ABSTRACT

SangPil Hwang. Dynamic Time Series Analysis Using Logistic Function. (Under the direction of David A. Dickey.)

This paper investigates a set of autoregressive time series models whose coefficients have the form of a logistic function. The transfer function type models give additional flexibility over the fixed coefficients models and include them as a special case. NLAR models with the AR(1) coefficient being a hyperbolic tangent function work well for modeling series which have asymmetric stochastic volatility or changing amplitude around 0 with a persistent autocorrelation and locally nonstationary behavior.

Dynamic Time Series Analysis Using Logistic Function

by

SangPil Hwang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

STATISTICS

Raleigh

2004

APPROVED BY:

David A. Dickey Chair of Advisory Comittee	Sastry G. Pantula
Bibhuti B. Bhattacharyya	Matt Holt

To my family

Biography

SangPil Hwang was born in ChangWon, the Republic of Korea on Feburary 1, 1967. He entered Department of Economics at Seoul National University in 1985 and took his B.S. in Economics in 1989. He worked for the Bank of Korea from 1992 to 2000. He joined Department of Statistics, North Carolina State University for the pursuit of Ph.D. degree in May, 2000 and obtained his M.S. in Statistics in May, 2002.

Acknowledgements

I would like to express my deepest gratitude and biggest appreciation to my advisor Dr. David A. Dickey. I am also grateful for precious instruction and helpful suggestions from Dr. Sastry G. Pantula, Dr. Bibhuti B. Bhattacharyya and Dr. Matt Holt. I deeply thank all of them.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Dynamic time series analysis based on nonlinear models has become a topic of interest lately. Here we will analyze time series whose coefficients involve the logistic function. The models studied include a transfer function type model and a nonlinear extension of the autoregressive model.

The usual transfer function has the form

$$\begin{aligned} y_t &= \sum_{i=0}^{\infty} \gamma_i x_{t-i} + z_t \\ &= \gamma(B) x_t + z_t \end{aligned}$$

where $\sum_i |\gamma_i| < \infty$.

We assume the input process x_t and noise process z_t are both stationary and are mutually independent. The coefficients $\gamma_0, \gamma_1, \cdots$ describe the weights assigned to past values of x_t used in predicting y_t and

$$\gamma(B) = \sum_{i=0}^{\infty} \gamma_i B^i$$

where B is the backshift operator $B(x_t) = x_{t-1}$(Shumway and Stoffer, 1999).

The model allows y_t to depend on current and past values of the explanatory variable X. For this standard model, the coefficients are fixed, neither they nor their estimates change with time. We extend this to a model which has flexible weights, reflecting a different impact of explanatory variables at different times.

We first consider a simple nonlinear transfer function model where the weight is expressed using a logistic function. Our model is

$$y_t = \gamma\rho(x_{t-1})x_{t-1} + e_t$$

where $\rho(x_{t-1}) = \frac{1}{\exp(\alpha+\beta x_{t-1})+1}$.

Thus, $\rho(x_{t-1})$ is a function of the explanatory variable X. The model can be extended to, for example,

$$y_t = \gamma_1\rho(x_{t-d})x_{t-1} + \gamma_2 (1 - \rho(x_{t-d}))\, x_{t-2} + e_t$$

where $\rho(x_{t-d}) = \frac{1}{\exp(\alpha+\beta x_{t-d})+1}$ and $d = 1, 2, \cdots$, or

$$\begin{aligned} y_t &= \gamma_1[1 - \rho_1(x_{t-d}) - \rho_2(x_{t-d})]x_t + \gamma_2\rho_1(x_{t-d})x_{t-1} \\ &+ \gamma_3\rho_2(x_{t-d})x_{t-2} + e_t \end{aligned}$$

where $\rho_1(x_{t-d}) = \frac{\exp(\alpha_1+\beta_1 x_{t-d})}{\sum_{i=1}^{2}\exp(\alpha_i+\beta_i x_{t-d})+1}$ and $\rho_2(x_{t-d}) = \frac{\exp(\alpha_2+\beta_2 x_{t-d})}{\sum_{i=1}^{2}\exp(\alpha_i+\beta_i x_{t-d})+1}$. $d = 0, 1, 2, \cdots$.

Also, variants of

$$y_t = \rho(x_{t-d})[\gamma_0 x_t + \gamma_1 x_{t-1} + \cdots + \gamma_k x_{t-k}] + e_t$$

where $\rho(x_{t-d}) = \frac{1}{\exp(\alpha+\beta x_{t-d})+1}$ and $d = 0, 1, 2, \cdots$, are considered.

The idea of the model with two lagged Xs is to allow the relative importance of x_{t-1} and x_{t-2} to vary with the magnitude of X. For example, it may be that when X is small, the coefficient of x_{t-2} receives relatively little weight as compared to x_{t-1}.

This model is a simple variant of smooth transition regression(STR) model introduced in the literature(Bacon and Watts, 1971; Granger and Teräsvirta, 1993).

$$\begin{aligned} y_t &= (c_{11} + \phi_{10}x_t + \phi_{11}x_{t-1} + \cdots + \phi_{1p}x_{t-p}) \\ &+ (c_{12} + \phi_{20}x_t + \phi_{21}x_{t-1} + \cdots + \phi_{2p}x_{t-p})F(x_{t-d}) + e_t \end{aligned}$$

where $F(x_{t-d})$ is a continuous function which may be either even or odd and $d = 0, 1, 2, \cdots$. For example, $F(x_{t-d})$ may be a cumulative distribution function such as a $N(\mu, \sigma^2)$, or an odd, monontonically increasing function with $F(-\infty) = 0$ and $F(\infty) = 1$. Where $F(x_{t-d})$ changes very sharply, it could be viewed as a threshold model with 2 regimes.

Here, we focus on the changing behavior of the coefficients rather than the regime shifting. In this way, the nonlinear estimation could be more efficient and the model is easily extended to three or more weight models. The estimation with more weights is implemented easily as well.

The model has worked well in fitting to a string of log transformed daily flows for the Neuse River in North Carolina in which Y is downstream flow and X is a flow at an upstream location. In a period of high upstream flow, water would move downstream faster. On the other hand, the water upstream could take longer to clear out by virtue of the high volume. Our model allows the data to inform us on this issue.

Another goal of this paper is to investigate the possibility of allowing the second moment properties of a univariate time series to change dynamically by using constant variance innovations but dynamically changing difference equation coefficients. This also can be accomplished through the use of the logistic function. This gives somewhat different dynamic effects than the well known ARCH models in that the conditional innovation variance changes in ARCH models.

ARIMA models are the most common time series models for data fitting and forecasting. In these models y_t represents the observation at time t, μ represents the long term mean and e_t represents the innovation, that is, e_t is an uncorrelated mean 0, variance σ^2 sequence consisting of the portion of y_t that can not be forecast from the past. The general model is

$$y_t - \mu - \phi_1(y_{t-1} - \mu) - \cdots - \phi_{p+d}(y_{t-p-d} - \mu) = e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$$

where $\phi_1, \cdots, \phi_{p+d}, \theta_1, \cdots, \theta_q$ are fixed but unknown parameters. They are not random. The behavior of the data generated from this model is dependent on the roots of the "characteristic equation" $m^{p+d} - \phi_1 m^{p+d-1} - \cdots - \phi_{p+d} = 0$. It is assumed that d of these roots are 1 and that the remaining p roots are less than 1 in magnitude. If $d = 0$, the model is called ARMA and has the property that a convergent weighted sum of the e_t sequence exists which, when used as $y_t - \mu$, solves the difference equation above.

Furthermore, if $d = 0$, estimates of the parameters based on linear (if $q = 0$) or nonlinear least square estimates have asymptotic multivariate normal distributions.

If $d = 0$, the series is typically referred to as "stationary" meaning that $y_t - \mu$ has mean 0 and lag j covariance that is a function of j only. If d and q are both 0, the model is autoregressive of order p,AR(p). Low order autoregressive models are often used for modeling.

Stationary models have autocorrelations that are bounded by an exponentially decaying function of the lag number. This quickly decaying correlation seems inconsistent with many observed time series. The traditional method of dealing with this has been to difference the data at least once, and then fit an ARMA model to the differences, that is d is taken to be 1 or more. In the class of ARIMA(p,d,q) models, one linear difference equation with fixed parameters is assumed to govern the behavior of the series at all times.

Over the past 20 years, various models that allow more flexibility have been introduced. They can be divided into nonparametric, semiparametric, and parametric groups in general. In the parametric group, where a specific functional form is assumed, usually with some parameters to be estimated, we have

(i) the polynomial model, for example, the quadratic

$$y_t = \delta + \beta' Z_t + Z_t' C Z_t + e_t$$

where C is a symmetric matrix of parameters. Z_t consists of independent explanatory variables or lagged variables,

(ii) the smooth transition regression(STR) model

$$y_t = \beta_1' Z_t + F(Z_t)\beta_2' Z_t + e_t$$

where F is a function for capturing the transition aspect of the model such as a normal CDF or a logistic function,

(iii) the flexible Fourier form

$$y_t = \delta + \beta' Z_t + Z_t' C Z_t + \sum_{j=1}^{q} \{c_j \sin(j(\gamma' Z_t)) + d_j \cos(j(\gamma' Z_t))\} + e_t$$

which is the polynomial model with sine and cosine terms added,

(iv) neural networks

$$y_t = \alpha + \sum_{j=1}^{q} \beta_j \phi(\gamma_j' Z_t) + e_t$$

where ϕ is a squashing function, such as a cumulative distribution function or a logistic function and Z_t consists of independent explanatory variables or lagged variables.

(Granger and Teräsvirta, 1993).

In the area of the nonlinear autoregressive(NLAR) models, amplitude-dependent exponential autoregressive(EXPAR) models were independently introduced by Jones (1976) and Ozaki and Oda(1978), and have become widely known. The EXPAR model is

$$y_t = \sum_{j=1}^{q} [\alpha_j + \beta_j \exp(-\delta y_{t-1}^2)] y_{t-j} + e_t$$

where $\delta > 0$.

Also, the autoregressive conditionally heteroscedastic(ARCH) model of Engle(1982) and subsequent variants GARCH, EGARCH etc, are very popular. These models allow the variance to change in a dynamic way by letting the variance at time t satisfy a difference equation whose innovations are squared residuals.

In this paper, we begin by proposing a minor adjustment to the AR(1) model. This adjustment appears to provide quite a bit of flexibility in terms of the types of data structure it can provide.

The usual autoregressive order 1 model with mean 0 satisfies

$$y_t = \rho y_{t-1} + e_t$$

where ρ and σ^2, the variance of e_t, are constant.

Our proposed model retains the constant innovations variance and uses dynamic coefficients to model changing variances in the observations. The modification is to replace ρ by a modified logistic function of past Y values, namely, the random variable $\gamma\rho(y_{t-1})$.

$$y_t = \gamma\rho(y_{t-1})y_{t-1} + e_t$$

where $\rho(y_{t-1}) = \frac{\exp(\alpha+\beta f(y_{t-1}))-1}{\exp(\alpha+\beta f(y_{t-1}))+1}$. Here $f(y) = |y|$ or $f(y) = y$, $|\gamma| \leq 1$, and $\beta > 0$.

The modified logistic function we are using is also called a hyperbolic tangent.

$$\tanh z = \frac{\sinh z}{\cosh z} = \frac{\exp(2z) - 1}{\exp(2z) + 1}$$

where $z = \frac{1}{2}(\alpha + \beta f(y))$. The range of $\tanh z$ is (-1,1).

Notice that $|\gamma\rho(t)| < |\gamma|$. So, the model can produce local autocorrelation coefficients quite close to ± 1 if $|\gamma|$ is near 1. This allows the model to generate data which is locally nonstationary in appearence but in the long term tends to be mean reverting. The model is useful for explaining series which have asymmetric stochastic volatility or changing amplitude around 0 with a more or less persistent autocorrelation rather than an exponential decay. A model where $\rho(y_{t-1})$ has the form $\frac{\exp(\alpha+\beta f(y_{t-1}))-1}{\exp(\alpha+\beta f(y_{t-1}))+1}$ gives more interesting features such as mentioned previously, with 2 or more regimes, than does a model using the form $\frac{1}{\exp(\alpha+\beta f(y_{t-1}))+1}$.

The model can be considered as a variant of the logistic smooth threshold autoregressive(LSTAR) model such that

$$\begin{aligned} y_t &= \gamma\rho(y_{t-1})y_{t-1} + e_t \\ &= \gamma\frac{\exp(\alpha+\beta f(y_{t-1}))-1}{\exp(\alpha+\beta f(y_{t-1}))+1}y_{t-1} + e_t \\ &= \gamma y_{t-1} - 2\gamma y_{t-1}\frac{1}{\exp(\alpha+\beta f(y_{t-1}))+1} + e_t. \end{aligned}$$

A gradual transition between the different regimes is obtained by a smoothly changing logistic function, which changes from 0 to 1 depending on $f(y_{t-1})$. This can also be thought of as a very simple form of the single neural network model mentioned previously.

The STAR model has been used a lot to analysis regime-switching behavior of series. The STAR model for a univariate y_t, which is observed at $t = 1-p, 1-(p-1), \cdots -1, 0, 1, \cdots, T-1, T$, is given by

$$y_t = \phi_1' x_t[1 - G(s_t;\gamma,c)] + \phi_2' x_t G(s_t;\gamma,c) + \epsilon_t$$

where $x_t = (1, \tilde{x}_t')'$ with $\tilde{x}_t = (y_{t-1}, \cdots, y_{t-p})'$ and $\phi_i = (\phi_{i,0}, \phi_{i,1}, \cdots \phi_{i,p})', i = 1, 2.$ $t = 1, \cdots, T$. The model allows exogenous variables $z_{1t}, \cdots, z_{kt}$ as additional regressors. The ϵ_t's are assumed to be a martingale difference sequence with $E[\epsilon_t|\Omega_{t-1}] = 0$ and $E[\epsilon_t^2|\Omega_{t-1}] = \sigma^2 . \Omega_{t-1} = \{y_{t-1}, y_{t-2}, \cdots, y_{1-p-1}, y_{1-p}\}$.

The transition function $G(s_t;\gamma,c)$ is a continuous function that is bounded between 0 and 1. Usually, the logistic function

$$G(s_t;\gamma,c) = (1+\exp\{-\gamma(s_t-c)\})^{-1}$$

where $\gamma > 0$ and the exponential function

$$G(s_t;\gamma,c) = 1-\exp\{-\gamma(s_t-c)^2\}$$

where $\gamma > 0$ are used. s_t could be a lagged endogenous variable, an exogenous variable, a time trend or a function of them. The resultant models are called the logistic STAR(LSTAR) and the exponential STAR(ESTAR) model respectively.

The conditions under which STAR models generate series that are stationary are not well known(Chan and Tong, 1986; Tong, 1990; Franses and van Dijk, 2000). The stationarity and ergodicity of the series are generally pre-assumed. Testing unit roots in TAR models has been discussed in Enders and Granger(1998). Tong(1990) has proved that the nonlinear least square(NLS) estimates of the stationary and ergodic LSTAR model are consistent and asymptotically normal. Specification, estimation, and evaluation of STAR models are introduced in detail in Teräsvirta(1994) and Eitrheim and Teräsvirta(1996).

Recently, various extensions of the basic STAR model have been suggested. The multiple STAR(MRSTAR) model is obtained by encapsulating two different two regime STAR models(van Dijk and Franses, 1999).

$$\begin{aligned} y_t &= [\phi_1' x_t(1-G_1(s_{1t};\gamma_1,c_1)+\phi_2' x_t G_1(s_{1t};\gamma_1,c_1)][1-G_2(s_{2t};\gamma_2,c_2)] \\ &+ [\phi_3' x_t(1-G_1(s_{1t};\gamma_1,c_1)+\phi_4' x_t G_1(s_{1t};\gamma_1,c_1)]G_2(s_{2t};\gamma_2,c_2)+\epsilon_t \end{aligned}$$

The model allows for a maximum of four different regimes. It can be extended to contain 2^m regimes with $m > 2$.

The flexible coefficient STAR model(Medeiros and Veiga, 2000) can be derived from the MRSTAR. It is obtained by assuming the transition variables s_{1t} and s_{2t} are linear combinations of lagged dependent variables, i.e., $s_{it} = \alpha_t' \tilde{x}_t, i = 1, 2$, and imposing the restriction $\phi_{1,j} - \phi_{2,j} - \phi_{3,j} + \phi_{4,j} = 0$ for $j = 1, \cdots, p$. The model can be rewritten as

$$y_t = \phi_0^{*\prime} x_t + \phi_1^{*\prime} x_t G_1(\alpha_1' \tilde{x}_t; \gamma_1, c_1) + \phi_2^{*\prime} x_t G_2(\alpha_2' \tilde{x}_t; \gamma_2, c_2) + \epsilon_t$$

where $\phi_0^* = \phi_1, \phi_1^* = \phi_2 - \phi_1$ and $\phi_2^* = \phi_3 - \phi_1$(van Dijk, Ter$\ddot{a}$svirta, and Franses, 2002).

If one of the transition variables is taken to be time t, MRSTAR leads to time-varing STAR(TVSTAR) model. The model is useful for analyzing the time series which display both nonlinearity and structural instability(Franses and van Dijk, 2000; van Dijk, Ter$\ddot{a}$svirta, and Franses, 2002).

In addition, the fractionally integrated STAR(FISTAR) model which combines the features of long memory and nonlinearity into a single model has been suggested(van Dijk and Franses, 2000) and STAR-GARCH model where the errors have GARCH structure have been used for forecasting(Chan and McAleer, 2002).

We pay attention to the first order LSTAR model

$$y_t = \phi_2 y_{t-1} + (\phi_1 - \phi_2) y_{t-1} \frac{1}{\exp(\alpha + \beta f(y_{t-1}) + 1)} + e_t$$

where $\phi_1 = -\gamma$ and $\phi_2 = \gamma$ with $|\gamma| < 1$. A series with a strong persistent auto-correlation like a long memory process is generated when $|\gamma|$ is near 1 and we prove ergodicity of the series, and consistency and asymptotic distributions of parameter estimates for this model.

We can incorporate serially correlated errors into the model easily and still get the properties above.

$$y_t = \gamma\rho(y_{t-1})y_{t-1} + \eta_t,$$

and

$$\eta_t = \delta_1\eta_{t-1} + \cdots + \delta_k\eta_{t-k} + e_t.$$

NLAR(1) with serially correlated errors can be displayed as a LSTAR with many regimes.

Usually, for threshold models, it is not easy to estimate many regimes at once including the identification of order p and the delay factor d. The parameters of our suggested model can be easily estimated by using the Gauss-Newton algorithm, and one-step-ahead prediction is easily obtained from the fitted model. There is a possibility that the usual STAR fitting process, because of its generality, does not work well for the data generated by our suggested NLAR models(Granger and Teräsvirta, 1993; Teräsvirta,1994).

Estimation problems(non-convergence) are fairly common for some parameter settings. However, we find that setting γ set to a value near, but less than 1 gives good convergence and provides good prediction mean square errors, even if the true γ is not near 1. It will be shown through simulation that the fitting is fairly robust with respect to the assumed γ.

Finally, we deal with the model with $\gamma = 1$.

$$y_t = \rho(y_{t-1})y_{t-1} + e_t$$

where $\rho(y_{t-1}) = \frac{\exp(\alpha+\beta f(y_{t-1}))-1}{\exp(\alpha+\beta f(y_{t-1}))+1}$. When the parameter space of $|\gamma|$ is not restricted to $|\gamma| < 1$, e.g., $\gamma = 1$, it is not easy to obtain the theoretical distribution of parameter estimates. In fact, it appears that no single distribution applies, even for large samples, across the full range of possible (α, β) values. The Monte Carlo study suggests that similar distributional results to those of parameter estimates with $|\gamma|$ near but less than 1 are obtained using only those series which reject a unit root, and we find a region in which the normal approximation works reasonably well. The region is analogous to the stationarity region in standard ARIMA models in which standard behavior of estimators seems to hold. This extension to the case $\gamma = 1$ distinguishes our investigation from other nonlinear approaches of which we are aware.

Chapter 2

Transfer function type models

2.1 some nonlinear models

We will study some time series models of the transfer function type here. A simple case to begin with is

$$y_t = \gamma\rho(x_{t-1})x_{t-1} + e_t \tag{2.1}$$

where $\rho(x_{t-1}) = \frac{1}{\exp(\alpha+\beta x_{t-1})+1}$. In this model, x_t could be a sequence of fixed values or a random process. The e_t are independent draws from a $N(0, \sigma^2)$ distribution and are independent of X. Here, γ is a scale adjustment. An intercept could be added and x_{t-1} replaced by $(x_{t-1} - \mu)$. Notice that $\rho(x_{t-1})$ depends on the explanatory variable X(see Figure 2.1 and 2.2). This model can be estimated by nonlinear least squares.

Let an estimator of the unknown θ_0 be the θ that minimizes $Q_n(\theta)$, where $Q_n(\theta)$

is a function of the observations and θ. The least squares estimator of θ_0 is the θ that minimizes

$$Q_{nl}(\theta) = \frac{1}{n}\sum_{t=1}^{n}(y_t - f_t(\theta))^2$$

where $f_t(\theta) = \gamma\rho(x_{t-1})x_{t-1}$. Thus, at $\theta = \theta_0$

$$\frac{1}{n}\sum_{t=1}^{n}2(y_t - f_t(\theta))f_t'(\theta) = 0.$$

The function $Q_n(\theta)$, up to a scalar constant, is the negative of the logarithm of the likelihood in the case of maximum likelihood estimation for the model with normal independent e_t.

The Gauss-Newton algorithm gives a one step adjustment based on an initial estimate $\hat{\theta}_a$.

$$\hat{\theta}_{(a+1)} = \hat{\theta}_a + (F_{nk}'(\hat{\theta}_a)F_{nk}(\hat{\theta}_a))^{-1}F_{nk}'(\hat{\theta}_a)\hat{e}_a$$

where $F_t(\theta)$ is the k dimensional vector of first derivatives of $f_t(\theta)$,$t = 1, 2, \cdots, n$, and

$$F_{nk}(\theta) = [F_1'(\theta), F_2'(\theta), \cdots, F_n'(\theta)]'$$

is the $n \times k$ matrix of first derivatives. $\hat{e}_a = \hat{e}_a(\hat{\theta}_a)$ is a vector of residuals using current estimates.

For $\theta' = (\gamma, \alpha, \beta)$, $F_{nk}(\theta)$ will have the form

$$\begin{pmatrix} \rho(x_0)x_0 & -\gamma\rho(x_0)(1-\rho(x_0))x_0 & -\gamma\rho(x_0)(1-\rho(x_0))x_0^2 \\ \rho(x_1)x_1 & -\gamma\rho(x_1)(1-\rho(x_1))x_1 & -\gamma\rho(x_1)(1-\rho(x_1))x_1^2 \\ \rho(x_2)x_2 & -\gamma\rho(x_2)(1-\rho(x_2))x_2 & -\gamma\rho(x_2)(1-\rho(x_2))x_2^2 \\ \vdots & \vdots & \vdots \\ \rho(x_{n-2})x_{n-2} & -\gamma\rho(x_{n-2})(1-\rho(x_{n-2}))x_{n-2} & -\gamma\rho(x_{n-2})(1-\rho(x_{n-2}))x_{n-2}^2 \\ \rho(x_{n-1})x_{n-1} & -\gamma\rho(x_{n-1})(1-\rho(x_{n-1}))x_{n-1} & -\gamma\rho(x_{n-1})(1-\rho(x_{n-1}))x_{n-1}^2 \end{pmatrix}.$$

So, $F'_{nk}(\theta)F_{nk}(\theta)$ will be

$$\sum_{t=1}^{n}\begin{pmatrix} \rho^2(x_{t-1})x_{t-1}^2 & \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^2 & \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^3 \\ \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^2 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^2 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^3 \\ \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^3 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^3 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^4 \end{pmatrix}.$$

We need $F_{nk}(\theta)$ to be nonsingular everywhere in some neighborhood of the true parameters. Obviously, we have to exclude the possibility of $\gamma = 0$. If $\beta = 0$, $F_{nk}(\theta)$ is singular and it violates the rank qualification(Gallant, 1986). Thus, in practice,a failure of the estimates to converge could be caused by γ or β being 0. When β is 0, $\rho(x_{t-1})$ is constant and it becomes impossible to break the product of constants $\gamma\rho(x_{t-1})$ into meaningful components based on observed data. Column 2 of $F_{nk}(\theta)$ becomes a constant multiple of column 1. Of course, we must also assume that X takes on enough values so that X and X^2 are not linearly dependent. Note also that, as a practical matter, the logistic function, for certain α and β, and range of X, can be almost flat(i.e., constant) so that an analyst should always hold the constant coefficient model as a possible model when non-convergence is encountered. Threshold models with 2 regimes could be considered as competitors when β approaches infinity.

When $\beta \neq 0$, the matrix will satisfy the rank qualification. Thus, there will be $\frac{1}{n}F'_{nk}(\theta)F_{nk}(\theta)$ and a limit matrix provided $\frac{1}{n}\sum_{t=1}^{n} x_{t-1}^j$ converges for $j = 2, 3, 4$. Making this assumption, we write the limit as

$$B(\theta_0) = \lim_{n\to\infty}\frac{1}{n}F'_{nk}(\theta_0)F_{nk}(\theta_0)$$

and assume that $B(\theta_0)$ is nonsingular where θ_0 is the true value of the parameter θ.

Thus, we are assuming $\beta_0 \neq 0$.

Also, in our model, we have only one local minimum at $\theta = \theta_0$. This is called identification condition in Gallant(1986). For the uniqueness of θ_0, we show

$$S(\theta) = \lim_{n \to \infty} \frac{1}{n} \sum (f_t(\theta) - f_t(\theta_0))^2$$

has a unique minimum at $\theta = \theta_0$.

Note

$$\frac{1}{\exp(\alpha + \beta x) + 1} = \frac{1}{\exp(\alpha_0 + \beta_0 x) + 1}.$$

only if

$$\exp(\alpha + \beta x) + 1 = \exp(\alpha_0 + \beta_0 x) + 1.$$

Thus, if $\beta = \beta_0$, α must equal α_0. If $\beta \neq \beta_0$, then x must be $\frac{\alpha - \alpha_1}{\beta - \beta_1}$, that is, the curves cross at only one x. So, if $\gamma = \gamma_0$, the curves,

$$f(x) = \gamma \frac{1}{\exp(\alpha + \beta x) + 1}$$

and

$$f_0(x) = \gamma_0 \frac{1}{\exp(\alpha_0 + \beta_0 x) + 1}$$

cross at only one x, that is, the curves are equal only if $\alpha = \alpha_0$ and $\beta = \beta_0$. Note if $\gamma = \gamma_0 = 0$, then they are equal even if $\alpha \neq \alpha_0$, etc.

Setting $f(x) = f_0(x)$, we have, assuming $\gamma_0 \neq 0$,

$$\frac{\gamma}{\gamma_0} = \frac{\exp(\alpha + \beta x) + 1}{\exp(\alpha_0 + \beta_0 x) + 1}.$$

On the left is a constant. On the right is a function of x which can only be constant if $\alpha = \alpha_0$, and $\beta = \beta_0$. Thus, we must have $\alpha = \alpha_0$, $\beta = \beta_0$, in which case the right is the constant 1 and $\gamma = \gamma_0$.

In conclusion, under all combinations of $\theta = (\gamma, \alpha, \beta)$ that have $\beta \neq 0$ and $\gamma \neq 0$, $S(\theta) = 0$ is obtained only when $\alpha = \alpha_0, \gamma = \gamma_0$ and $\beta = \beta_0$. Hence a unique minimum exists at $\theta = \theta_0$.

To begin with, we suppose x_t to be a sequence of fixed values or conditionally fixed as in classical theory. Under the conditional approach, x_t is ancillary in the sense that the joint distribution of x_t does not depend on model parameters. In the case of fixed Xs, it is reasonable to assume that $|x_t| < c < \infty$.

The consistency of $\hat{\theta}_n$ is obtained using theorem 6 in Jennrich(1969). With x_t fixed and e_t from an IID $(0, \sigma^2)$, if a unique minimum exists at $\theta = \theta_0$, then $\hat{\theta}_n$ is a strongly consistent estimator of θ_0.

We quote theorem 6 from Jennrich(1969) in our notation.

Assume

(i) A sequence of real valued responses y_t has the structure

$$y_t = f_t(\theta_0) + e_t$$

where the f_t are known continuous functions, and θ_0 is in a compact subset Θ of a Euclidean space and the e_t are IID $(0, \sigma^2)$.

(ii) The tail cross product $[f, f]$ of $f = (f_t)$ with itself exists and

$$\int (f(\theta) - f(\theta_0))^2 dF(x)$$

has a unique minimum at $\theta = \theta_0$. We define $[f, f]$ and x below.

Let $\hat{\theta}_n$ be a sequence of least squares estimators. Under assumptions (i) and (ii), $\hat{\theta}_n$ and $\hat{\sigma^2}_n$ are strongly consistent sequences of estimators for θ_0 and σ^2(Jennrich, 1969).

For the assumptions, let $x = (x_t)$ and $y = (y_t)$ be two sequences of real numbers and let $(x, y)_n = n^{-1} \sum_{t=1}^{n} x_t y_t$. If $(x, y)_n$ converges to a real number, its limit (x, y) will be called the tail product of x and y. Let g and h be two sequence valued functions on Θ. If $(g(\alpha), h(\beta))_n \to (g(\alpha), h(\beta))$ uniformly for all α and β in Θ, let $[g, h]$ denote the function on $\Theta \times \Theta$ which takes $< \alpha, \beta >$ into $(g(\alpha), h(\beta))$. This function will be called the tail cross product of g and h (Jennrich, 1969).

Jennrich(1969) proved if $\tilde{g}$ and $\tilde{h}$ are bounded and continuous functions on $X \times \Theta$, $g_t = \tilde{g}(x_t, \theta)$ and $h_t = \tilde{h}(x_t, \theta)$, where $x_1, x_2, \cdots$ is a sequence of vectors in X whose sample distribution function F_n approaches a distribution function F completely, then

$$(g(\alpha), h(\beta))_n \to \int \tilde{g}(x, a)\tilde{h}(x, a) dF(x)$$

uniformly for all α and β in Θ. Hence the tail cross product $[g, h]$ exists.

In our case, f is a bounded and continuous function on $X \times \Theta$ where x_t is a bounded sequence of real numbers and $|\gamma| < M < \infty$.

$$|f_t(\theta)| = |\gamma \rho(x_{t-1}) x_{t-1}| < |\gamma x_{t-1}| \leq |\gamma||x_{t-1}| < \infty.$$

We assume the sample distribution function $F_n(x)$ approaches a distribution function $F(x)$ completely. Thus the tail cross product $[f, f]$ exists. With the identification condition shown before, the assumptions in Jennrich's theorem are all satisfied. Hence, $\hat{\theta}_n$ is a strongly consistent sequence of estimators for θ_0.

For the asymptotic distribution of $\hat{\theta}_n$, theorem 7 in Jennrich(1969) could be used.

In addition to assumption (i) and (ii) from theorem 6, assume

(iii) For $\theta = (\theta_1, \cdots, \theta_k)$, the derivatives $[f_t^{(1)}(\theta), \cdots, f_t^{(k)}(\theta), f_t^{(11)}(\theta), \cdots, f_t^{(1k)}(\theta), f_t^{(k1)}(\theta), \cdots, f_t^{(kk)}(\theta)]$ exist and are continuous on Θ and that all tail cross products of the form $[g, h]$ where $g, h = f, f^{(j)}, f^{(jr)}$, exist.

(iv) The true parameter vector θ_0 is an interior point of Θ and the matrix $B(\theta_0)$ is nonsingular, where $B(\theta_0) = \lim_{n\to\infty} \frac{1}{n} F'_{nk}(\theta_0) F_{nk}(\theta_0)$.

Then, under assumption (i),(ii),(iii) and (iv),

$$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N[0, B^{-1}(\theta_0)\sigma^2]$$

(Jennrich, 1969).

We have already shown assumptions (i) and (ii) are satisfied.

With $L_t = \alpha + \beta x_{t-1}$, we have

$$\begin{aligned}
\frac{\partial f_t(\theta)}{\partial \gamma} &= \frac{1}{\exp(L_t) + 1} x_{t-1}, \\
\frac{\partial f_t(\theta)}{\partial \alpha} &= -\gamma \frac{\exp(L_t)}{(\exp(L_t) + 1)^2} x_{t-1}, \\
\frac{\partial f_t(\theta)}{\partial \alpha^2} &= -\gamma \frac{\exp(L_t)(1 - \exp(L_t))}{(\exp(L_t) + 1)^3} x_{t-1},
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial f_t(\theta)}{\partial \beta} &= -\gamma \frac{\exp(L_t)}{(\exp(L_t)+1)^2} x_{t-1}^2 \\
\frac{\partial f_t(\theta)}{\partial \beta^2} &= -\gamma \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} x_{t-1}^3, \\
\frac{\partial f_t(\theta)}{\partial \gamma \partial \alpha} &= -\frac{\exp(L_t)}{(\exp(L_t)+1)^2} x_{t-1}, \\
\frac{\partial f_t(\theta)}{\partial \gamma \partial \beta} &= -\frac{\exp(L_t)}{(\exp(L_t)+1)^2} x_{t-1}^2, \\
\frac{\partial f_t(\theta)}{\partial \alpha \partial \beta} &= -\gamma \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} x_{t-1}^2
\end{aligned}
$$

where the derivatives are not 0.

All derivatives are bounded and continuous functions where x_t is a bounded sequence of real numbers and $|\gamma| < M < \infty$. Thus all tail cross products exist and assumption (iii) is satisfied. Assumption (iii) is needed to get asymptotic normality of $\hat{\theta}_n$ using the central limit theorem. For assumption (iv), our conditions $\gamma \neq 0$ and $\beta \neq 0$ ensure the existence of a nonsingular $B(\theta_0)$. Hence, according to theorem 7 in Jennrich, the sequence of least squares estimators will be asymptotically normally distributed.

Fuller(1996) showed one-step Gauss-Newton estimator has a limiting normal distribution. We introduce theorem 5.5.4 in Fuller(1996) in our notation.

We write the model as

$$y_t = f_t(\theta_0) + e_t$$

where the e_t are IID $(0, \sigma^2)$ random variables or are $(0, \sigma^2)$ martingale differences.

For the model, the vector sequence $\{[F_t(\theta_0)e_t, e_t]\}$ satisfies

$$E\{[F_t(\theta_0)e_t, e_t, e_t^2] | A_{t-1}\} = [\mathbf{0}, 0, \sigma^2],$$

$$E\{|[F_t(\theta_0)e_t, e_t]|^{2+\nu}|A_{t-1}\} < M_F < \infty,$$

$$E\{[F_t'(\theta_0)F_t(\theta_0)e_t^2]|A_{t-1}\} = F_t'(\theta_0)F_t(\theta_0)\sigma^2$$

a.s., for all t, where $\delta > 0$ and M_F is a positive constant.

A_{t-1} is the σ-field generated by

$$\{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_t}, e_1, e_2, ..., e_{t-1}\}.$$

$\mathbf{x_t}$ could be fixed or random.

Let $\tilde{\theta}$ be the one-step Gauss-Newton estimator of θ_0 given by

$$\tilde{\theta} = \hat{\theta} + (F_{nk}'(\hat{\theta})F_{nk}(\hat{\theta}))^{-1}F_{nk}'(\hat{\theta})\hat{e}.$$

Assume

(i) There is an open set S such that S is in $\Theta, \theta_0 \in S$, and

$$p\lim_{n\to\infty} \frac{1}{n} F'(\theta)F(\theta) = B(\theta)$$

is nonsingular for all θ in S, where $F(\theta)$ is the $n \times k$ matrix with tj-th element given by $f_t^{(j)}(\theta)$.

(ii) $p\lim_{n\to\infty} \frac{1}{n} G'(\theta)G(\theta) = L(\theta)$ uniformly in θ on the closure $\bar{S}$ of S, where the elements of $L(\theta)$ are continuous functions of θ on $\bar{S}$, and $G(\theta)$ is an $n \times (1 + k + k^2 + k^3)$ matrix with t-th row given by

$$\Big[f_t(\theta), f_t^{(1)}(\theta), \cdots, f_t^{(k)}(\theta), f_t^{(11)}(\theta), \cdots, f_t^{(1k)}(\theta),$$
$$f_t^{(k1)}(\theta), \cdots, f_t^{(kk)}(\theta), f_t^{(111)}(\theta), \cdots, f_t^{(kkk)}(\theta)\Big]$$

where $f_t^{(j)}(\theta), f_t^{(jr)}(\theta), f_t^{(jrs)}(\theta)$ denote the first, second and third partial derivative of $f_t(\theta)$ with respect to j, jr, jrs-th element of θ.

(iii) The initial estimator of θ_0, say, $\hat{\theta}$, satisfies $(\hat{\theta}-\theta_0) = O_p(a_n)$, where $\lim_{n\to\infty} a_n = 0$.

(iv) $n^{-1/2}\sum_{t=1}^{n} F_t'(\theta_0)e_t \xrightarrow{d} N[0, B(\theta_0)\sigma^2]$ and $a_n^2 = o(n^{-1/2})$.

Then

$$\tilde{\theta} - \theta_0 = (F'(\theta_0)F(\theta_0))^{-1}F'(\theta_0)e + O_p\{max(a_n^2, a_n n^{-1/2})\}$$

and

$$n^{1/2}(\tilde{\theta} - \theta_0) \xrightarrow{d} N[0, B^{-1}(\theta_0)\sigma^2].$$

(Fuller, 1996).

We know $\hat{\theta}_n$ is a strongly consistent estimator of θ_0. Now, we assume the initial estimator to be consistent where $(\hat{\theta} - \theta_0) = O_p(n^{-1/2})$(condition (iii)). An initial consistent estimator can be obtained through a random search of n values of θ for the one θ_n^* which minimizes $Q_n(\theta)$(Jennrich, 1969).

In our case, $f_t(\theta) = \gamma\rho(x_{t-1})x_{t-1}$ is continuously differentiable with respect to θ, and $F_{nk}'(\theta)F_{nk}(\theta)$ is nonsingular except for the case $\gamma = 0$ or $\beta = 0$. We assume the true θ, θ_0 has $\gamma \neq 0$ and $\beta \neq 0$ as well.

As before, we suppose x_t to be a sequence of fixed values or conditionally fixed as in classical theory. Then, all sample variation enters only via the random variables $\{e_1, e_2, \cdots, e_{t-1}\}$.

Under this assumption, $F_t(\theta_0)$ has (conditionally) all fixed values and the conditions for the model are satisfied with e_t from a $N(0, \sigma^2)$ distribution. $E(|e_t|^{2+\nu}) < M < \infty$ is enough for e_t.

Specifically,

$$\begin{aligned} F_t'(\theta_0) &= [\rho_0(x_{t-1})x_{t-1}, -\gamma_0\rho_0(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}, \\ & \quad -\gamma_0\rho_0(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}^2] \end{aligned}$$

where $\rho_0(x_{t-1}) = \frac{1}{\exp(\alpha_0+\beta_0 x_{t-1})+1}$.

Thus, under given A_{t-1},

$$\begin{aligned} E\{[F_t(\theta_0)e_t, e_t, e_t^2]|A_{t-1}\} &= [\mathbf{0}, 0, \sigma^2]. \\ E\{|[F_t(\theta_0)e_t, e_t]|^{2+\nu}|A_{t-1}\} &= E\{[\rho_0^2(x_{t-1})x_{t-1}^2 \\ &+ \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2 x_{t-1}^2 \\ &+ \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2 x_{t-1}^4 \\ &+ 1]^{(2+\nu)/2}[e_t^2]^{(2+\nu)/2}|A_{t-1}\} \\ &= [\rho_0^2(x_{t-1})x_{t-1}^2 + \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2 x_{t-1}^2 \\ &+ \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2 x_{t-1}^4 + 1]^{(2+\nu)/2} \\ & \quad E\{[e_t^2]^{(2+\nu)/2}|A_{t-1}\}. \end{aligned}$$

With $E(|e_t|^{2+\nu}) < M < \infty$, $E\{|[F_t(\theta_0)e_t, e_t]|^{2+\nu}|A_{t-1}\} < M_F < \infty$.

Now, for condition (i), $p\lim_{n\to\infty} \frac{1}{n}F'(\theta)F(\theta) = B(\theta) =$

$$\frac{1}{n}\sum \begin{pmatrix} \rho_0^2(x_{t-1})x_{t-1}^2 & \gamma_0\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}^2 & \gamma_0\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}^3 \\ \gamma_0\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}^2 & \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2 x_{t-1}^2 & \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2 x_{t-1}^3 \\ \gamma_0\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}^3 & \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2 x_{t-1}^3 & \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2 x_{t-1}^4 \end{pmatrix}$$

where $\rho_0(x_{t-1}) = \frac{1}{\exp(\alpha_0+\beta_0 x_{t-1})+1}$. In the case of fixed Xs, it is reasonable to assume that $|x_t| < c < \infty$, and that $n^{-1}\sum x_t^j, j = 2,3,4$, converges to some appropriate nonzero constant.

Condition (ii) is also satisfied with X fixed.

With $L_t = \alpha + \beta x_{t-1}$, we have

$$
\begin{aligned}
\frac{\partial f_t(\theta)}{\partial \gamma} &= \frac{1}{\exp(L_t)+1} x_{t-1}, \\
\frac{\partial f_t(\theta)}{\partial \alpha} &= -\gamma \frac{\exp(L_t)}{(\exp(L_t)+1)^2} x_{t-1}, \\
\frac{\partial f_t(\theta)}{\partial \alpha^2} &= -\gamma \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} x_{t-1}, \\
\frac{\partial f_t(\theta)}{\partial \alpha^3} &= -\gamma \frac{C\exp(L_t)}{(\exp(L_t)+1)^4} x_{t-1}, \\
\frac{\partial f_t(\theta)}{\partial \beta} &= -\gamma \frac{\exp(L_t)}{(\exp(L_t)+1)^2} x_{t-1}^2 \\
\frac{\partial f_t(\theta)}{\partial \beta^2} &= -\gamma \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} x_{t-1}^3, \\
\frac{\partial f_t(\theta)}{\partial \beta^3} &= 2\gamma \frac{C\exp(L_t)}{(\exp(L_t)+1)^4} x_{t-1}^4, \\
\frac{\partial f_t(\theta)}{\partial \gamma \partial \alpha} &= -\frac{\exp(L_t)}{(\exp(L_t)+1)^2} x_{t-1}, \\
\frac{\partial f_t(\theta)}{\partial \gamma \partial \beta} &= -\frac{\exp(L_t)}{(\exp(L_t)+1)^2} x_{t-1}^2, \\
\frac{\partial f_t(\theta)}{\partial \alpha \partial \beta} &= -\gamma \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} x_{t-1}^2, \\
\frac{\partial f_t(\theta)}{\partial \alpha^2 \partial \gamma} &= -\frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} x_{t-1}, \\
\frac{\partial f_t(\theta)}{\partial \alpha^2 \partial \beta} &= -\gamma \frac{C\exp(L_t)}{(\exp(L_t)+1)^4} x_{t-1}^2, \\
\frac{\partial f_t(\theta)}{\partial \beta^2 \partial \gamma} &= -\frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} x_{t-1}^3, \\
\frac{\partial f_t(\theta)}{\partial \beta^2 \partial \alpha} &= -\gamma \frac{C\exp(L_t)}{(\exp(L_t)+1)^4} x_{t-1}^3,
\end{aligned}
$$

$$\frac{\partial f_t(\theta)}{\partial\gamma\partial\alpha\partial\beta} = 2\frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3}x_{t-1}^2$$

where the derivatives are not 0 and where $C = 1 - 4\exp(L_t) + \exp(2L_t)$. Thus we have $L(\theta)$ such that $p\lim_{n\to\infty}\frac{1}{n}G'(\theta)G(\theta)$ converges uniformly in θ on the closure $\bar{S}$ of S as in condition (i). All tail cross products exist for each term in the matrix.

To show condition (iv) is satisfied, we need to show that for any k-dimensional vector $\lambda, \lambda \neq 0$,

$$\lambda'[n^{-1/2}F'_{nk}(\theta_0)e_n] = \sum_{t=1}^{n} n^{-1/2}\Big[\sum_{j=1}^{k}\lambda_j f_t^{(j)}(\theta_0)e_t\Big]$$

converges in law to a univariate normal distribution. Fuller(1996) proved it using theorem 5.3.4 for general nonlinear model.

Theorem 5.3.4 in Fuller(1996) gives a central limit theorem for martingale differences as follows.

