2\right)|\Sigma|^2}\sum_{t=2}^n\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}G_{t-1}$$

$$-\frac{1}{|\Sigma|}\left[\sigma_x^2\left(\frac{1}{\sigma^2}+\frac{\rho^2}{1-\rho^2}\frac{1}{D}\right)+\frac{1}{1+\beta_1^2}\right]\sum_{t=2}^n\big(\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\big)F_{t-1}$$

$$-\frac{1}{|\Sigma|}\rho\sigma_x^2\frac{A}{D^2}\left(\frac{A}{D}+\frac{2\rho^2}{(1-\rho^2)}\frac{AC}{D^2}\right)\sum_{t=2}^n G_{t-1}^2.$$

Other second order partial derivatives are

$$\frac{\partial^2\ell}{\partial(\beta_0)^2}=-\frac{1}{|\Sigma|}(n-1)\left\{\frac{1}{1+\beta_1^2}\frac{|\Sigma|}{\sigma^2}(1-\rho)^2+\frac{\beta_1^2}{1+\beta_1^2}\sigma^2\left(1-\rho\frac{A}{D}\right)^2\right\}$$

$$\frac{\partial^2\ell}{\partial\beta_0\partial\beta_1}=2\frac{\beta_1}{\left(1+\beta_1^2\right)^2}\frac{(1-\rho)}{\sigma^2}\sum_{t=2}^n\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\}-\frac{1}{1+\beta_1^2}\frac{(1-\rho)}{\sigma^2}\sum_{t=2}^n(X_t-\rho X_{t-1})$$

$$+\frac{\sigma^4\left\{\frac{|\Sigma|}{\sigma^2}\left[\left(1-\beta_1^2\right)\left(1-\rho\frac{A}{D}\right)+2\beta_1^2\rho\frac{AC}{D^2}\right]-2\beta_1^2C\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\left(1-\rho\frac{A}{D}\right)\right\}\sum_{t=2}^n\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}}{\left(1+\beta_1^2\right)^2|\Sigma|^2}$$

$$+\frac{\sigma^2}{|\Sigma|}\frac{\beta_1}{1+\beta_1^2}\left(1-\rho\frac{A}{D}\right)\sum_{t=2}^{n}\left\{\frac{\rho A\left[(Y_{t-1}-\beta_0)D-2\beta_1\left(A\mu_x+\sigma_x^2 E_{t-1}\right)\right]}{D^2}+Y_t-\beta_0-2\beta_1\mu_x\right\},$$

$$\frac{\partial^2 \ell}{\partial\beta_0\partial\rho}=-\frac{\beta_1}{1+\beta_1^2}\frac{A}{D}\frac{\left\{2\sigma^2\frac{\rho AC(C+\sigma^2)\left(1-\rho\frac{A}{D}\right)}{D}+\sigma^2|\Sigma|+2\rho^2\frac{AC}{D}|\Sigma|\right\}}{|\Sigma|^2}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}$$

$$+(1-\rho)\frac{1}{1+\beta_1^2}\frac{1}{\sigma^2}\sum_{t=2}^{n}F_{t-1}+\frac{1}{1+\beta_1^2}\frac{1}{\sigma^2}\sum_{t=2}^{n}\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\}$$

$$-\frac{1}{|\Sigma|}\frac{\beta_1}{1+\beta_1^2}\sigma^2\left(\frac{A}{D}+\frac{2\rho^2}{(1-\rho^2)}\frac{AC}{D^2}\right)\left(1-\rho\frac{A}{D}\right)\sum_{t=2}^{n}G_{t-1},$$

$$\frac{\partial^2 \ell}{\partial(\beta_1)^2}=(n-1)\sigma_x^2\sigma^2\frac{1}{|\Sigma|}\left\{\frac{1}{|\Sigma|}2\beta_1^2\sigma_x^2\sigma^2\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)^2-\left(1+\rho^2\left(\frac{A}{D}\right)^2-4\beta_1^2\sigma_x^2\rho^2\frac{A^2}{D^3}\right)\right\}$$

$$+\frac{(1-3\beta_1^2)}{(1+\beta_1^2)^3\sigma^2}\sum_{t=2}^{n}\left(\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\right)^2-\frac{1}{1+\beta_1^2}\frac{1}{\sigma^2}\sum_{t=2}^{n}(X_t-\rho X_{t-1})^2$$