Let $\{Z_{tn} : 1 \leq t \leq n, n \geq 1\}$ denote a triangular array of random variables defined on the probability space (Ω, A, p), and let $\{A_{tn} : 0 \leq t \leq n, n \geq 1\}$ be any triangular array of sub-σ-fields of A such that for each n and $1 \leq t \leq n$, Z_{tn} is A_{tn}-measurable and $A_{t-1,n}$ is contained in A_{tn}. For $1 \leq k \leq n, 1 \leq j \leq n$, and $n \geq 1$, let

$$\begin{aligned} S_{kn} &= \sum_{t=1}^{k} Z_{tn}, \\ \delta_{tn}^2 &= E\{Z_{tn}^2 | A_{t-1,n}\}, \\ V_{jn}^2 &= \sum_{t=1}^{j}\delta_{tn}^2, \end{aligned}$$

and

$$s_{nn}^2 = E\{V_{nn}^2\}.$$

Assume

(i) $E(Z_{tn}|A_{t-1,n}) = 0$ almost surely for $1 \leq t \leq n$,

(ii) $V_{nn}^2 s_{nn}^{-2} \xrightarrow{p} 1$,

(ii) $\lim_{n\to\infty} s_{nn}^{-2} \sum_{j=1}^{n} E\{Z_{jn}^2 I(|Z_{jn}| \geq \epsilon s_{nn})|A_{t-1,n}\} = 0$ for all $\epsilon > 0$. $I(A)$ denotes the indicator function of a set A.

Then, as $n \to \infty$,

$$s_{nn}^{-1} S_{nn} \xrightarrow{d} N(0,1)$$

(Fuller, 1996).

Note that $n^{-1/2} F_{nk}'(\theta_0) e_n$ is already normal with e_t from an IID $N(0, \sigma^2)$ distribution. We quote the proof in our notation. Let

$$Z_{tn} = n^{-1/2}\Big[\sum_{j=1}^{k} \lambda_j f_t^{(j)}(\theta_0) e_t\Big]$$

and let $A_{t,n}$ be $A_t, 0 \leq t \leq n, n \geq 1$. Then

$$E\{Z_{tn}|A_{t-1,n}\} = 0$$

almost surely and

$$V_{nn}^2 = \sum_{t=1}^{n} E\{Z_{tn}^2 | A_{t-1,n}\} = \sigma^2 \lambda' [n^{-1} F_{nk}'(\theta_0) F_{nk}(\theta_0)] \lambda.$$

Also

$$Var\{n^{-1/2} \lambda' F_{nk}'(\theta_0) e_n\} = \sigma^2 \lambda' E\{[n^{-1} F_{nk}'(\theta_0) F_{nk}(\theta_0)]\} \lambda$$

and

$$s_{nn}^2 = E\{V_{nn}^2\} = \sigma^2 \lambda' E\{[n^{-1} F_{nk}'(\theta_0) F_{nk}(\theta_0)]\}\lambda$$

.

We have

$$\begin{aligned} p \lim_{n\to\infty} \frac{1}{n} F_{nk}'(\theta_0) F_{nk}(\theta_0) &= \lim_{n\to\infty} \frac{1}{n} E\{F_{nk}'(\theta_0) F_{nk}(\theta_0)\} \\ &= B(\theta_0) \end{aligned}$$

from condition (i).

Thus,

$$\begin{aligned} \sigma^{-2}(V_{nn}^2 - s_{nn}^2) &= \lambda'[n^{-1} F_{nk}'(\theta_0) F_{nk}(\theta_0)]\lambda \\ &- \lambda' E\{[n^{-1} F_{nk}'(\theta_0) F_{nk}(\theta_0)]\}\lambda \end{aligned}$$

converges to zero in probability.

$$s_{nn}^2 \to \sigma^2 \lambda' B(\theta_0) \lambda > 0$$

and

$$(V_{nn}^2 - s_{nn}^2) s_{nn}^{-2} \to 0$$

as $n \to \infty$ by the Slutsky theorem.

Now for arbitrary $\epsilon > 0$,

$$s_{nn}^{-2} \sum_{j=1}^{n} E\{Z_{jn}^2 I(|Z_{jn}| \geq \epsilon s_{nn}) | A_{t-1,n}\}$$

$$
\begin{aligned}
&\leq \; s_{nn}^{-2}\sum_{j=1}^{n}(\epsilon s_{nn})^{-\nu}E\{|Z_{jn}|^{2+\nu}I(|Z_{jn}|\geq \epsilon s_{nn})|A_{t-1,n}\} \\
&\leq \; \epsilon^{-\nu}s_{nn}^{-(2+\nu)}n^{-(1+\nu/2)} \\
&\quad \times\sum_{t=1}^{n}E\Big\{\Big|\sum_{j=1}^{k}\lambda_j f_t^{(j)}(\theta_0)e_t\Big|^{2+\nu}I(|Z_{jn}|\geq \epsilon s_{nn})|A_{t-1,n}\Big\} \\
&\leq \; \epsilon^{-\nu}s_{nn}^{-(2+\nu)}n^{-(1+\nu/2)}\sum_{t=1}^{n}E\Big\{\Big|\sum_{j=1}^{k}\lambda_j f_t^{(j)}(\theta_0)e_t\Big|^{2+\nu}|A_{t-1,n}\Big\} \\
&\leq \; \epsilon^{-\nu}s_{nn}^{-(2+\nu)}n^{-(\nu/2)}M_F
\end{aligned}
$$

using the previously defined bound M_F, and

$$
\lim_{n\to\infty} s_{nn}^{-2}\sum_{j=1}^{n}E\{Z_{jn}^2 I(|Z_{jn}|\geq \epsilon s_{nn})|A_{t-1,n}\}=0
$$

for all $\epsilon > 0$(Fuller, 1996).

Hence, the conditions for theorem 5.3.4 in Fuller are all satisfied and

$$
n^{-1/2}\sum_{t=1}^{n}F_t'(\theta_0)e_t \xrightarrow{d} N[0, B(\theta_0)\sigma^2].
$$

Thus, according to Fuller's theorem, we can conclude that the asymptotic distribution of $\tilde{\theta}$ converges to a normal distribution.

$$
n^{1/2}(\tilde{\theta}-\theta_0) \xrightarrow{d} N[0, B^{-1}(\theta_0)\sigma^2]
$$

where $B(\theta_0)$ =$\text{plim}_{n\to\infty}\,\frac{1}{n}F_{nk}'(\theta_0)F_{nk}(\theta_0)$.

If we know x_t is a finite random sample from a $(0,\sigma^2)$ distribution or a strictly stationary and ergodic sequence such as a a finite order Gaussian ARMA series, we can obtain $B(\theta_0)$ that has the form of an expectation.

The consistency of $\hat{\theta}_n$ can be shown using lemma 5.5.2 in Fuller(1996), which is more general than that of Jennrich(1969). We state the lemma in our notation for completeness.

Let $\hat{\theta}_n$ in Θ be a measurable function that minimizes an objective function $Q_n(\theta)$ on Θ almost surely. In addition, given (Ω, A, p) and a compact set Ω that is a subset of R^k, let $Q_n : \Omega \times \Theta \rightarrow R$ be a random function continuous on Θ a.s., for $n = 1, 2, \cdots$. Suppose there exists a function $\bar{Q}_n : \Theta \rightarrow R$ such that

$$Q_n(\theta) - \bar{Q}_n(\theta) \xrightarrow{a.s.} 0$$

uniformly on Θ and assume that for any $\eta > 0$,

$$\lim_{n\rightarrow\infty}\inf \left\{ \inf_{|\theta-\theta^0|\geq\eta} [\bar{Q}_n(\theta) - \bar{Q}_n(\theta_0)] \right\} > 0.$$

Then $\hat{\theta}_n - \theta_0 \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$(Fuller, 1996).

The first condition is related to a uniform law of large numbers and the second condition is the identification condition.

Where $Q_n(\theta) = n^{-1}\sum_{t=1}^{n}(y_t - f_t(\theta))^2$, a natural choice for $\bar{Q}_n(\theta)$ is

$$\bar{Q}_n(\theta) = n^{-1}\sum_{t=1}^{n} E[(y_t - f_t(\theta))^2]$$

(Fuller, 1996).

Here,

$$\begin{aligned}
\bar{Q}_n(\theta) &= n^{-1}\sum_{t=1}^{n} E\Big[(y_t - f_t(\theta_0) + f_t(\theta_0) - f_t(\theta))^2\Big] \\
&= n^{-1}\sum_{t=1}^{n}\Big[E(e_t^2) + 2E(e_t(f_t(\theta_0) - f_t(\theta))) + E((f_t(\theta_0) - f_t(\theta))^2)\Big] \\
&= \sigma^2 + E((f_t(\theta_0) - f_t(\theta))^2)
\end{aligned}$$

where X and e are independent.

Now

$$
\begin{aligned}
Q_n(\theta) &= n^{-1}\sum_{t=1}^{n}\Big[(y_t - f_t(\theta_0) + f_t(\theta_0) - f_t(\theta))^2\Big] \\
&= n^{-1}\sum_{t=1}^{n}\Big[e_t^2 + 2e_t(f_t(\theta_0) - f_t(\theta)) + (f_t(\theta_0) - f_t(\theta))^2\Big].
\end{aligned}
$$

Note that if x_t is a finite order Gaussian ARMA series and e_t is IID $N(0, \sigma^2)$, then x_t and e_t are jointly strictly stationary and ergodic. So, any measurable function of the series is also strictly stationary and ergodic(Stout, 1974;Taniguchi and Kakizawa, 2000). We quote theorem 1.3.3 from Taniguchi and Kakizawa(2000) in our notation.

Suppose that a vector process $\{\mathbf{Z_t} : t \in Z\}$ is strictly stationary and ergodic, and that there is a measurable function $\phi : R^\infty \to R^k$. Let $\mathbf{Y_t} = \phi(\mathbf{Z_t}, \mathbf{Z_{t-1}}, \cdots)$. Then $\{\mathbf{Y_t} : t \in Z\}$ is strictly stationary and ergodic.

Thus $e_t^2, e_t(f_t(\theta_0) - f_t(\theta))$ and $(f_t(\theta_0) - f_t(\theta))^2$ are strictly stationary and ergodic. In addition, the ergodic theorem says if $\{\mathbf{Y_t} : t \in Z\}$ is strictly stationary and ergodic and $E||\mathbf{Y_t}|| < \infty$, then

$$
\frac{1}{n}\sum_{t=1}^{n}\mathbf{Y_t} \xrightarrow{a.s.} E(\mathbf{Y_1})
$$

(Stout, 1974; Davidson, 1994; Taniguchi and Kakizawa, 2000).

So,

$$
n^{-1}\sum_{t=1}^{n} e_t^2 \xrightarrow{a.s.} E(e_t^2) = \sigma^2
$$

with $E(e_t^2) < \infty$.

$$
\begin{aligned}
& E(|e_t(f_t(\theta_0) - f_t(\theta))|) \\
=\ & E(|e_t||f_t(\theta_0) - f_t(\theta)|) \\
=\ & E(|e_t|)E(|f_t(\theta_0) - f_t(\theta)|).
\end{aligned}
$$

With $E(e_t^2) < \infty$ and $E(|x_t|) < \infty$,

$$E(|e_t(f_t(\theta_0) - f_t(\theta))|) < \infty.$$

Thus, by the ergodic theorem,

$$n^{-1}\sum_{t=1}^{n} e_t(f_t(\theta_0) - f_t(\theta)) \xrightarrow{a.s.} E(e_t(f_t(\theta_0) - f_t(\theta))) = 0.$$

In this way, with $E(x_t^2) < \infty$,

$$n^{-1}\sum_{t=1}^{n}(f_t(\theta_0) - f_t(\theta))^2 \xrightarrow{a.s.} E((f_t(\theta_0) - f_t(\theta))^2)$$

as well.

That is,

$$Q_n(\theta) = n^{-1}\sum_{t=1}^{n}(y_t - f_t(\theta))^2 \xrightarrow{a.s.} \bar{Q}_n(\theta) = \sigma^2 + E((f_t(\theta_0) - f_t(\theta))^2).$$

Hence,

$$Q_n(\theta) - \bar{Q}_n(\theta) \xrightarrow{a.s.} 0.$$

To show

$$\lim_{n\to\infty}\inf \Big\{ \inf_{|\theta-\theta^0|\geq\eta} [\bar{Q}_n(\theta) - \bar{Q}_n(\theta_0)] \Big\} > 0,$$

we notice that

$$\begin{aligned}
\bar{Q}_n(\theta) - \bar{Q}_n(\theta_0) &= n^{-1}\sum_{t=1}^{n} E\Big[(y_t - f_t(\theta))^2 - (y_t - f_t(\theta_0))^2\Big] \\
&= n^{-1}\sum_{t=1}^{n} E((f_t(\theta_0) - f_t(\theta))^2).
\end{aligned}$$

Thus, if a unique minimum exists at $\theta = \theta_0$, the condition is satisfied. Hence, $\hat{\theta}_n - \theta_0 \xrightarrow{a.s.} 0$ as $n \to \infty$.

For the asymptotic distribution of $\hat{\theta}_n$, we use Fuller's theorem 5.5.4 here again.

In this case, the variations enter into Y through both X and e. The Xs are assumed to be independent of the errors and A_{t-1} is generated by $\{x_1, x_2, ..., x_{t-1}, e_1, e_2, ..., e_{t-1}\}$.

Given A_{t-1} and e_t from a $N(0, \sigma^2)$ distribution, the vector sequence $\{[F_t(\theta_0)e_t, e_t]\}$ satisfies

$$E\{[F_t(\theta_0)e_t, e_t, e_t^2]|A_{t-1}\} = [\mathbf{0}, 0, \sigma^2],$$

$$E\{|[F_t(\theta_0)e_t, e_t]|^{2+\nu}|A_{t-1}\} < M_F < \infty,$$

$$E\{[F_t'(\theta_0)F_t(\theta_0)e_t^2]|A_{t-1}\} = F_t'(\theta_0)F_t(\theta_0)\sigma^2$$

a.s., for all t.

For condition (i), $B(\theta)$ is obtained by the ergodic theorem.

If x_t is an IID sequence or a strictly stationary and ergordic series, then any function of the series is also strictly stationary and ergodic.

So, by theorem 1.3.3 in Taniguchi and Kakizawa(2000), we know $\rho^2(x_{t-1})x_{t-1}^2$, $\gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^j$ and $\gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^j$ where $j = 2, 3, 4$, are strictly stationary and ergodic.

Also,

$$\left|\rho^2(x_{t-1})x_{t-1}^2\right| < |x_{t-1}^2|.$$

For $j = 2, 3, 4$,

$$\left|\gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^j\right| < |\gamma||x_{t-1}^j|,$$

$$\left|\gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^j\right| < |\gamma^2||x_{t-1}^j|.$$

Assuming a finite 4th moment for X, all terms in $F'_{nk}(\theta)F_{nk}(\theta)$ converge to finite expectations by the ergodic theorem.

Thus,

$$\frac{1}{n}\sum_{t=1}^{n}\rho^2(x_{t-1})x_{t-1}^j \xrightarrow{a.s.} E[\rho^2(x_{t-1})x_{t-1}^j].$$

For $j = 2, 3$,

$$\frac{1}{n}\sum_{t=1}^{n}\gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^j \xrightarrow{a.s.} E[\gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^j],$$

and for $j = 2, 3, 4$,

$$\frac{1}{n}\sum_{t=1}^{n}\gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^j \xrightarrow{a.s.} E[\gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^j].$$

Recall that our assumptions imply $E(\phi(x_{t-1}))$ is a constant function of t.

This shows that $\frac{1}{n}F'_{nk}(\theta)F_{nk}(\theta) =$

$$\frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix} \rho^2(x_{t-1})x_{t-1}^2 & \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^2 & \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^3 \\ \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^2 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^2 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^3 \\ \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^3 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^3 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^4 \end{pmatrix}$$

converges to

$$E\begin{pmatrix} \rho^2(x_{t-1})x_{t-1}^2 & \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^2 & \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^3 \\ \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^2 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^2 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^3 \\ \gamma\rho^2(x_{t-1})(1-\rho(x_{t-1}))x_{t-1}^3 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^3 & \gamma^2\rho^2(x_{t-1})(1-\rho(x_{t-1}))^2x_{t-1}^4 \end{pmatrix}$$

in probability and this is continuous at θ_0.

Using theorem 1.3.3 from Taniguchi and Kakizawa(2000) and the ergodic theorem, condition (ii) is also obtained uniformly in θ on the closure $\bar{S}$ of S with a finite 8th moment for X.

Condition(iv) is satisfied irrespective of X being fixed or random.

Hence, according to Fuller's theorem, we can conclude that the asymptotic distribution of $\tilde{\theta}$ converges to a normal distribution.

$$n^{1/2}(\tilde{\theta} - \theta_0) \xrightarrow{d} N[0, B^{-1}(\theta_0)\sigma^2]$$

and $B(\theta_0)$ has the form

$$E\begin{pmatrix} \rho_0^2(x_{t-1})x_{t-1}^2 & \gamma_0\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}^2 & \gamma_0\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}^3 \\ \gamma_0\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}^2 & \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2x_{t-1}^2 & \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2x_{t-1}^3 \\ \gamma_0\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}^3 & \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2x_{t-1}^3 & \gamma_0^2\rho_0^2(x_{t-1})(1-\rho_0(x_{t-1}))^2x_{t-1}^4 \end{pmatrix}$$

where $\rho_0(x_{t-1}) = \frac{1}{\exp(\alpha_0+\beta_0 x_{t-1})+1}$(see section 2.2).

Serially correlated errors are easily incorporated into this model. The approach is to assume the process $\{\eta_t\}_{t=-\infty}^{\infty}$ generating the realized disturbances $\{\eta_t\}_{t=1}^{n}$ is covariance stationary in the following model.

$$y_t = f(x_t; \theta) + \eta_t$$

and

$$\eta_t = \delta_1\eta_{t-1} + \cdots + \delta_k\eta_{t-k} + e_t.$$

Then, we get the autocovariance function of the process

$$\gamma(h) = cov(\eta_t, \eta_{t+h})$$

where $h = 0, \pm 1, \pm 2, \cdots$.

Now, with known Γ_n, we would estimate θ_0 by the value of θ which minimizes

$$[y - f(\theta)]'\Gamma_n^{-1}[y - f(\theta)]$$

where $y = (y_1, y_2, \cdots, y_n)'(n \times 1)$ and

$$f(\theta) = [f(x_1; \theta), f(x_2; \theta), \cdots, f(x_n; \theta)]'(n \times 1).$$

If we suppose that Γ_n^{-1} can be factored as $\Gamma_n^{-1} = cP'P$ where c is scalar and we put $z = Py, h(\theta) = Pf(\theta)$, and $e = P\eta$, then the newly created model will be

$$z = h(\theta) + e$$

where $E(e) = 0$ and $Var(e) = \sigma^2 I$.

For example, if the errors have AR(1) structure with coefficient δ, then

$$\hat{P} = \begin{pmatrix} \sqrt{1-\delta^2} & 0 & 0 & 0 & \cdots & 0 & 0 \\ -\delta & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\delta & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\delta & 1 \end{pmatrix}$$

can be considered.

Here, $z_t = Py_t$ has the form

$$\begin{pmatrix} \sqrt{1-\delta^2}y_1 \\ -\delta y_1 + y_2 \\ -\delta y_2 + y_3 \\ \\ \vdots \\ \\ -\delta y_{n-2} + y_{n-1} \\ -\delta y_{n-1} + y_n \end{pmatrix}$$

and $h_t(\theta) = Pf_t(\theta)$ has the form

$$\begin{pmatrix} \sqrt{1-\delta^2}\gamma\rho(x_0)x_0 \\ -\delta\gamma\rho(x_0)x_0 + \gamma\rho(x_1)x_1 \\ -\delta\gamma\rho(x_1)x_1 + \gamma\rho(x_2)x_2 \\ \\ \vdots \\ \\ -\delta\gamma\rho(x_{n-3})x_{n-3} + \gamma\rho(x_{n-2})x_{n-2} \\ -\delta\gamma\rho(x_{n-2})x_{n-2} + \gamma\rho(x_{n-1})x_{n-1} \end{pmatrix}.$$

Now the estimators are the minimizer of

$$Q_{nl}(\theta) = \frac{1}{n}\sum_{t=2}^{n}[z_t - \delta\gamma\rho(x_{t-2})x_{t-2} - \gamma\rho(x_{t-1})x_{t-1}]^2$$

with resepct to $\theta = (\gamma, \alpha, \beta)$ with δ known. We can ignore the first observation. Then the model goes back to the standard case. This justifies the nonlinear least squares estimator and associated inference procedure(Gallant, 1986). The method would be the general nonlinear least squares estimator.

In reality, Γ_n^{-1} is not known. It can be replaced by $\hat{\Gamma}_n^{-1}$ which would be obtained applying an AR(p) model to estimated residuals(Gallant and Goebel, 1976; Gallant, 1986). Gallant(1986) provides the theoretical justification for this approach. Then the model is estimated using the Gauss-Newton algorithm with $\hat{\Gamma}_n^{-1}$.

The estimation procedure may be iterated getting autocorrelation in residuals from the previously fitted model. The asymptotic properties of the estimator by iteration do not differ and $n^{-1/2}(\hat{\theta}_n - \theta_0)$ is asymptotically normally distributed under appropriate regularity condition(Gallant and Goebel, 1976).

The consistency of $\hat{\theta}_n$ is still obtained because a unique minimum exists at $\theta = \theta_0$. Note that we assume $0 < \delta < 1$. Using theorem 5.5.4 in Fuller again, we show the one-step Gauss-Newton estimator converges to a normal distribution.

Let $H_t(\theta)$ is the k dimensional vector of first derivatives of $h_t(\theta)$,$t = 1, 2, \cdots, n$, and

$$H_{nk}(\theta) = [H_1'(\theta), H_2'(\theta), \cdots, H_n'(\theta)]'$$

is the $n \times k$ matrix of first derivatives.

Now the Gauss-Newton estimator will be

$$\begin{aligned}
\tilde{\theta} &= \hat{\theta} + (H_{nk}'(\hat{\theta})H_{nk}(\hat{\theta}))^{-1}H_{nk}'(\hat{\theta})\hat{e} \\
&= \hat{\theta} + (F_{nk}'(\hat{\theta})\hat{\Gamma}_n^{-1}F_{nk}(\hat{\theta}))^{-1}F_{nk}'(\hat{\theta})\hat{\Gamma}_n^{-1}\hat{e}.
\end{aligned}$$

We assume x_t is conditionally fixed and the error structure η_t is covariance stationary.

To begin with,

$$E\{[H_t(\theta_0)e_t, e_t, e_t^2]|A_{t-1}\} = [\mathbf{0}, 0, \sigma^2],$$

$$E\{|[H_t(\theta_0)e_t, e_t]|^{2+\nu}|A_{t-1}\} < M_H < \infty,$$

$$E\{[H_t'(\theta_0)H_t(\theta_0)e_t^2]|A_{t-1}\} = H_t'(\theta_0)H_t(\theta_0)\sigma^2$$

a.s., for all t.

Assuming the errors follow an AR(1) process with parameter δ, and ignoring the first observation,

$$\begin{aligned}
H_t'(\theta_0) = {} & [-\hat{\delta}\rho_0(x_{t-2})x_{t-2} + \rho_0(x_{t-1})x_{t-1}, \\
& -\hat{\delta}\gamma_0\rho_0(x_{t-2})(1-\rho_0(x_{t-2}))x_{t-2} + \gamma_0\rho_0(x_{t-2})(1-\rho_0(x_{t-2}))x_{t-2}, \\
& -\hat{\delta}\gamma_0\rho_0(x_{t-2})(1-\rho_0(x_{t-2}))x_{t-2}^2 + \gamma_0\rho_0(x_{t-1})(1-\rho_0(x_{t-1}))x_{t-1}^2]
\end{aligned}$$

where $\rho_0(x_{t-1}) = \frac{1}{\exp(\alpha_0+\beta_0 x_{t-1})+1}$.

A_{t-1} is generated by $\{e_1, e_2, \cdots, e_{t-1}\}$, and under a given A_{t-1}, the assumptions for the vector sequence $\{[H_t(\theta_0)e_t, e_t]\}$ are all satisfied with e_t from a $N(0, \sigma^2)$ distribution.

Now condition (i) and (ii) are satisfied where

$$\begin{aligned} p\lim_{n\to\infty}\frac{1}{n}H'(\theta)H(\theta) &= p\lim_{n\to\infty}\frac{1}{n}(\hat{P}F(\theta))'(\hat{P}F(\theta)) \\ &= p\lim_{n\to\infty}\frac{1}{n}F'(\theta)\hat{P}'\hat{P}F(\theta) \\ &= p\lim_{n\to\infty}\frac{1}{n}F'(\theta)\hat{\Gamma}^{-1}F(\theta) \\ &= B(\theta) \end{aligned}$$

where $\hat{\Gamma}^{-1}$ is an $n \times n$ matrix and

$$\begin{aligned} p\lim_{n\to\infty}\frac{1}{n}J'(\theta)J(\theta) &= p\lim_{n\to\infty}\frac{1}{n}G'(\theta)\hat{\Gamma}^{-1}G(\theta) \\ &= L(\theta) \end{aligned}$$

where $J(\theta)$ is an $n \times (1 + k + k^2 + k^3)$ matrix with t-th row given by

$$\Big[h_t(\theta), h_t^{(1)}(\theta), \cdots, h_t^{(k)}(\theta), h_t^{(11)}(\theta), \cdots, h_t^{(1k)}(\theta),$$
$$h_t^{(k1)}(\theta), \cdots, h_t^{(kk)}(\theta), h_t^{(111)}(\theta), \cdots, h_t^{(kkk)}(\theta)\Big].$$

For condition (iv), the proof is exactly the same as that of the standard case except that

$$Z_{tn} = n^{-1/2}\Big[\sum_{j=1}^{k}\lambda_j h_t^{(j)}(\theta_0)e_t\Big],$$

$$E\{Z_{tn}|A_{t-1,n}\} = 0$$

almost surely and

$$V_{nn}^2 = \sum_{t=1}^{n} E\{Z_{tn}^2|A_{t-1,n}\} = \sigma^2[n^{-1}F'_{nk}(\theta_0)\hat{\Gamma}_n^{-1}F_{nk}(\theta_0)].$$

Thus, according to theorem 5.5.4 in Fuller, the asymptotic distribution of $\tilde{\theta}$ converges to a normal distribution.

$$n^{1/2}(\tilde{\theta} - \theta_0) \xrightarrow{d} N[0, B^{-1}(\theta_0)\sigma^2]$$

where $B(\theta_0) = \text{plim}_{n\to\infty} \frac{1}{n} F'_{nk}(\theta_0)\hat{\Gamma}_n^{-1} F_{nk}(\theta_0)$.

Before building a nonlinear model, it is advisable to find out if indeed a nonlinear model would adequately characterize the relationship under analysis. Various nonlinearity tests have been suggested in the literature(Tong, 1990; Granger and Ter$\ddot{a}$svirta, 1993). They largely consist of tests derived against a specific nonlinear alternative that one has in mind or tests against a general nonlinear model.

Polynomial regressions can be used to test for nonlinearity in conditional expectation(Fuller, 1996). Let

$$\begin{aligned} y_t &= f(x_t; \theta) + e_t \\ &= \gamma\rho(x_{t-1})x_{t-1} + e_t \end{aligned}$$

where $\rho(x_{t-1}) = \frac{1}{\exp(\alpha+\beta x_{t-1})+1}$ and $\theta = (\gamma, \alpha, \beta)$.

Suppose X comes from a stationary time series with a $N(0, \sigma^2)$ distribution. For such series, theorem 8.6.1 in Fuller(1996) suggests that we fit a polynomial regression using $1, x_{t-1}, x_{t-1}^2$, $x_{t-1}^3, \cdots$, as explanatory variables.

Intuitively, this simply says that we expect curvature in the plot of y_t versus x_{t-1} when the coefficient on x_{t-1} is a nonconstant function of x_{t-1}. We can approximate the nonlinear function by, say, a 4th degree polynomial terms of x_{t-1}. Then, the null distribution of the test statistic F is approximately that of Snedecor's F with 3 and $(n-5)$ degrees of freedom, according to Fuller(1996).

Now we can extend our results to the model below.

$$y_t = \gamma_1 \rho(x_{t-d}) x_{t-1} + \gamma_2 \left(1 - \rho(x_{t-d})\right) x_{t-2} + e_t \tag{2.2}$$

where $\rho(x_{t-d}) = \frac{1}{\exp(\alpha+\beta x_{t-d})+1}$ and $d = 1, 2, \cdots$.

In this model, $\rho(x_{t-d})$ reflects a weight that allows X to dictate the relative influence of 1 and 2 lags. A weight could be dependent on the other lagged X variables.

Variants of the model can also be considered. One variant is

$$y_t = \rho(x_{t-d})[\gamma_0 x_t + \gamma_1 x_{t-1} + \cdots + \gamma_k x_{t-k}] + e_t \tag{2.3}$$

where $\rho(x_{t-d}) = \frac{1}{\exp(\alpha+\beta x_{t-d})+1}$ and $d = 0, 1, 2, \cdots$. $f(x)$ is a function of X. The model gives smoothly changing coefficients depending on lagged X.

In a similar way to that shown previously, we obtain the parameter estimates and the asymptotic distributions.

2.2 simulation

For model (2.1), 5,000 draws of data based on $(\gamma, \alpha, \beta) = (1.0, 0.8, 0.5)$ have been generated where (i) x_t and e_t come from an IID $N(0, 1)$ and an IID $N(0, 0.04)$ respectively, and (ii) x_t is AR(1) with the coefficient 0.6 and innovations $N(0, 1)$ and e_t comes from an IID $N(0, 0.04)$. x_t and e_t are independent.

The Gauss-Newton algorithm has been employed to estimate each triplet of parameters (γ, α, β). The maximum number of iterative updates of the estimation algorithm is chosen to be 100 and if the iteration continues up to this number, each program is stopped and it is declared that no convergence was reached. Also, if 2 successive iterates $\hat{\theta}_{(a+1)}$ and $\hat{\theta}_a$ have a relative difference of less than 10^{-8}, we stop the program and declare $\hat{\theta}_{(a+1)}$ to be solution. That is, we stop if

$$max_{l=1,\ldots,p}|\hat{\theta}_{l,(a+1)} - \hat{\theta}_{l,a}|/|\hat{\theta}_{l,a}| < 10^{-8}.$$

Without any further mention, all nonlinear least squares estimation in this paper will be done using the process above. Also, for all simulations made here, the initial values for generating the series are set to be 0 and the first 100 observations are discarded to eliminate the initialization effects. From these 5,000 nonlinear estimations, we have the empirical distribution of the parameters.

As shown in Table 2.1, the biases get smaller with increasing n. The figures in parentheses in the table show the theoretical standard errors of the parameter estimates. They are computed from the asymptotic variance-covariance matrix $B^{-1}(\theta_0)\sigma^2$ and are numerically evaluated using "SAS[1] IML". The standard deviations(STD) of $\hat{\gamma}$ are bigger than their theoretical values mainly due to a few extreme estimates.

[1]SAS is the registered trademark of SAS Institute, Cary, N.C.

JB stands for the Jarque-Bera statistic for testing whether the data are normally distributed. The test statistic measures the difference of the skewness and kurtosis of the data from those of the normal distribution. Under the null hypothesis of a normal distribution, the Jarque-Bera statistic is distributed as χ^2 with 2 degrees of freedom($\chi^2_{0.05} = 5.99$).

JB statistics soundly reject normality. However, it is seen that the skewness is not so bad. The non-normality appears to be due to a few extreme tail values. Often $\alpha = 0.05$ is used in constructing tests and confidence intervals. To see how the outliers affect this, a t statistic is computed by taking the deviation of each parameter estimate from the true value, then dividing by the estimated standard error. The proportion of t statistics that exceed the standard $Z_{0.025}$ is also reported in Table 2.1. They are close to the nominal 0.05 with $\hat{\gamma}$ showing the largest departure(0.0822 vs 0.05) where $n = 1,000$. We conclude that the departure from normality would have a rather small impact on the level in a nominal 5% level test.

The estimation did not converge in 100 iterations for a few replicates and thus the number of replications R in the table is not 5,000 in some cases.

2.3 application

A transfer function model, which has the form of model (2.2), having coefficients that are logistic functions of an observed variable, is fit to a string of log transformed daily flows for the Neuse River in North Carolina. The flows are measured at Goldsboro and downstream at Kinston North Carolina. The data consist of 400 daily measurements from Oct 1, 1970 to Nov 4, 1971(see Figure 2.3 (a)). The data have previously been analyzed using a standard, fixed transfer function methodology(Brocklebank and Dickey, 2003).

Their model is

$$\begin{aligned}
\nabla y_t &= \underset{(0.0185)}{0.4954}(1+\underset{(0.0454)}{0.5503}B)\nabla x_{t-1} \\
&+ (1-\underset{(0.0351)}{0.8878}B)/(1-\underset{(0.0505)}{1.1632}B+\underset{(0.0456)}{0.4796}B^2)e_t
\end{aligned}$$

where y_t and x_t are the log transformed series at the two stations and the mean square error for this model is 0.00584. The figures in parentheses indicate the standard errors for parameter estimates.

Since these are two stations on the same stream and the response is a flow, it would seem sensible to propose a model in which the lag structure between the stations is a function of the previous flow at the upstream station. During a high flow period, a percentage change in flow(additive change in log(flow)) at the upstream location is adding more water into the system than during a low flow period so that the effects might be felt downstream for a longer time, that is, the lag structure might extend back more days during a high flow period. On the other hand, the increased volume produces faster flows and thus any given parcel of water should arrive downstream in less time during a high flow period. Figure 2.4 shows that the prewhitened crosscorrelation imply that lags 1 and 2 capture the majority of the lag relationship.

Experiments show that the prewhitened crosscorrelation could be used for selecting lagged X variables, even if there is serial correlation in the errors. We get a similar pattern of the crosscorrelation function using the level data of two variables.

We estimate a model that allows differing lag structures and let the data settle the question of how these structures relate to flow rates. Because two lags appear to capture the majority of the transfer activity, we fit model (2.2) using two lags.

We have done a linearity test based on Theorem 8.6.1 of Fuller(1996). If we estimate a quadratic model as a first approximation, we obtain

$$\begin{aligned} y_t &= \underset{(0.3790)}{1.8734} + \underset{(0.3003)}{1.9046}x_{t-1} - \underset{(0.3000)}{1.2672}x_{t-2} + \underset{(0.0736)}{0.0512}x_{t-1}^2 - \underset{(0.1349)}{0.2985}x_{t-1}x_{t-2} \\ &+ \underset{(0.0681)}{0.2659}x_{t-2}^2 + \eta_t. \end{aligned}$$

The F statistic for testing the hypothesis that all quadratic terms are zero is $12.48(> F_\infty^3 \approx 2.60, \alpha = 0.05)$. The hypothesis of all zero coefficients is rejected.

Note that the estimate of x_{t-1}^2 is not significant. We can remove x_{t-1}^2 from the quadratic model. This leaves $x_{t-1}x_{t-2}$ and x_{t-2}^2. Because x_{t-2} appears in both, we use x_{t-2} as the variable in our logistic weight function.

With many parameters to be estimated, the nonlinear estimation algorithm converges well where the mean of X variable is subtracted from $x_t(dx_t = x_t - \bar{x})$. The resulting model is

$$
\begin{aligned}
y_t &= 0.8942\frac{\exp(-0.1570 - 0.4069dx_{t-2})}{\exp(-0.1570 - 0.4069dx_{t-2}) + 1}dx_{t-1} \\
&\quad (0.1008) \qquad (0.2069) \quad (0.1029) \\
&+ 0.5523\Big[1 - \frac{\exp(-0.1570 - 0.4069dx_{t-2})}{\exp(-0.1570 - 0.4069dx_{t-2}) + 1}\Big]dx_{t-2} \\
&\quad (0.0658) \\
&+ 7.5852 + \eta_t, \\
&\quad (0.0862)
\end{aligned}
$$

and

$$
\begin{aligned}
\eta_t &= 1.3220\eta_{t-1} - 0.5054\eta_{t-2} + 0.1371\eta_{t-4} + e_t. \\
&\quad (0.0473) \qquad (0.0586) \qquad (0.0306)
\end{aligned}
$$

The "SAS" procedure "PROC NLIN" was used for nonlinear estimation. We estimate the parameters by minimizing directly

$$\frac{1}{n}\sum_{t=2}^{n}[y_t - \delta y_{t-1} - \gamma\rho(x_{t-1})x_{t-1} + \delta\gamma\rho(x_{t-2})x_{t-2}]^2$$

with resepct to $\theta = (\delta, \gamma, \alpha, \beta)$, for example, where the model is

$$y_t = \gamma\rho(x_{t-1})x_{t-1} + \eta_t$$

and

$$\eta_t = \delta\eta_{t-1} + e_t,$$

which is easy to implement(Seber and Wild, 1988).

Note that the cross products matrix of the first derivatives for $\theta = (\delta, \gamma, \alpha, \beta)$ is block-diagonal, implying that the serial correlation can be estimated separately.

The lag p for an AR(p) model could be determined by observing autocorrelation in residuals from the fitted model where serial correlation is not considered.

We have estimated the same model for this data by using the stepwise generalized least squares method suggested by Gallant(1986). The following table shows 20 Gauss-Newton iterations. The last row shows almost the same estimates as given before(the first row($*$) in the Table). The final estimates for an AR(4) model in residuals are very similar as well.

	α	β	γ_0	γ_1	γ_2	MSE
*	-0.1570	-0.4069	7.5852	0.8942	0.5523	0.00584
1	0.5260	-1.7870	7.4819	0.8626	0.9534	0.02490
2	0.4229	-0.3564	7.4995	0.8155	1.0019	0.00663
3	0.3973	-0.2914	7.5066	0.8261	0.9622	0.00600
4	0.3588	-0.2935	7.5097	0.8345	0.9306	0.00599
5	0.3286	-0.2964	7.5122	0.8407	0.9061	0.00598
6	0.3004	-0.2994	7.5145	0.8464	0.8838	0.00598
7	0.2720	-0.3027	7.5170	0.8520	0.8618	0.00597
8	0.2409	-0.3066	7.5197	0.8578	0.8382	0.00597
9	0.2044	-0.3116	7.5230	0.8644	0.8109	0.00596
10	0.1590	-0.3187	7.5274	0.8722	0.7776	0.00595
11	0.1138	-0.3260	7.5320	0.8801	0.7453	0.00594
12	0.0476	-0.3400	7.5396	0.8881	0.6978	0.00591
13	-0.0286	-0.3593	7.5501	0.8950	0.6445	0.00587
14	-0.0920	-0.3806	7.5624	0.8966	0.5988	0.00583
15	-0.1289	-0.3970	7.5734	0.8938	0.5715	0.00581
16	-0.1443	-0.4058	7.5805	0.8907	0.5593	0.00581
17	-0.1495	-0.4094	7.5838	0.8890	0.5549	0.00581
18	-0.1515	-0.4108	7.5852	0.8882	0.5532	0.00581
19	-0.1521	-0.4113	7.5856	0.8879	0.5527	0.00581
20	-0.1524	-0.4115	7.7858	0.8878	0.5524	0.00581

All estimates except that for α are significant at the 5% level and there seems to exist no significant autocorrelation of residuals. The residuals from this fit have been passed to "SAS PROC ARIMA" to check the white noise assumption. Note that the Ljung-Box theory was developed for standard ARIMA models, and besides, degrees of freedom would have to be adjusted for the model fitting. The Ljung-Box statistic may be conservative when the test is applied to estimated residuals of a nonlinear model with small sample size(Eitrheim and Ter$\ddot{a}$svirta, 1996).

However, ARCH effects are detected here using the standard statistical tests. The ARCH test is used to check autoregressive conditional heteroskedasticity(ARCH) in the residuals and is motivated by the fact that the magnitude of residuals often appears to be related to the magnitude of recent residuals, especially in financial time series(Engle, 1982). "SAS PROC AUTOREG" provides Q and Lagrange multipler(LM) statistics that test for the absence of ARCH effects. Ignoring the ARCH effects may result in loss of efficiency of estimators and the test on residuals from nonlinear model fitting is often used to judge model misspecification(Granger and Ter$\ddot{a}$svirta, 1993).

Because there seems to be heteroscedasticity in residuals, we get the heteroscedastic invariant variance estimates. These are valid even if the innovations are not IID Gaussian(White, 1982).

Let $\hat{\theta}$ be the least squares estimator, which minimizes

$$\frac{1}{n}\sum_{t=1}^{n}(y_t - f_t(\theta))^2.$$

Then, the heteroscedastic invariant variance estimates are obtained as

$$\frac{1}{n}[A_n(\hat{\theta})B_n^{-1}(\hat{\theta})A_n(\hat{\theta})]^{-1}$$

where

$$A_n(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n}[f_t'(\hat{\theta})f_t'(\hat{\theta})^T - (y_t - f_t(\theta))f_t''(\theta)],$$

or simpler

$$A_n(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n} f_t'(\hat{\theta}) f_t'(\hat{\theta})^T,$$

and

$$B_n(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n}(y_t - f_t(\hat{\theta}))^2 f_t'(\hat{\theta}) f_t'(\hat{\theta})^T$$

(White, 1980, 1982; Gallant, 1986).

Using the parameter estimates obtained from the least squares estimation above, we get the robust standard errors for parameters. A simpler $A_n(\hat{\theta})$ is used for calculating variance estimates. All estimates are still significant except for that of α.

The parameter estimates and robust standard errors

α	β	γ_0	γ_1	γ_2	δ_1	δ_2	δ_4
-0.1570	-0.4069	7.5852	0.8942	0.5523	1.3220	-0.5054	0.1371
(0.2316)	(0.1138)	(0.0950)	(0.1411)	(0.0709)	(0.1266)	(0.1910)	(0.0792)

Note: The figures in parentheses indicate the robust standard errors.

We can think of $p = \frac{\exp(-0.1570 - 0.4069 dx_{t-2})}{\exp(-0.1570 - 0.4069 x_{t-2}) + 1}$ as a weight that allows the upstream flow dx_{t-2} to allocate the transfer relationship between 1 and 2 lags. Notice that if β in model (2.2) is 0, then p is a redundant constant and this is just the usual fixed coefficient transfer function model. Our β confidence interval lies well away from $\beta = 0$.

The mean square error 0.00585 is about the same as 0.00584 found with the fixed coefficient transfer function model. Notice that a unit root has been imposed on the error term in the fixed coefficient model, a rather difficult result to justify in stream flows, and there is some evidence(moving average coefficient 0.88) against unit roots. Mean square error comparisons may not be critically important here. Using the level variables in the standard transfer function model, we have not found a satisfactory model in terms of the cross-correlation of residuals with the input x_t and the autocorrelation in residuals. The obtained MSEs from those models are around 0.0063.

In Figure 2.3 (b), the observed and predicted series from the nonlinear transfer function are overlaid and are almost indistinguishable. The predictions seem to be excellent. Also, the weights for dx_{t-1} and dx_{t-2} are plotted against time in Figure 2.5. Notice that the logistic slope, $\hat{\beta}$ is significant and negative, indicating as do the graphs that larger flows at Goldsboro are associated with longer lag structures. That is, a percentage change in flow during a high flow period seems to result in a longer lasting downstream effect. Higher flows at Goldsboro put more weight on the second coefficient, resulting in a slightly larger and more prolonged effect at Kinston.

The range of $\hat{\rho}(dx_{t-2})$ is from about 0.26 to 0.69 in this series. We have analyzed the phase spectra of two stream flows at each of these extremes. To illustrate the results, we use the observed x_t series and generate two $\hat{y}_t$ series by computing

$$\hat{y}_t = \hat{\gamma}_0 + \hat{\gamma}_1 \hat{\rho}_i dx_{t-1} + \hat{\gamma}_2 (1 - \hat{\rho}_i) dx_{t-2}$$

where $\hat{\rho}_1 = 0.26$ and $\hat{\rho}_2 = 0.69$. We then calculate the two phase spectra and put them in Figure 2.6. Note that if

$$y_t = \gamma_0 + \gamma_1 \rho_i dx_{t-1} + \gamma_2 (1 - \rho_i) dx_{t-2} + \eta_t,$$

then under the standard independence assumption, the correlation between y_t and dx_{t-j} is the same as that between $\hat{y}_t$ and dx_{t-j} and thus the phase spectrum can be computed using $\hat{y}_t$.

The lag effects can be seen in the phase spectrum. The slope of the phase spectrum at low frequencies measures the time delay as a function of frequency. The phase spectra at our two extreme regimes give an estimated 1.55 day lag effect at a high flow period, whereas 1.14 day lag effect is observed at a low flow period using a few spectrum estimates at low frequencies(see Figure 2.6). Our nonlinear model estimation result coincides with the phase spectrum analysis.

We get the impulse response function based on the fitted model. The responses are obtained for a high and low stream flow scenario respectively. The stream flow series in Goldsboro extends from about -2.35(c_1) to 2.15(c_2). At both extreme levels, we impose a one unit increase($\Delta dx_{t-2} = 1$) and calculate the stream flow changes in Kinston.

$$\begin{aligned} y_t &= 0.8942\frac{\exp(-0.1570 - 0.4069(c_i + \Delta dx_{t-2}))}{(\exp(-0.1570 - 0.4069(c_i + \Delta dx_{t-2})) + 1}(c_i + \Delta dx_{t-1}) \\ &+ 0.5523\Big[1 - \frac{\exp(-0.1570 - 0.4069(c_i + \Delta dx_{t-2}))}{\exp(-0.1570 - 0.4069(c_i + \Delta dx_{t-2})) + 1}\Big](c_i + \Delta dx_{t-2}) \\ &+ 7.5852 + \eta_t \end{aligned}$$

where $i = 1, 2$. Δdx_t is given by AR models.

In terms of the estimation results for the log-transformed Goldsboro stream flow series, AR(5) with some insignificant coefficients is chosen based on the AIC. An AR(2) is selected by the SBC. We generate values following the impulse based on the estimated AR(2) and AR(5) models as well as an AR(1) model. Figure 2.7 shows the responses following each impulse function. The response deviations and the changes in $\rho(dx_{t-2})$ from the equilibrium state are displayed. The changes are more dramatic at a low stream flow at Goldsboro.

The model could even be extended to

$$\begin{aligned} y_t &= \gamma_1[1-\rho_1(x_{t-d})-\rho_2(x_{t-d})]x_t+\gamma_2\rho_1(x_{t-d})x_{t-1} \\ &+ \gamma_3\rho_2(x_{t-d})x_{t-2}+e_t \end{aligned}$$

where $\rho_1(x_{t-d}) = \frac{\exp(\alpha_1+\beta_1 x_{t-d})}{\sum_{i=1}^{2}\exp(\alpha_i+\beta_i x_{t-d})+1}$ and $\rho_2(x_{t-d}) = \frac{\exp(\alpha_2+\beta_2 x_{t-d})}{\sum_{i=1}^{2}\exp(\alpha_i+\beta_i x_{t-d})+1}$. $d = 0, 1, 2, \cdots$.

We use 3 logistic weights, one for each of x_t, x_{t-1} and x_{t-2}. As an example, we analyse flows from the same stream between Kinston and downstream Fort Barnwell in North Carolina. 1,003 daily measurements from Oct 1, 1996 to Jun 30, 1999 are used. The data are log transformed and differenced as before. Figure 2.8 (a) shows two series.

Using an ARMA(3,1) for prewhitening the series(∇x_t) in Kinston, three lags appear to capture the majority of the transfer activity in this case(see Figure 2.9).

By a standard transfer model, we get the fitted model

$$\begin{aligned} \nabla y_t &= \underset{(0.0224)}{0.3042}(1+\underset{(0.1590)}{0.6175}B-\underset{(0.1743)}{0.5939}B^2)/(1-\underset{(0.0629)}{0.6639}B)\nabla x_t \\ &+ (1-\underset{(0.0873)}{0.4402}B-\underset{(0.0875)}{0.5550}B^2)/(1-\underset{(0.0989)}{0.5448}B-\underset{(0.0914)}{0.2421}B^2)e_t \end{aligned}$$

The obtained MSE is 0.00467.

However, there seems to be a near unit root in the estimated coefficients of the numerator factor for e_t suggesting possible overdifferencing. Considering we have rather a long series here, we estimate the transfer function model for level data as well.

$$
\begin{aligned}
y_t &= 0.3042(1 + 0.6126B - 0.6049B^2)/(1 - 0.6699B)x_t \\
&\quad\;\; (0.0224) \quad (0.1586) \quad (0.1733) \qquad\qquad (0.0614) \\
&+ (1 + 0.5570B)/(1 - 0.5477B - 0.2424B^2)e_t + 0.9228. \\
&\qquad\quad (0.0878) \qquad (0.0991) \qquad (0.0913) \qquad (0.1215)
\end{aligned}
$$

The model seems to fit well and the obtained MSE is 0.00465.

We fit the data using a 3 weight model considering that 3 major lags are found in the cross-correlation function. We have done a linearity test for the levels y_t using the explanatory variables x_t, x_{t-1}, and x_{t-2} and polynomials in these up through quadratic terms. The F statistic is 8.13. The linearity hypothesis is rejected at the 5% significance level($F_\infty^6 \approx 2.10, \alpha = 0.05$).

Because the nonlinear estimation algorithm in the suggested model does not converge well with many parameters estimated, the mean of the X variable is subtracted from $x_t(dx_t = x_t - \bar{x})$ and we set $\alpha = \alpha_1 = \alpha_2$. Nonlinear parameter estimates are well obtained and the estimates are all significant at the 5% level except for that of α(t-ratio: -1.93).

$$
\begin{aligned}
y_t &= 0.8020 \frac{1}{\exp(-0.6106 - 0.6877dx_{t-2}) + \exp(-0.6106 + 0.6823dx_{t-2}) + 1} dx_t \\
&\quad (0.1273) \quad (0.3150) \quad (0.1571) \qquad\qquad\qquad\qquad (0.1771) \\
&+ 0.8235 \frac{\exp(-0.6106 - 0.6877dx_{t-2})}{\exp(-0.6106 - 0.6877dx_{t-2}) + \exp(-0.6106 + 0.6823dx_{t-2}) + 1} dx_{t-1} \\
&\quad (0.1004) \\
&+ 1.0727 \frac{\exp(-0.6106 + 0.6823dx_{t-2})}{\exp(-0.6106 - 0.6877dx_{t-2}) + \exp(-0.6106 + 0.6823dx_{t-2}) + 1} dx_{t-2} \\
&\quad (0.1178)
\end{aligned}
$$

$$+ \quad 7.9439 + \eta_t,$$
$$(0.0205)$$

and

$$\eta_t = 1.0868\eta_{t-1} - 0.3440\eta_{t-2} + 0.1255\eta_{t-3} + e_t.$$
$$(0.0319) \qquad (0.0458) \qquad (0.1255)$$

The MSE is 0.00471 and no significant serial correlation in the residuals remains. The MSE is slightly worse than that of the transfer function in this example.

One possible reason the nonlinear models are not superior is failure to include enough lagged X terms in the explanatory variables. The term $(1 - 0.6699B)^{-1}x_t$ in the standard transfer function model includes more lagged X variables and three lags may not be enough to approximate this.

Including dx_{t-4} and dx_{t-6} in the explanatory variables, we have the following result(Originally dx_{t-3} and dx_{t-5} were included, but omitted because they had $|t|$-ratio less than 1).

$$\begin{aligned}
y_t &= 0.8066\frac{1}{\exp(-0.5814 - 0.7231dx_{t-2}) + \exp(-0.5814 + 0.7617dx_{t-2}) + 1}dx_t \\
&\quad (0.1270) \quad (0.3155) \quad (0.1572) \qquad\qquad\qquad\qquad (0.2002) \\
&+ 0.7629\frac{\exp(-0.5814 - 0.7231dx_{t-2})}{\exp(-0.5814 - 0.7231dx_{t-2}) + \exp(-0.5814 + 0.7617dx_{t-2}) + 1}dx_{t-1} \\
&\quad (0.0882) \\
&+ 0.9121\frac{\exp(-0.5814 + 0.7617dx_{t-2})}{\exp(-0.5814 - 0.7231dx_{t-2}) + \exp(-0.5814 + 0.7617dx_{t-2}) + 1}dx_{t-2} \\
&\quad (0.0916) \\
&+ 7.9572 + 0.0366dx_{t-4} + 0.0596dx_{t-6} + \eta_t, \\
&\quad (0.0206)(0.0211) \qquad (0.0181)
\end{aligned}$$

and

$$\eta_t = 1.0665\eta_{t-1} - 0.3256\eta_{t-2} + 0.1287\eta_{t-3} + e_t.$$
$$\quad (0.0319) \qquad (0.0455) \qquad (0.0319)$$

Now the obtained MSE is 0.00457 and the estimates are significant at the 5% significance level except for those of α(t-ratio: -1.84) and dx_{t-4}(t-ratio: 1.73). There seems to be no serial correlation left in residuals as before. A few added X terms have brought the MSE to the smallest value yet.

It is reasonable that we observe the nonlinear behavior at a few low lag terms rather than through all lags.

Figure 2.8 (b) shows the predicted stream flows of Fort Barnwell and Figure 2.10 shows the weights for dx_t, dx_{t-1} and dx_{t-2} respectively. A large X at the upstream gives more weight to the coefficients of lag 2. The coefficients from the two nonlinear models estimated above give the similar patterns.

ARCH effects are detected in the residuals of all fitted models including the fixed coefficient model. Using robust standard error estimates, the nonlinear estimates for the model, where dx_{t-4} and dx_{t-6} are included, are all significant at the 5% significance level except for those of α(t-ratio: -1.83) and dx_{t-4}(t-rtaio: 1.51).

The parameter estimates and robust standard errors

α	β_1	β_2	γ_0	γ_1	γ_2
-0.5814	-0.7231	0.7617	7.9572	0.8066	0.7629
(0.3178)	(0.2278)	(0.2607)	(0.0197)	(0.1341)	(0.1272)
γ_3	γ_4	γ_6	δ_1	δ_2	δ_3
0.9121	0.0366	0.0596	1.0665	-0.3256	0.1287
(0.0768)	(0.0242)	(0.0194)	(0.0631)	(0.0752)	(0.0384)

Note: The figures in parentheses indicate the robust standard errors.

This kind of nonlinear model adds insights unavailable with the fixed coefficient model.

Next, we analyze the relationship between the southern oscillation index(SOI) and associated recruitment(number of new fish) introduced in Shumway and Stoffer(1999). The data consist of 453 monthly observations ranging over the years 1950-1987(see Figure 2.11). The SOI measures changes in air pressure, related to sea surface temperatures in the central Pacific. The data have been analyzed using a standard transfer function model or frequency domain method in that text. We notice nonlinear behavior, as the relation tends to flatten out at both extremes at some lags(see Figure 2.12). We attempt to find a nonlinear relationship using model (2.3).

The cross-correlation of the two prewhitened series shows an apparent shift of $d = 5$ months and exponential decrease thereafter(see Figure 2.13). The SOI series, x_t, is detrended by a linear function of time as in Shumway and Stoffer(1999), and the recruitment series, y_t, is standardized. Both series are prewhitened using an AR(1) model.

Using a standard transfer function model, we obtain

$$\begin{aligned} y_t &= \underset{(0.0393)}{-0.7536}(1 - \underset{(0.0230)}{0.8157}B)^{-1}x_{t-5} + (1 - \underset{(0.0432)}{1.2647}B + \underset{(0.0432)}{0.4105}B^2)^{-1}e_t. \end{aligned}$$

The estimates are all significant and the mean square error is 0.0638.

As noted previously, there appear to exist nonlinear relationships at some lags such as $h = 5, \cdots, 10$. We suspect that there might be a different impact for y_t depending on the magnitude of x_{t-h} at a certain h. The estimated results using model (2.3) where $f(x) = x_{t-5}$ are quite satisfactory with mean square error 0.0595 and all coefficients significant at the 5% level. We have found a logistic relationship for the response of y_t which is consistent with the scatterplots of y_t and $x_{t-h}, h = 5, \cdots, 10$.

The estimated model is

$$
\begin{aligned}
y_t &= \underset{}{\frac{1}{\exp(\underset{(0.7601)}{-4.0422} - \underset{(0.8640)}{3.8784}x_{t-5}) + 1}}\Big[\underset{(0.0448)}{-0.8049}x_{t-5} \\
&- \underset{(0.0490)}{0.5647}x_{t-6} - \underset{(0.0519)}{0.4818}x_{t-7} - \underset{(0.0518)}{0.4771}x_{t-8} - \underset{(0.0505)}{0.3858}x_{t-9} \\
&- \underset{(0.0481)}{0.2596}x_{t-10} - \underset{(0.0387)}{0.1512}x_{t-11}\Big] + \eta_t,
\end{aligned}
$$

and

$$
\begin{aligned}
\eta_t &= \underset{(0.0434)}{1.3541}\eta_{t-1} - \underset{(0.0434)}{0.5036}\eta_{t-2} - \underset{(0.0634)}{0.1357}\eta_{t-4} \\
&+ \underset{(0.0797)}{0.3156}\eta_{t-5} - \underset{(0.0483)}{0.1830}\eta_{t-6}
\end{aligned}
$$

Figure 2.14 shows the behavior of $\rho(x_{t-5})$. The residuals from this fit have been passed to "SAS PROC ARIMA" to check the white noise assumption as before. We suggest subtracting 5 degrees of freedom for the χ^2 test corresponding to 5 autoregressive noise parameters. Unlike the constant coefficient results, we do not reject the white noise null hypothesis. Also, there does not seem to exist a serious ARCH effect as investigated by the ARCH test for residuals unlike the standard transfer function model.

If we add some linear lagged terms from x_{t-12} up to x_{t-15} into the nonlinear model, considering the slowly decaying term $(1 - 0.8157B)^{-1}x_{t-5}$ in the standard transfer function model, then we obtain the MSE 0.0554.

The following table shows the fitting results where "nlin (i) and (ii)" are the nonlinear models without and with added linear lagged terms in the model respectively. A few added X terms have resulted in the considerable improvement of the MSE. No serious serial correlation or ARCH effects are found in the residuals.

It seems to be reasonable that the nonlinear behavior is found in the first few cross-correlation lags rather than through all lags.

	α	β	MSE	ARCH(1)	ARCH(4)	ARCH(6)
trans			0.0638	0.0451	0.0123	0.0229
nlin (i)	-4.0442	-3.8784	0.0595	0.3935	0.3635	0.1195
	(0.7601)	(0.8640)				
nlin (ii)	-3.9763	-3.7706	0.0554	0.2852	0.3775	0.4853
	(0.6518)	(0.7469)				

Note: The figures in parentheses indicate the standard errors for the parameter estimates and ARCH(q) is the p-value for the ARCH LM test of no ARCH effects up to order q. 'trans' means the standard transfer function model and 'nlin' is the nonlinear model where more lagged X terms are not included.

Now, we analyze the relationship between precipitation and stream flow. The data consist of daily records of precipitation and stream flow at Tarboro and Kinston North Carolina from Aug 1, 1948 to July 31, 2002. There is a total of 19,723 paired observations of rainfall and stream flow for each region.

Averages of yearly precipitation at Tarboro and Kinston are 46.63 and 50.62 inches respectively(see Table 2.2 and 2.3). At Tarboro, the maximum yearly precipitation was 63.66 inches in 1999 when Hurricane Floyd passed through North Carolina. The minimum precipitation was 32.21 inches in 1988. For Kinston, 1971 has the biggest(67.10 inches) yearly precipitation and 1954 has the smallest(33.36 inches).

Both areas have about 110 days of rain per year. The year 1989 has the maximum number of rainy days in both areas. The average daily flows are 2,211.48 and 2,845.66 cubic feet per second respectively in Tarboro and Kinston. In Tarboro, the year of maximum average daily stream flow is 1999 with 4,276.09 cubic feet per second. The minimum, 614.09 cubic feet per second was in 1981. For Kinston, maximum and minimum stream flows are 4,860.86 and 1,066.71 cubic feet per second in 1989 and in 1986 respectively.

Seasonal patterns in environmental data such as precipitation and stream flow series are quite common. Figure 2.15 and 2.16 show the monthly pattern of rainfall and stream flow in both regions. As shown in the figures, there seem to be clear seasonal trends at both sites. The rainy days are frequent around summer. The average stream flows show relatively higher values in the spring and in September. Also, we find a very similar patten of yearly rainfall and stream flow in both areas(see Figure 2.17).

Statistical models typically require independent errors with homogeneous variance. This can not be reasonably assumed for some data including those discussed here. In our analysis, a logarithmic transformation has been made for stream flow data, and a square root transformation for precipitation data. Figure 2.18 shows QQ plots for transformed data. Most of the extreme deviations from a straight line for both rainfall and stream flows are caused by the observations in 1999 when Hurricane Floyd passed through North Carolina. The smaller outliers for stream flows occur mainly in 1954 and in 1968. The effect of having many rainless days is evident in the rainfall plots of Figure 2.18.

As noticed previously, there exist clear seasonal trends in all series. Periodogram analyses show that there exist cycles at about 182.5 day and 365 days for both transformed rainfall and stream flow series(see Figure 2.19 and 2.20). These are modeled by sinusoids providing deseasonalized residuals for both data sets. Using the deseasonalized data, we investigate the lag structure between rainfall and stream flows by calculating the cross-correlation of these two series.

Figure 2.21 shows the cross-correlations of deseasonalized stream flow and rainfall in Tarboro and Kinston. They are obtained from the prewhitened series(ARIMA (1,0,2) for Tarboro and ARIMA(2,0,1) for Kinston). As expected, stream flow is affected by current and lagged rain. The effects of rainfall on flow persist for several days. They initially increase then decrease as time passes by. The highest cross-correlations are at lag 4 and at lag 6 for Tarboro and for Kinston respectively. The lag effects can be shown in the phase spectra of rainfall and stream flow as well. Figure 2.22 shows the phase spectra of Tarboro and Kinston. In a pure delay model, the slope of the phase spectrum at low frequencies measures the time delay as a function of frequency, where the rainfall precedes the stream flow in this case. The slopes as frequency goes from 0 to 0.6 appear to be in the -3 to -4 range, a 3 or 4 day lag, for Tarboro and -5 to -6 range, a 5 or 6 day lag for Kinston, which seems consistent with our observed maximum cross-correlations.

Keeping these facts in mind, we analyze the data using a transfer function model. The goal is to predict deseasonalized stream flows S_t from deseasonalized rainfall R_t.

The estimated transfer function model for the Tarboro data is

$$\begin{aligned} S_t &= 0.2147(1 - 1.3690B - 0.4698B^2 + 0.6196B^3 + 1.0143B^4 \\ &- 0.7838B^5 - 0.1361B^6 - 0.1099B^7 + 0.3889B^8 - 0.1517B^9)^{-1} \\ & \quad (1 + 0.5361B - 0.6315B^2 - 0.7246B^3)R_t \\ &+ (1 - 2.1038B + 1.3539B^2 - 0.2487B^3)^{-1} \\ & \quad (1 + 0.5361B - 0.6315B^2 - 0.7246B^3)z_t, \end{aligned}$$

and for Kinston, it is

$$\begin{aligned} S_t &= 0.1027(1 - 2.5422B + 2.8076B^2 - 2.0561B^3 + 1.4405B^4 \\ &- 1.0998B^5 + 0.6949B^6 - 0.3704B^7 + 0.2331B^8 - 0.0985B^9)^{-1} \\ & \quad (1 - 0.6621B + 0.3902B^2 + 0.3532B^3)R_t \\ &+ (1 - 1.3203B + 0.3662B^2)^{-1}(1 + 0.3143B)z_t. \end{aligned}$$

The obtained mean square errors(MSEs) are 0.02661 and 0.01025 for Tarboro and Kinston respectively. The estimates are all significant at the 5% level.

On the other hand, we can also approximate the relationship using a typical linear regression.

$$S_t = \phi_0 R_t + \phi_1 R_{t-1} + \cdots + \phi_k R_{t-k} + \eta_t \tag{2.4}$$

where R_{t-k} is uncorrelated with η_t.

If the input, R_t, is white noise or prewhitened, then the cross-correlation function between the prewhitened input and correspondingly transformed output is directly proportional to the ϕ_js. Simply regressing S_t on R_t and its lags is an alternative, approximate method of identifying and estimating the relationship between S and R. Given that the weights ϕ_j decline quickly and sufficient lags are used in the regression, we get regression coefficients that approximate the impulse response function(Box, Jenkins and Reinsel, 1994).

For identifying the impulse response function here(see Figure 2.21), we have used ARIMA(1,0,2) and ARIMA(2,0,1) for prewhitening in each region. The coefficients are such that these can be closely approximated by an MA(1) with coefficient 0.174 and 0.168 for Tarboro and for Kinston respectively. These amounts of autocorrelation do not much affect the relationship between rainfall and stream flow.

The model requires selection of the point k beyond which the cross-correlation is effectively zero and η_t can be serially correlated. We estimate the series using a typical linear regression method.

Based on the cross-correlation structure of the prewhitened series, the Akaike information criterion(AIC), and the Schwartz Bayesian criterion(SBC), k is chosen to be 24 for Tarboro and 29 for Kinston. The residuals from the initial models in both areas are serially correlated so refined estimates are obtained adjusting for autocorrelation of residuals. The patterns of the estimated coefficients look like the cross-correlations between the two series.

As mentioned previously, there might be a different lag structure when there are heavy rains. We can estimate a different response for different input levels just by adding some indicator variables. We expect the relationship between rainfall and stream flow will be strong with heavy rains and add indicator variables to catch the additional effects of rainfall on stream flow. To capture this effect, the model we consider is

$$\begin{aligned} S_t &= \phi_0 R_t + \phi_1 R_{t-1} + \cdots + \phi_k R_{t-k} \\ &+ \gamma_0 IR_t + \gamma_1 IR_{t-1} + \cdots + \gamma_q IR_{t-q} + \eta_t \end{aligned} \tag{2.5}$$

where $IR_t = R_t - R_0$ for a "heavy rain" period, and $IR_t = 0$ for a "low rain" period.

Heavy rain is arbitrarily defined by letting R_0 be the $95th$ percentile of the precipitation data. Most of the major Hurricanes and tropical storms which pass through both regions are included here. Rainfall data on days of high rainfall and the subsequent 24 days for Tarboro and 29 days for Kinston are used for a "heavy rain" period, so that immediate and lagged effects can be considered.

The estimated coefficients for Tarboro are all significant at the 5% level and the MSE is 0.02725 without indicator variables(model (2.4)) and 0.02595 with indicator variables(model (2.5)). Table 2.4 shows the coefficients from the fitted models. Elimination of insignificant indicator variables leads to a model with $q = 8$. The MSE is less than that of the previous transfer function model, 0.02661 with indicator variables included. The rainfall coefficient is highest at lag 4 and the "heavy rain" period shows some difference up to $q = 8$ days after rainfall depending on the size of $(R_t - R_0)$.

We test $H_0 : \gamma_0 = \gamma_1 = \cdots = \gamma_{q=8} = 0$ using F statistics.

$$F = \frac{(SSE_{red} - SSE_{ful})/(dfe_{red} - dfe_{ful})}{SSE_{ful}/dfe_{ful}}.$$

The calculated F statistic is 90.85 and the p-value is < 0.0001. Thus we reject H_0 at the 5% significance level.

For Kinston, all coefficients are also significant and different effects of rainfall exist up to $q = 4$ days after "heavy" precipitation. The highest coefficient is at lag 6. The MSEs are 0.01035 without indicator variables(model (2.4)) and 0.01000 with indicator variables(model (2.5)). The F statistic and p-value for Kinston are 114.90 and < 0.0001 respectively. Table 2.4 shows the coefficients from the fitted models.

The previous analysis shows that there exist different effects depending on the period of rainfall. We expect that the responses of stream flows will differ linearly or nonlinearly based on the amount of rainfall several days prior to the date of measured stream flow.

We did a linearity test based on Fuller(1996) using the quadratic terms up to lag 3 as explanatory variables as well as linear terms, expecting that the nonlinear behavior will be centered on a few low lag terms rather than through all lags. The F statistic for the test that the coefficients of all quadratic terms are zero is 37.50 for Tarboro. The F statistic for Kinston is 10.19. The linearity hypothesis is rejected at the 5% level in both areas.

Now we fit the following nonlinear model

$$S_t = \rho(t)[R_t + \phi_1 R_{t-1} + \cdots + \phi_q R_{t-q}] + \phi_{q+1} R_{t-q-1} + \cdots + \eta_t$$

where $\rho(t) = \frac{1}{\exp(\alpha+\beta f(R))+1}$ for the series. Our intent is to replace the lag $0, \cdots, q$ coefficients in model (2.5), those for which our indicators have been significant, with nonlinear coefficients.

Where $q = 1$, the nonlinear model with a smaller MSE is obtained. The variable

R_t has worked well as a transition variable in the logistic function, i.e., $f(R) = R_t$. The estimates are all significant at the 5% level and the obtained MSE is 0.02683 for Tarboro.

$$\begin{aligned} S_t &= \frac{1}{\exp(1.5172 - 0.2143R_t) + 1}\Big[R_t + 2.1693R_{t-1}\Big] \\ &+ 0.5053R_{t-2} + \cdots + 0.00913R_{t-23} + \eta_t, \end{aligned}$$

and

$$\begin{aligned} \eta_t &= 1.3230\eta_{t-1} - 0.4810\eta_{t-2} + 0.1015\eta_{t-3} - 0.0293\eta_{t-5} \\ &+ 0.0487\eta_{t-6} + e_t. \end{aligned}$$

Table 2.4 shows all estimated coefficients from the fitted model.

For Kinston, we have obtained

$$\begin{aligned} S_t &= \frac{1}{\exp(2.3351 - 0.1559R_t) + 1}\Big[R_t + 2.0907R_{t-1}\Big] \\ &+ 0.2322R_{t-2} + \cdots + 0.00869R_{t-25} + \eta_t, \end{aligned}$$

and

$$\begin{aligned} \eta_t &= 1.6445\eta_{t-1} - 0.8831\eta_{t-2} + 0.2669\eta_{t-3} - 0.0861\eta_{t-4} \\ &+ 0.0346\eta_{t-6} + e_t. \end{aligned}$$

The estimates, which are shown in Table 2.4, are all significant at the 5% level and the MSE is 0.010207.

We have found a nonlinear relationship for the stream flows depending on the amount of rainfall a few days prior to that of the measured stream flow. However, the performance of the nonlinear models does not seem to be superior to some other suggested models in these examples.

Table 2.1: Distribution of $\hat{\gamma}, \hat{\alpha}$, and $\hat{\beta}$ for model (2.1)

(i) $\gamma = 1.0, \alpha = 0.8$, $\beta = 0.5$. $x_t \sim N(0,1)$ and $e_t \sim N(0, 0.04)$

		$n = 1,000$	$n = 2,000$	$n = 3,000$
$\hat{\gamma}$	bias	0.0916	0.0301	0.0178
	STD	0.4451	0.1709	0.1284
		(0.1843)	(0.1303)	(0.1064)
	skewness	8.6066	2.2637	1.5333
	JB	>1,000	>1,000	>1,000
	$Pr(\|t\| > z_{0.025})$	0.0878	0.0640	0.0604
$\hat{\alpha}$	bias	0.0463	0.0181	0.0094
	STD	0.3713	0.2286	0.1846
		(0.2934)	(0.2075)	(0.1694)
	skewness	1.2511	0.6887	0.5556
	JB	>1,000	855.62	453.98
	$Pr(\|t\| > z_{0.025})$	0.0459	0.0416	0.0488
$\hat{\beta}$	bias	0.0034	0.0020	0.0016
	STD	0.0672	0.0452	0.0365
		(0.0614)	(0.0434)	(0.0354)
	skewness	0.2273	0.1430	0.0262
	JB	53.73	19.82	0.62
	$Pr(\|t\| > z_{0.025})$	0.0529	0.0536	0.0502
	R	4,991	5,000	5,000

(ii) $\gamma = 1.0, \alpha = 0.8,\ \beta = 0.5.$

$x_t = 0.6x_{t-1} + \mu_t, \mu_t \sim N(0,1)$ and $e_t \sim N(0, 0.04)$

		$n = 1,000$	$n = 2,000$	$n = 3,000$
$\hat{\gamma}$	bias	0.0168	0.0040	0.0031
	STD	0.1193	0.0715	0.0564
		(0.0926)	(0.0655)	(0.0535)
	skewness	1.8222	0.7375	0.4790
	JB	>1,000	786.01	298.92
	$Pr(\|t\| > z_{0.025})$	0.0582	0.0488	0.0546
$\hat{\alpha}$	bias	0.0097	0.0001	0.0006
	STD	0.1773	0.1152	0.0924
		(0.1540)	(0.1089)	(0.0889)
	skewness	0.4923	0.2665	0.1031
	JB	586.46	89.68	22.25
	$Pr(\|t\| > z_{0.025})$	0.0454	0.0432	0.0508
$\hat{\beta}$	bias	0.0017	0.0018	0.0011
	STD	0.0394	0.0268	0.0218
		(0.0367)	(0.0259)	(0.0212)
	skewness	0.1238	0.0738	0.0876
	JB	19.85	4.78	12.88
	$Pr(\|t\| > z_{0.025})$	0.0492	0.0466	0.0518
	R	5,000	5,000	5,000

Note: The figures in parentheses show the theoretical standard error of parameters in $B^{-1}(\theta_0)\sigma^2$ which is numerically evaluated using SAS integration procedure.

Table 2.2: The summary of yearly precipitation and stream flow(1949-2001)

		Tarboro	Kinston
Precipitation (inches per year)	mean	46.63	50.62
	std	7.35	7.03
	min	32.21	33.36
	max	63.66	67.10
Rainy days (per year)	mean	112.72	109.70
	std	13.02	17.52
	min	88.00	67.00
	max	144.00	142.00
Stream flow	mean	2211.48	2845.66
	std	770.79	940.02
	min	614.09	1066.71
	max	4276.09	4860.86

Note: Stream flows are recorded as daily cubic feet per second.

Table 2.3: Major Hurricane or tropical storm passing through both regions

			Tarboro		Kinston	
year	period	name	excess	max	excess	max
1954	Oct.5-18	Hurricane HAZEL	0.14	1.20	0.39	1.28
1955	Aug.7-21	Hurricane DIANE	0.77	4.75	1.30	7.83
	Sep.10-24	Hurricane IONE	0.64	3.10	1.79	4.31
1956	Sep.21-30	Hurricane FLOSSY	0.84	3.40	-0.13	1.06
1959	Jul.5-12	Hurricane CINDY	0.09	0.70	-0.13	1.16
1960	Jul.28-Aug.1	Hurricane BRENDA	1.87	4.49	1.27	3.98
	Aug.29-Sep.14	Hurricane DONNA	1.10	4.78	1.11	6.82
1964	Aug.20-Sep.5	Hurricane CLEO	-0.04	1.21	0.21	1.55
1965	Jun.11-18	Tropical storm 1	0.67	2.89	0.33	2.01
1971	Sep.6-Oct.5	Hurricane GINGER	0.51	5.09	0.41	4.05
1972	Jun.14-23	Hurricane AGNES	0.02	0.94	0.30	1.36
1981	Aug.7-22	Hurricane DENIS	0.24	2.31	0.31	3.44
1984	Sep.8-16	Hurricane DIANA	0.50	4.46	0.16	2.20
1996	Jul.5-17	Hurricane BERTHA	0.33	4.00	0.40	4.58
	Aug.23-Sep.10	Hurricane FRAN	0.38	2.21	0.45	5.69
1997	Jul.16-27	Hurricane DANNY	-0.02	1.73	0.31	2.07
1998	Aug.31-Sep.8	Hurricane EARL	0.02	1.50	0.33	1.97
1999	Aug.24-Sep.8	Hurricane DENNIS	0.39	2.81	0.32	5.50
	Sep.7-Sep.19	Hurricane FLOYD	2.61	7.11	1.74	11.80
	Oct.12-19	Hurricane IRENE	1.08	4.29	0.94	4.02
2000	Sep.15-25	Tropical storm HELENE	-0.01	1.05	0.34	1.62

Note: "excess" means the average excess rainfall by the day over the monthly average rainfall by the day where precipitation is bigger than 0 and "max" shows the maximum daily precipitation over the period.

Table 2.4: The coefficient estimates from the fitted models

	Tarboro			Kinston		
	reg(i)	reg(ii)	nlin	reg(i)	reg(ii)	nlin
α			1.5172			2.3351
β			-0.2143			-0.1559
ϕ_0	0.2098	0.1189		0.1020	0.0619	
ϕ_1	0.3992	0.3030	2.1693	0.1906	0.1463	2.0907
ϕ_2	0.5092	0.4402	0.5053	0.2388	0.2056	0.2322
ϕ_3	0.6005	0.5480	0.5959	0.3168	0.3001	0.3080
ϕ_4	0.6019	0.5574	0.5979	0.3793	0.3726	0.3699
ϕ_5	0.5473	0.5065	0.5434	0.4033	0.4008	0.3934
ϕ_6	0.4812	0.4473	0.4757	0.4035	0.4009	0.3929
ϕ_7	0.4106	0.3830	0.4044	0.3867	0.3845	0.3753
ϕ_8	0.3484	0.3323	0.3413	0.3575	0.3560	0.3450
ϕ_9	0.3015	0.2956	0.2934	0.3269	0.3251	0.3128
ϕ_{10}	0.2643	0.2571	0.2554	0.2948	0.2927	0.2794
ϕ_{11}	0.2263	0.2188	0.2166	0.2658	0.2627	0.2485
ϕ_{12}	0.2064	0.1991	0.1961	0.2375	0.2346	0.2189
ϕ_{13}	0.1814	0.1750	0.1717	0.2122	0.2093	0.1921
ϕ_{14}	0.1584	0.1538	0.1489	0.1880	0.1854	0.1665
ϕ_{15}	0.1421	0.1377	0.1323	0.1684	0.1662	0.1452
ϕ_{16}	0.1231	0.1189	0.1133	0.1519	0.1497	0.1277
ϕ_{17}	0.1066	0.1021	0.0967	0.1360	0.1333	0.1103
ϕ_{18}	0.0890	0.0854	0.0798	0.1224	0.1206	0.0961
ϕ_{19}	0.0741	0.0710	0.0652	0.1135	0.1121	0.0861
ϕ_{20}	0.0648	0.0634	0.0561	0.1021	0.1005	0.0734

Note: "reg(i)" and "reg(ii)" are the regression models without and with the indicator variables respectively. "nlin" means the nonlinear model.

	Tarboro			Kinston		
	reg(i)	reg(ii)	nlin	reg(i)	reg(ii)	nlin
ϕ_{21}	0.0538	0.0522	0.0435	0.0898	0.0876	0.0590
ϕ_{22}	0.0429	0.0405	0.0321	0.0791	0.0778	0.0463
ϕ_{23}	0.0209	0.0186	0.00913	0.0661	0.0654	0.0312
ϕ_{24}	0.009428	0.008867		0.0568	0.0564	0.0206
ϕ_{25}				0.0452	0.0455	0.00869
ϕ_{26}				0.0333	0.0330	
ϕ_{27}				0.0231	0.0223	
ϕ_{28}				0.0141	0.0133	
ϕ_{29}				0.006109	0.005778	
γ_0		0.5413			0.2319	
γ_1		0.5876			0.2645	
γ_2		0.4132			0.1876	
γ_3		0.3049			0.0829	
γ_4		0.2500			0.0231	
γ_5		0.2239				
γ_6		0.1759				
γ_7		0.1370				
γ_8		0.0690				
MSE	0.02725	0.02595	0.02683	0.01035	0.01000	0.01021

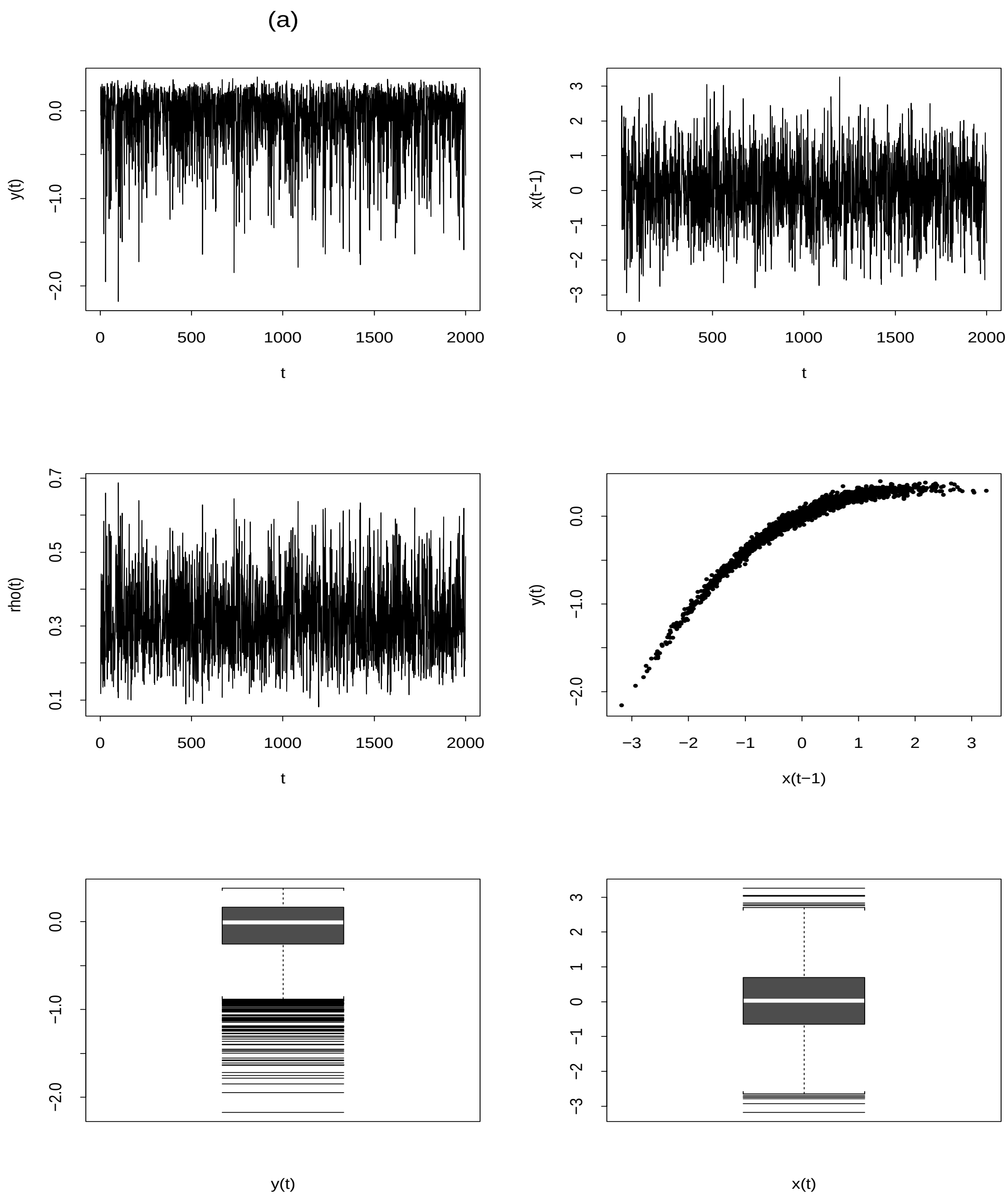


Figure 2.1: y_t and x_t where (a) $(\gamma, \alpha, \beta, \sigma) = (1.0, 0.8, 0.5, 0.04)$

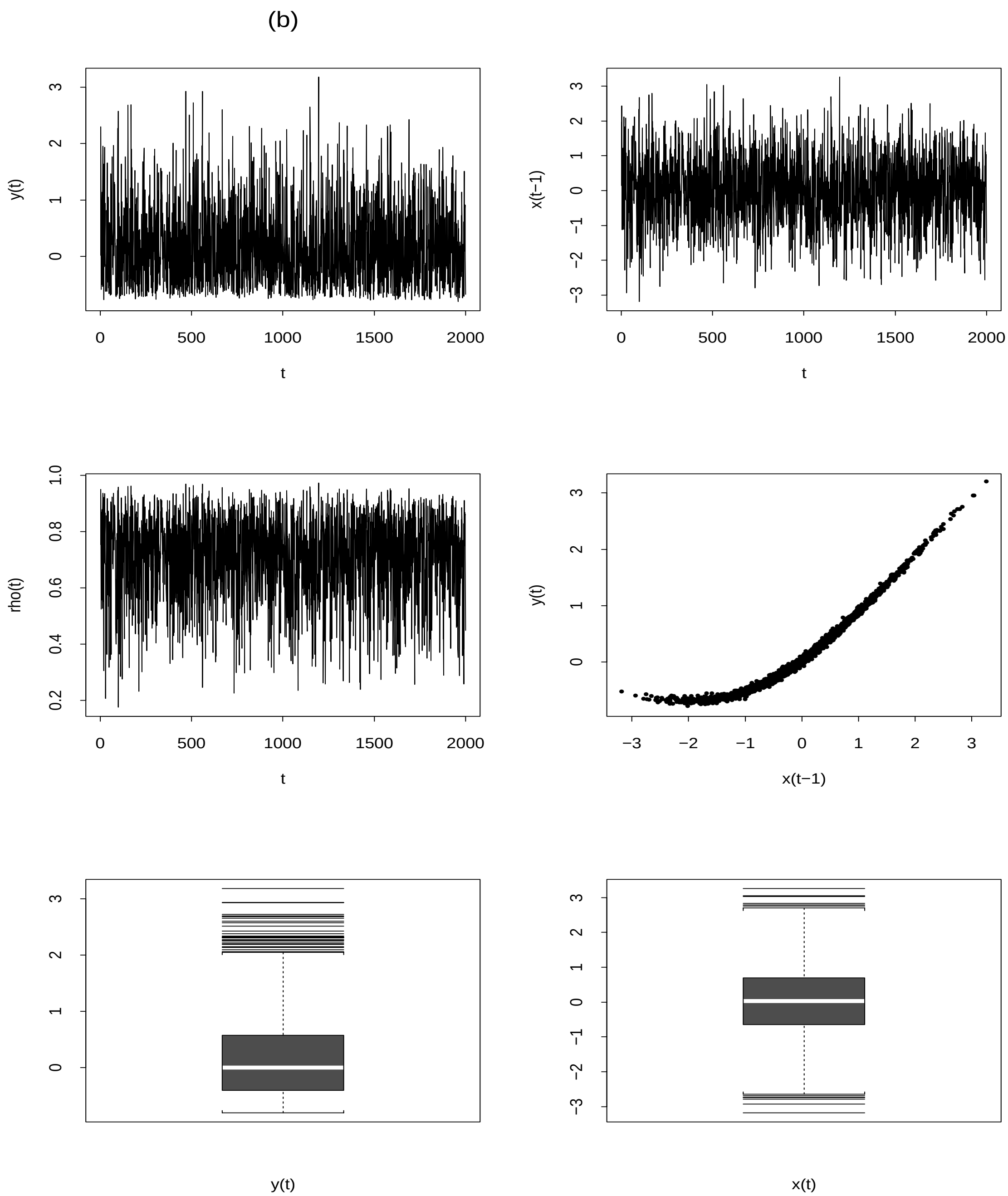
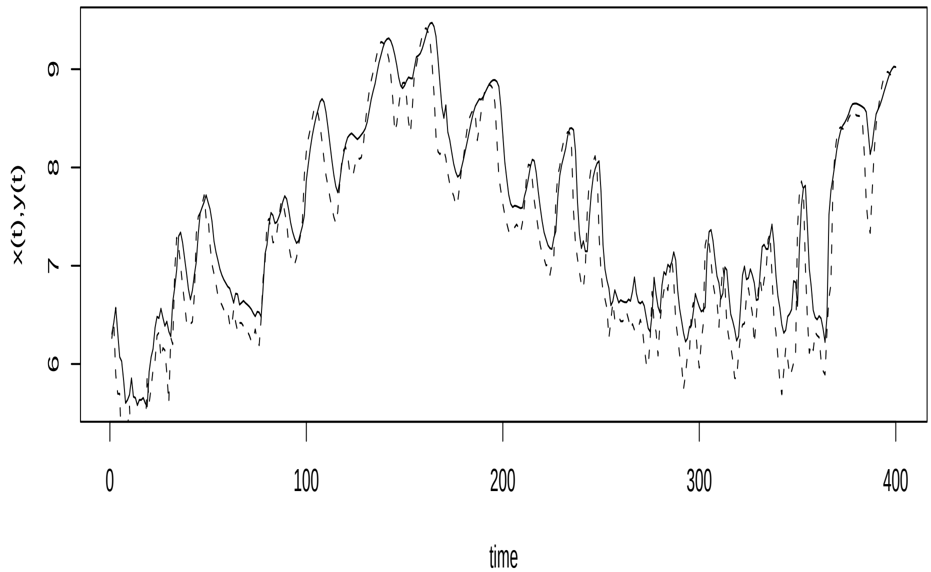


Figure 2.2: y_t and x_t where (b) $(\gamma, \alpha, \beta, \sigma) = (1.0, -1.0, -0.8, 0.04)$

(a)



(b)

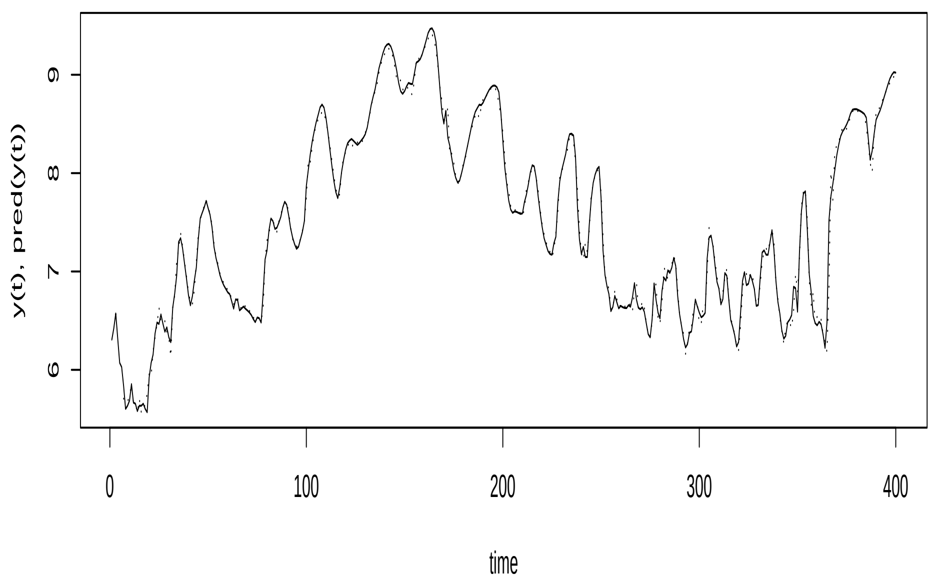


Figure 2.3: (a) the log transformed stream flow series of Goldsboro(dotted line) and Kinston(solid line), (b) prediction(dotted line) and actual series(solid line) for Kinston