$$+\frac{\sigma^4\left\{(1-3\beta_1^2)\left\{C\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)+\frac{|\Sigma|}{\sigma^2}\right\}-4\beta_1^2C^2\left\{\frac{\sigma^2}{|\Sigma|}\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)^2+\rho^2\frac{A^2}{D^3}\right\}\right\}\sum_{t=2}^{n}\left\{\left((X_t-\mu_1)+\beta_1(Y_t-\mu_2)\right)^2\right\}}{(1+\beta_1^2)^3|\Sigma|^2}$$

$$+4\beta_1\frac{1}{(1+\beta_1^2)^2\sigma^2}\sum_{t=2}^{n}\{\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\}(X_t-\rho X_{t-1})$$

$$+\frac{4\beta_1\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}\left\{\frac{\rho A\left[(Y_{t-1}-\beta_0)D-2\beta_1\left(A\mu_x+\sigma_x^2E_{t-1}\right)\right]}{D^2}+Y_t-\beta_0-2\beta_1\mu_x\right\}}{(1+\beta_1^2)^2|\Sigma|\sigma^{-2}\left\{\frac{\sigma^2}{|\Sigma|}C\left[1+\rho^2\left(\frac{A}{D}\right)^2\right]+1\right\}^{-1}}$$

$$-\frac{1}{1+\beta_1^2}\frac{\sigma^2}{|\Sigma|}\sum_{t=2}^{n}\left\{\frac{\rho A\left[(Y_{t-1}-\beta_0)D-2\beta_1\left(A\mu_x+\sigma_x^2E_{t-1}\right)\right]}{D^2}+Y_t-\beta_0-2\beta_1\mu_x\right\}^2$$

$$-2\frac{\sigma^2\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}\left\{\left(1-\frac{4\beta_1^2\sigma_x^2}{D}\right)\frac{\rho A\left(A\mu_x+\sigma_x^2E_{t-1}\right)}{D^2}+\frac{2\rho\beta_1\sigma_x^2A(Y_{t-1}-\beta_0)}{D^2}-\mu_x\right\}}{(1+\beta_1^2)|\Sigma|},$$

$$\frac{\partial^2 \ell}{\partial\beta_1\partial\rho}=2(n-1)\rho\beta_1\sigma_x^2\left\{\frac{1}{|\Sigma|^2}\sigma^2\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)C(C+\sigma^2)\frac{A^2}{D^2}-\frac{1}{|\Sigma|}A^2\left(2\frac{AC}{D^3}+\sigma^2\frac{D-2C}{D^3}\right)\right\}$$

$$+\frac{2\rho\beta_1A^2\sigma^2\sigma_x^4\left\{-2\frac{\sigma^2}{|\Sigma|}\left(1+\rho^2\left(\frac{A}{D}\right)^2\right)\frac{(C+\sigma^2)}{D^2}+\frac{D-2(C+\sigma^2)}{D^3}\right\}}{|\Sigma|^2}\sum_{t=2}^{n}\{(X_t-\mu_1)+\beta_1(Y_t-\mu_2)\}^2$$

$$-2\beta_1 \frac{1}{\sigma^2} \frac{1}{\left(1+\beta_1^2\right)^2} \sum_{t=2}^{n} \left(\beta_1 (X_t - \mu_1) - (Y_t - \mu_2)\right) F_{t-1}$$

$$-2\, \frac{\beta_1 \sigma^4 \frac{A}{D} \left\{ \frac{|\Sigma|}{\sigma^2} \left(1 + \frac{C}{D} + \frac{4\rho^2 \quad C^2}{(1-\rho^2) D^2}\right) + C\left(1 + \rho^2 \left(\frac{A}{D}\right)^2\right)\left(1 + \frac{2\rho^2 \quad C}{(1-\rho^2) D}\right)\right\} \sum_{t=2}^{n} \left((X_t - \mu_1) + \beta_1 (Y_t - \mu_2)\right) G_{t-1}}{\left(1+\beta_1^2\right)^2 |\Sigma|^2}$$