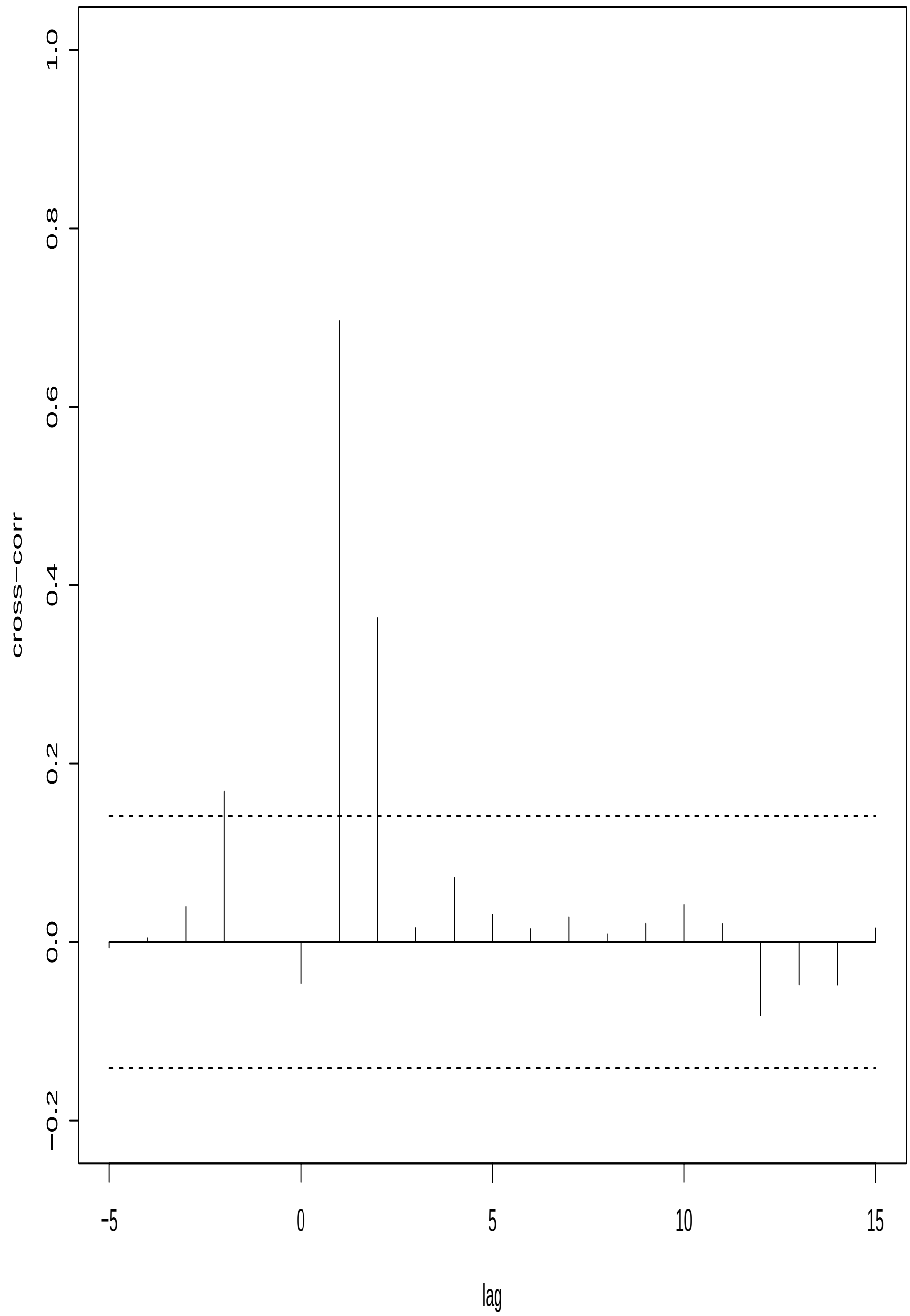


Figure 2.4: The cross-correlation of ∇y_t and ∇x_t

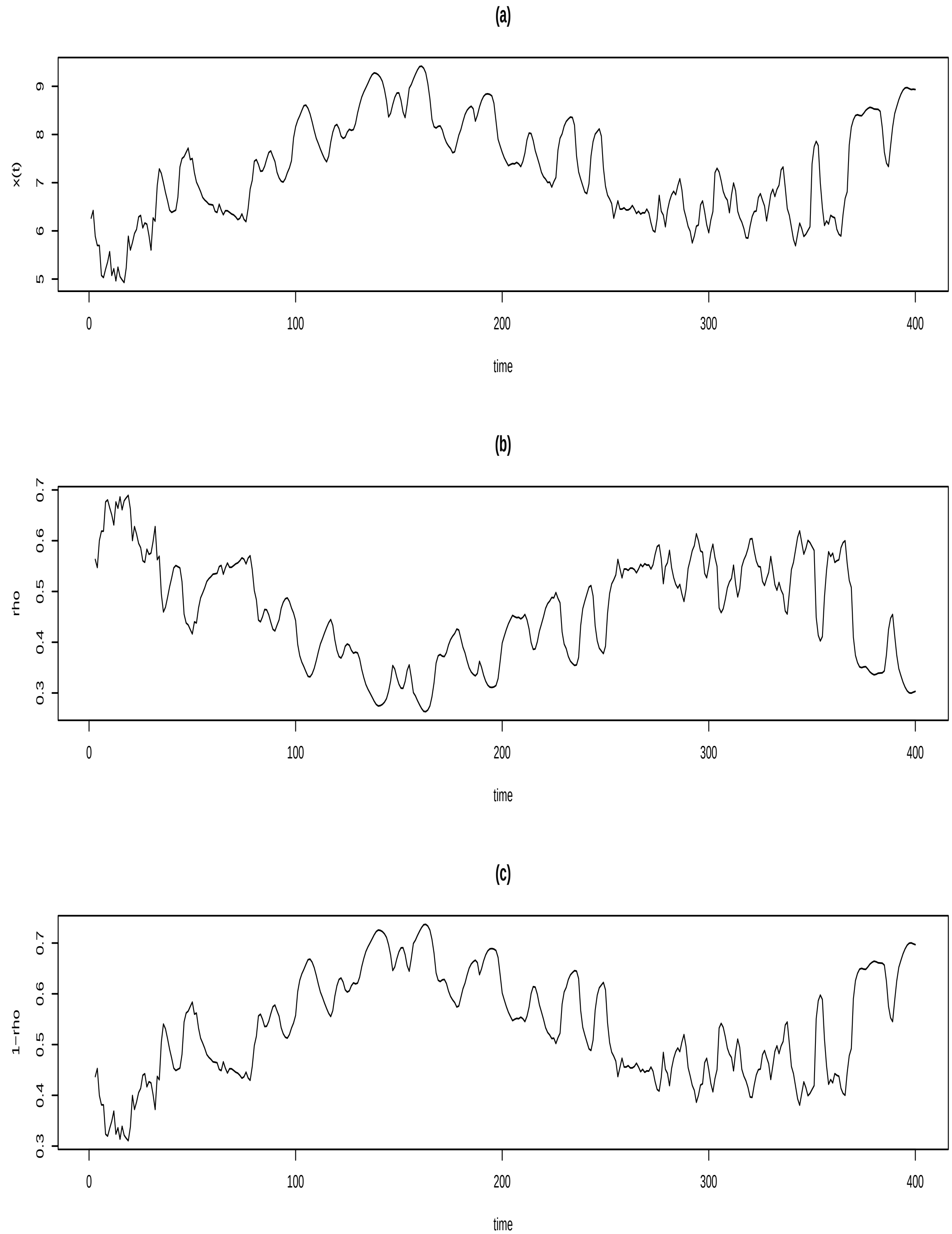


Figure 2.5: (a) the log transformed stream flow series of Goldsboro, (b) $\rho(dx_{t-2})$ vs time, (c)$1-\rho(dx_{t-2})$ vs time

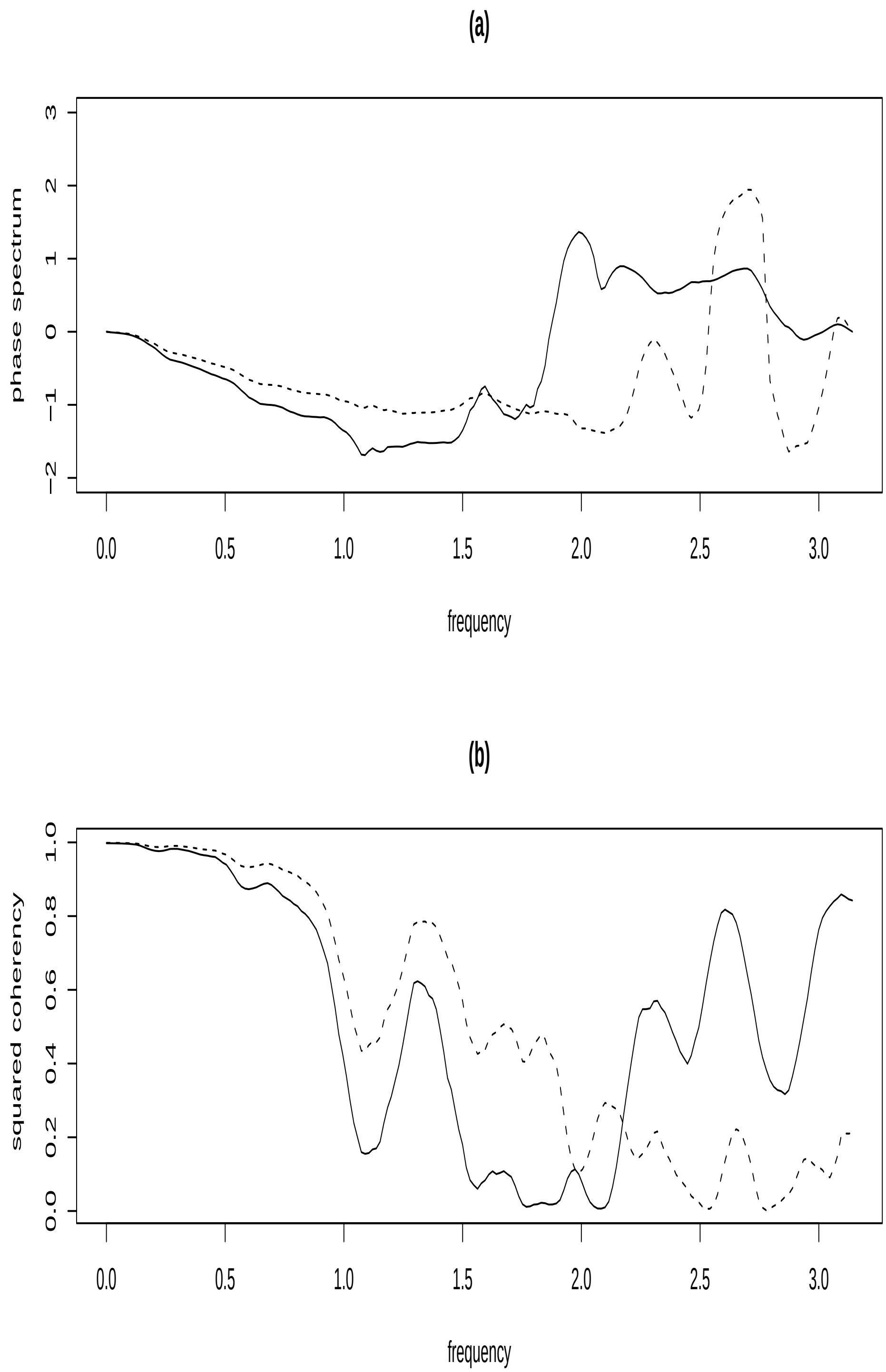


Figure 2.6: (a) phase spectrum and (b) squared coherency of two log transformed stream flow series at a high flow period(solid line) and a low flow period(dotted line)

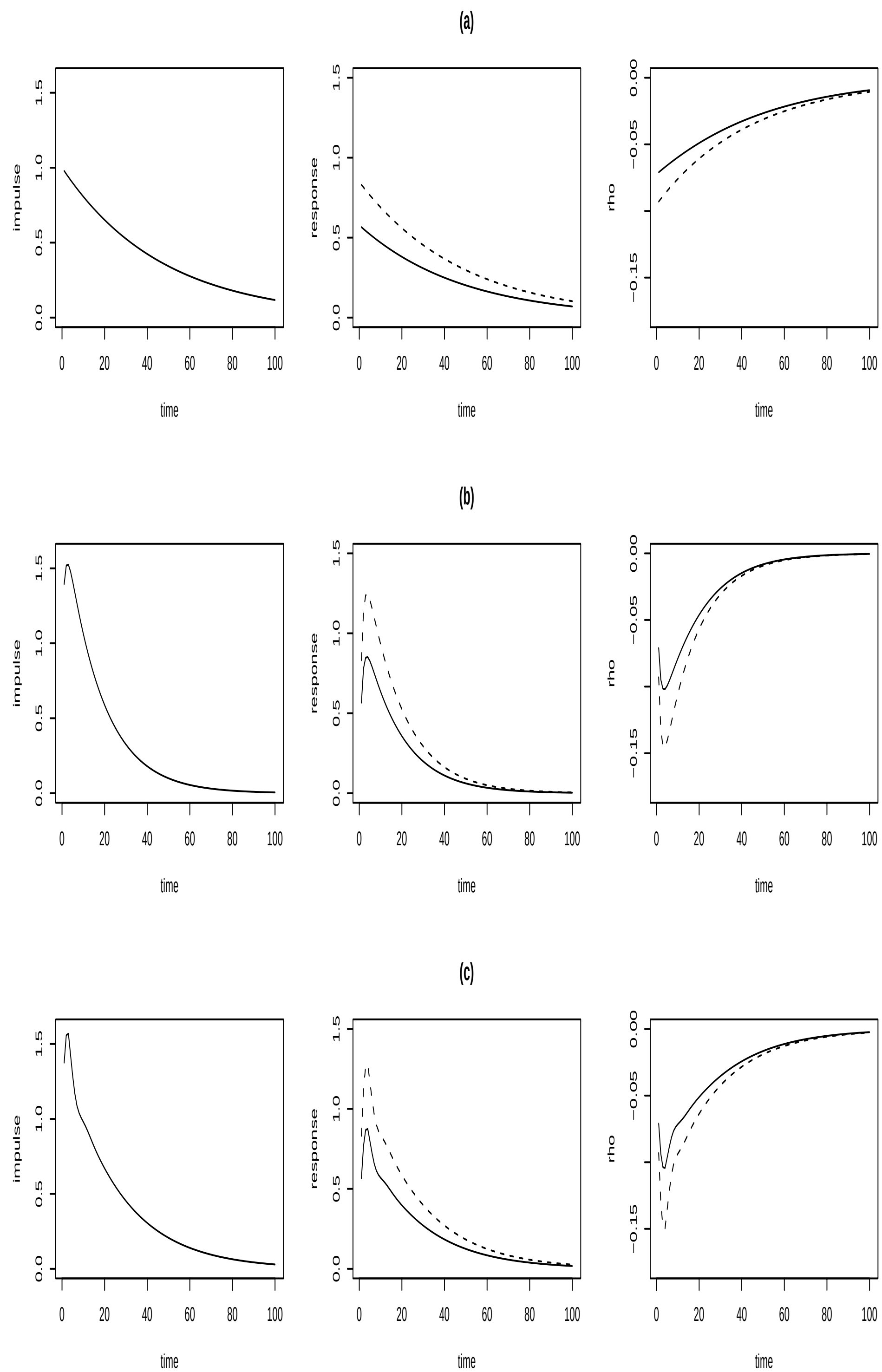


Figure 2.7: The response deviations and the changes in $\rho(dx_{t-2})$ from the equilibrium state by the given impulse of (a) AR(1), (b) AR(2), and (c) AR(5), at a high flow(solid line) and a low flow(dotted line)

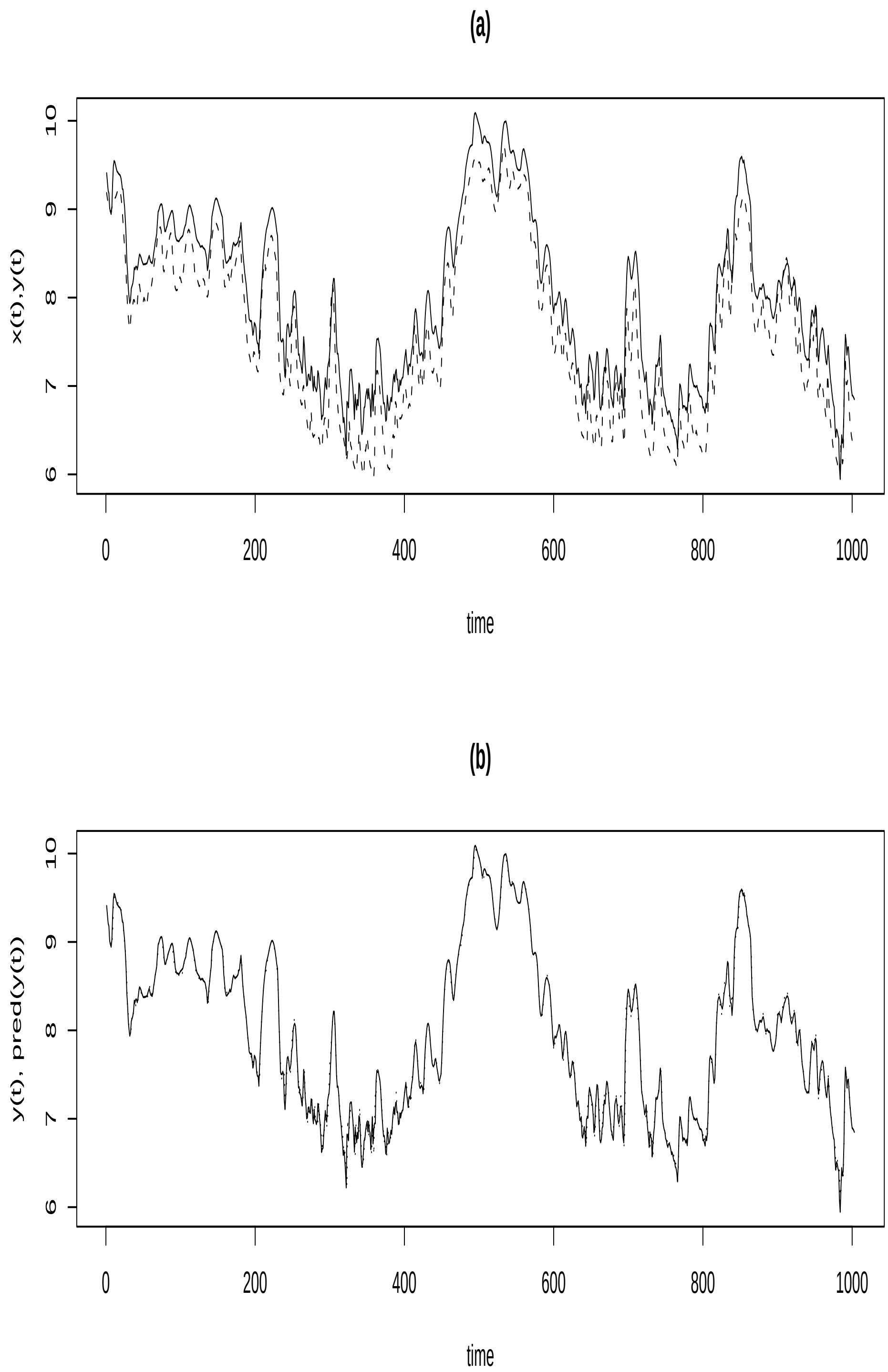


Figure 2.8: (a) the log transformed stream flow series of Kinston(dotted line) and Fort Barnwell(solid line), (b) prediction(dotted line) and actual series(solid line) for Fort Barnwell

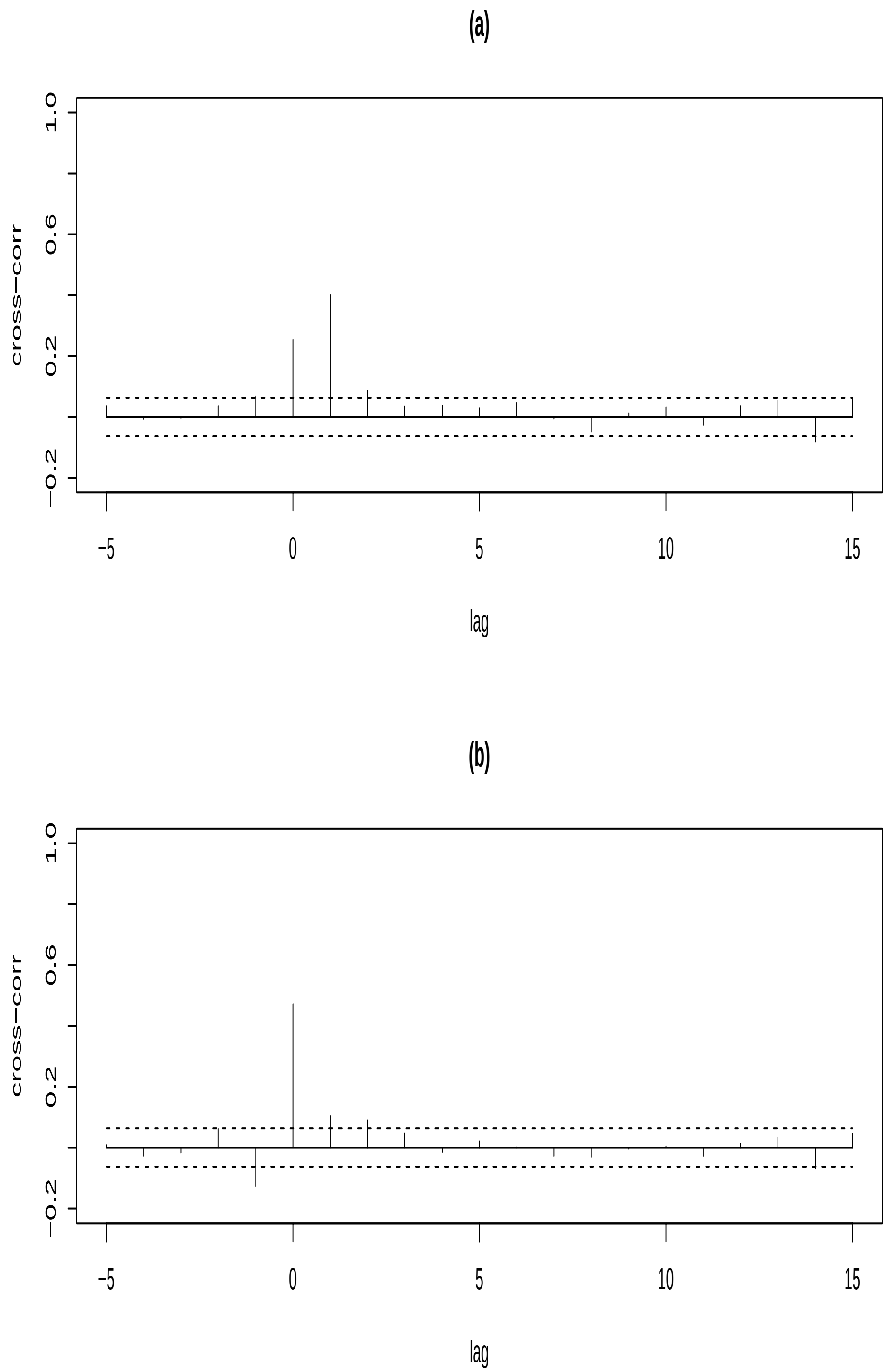


Figure 2.9: (a) the cross-correlation of ∇y_t and ∇x_t, (b) the cross-correlation of y_t and x_t

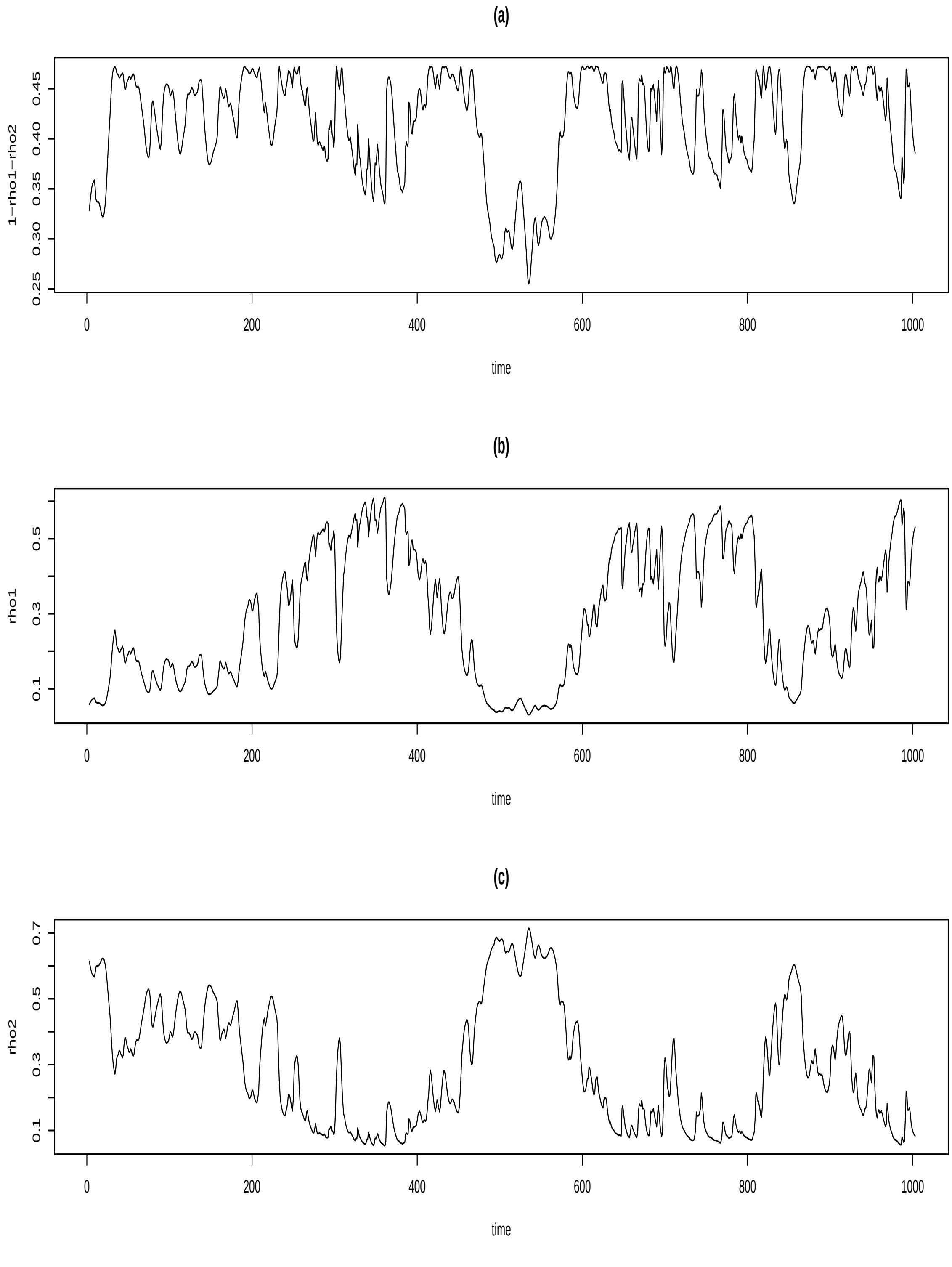


Figure 2.10: (a) $1-\rho_1(dx_{t-2})-\rho_2(dx_{t-2})$ vs time, (b) $\rho_1(dx_{t-2})$ vs time, (c)$\rho_2(dx_{t-2})$ vs time

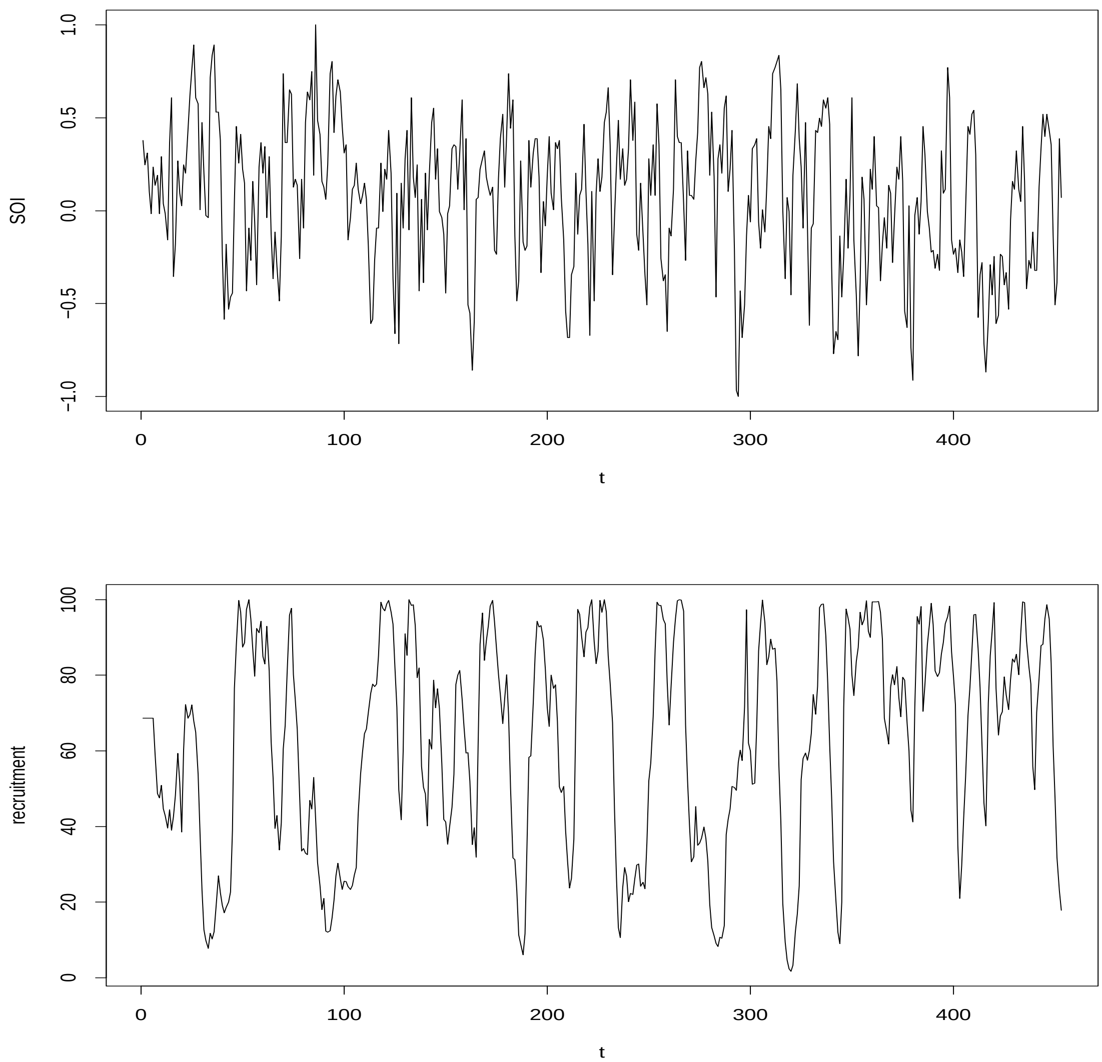


Figure 2.11: SOI and recruitment series

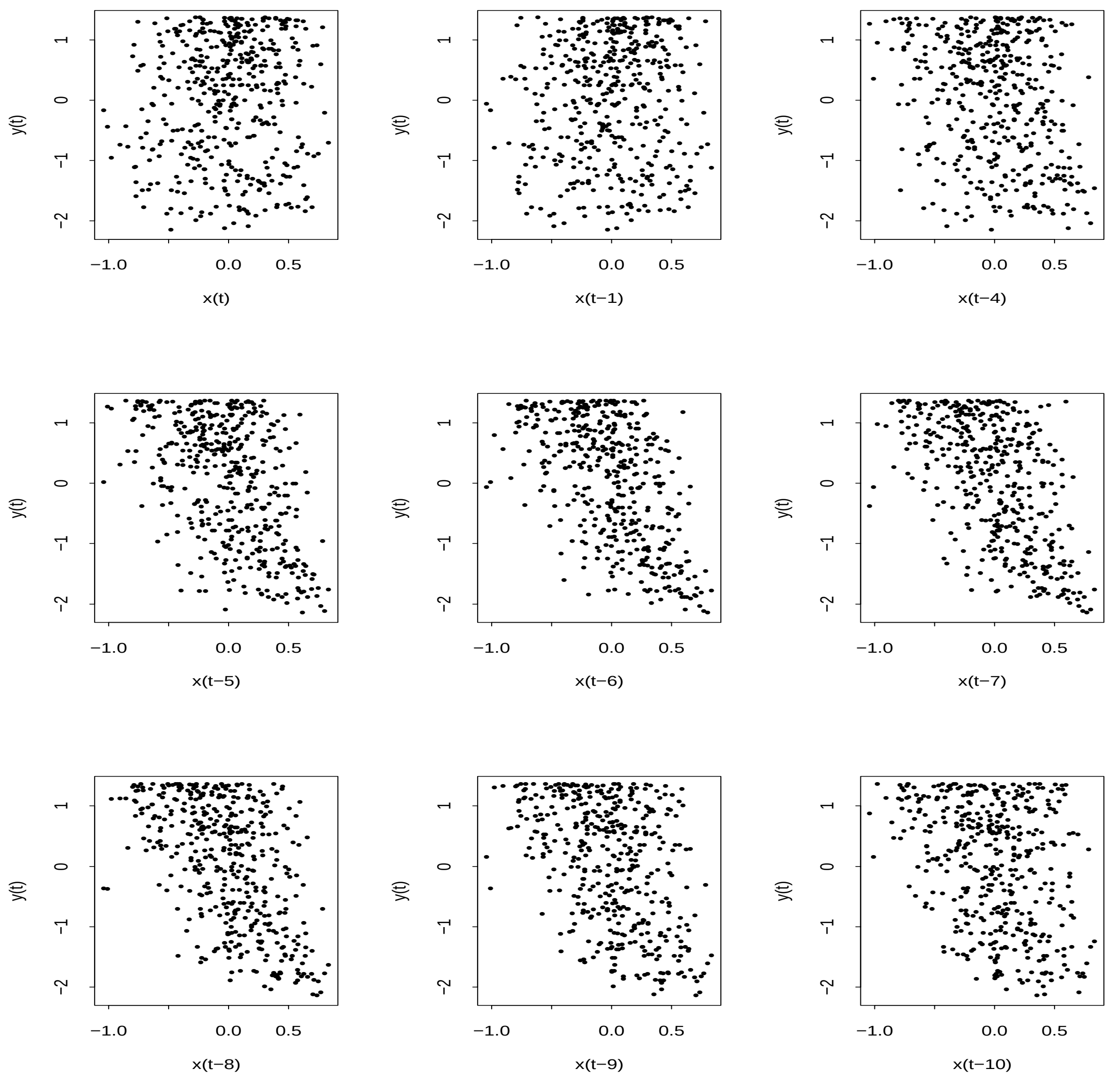


Figure 2.12: Scatterplot of recruitment y_t with lagged SOI data, $x_{t-h}, h = 0, 1, 4 \cdots, 10$

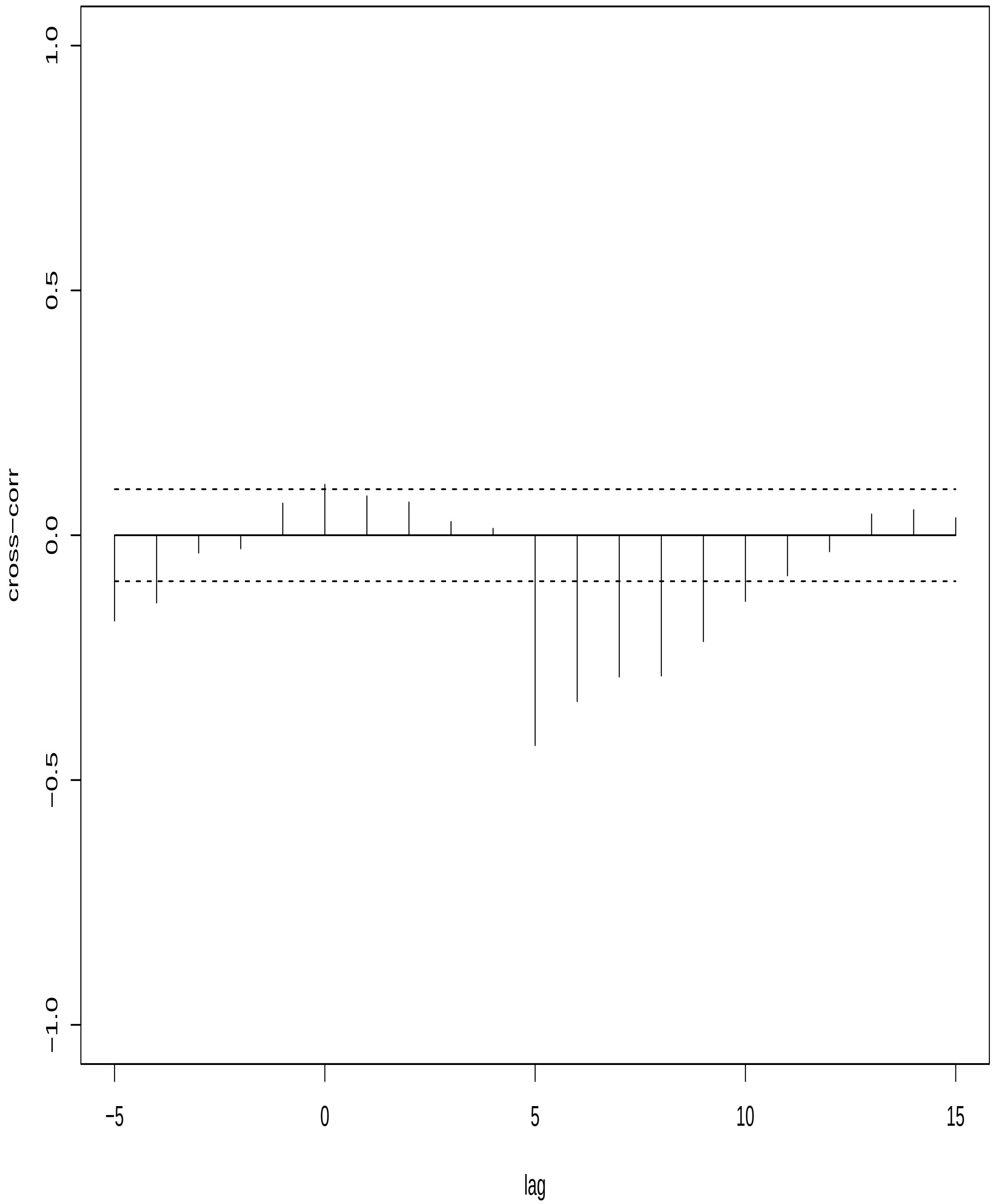
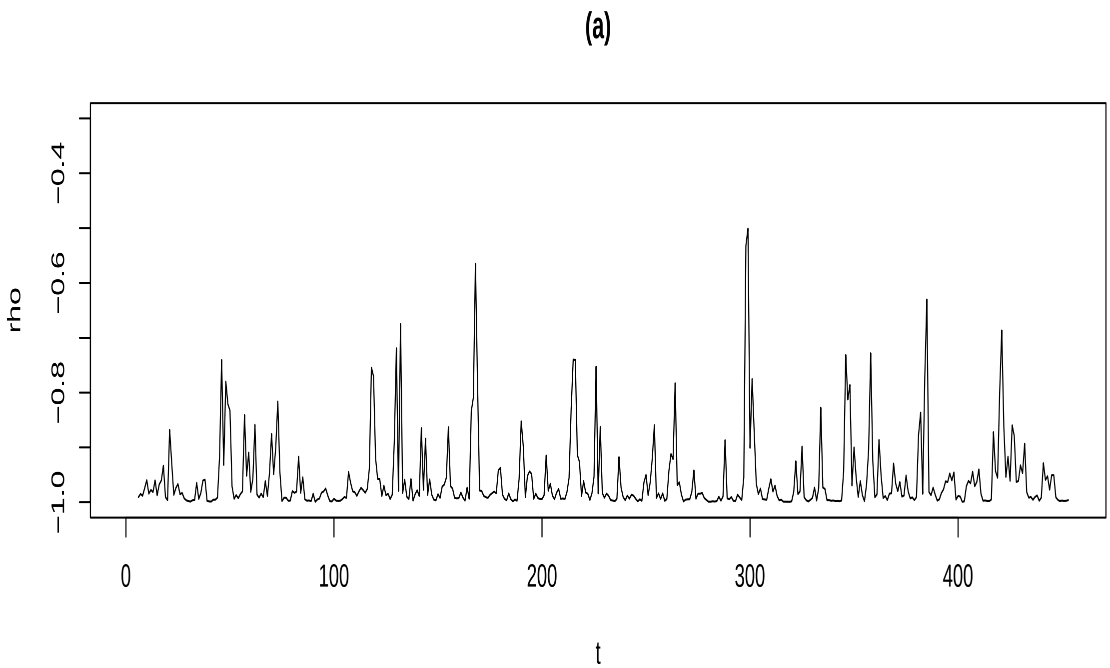


Figure 2.13: The cross-correlation of y_t and x_t



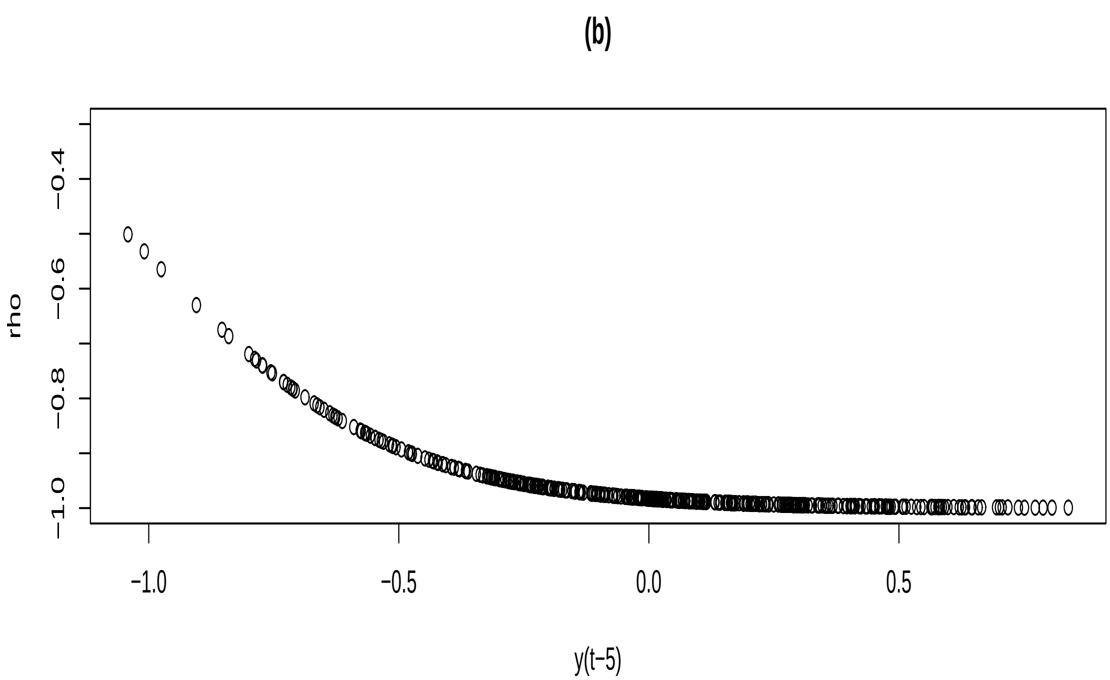


Figure 2.14: (a) $\rho(x_{t-5})$ vs time, (b) $\rho(x_{t-5})$ vs x_{t-5}

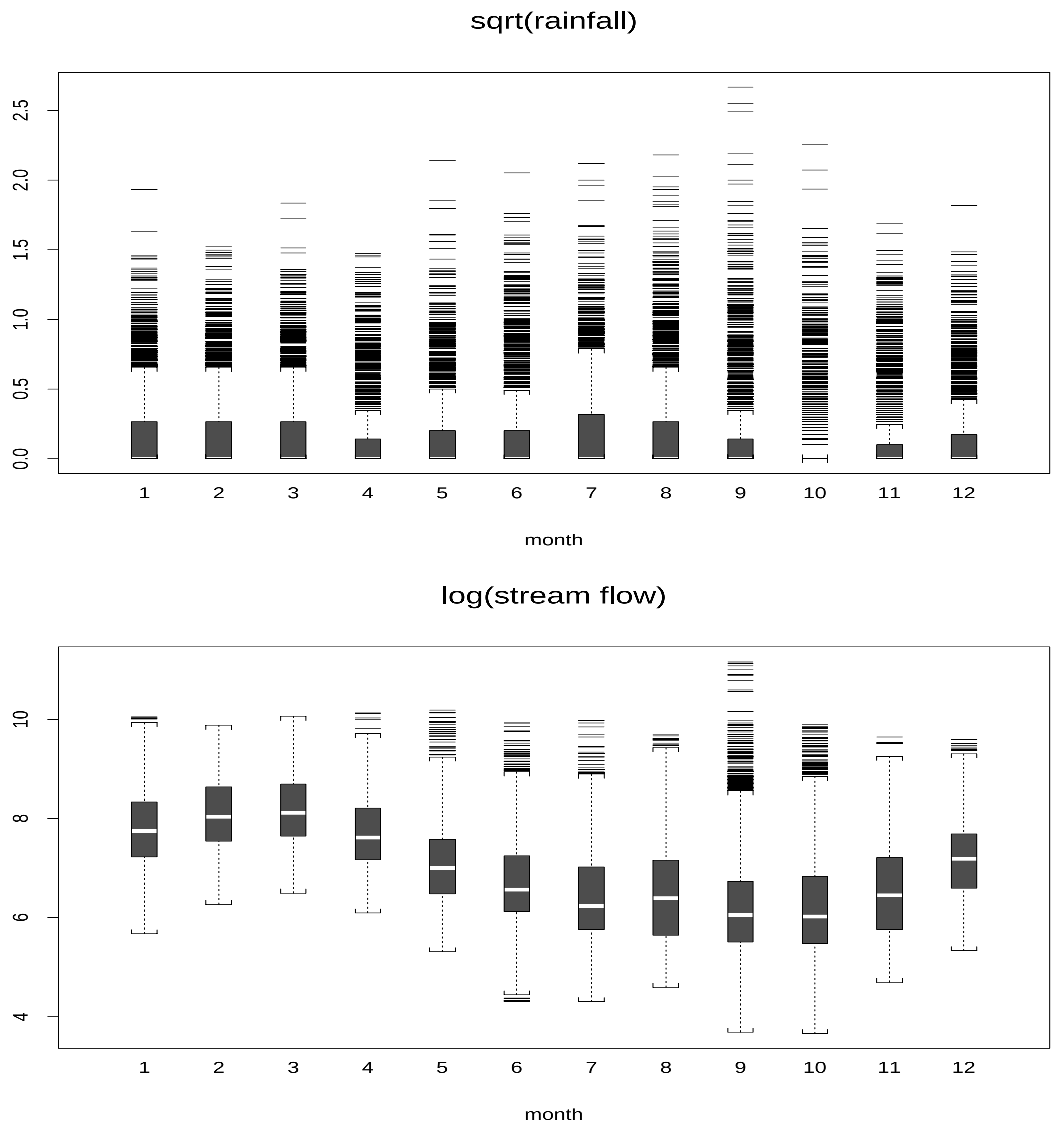


Figure 2.15: Boxplot of rainfall and stream flow in Tarboro(on a daily basis)

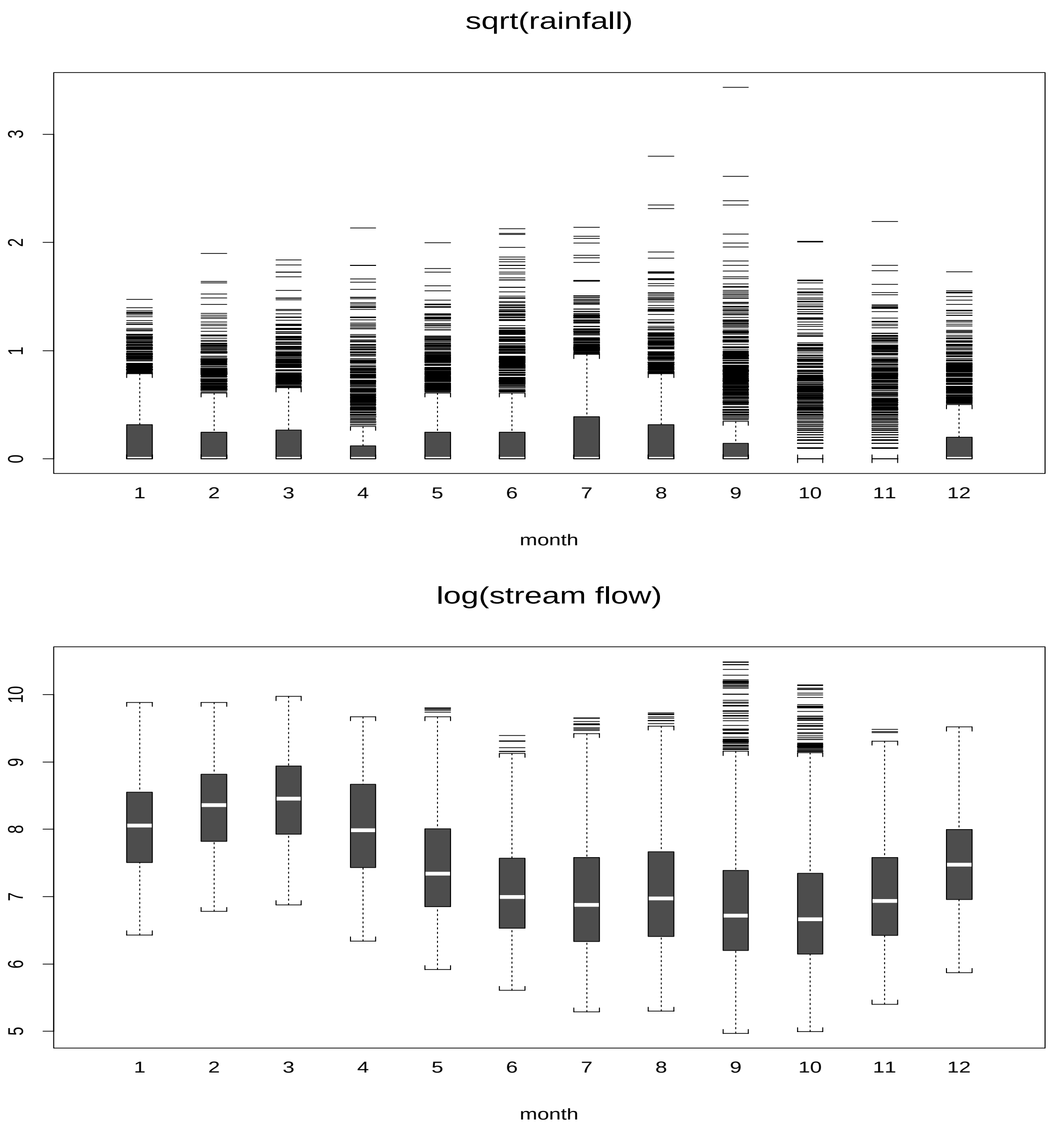


Figure 2.16: Boxplots of rainfall and stream flow in Kinston(on a daily basis)

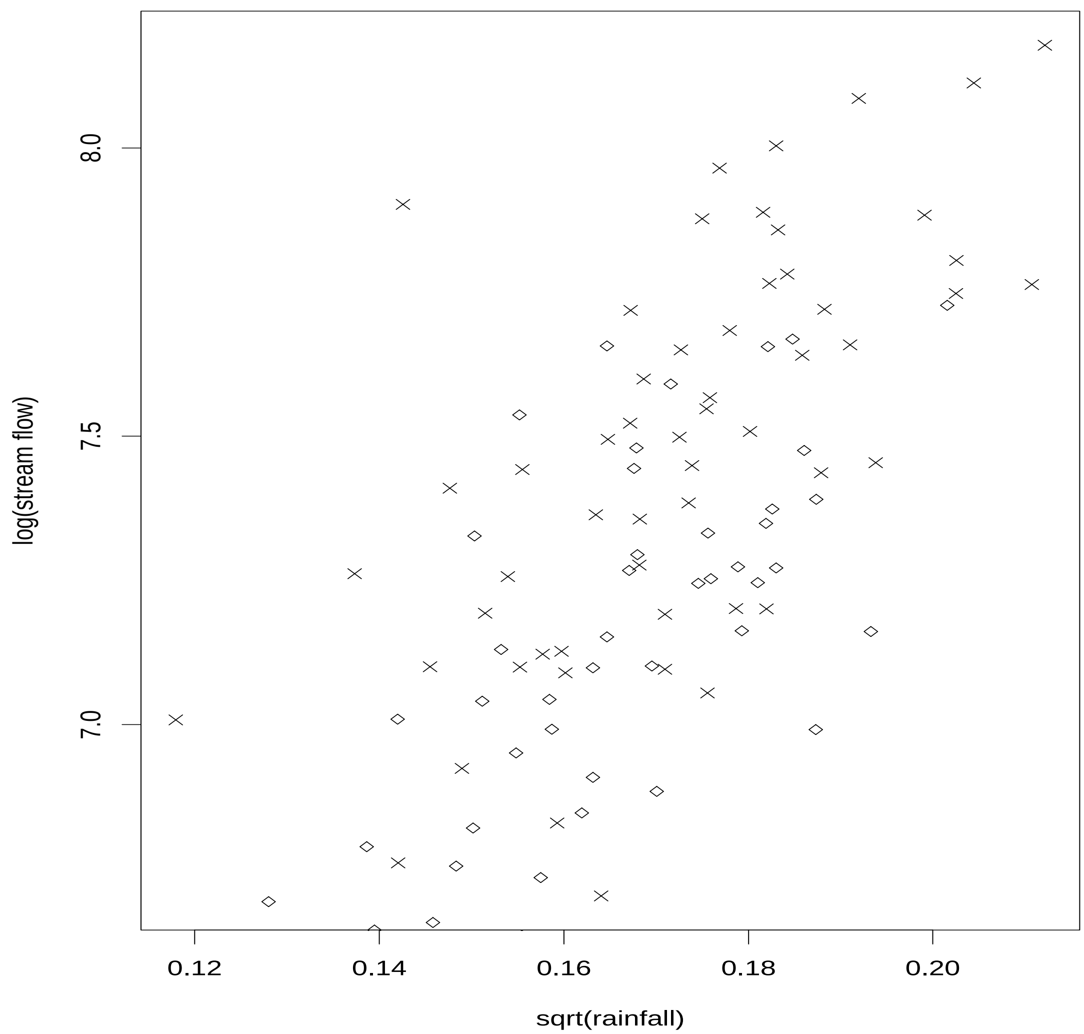


Figure 2.17: Scatterplot of yearly rainfall and stream flow in Tarboro and Kinston(daily average): $\diamondsuit$ Tarboro and $\times$ Kinston

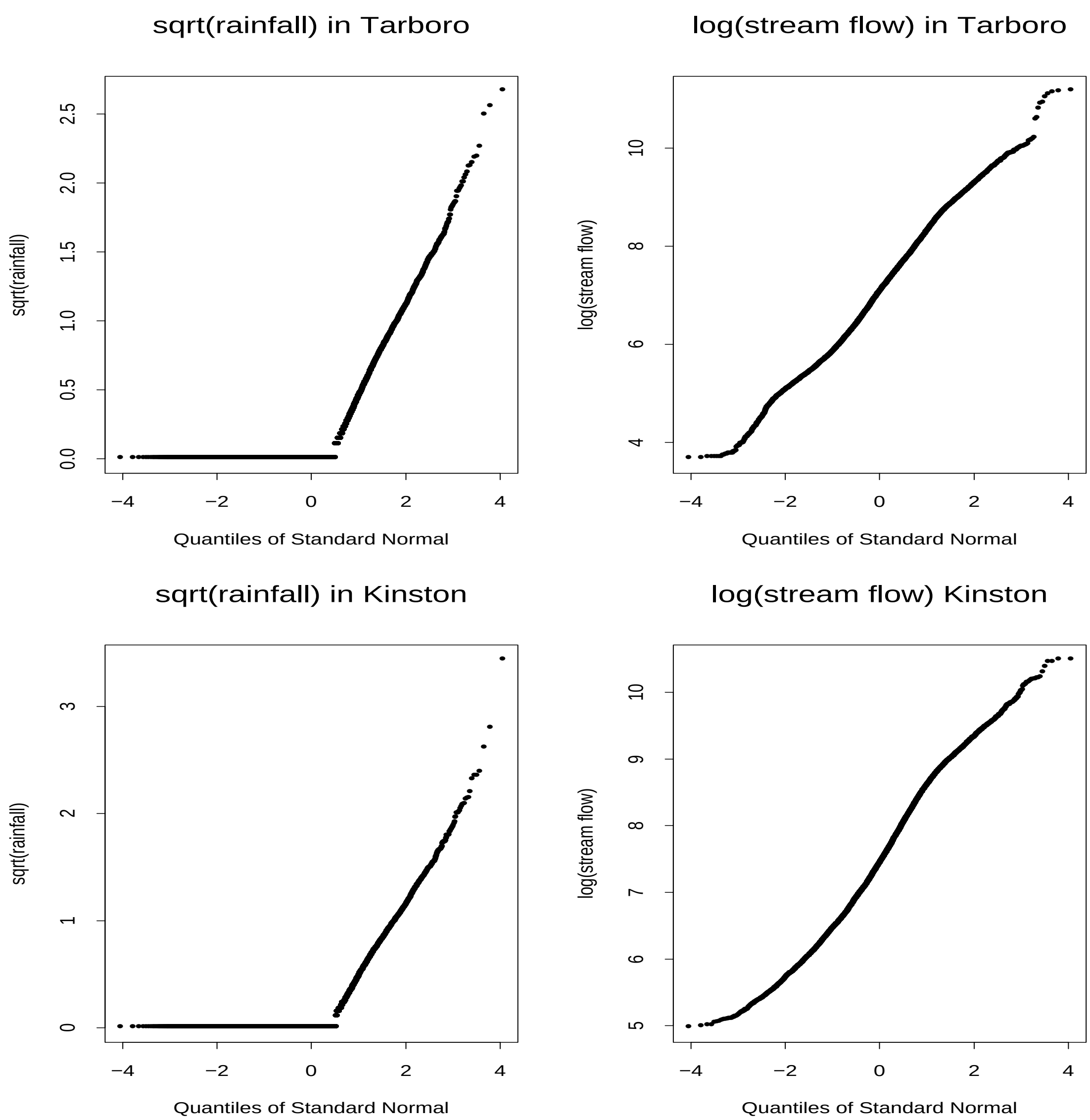


Figure 2.18: QQ plots of rainfall and stream flows in Tarboro and Kinston

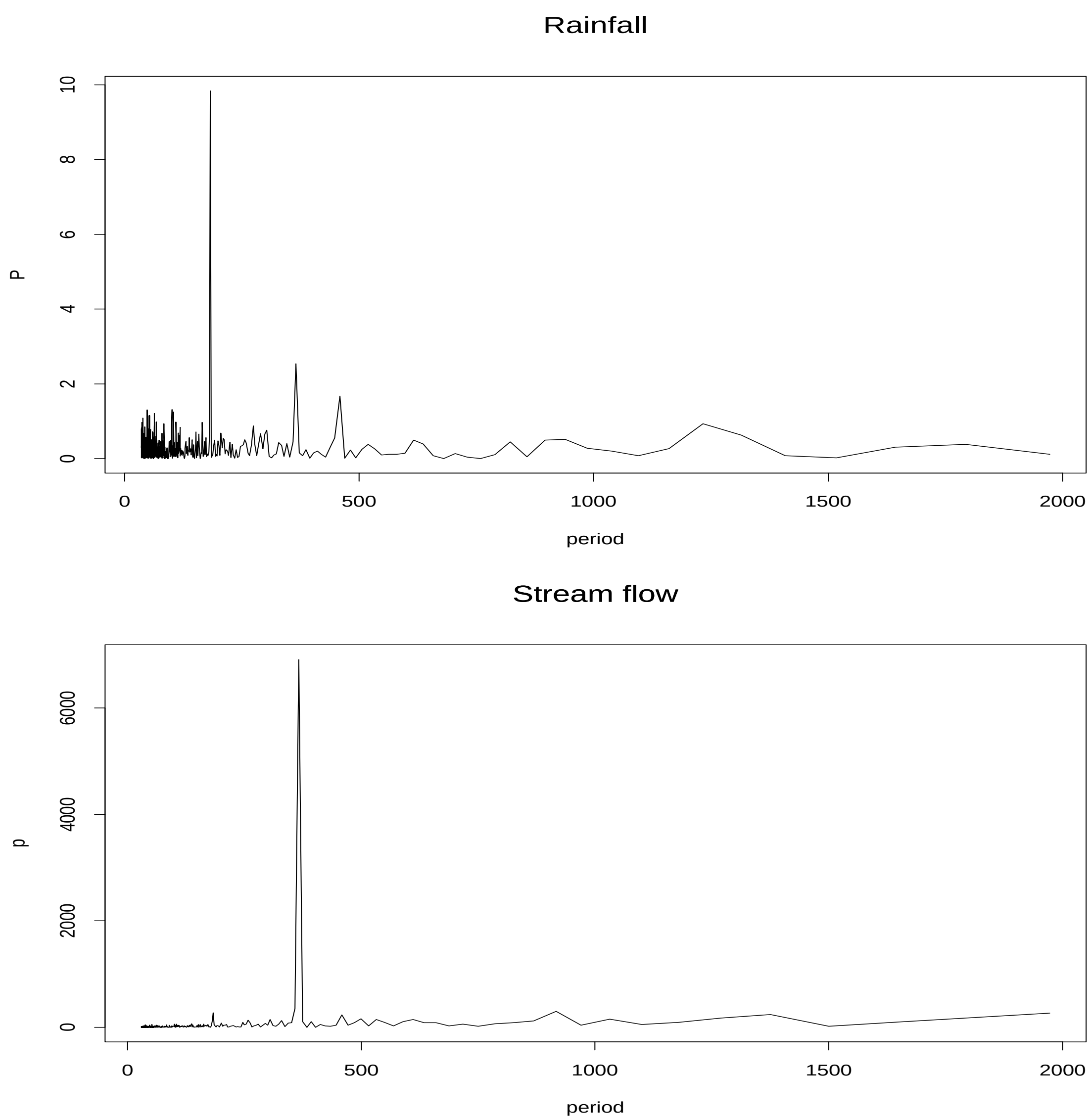


Figure 2.19: Periodograms of rainfall and stream flows in Tarboro

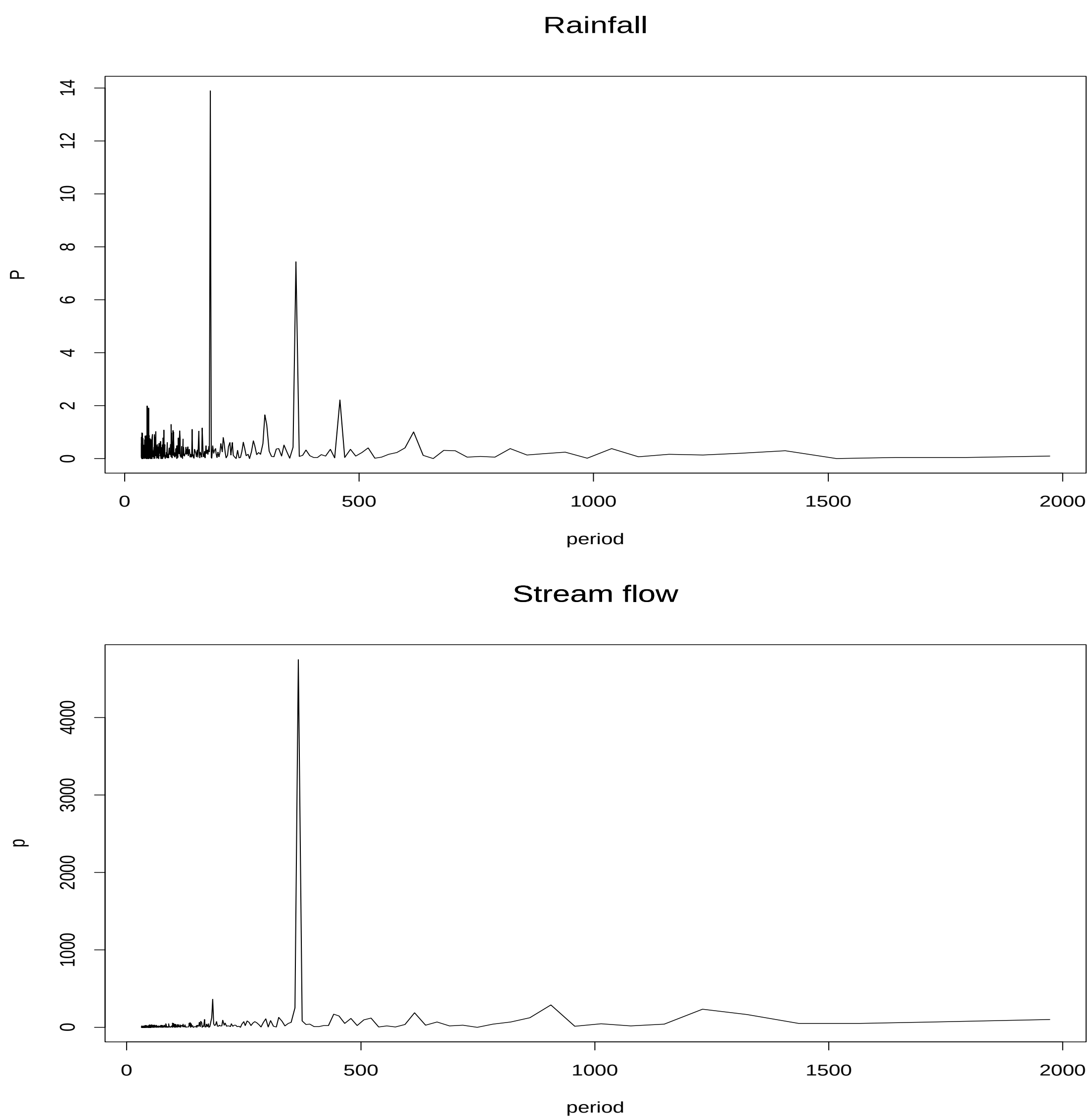


Figure 2.20: Periodograms of rainfall and stream flows in Kinston

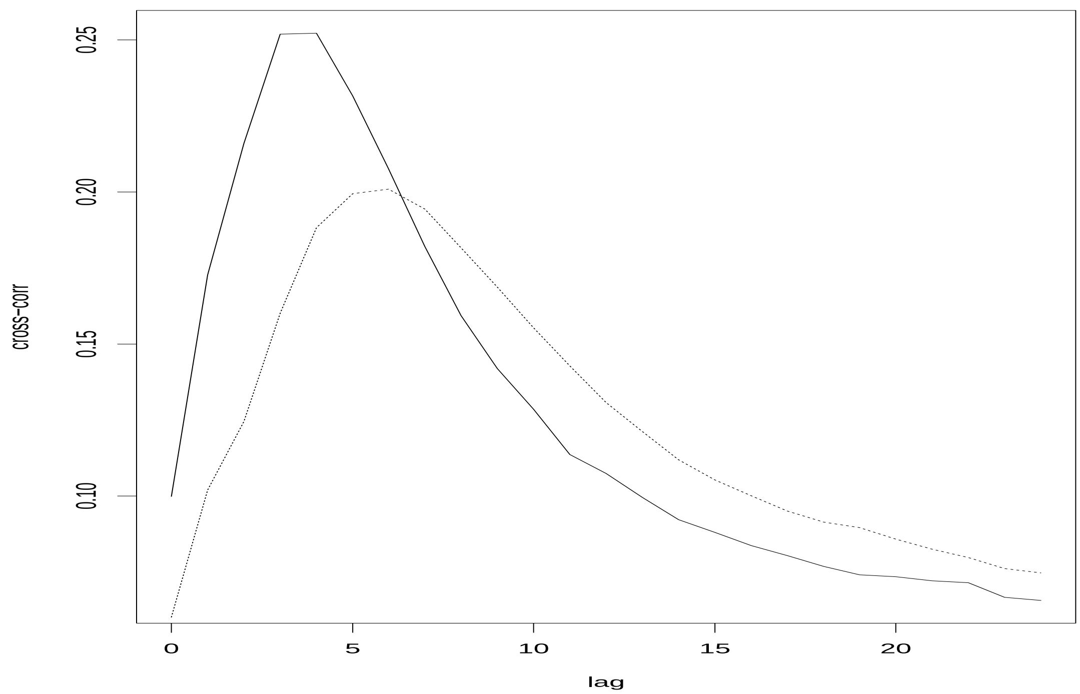


Figure 2.21: The cross-correlations of rainfall and stream flows in Tarboro(solid line) and Kinston(dotted line)

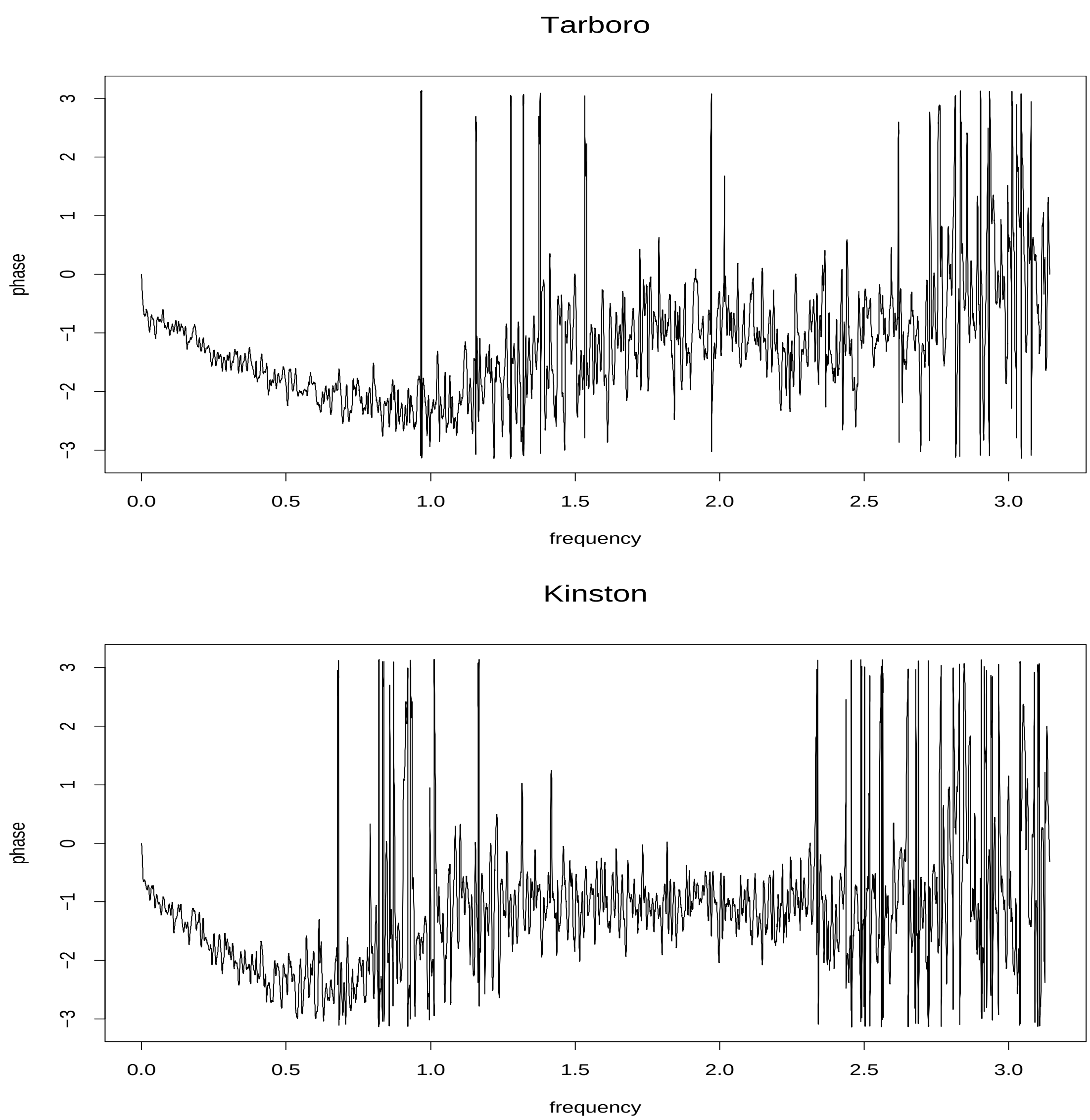


Figure 2.22: The phase spectrum of rainfall and stream flows in Tarboro and Kinston

Chapter 3

Nonlinear autoregressive model

3.1 NLAR(1) model

We now consider a nonlinear autoregressive model with no intercept.

$$y_t = \gamma\rho(y_{t-1})y_{t-1} + e_t \tag{3.1}$$

where $\rho(y_{t-1}) = \frac{\exp(\alpha+\beta f(y_{t-1}))-1}{\exp(\alpha+\beta f(y_{t-1}))+1}$. Here $f(y) = |y|$ or $f(y) = y$, $|\gamma| < 1$, $\beta > 0$, and e_t comes from an IID $(0, \sigma^2)$ distribution.

Models with lagged dependent variables are classified as dynamic models(Gallant 1986). This model is a minor adjustment to the usual constant coefficient AR(1) model and a specific form of LSTAR. This adjustment appears to provide quite a bit of flexibility in terms of the types of data structure it can provide. The model is appropriate for series with asymmetric stochastic volatility or changing amplitude around 0 and a more or less persistent autocorrelation at long lags where $|\gamma|$ is near 1. The difference between the series generated using $f(y) = |y|$ and $f(y) = y$ is that the former appear symmetric in the long run with locally asymmetric features, while some of the latter can stay asymmetric for a long time(see Figure 3.1,$\cdots$,3.6).

Notice that $|\gamma\rho(y_{t-1})| < |\gamma|$. The model can produce local autocorrelation coefficients quite close to ± 1 if $|\gamma|$ is near 1. This allows the model to generate data that is locally nonstationary in appearance but in the long term tends to be mean-reverting. The restriction $|\gamma| < 1$ holds $\gamma\rho(y_{t-1})$ within the $(-1, 1)$ interval. With this restriction, $F'_{nk}(\theta)F_{nk}(\theta)$ satisfies the rank qualification and the identification condition with $\beta > 0$ and $\gamma \neq 0$ as shown in chapter 2 for model (2.1).

We restrict the range of β to $\beta > 0$. Notice that for $\theta' = (\gamma, \alpha, \beta)$ and $(-\gamma, -\alpha, -\beta)$,

$$\gamma \frac{\exp(\alpha + \beta f(y_{t-1})) - 1}{\exp(\alpha + \beta f(y_{t-1})) + 1} = -\gamma \frac{\exp(-\alpha - \beta f(y_{t-1})) - 1}{\exp(-\alpha - \beta f(y_{t-1})) + 1}.$$

Thus, there will be multiple optimizing solutions. But we always find a unique minimum by restricting $\beta > 0$.

For dynamic models, it is not easy to get consistency and asymptotic normality of parameter estimates. There has been much discussion about that in the literature(Gallant, 1986; Tjϕstheim, 1986; Tong,1990; Taniguchi and Kakizawa, 2000). Tong(1990) introduces general conditions for consistency and asymptotic normality. Also, he proved consistency and asymptotic normality of the parameter estimates in the STAR model under the assumption that the series is strictly stationary and ergodic. The conditions under which STAR models generate series that are stationary are not well known(Chan and Tong, 1986; Tong, 1990; Franses and van Dijk, 2000).

We quote theorem A1.10 and theorem 4.3 from Tong(1990). These prove the geometric ergodicity of series. If a series satisfies the conditions from the theorems, it is called geometrically ergodic and consistency and asymptotic normal distribution of parameter estimates can be obtained under some restrictions. We state them using our notation.

Consider the following nonlinear autoregressive(NLAR(m)) model

$$Y_t = \alpha_1 f_1(Y_{t-1}; \theta_1) + \alpha_2 f_2(Y_{t-2}; \theta_2) + \cdots + \alpha_m f_m(Y_{t-m}; \theta_m) + e_t.$$

We assume that

(i) For all θ_i, $f_i(.; \theta_i)$ is a fixed function bounded over bounded subsets of R.

(ii) e_t is IID, $E(|e_t|) < \infty$, and e_1 admits a positive and continuous p.d.f.

(iii) For all i, there exist $\phi_i(\alpha_i, \theta_i)$ such that $\alpha_i f_i(y; \theta_i) - \phi_i y$ is a bounded function.

Now we write the model as

$$\begin{aligned} \mathbf{Y_t} &= T(\mathbf{Y_{t-1}}) + S(\mathbf{Y_{t-1}}, e_t) \\ &= \mathbf{\Phi Y_{t-1}} + S(\mathbf{Y_{t-1}}, e_t) \end{aligned}$$

where

$$\mathbf{\Phi} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{m-1} & \phi_m \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

and

$$S(\mathbf{Y_{t-1}}, e_t) = \mathbf{Y_t} - \mathbf{\Phi Y_{t-1}}.$$

The associated deterministic difference equation is

$$\mathbf{y_t} = T(\mathbf{y_{t-1}}) = \Phi \mathbf{y_{t-1}}.$$

Under the assumptions above, if the following conditions hold, then $\mathbf{Y_t}$ is geometrically ergodic. Note that y_t is the solution to the deterministic equation and Y_t is a random variable.

(i) $\mathbf{0} \in R^m$ is an equilibrium state for $\mathbf{y_t} = T(\mathbf{y_{t-1}})$, that is, $\mathbf{0} = T(\mathbf{0})$, and is exponentially asymptotically stable in the large, that is, there exists a K and $c > 0$ such that for all $t \geq 0$, and starting with $\mathbf{y_0} \in R^m, ||\mathbf{y_t}|| \leq Ke^{-ct}||\mathbf{y_0}||$, where $||.||$ denotes the Euclidean norm in R^m.

(ii) T is Lipschitz continuous over R^m, that is, there exists a $M > 0$ such that for all $\mathbf{y}, \mathbf{x} \in R^m, ||T(\mathbf{y}) - T(\mathbf{x})|| \leq M||\mathbf{y} - \mathbf{x}||$.

(iii) For some $\tau > 0, E[||S(\mathbf{Y_{t-1}}, e_t)||$ given $\mathbf{Y_{t-1}} = \mathbf{y}] \leq \tau$ for all $\mathbf{y} \in R^m$.

(end of theorem A1.10).

Notice that this theorem expresses $\mathbf{Y_t}$ in a form reminiscent of the AR(m) model

$$\mathbf{Y_t} = T(\mathbf{Y_{t-1}}) + \mathbf{e_t}.$$

Stationary conditions on the roots of $\mathbf{T}$ are sufficient for (i) and (ii).

For our model where $m = 1$ and $f(y) = |y|$,

$$\begin{aligned} y_t &= \gamma\rho(y_{t-1})y_{t-1} + e_t \\ &= \gamma\frac{\exp(\alpha + \beta|y_{t-1}|) - 1}{\exp(\alpha + \beta|y_{t-1}|) + 1}y_{t-1} + e_t \\ &= \gamma y_{t-1} - \gamma\frac{2}{\exp(\alpha + \beta|y_{t-1}|) + 1}y_{t-1} + e_t. \end{aligned}$$

It has the origin as the exponentially asymptotically stable equilibrium because the root of the characteristic equation of

$$y_t = \gamma y_{t-1}$$

lies inside the unit circle as long as $|\gamma| < 1$ and

$$S(y_{t-1}, e_t) = -\gamma \frac{2}{\exp(\alpha + \beta|y_{t-1}|) + 1} y_{t-1} + e_t.$$

Note that $|\gamma \frac{2}{\exp(\alpha+\beta|y|)} y| \leq c < \infty$ where $\beta > 0$ and $E(|e_t|) = c' < \infty$ for some c and c'. Thus, condition (i) and (iii) are satisfied. Clearly, T is Lipschitz continuous over R. Hence, the data generated by the model above are geometrically ergodic whenever $|\gamma| < 1$.

For $f(y) = y$, $|\gamma \frac{2}{\exp(\alpha+\beta y)} y|$ is not bounded. So, we may not use theorem A1.10. We employ another theorem 4.3 by Tong(1990) which proves geometric ergodicity for this case with additive noise.

Under the Markov chain defined by

$$\mathbf{Y_t} = T(\mathbf{Y_{t-1}}) + \mathbf{e_t}$$

where $t \geq 1$ and $T : R^m \rightarrow R^m$, if $\mathbf{Y_t}$ satisfies

(i) The same as condition (i) in theorem A1.10.

(ii) Either $\mathbf{e_t}$ are IID, with marginal distribution function absolutely continuous, with an everywhere positive probability density function over R^m, and with $E||\mathbf{e_t}|| < \infty$, or $\mathbf{e_t} = (e_t, 0, \cdots, 0)'$ with e_t being IID, each having an absolutely continuous distribution with an everywhere positive probability density function over R and $E(|e_t|) < \infty$.

(iii) The same as condition (ii) in theorem A1.10.

then Y_t is geometrically ergodic(end of theorem 4.3).

Where $m = 1$,

$$\begin{aligned} y_t &= T(y_{t-1}) \\ &= \gamma\rho(y_{t-1})y_{t-1}. \end{aligned}$$

So,

$$\begin{aligned} y_1 &= \gamma\rho(y_0)y_0, \\ y_2 &= \gamma\rho(y_1)y_1 \\ &= \gamma^2\rho(y_1)\rho(y_0)y_0, \\ y_3 &= \gamma\rho(y_2)y_2 \\ &= \gamma^3\rho(y_2)\rho(y_1)\rho(y_0)y_0, \\ y_4 &= \gamma\rho(y_3)y_3 \\ &= \gamma^4\rho(y_3)\rho(y_2)\rho(y_1)\rho(y_0)y_0. \end{aligned}$$

In general,

$$\begin{aligned} y_t &= \gamma\rho(y_{t-1})y_{t-1} \\ &= \gamma^t\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_1)\rho(y_0)y_0. \end{aligned}$$

Clearly,

$$|\gamma^t\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_1)\rho(y_0)y_0| \leq |\gamma^t y_0| \leq |\gamma^t||y_0|.$$

Thus, there exists a K and $c > 0$ such that

$$|y_t| \leq |\gamma^t||y_0| \leq Ke^{-ct}|y_0|.$$

Condition (i) is satisfied.

The function $\frac{\exp(\alpha+\beta y)-1}{\exp(\alpha+\beta y)+1}y$ is Lipschitz continuous over R. To begin with, we assume $\alpha = 0, \beta > 0, \gamma = 1$, and $y > 0$. So,

$$g(y) = \rho(y)y = \frac{\exp(\beta y) - 1}{\exp(\beta y) + 1}y.$$

Note that $g(y) = g(-y)$, because

$$\frac{\exp(\beta y) - 1}{\exp(\beta y) + 1}y = \frac{\exp(\beta(-y)) - 1}{\exp(\beta(-y)) + 1}(-y).$$

We pick $y_1 > 0$, $y_2 = y_1 + \delta$ where $\delta > 0$ and want to show there exist a M such that

$$\frac{|g(y+\delta) - g(y)|}{\delta} < M$$

for all y and δ.

The mean value theorem says

$$g(y+\delta) - g(y) = \delta g'(\dot{y})$$

for $y \leq \dot{y} \leq y + \delta$.

$$g'(y) = \rho'(y)y + \rho(y)$$

and

$$\rho'(y) = \frac{2\beta\exp(\beta y)}{(\exp(\beta y)+1)^2} = 2\beta\frac{1}{(\exp(\beta y)+1)}\frac{\exp(\beta y)}{(\exp(\beta y)+1)}.$$

Here, $\rho'(y)y = \frac{2\beta\exp(\beta y)}{(\exp(\beta y)+1)^2}y \to 0$ exponentially fast and thus $|\rho'(y)y|$ takes on a maxium value c.

$$|g'(y)| \leq c + 1$$

for all $y > 0$.

Hence,

$$|g(y+\delta) - g(y)| = |\delta g'(\dot{y})| < \delta |c+1|$$

where $M = c+1$. $g(y)$ is Lipschitz continuous.

Now, for $-\infty < y < \infty$, we take any $y_2 > y_1$.

$$|y_2 - y_1| \geq ||y_2| - |y_1||$$

and

$$g(y) = g(|y|)$$

by symmetry. Thus,

$$\frac{|g(y_2) - g(y_1)|}{|y_2 - y_1|} < \frac{|g(|y_2|) - g(|y_1|)|}{||y_2| - |y_1||} < M$$

using the result above.

We move to the region $\beta < 0$. With $\tau = -\beta > 0$,

$$g(y) = \frac{\exp(-\tau y) - 1}{\exp(-\tau y) + 1} y$$

and

$$-g(y) = \frac{\exp(\tau y) - 1}{\exp(\tau y) + 1} y.$$

Thus, $-g(y)$ is Lipschitz continuous. Note that

$$| - g(y_2) - (-g(y_1))| = |g(y_2) - g(y_1)|.$$

It follows that if $-g(y)$ is Lipschitz continuous, then $g(y)$ is also Lipschitz continuous.

Where $\alpha \neq 0$, we write

$$g(y) \;\; = \;\; \frac{\exp(\alpha + \beta y) - 1}{\exp(\alpha + \beta y) + 1} y$$

$$
\begin{aligned}
&= \frac{\exp(\beta(y+\frac{\alpha}{\beta}))-1}{\exp(\beta(y+\frac{\alpha}{\beta}))+1}(y+\frac{\alpha}{\beta}) - \frac{\alpha}{\beta}\frac{\exp(\beta(y+\frac{\alpha}{\beta}))-1}{\exp(\beta(y+\frac{\alpha}{\beta}))+1} \\
&= \frac{\exp(\beta z)-1}{\exp(\beta z)+1}z - \frac{\alpha}{\beta}\frac{\exp(\beta z)-1}{\exp(\beta z)+1}
\end{aligned}
$$

with $z = y + \frac{\alpha}{\beta}$. Then, $z_2 - z_1 = y_2 - y_1$ and the first function $\frac{\exp(\beta z)-1}{\exp(\beta z)+1}z$ is Lipschitz continuous. The second function $\frac{\alpha}{\beta}\frac{\exp(\beta z)-1}{\exp(\beta z)+1}$ is also Lipschitz continuous, because its derivative $\frac{2\alpha\exp(\beta z)}{(\exp(\beta z)+1)^2} = 2\alpha\frac{1}{(\exp(\beta z)+1)}\frac{\exp(\beta z)}{(\exp(\beta z)+1)}$ is bounded. So again, by the mean value theorem, we have the result. In this way, condition (iii) is satisfied.

We assume e_t is an IID $(0, \sigma^2)$ sequence, so this satisfies condition (ii). Hence, y_t is geometrically ergodic. The previous case $f(y) = |y|$ can be shown to be geometrically ergodic by using this approach, too.

Tjϕstheim(1986) proved the consistency and asymptotic normality of parameter estimates where Y_t is strictly stationary and ergodic with some restrictions. Let $\{Y_t, t \in I\}$ be a discrete time stochastic process taking values in R^d and defined on a probability space (Ω, A, P). Observations $(Y_1, \cdots, Y_n)$ are available and $A^Y_{t-1}(p)$ is the σ-field generated by $\{Y_s, t-p \le s \le t-1\}$. For NLAR($p$), we have $E(Y_t|A^Y_{t-1}) = E(Y_t|A^Y_{t-1}(p))$ where $t \ge p+1$ and A^Y_{t-1} is the sub σ-field of A generated by $\{Y_s, s < t\}$. We quote theorem 3.1 and 3.2 of Tjϕstheim(1986) in our notation.

Assume that Y_t is a d-dimensional strictly stationary ergodic process with $E(Y_t^2) < \infty$ and such that $Y_{t|t-1}(\theta) = E(Y_t|A^Y_{t-1}(p))$ is almost surely three times continuously differentiable in an open set Θ containing θ_0. Moreover, suppose that

(i) $E\Big(\Big|\frac{\partial Y_{t|t-1}}{\partial\theta_i}(\theta_0)\Big|^2\Big) < \infty$ and $E\Big(\Big|\frac{\partial^2 Y_{t|t-1}}{\partial\theta_i\partial\theta_j}(\theta_0)\Big|^2\Big) < \infty$ for $i, j = 1, \cdots, r$.