$$+ \frac{2\rho A^2 (C+\sigma^2) \sigma^2 \sigma_X^2 \sum_{t=2}^{n} \{(X_t - \mu_1) + \beta_1 (Y_t - \mu_2)\} \left\{ \frac{\rho A[(Y_{t-1} - \beta_0) D - 2\beta_1 (A\mu_x + \sigma_X^2 E_{t-1})]}{D^2} + Y_t - \beta_0 - 2\beta_1 \mu_x \right\}}{D^2 |\Sigma|^2}$$

$$+ \frac{1}{1+\beta_1^2} \frac{1}{\sigma^2} \sum_{t=2}^{n} (X_t - \rho X_{t-1}) F_{t-1} + \frac{1}{1+\beta_1^2} \frac{1}{\sigma^2} \sum_{t=2}^{n} \{\beta_1 (X_t - \mu_1) - (Y_t - \mu_2)\} X_{t-1}$$

$$+ \frac{1}{1+\beta_1^2} \frac{A}{D} \left(\sigma^2 + \frac{2\rho^2 AC}{D}\right) \frac{\sum_{t=2}^{n} \left\{ \frac{\rho A[(Y_{t-1} - \beta_0) D - 2\beta_1 (A\mu_x + \sigma_X^2 E_{t-1})]}{D^2} + Y_t - \beta_0 - 2\beta_1 \mu_x \right\} G_{t-1}}{|\Sigma|}$$

$$+ \frac{1}{|\Sigma|} \frac{1}{1+\beta_1^2} \frac{A}{D} \left(\sigma^2 + \frac{2\rho^2 AC}{D}\right) \sum_{t=2}^{n} \{(X_t - \mu_1) + \beta_1 (Y_t - \mu_2)\}\{(Y_{t-1} - \beta_0) - 2\beta_1 \mu_x\},$$

$$\frac{\partial^2 \ell}{\partial \rho^2} = (n-1) C (C + \sigma^2) \left(\frac{A}{D}\right)^2 \left\{ \frac{1}{|\Sigma|^2} 2\rho^2 C (C + \sigma^2) \left(\frac{A}{D}\right)^2 - \frac{1}{|\Sigma|} \left(1 + 4 \frac{\rho^2}{1-\rho^2} \frac{C}{D}\right) \right\}$$

$$+ \frac{\sigma^2 \sigma_X^2}{|\Sigma|^2} \frac{A^2 (C+\sigma^2)}{D^2} \left\{ \frac{4\rho^2 AC}{D|\Sigma|} \left(\frac{|\Sigma|}{\sigma^2} - \frac{A(C+\sigma^2)}{D}\right) + 1 \right\} \sum_{t=2}^{n} \{(X_t - \mu_1) + \beta_1 (Y_t - \mu_2)\}^2$$

$$+2 \frac{\rho A^2}{D^2} \frac{\sigma_X^2 \left\{ \frac{|\Sigma|}{\sigma^2} \left(\sigma^2 + 2(1+\rho^2) A - \frac{4\rho^2 A^2}{D}\right) - \frac{2A(C+\sigma^2)}{D} \left(\sigma^2 + \frac{2\rho^2 AC}{D}\right) \right\} \sum_{t=2}^{n} \{(X_t - \mu_1) + \beta_1 (Y_t - \mu_2)\} G_{t-1}}{|\Sigma|^2}$$

$$- \frac{1}{1+\beta_1^2} \frac{1}{\sigma^2} \sum_{t=2}^{n} F_{t-1}^2 - \frac{1}{|\Sigma|} \frac{1}{1+\beta_1^2} \sigma^2 \left(\frac{A}{D} + \frac{2\rho^2}{(1-\rho^2)} \frac{AC}{D^2}\right)^2 \sum_{t=2}^{n} G_{t-1}^2.$$