(ii) The vectors $\partial Y_{t|t-1}(\theta_0)/\partial\theta_i, i = 1, \cdots, r$, are linearly independent in the sense that if $a_1, \cdots, a_r$ are arbitrary real numbers such that

$$E\Big(\Big|\sum_{i=1}^{r} a_i \frac{\partial Y_{t|t-1}}{\partial \theta_i}(\theta_0)\Big|^2\Big) = 0,$$

then $a_1 = a_2 = \cdots = a_r = 0$.

(iii) For $\theta \in \Theta$, there exist functions $G_{t-1}^{ijk}(Y_1, \cdots, Y_{t-1})$ and H_t^{ijk} $(Y_1, \cdots, Y_{t-1})$ such that

$$\Big|\frac{\partial Y_{t|t-1}}{\partial \theta_i}(\theta)\frac{\partial^2 Y_{t|t-1}}{\partial \theta_j \partial \theta_k}(\theta)\Big| \leq G_{t-1}^{ijk}, E(G_{t-1}^{ijk}) < \infty,$$

$$\Big|(Y_t - Y_{t|t-1}(\theta))\frac{\partial^3 Y_{t|t-1}}{\partial \theta_i \partial \theta_j \partial \theta_k}(\theta)\Big| \leq H_t^{ijk}, E(H_t^{ijk}) < \infty,$$

for $i, j, k = 1, \cdots, r$.

Then there exists a sequence of estimators $\hat{\theta}_n$ minimizing

$$Q_n(\theta) = \sum_{t=p+1}^{n} (Y_t - Y_{t|t-1}(\theta))^2.$$

for which, $\hat{\theta}_n \to \theta_0$ almost surely as $n \to \infty$ and for $\epsilon > 0$, there is an event E in (Ω, A, P) with $P(E) > 1 - \epsilon$ and an n_0 such that on E and for $n > n_0, \partial Q_n(\hat{\theta}_n)/\partial \theta_i = 0, i = 1, \cdots, r$, and Q_n attains a relative minimum at $\hat{\theta}_n$(end of theorem 3.1). Tjϕstheim's theorem 3.1 establishes almost sure convergence of a sequence $(\hat{\theta}_n)$. His theorem 3.2 establishes the limit distribution of the normalized sequence $n^{1/2}(\hat{\theta}_n - \theta_0)$. It includes some additional assumptions as follows.

(i)

$$E(Y_t | A_{t-1}^Y) = E(Y_t | A_{t-1}^Y(p))$$

almost surely, and

$$\begin{aligned} f_{t|t-1} &= E((Y_t - Y_{t|t-1})(Y_t - Y_{t|t-1})^T | A_{t-1}^Y) \\ &= E((Y_t - Y_{t|t-1})(Y_t - Y_{t|t-1})^T | A_{t-1}^Y(p)) \end{aligned}$$

almostly surely.

(ii) The same as the condition (i),(ii),and (iii) in theorem 3.1.

(iii)

$$R = E\Big(\frac{\partial Y^T_{t|t-1}}{\partial \theta}(\theta_0) f_{t|t-1}(\theta_0) \frac{\partial Y_{t|t-1}}{\partial \theta}(\theta_0)\Big) < \infty.$$

Then $\hat{\theta}_n$ which is the estimator obtained by minimizing $Q_n(\theta)$ converges to a normal distribution.

$$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, U^{-1}RU^{-1})$$

where

$$U = E\Big(\frac{\partial Y^T_{t|t-1}}{\partial \theta}(\theta_0) \frac{\partial Y_{t|t-1}}{\partial \theta}(\theta_0)\Big).$$

Note that U and R are $r \times r$ matrices(end of theorem 3.2).

Condition (i) is trivially satisfied for nonlinear AR processes and condition (iii) is implied by condition (i) of theorem 3.1 in general time series where $Y_t - Y_{t|t-1}(\theta_0)$ is independent of A^Y_{t-1}.

Thus, it is enough to check conditions (i)-(iii) in theorem 3.1 to allow application of both theorems in our case. For this, we assume $E(e_t^4) < \infty$, which implies $E(y_t^4) < \infty$.

With $y_0 = 0$,

$$\begin{aligned}
y_1 &= e_1, \\
y_2 &= \gamma\rho(y_1)e_1 + e_2, \\
y_3 &= \gamma\rho(y_2)y_2 + e_3 \\
&= \gamma\rho(y_2)(\gamma\rho(y_1)e_1 + e_2) + e_3 \\
&= \gamma^2\rho(y_2)\rho(y_1)e_1 + \gamma\rho(y_2)e_2 + e_3,
\end{aligned}$$

$$\begin{aligned}
y_4 &= \gamma\rho(y_3)y_3 + e_4 \\
&= \gamma\rho(y_3)(\gamma^2\rho(y_2)\rho(y_1)e_1 + \gamma\rho(y_2)e_2 + e_3) + e_4 \\
&= \gamma^3\rho(y_3)\rho(y_2)\rho(y_1)e_1 + \gamma^2\rho(y_3)\rho(y_2)e_2 + \gamma\rho(y_3)e_3 + e_4.
\end{aligned}$$

In this way,

$$\begin{aligned}
y_t &= \gamma^{t-1}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_2)\rho(y_1)e_1 \\
&+ \gamma^{t-2}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_3)\rho(y_2)e_2 \\
&+ \cdots + \gamma^2\rho(y_{t-1})\rho(y_{t-2})e_{t-2} + \gamma\rho(y_{t-1})e_{t-1} + e_t.
\end{aligned}$$

Clearly,

$$\begin{aligned}
|\gamma^{t-1}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_2)\rho(y_1)e_1| &\leq |\gamma^{t-1}e_1|, \\
|\gamma^{t-2}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_3)\rho(y_2)e_2| &\leq |\gamma^{t-2}e_2|, \\
\vdots &\leq \vdots, \\
|\gamma^2\rho(y_{t-1})\rho(y_{t-2})e_{t-2}| &\leq |\gamma^2 e_{t-2}|, \\
|\gamma\rho(y_{t-1})e_{t-1}| &\leq |\gamma e_{t-1}|.
\end{aligned}$$

Thus,

$$\begin{aligned}
&|\gamma^{t-1}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_2)\rho(y_1)e_1| + |\gamma^{t-2}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_3)\rho(y_2)e_2| \\
&\qquad + \cdots + |\gamma^2\rho(y_{t-1})\rho(y_{t-2})e_{t-2}| + |\gamma\rho(y_{t-1})e_{t-1}| + |e_t| \\
&\qquad \leq |\gamma^{t-1}e_1| + |\gamma^{t-2}e_2| + \cdots + |\gamma^2 e_{t-2}| + |\gamma e_{t-1}| + |e_t|.
\end{aligned}$$

With $E(e_t^2) \leq M < \infty$ for all t and $\sum_{i=0}^{t-1}|\gamma^i| < \infty$,

$$\sum_{i=0}^{t-1} E(|\gamma^i e_{t-i}|) = \sum_{i=0}^{t-1} |\gamma^i| E(|e_{t-i}|) \leq \sum_{i=0}^{\infty} |\gamma^i| \sqrt{(M+1)} \leq \infty.$$

The right-hand side is absolutely summable. Therefore, the left-handed side is absolutely summable and

$$\begin{aligned}
E(y_t) &= E(\gamma^{t-1}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_2)\rho(y_1)e_1) \\
&+ E(\gamma^{t-2}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_3)\rho(y_2)e_2) \\
&+ \cdots + E(\gamma^2\rho(y_{t-1})\rho(y_{t-2})e_{t-2}) + E(\gamma\rho(y_{t-1})e_{t-1}) + E(e_t).
\end{aligned}$$

Again,

$$\begin{aligned}
y_t^2 &= \gamma^{2(t-1)}\rho^2(y_{t-1})\rho^2(y_{t-2})\cdots\rho^2(y_2)\rho^2(y_1)e_1^2 \\
&+ \gamma^{2(t-2)}\rho^2(y_{t-1})\rho^2(y_{t-2})\cdots\rho^2(y_3)\rho^2(y_2)e_2^2 \\
&+ \cdots + \gamma^4\rho^2(y_{t-1})\rho^2(y_{t-2})e_{t-2}^2 + \gamma^2\rho^2(y_{t-1})e_{t-1}^2 + e_t^2 \\
&+ 2\gamma^{t-1}\gamma^{t-2}\rho^2(y_{t-1})\rho^2(y_{t-2})\cdots\rho^2(y_2)\rho(y_1)e_1e_2 \\
&+ \cdots + 2\gamma\rho(y_{t-1})e_{t-1}e_t.
\end{aligned}$$

Clearly,

$$\begin{aligned}
&|\gamma^{2(t-1)}\rho^2(y_{t-1})\rho^2(y_{t-2})\cdots\rho^2(y_2)\rho^2(y_1)e_1^2| \\
&+|\gamma^{2(t-2)}\rho^2(y_{t-1})\rho^2(y_{t-2})\cdots\rho^2(y_3)\rho^2(y_2)e_2^2| \\
&+\cdots+|\gamma^4\rho^2(y_{t-1})\rho^2(y_{t-2})e_{t-2}^2| + |\gamma^2\rho^2(y_{t-1})e_{t-1}^2| + |e_t^2| \\
&+2|\gamma^{t-1}\gamma^{t-2}\rho^2(y_{t-1})\rho^2(y_{t-2})\cdots\rho^2(y_2)\rho(y_1)e_1e_2| \\
&+\cdots+2|\gamma\rho(y_{t-1})e_{t-1}e_t| \\
&\leq |\gamma^{2(t-1)}e_1^2| + |\gamma^{2(t-2)}e_2^2| + \cdots + |\gamma^4e_{t-2}^2| + |\gamma^2e_{t-1}^2| + |e_t^2| \\
&+2|\gamma^{t-1}\gamma^{t-2}e_1e_2| + \cdots + 2|\gamma e_{t-1}e_t|.
\end{aligned}$$

The expected value of right-hand side is absolutely summable provided $E(e_t^2) \leq$

$M < \infty$ for all t. This follows from

$$E(|e_{t-i}e_{t-j}|) \le [E(e_{t-i}^2)]^{1/2}[E(e_{t-j}^2)]^{1/2}$$

for $i \neq j$. Thus, the left-hand side is also absolutely summable and $E(y_t^2) < \infty$ exists.

For $E(y_t^4)$, each term of y_t^4 expressed as a combination of $e_{t-i}, i = 1, 2, ..., t-1$ will be less than

$$\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}\sum_{k=0}^{t-1}\sum_{l=0}^{t-1}|\gamma^i\gamma^j\gamma^k\gamma^l||e_{t-i}e_{t-j}e_{t-k}e_{t-l}|$$

$$\begin{aligned} &= \sum_{i=j=k=l}|\gamma^{4i}||e_{t-i}^4| \\ &+ \sum_{i=j=k\neq l}|\gamma^{3i}\gamma^l||e_{t-i}^3e_{t-l}| \\ &+ \sum_{i=j\neq k=l}|\gamma^{2i}\gamma^{2l}||e_{t-i}^2e_{t-k}^2| \\ &+ \sum_{i\neq j\neq k\neq l}|\gamma^i\gamma^j\gamma^k\gamma^l||e_{t-i}e_{t-j}e_{t-k}e_{t-l}|. \end{aligned}$$

With $E(e_t^4) \le M < \infty$ for all t and

$$\begin{aligned} E(|e_{t-i}^2e_{t-k}^2|) &\le [E(e_{t-i}^4)]^{1/2}[E(e_{t-j}^4)]^{1/2} \\ E(|e_{t-i}^3e_{t-l}|) &\le [E(e_{t-i}^4)]^{3/4}[E(e_{t-j}^4)]^{1/4} \\ E(|e_{t-i}e_{t-j}e_{t-k}e_{t-l}|) &\le [E(e_{t-i}^2e_{t-j}^2)]^{1/2}[E(e_{t-k}^2e_{t-l}^2)]^{1/2}, \end{aligned}$$

both sides are absolutely summable and $E(y_t^4) < \infty$ exists as long as $|\gamma| < 1$. Thus, $E(y_t^4) < \infty$ provided $E(e_t^4) \le M < \infty$. This result holds for both $f(y) = y$ and $f(y) = |y|$.

Now, with $L_t = \alpha + \beta y_{t-1}$, we have

$$\frac{\partial y_{t|t-1}}{\partial \gamma} = \frac{\exp(L_t) - 1}{\exp(L_t) + 1}y_{t-1},$$

$$
\begin{aligned}
\frac{\partial y_{t|t-1}}{\partial \alpha} &= 2\gamma \frac{\exp(L_t)}{(\exp(L_t)+1)^2} y_{t-1}, \\
\frac{\partial y^2_{t|t-1}}{\partial \alpha^2} &= 2\gamma \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} y_{t-1}, \\
\frac{\partial y^3_{t|t-1}}{\partial \alpha^3} &= 2\gamma \frac{C\exp(L_t)}{(\exp(L_t)+1)^4} y_{t-1}, \\
\frac{\partial y_{t|t-1}}{\partial \beta} &= 2\gamma \frac{\exp(L_t)}{(\exp(L_t)+1)^2} y^2_{t-1} \\
\frac{\partial y^2_{t|t-1}}{\partial \beta^2} &= 2\gamma \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} y^3_{t-1}, \\
\frac{\partial y^3_{t|t-1}}{\partial \beta^3} &= 2\gamma \frac{C\exp(L_t)}{(\exp(L_t)+1)^4} y^4_{t-1}, \\
\frac{\partial y^2_{t|t-1}}{\partial \gamma \partial \alpha} &= 2 \frac{\exp(L_t)}{(\exp(L_t)+1)^2} y_{t-1}, \\
\frac{\partial y^2_{t|t-1}}{\partial \gamma \partial \beta} &= 2 \frac{\exp(L_t)}{(\exp(L_t)+1)^2} y^2_{t-1}, \\
\frac{\partial y^2_{t|t-1}}{\partial \alpha \partial \beta} &= 2\gamma \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} y^2_{t-1}, \\
\frac{\partial y^3_{t|t-1}}{\partial \alpha^2 \partial \gamma} &= 2 \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} y_{t-1}, \\
\frac{\partial y^3_{t|t-1}}{\partial \alpha^2 \partial \beta} &= 2\gamma \frac{C\exp(L_t)}{(\exp(L_t)+1)^4} y^2_{t-1}, \\
\frac{\partial y^3_{t|t-1}}{\partial \beta^2 \partial \gamma} &= 2 \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} y^3_{t-1}, \\
\frac{\partial y^3_{t|t-1}}{\partial \beta^2 \partial \alpha} &= 2\gamma \frac{C\exp(L_t)}{(\exp(L_t)+1)^4} y^3_{t-1}, \\
\frac{\partial y^3_{t|t-1}}{\partial \gamma \partial \alpha \partial \beta} &= 2 \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} y^2_{t-1}
\end{aligned}
$$

where the derivatives are not 0 and where $C = 1 - 4\exp(L_t) + \exp(2L_t)$. These derivatives are bounded by $|y_{t-1}|, |y^2_{t-1}|, |y^3_{t-1}|$, and $|y^4_{t-1}|$. Thus, with $E(y^4_t) < \infty$, condition (i) is satisfied.

For condition (iii),

$$\begin{aligned}
|y_t - y_{t|t-1}| &= |y_t - \gamma \frac{\exp(L_t)-1}{\exp(L_t)+1} y_{t-1}| \\
&\leq |y_t| + |y_{t-1}|.
\end{aligned}$$

Thus, for example,

$$\Big|(y_t - y_{t|t-1}) \frac{\partial^3 y_{t|t-1}}{\partial\gamma\partial\alpha\partial\beta}\Big|$$

$$\begin{aligned}
&= \Big|2(y_t - y_{t|t-1}) \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} y_{t-1}^2\Big| \\
&\leq |y_t - y_{t|t-1}|\Big|2 \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} y_{t-1}^2\Big| \\
&\leq |y_t|\Big|2 \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} y_{t-1}^2\Big| \\
&+ |y_{t-1}|\Big|2 \frac{\exp(L_t)(1-\exp(L_t))}{(\exp(L_t)+1)^3} y_{t-1}^2\Big| \\
&\leq |y_t||y_{t-1}^2| + |y_{t-1}^3| \\
&\leq (y_t^2)^{1/2}(y_{t-1}^4)^{1/2} + |y_{t-1}^3|,
\end{aligned}$$

and

$$E((y_t^2)^{1/2}(y_{t-1}^4)^{1/2} + |y_{t-1}^3|) < \infty$$

with $E(y_t^4) < \infty$. We know that condition (iii) is satisfied in this way.

Finally, let a_1, a_2, and a_3 be three arbitrary real numbers. Then

$$E\Big(\Big|a_1 \frac{\partial y_{t|t-1}}{\partial\gamma} + a_2 \frac{\partial y_{t|t-1}}{\partial\alpha} + a_3 \frac{\partial y_{t|t-1}}{\partial\beta}\Big|^2\Big) = 0$$

implies

$$a_1 \frac{\exp(L_t)-1}{\exp(L_t)+1} y_{t-1} + 2a_2\gamma \frac{\exp(L_t)}{(\exp(L_t)+1)^2} y_{t-1} + 2a_3\gamma \frac{\exp(L_t)}{(\exp(L_t)+1)^2} y_{t-1}^2 = 0$$

almost surely. $E(y_t^2) \geq E(e_t^2) > 0$. Y takes on enough values that each derivative is not linearly dependent on any other and $a_1 = a_2 = a_3 = 0$ follows.

In detail, it should be

$$a_1(\exp(2L_t) - 1)y_{t-1} + 2a_2\gamma \exp(L_t)y_{t-1} + 2a_3\gamma \exp(L_t)y_{t-1}^2 = 0$$

from the equation above. Note that y and y^2 are linear independent functions of y. So, $y \exp(L)$ and $y^2 \exp(L)$ must be linear independent. In addition, neither these nor $y \exp(2L)$ is a linear function of y. Thus, $y \exp(2L) - y$ must be linearly independent of $y \exp(L)$ and $y^2 \exp(L)$. Unless $\gamma = 0$, which we assume does not happen since $\gamma = 0$ implies α and β not identified, the linear independence is thus proved. Condition (ii) is also satisfied and the consistency and asymptotic normality of parameter estimates are obtained.

Hence,

$$\hat{\theta}_n \to \theta_0$$

almost surely, and

$$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N[0, U^{-1}RU^{-1}].$$

Here, $U^{-1}RU^{-1} =$

$$\left[E\begin{pmatrix} (1-2\rho_0)^2 y_{t-1}^2 & 2\gamma_0\rho_0(1-\rho_0)(1-2\rho_0)y_{t-1}^2 & 2\gamma_0\rho_0(1-\rho_0)(1-2\rho_0)y_{t-1}^3 \\ 2\gamma_0\rho_0(1-\rho_0)(1-2\rho_0)y_{t-1}^2 & 4\gamma_0^2\rho_0^2(1-\rho_0)^2 y_{t-1}^2 & 4\gamma_0^2\rho_0^2(1-\rho_0)^2 y_{t-1}^3 \\ 2\gamma_0\rho_0(1-\rho_0)(1-2\rho_0)y_{t-1}^3 & 4\gamma_0^2\rho_0^2(1-\rho_0)^2 y_{t-1}^3 & 4\gamma_0^2\rho_0^2(1-\rho_0)^2 y_{t-1}^4 \end{pmatrix}\right]^{-1} \sigma^2$$

where $\rho_0 = \frac{1}{\exp(\alpha_0 + \beta_0 y_{t-1}) + 1}$.

We can show this where $f(y_{t-1}) = |y_{t-1}|$ as well. Then, $U^{-1}RU^{-1} =$

$$\left[E\begin{pmatrix} (1-2\rho_0)^2 y_{t-1}^2 & 2\gamma_0\rho_0(1-\rho_0)(1-2\rho_0)y_{t-1}^2 & 2\gamma_0\rho_0(1-\rho_0)(1-2\rho_0)|y_{t-1}|y_{t-1}^2 \\ 2\gamma_0\rho_0(1-\rho_0)(1-2\rho_0)y_{t-1}^2 & 4\gamma_0^2\rho_0^2(1-\rho_0)^2 y_{t-1}^2 & 4\gamma_0^2\rho_0^2(1-\rho_0)^2 |y_{t-1}|y_{t-1}^2 \\ 2\gamma_0\rho_0(1-\rho_0)(1-2\rho_0)|y_{t-1}|y_{t-1}^2 & 4\gamma_0^2\rho_0^2(1-\rho_0)^2 |y_{t-1}|y_{t-1}^2 & 4\gamma_0^2\rho_0^2(1-\rho_0)^2 y_{t-1}^4 \end{pmatrix}\right]^{-1} \sigma^2$$

where $\rho_0 = \frac{1}{\exp(\alpha_0 + \beta_0 |y_{t-1}|) + 1}$.

Persistent autocorrelation as in model (3.1) can appear along with serial autocorrelation of residuals under the same model(see Figure 3.7). We extend our model to the one with serially correlated errors.

$$y_t = \gamma\rho(y_{t-1})y_{t-1} + \eta_t,$$

and

$$\eta_t = \delta_1\eta_{t-1} + \cdots + \delta_k\eta_{t-k} + e_t.$$

where e_t comes from an IID $(0, \sigma^2)$ distribution. The series generated by this model is still geometrically ergodic.

For example, k=1, where $f(y) = |y|$

$$\begin{aligned} y_t &= \gamma\rho(y_{t-1})y_{t-1} + \eta_t, \\ \eta_t &= \delta\eta_{t-1} + e_t. \end{aligned}$$

So,

$$\begin{aligned} y_t &= \gamma\rho(y_{t-1})y_{t-1} + \eta_t, \\ \delta y_{t-1} &= \delta\gamma\rho(y_{t-2})y_{t-2} + \delta\eta_{t-1}. \end{aligned}$$

Then,

$$\begin{aligned} y_t &= \Big[\delta + \gamma\frac{\exp(\alpha+\beta|y_{t-1}|)-1}{\exp(\alpha+\beta|y_{t-1}|)+1}\Big]y_{t-1} \\ &- \delta\gamma\frac{\exp(\alpha+\beta|y_{t-2}|)-1}{\exp(\alpha+\beta|y_{t-2}|)+1}y_{t-2} + e_t \\ &= (\delta+\gamma)y_{t-1} - \delta\gamma y_{t-2} \\ &- \gamma\frac{2}{\exp(\alpha+\beta|y_{t-1}|)+1}y_{t-1} \\ &+ \delta\gamma\frac{2}{\exp(\alpha+\beta|y_{t-2}|)+1}y_{t-2} + e_t. \end{aligned}$$

Notice that this can be thought as a LSTAR with four regimes. NAR(1) with serially correlated errors is one variant of LSTAR with many regimes.

We have the skeleton

$$\begin{aligned} \mathbf{y_t} &= T(\mathbf{y_{t-1}}) \\ &= \begin{pmatrix} (\delta+\gamma)y_{t-1} - \delta\gamma y_{t-2} \\ y_{t-1} \end{pmatrix}, \end{aligned}$$

and

$$S(\mathbf{Y_{t-1}}, \mathbf{e_t}) = \begin{pmatrix} -\gamma\frac{2}{\exp(\alpha+\beta|y_{t-1}|)+1}y_{t-1} + \delta\gamma\frac{2}{\exp(\alpha+\beta|y_{t-2}|)+1}y_{t-2} + e_t \\ 0 \end{pmatrix}.$$

Two roots of the characteristic equation

$$m^2 - (\delta+\gamma)m + \delta\gamma = 0$$

lie inside of unit circle under $|\delta| < 1$ and $|\gamma| < 1$ and each of the terms of $S(\mathbf{Y_{t-1}}, \mathbf{e_t})$ is bounded. Following theorem A1.10, the series is geometrically ergodic.

For $f(y) = y$, we use theorem 4.3 again.

For $m = 2$,

$$\begin{aligned} \mathbf{y_t} &= \mathbf{T}(\mathbf{y_{t-1}}) \\ &= \begin{pmatrix} \left[\delta + \gamma\frac{\exp(\alpha+\beta y_{t-1})-1}{\exp(\alpha+\beta y_{t-1})+1}\right]y_{t-1} - \delta\gamma\frac{\exp(\alpha+\beta y_{t-2})-1}{\exp(\alpha+\beta y_{t-2})+1}y_{t-2} \\ y_{t-1} \end{pmatrix}, \end{aligned}$$

and

$$\mathbf{e_t} = \begin{pmatrix} e_t \\ 0 \end{pmatrix}.$$

$$y_t = (\delta + \gamma\rho(y_{t-1}))y_{t-1} - \delta\gamma\rho(y_{t-2})y_{t-2} + e_t$$

where $\rho(y_{t-1}) = \frac{\exp(\alpha+\beta y_{t-1})-1}{\exp(\alpha+\beta y_{t-1})+1}$.

So,

$$
\begin{aligned}
y_1 &= (\delta + \gamma\rho(y_0))y_0, \\
y_2 &= (\delta + \gamma\rho(y_1))y_1 - \delta\gamma\rho(y_0)y_0 \\
&= (\delta^2 + \delta\gamma\rho(y_1) + \gamma^2\rho(y_1)\rho(y_0))y_0, \\
y_3 &= (\delta + \gamma\rho(y_2))y_2 - \delta\gamma\rho(y_1)y_1 \\
&= (\delta^3 + \delta^2\gamma\rho(y_2) + \delta\gamma^2\rho(y_2)\rho(y_1) \\
&+ \gamma^3\rho(y_2)\rho(y_1)\rho(y_0)y_0, \\
y_4 &= (\delta + \gamma\rho(y_3))y_3 - \delta_1\gamma\rho(y_2)y_2 \\
&= (\delta^4 + \delta^3\gamma\rho(y_3) + \delta^2\gamma^2\rho(y_3)\rho(y_2) \\
&+ \delta\gamma^3\rho(y_3)\rho(y_2)\rho(y_1) \\
&+ \gamma^4\rho(y_3)\rho(y_2)\rho(y_1)\rho(y_0))y_0.
\end{aligned}
$$

In general,

$$
\begin{aligned}
y_t &= (\delta + \gamma\rho(y_{t-1}))y_{t-1} - \delta\gamma\rho(y_{t-2})y_{t-2} \\
&= (\delta^t + \delta^{t-1}\gamma\rho(y_{t-1}) + \delta^{t-2}\gamma^2\rho(y_{t-1})\rho(y_{t-2}) \\
&+ \delta^{t-3}\gamma^3\rho(y_{t-1})\rho(y_{t-2})\rho(y_{t-3}) \\
&+ \cdots \\
&+ \delta^2\gamma^{t-2}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_3)\rho(y_2) \\
&+ \delta\gamma^{t-1}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_2)\rho(y_1) \\
&+ \gamma^t\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_1)\rho(y_0))y_0.
\end{aligned}
$$

Clearly,

$$
\begin{aligned}
|(\delta^t + \delta^{t-1}\gamma\rho(y_{t-1}) + \delta^{t-2}\gamma^2\rho(y_{t-1})\rho(y_{t-2}) \\
+\delta^{t-3}\gamma^3\rho(y_{t-1})\rho(y_{t-2})\rho(y_{t-3}) + \cdots \\
+\delta^2\gamma^{t-2}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_3)\rho(y_2) \\
+\delta\gamma^{t-1}\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_2)\rho(y_1) \\
+\gamma^t\rho(y_{t-1})\rho(y_{t-2})\cdots\rho(y_1)\rho(y_0))y_0| \\
\leq (|\delta^t| + |\delta^{t-1}\gamma| + |\delta^{t-2}\gamma^2| + \cdots \\
+|\delta^2\gamma^{t-2}| + |\delta^1\gamma^{t-1}| + |\gamma^t|)|y_0|.
\end{aligned}
$$

Under $|\delta| < 1$ and $|\gamma| < 1$, There exists a K and $c > 0$ such that

$$|y_t| \leq Ke^{-ct}|y_0|.$$

Hence, according to theorem 4.3, the series is geometrically ergodic.

The series with serially correlated errors is geometrically ergodic and the parameter estimates are normally distributed based on the theorem 3.1 and 3.2 of Tjϕstheim(1986).

For forecasting, consider the model

$$y_t = F(y_{t-1}; \theta) + e_t$$

for some nonlinear function $F(y_{t-1}; \theta)$. Using a least square criterion, the optimal point forecasts of future values of the time series are given by their conditional expectation(Frances and van Dijk, 2000). Thus, the optimal h-step-ahead forecast of y_{t+h} at time t is obtained by

$$\hat{y}_{t+h|t} = E[y_{t+h}|\Omega_t]$$

where Ω_t denotes the history of the time series up to and including the observation at time t.

Using the fact that $E[e_{t+1}|\Omega_t] = 0$, the one-step-ahead forecast in our model is easily obtained as

$$\begin{aligned}
\hat{y}_{t+1|t} &= E[y_{t+1}|\Omega_t] \\
&= E[F(y_t;\theta) + e_{t+1}|\Omega_t] \\
&= F(y_t;\theta) \\
&= \gamma\rho(y_t)y_t
\end{aligned}$$

where $\rho(t) = \frac{\exp(\alpha+\beta f(y_t))-1}{\exp(\alpha+\beta f(y_t))+1}$.

The forecast at $h = 2$ is given by

$$\begin{aligned}
\hat{y}_{t+2|t} &= E[y_{t+2}|\Omega_t] \\
&= E[F(y_{t+1};\theta) + e_{t+2}|\Omega_t] \\
&= E[F(y_{t+1};\theta)|\Omega_t] \\
&= E[F(F(y_t;\theta) + e_{t+1};\theta)|\Omega_t] \\
&= E[F(\hat{y}_{t+1|t} + e_{t+1};\theta)|\Omega_t].
\end{aligned}$$

(Lin and Granger, 1994; Franses and van Dijk, 2000).

Notice that

$$E[F(y_{t+1};\theta)|\Omega_t] \neq F(E[y_{t+1}|\Omega_t];\theta) = F(\hat{y}_{t+1|t};\theta).$$

The expected value of a function is generally not equal to the function of the expected value. The forecast will be biased in general and will not go to zero as the sample size becomes large(Brown and Mariano, 1989; Lin and Granger, 1994). Various methods to obtain more desirable multiple-step-ahead forecasts have been discussed. The Monte Carlo and bootstrap methods work well compared to other methods(Lin and Granger, 1994; Franses and van Dijk, 2000).

The 2-step-ahead Monte Carlo forecast is given by

$$\hat{y}_{t+2|t} = \frac{1}{k}\sum_{i=1}^{k} F(\hat{y}_{t+1|t} + e_i; \theta)$$

where k is some large number and e_i comes from the presumed distribution of e_{t+1}. The bootstrap forecast is given by

$$\hat{y}_{t+2|t} = \frac{1}{k}\sum_{i=1}^{k} F(\hat{y}_{t+1|t} + \hat{e}_i; \theta).$$

The residuals from the estimated model $\hat{e}_t, t = 1, \cdots, n$ are used with no assumption of the distribution of e_{t+1}, which is one advantage over the Monte Carlo method.

3.2 simulation

We look at data sets generated recursively from the model (3.1).

$$y_t = \gamma\rho(y_{t-1})y_{t-1} + e_t$$

where $|\gamma| < 1$ and $\rho(y_{t-1}) = \frac{\exp(\alpha+\beta f(y_{t-1}))-1}{\exp(\alpha+\beta f(y_{t-1}))+1}$ with $y_0 = 0$. We first deal with the case $f(y) = y$.

For this kind of nonlinear regression model, it is not easy to get a good asymptotic approximation to the finite sample behavior(Granger and Teräsvirta, 1993). Our case also shows the need for a large sample size to obtain a nearly normal distribution of $(\hat{\gamma}, \hat{\alpha}, \hat{\beta})$. The convergence rate of the distribution $(\hat{\gamma}, \hat{\alpha}, \hat{\beta})$ to a normal density seems to be dependent on the relative size of parameters. Also, we have much interest in the case where $|\gamma|$ is near 1. For estimation of parameters, we suggest γ be assumed known and not estimated rather than estimating (γ, α, β) at the same time. By doing that, we can use normal approximation for $\hat{\alpha}$ and $\hat{\beta}$ with moderate sample size. We will demonstrate that estimating α and β with γ set to near 1 gives good prediction one-step-ahead forecasting error.

The modified logistic function we are using is also called a hyperbolic tangent.

$$\tanh z = \frac{\sinh z}{\cosh z} = \frac{\exp(2z) - 1}{\exp(2z) + 1}$$

where $z = \frac{1}{2}(\alpha + \beta y)$. Because this varies between 1 and -1, the case where $\gamma = 1$ allows transitions between seemingly stationary and nonstationary behavior. Thus, fixing γ at 1 and estimating α and β would be of some practical interest. However, it appears that the distributional results here can only be obtained through simulation.

For $|\gamma| < 1$, theoretical results are obtainable as shown in section (3.1). We will consider cases where the true parameter is assumed and cases where the true γ is not the one assumed. The idea is that a hyperbolic tangent with amplitude γ over some observed domain might be near another with amplitude γ_1 provided α and β are adjusted appropriately. For example, if β is near 0 and $\alpha = \log 2$, then $\gamma \frac{\exp(\alpha+\beta f(y))-1}{\exp(\alpha+\beta f(y))+1}$ is close to $\frac{1}{3}\gamma$. Note if $\gamma_1 = 2\gamma$ and $\alpha = \log 1.4$, we again have $\gamma_1 \frac{\exp(\alpha+\beta f(y))-1}{\exp(\alpha+\beta f(y))+1} = \frac{1}{3}\gamma$. Clearly, this sort of computation would not always be possible. If the hyperbolic tangent is near 1 and γ is reduced to $\gamma_1 < \gamma$, it would not likely be possible to make up for the decrease by adjusting (α, β). The maximum value for $\gamma_1 \frac{\exp(\alpha+\beta f(y))-1}{\exp(\alpha+\beta f(y))+1}$ is $\gamma_1 < \gamma$. Thus, if we assume a known γ, it is better to err on the high size. With this in mind, we will investigate the case where a value of γ near but less than 1 is assumed and the true γ is less than assumed.

For showing the usefulness of the estimation with γ fixed at certain values, 5,000 data sets have been generated based on $(\gamma, \alpha, \beta) = (0.5, 2.0, 0.8)$, $(-0.5, 1.0, 0.6)$ and $(0.5, -3.0, 0.2)$ respectively where e_t comes from IID $N(0, 1)$. Then we estimate the parameters α and β with γ unknown, with γ fixed at the true value(*), and γ fixed at 0.99(*). The simulation results are shown in Table 3.1. "SAS PROC NLIN" is used for this particular nonlinear estimation.

Entries in Table 3.1 are explained in the following paragraphs. Notice that we have restricted $\beta > 0$ to ensure the geometric ergodicity of y_t using theorem A1.10 of Tong(1990) in section (3.1), and for $\theta' = (\gamma, \alpha, \beta)$ and $(-\gamma, -\alpha, -\beta)$,

$$\gamma \frac{\exp(\alpha + \beta f(y_{t-1})) - 1}{\exp(\alpha + \beta f(y_{t-1})) + 1} = -\gamma \frac{\exp(-\alpha - \beta f(y_{t-1})) - 1}{\exp(-\alpha - \beta f(y_{t-1})) + 1}.$$

Restricting $\beta > 0$ is the same as specifying the sign of γ. Hence, we suggest γ

have a positive sign as well as assuming γ known. Furthermore, under this estimation assumption, we can estimate the model where β is known to be 0 by obtaining nonsingularity of $F_t(\theta)$. When $\beta = 0$, $\gamma\rho(y_{t-1})$ becomes

$$\gamma\rho(y_{t-1}) = \gamma\frac{\exp(\alpha) - 1}{\exp(\alpha) + 1}$$

which is assumed to be $0.99\frac{\exp(\alpha)-1}{\exp(\alpha)+1}$ here. So, if the true γ is $\gamma_0 < 0.99$, one can find α to exactly make $0.99\frac{\exp(\alpha)-1}{\exp(\alpha)+1} = \gamma_0\frac{\exp(\alpha_0)-1}{\exp(\alpha_0)+1}$ for given γ_0 and α_0.

The series generated by smaller γ can be estimated effectively by fixing γ at 0.99. Doing this, we can still get an approximate normal distribution of $\hat{\alpha}$ and $\hat{\beta}$. For example, in the case of $(\gamma, \alpha, \beta) = (0.5, 2.0, 0.8)$, we get approximate normal distribution of $(\gamma, \hat{\alpha}, \hat{\beta})$ with Monte Carlo mean $(0.99, 0.6544, 0.2051)$. That is, our simulations indicate these numbers as means for α and β under $\gamma = 0.99$. The distributions of $\hat{\alpha}$ and $\hat{\beta}$ with γ fixed at 0.99 appear to be approximately normal while even the estimation of α and β with γ fixed at its true value rejects the normality in terms of JB statistics. Note, however, these distributions are not centered on the true α and β. While it may seem unusual to fix γ at some value, notice that the same thing is done is standard time series analysis. When one differences a series, they are assuming that a certain parameter is 1, rather than estimating it. This is common practice even though it is a well known fact that all parameters can be estimated consistently using, for example, least squares.

Of course, the key question is whether this scheme leads to good one-step-ahead forecasts. We address this by studying the mean squared error(MSE). The MSE ratios of the estimation with γ fixed at 0.99 over γ fixed at the true value and γ unknown are almost 1 showing the effectiveness of this nonlinear estimation.

The results could apply to the model (3.1) with serially correlated errors. We performed 5,000 nonlinear estimations for data sets generated based on $(\gamma, \alpha, \beta) = (0.5, 2.0, 0.8)$ and $(0.5, -3.0, 0.2)$ where η_t comes from AR(1) with coefficient(δ) 0.95, giving similar features(see (vi) and (vii) in Table 3.1).

We need to mention that the distribution of y_t of STAR is not well known. In general, one has to resort to numerical procedures or simulation to evaluate the distribution of y_t(Tong, 1990; Franses and van Dijk, 2000).

We are interested in whether a model with an assumed γ will produce similar low order moments to a model using the true (γ, α, β). To that end, we use the means of $(\hat{\alpha}, \hat{\beta})$ from our previous simulations. Based on simulation, y_t generated based on $(\gamma, \alpha, \beta) = (0.5, 2.0, 0.8)$ and $(0.99, 0.6544, 0.2051)$ both have estimated mean($E(y_t)$) 0.15 and standard deviation($Std(y_t)$) 1.07 with $n = 1,000$. For $(\gamma, \alpha, \beta) = (-0.5, 1.0, 0.6)$ and $(0.99, -0.4097, -0.2447)$, the mean is -0.10 and standard deviation is 1.03 for both. The mean of y_t is generally not 0 as the examples above indicate. Figure 3.8 and 3.9 show estimates of $E(\bar{y}_t)$, each obtained through 500 generated data sets, for various α and β with γ fixed at 0.99, $\sigma = 1.0$ and $\sigma = 0.5$ respectively.

We made another simulation about the distributions of $\hat{\alpha}$ and $\hat{\beta}$. For this simulation, we fix γ at 0.99 and $\alpha = 1.0$. 5,000 data sets have been generated for this and n is set to 1,000, 2,000, 3,000 respectively. e_t comes from $N(0, 1)$. β takes on the values -0.9 to 0.9 by steps of 0.3. For each $\hat{\alpha}$ and $\hat{\beta}$, 5,000 nonlinear regressions are run. The results are shown in Table 3.2.

Judging from JB statistics, the normality null hypothesis is rejected in most cases.

But, $Pr(|t| > z_{0.025})$, the proportion of exceedences combining the two tails seems reasonably consistent with the normal distribution for both estimators.

Notice that there seems to exist a trend in skewness for β.

The range of α and β such that a normal approximation holds depends on the sample size and variation scale of e_t. Figure 3.10 show a triangular area of true (α_0, β_0) parameters for which the t tests of $\alpha = \alpha_0$ and $\beta = \beta_0$ have empirical rejection rates not significantly different from 0.05, based on binomial test of $H_0 : p = 0.05$.

A vertical shifting term is easily incorporated by putting κ into the model.

$$y_t = \kappa + \gamma\rho(y_{t-1})(y_{t-1} - \kappa) + \eta_t$$

where

$$\rho(y_{t-1}) = \frac{\exp(\alpha + \beta f(y_{t-1} - \kappa)) - 1}{\exp(\alpha + \beta f(y_{t-1} - \kappa)) + 1}.$$

The distributional results of other parameters do not change much by adding the term.

Finally, we compare the nonlinear estimation with ARMA fits. The data have been generated using the same random number sequence $e_t \sim N(0, 1)$ for both models where $\gamma = 0.5$ or $\gamma = 0.99$ and $n = 2,000$. The series are estimated using an ARMA model and our NLAR(1) model with $\gamma = 0.99$ fixed. Interestingly, an ARMA fits each of the series generated by the model above quite well with no indication of lack of fit using the Ljung-Box statistic on the obtained residuals. The ARMA models have been chosen based on the Akaike information criterion(AIC). Table 3.3 shows that most of the χ^2 p values are 0.2 or higher, checking up to 48 lags, and most of the ARMA parameters have p values less than 0.001, so the fits are excellent by all the standard measures. Thus if such a NLAR model underlies a set of observed data, that fact would not easily be revealed by standard diagnostics. Notice that most of the ARMA fits give error variances bigger than the true innovations variance 1, while the nonlinear least square estimators, with $\gamma = 0.99$ fixed, result in estimates slightly under 1. The improvement in MSE is bigger where the data have been generated by $\gamma = 0.99$. There we have seen more asymmetric features for some series, especially as β increases. It is seen that fitting the true logistic autocorrelation by nonlinear least squares results in a nonnegligible improvement in the one-step-ahead prediction error variance versus ARMA models.

We do the same thing for $f(y) = |y|$ in model (3.1) and obtain similar results. One of the differences from the previous case where $f(y) = y$ is that the mean of the data generated by this is asymptotically 0 regardless of the parameter values(see Table 3.4,3.5,3.6 and Figure 3.11).

Nonlinear estimation can often be a more efficient way of estimation and prediction where the series have asymmetric volatility or severely changing amplitude with a rather persistent autocorrelation.

3.3 further analysis

The model with $\gamma = 1$ will be introduced here.

$$y_t = \rho(y_{t-1})y_{t-1} + e_t \tag{3.2}$$

where $\rho(y_{t-1}) = \frac{\exp(\alpha+\beta f(y_{t-1}))-1}{\exp(\alpha+\beta f(y_{t-1}))+1}$. Here $f(y) = |y|$ or $f(y) = y$, and e_t comes from an IID $N(0, \sigma^2)$ distribution. When the parameter space of $|\gamma|$ is not restricted, i.e., $|\gamma| = 1$, it is not easy to obtain the theoretical distribution of parameters. We can not say they are geometrically ergodic anymore except where $\beta = 0$.

The process generated by the model above could show a diverging behavior for a long period depending on e_t. For example, suppose e_t is replaced by a constant c for $t = n+1, n+2, \cdots$. It does not meet the condition of e_t by Tong(1990) for the geometric ergodicity of the series. But it can give a hint for the dynamics of y_t in a rather extreme way. For example, in a stationary AR(1) with parameter ρ, we would get convergence to the constant $\frac{c}{1-\rho}$, while a random walk would produce the linear sequence ct.

Figure 3.12 and 3.13 show the trend of y_t and $\rho(y_{t-1}) = \frac{\exp(1.0+0.6y_{t-1})-1}{\exp(1.0+0.6y_{t-1})+1}$ with c fixed at 4.0, 1.0, 0.0, -0.2, -1.0, -4.0 respectively. While the series always gets stable at $|\gamma| < 1$, y_t diverges sometimes where $\gamma = 1$. If $\rho(y_{t-1})$ hits near 1 very often, this could result in the instability of the series. Where $\beta = 0$, y_t always converges.

In general, there exist no constant K and $c > 0$ such that

$$|y_t| \leq Ke^{-ct}|y_0|.$$

Thus, we do not satisfy the assumptions of the theorems by Tong(1990). That is $\mathbf{0} = \mathbf{T}(\mathbf{0})$ is not exponentially asymptotically stable(Cline and Pu, 1999). Where

$\beta = 0$, we can always find a K and $c > 0$ satisfying the condition above.

Most discussions for the geometric ergodicity of the series, including those by Tong(1990), give sufficient conditions(Tweedie, 1975; Chan and Tong, 1985, 1994; Tjϕstheim, 1990; Meyn and Tweedie, 1994; An and Huang, 1996; Cline and Pu, 1999). The necessary and sufficient conditions seem to be found only for a very special kind of threshold AR(1) model(Petruccelli and Woolford, 1984; Chan, Petruccelli, Tong and Woolford, 1985; Chen and Say, 1991; Guo and Petruccelli, 1991).

The suggested model could be viewed as a variant of the STAR model. Chan and Tong(1986) have showed that in the STAR model,

$$\begin{aligned} y_t &= \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} \\ &+ (\phi_0' + \phi_1' y_{t-1} + \cdots + \phi_p' y_{t-p}) F((y_{t-d} - \gamma)/z) + e_t \end{aligned}$$

where $d \geq 0$ and $p \geq 1$, if (i) either $p = 1, d = 1, \phi_1 < 1, \phi_1 + \phi_1' < 1$, and $\phi_1(\phi_1 + \phi_1') < 1$ or (ii)$\sup_{0 \leq \theta \leq 1} \left(\sum_{i=1}^{p} |\phi_i + \theta \phi_i'| \right) < 1$, then y_t is ergodic and there is a unique stationary process satisfying the STAR difference equation. F is the standard Gaussian distribution. e_t are IID random variables independent of $y_s, s < t$ and is assumed to have finite second moment and zero mean.