According to the assumptions of the model, the expectations of terms in the above second order partial derivatives are

$$E \frac{\sum_{t=2}^{n} \{\beta_1 (X_t - \mu_1) - (Y_t - \mu_2)\}}{n-1} = E \frac{\sum_{t=2}^{n} \{(X_t - \mu_1) + \beta_1 (Y_t - \mu_2)\}}{n-1} = 0,$$

$$E \frac{1}{n-1} \sum_{t=2}^{n} \{\beta_1 (X_t - \mu_1) - (Y_t - \mu_2)\}^2 = (1 + \beta_1^2) \sigma^2,$$

$$E \frac{1}{n-1} \sum_{t=2}^{n} \{(X_t - \mu_1) + \beta_1 (Y_t - \mu_2)\}^2 = (1 + \beta_1^2) \frac{|\Sigma|}{\sigma^2},$$

$$E \frac{\sum_{t=2}^{n} \{(X_t - \mu_1) + \beta_1 (Y_t - \mu_2)\} G_{t-1}}{n-1} = E \frac{1}{n-1} \sum_{t=2}^{n} (Y_t - \mu_2) G_{t-1} = 0,$$

$$E\frac{1}{n-1}\sum_{t=2}^{n} G_{t-1}^2 = E\frac{1}{n-1}\sum_{t=2}^{n} E_{t-1}G_{t-1} = (1+\beta_1^2)D,$$

$$E\frac{\sum_{t=2}^{n}\big((X_t-\mu_1)+\beta_1(Y_t-\mu_2)\big)(Y_{t-1}-\beta_0)}{n-1} = E\frac{\sum_{t=2}^{n}\big((X_t-\mu_1)+\beta_1(Y_t-\mu_2)\big)E_{t-1}}{n-1} = 0,$$

$$E\frac{\sum_{t=2}^{n}\big(\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\big)X_{t-1}}{n-1} = E\frac{\sum_{t=2}^{n}\big(\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\big)E_{t-1}}{n-1} = 0,$$

$$E\frac{1}{n-1}\sum_{t=2}^{n}\big((X_t-\mu_1)+\beta_1(Y_t-\mu_2)\big)(Y_t-\mu_2) = \beta_1\frac{|\Sigma|}{\sigma^2},$$

$$E\frac{1}{n-1}\sum_{t=2}^{n}(Y_{t-1}-\beta_0)G_{t-1} = \beta_1 D,$$

$$E\frac{1}{n-1}\sum_{t=2}^{n}\big(\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\big)F_{t-1} = E\frac{1}{n-1}\sum_{t=2}^{n}(X_t-\rho X_{t-1})F_{t-1} = 0,$$

$$E\frac{1}{n-1}\sum_{t=2}^{n}\big(\beta_1(X_t-\mu_1)-(Y_t-\mu_2)\big)(X_t-\mu_1) = \beta_1\sigma^2,$$

$$E\frac{1}{n-1}\sum_{t=2}^{n} F_{t-1}^2 = (1+\beta_1^2)A,$$

$$E\frac{1}{n-1}\sum_{t=2}^{n}(X_t-\rho X_{t-1})^2 = (1+\rho^2)\sigma_x^2 + (1-\rho)^2\mu_x^2 + \sigma^2,$$

$$E\frac{1}{n-1}\sum_{t=2}^{n}\left(\frac{\rho A\big[(Y_{t-1}-\beta_0)D-2\beta_1\big(A\mu_x+\sigma_x^2E_{t-1}\big)\big]}{D^2} + Y_t - \beta_0 - 2\beta_1\mu_x\right)^2$$

$$= \beta_1^2\sigma_x^2 + \rho^2\left(\frac{AC^2}{D^2} + \beta_1^2\frac{A^2}{D^2}\sigma_x^2 + 4\beta_1^2\frac{A^2C}{D^3}\sigma_x^2\right) + \sigma^2 + \beta_1^2\left(1-\rho\frac{A}{D}\right)^2\mu_x^2.$$

By substituting these into second order partial derivatives and setting ρ equal to zero, we have information matrix under $H_0\colon \rho = 0$ as

$$I(\theta) = \begin{pmatrix} M & 0 \\ 0 & (n-1)\left(1+\frac{\sigma^4}{D_0^2}\right)\end{pmatrix},$$

where

$$D_0 = (1+\beta_1^2)\sigma_x^2 + \sigma^2,$$

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22}\end{pmatrix},$$

$$M_{11} = (n-1)\begin{pmatrix} \frac{(1+\beta_1^2)}{D_0} & 0 & 0 \\ 0 & \frac{(1+\beta_1^2)^2}{2D_0^2} & \frac{(1+\beta_1^2)}{2D_0^2} \\ 0 & \frac{(1+\beta_1^2)}{2D_0^2} & \frac{1}{2}\left(\frac{1}{\sigma^4} + \frac{1}{D_0^2}\right) \end{pmatrix},$$

$$M_{12} = (n-1)\begin{pmatrix} \frac{\beta_1}{D_0} & \frac{\beta_1}{D_0}\mu_x \\ 0 & \frac{\beta_1 C}{D_0^2} \\ 0 & \frac{\beta_1 \sigma_x^2}{D_0^2} \end{pmatrix},$$

and

$$M_{22} = (n-1)\begin{pmatrix} \frac{\sigma_x^2+\sigma^2}{\sigma^2 D_0} & \frac{\sigma_x^2+\sigma^2}{\sigma^2 D_0}\mu_x \\ \frac{\sigma_x^2+\sigma^2}{\sigma^2 D_0}\mu_x & \frac{1}{|\Sigma_0|}\left\{\sigma_x^4 + 2\beta_1^2 \frac{\sigma^4\sigma_x^4}{|\Sigma_0|} + (\sigma_x^2+\sigma^2)\mu_x\right\} \end{pmatrix}.$$

APPENDIX C

DERIVATIVES AND INFORMATION MATRIX IN FUNCTIONAL CASE

FOR $H_0: \rho = 0$

Taking the second order partial derivatives of functional case log-likelihood function with respect to $\psi = (x_1, x_2, \dots, x_n, \beta_0, \beta_1, \sigma^2, \rho)$ and setting ρ equal to zero, we have

$$\frac{\partial^2 \ell}{\partial x_t^2} = -\sigma^{-2}(1+\beta_1^2), t = 2, \dots, n$$

$$\frac{\partial^2 \ell}{\partial x_t \partial \beta_0} = -\sigma^{-2}\beta_1, \qquad t = 2, \dots, n$$

$$\frac{\partial^2 \ell}{\partial x_t \partial \beta_1} = \sigma^{-2}(Y_t - y_t - \beta_1 x_t), \qquad t = 2, \dots, n$$

$$\frac{\partial^2 \ell}{\partial x_t \partial \sigma^2} = -\sigma^{-4}\{(X_t - x_t) + \beta_1(Y_t - y_t)\}, \quad t = 2, \dots, n$$

$$\frac{\partial^2 \ell}{\partial x_t \partial \rho} = -\frac{\{(X_{t-1}-x_{t-1})+\beta_1(Y_{t-1}-y_{t-1})+(X_{t+1}-x_{t+1})+\beta_1(Y_{t+1}-y_{t+1})\}}{\sigma^2}, \quad t = 2, \dots, n-1$$

$$\frac{\partial^2 \ell}{\partial x_n \partial \rho} = -\sigma^{-2}\{(X_{n-1} - x_{n-1}) + \beta_1(Y_{n-1} - y_{n-1})\},$$

$$\frac{\partial^2 \ell}{\partial \beta_0^2} = -(n-1)\sigma^{-2},$$

$$\frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} = -\sigma^{-2}\sum_{t=2}^{n} x_t,$$

$$\frac{\partial^2 \ell}{\partial \beta_0 \partial \sigma^2} = -\sigma^{-4}\sum_{t=2}^{n}(Y_t - y_t),$$

$$\frac{\partial^2 \ell}{\partial \beta_0 \partial \rho} = -\sigma^{-2}\sum_{t=2}^{n}\{(Y_{t-1} - y_{t-1}) + (Y_t - y_t)\},$$

$$\frac{\partial^2 \ell}{\partial \beta_1^2} = -\sigma^{-2}\sum_{t=2}^{n} x_t^2,$$

$$\frac{\partial^2 \ell}{\partial \beta_1 \partial \sigma^2} = -\sigma^{-4} \sum_{t=2}^{n} (Y_t - y_t) x_t,$$

$$\frac{\partial^2 \ell}{\partial \beta_1 \partial \rho} = -\sigma^{-2} \sum_{t=2}^{n} \{(Y_{t-1} - y_{t-1}) x_t + (Y_t - y_t) x_{t-1}\},$$