Further, they have said that the condition (i) is 'almost' necessary and sufficient for the ergodicity of y_t because if $\phi_1 > 1, \phi_1 + \phi_1' > 1$, and $\phi_1(\phi_1 + \phi_1') > 1$ where $p = 1, d = 1$, then y_t is not ergodic(Chan and Tong, 1986).

The series generated with $|\gamma| = 1$ shows very similar features as with $|\gamma|$ near 1. However, it sometimes causes apparent nonstationarity which is the main reason for failure of convergence of any nonlinear estimation algorithm(see Figure 3.14 and 3.15). That is, for some parameter values, y_t eventually moves into a region where

$\rho(y_{t-1})$ is nearly 1. The series behaves like a random walk and thus wanders around in a region producing $\rho(y_{t-1})$ near 1 repeatedly. The series can remain "stuck" in this random-walk-like behavior for relatively long periods. This suggests that we can not make unambiguous conclusions about the geometric ergodicity of the series in the general parameter space of (α, β) under given innovations. Noise having finite support could be studied as suggested in Chan and Tong(1994).

In contrast, for $\beta = 0$, the model reduces to the usual AR(1) model and thus the asymptotic distribution of $\hat{\theta}_n$ can be proved clearly.

Under $H_0 : \beta = 0$ with $f(y) = y$ for model (3.2), the true $\rho(y_{t-1})$ is the constant $\rho_0 = \frac{\exp(\alpha_0)-1}{\exp(\alpha_0)+1}$. Suppose we estimate α and β. $F'_{nk}(\hat{\theta})F_{nk}(\hat{\theta})$ for the Gauss-Newton algorithm has the form

$$\begin{pmatrix} \frac{4\exp(2L_t)}{(\exp(L_t)+1)^4} y_{t-1}^2 & \frac{4\exp(2L_t)}{(\exp(L_t)+1)^4} y_{t-1}^3 \\ \frac{4\exp(2L_t)}{(\exp(L_t)+1)^4} y_{t-1}^3 & \frac{4\exp(2L_t)}{(\exp(L_t)+1)^4} y_{t-1}^4 \end{pmatrix}$$

where $L_t = \alpha + \beta y_{t-1}$.

Now assume we have initial consistent estimates

$$\hat{\alpha}_n - \alpha_0 = O_p(\frac{1}{\sqrt{n}}),$$

$$\hat{\beta}_n - \beta_0 = O_p(\frac{1}{\sqrt{n}}).$$

Because y_t is a simple AR(1) under H_0, the matrix $B(\theta_0)$ to which $\frac{1}{n}F'_{nk}(\theta)F_{nk}(\theta)$ converges in probability with $H_0 : \beta = 0$ can be found as

$$B(\theta_0) = \begin{pmatrix} \frac{4\exp(2\alpha_0)}{(\exp(\alpha_0)+1)^4} \frac{\sigma^2}{1-\rho_0^2} & 0 \\ 0 & \frac{4\exp(2\alpha_0)}{(\exp(\alpha_0)+1)^4} \frac{3\sigma^4}{(1-\rho_0^2)^2} \end{pmatrix}$$

and

$$n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N[0, B^{-1}(\theta_0)\sigma^2].$$

This result holds irrespective of the form of $f(y)$. Table 3.7 shows the distribution of $\hat{\alpha}$ and $\hat{\beta}$ coincides with the theoretical results where $\beta = 0$, even though the normality is not verified sometimes in terms of JB statistics at $|\alpha|$ far away from 0. For this, β has been fixed at 0 and α has changed from -3.6 to 3.6 by the step of 1.2. $n = 1,000$. We have estimated (α, β) for the generated data using the usual nonlinear least squares estimation algorithm. Other than this special case, there seems to be no theoretical way to get asymptotic distributions of parameters that hold over the whole parameter space.

Simulation indicates that certain parameter combinations result in behavior reminiscent of a unit process and in those cases parameters are difficult to obtain(convergence problems) or they have unusual distributions. To avoid such cases, we propose doing an initial unit root test, then fitting

$$y_t = \rho(y_{t-1})y_{t-1} + e_t,$$

only if we reject the unit root null hypothesis. In this way, we should avoid (α, β) settings that cause unit root behavior and resulting estimation problems.

Figure 3.16 and 3.17 show the distribution of $\hat{\alpha}$ and $\hat{\beta}$ where the series are generated by

$$y_t = \frac{\exp(1.0 + 0.6y_{t-1}) - 1}{\exp(1.0 + 0.6y_{t-1}) + 1} y_{t-1} + e_t.$$

(a) and (b) in Figures are obtained using all generated series and using only those series which reject a unit root hypothesis by the Augmented Dickey-Fuller(ADF) test respectively. Even if the sample size increases from $n = 1,000$ to $n = 3,000$, the distributions of $\hat{\alpha}$ and $\hat{\beta}$ in (a) do not seem close to normal or some other familiar density.

Now, as in Table 3.2 and 3.5 of section (3.2), we generated 5,000 data sets. n is set to 1,000, 2,000, 3,000 respectively and e_t comes from $N(0,1)$. β takes on the values -0.9 to 0.9 by steps of 0.3. Each series has been tested by the ADF test and R shows the number of replications which appear stationary based on that test and are convergent using the nonlinear estimation procedure. In this experiment, every series that appeared stationary has produced estimates, that is, convergence is obtained.

The simulation results are similar to those of $\gamma = 0.99$ fixed cases(see Table 3.8 and 3.9). In terms of JB, normality of the distributions of $\hat{\alpha}$ and $\hat{\beta}$ is rejected in most cases. But, for the 5% hypothesis tests, the normal approximation can be reasonably used under a certain range of parameter values, which is similar to that found in the $\gamma = 0.99$ fixed case. The region of parameters where the normal approximation is available based on $Pr(|t| > z_{0.025})$ seems to be somewhat reduced(see Figure 3.18 and 3.19).

Though the distributions of α and β look similar, the situation is different from the model with $|\gamma|$ near 1. The mean($E(\bar{y}_t)$) of series with $\gamma = 1$ does not get stable and the stationarity ratio does not improve with bigger sample size in certain cases. For example, 5,000 data sets with $(\alpha, \beta) = (1.0, 0.8)$ are generated for each of our sample sizes and the means for those series which reject unit roots are obtained(see Table 3.10). When $\gamma = 0.99$ or 0.999, the stationarity ratios increase and these means appear to approach some fixed value as n increases. But, with $\gamma = 1$, the stationarity ratios tend to decrease and the means appear to continue to rise even when $n = 13,000$. This means that the series generated based on $|\gamma| < 1$ and $|\gamma| = 1$ are quite different in nature even though the models look very similar.

For practical use, after ADF test, estimating with γ set to near 1 can be considered. As shown previously, the series generated with $\gamma = 1$ could lead to nonstationary behavior in the long run and severe local nonstationarity. That is, $\rho(y_{t-1})$ could be so close to 1 that the series behaves like a random walk for very long periods. Note that the series with $|\gamma|$ near 1 show very similar behavior to the series at $\gamma = 1$ after unit root hypothesis has been rejected.

We have generated 5,000 series based on $\gamma = 1.0$ and (α, β) as given in Table 3.11 and 3.12. Then, we estimate the parameters α and β with γ unknown, with γ fixed at the true value(*), and γ fixed at 0.99(*) for only the series which reject unit roots. Table 3.11 and 3.12 show the MSE ratios of the estimation with γ fixed at 0.99 over γ fixed at 1 and γ unknown are almost 1.

To summarize all simulation results for our suggested models, where $\gamma = 1$, we can obtain good parameter estimates by nonlinear least squares estimation using only those series which reject unit roots hypothesis by the ADF test. Estimating with γ set to near 1 is helpful to get better convergence and distributional results of parameter estimates. The one-step-ahead forecasting errors seem good for the estimation with γ fixed near 1. The estimates obtained in this way could be used as initial values for the estimation of all parameters at the same time.

3.4 application

As mentioned in section (3.2), the model (3.1) with serially correlated errors can be displayed as a LSTAR model with many regimes. It is not easy to estimate many regimes at once including the identification of order p and delay factor d. There is a possibility that the usual STAR fitting process, because of its generality, does not work well for the data generated by our suggested NLAR models. We show this through an artificially generated data set. The data are generated based on

$$y_t = 0.99\frac{\exp(0.8 + 0.45y_{t-1}) - 1}{\exp(0.8 + 0.45y_{t-1}) + 1}y_{t-1} + \eta_t,$$

and

$$\eta_t = 0.95\eta_{t-1} + e_t.$$

The series has a strong persistent autocorrelation as shown in Figure 3.7 (b).

We fit the model by

$$y_t = 0.99\frac{\exp(0.6906 + 0.4867y_{t-1}) - 1}{\exp(0.6906 + 0.4867y_{t-1}) + 1}y_{t-1} + \eta_t,$$
$$(0.0429)(0.0342)$$

and

$$\eta_t = 0.9593\eta_{t-1} + e_t.$$
$$(0.0056)$$

We want to fit this data using a LSTAR. The estimation procedure of STAR models is well introduced in Teräsvirta(1994). The major steps are

(i) specification of a linear AR(p) model,

(ii) testing linearity for different values of the delay parameter d and, if it is rejected, simultaneously determining d,

(iii) choosing between LSTAR and ESTAR

(Granger and Teräsvirta, 1993; Teräsvirta, 1994).

The AIC or the SBC may be used to find an appropriate linear model in the step (i). The Ljung-Box test statistic is also considered to avoid the effects of any remaining serial correlation on the linearity tests(Granger and Teräsvirta, 1993; Hall, Skalin, and Teräsvirta, 2001).

Judging from the AIC and the Ljung-Box statistics, AR(8) seems to be appropriate for order p of the model.

Some statistics for determining p in a linear AR model

	AIC	SBC	p(lag 6)	p(lag 12)
$p = 1$	-3460.00	-3454.00	0.00	0.00
$p = 2$	-4575.96	-4563.94	0.00	0.00
$p = 3$	-4653.20	-4635.18	0.00	0.00
$p = 4$	-4677.38	-4653.36	0.03	0.00
$p = 5$	-4684.16	-4654.12	0.02	0.00
$p = 6$	-4690.48	-4654.44	-	0.01
$p = 7$	-4691.40	-4649.36	-	0.01
$p = 8$	-4704.44	-4656.39	-	0.36
$p = 9$	-4702.52	-4648.46	-	0.21
$p = 10$	-4700.52	-4640.46	-	0.10

Note: p means p-values of the Ljung-Box statistics at lag 6 and lag 12.

Notice that the true series is the one such that the order p is 2 and it consists of two logistic functions with $d = 1$ and $d = 2$. So, the model will be misspecified in this case as we know $p = 2$.

$$\begin{aligned} y_t &= 0.99\frac{\exp(0.8+0.45y_{t-1})-1}{\exp(0.8+0.45y_{t-1})+1}y_{t-1} \\ &+ 0.95(y_{t-1} - 0.99\frac{\exp(0.8+0.45y_{t-2})-1}{\exp(0.8+0.45y_{t-2})+1}y_{t-2}) + e_t \\ &= 1.94y_{t-1} - 0.9405y_{t-2} - 1.98y_{t-1}\frac{1}{\exp(0.8+0.45y_{t-1})+1} \\ &+ 1.881y_{t-2}\frac{1}{\exp(0.8+0.45y_{t-2})+1} + e_t. \end{aligned}$$

For the linearity test and the search for the delay factor d, we run the following auxiliary regression.

$$\begin{aligned} y_t &= \beta_0 + \sum_{j=1}^{p}\beta_{1j}y_{t-j} + \sum_{j=1}^{p}\beta_{2j}y_{t-j}y_{t-d} \\ &+ \sum_{j=1}^{p}\beta_{3j}y_{t-j}y_{t-d}^2 + \sum_{j=1}^{p}\beta_{4j}y_{t-j}y_{t-d}^3 + \eta_t \end{aligned}$$

and test $H_0 : \beta_{2j} = \beta_{3j} = \beta_{4j} = 0, j = 1, \cdots, p = 8$.

LM-test statistics, approximately F distributed, are 23.09, 24.23, 23.38, 22.44, 20.95, 19.61, 18.92, and 18.43 for each $d = 1, \cdots, 8$. This test statistic is used for linearity test against STAR under the null hypothesis of linearity(Luukkonen, Saikkonen and Teräsvirta, 1988; Teräsvirta, 1994; Franses and van Dijk, 2000). All F statistics are bigger than $1.52(\approx F_{n-4p-1=\infty}^{3p=24}, \alpha = 0.05)$. Linearity against STAR is rejected at the 5% significance level.

Also, the p-values of F statistics have the smallest value at $d = 2$. The delay factor seems to be $d = 2$ in the sense that the test should have the maximum power, if the alternative model is correctly specified, i.e., if the correct transition variable is used(van Dijk, Teräsvirta, and Franses, 2002). Hence, the model could be erroneously estimated as a STAR with $p = 8$ and $d = 2$ using the usual STAR estimation procedure.

Ten series are randomly generated to see how p is misspecified using the usual STAR fitting process. The true models are

(i)

$$y_t = 0.99\frac{\exp(1.0 + 0.8y_{t-1}) - 1}{\exp(1.0 + 0.8y_{t-1}) + 1}y_{t-1} + e_t,$$

and

(ii)

$$y_t = 0.99\frac{\exp(0.8 + 0.45y_{t-1}) - 1}{\exp(0.8 + 0.45y_{t-1}) + 1}y_{t-1} + \eta_t$$

with

$$\eta_t = 0.95\eta_{t-1} + e_t.$$

The p which gives the minimum AIC and no serious autocorrelation by the Ljung-Box statistics in each series is used to select a linear AR approximation. As the table shows, p is misspecified in all cases. However, the delay factor d seems to be well specified in this experiment.

		1	2	3	4	5	6	7	8	9	10
model (i)	p	5	6	9	4	5	3	4	3	7	5
	d	1	1	1	1	1	1	1	1	1	1
model (ii)	p	4	10	4	5	4	3	5	4	8	10
	d	2	1	2	3	2	1	1	2	2	1

Now, back to the series introduced in the beginning, we already know that p is 2 and d is at most 2 in the true generated series. For comparison, we have estimated the model with $p = 2$ and one logistic function $d = 1$ or $d = 2$ assuming wrongly a LSTAR with only 2 regimes.

For $d = 1$, we get

$$
\begin{aligned}
y_t &= \underset{(0.0271)}{2.0095}y_{t-1} - \underset{(0.0271)}{1.0118}y_{t-2} \\
&+ \frac{1}{\exp(\underset{(0.4424)}{-0.1831} + \underset{(0.2717)}{1.2120}y_{t-d}) + 1}[\underset{(0.2575)}{-1.3277}y_{t-1} + \underset{(0.2515)}{1.2826}y_{t-2}] + e_t.
\end{aligned}
$$

For $d = 2$,

$$
\begin{aligned}
y_t &= \underset{(0.0268)}{2.0047}y_{t-1} - \underset{(0.0268)}{1.0068}y_{t-2} \\
&+ \frac{1}{\exp(\underset{(0.4724)}{0.0866} + \underset{(0.2428)}{1.0886}y_{t-d}) + 1}[\underset{(0.3707)}{-1.5346}y_{t-1} + \underset{(0.3594)}{1.4835}y_{t-2}] + e_t.
\end{aligned}
$$

The MSEs for the last two models are bigger than that based on the true model. The MSE based on the true model is 0.0102 compared with 0.0104 and 0.0103 respectively. For the last two models, we can proceed to a model with more regimes through the remaining nonlinearity test. The remaining nonlinearity tests against 3 or 4 regimes have been suggested in Eitrheim and Teräsvirta(1996), and van Dijk and Franses(1999). However, the procedure could be quite complicated for higher order serially correlated errors.

For real data analysis, we analyze the stream flow series in Goldsboro and Kinston North Carolina. The data have been introduced in Chapter 2 for a transfer function analysis. Here, we study two series using our suggested models.

First, we deal with the stream flow series in Goldsboro. The series is log-transformed as in Chapter 2. The unit root hypothesis is rejected for this series at the significance level $\alpha = 0.05$ with the mean term added. The stream flows are expected to have a slowly decaying autocorrelation like a long memory process. As shown in Figure 3.20, the series has an autocorrelation function which decreases slowly and stays, more or less, at a constant and significant level for a long time. This characteristic coincides with the feature of autocorrelation functions by our suggested NLAR model.

To begin with, we estimate the series using an ARMA model. The fitted ARMA model is

$$\begin{aligned} y_t &= 7.2940 + (1 - 1.3948B + 0.4262B^2)^{-1}e_t. \\ & \quad (0.3083) \qquad\quad (0.0452) \quad (0.0453) \end{aligned}$$

There does not seem to exist serial correlation in residuals and the MSE is 0.04078.

We have done a linearity test based on Theorem 8.6.1 of Fuller(1996). It is important to determine whether a nonlinear model is an adequate representation of the process generating the data before building a nonlinear time series model(Teräsvirta, 1994). We have estimated y_t using quadratic polynomial terms y_{t-1}^2, y_{t-2}^2 and $y_{t-1}y_{t-2}$ as explanatory variables based on the fitted result above and found that the F statistic for testing the hypothesis that the coefficient of the quadratic term is zero is $11.36 (> F_\infty^3 \approx 2.60, \alpha = 0.05)$. The hypothesis of a zero coefficient is rejected.

We have fitted the series using an autoregressive nonlinear model with a hyperbolic tangent function and obtained a better result than a standard ARMA model.

The fitted model using a NLAR model is

$$
\begin{aligned}
y_t &= 0.99\frac{\exp(0.5058 + 0.3783(y_{t-1} - 7.1295)) - 1}{\exp(0.5058 + 0.3783(y_{t-1} - 7.1295)) + 1}(y_{t-1} - 7.1295) \\
&\qquad\quad (0.3580) \quad (0.0713) \qquad (0.2831) \\
&+ 7.1295 + \eta_t,
\end{aligned}
$$

and

$$
\begin{aligned}
\eta_t &= 1.2766\eta_{t-1} - 0.3234\eta_{t-2} + e_t. \\
&\quad (0.1366) \qquad (0.1295)
\end{aligned}
$$

The estimates are all significant at the 5% level except for that of α and the MSE is 0.03656. The residuals seem to have no significant autocorrelation left and no ARCH effects are detected(ARCH(1)=0.8271, ARCH(4)=0.1362, ARCH(6)=0.3159, ARCH(q) is the p-value for the ARCH LM test of no ARCH effects up to order q). ARCH test for residuals is often used for a diagnostic check for the fitted nonlinear model and the standard tests of constant conditional variance against ARCH have power against nonlinearity in the conditional mean(Granger and Teräsvirta, 1993; van Dijk, Teräsvirta, and Franses, 2002). Figure 3.20 shows the predicted values of the stream flow series and $\rho(y_{t-1})$ against y_{t-1}.

Now, we fit the stream flow series in Kinston. The series is log-transformed as well and the unit root hypothesis is rejected at the 5% significance level with the mean term added. Figure 3.21 shows the features of the series. Using an ARMA(5,1), we have obtained the following result.

$$
y_t = (1 - 2.4248B + 2.1704B^2 - 1.0866B^3 + 0.5338B^4 - 0.1909B^5)^{-1}
$$

$$
\begin{array}{l}
\qquad (0.1165) \quad (0.2128) \qquad (0.1755) \qquad (0.1285) \qquad (0.0506) \\
(1 - 0.7980B)e_t + 7.5012. \\
\qquad (0.1116) \qquad (0.5750)
\end{array}
$$

There does not seem to exist serial correlation in residuals and the MSE is 0.01526. Notice that a near unit root seems to be found in the fitted AR coefficients, which is difficult to justify in stream flow series. A first differenced series could be modeled.

We fit our suggested NLAR model again for this series. The linearity hypothesis(based on Fuller(1996) using all quadratic polynomial terms of $y_{t-1}, \cdots, y_{t-5}$) is rejected at the 5% significance level($F = 2.83 > F_{\infty}^{15} \approx 1.67, \alpha = 0.05$). The estimation result is as follows. All variables with t-ratio bigger than 1 are included in this model.

$$
\begin{aligned}
y_t &= 0.99 \frac{\exp(1.3315 + 0.4170(y_{t-1} - 7.4773)) - 1}{\exp(1.3315 + 0.4170(y_{t-1} - 7.4773)) + 1}(y_{t-1} - 7.4773) \\
&\qquad\quad (0.4712) \quad (0.1200) \qquad (0.3419) \\
&+ 7.4773 + \eta_t,
\end{aligned}
$$

and

$$
\begin{aligned}
\eta_t &= 1.1460\eta_{t-1} - 0.3503\eta_{t-2} + 0.2292\eta_{t-3} - 0.0998\eta_{t-4} \\
&\quad (0.1560) \qquad (0.1330) \qquad (0.0775) \qquad (0.0640) \\
&- 0.0975\eta_{t-7} + 0.1264\eta_{t-8} + e_t. \\
&\quad (0.0550) \qquad (0.0492)
\end{aligned}
$$

The estimates for α and β are significant at the 5% significance level. No serial correlation is found in residuals and serious ARCH effects do not seem to be detected unlike the ARMA model(ARCH(1)=0.0122, ARCH(4)=0.1321, ARCH(6)=0.2837). The obtained MSE is 0.01439. Figure 3.21 shows the predicted y_t and $\rho(y_{t-1})$ against y_{t-1}.

Based on the fitting results so far, NLAR(1) models with serially correlated errors seem to yield competitive one-step-ahead forecasting errors for the stream flow series in both regions. As a reference, the MSE obtained here for Kinston is almost 2.5 times bigger than in the transfer function type model where another explanatory variable is added rather than its lagged own terms.

We fit the weekly Soybean price series in North Carolina from March 1, 1982 to March 21, 1999, which consists of the 899 observations as an another example. There seems no noticeable seasonality in the series judging from the periodgram analysis(see Figure 3.22 (b)). The unit root hypothesis is rejected with the mean term added. Figure 3.22 shows that the series has a persistent autocorrelation and a strong partial antocorrelation at the first lag.

As usual, we fit the series using an ARMA model. AR(1) seems appropriate, but the Ljung-Box statistics show that there exists a significant autocorrelation in residuals at some low lags. The obtained MSE is 0.042482.

$$
\begin{aligned}
y_t &= 6.3323 + 0.9791(y_{t-1} - 6.3323) + e_t. \\
& (0.3114)(0.0068)
\end{aligned}
$$

We did a linearity test using the terms y_{t-1} and y_{t-1}^2 as explanatory variables. The F statistic for the test that the coefficient of the quadratic term is zero shows 7.87, which is bigger than $F_{\infty}^1 \approx 3.84$ at $\alpha = 0.05$. The linearity hypothesis is rejected at the 5% significance level.

We have analyzed the data using our suggested model. Using the values obtained with γ set to 0.99 as initial values for nonlinear least squares estimation, we get the following result.

$$
\begin{aligned}
y_t &= \underset{(0.0060)}{0.9976} \frac{\exp(\underset{(2.2349)}{4.7507} - \underset{(0.6218)}{1.3765}(y_{t-1} - \underset{(0.7190)}{7.7129})) - 1}{\exp(4.7507 - 1.3765(y_{t-1} - 7.7129)) + 1}(y_{t-1} - 7.7129) \\
&+ 7.7129 + e_t.
\end{aligned}
$$

The estimates are all significant at the 5% level and the obtained MSE is 0.041701, much reduced over that of the AR(1). However, there still seems to be autocorrelation left in residuals adjusting the χ^2 degrees of freedom. Removing the autocorrelation in residuals up to lag 6, we obtain the fitting model such as

$$
\begin{aligned}
y_t &= \underset{(0.0066)}{0.9980} \frac{\exp(\underset{(1.9499)}{4.6409} - \underset{(0.5817)}{1.2688}(y_{t-1} - \underset{(0.6500)}{7.6877})) - 1}{\exp(4.6409 - 1.2688(y_{t-1} - 7.6877)) + 1}(y_{t-1} - 7.6877) \\
&+ 7.6877 + \eta_t,
\end{aligned}
$$

and

$$
\eta_t = \underset{(0.0345)}{0.0693}\eta_{t-2} - \underset{(0.0345)}{0.0727}\eta_{t-5} + e_t. \tag{3.3}
$$

Now the MSE is 0.041572, least of all and there seems to be no significant autocorrelation left at low lags, even though we have adjusted the degrees of freedom for χ^2 statistics. We have estimated the series removing the autocorrelation up to lag 11, but no particular difference is observed in parameter estimates. The MSE there is 0.041153.

As mentioned previously, significant serial correlation in residuals is not found at low lags, but ARCH effects are detected for all models. The linearity of the error process could be tested against STAR to see whether the fitted model has captured all of the nonlinearity(Granger and Ter$\ddot{a}$svirta,1993). Because no serious correlation is left, $p = 1$ is used as the AR order for the tests. F statistics are 0.35, 8.41, 2.53, 2.78, 1.20, 5.07 and 1.72 for each $d = 1, \cdots, 7$. There seems to be no remaining nonlinearity at most of the delay factors($\mathrm{F}^3_\infty \approx 2.60, \alpha = 0.05$). But some nonlinearity is found at $d = 2, 4$ and 6.

Notice that neglected heteroscedasticity of the series may result in the spurious rejection of the linearity hypothesis(Granger and Ter$\ddot{a}$svirta, 1993; Franses and van Dijk, 2000).

For example, we did a nonlinearity test against STAR for a GARCH generated data by

$$\epsilon_t = z_t \sqrt{h_t},$$

and

$$h_t = 0.001 + 0.5\epsilon^2_{t-1} + 0.4h_{t-1}$$

where z comes from $N(0, 1)$ and $n = 3,000$. Ten series with no serial correlation are chosen for this experiment. Since there is no serial correlation left in the series, $p = 1$ is used as the AR order for the tests and the delay factor d changes from 1 to 8. The test statistics are F distributed($F^3_\infty \approx 2.60, \alpha = 0.05$).

The table below shows that the linearity hypothesis against STAR is rejected many times for the nonlinearity generated by GARCH. This implies that the nonlinearity test against STAR is sensitive to GARCH effects.

	$d = 1$	2	3	4	5	6	7	8
1	13.93	1.67	1.13	10.54	7.53	3.31	0.34	2.92
2	2.52	8.07	6.88	6.11	1.48	5.48	4.02	2.00
3	14.52	37.33	4.28	13.51	2.42	6.86	1.07	1.42
4	4.18	1.57	3.17	1.38	4.33	8.08	2.62	1.66
5	9.05	4.09	9.95	1.70	10.86	5.57	3.21	9.82
6	4.52	16.52	3.16	4.49	0.80	7.38	1.59	3.34
7	9.87	5.78	1.90	3.45	19.49	7.62	8.13	0.26
8	13.04	0.49	9.09	4.27	0.93	1.82	1.08	4.40
9	5.70	18.58	15.77	7.77	6.26	2.53	3.02	0.45
10	2.18	3.30	23.00	17.25	2.13	5.69	8.58	6.67

For this, robust tests for nonlinearity under heteroscedastity have been suggested (Granger and Ter$\ddot{a}$svirta, 1993; Franses and van Dijk, 2000).

The procedure against STAR is that we regress y_t on $\mathbf{s_t}= (1, y_{t-1}, \cdots, y_{t-p})$ and get the residuals $\hat{w}_t, t = 1, \cdots, n$. We regress the auxiliary regressors $(y_{t-1}y_{t-d}, \cdots, y_{t-p}y_{t-d}, y_{t-1}^2 y_{t-d}, \cdots, y_{t-p}^2 y_{t-d}, y_{t-1}^3 y_{t-d}, \cdots, y_{t-p}^3 y_{t-d})$ on $\mathbf{s_t}$ and compute the residuals $\hat{r}_t$. Then we weight the residuals $\hat{r}_t$ by $\hat{w}_t$ and regress $\mathbf{1}$ on the weighted residuals. The explained sum of squares from this regression is the test statistic. It is χ^2_{3p} distributed. This robust test is not recommended to find and model any nonlinearity of the original series, because the robustification often weakens the power of linearity tests, and leads to the failure of existing nonlinearity detection. It is rather considered at the evaluation stage of model building(Lundbergh and Ter$\ddot{a}$svirta, 1998; van Dijk, Ter$\ddot{a}$svirta and Franses, 2002).

We use this robust test as a diagnostic check for the nonlinearity of residuals. Applying a heteroscedasticity-consistent variant of the LM-type test statistic against STAR, we get χ^2 statistics 0.31, 5.37, 1.98, 3.16, 1.75, 2.84 and 1.11 for $d = 1, \cdots, 7$ respectively($\chi^2_3 \approx 7.81, \alpha = 0.05$). Nonlinearity from STAR does not seem to exist in residuals.

However, considering ARCH effects are detected, we try to fit a STAR model, which is more flexible. We did linearity tests against STAR. $p = 1$ seems reasonable for this. Changing the delay factor d from 1 to 7 in sequence, we get the F statistics 6.80, 3.70, 4.95, 3.71, 4.00, 1.88 and 1.64. The linearity hypothesis against STAR is rejected at $d = 1, \cdots, 5 (F^3_\infty \approx 2.60, \alpha = 0.05)$, and $d = 1$ is chosen for the delay factor.

For the choice of LSTAR or ESTAR, two methods are generally used. One way suggested by Teräsvirta(1994) is to use the linearity test against STAR model.

$$\begin{aligned} y_t &= \beta_0 + \sum_{j=1}^{p} \beta_{1j} y_{t-j} + \sum_{j=1}^{p} \beta_{2j} y_{t-j} y_{t-d} \\ &+ \sum_{j=1}^{p} \beta_{3j} y_{t-j} y_{t-d}^2 + \sum_{j=1}^{p} \beta_{4j} y_{t-j} y_{t-d}^3 + \eta_t. \end{aligned}$$

The hypothesis tested within the auxiliary regression is $H_{01} : \beta_{4j} = 0, H_{02} : \beta_{3j} = 0 | \beta_{4j} = 0, H_{03} : \beta_{2j} = 0 | \beta_{3j} = \beta_{4j} = 0$. If p-value of the test corresponding to H_{02} is smallest, then an ESTAR is selected, while a LSTAR is selected in other cases. As mentioned previously, the test $H_{LS} : \beta_{2j} = \beta_{3j} = \beta_{4j} = 0$ is used for testing linearity against STAR model, especially the LSTAR model(van Dijk, Teräsvirta, and Franses, 2002).

Escribano and Jordá(1999) have introduced another method. For the auxiliary regression

$$\begin{aligned} y_t &= \beta_0 + \sum_{j=1}^{p} \beta_{1j} y_{t-j} + \sum_{j=1}^{p} \beta_{2j} y_{t-j} y_{t-d} \\ &+ \sum_{j=1}^{p} \beta_{3j} y_{t-j} y_{t-d}^2 + \sum_{j=1}^{p} \beta_{4j} y_{t-j} y_{t-d}^3 \\ &+ \sum_{j=1}^{p} \beta_{5j} y_{t-j} y_{t-d}^4 + \eta_t, \end{aligned}$$

H_{0E} : $\beta_{3j} = \beta_{5j} = 0$ and H_{0L} : $\beta_{2j} = \beta_{4j} = 0$ are tested. If H_{0E} is more strongly rejected, then an ESTAR is selected. The test H_{ES} : $\beta_{2j} = \beta_{3j} = \beta_{4j} = \beta_{5j} = 0$ is used for testing linearity against ESTAR model(van Dijk, Teräsvirta, and Franses, 2002). However, there is no clear cut preferred method for this and both can be well fitted if both functions have similar shape within the range of the transition variable.

We have done the tests for choosing a STAR model. The AIC statistics show the local minimum at lag 1,6,11,16 reflecting the strong partial autocorrelation at lag 1 and some significant partial autocorrelations at other lags(see Figure 3.22 (f)). As a linear basis for AR(p), $p = 6, 11$ and 16 as well as $p = 1$ are also considered. Both tests choose an ESTAR as an appropriate model at $p = 1, 6, 11$ and 16.

		test (I) H_{01}	H_{02}	H_{03}		test (II) H_{0E}	H_{0L}		H_{LS}	H_{ES}
$p = 1$	$d = 1$	1.13	11.32	7.87	E	0.47	0.24	E	6.80(0.00)	5.13(0.00)
	$d = 2$	2.34	8.17	0.58	E	2.89	2.38	E	3.70(0.01)	3.82(0.00)
	$d = 3$	3.23	6.53	5.03	E	1.58	1.04	E	4.95(0.00)	3.86(0.00)
	$d = 4$	1.79	5.61	3.69	E	0.74	0.39	E	3.71(0.01)	2.81(0.02)
	$d = 5$	1.11	6.13	4.72	E	0.85	0.63	E	4.00(0.01)	3.23(0.01)
	$d = 6$	0.94	3.32	1.39	E	7.56	7.89	L	1.88(0.13)	4.91(0.00)
	$d = 7$	1.79	1.93	1.20	E	4.83	5.11	L	1.64(0.18)	3.14(0.01)
$p = 6$	$d = 4$	6.91	7.24	3.46	E	6.09	5.79	E	6.07(0.00)	6.02(0.00)
$p = 11$	$d = 4$	4.78	7.21	3.06	E	3.90	3.69	E	5.27(0.00)	4.84(0.00)
$p = 16$	$d = 4$	4.12	6.29	2.92	E	3.84	3.70	E	4.73(0.00)	4.60(0.00)

Note: "E" and "L" mean that an ESTAR or a LSTAR will be selected based on the test results. The figures in parenthesis show the p-values of F statistics.

Here, we need to mention that the tests for choosing STAR models, i.e., ESTAR or LSTAR, could misspecify the appropriate model for the NLAR(1) with serially correlated errors. Some experiments with ten randomly generated series using the coefficients of the fitted model (3.3), where the shifting term is zero and σ^2 is 0.04, show that H_{02} is more strongly rejected in favor of an ESTAR in many cases following the test procedure suggested by Teräsvirta(1994). Escribano and Jordá(1999) chooses a LSTAR.

As in Figure 3.22 (f), every generated series shows a strong partial autocorrelation at lag 1 and minor significant partial autocorrelation at some other lags. Considering this, we choose $p = 1$ and find p which gives the minimum AIC with no serious autocorrelation by the Ljung-Box statistics. For each series, this is our basis for a linear AR model. At both ps, we find the delay factor d and make tests for choosing an ESTAR or a LSTAR. The following table shows the results.

			test H_{01}	(I) H_{02}	H_{03}		test H_{0E}	(II) H_{0L}	
1	$p=1$	$d=5$	2.62	17.88	2.92	E	3.86	8.65	L
	$p=3$	$d=5$	2.50	6.02	3.82	E	3.01	4.59	L
2	$p=1$	$d=1$	4.41	4.34	9.50	L	5.09	7.83	L
	$p=3$	$d=4$	2.51	2.15	4.69	L	3.09	4.61	L
3	$p=1$	$d=3$	9.29	9.44	0.26	E	8.03	10.13	L
	$p=4$	$d=4$	2.66	3.30	0.49	E	1.67	1.85	L
4	$p=1$	$d=3$	3.15	9.40	4.10	E	3.52	7.56	L
	$p=7$	$d=1$	1.26	2.53	1.21	E	1.38	1.66	L
5	$p=1$	$d=3$	2.51	8.61	0.37	E	5.08	6.98	L
	$p=3$	$d=2$	0.74	2.18	1.07	E	1.10	1.75	L
6	$p=1$	$d=1$	4.07	12.06	1.74	E	6.52	9.56	L
	$p=8$	$d=4$	1.58	2.70	0.86	E	1.65	2.07	L
7	$p=1$	$d=3$	8.49	22.33	8.38	E	9.61	17.18	L
	$p=6$	$d=1$	1.78	3.93	2.55	E	1.82	3.15	L
8	$p=1$	$d=2$	2.16	4.07	1.19	E	4.18	5.41	L
	$p=6$	$d=2$	1.45	1.83	1.51	E	1.27	1.49	L
9	$p=1$	$d=3$	6.65	9.68	1.49	E	9.40	11.88	L
	$p=6$	$d=4$	3.54	1.92	0.61	L	2.69	2.80	L
10	$p=1$	$d=1$	0.91	10.08	5.91	E	4.32	4.51	L
	$p=6$	$d=1$	1.45	2.52	1.65	E	2.03	2.40	L

Note: “E” and “L” mean that an ESTAR or a LSTAR will be selected based on the test results.

Two tests are different most of time in their results. Also, notice that the model used to generate the series is the MRLSTAR with $p = 6$ and $d = 1, 3, 6$. In this kind of model, d as well as p are often misspecified.

Now, back to the soybean series, an ESTAR is selected as the appropriate STAR model using two sets of tests. However, it could be a LSTAR judging from the experiment results. Also, H_{LS} is more strongly rejected than H_{ES} at $p = 1$ and $d = 1$. We have tried several STAR models including LSTAR at $p = 1, d = 1$ and $p = 6, d = 4$, in which the coefficients are more flexible, but no satisfactory results have not been found than our suggested models.

Figure 3.23 shows the pattern of the predicted values and $\rho(y_{t-1})$ using our suggested model. $\rho(y_{t-1})$ changes from about 0.73 to 1 within the range of y_t. If 1 is used as a value of $\rho(y_{t-1})$ in model (3.3), the resulting constant coefficient AR model has a biggest root 0.9980, while substitution of 0.73 for $\rho(y_{t-1})$ shows a biggest root $0.7285 (= 0.998 \times 0.73)$. The logistic function produces an interesting behavior in prices. When the level gets unusually high(June 19, 1988, around time 340, for example), the correlation becomes smaller and the next observation tends to drop toward the mean somewhat quickly. That initial drop implies an increased autocorrelation, thus slowing the rate of descent. An asymmetrical characteristic of price behavior can be well explained.

Even though, there seems to be no remaining nonlinearity in the estimated model, ARCH effects still remain in the residuals. Also, the distribution of the residuals has a higher kurtosis than normally distributed data.

STAR-GARCH models have been suggested lately(Lundbergh and Teräsvirta, 1998; Chan and McAleer, 2002). The assumption that the error sequence in STAR models has a constant conditional variance is normally not realistic when modeling high-frequency financial series(Lundbergh and Teräsvirta, 1998).

The STAR-GARCH model allows ϵ_t to follow a GARCH process.

$$\begin{aligned} y_t &= \Big(\phi_{10} + \sum_{i=1}^{r} \phi_{1i} y_{t-i+1}\Big)(1 - G(s_t;\gamma,c)) \\ &+ \Big(\phi_{20} + \sum_{i=1}^{r} \phi_{2i} y_{t-i+1}\Big) G(s_t;\gamma,c) + \epsilon_t, \end{aligned}$$

and

$$\epsilon_t = z_t \sqrt{h_t}$$

where

$$z_t \sim i.i.d.(0,1),$$

$$h_t = w + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i}.$$

The transition function $G(s_t;\gamma,c)$ is a continuous function that is bounded between 0 and 1. Usually, the logistic function

$$G(s_t;\gamma,c) = (1 + \exp\{-\gamma(s_t - c)\})^{-1}$$

where $\gamma > 0$ and the exponential function

$$G(s_t;\gamma,c) = 1 - \exp\{-\gamma(s_t - c)^2\}$$

where $\gamma > 0$ are used. s_t could be a lagged endogenous variable, an exogenous variable, a time trend and a function of them. The resultant models are called the logistic STAR(LSTAR) and the exponential STAR(ESTAR) model respectively.

The regularity conditions for stationarity of the GARCH component or for the existence of its moments, and statistical properties relating to the GARCH component are not well known(Chan and McAleer, 2002). Chan and McAleer(2002,2003) have proved that under the following assumptions regarding the structural and asymptotic properties of the STAR model,

(i) the process generating the STAR model is strictly stationary and ergodic,

(ii) the necessary and sufficient conditions for the existence of moments are satisfied,

(iii) the maximum likelihood estimators of parameters in the STAR model are consistent and asymptotically normal,

the STAR-GARCH model given above has a unique, second-order stationary solution, and $\{y_t, \epsilon_t, h_t\}$ are strictly stationary and ergodic. It extends to MRSTAR, too.

In addition, where $p = 1$ and $q = 1$ in the GARCH lags, if

$$E(\log(\alpha_1 z_t^2 + \beta_1)) < 0$$

then, $\hat{\theta}$ which maximizes the likelihood function $l(\theta)$

$$l(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\Big(\log h_t + \frac{\epsilon_t^2}{h_t}\Big)$$

for the STAR-GARCH model defined, is consistent for θ_0 and asymptotically normal(Chan and McAleer, 2002, 2003).

The log-moment condition can be replaced by the condition $E(\epsilon_t^2) < \infty$. Empirically, the moment conditions are obtained by

$$T^{-1}\sum_{t=1}^{T}\log(\hat{\alpha}_1 \hat{z}_t^2 + \hat{\beta}_1) < 0,$$

and

$$\hat{\alpha}_1 + \hat{\beta}_1 < 1,$$

respectively.

We suggest the conditional mean of the STAR-GARCH model be replaced by our NLAR models. In NLAR models, the generating process is strictly stationary and ergodic where $|\gamma| < 1$. We assume that processes for which unit roots are rejected with a standard test are stationary and ergodic. Note that NLAR(1) with serially correlated errors can be considered as a special case of MRLSTAR. We suppose the other conditions (ii) and (iii) be satisfied as in the STAR-GARCH models.

We have analyzed the series using NLAR-GARCH model. It is preferable to estimate the parameters for the conditional mean and the conditional variance at the same time(Lundbergh and Teräsvirta, 1998). However, the information matrix of a STAR-GARCH model is block-diagonal if z_t follows a symmetric distribution. Thus, the conditional mean could be estimated at the first stage by nonlinear least squares, and using the obtained residuals, the conditional variance could be estimated without loss of asymptotic efficiency(Lundbergh and Teräsvirta, 1998; Chan and McAleer, 2002).

The GARCH for residuals where $p = 1$ and $q = 1$ is well fitted using "SAS PROC AUTOREG" procedure. The maximum likelihood estimation is employed there.

$$\begin{aligned} h_t &= 0.001495 + 0.2090\epsilon_{t-1}^2 + 0.7627h_{t-1}. \\ & \quad (0.0003) \quad (0.0258) \qquad (0.0225) \end{aligned}$$

The estimates are all significant at the 5% significance level, and $\hat{\alpha}_1 + \hat{\beta}_1 = 0.9717 < 1$. The second moment condition is satisfied empirically.

z_t will follow $i.i.d.(0,1)$ theoretically. Using the estimated standard errors, we get $\hat{z}_t = \hat{\epsilon}_t \hat{h}_t^{-1/2}$. One important assumption in a GARCH model is that z_t are independent and identically distributed. So, if the model is well specified, $\hat{z}_t$ will possess the properties such as constant variance and lack of serial correlation(Franses and van Dijk, 2000). We investigate the properties of $\hat{z}_t$.

First, there seems to be no autocorrelation left for $\hat{z}_t$ and $\hat{z}_t^2$.

For independence test, we use two tests suggested in Brockwell and Davis(1991). Let $y_1, \cdots, y_n$ be a sequence of the observations. If $y_{i-1} < y_i$ and $y_i > y_{i+1}$ or $y_{i-1} > y_i$ and $y_i < y_{i+1}$, the data has a turning point at time $i, 1 < i < n$. Define T to be the number of turning points of the sequence $y_1, \cdots, y_n$. If $y_1, \cdots, y_n$ are observations of an IID sequence, then,

$$\mu_T = E(T) = 2(n-2)/3,$$

and

$$\sigma_T^2 = Var(T) = (16n - 29)/90.$$

T follows $AN(\mu_T, \sigma_T^2)$ and $|T - \mu_T|/\sigma_T$ is used as a test statistic.

Also, we count the number(S) of values of i such that $y_i > y_{i-1}, i = 2, \cdots, n$ or equivalently the number of times the $y_i - y_{i-1} > 0$. Then,

$$\mu_S = E(S) = (n-1)/2,$$

and

$$\sigma_S^2 = Var(S) = (n+1)/12.$$

S follows $AN(\mu_S, \sigma_S^2)$.

We calculate $|T - \mu_T|/\sigma_T$ and $|S - \mu_S|/\sigma_S$ for $\hat{z}_t$ and $\hat{z}_t^2$.

	T	μ_T	$\|T - \mu_T\|/\sigma_T$	S	μ_S	$\|S - \mu_S\|/\sigma_S$
$\hat{z}_t$	598	593.3	0.37	451	445.5	0.64
$\hat{z}_t^2$	596	593.3	0.21	453	445.5	0.87

Both $\hat{z}_t$ and $\hat{z}_t^2$ are shown to be random sequences, and estimated skewness and excess kurtosis for $\hat{z}_t$ are 0.0729 and 1.0815. The model seems to be fitted well.