$$\frac{\partial^2 \ell}{\partial (\sigma^2)^2} = -\sigma^{-6} \{\sum_{t=2}^{n} [(X_t - x_t)^2 + (Y_t - y_t)^2] - (n-1)\sigma^2\},$$

$$\frac{\partial^2 \ell}{\partial \sigma^2 \partial \rho} = -\sigma^{-4} \sum_{t=2}^{n} \{(X_t - x_t)(X_{n-1} - x_{n-1}) + (Y_t - y_t)(Y_{t-1} - y_{t-1})\},$$

and

$$\frac{\partial^2 \ell}{\partial \rho^2} = -\sigma^{-4} \sum_{t=2}^{n} \{(X_{t-1} - x_{t-1})^2 + (Y_{t-1} - y_{t-1})^2\}.$$

Therefore, the Hessians matrix under the null hypothesis is

$$H(\psi)|_{\rho=0} = \begin{pmatrix} D & B \\ B^T & A \end{pmatrix},$$

where

$$A = \frac{-1}{\sigma^2} \begin{bmatrix} (n-1) & \sum_{t=2}^{n} x_t & \frac{1}{\sigma^2} \sum_{t=2}^{n} (Y_t - y_t) & \sum_{t=2}^{n} \begin{bmatrix} (Y_t - y_t) \\ +(Y_{t-1} - y_{t-1}) \end{bmatrix} \\ \sum_{t=2}^{n} x_t & \sum_{t=2}^{n} x_t^2 & \frac{1}{\sigma^2} \sum_{t=2}^{n} (Y_t - y_t) x_t & \sum_{t=2}^{n} \begin{bmatrix} (Y_t - y_t) x_{t-1} \\ +(Y_{t-1} - y_{t-1}) x_t \end{bmatrix} \\ \frac{1}{\sigma^2} \sum_{t=2}^{n} (Y_t - y_t) & \frac{1}{\sigma^2} \sum_{t=2}^{n} (Y_t - y_t) x_t & \frac{\begin{Bmatrix} \sum_{t=2}^{n} \begin{bmatrix} (X_t - x_t)^2 \\ +(Y_t - y_t)^2 \end{bmatrix} \\ -(n-1)\sigma^2 \end{Bmatrix}}{\sigma^4} & \frac{\sum_{t=2}^{n} \begin{bmatrix} (X_t - x_t)(X_{t-1} - x_{t-1}) \\ +(Y_t - y_t)(Y_{t-1} - y_{t-1}) \end{bmatrix}}{\sigma^2} \\ \sum_{t=2}^{n} \begin{bmatrix} (Y_t - y_t) \\ +(Y_{t-1} - y_{t-1}) \end{bmatrix} & \sum_{t=2}^{n} \begin{bmatrix} (Y_t - y_t) x_{t-1} \\ +(Y_{t-1} - y_{t-1}) x_t \end{bmatrix} & \frac{\sum_{t=2}^{n} \begin{bmatrix} (X_t - x_t)(X_{t-1} - x_{t-1}) \\ +(Y_t - y_t)(Y_{t-1} - y_{t-1}) \end{bmatrix}}{\sigma^2} & \sum_{t=2}^{n} \begin{bmatrix} (X_{t-1} - x_{t-1})^2 \\ +(Y_{t-1} - y_{t-1})^2 \end{bmatrix} \end{bmatrix},$$

$$B = \frac{-1}{\sigma^2} \begin{bmatrix} 0 & 0 & 0 & [(X_2 - x_2) + (Y_2 - y_2)\beta_1] \\ \beta_1 & [x_2 \beta_1 - (Y_2 - y_2)] & \sigma^{-2} [(X_2 - x_2) + (Y_2 - y_2)\beta_1] & \begin{bmatrix} (X_1 - x_1) + (Y_1 - y_1)\beta_1 \\ +(X_3 - x_3) + (Y_3 - y_3)\beta_1 \end{bmatrix} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_1 & [x_n \beta_1 - (Y_n - y_n)] & \sigma^{-2} [(X_n - x_n) + (Y_n - y_n)\beta_1] & [(X_{n-1} - x_{n-1}) + (Y_{n-1} - y_{n-1})\beta_1] \end{bmatrix},$$

and

$$D = \begin{pmatrix} 0 & 0_{1\times(n-1)} \\ 0_{(n-1)\times 1} & -\sigma^{-2}\left(1+\beta_1^2\right) I_{(n-1)\times(n-1)} \end{pmatrix}.$$

Taking the expectation of the Hessians matrix, we have

$$E(D) = D,$$

$$E(B) = -\sigma^{-2} \begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_1 & x_2\beta_1 & 0 & 0 \\ \vdots & \vdots & & \\ \beta_1 & x_n\beta_1 & 0 & 0 \end{bmatrix},$$

and

$$E(A) = -\sigma^{-2} \begin{bmatrix} n-1 & \sum_{t=2}^{n} x_t & 0 & 0 \\ \sum_{t=2}^{n} x_t & \sum_{t=2}^{n} x_t^2 & 0 & 0 \\ 0 & 0 & (n-1)\sigma^{-2} & 0 \\ 0 & 0 & 0 & 2(n-1)\sigma^2 \end{bmatrix}.$$

Thus, the information matrix is

$$I(\psi)|_{\rho=0} = -\begin{pmatrix} E(D) & E(B) \\ E(B^T) & E(A) \end{pmatrix}.$$

APPENDIX D

DERIVATIVES AND INFORMATION MATRIX IN FUNCTIONAL CASE

FOR $H_0: (\beta_0, \beta_1) = (b_0, b_1)$

Taking the second order derivatives of functional case log-likelihood function with respect to $(\lambda_2, \ldots, \lambda_n, \beta_0, \beta_1, \sigma^2, \rho)$ and setting $(\beta_0, \beta_1) = (b_0, b_1)$, we have

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda_t^2} = -\sigma^{-2}\left(1 + b_1^2\right), \qquad t = 2, \ldots, n$$

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda_t \partial \lambda_{t+1}} = 0,$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_0^2} = -(n-1)\sigma^{-2}(1-\rho)^2,$$

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda_t \partial \beta_0} = -b_1 \sigma^{-2}(1-\rho), \qquad t = 2, \ldots, n$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_1^2} = -\sigma^{-2} \sum_{t=2}^{n} \lambda_t^2,$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_0 \partial \beta_1} = -\sigma^{-2}(1-\rho) \sum_{t=2}^{n} \lambda_t,$$

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda_t \partial \beta_1} = \sigma^{-2}\left\{\left(Y_t - \gamma_t - \rho Y_{t-1}\right) - b_1 \lambda_t\right\}, \quad t = 2, \ldots, n$$

$$\frac{\partial^2 \mathcal{L}}{\partial \left(\sigma^2\right)^2} = (n-1)\sigma^{-4} - \sigma^{-6} \sum_{t=2}^{n} \left\{\left(X_t - \lambda_t - \rho X_{t-1}\right)^2 + \left(Y_t - \gamma_t - \rho Y_{t-1}\right)^2\right\},$$

$$\frac{\partial^2 \mathcal{L}}{\partial \sigma^2 \partial \lambda_t} = -\sigma^{-4}\left\{\left(X_t - \lambda_t - \rho X_{t-1}\right) + b_1\left(Y_t - \gamma_t - \rho Y_{t-1}\right)\right\}, \qquad t = 2, \ldots, n$$

$$\frac{\partial^2 \mathcal{L}}{\partial \sigma^2 \partial \beta_0} = -\sigma^{-4}\left(1-\rho\right)\sum_{t=2}^{n}\left(Y_t - \gamma_t - \rho Y_{t-1}\right),$$

$$\frac{\partial^2 \mathcal{L}}{\partial \sigma^2 \partial \beta_1} = -\sigma^{-4}\sum_{t=2}^{n}\left\{\left(Y_t - \gamma_t - \rho Y_{t-1}\right)\lambda_t\right\},$$

$$\frac{\partial^2 \mathcal{L}}{\partial \rho^2} = -\sigma^{-2}\sum_{t=2}^{n}\left\{\left(X_{t-1} - x_{t-1}\right)^2 + \left(Y_{t-1} - y_{t-1}\right)^2\right\},$$

$$\frac{\partial^2 \mathcal{L}}{\partial \rho \partial \lambda_t} = \upsilon^{-2}\left\{\left(x_{t-1} - X_{t-1}\right) + b_1\left(y_{t-1} - Y_{t-1}\right)\right\}, \qquad t = 2,\ldots,n$$

$$\frac{\partial^2 \mathcal{L}}{\partial \rho \partial \beta_0} = -\sigma^{-2}\sum_{t=2}^{n}\left(Y_t - \gamma_t - \rho Y_{t-1}\right) + \sigma^{-2}\left(1-\rho\right)\sum_{t=2}^{n}\left(y_{t-1} - Y_{t-1}\right),$$

$$\frac{\partial^2 \mathcal{L}}{\partial \rho \partial \beta_1} = \upsilon^{-2}\sum_{t=2}^{n}\left\{\left(y_{t-1} - Y_{t-1}\right)\lambda_t + \left(Y_t - \gamma_t - \rho Y_{t-1}\right)\left(-x_{t-1}\right)\right\},$$

and

$$\frac{\partial^2 \mathcal{L}}{\partial \rho \partial \sigma^2} = -\sigma^{-4}\sum_{t=2}^{n}\left\{\left(X_t - \lambda_t - \rho X_{t-1}\right)\left(X_{t-1} - x_{t-1}\right) + \left(Y_t - \gamma_t - \rho Y_{t-1}\right)\left(Y_{t-1} - y_{t-1}\right)\right\}.$$

The information matrix which is the expectation of negative of Hessians matrix is

$$I = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix},$$

where

$$A = \sigma^{-2}(1 + b_1^2) I_{(n-1)\times(n-1)},$$

$$B = \begin{pmatrix} b_1\sigma^{-2}(1-\rho) & b_1\sigma^{-2}\lambda_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ b_1\sigma^{-2}(1-\rho) & b_1\sigma^{-2}\lambda_{n-1} & 0 & 0 \\ b_1\sigma^{-2}(1-\rho) & b_1\sigma^{-2}\lambda_n & 0 & 0 \end{pmatrix},$$

and

$$C = \begin{pmatrix} (n-1)\sigma^{-2}(1-\rho)^2 & (1-\rho)\frac{\sum_{t=2}^{n}\lambda_t}{\sigma^2} & 0 & 0 \\ (1-\rho)\frac{\sum_{t=2}^{n}\lambda_t}{\sigma^2} & \frac{\sum_{t=2}^{n}\lambda_t^2}{\sigma^2} & 0 & 0 \\ 0 & 0 & (n-1)\sigma^{-4} & 0 \\ 0 & 0 & 0 & 2(n-1)(1-\rho^2)^{-1} \end{pmatrix}.$$

There portion of inverse of information matrix corresponding to $(\beta_0, \beta_1, \sigma^2, \rho)$ is

$$(C - B^T A^{-1} B)^{-1} = \begin{bmatrix} \frac{(n-1)(1-\rho)^2}{(1+b_1^2)\sigma^2} & \frac{(1-\rho)\sum_{t=2}^{n}\lambda_t}{(1+b_1^2)\sigma^2} & 0 & 0 \\ \frac{(1-\rho)\sum_{t=2}^{n}\lambda_t}{(1+b_1^2)\sigma^2} & \frac{\sum_{t=2}^{n}\lambda_t^2}{(1+b_1^2)\sigma^2} & 0 & 0 \\ 0 & 0 & (n-1)\sigma^{-4} & 0 \\ 0 & 0 & 0 & 2(n-1)(1-\rho^2)^{-1} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{\sum_{t=2}^{n}\lambda_t^2}{D} & \frac{-(1-\rho)\sum_{t=2}^{n}\lambda_t}{D} & 0 & 0 \\ \frac{-(1-\rho)\sum_{t=2}^{n}\lambda_t}{D} & \frac{(n-1)(1-\rho)^2}{D} & 0 & 0 \\ 0 & 0 & \frac{\sigma^4}{(n-1)} & 0 \\ 0 & 0 & 0 & \frac{(1-\rho^2)}{2(n-1)} \end{bmatrix}^{-1},$$

where

$$D = \frac{(1-\rho)^2}{(1+b_1^2)\sigma^2}\{(n-1)\sum_{t=2}^{n}\lambda_t^2 - (\sum_{t=2}^{n}\lambda_t)^2\}.$$