The fitted NLAR-GARCH model explains the change of the conditional innovation variance as well as the dynamically changing second moments following the constant innovations with dynamically moving difference equation coefficients. Figure 3.24 shows the estimated conditional standard errors($\sqrt{\hat{h}_t}$).

From the examples, we can see that NLAR models using hyperbolic tangent function, of which the estimates are easily obtained through Gauss-Newton algorithm, could be one alternative to deal with the series with a rather persistent autocorrelation and nonlinearity.

Table 3.1: Estimation of α and β for model (3.1) where $f(y) = y$

(i) $\gamma = 0.5, \alpha = 2.0$, and $\beta = 0.8$, $n = 1,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.5549	0.4213	21.34	>1,000	4,809	
$\hat{\alpha}$	2.2684	1.4691	14.71	>1,000		
$\hat{\beta}$	0.9487	0.6929	15.70	>1,000		
$\hat{\alpha}^{\star}$	2.1106	0.5511	1.4475	>1,000	4,995	0.9999(4,804)†
$\hat{\beta}^{\star}$	0.8647	0.2929	1.3208	>1,000		
$\hat{\alpha}^{*}$	0.6544	0.0671	0.0540	3.07	5,000	1.0028(4,809)†
$\hat{\beta}^{*}$	0.2051	0.0373	-0.0583	3.60		1.0027(4,995)‡

(ii) $\gamma = -0.5, \alpha = 1.0$, and $\beta = 0.6$, $n = 1,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	-0.5993	0.6245	-11.84	>1,000	4,013	
$\hat{\alpha}$	1.4153	1.8418	26.17	>1,000		
$\hat{\beta}$	0.8435	1.1021	24.56	>1,000		
$\hat{\alpha}^{\star}$	1.0291	0.2406	0.8638	>1,000	5,000	0.9998(4,013)†
$\hat{\beta}^{\star}$	0.6138	0.1377	0.7712	>1,000		
$\hat{\alpha}^{*}$	-0.4097	0.0696	-0.0217	4.19	5,000	1.0007(4,013)†
$\hat{\beta}^{*}$	-0.2447	0.0407	-0.0353	1.96		1.0005(5,000)‡

(iii) $\gamma = 0.5, \alpha = -3.0$, and $\beta = 0.2$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.5012	0.1898	19.19	>1,000	2,862	
$\hat{\alpha}$	<-1,000	>1,000	-53.50	>1,000		
$\hat{\beta}$	>1,000	>1,000	53.50	>1,000		
$\hat{\alpha}^{\star}$	-3.2288	0.9860	-10.61	>1,000	4,991	$0.9999(2{,}859)^{\dagger}$
$\hat{\beta}^{\star}$	0.2471	0.3636	5.3677	>1,000		
$\hat{\alpha}^{*}$	-0.9809	0.0415	0.0056	3.53	5,000	$1.0001(2{,}862)^{\dagger}$
$\hat{\beta}^{*}$	0.0236	0.0216	0.0166	1.28		$1.0000(4{,}991)^{\ddagger}$

(iv) $\gamma = 0.99, \alpha = 1.0$, and $\beta = 0.8$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.9874	0.0060	-1.6445	>1,000	5,000	
$\hat{\alpha}$	1.0121	0.0761	0.1923	38.96		
$\hat{\beta}$	0.8125	0.0463	0.2678	62.52		
$\hat{\alpha}^{\star}$	1.0021	0.0729	0.1795	32.86	5,000	$1.0000(5{,}000)^{\dagger}$
$\hat{\beta}^{\star}$	0.8039	0.0436	0.3218	97.88		

(v) $\gamma = 0.99, \alpha = -3.0$, and $\beta = 0.1$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	1.0176	0.2574	9.0092	>1,000	3,810	
$\hat{\alpha}$	-3.6420	1.6396	-13.80	>1,000		
$\hat{\beta}$	0.2055	0.2225	18.96	>1,000		
$\hat{\alpha}^{\star}$	-3.0040	0.1059	-0.2451	54.54	5,000	$0.9999(3{,}810)^{\dagger}$
$\hat{\beta}^{\star}$	0.1010	0.0269	0.1055	12.35		

(vi) $\gamma = 0.5, \alpha = 2.0,\ \beta = 0.8,$ and $\delta = 0.95,\ n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.5430	0.2496	14.51	>1,000	4,709	
$\hat{\alpha}$	2.1892	0.8390	1.4027	>1,000		
$\hat{\beta}$	1.0015	0.5989	2.3510	>1,000		
$\hat{\delta}$	0.9492	0.0066	-0.4293	200.72		
$\hat{\alpha}^{\star}$	2.0341	0.2240	0.6465	576.42	5,000	1.0000(4,709)†
$\hat{\beta}^{\star}$	0.8316	0.2200	0.6567	585.97		
$\hat{\delta}^{\star}$	0.9492	0.0066	-0.4310	216.12		
$\hat{\alpha}^{*}$	0.7755	0.0446	0.0065	0.84	5,000	1.0006(4,709)†
$\hat{\beta}^{*}$	0.1899	0.0392	0.0227	2.49		1.0006(5,000)‡
$\hat{\delta}^{*}$	0.9493	0.0066	-0.4258	209.81		

(vii) $\gamma = 0.5, \alpha = -3.0,\ \beta = 0.2,$ and $\delta = 0.95,\ n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.4798	0.1326	23.11	>1,000	2,420	
$\hat{\alpha}$	<1,000	>1,000	-49.19	>1,000		
$\hat{\beta}$	>1,000	>1,000	49.19	>1,000		
$\hat{\delta}$	0.9495	0.0060	-0.4028	75.33		
$\hat{\alpha}^{\star}$	-5.5231	133.20	-66.82	>1,000	4,993	0.9998(2,420)†
$\hat{\beta}^{\star}$	3.0704	165.77	66.56	>1,000		
$\hat{\delta}^{\star}$	0.9494	0.0060	-0.3779	124.87		
$\hat{\alpha}^{*}$	-0.9853	0.0431	-0.0155	3.29	5,000	0.9999(2,420)†
$\hat{\beta}^{*}$	0.0234	0.0960	-0.0485	7.27		1.0000(4,993)‡
$\hat{\delta}^{*}$	0.9493	0.0060	-0.3838	130.58		

Note: * and * indicate the estimation of α and β with γ fixed at true value and with γ fixed at 0.99. † and † show MSE ratio over the estimation of γ unknown and of γ fixed at true value respectively.

Table 3.2: Distribution of $\hat{\alpha}$ and $\hat{\beta}$ for model (3.1) where $f(y) = y$

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
-0.9	$\hat{\alpha}$	bias	0.0134	0.0051	0.0040
		STD	0.1736	0.1069	0.0683
		skewness	1.9798	0.3378	0.3067
		JB	>1,000	168.21	92.26
		$Pr(\|t\| > z_{0.025})$	0.0424	0.0446	0.0488
	$\hat{\beta}$	bias	-0.0195	-0.0098	-0.0054
		STD	0.1150	0.0666	0.0537
		skewness	-2.6314	-0.5830	-0.4404
		JB	>1,000	422.71	208.81
		$Pr(\|t\| > z_{0.025})$	0.0476	0.0454	0.0482
		R	4,998	5,000	5,000
-0.6	$\hat{\alpha}$	bias	-0.0012	0.0003	-0.0016
		STD	0.0965	0.0673	0.0545
		skewness	0.0843	0.0772	0.0504
		JB	8.28	8.31	2.15
		$Pr(\|t\| > z_{0.025})$	0.0496	0.0482	0.0444
	$\hat{\beta}$	bias	-0.0019	-0.0008	-0.0004
		STD	0.0523	0.0373	0.0294
		skewness	-0.2592	-0.1462	-0.0814
		JB	65.42	18.13	6.40
		$Pr(\|t\| > z_{0.025})$	0.0490	0.0480	0.0508
		R	5,000	5,000	5,000

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
-0.3	$\hat{\alpha}$	bias	-0.0044	-0.0030	-0.0006
		STD	0.0757	0.0542	0.0443
		skewness	0.0596	-0.0055	-0.0177
		JB	2.97	0.50	1.70
		$Pr(\lvert t\rvert > z_{0.025})$	0.0482	0.0486	0.0480
	$\hat{\beta}$	bias	0.0004	0.0010	0.0004
		STD	0.0391	0.0276	0.0228
		skewness	0.0262	-0.0250	0.0694
		JB	0.58	2.52	4.40
		$Pr(\lvert t\rvert > z_{0.025})$	0.0518	0.0464	0.0428
		R	5,000	5,000	5,000
0.0	$\hat{\alpha}$	bias	-0.0046	-0.0023	-0.0023
		STD	0.0729	0.0514	0.0420
		skewness	0.0516	-0.0827	0.0547
		JB	2.31	5.72	2.66
		$Pr(\lvert t\rvert > z_{0.025})$	0.0488	0.0512	0.0462
	$\hat{\beta}$	bias	-0.0002	0.0000	-0.0003
		STD	0.0381	0.0262	0.0213
		skewness	0.0204	0.0174	0.0755
		JB	1.83	1.73	4.90
		$Pr(\lvert t\rvert > z_{0.025})$	0.0522	0.0530	0.0502
		R	5,000	5,000	5,000

β			$n=1,000$	$n=2,000$	$n=3,000$
0.3	$\hat{\alpha}$	bias	-0.0052	-0.0016	-0.0020
		STD	0.0773	0.0545	0.0443
		skewness	0.0461	-0.0408	0.0512
		JB	4.28	3.67	2.96
		$Pr(\|t\| > z_{0.025})$	0.0514	0.0498	0.0550
	$\hat{\beta}$	bias	-0.0024	-0.0001	0.0004
		STD	0.0397	0.0276	0.0224
		skewness	-0.0264	-0.0186	-0.0442
		JB	1.39	0.84	1.74
		$Pr(\|t\| > z_{0.025})$	0.0544	0.0492	0.0524
		R	5,000	5,000	5,000
0.6	$\hat{\alpha}$	bias	-0.0008	0.0007	-0.0005
		STD	0.0954	0.0681	0.0547
		skewness	0.1878	0.0453	0.0693
		JB	38.57	2.38	4.51
		$Pr(\|t\| > z_{0.025})$	0.0446	0.0546	0.0544
	$\hat{\beta}$	bias	0.0016	0.0002	0.0004
		STD	0.0522	0.0366	0.0293
		skewness	0.2445	0.1667	0.0844
		JB	73.55	25.68	6.26
		$Pr(\|t\| > z_{0.025})$	0.0504	0.0570	0.0486
		R	5,000	5,000	5,000

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
0.9	$\hat{\alpha}$	bias	0.0149	0.0048	0.0056
		STD	0.1659	0.1083	0.0873
		skewness	1.6183	0.3888	0.2989
		JB	>1,000	168.53	100.93
		$Pr(\|t\| > z_{0.025})$	0.0482	0.0458	0.0548
	$\hat{\beta}$	bias	0.0165	0.0074	0.0051
		STD	0.1078	0.0667	0.0543
		skewness	2.4445	0.6480	0.4511
		JB	>1,000	601.59	209.04
		$Pr(\|t\| > z_{0.025})$	0.0540	0.0440	0.0468
		R	5,000	5,000	5,000

Table 3.3: Comparison of ARMA and nonlinear estimation for model (3.1) where $f(y) = y$

(i) data generated by $\gamma = 0.5$

	Q(6)	ARMA(μ:p:q)	MSE	NLIN($\alpha : \beta$)	MSE
$\alpha = 1$ $\beta = 0.3$	$p = 0.79$	0.11:1.07 -0.14 -0.06:0.87 (4.83:12.75 -3.71 -2.56:10.64)	0.974	0.41:0.15 (0.05:0.03)	0.971
$\alpha = 1$ $\beta = -0.3$	$p = 0.28$	-0.08:-2.74:- (0.21:9.80:-)	0.972	0.43:-0.10 (0.05:0.03)	0.969
$\alpha = 1$ $\beta = 0.6$	$p = 0.88$	0.15:0.22:- (5.23:10.10:-)	1.033	0.42:0.20 (0.05:0.03)	1.017
$\alpha = 1$ $\beta = 0.9$	$p = 0.06$	0.25:0.26:- (8.39:12.11:-)	0.989	0.47:0.28 (0.05:0.03)	0.970
$\alpha = 1$ $\beta = -0.9$	$p = 0.22$	-0.20:0.21:- (-6.83:9.42:-)	1.058	0.32:-0.34 (0.05:0.03)	0.999
$\alpha = 1$ $\beta = 1.2$	$p = 0.30$	0.27:0.21:- (9.13:9.80:-)	1.053	0.25:0.40 (0.05:0.03)	0.989
$\alpha = 1$ $\beta = 2.0$	$p = 0.34$	0.35:0.20:- (12.03:9.09:-)	1.078	0.12:0.50 (0.06:0.03)	0.996
$\alpha = 1$ $\beta = -2.0$	$p = 0.99$	-0.35:0.19 0.05 0.05:- (-10.54:8.54 2.23 2.27:-)	1.072	0.15:-0.52 (0.05:0.03)	0.977
$\alpha = 3.0$ $\beta = 0.1$	$p = 0.86$	0.01:0.45 0.05:- (0.29:20.16 2.06:-)	0.966	1.04:-0.02 (0.05:0.03)	0.968
$\alpha = -3.0$ $\beta = 0.1$	$p = 0.24$	0.00:-0.46:- (0.30:-23.15:-)	0.969	-1.01:0.00 (0.05:0.03)	0.969

(ii) data generated by $\gamma = 0.99$

	Q(6)	ARMA(μ:p:q)	MSE	NLIN($\alpha : \beta$)	MSE
$\alpha = 1$ $\beta = 0.3$	$p = 0.82$	0.33:1.31 -0.34 -0.08:0.83 (8.60:12.23 -5.04 -3.27:7.83)	1.003	0.95:0.31 (0.05:0.03)	0.972
$\alpha = 1$ $\beta = -0.3$	$p = 0.18$	-0.28:0.50:- (-6.32:25.51:-)	0.998	1.00:-0.27 (0.05:0.03)	0.969
$\alpha = 1$ $\beta = 0.6$	$p = 0.51$	0.96:1.40 -0.43:0.70 (3.44:21.05 -7.03:12.37)	1.227	1.02:0.58 (0.07:0.03)	1.015
$\alpha = 1$ $\beta = 0.9$	$p = 0.01$	2.20:1.47 -0.49:0.60 (4.52:9.81 -3.40:4.32)	1.118	1.22:0.98 (0.12:0.08)	0.970
$\alpha = 1$ $\beta = -0.9$	$p = 0.41$	-1.89:1.57 -0.51 -0.07:0.80 (-5.64:12.08 -5.20 -2.01:6.30)	1.201	1.09:-0.90 (0.10:0.06)	0.996
$\alpha = 1$ $\beta = 1.2$	$p = 0.14$	2.55:1.48 -0.49:0.61 (4.14:10.76 -3.70:4.80)	1.157	0.98:1.24 (0.16:0.12)	0.986
$\alpha = 1$ $\beta = 2.0$	$p = 0.29$	2.91:1.59 -0.60:0.68 (4.43:9.61 -3.73:4.42)	1.105	1.08:2.29 (0.42:0.47)	0.986
$\alpha = 1$ $\beta = -2.0$	$p = 0.28$	-2.77:0.88 0.09:- (-3.76:39.53 4.13:-)	1.105	0.93:-1.95 (0.33:0.33)	0.969
$\alpha = 3.0$ $\beta = 0.1$	$p = 0.32$	0.46:0.91:- (1.97:96.45:-)	0.973	3.04:0.11 (0.12:0.03)	0.969
$\alpha = -3.0$ $\beta = 0.1$	$p = 0.30$	0.02:-0.89:- (2.04:-89.05:-)	0.968	-3.00:0.07 (0.12:0.03)	0.968

Note: The figures in parentheses indicate t values in ARMA and standard errors in NLIN for estimates. Q(6) figures show BOX-Ljung χ^2 statistic at lag 6.

Table 3.4: Estimation of α and β for model (3.1) where $f(y) = |y|$

(i) $\gamma = 0.5, \alpha = 2.0$, and $\beta = 0.8$, $n = 1,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.5376	0.1651	5.5423	>1,000	2,817	
$\hat{\alpha}$	5.4372	49.40	10.98	>1,000		
$\hat{\beta}$	2.8633	16.80	-3.14	>1,000		
$\hat{\alpha}^{\star}$	5.3901	40.33	13.59	>1,000	4,766	0.9996(2,767)†
$\hat{\beta}^{\star}$	0.6139	11.59	-9.4138	>1,000		
$\hat{\alpha}^{*}$	0.9143	0.1924	0.0549	2.78	5,000	1.0002(2,817)†
$\hat{\beta}^{*}$	0.0568	0.0995	0.0090	0.15		1.0002(4,766)‡

(ii) $\gamma = -0.5, \alpha = 1.0$, and $\beta = 0.6$, $n = 1,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	-0.5136	0.3104	-8.6621	>1,000	3,031	
$\hat{\alpha}$	1.0088	21.07	18.90	>1,000		
$\hat{\beta}$	2.7353	11.20	5.6212	>1,000		
$\hat{\alpha}^{\star}$	1.0113	0.7424	-0.0771	>1,000	4,995	0.9997(3,029)†
$\hat{\beta}^{\star}$	0.6455	0.6446	11.32	>1,000		
$\hat{\alpha}^{*}$	-0.5700	0.1778	-0.0107	0.18	5,000	1.0002(3,031)†
$\hat{\beta}^{*}$	-0.1363	0.0968	-0.0285	0.86		1.0001(4,995)‡

(iii) $\gamma = 0.5, \alpha = -3.0$, and $\beta = 0.1$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.4768	0.1929	21.55	>1,000	2,620	
$\hat{\alpha}$	<-1,000	>1,000	-51.19	>1,000		
$\hat{\beta}$	<-1,000	>1,000	-51.19	>1,000		
$\hat{\alpha}^{\star}$	-3.0705	1.3021	-31.55	>1,000	4,993	$0.9999(2{,}620)^{\dagger}$
$\hat{\beta}^{\star}$	0.1966	0.5003	10.38	>1,000		
$\hat{\alpha}^{*}$	-0.9979	0.1082	-0.0947	7.47	5,000	$1.0000(2{,}620)^{\dagger}$
$\hat{\beta}^{*}$	0.0337	0.0559	0.0115	0.11		$1.0000(4{,}993)^{\ddagger}$

(iv) $\gamma = 0.99, \alpha = 1.0$, and $\beta = 0.8$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.9880	0.0054	-1.7255	>1,000	5,000	
$\hat{\alpha}$	0.9782	0.2321	-0.0799	15.62		
$\hat{\beta}$	0.8308	0.1352	0.5845	433.34		
$\hat{\alpha}^{\star}$	1.0014	0.2291	-0.0480	12.59	5,000	$1.0000(5{,}000)^{\dagger}$
$\hat{\beta}^{\star}$	0.8065	0.1236	0.5149	333.50		

(v) $\gamma = 0.99, \alpha = -3.0$, and $\beta = 0.1$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.9285	0.2070	15.22	>1,000	3,015	
$\hat{\alpha}$	<-1,000	>1,000	-54.91	>1,000		
$\hat{\beta}$	94.47	>1,000	46.40	>1,000		
$\hat{\alpha}^{\star}$	-3.0251	0.2206	-0.2344	50.31	5,000	$1.0000(3{,}015)^{\dagger}$
$\hat{\beta}^{\star}$	0.1085	0.0612	0.1739	26.65		

(vi) $\gamma = 0.5, \alpha = 2.0$, $\beta = 0.8$, and $\delta = 0.95$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.5223	0.2626	16.50	>1,000	3,601	
$\hat{\alpha}$	2.5706	15.53	25.12	>1,000		
$\hat{\beta}$	2.6747	13.27	3.7943	>1,000		
$\hat{\delta}$	0.9493	0.0066	-0.4061	118.91		
$\hat{\alpha}^{\star}$	2.0371	0.5104	9.7250	>1,000	4,999	0.9999(3,600)†
$\hat{\beta}^{\star}$	0.8503	0.6938	2.9257	>1,000		
$\hat{\delta}^{\star}$	0.9493	0.0065	-0.4232	173.56		
$\hat{\alpha}^{*}$	0.8445	0.0719	0.0880	7.22	5,000	1.0002(3,601)†
$\hat{\beta}^{*}$	0.1119	0.0626	-0.0996	21.18		1.0001(4,999)‡
$\hat{\delta}^{*}$	0.9494	0.0065	-0.4042	157.64		

(vii) $\gamma = 0.5, \alpha = -3.0$, $\beta = 0.2$, and $\delta = 0.95$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.4706	0.0875	10.75	>1,000	2,076	
$\hat{\alpha}$	<1,000	>1,000	-45.56	>1,000		
$\hat{\beta}$	>1,000	>1,000	45.56	>1,000		
$\hat{\delta}$	0.9493	0.0060	-0.3074	32.65		
$\hat{\alpha}^{\star}$	-3.0860	0.9608	-5.4098	>1,000	4,995	0.9998(2,076)†
$\hat{\beta}^{\star}$	0.1362	2.4901	-9.6762	>1,000		
$\hat{\delta}^{\star}$	0.9493	0.0059	-0.3323	97.47		
$\hat{\alpha}^{*}$	-0.9856	0.0720	-0.0040	0.05	5,000	0.9999(2,076)†
$\hat{\beta}^{*}$	0.0231	0.1630	-0.0020	9.75		1.0000(4,995)‡
$\hat{\delta}^{*}$	0.9493	0.0059	-0.3320	97.07		

Note: * and * indicate the estimation of α and β with γ fixed at true value and with γ fixed at 0.99. † and † show MSE ratio over the estimation of γ unknown and of γ fixed at true value respectively.

Table 3.5: Distribution of $\hat{\alpha}$ and $\hat{\beta}$ for model (3.1) where $f(y) = |y|$

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
-0.9	$\hat{\alpha}$	bias	0.0014	0.0008	-0.0006
		STD	0.1846	0.1281	0.1039
		skewness	0.0190	-0.0607	0.0449
		JB	0.43	3.13	2.66
		$Pr(\|t\| > z_{0.025})$	0.0530	0.0520	0.0468
	$\hat{\beta}$	bias	0.0005	0.0007	0.0004
		STD	0.1122	0.0785	0.0634
		skewness	0.0012	0.0715	-0.0399
		JB	0.04	5.14	5.44
		$Pr(\|t\| > z_{0.025})$	0.0518	0.0496	0.0448
		R	5,000	5,000	5,000
-0.6	$\hat{\alpha}$	bias	-0.0033	-0.0005	0.0017
		STD	0.1777	0.1244	0.1005
		skewness	0.0135	0.0026	0.0378
		JB	0.27	0.48	1.22
		$Pr(\|t\| > z_{0.025})$	0.0500	0.0480	0.0392
	$\hat{\beta}$	bias	0.0022	0.0004	-0.0006
		STD	0.1013	0.0719	0.0579
		skewness	-0.0079	0.0462	-0.0020
		JB	0.08	1.78	0.13
		$Pr(\|t\| > z_{0.025})$	0.0434	0.0516	0.0424
		R	5,000	5,000	5,000

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
-0.3	$\hat{\alpha}$	bias	0.0044	0.0021	0.0028
		STD	0.1799	0.1267	0.1011
		skewness	0.1046	0.0638	0.0810
		JB	9.12	3.50	5.60
		$Pr(\|t\| > z_{0.025})$	0.0542	0.0508	0.0456
	$\hat{\beta}$	bias	-0.0018	-0.0012	-0.0015
		STD	0.0972	0.0684	0.0545
		skewness	-0.0711	0.0215	-0.0095
		JB	8.34	2.07	0.14
		$Pr(\|t\| > z_{0.025})$	0.0500	0.0532	0.0480
		R	5,000	5,000	5,000
0.0	$\hat{\alpha}$	bias	0.0112	0.0047	0.0041
		STD	0.1858	0.1354	0.1086
		skewness	0.0613	0.0880	0.0433
		JB	3.14	10.19	2.62
		$Pr(\|t\| > z_{0.025})$	0.0466	0.0562	0.0496
	$\hat{\beta}$	bias	-0.0066	-0.0033	-0.0019
		STD	0.0966	0.0693	0.0554
		skewness	-0.0131	-0.0476	-0.0142
		JB	0.20	5.44	0.17
		$Pr(\|t\| > z_{0.025})$	0.0494	0.0556	0.0464
		R	5,000	5,000	5,000

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
0.3	$\hat{\alpha}$	bias	0.0244	0.0118	0.0096
		STD	0.2081	0.1466	0.1191
		skewness	0.1104	0.0783	0.1390
		JB	14.09	6.05	16.41
		$Pr(\lvert t\rvert > z_{0.025})$	0.0532	0.0528	0.0532
	$\hat{\beta}$	bias	-0.0136	-0.0065	-0.0056
		STD	0.0985	0.0703	0.0569
		skewness	-0.0770	-0.0549	-0.1057
		JB	8.51	3.15	9.31
		$Pr(\lvert t\rvert > z_{0.025})$	0.0484	0.0546	0.0520
		R	5,000	5,000	5,000
0.6	$\hat{\alpha}$	bias	0.0197	0.0107	0.0070
		STD	0.2803	0.1900	0.1545
		skewness	0.0042	-0.0606	-0.0164
		JB	22.89	3.12	2.00
		$Pr(\lvert t\rvert > z_{0.025})$	0.0581	0.0538	0.0536
	$\hat{\beta}$	bias	-0.0049	-0.0036	-0.0022
		STD	0.1383	0.0937	0.0751
		skewness	0.4156	0.2893	0.1861
		JB	300.68	121.12	32.14
		$Pr(\lvert t\rvert > z_{0.025})$	0.0715	0.0670	0.0606
		R	4,992	5,000	5,000

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
0.9	$\hat{\alpha}$	bias	-0.0226	-0.0057	0.0027
		STD	0.5256	0.3477	0.2696
		skewness	-0.0114	-0.0493	-0.0557
		JB	>1,000	189.31	4.24
		$Pr(\lvert t\rvert > z_{0.025})$	0.0647	0.0575	0.0524
	$\hat{\beta}$	bias	0.0486	0.0219	0.0093
		STD	0.3616	0.2089	0.1572
		skewness	4.1322	1.1085	0.6294
		JB	>1,000	>1,000	500.67
		$Pr(\lvert t\rvert > z_{0.025})$	0.0792	0.0647	0.0570
		R	4,824	4,975	4,997

Table 3.6: Comparison of ARMA and nonlinear estimation for model (3.1) where $f(y) = |y|$

(i) data generated by $\gamma = 0.5$

	Q(6)	ARMA(μ:p:q)	MSE	NLIN($\alpha : \beta$)	MSE
$\alpha = 1$ $\beta = 0.3$	$p = 0.99$	-:0.29:- (-:13.50:-)	0.948	0.44:0.10 (0.12:0.07)	0.948
$\alpha = 1$ $\beta = -0.3$	$p = 0.99$	-:0.11:- (-:4.74:-)	0.949	0.40:-0.12 (0.12:0.07)	0.948
$\alpha = 1$ $\beta = 0.6$	$p = 0.87$	-:0.37:- (-:17.74:-)	1.004	0.52:0.15 (0.12:0.06)	1.002
$\alpha = 1$ $\beta = 0.9$	$p = 0.95$	-:0.41:- (-:20.13:-)	1.015	0.99:-0.06 (0.13:0.06)	1.015
$\alpha = 1$ $\beta = -0.9$	$p = 0.16$	-:-:0.10 (-:-:4.70)	1.042	0.45:-0.41 (0.12:0.07)	1.025
$\alpha = 1$ $\beta = 1.2$	$p = 0.72$	-:0.45:- (-:22.75:-)	1.069	0.74:0.13 (0.12:0.06)	1.067
$\alpha = 1$ $\beta = 2.0$	$p = 0.51$	-:0.46:- (-:23.42:-)	0.988	0.91:0.06 (0.13:0.07)	0.986
$\alpha = 1$ $\beta = -2.0$	$p = 0.42$	-:-0.36:- (-:-17.02:-)	1.009	0.08:-0.50 (0.13:0.07)	0.987
$\alpha = 3.0$ $\beta = 0.1$	$p = 0.48$	-:0.49:- (-:25.06:-)	0.924	1.15:-0.04 (0.14:0.08)	0.924
$\alpha = -3.0$ $\beta = 0.1$	$p = 0.81$	-:-0.48:- (-:-24.33:-)	1.053	-1.09:0.02 (0.14:0.07)	1.052

(ii) data generated by $\gamma = 0.99$

	Q(6)	ARMA(μ:p:q)	MSE	NLIN($\alpha : \beta$)	MSE
$\alpha = 1$ $\beta = 0.3$	$p = 0.78$	-:0.63:- (-:36.51:-)	1.056	1.16:0.17 (0.15:0.07)	1.052
$\alpha = 1$ $\beta = -0.3$	$p = 0.54$	-:0.27:- (-:12.41:-)	0.949	0.97:-0.25 (0.12:0.06)	0.942
$\alpha = 1$ $\beta = 0.6$	$p = 0.19$	-:0.88:- (-:82.77:-)	1.074	0.87:0.65 (0.18:0.09)	1.033
$\alpha = 1$ $\beta = 0.9$	$p = 0.63$	-:0.98:- (-:230.96:-)	1.085	0.56:1.05 (0.28:0.18)	0.987
$\alpha = 1$ $\beta = -0.9$	$p = 0.06$	-:-0.25:- (-:-11.51:-)	1.063	0.84:-0.80 (0.12:0.08)	0.983
$\alpha = 1$ $\beta = 1.2$	$p = 0.15$	-:0.99:- (-:263.69:-)	1.025	0.15:1.76 (0.54:0.48)	1.014
$\alpha = 1$ $\beta = 2.0$	$p = 0.26$	-:0.99:- (-:287.81:-)	0.991	2.05:0.94 (0.96:0.56)	0.989
$\alpha = 1$ $\beta = -2.0$	$p = 0.04$	-:-1.47 -0.47:-0.56 (-:-6.69 -2.22:-2.65)	1.134	0.88:-1.89 (0.29:0.25)	1.030
$\alpha = 3.0$ $\beta = 0.1$	$p = 0.72$	-:0.95:- (-:133.53:-)	0.986	3.14:0.14 (0.36:0.08)	0.984
$\alpha = -3.0$ $\beta = 0.1$	$p = 0.70$	-:-0.86:- (-:-74.41:-)	1.032	-2.92:0.09 (0.26:0.07)	1.029

Note: The figures in parentheses indicate t values in ARMA and standard errors in NLIN for estimates. Q(6) figures show BOX-Ljung χ^2 statistic at lag 6.

Table 3.7: The distribution of $\hat{\alpha}$ and $\hat{\beta}$ where $\beta = 0$ $n = 1,000$

α	parms	bias	STD	JB	$Pr(\|t\|)$	R
-3.6	α	-0.0054	0.2125(0.1965)	>1,000	0.0428	5,000
$[-0.9486]$	β	0.0006	0.0436(0.0365)	>1,000	0.0372	
-2.4	α	-0.0012	0.1154(0.1145)	2.43	0.0472	5,000
$[-0.8337]$	β	0.0001	0.0377(0.0365)	7.61	0.0460	
-1.2	α	-0.0035	0.0758(0.0750)	2.33	0.0524	5,000
$[-0.5371]$	β	-0.0006	0.0370(0.0365)	2.78	0.0482	
0.0	α	-0.0019	0.0636(0.0633)	0.05	0.0490	5,000
[0.0]	β	-0.0006	0.0367(0.0365)	1.23	0.0466	
1.2	α	-0.0080	0.0748(0.0750)	1.03	0.0516	5,000
[0.5371]	β	0.0004	0.0368(0.0365)	0.41	0.0450	
2.4	α	-0.0164	0.1139(0.1145)	2.79	0.0580	5,000
[0.8337]	β	0.0007	0.0377(0.0365)	0.66	0.0490	
3.6	α	-0.0600	0.2032(0.1965)	228.33	0.0848	5,000
[0.9486]	β	-0.0002	0.0414(0.0365)	116.26	0.0574	

Note: $Pr(|t)$ indicates $Pr(|t| > z_{0.025})$. The figures in parentheses [ ] show $\rho_0 = \frac{\exp(\alpha_0)-1}{\exp(\alpha_0)+1}$ and those in parenthesis ( ) are the theoetical standard error of parameters in $B^{-1}(\theta_0)\sigma^2$.

Table 3.8: Distribution of $\hat{\alpha}$ and $\hat{\beta}$ for model (3.2) where $f(y) = y$

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
-0.9	$\hat{\alpha}$	bias	-0.0138	-0.0081	-0.0026
		STD	0.1616	0.1148	0.1011
		skewness	2.0584	0.3771	0.5757
		JB	>1,000	90.70	88.28
		$Pr(\|t\| > z_{0.025})$	0.0430	0.0454	0.0550
	$\hat{\beta}$	bias	0.0082	0.0068	0.0063
		STD	0.1026	0.0727	0.0622
		skewness	-2.1514	-0.6653	-0.5103
		JB	>1,000	164.22	49.54
		$Pr(\|t\| > z_{0.025})$	0.0705	0.0717	0.0559
		R	1,418	1,256	1,091
-0.6	$\hat{\alpha}$	bias	-0.0087	-0.0046	-0.0032
		STD	0.0944	0.0696	0.0567
		skewness	0.1340	0.1374	0.1422
		JB	15.81	29.06	12.67
		$Pr(\|t\| > z_{0.025})$	0.0493	0.0536	0.0465
	$\hat{\beta}$	bias	0.0040	0.0022	0.0015
		STD	0.0514	0.0369	0.0298
		skewness	-0.2701	-0.2479	-0.2059
		JB	51.44	54.41	29.99
		$Pr(\|t\| > z_{0.025})$	0.0536	0.0531	0.0493
		R	4,014	3,843	3,634

β			$n=1,000$	$n=2,000$	$n=3,000$
-0.3	$\hat{\alpha}$	bias	-0.0047	-0.0023	-0.0011
		STD	0.0757	0.0540	0.0441
		skewness	0.0351	0.0498	0.0066
		JB	1.90	3.90	2.11
		$Pr(\|t\| > z_{0.025})$	0.0528	0.0534	0.0530
	$\hat{\beta}$	bias	0.0006	0.0001	0.0000
		STD	0.0381	0.0271	0.0221
		skewness	0.0166	-0.0039	0.0566
		JB	0.49	0.67	2.81
		$Pr(\|t\| > z_{0.025})$	0.0478	0.0498	0.0492
		R	5,000	5,000	5,000
0.0	$\hat{\alpha}$	bias	-0.0048	-0.0033	-0.0008
		STD	0.0712	0.0503	0.0414
		skewness	0.0098	-0.0274	-0.0051
		JB	0.55	0.63	0.02
		$Pr(\|t\| > z_{0.025})$	0.0472	0.0490	0.0484
	$\hat{\beta}$	bias	0.0002	0.0000	-0.0001
		STD	0.0366	0.0262	0.0214
		skewness	0.0288	-0.0469	0.0343
		JB	1.27	3.14	1.08
		$Pr(\|t\| > z_{0.025})$	0.0484	0.0500	0.0530
		R	5,000	5,000	5,000

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
0.3	$\hat{\alpha}$	bias	-0.0046	-0.0024	-0.0021
		STD	0.0749	0.0539	0.0438
		skewness	0.0100	0.0292	0.0251
		JB	1.46	1.23	2.35
		$Pr(\|t\| > z_{0.025})$	0.0462	0.0540	0.0506
	$\hat{\beta}$	bias	-0.0008	-0.0008	-0.0004
		STD	0.0390	0.0275	0.0218
		skewness	-0.0392	-0.0323	-0.0842
		JB	4.09	2.01	7.72
		$Pr(\|t\| > z_{0.025})$	0.0516	0.0524	0.0460
		R	5,000	5,000	5,000
0.6	$\hat{\alpha}$	bias	-0.0066	-0.0052	-0.0012
		STD	0.0984	0.0686	0.0571
		skewness	0.1087	0.0978	0.1091
		JB	26.39	6.22	9.86
		$Pr(\|t\| > z_{0.025})$	0.0563	0.0514	0.0490
	$\hat{\beta}$	bias	-0.0044	-0.0032	-0.0015
		STD	0.0519	0.0369	0.0303
		skewness	0.2709	0.2592	0.2997
		JB	64.37	45.89	61.18
		$Pr(\|t\| > z_{0.025})$	0.0548	0.0607	0.0490
		R	4,012	3,836	3,656

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
0.9	$\hat{\alpha}$	bias	-0.0195	-0.0123	-0.0119
		STD	0.1556	0.1257	0.0985
		skewness	0.5167	0.6795	0.3754
		JB	108.29	255.75	47.98
		$Pr(\|t\| > z_{0.025})$	0.0498	0.0520	0.0464
	$\hat{\beta}$	bias	-0.0109	-0.0028	-0.0062
		STD	0.0968	0.0761	0.0623
		skewness	0.8337	0.6915	0.6561
		JB	285.51	123.32	116.46
		$Pr(\|t\| > z_{0.025})$	0.0671	0.0594	0.0653
		R	1,505	1,230	1,057

Table 3.9: Distribution of $\hat{\alpha}$ and $\hat{\beta}$ for model (3.2) where $f(y) = |y|$

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
-0.9	$\hat{\alpha}$	bias	-0.0085	-0.0022	-0.0008
		STD	0.1876	0.1290	0.1064
		skewness	0.4426	0.5380	0.3850
		JB	>1,000	>1,000	>1,000
		$Pr(\|t\| > z_{0.025})$	0.0531	0.0452	0.0464
	$\hat{\beta}$	bias	0.0068	0.0018	0.0003
		STD	0.1155	0.0779	0.0656
		skewness	-0.7707	-0.2834	-0.4693
		JB	>1,000	>1,000	>1,000
		$Pr(\|t\| > z_{0.025})$	0.0549	0.0461	0.0465
		R	4,994	4,995	4,995
-0.6	$\hat{\alpha}$	bias	-0.0017	-0.0007	-0.0024
		STD	0.1751	0.1249	0.0999
		skewness	0.0720	0.0211	-0.0048
		JB	5.69	5.54	1.60
		$Pr(\|t\| > z_{0.025})$	0.0522	0.0516	0.0466
	$\hat{\beta}$	bias	0.0014	0.0002	0.0015
		STD	0.1015	0.0720	0.0568
		skewness	-0.0425	0.0173	0.0210
		JB	5.56	5.88	0.40
		$Pr(\|t\| > z_{0.025})$	0.0506	0.0520	0.0428
		R	5,000	5,000	5,000

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
-0.3	$\hat{\alpha}$	bias	0.0039	-0.0019	-0.0002
		STD	0.1750	0.1247	0.1023
		skewness	0.0707	-0.0103	-0.0150
		JB	4.28	0.33	1.30
		$Pr(\|t\| > z_{0.025})$	0.0468	0.0498	0.0542
	$\hat{\beta}$	bias	-0.0019	0.0017	0.0001
		STD	0.0955	0.0679	0.0549
		skewness	-0.0359	0.0178	0.0300
		JB	1.12	0.39	3.17
		$Pr(\|t\| > z_{0.025})$	0.0486	0.0504	0.0544
		R	5,000	5,000	5,000
0.0	$\hat{\alpha}$	bias	0.0082	0.0053	0.0029
		STD	0.1846	0.1324	0.1059
		skewness	0.0880	0.0858	0.0118
		JB	6.46	6.94	2.71
		$Pr(\|t\| > z_{0.025})$	0.0514	0.0506	0.0490
	$\hat{\beta}$	bias	-0.0046	-0.0034	-0.0016
		STD	0.0951	0.0670	0.0547
		skewness	-0.0286	-0.0280	0.0077
		JB	0.74	0.65	2.99
		$Pr(\|t\| > z_{0.025})$	0.0512	0.0468	0.0482
		R	5,000	5,000	5,000

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
0.3	$\hat{\alpha}$	bias	0.0234	0.0128	0.0073
		STD	0.2026	0.1420	0.1163
		skewness	0.1780	0.0444	0.1026
		JB	29.77	7.57	17.82
		$Pr(\lvert t\rvert > z_{0.025})$	0.0528	0.0488	0.0542
	$\hat{\beta}$	bias	-0.0136	-0.0083	-0.0042
		STD	0.0965	0.0675	0.0560
		skewness	-0.1593	-0.0622	-0.0757
		JB	22.94	4.37	13.94
		$Pr(\lvert t\rvert > z_{0.025})$	0.0502	0.0496	0.0544
		R	4,999	5,000	5,000
0.6	$\hat{\alpha}$	bias	0.0490	0.0221	0.0231
		STD	0.2850	0.1932	0.1585
		skewness	0.0296	-0.0882	-0.0431
		JB	4.94	5.03	1.08
		$Pr(\lvert t\rvert > z_{0.025})$	0.0714	0.0582	0.0564
	$\hat{\beta}$	bias	-0.0277	-0.0141	-0.0125
		STD	0.1317	0.0894	0.0732
		skewness	0.4797	0.3710	0.3530
		JB	230.50	120.53	84.31
		$Pr(\lvert t\rvert > z_{0.025})$	0.0912	0.0705	0.0654
		R	3,628	3,349	3,229

β			$n = 1,000$	$n = 2,000$	$n = 3,000$
0.9	$\hat{\alpha}$	bias	0.1888	0.0774	0.0497
		STD	0.6844	0.4031	0.3332
		skewness	1.5599	0.0463	-0.0049
		JB	>1,000	33.88	0.05
		$Pr(\lvert t\rvert > z_{0.025})$	0.1333	0.0791	0.0654
	$\hat{\beta}$	bias	-0.0931	-0.0494	-0.0267
		STD	0.3989	0.2300	0.1871
		skewness	6.7579	1.2877	0.8720
		JB	>1,000	588.82	185.24
		$Pr(\lvert t\rvert > z_{0.025})$	0.2339	0.1500	0.1090
		R	1,043	860	780

Table 3.10: The mean of model (3.1) with $\gamma = 0.99$, 0.999 and 1.00, $\alpha = 1.0$, and $\beta = 0.8$

(i) where $f(y) = y$

	$\gamma =$	0.99		$\gamma =$	0.999		$\gamma =$	1.00	
n	mean	STD	R	mean	STD	R	mean	STD	R
1,000	2.4756	1.0549	4,932	3.1134	1.3624	2,588	3.1478	1.4008	2,102
2,000	2.5590	0.8642	5,000	4.6198	1.8771	2,632	4.6271	1.9063	1,792
3,000	2.5290	0.6777	5,000	5.9081	2.2752	2,898	5.8740	2.2828	1,576
4,000	2.5438	0.5933	5,000	6.9463	2.5521	3,243	6.9624	2.6216	1,478
5,000	2.5491	0.5293	5,000	7.8862	2.8552	3,501	7.8079	2.7783	1,362
6,000	2.5762	0.4958	5,000	8.5895	3.0312	3,819	8.8218	3.1295	1,348
7,000	2.5544	0.4438	5,000	9.2507	3.2298	4,118	9.7184	3.2991	1,242
8,000	2.5523	0.4253	5,000	9.8554	3.3606	4,386	10.4749	3.5286	1,199
9,000	2.5598	0.3957	5,000	10.1754	3.4823	4,553	11.6832	3.5482	1,148
10,000	2.5543	0.3847	5,000	10.5641	3.5849	4,732	12.0972	3.9427	1,054
11,000	2.5562	0.3570	5,000	10.7806	3.6127	4,810	12.7711	4.0089	1,057
12,000	2.5499	0.3447	5,000	10.9710	3.6182	4,874	13.3524	4.1012	1,047
13,000	2.5555	0.3362	5,000	11.0867	3.6247	4,933	14.6505	4.4322	1,038

(ii) where $f(y) = |y|$

	$\gamma =$	0.99		$\gamma =$	0.999		$\gamma =$	1.00	
n	mean	STD	R	mean	STD	R	mean	STD	R
1,000	0.0163	1.5455	4,805	-0.0355	2.1025	1,959	0.0671	2.0906	1,553
2,000	0.0078	1.2231	5,000	0.0840	2.9303	2,108	0.1366	2.8985	1,349
3,000	0.0024	1.1068	5,000	0.0331	3.8890	2,442	-0.0981	3.6867	1,209
4,000	-0.0158	0.8990	5,000	-0.1331	4.4303	2,697	-0.0747	4.3820	1,121
5,000	0.0035	0.7915	5,000	0.0416	4.9563	2,970	0.3055	4.9907	977
6,000	0.0016	0.7334	5,000	0.1398	5.2765	3,395	-0.0198	5.6800	993
7,000	-0.0039	0.6751	5,000	-0.1955	5.6063	3,718	0.0726	5.9298	868
8,000	-0.0004	0.6241	5,000	-0.0425	5.9141	4,008	0.3187	6.5415	881
9,000	0.0217	0.6047	5,000	0.0144	6.0392	4,307	-0.2095	7.0508	848
10,000	0.0130	0.5549	5,000	-0.0309	6.2077	4,508	-0.0780	7.1496	785
11,000	0.0007	0.5259	5,000	-0.1225	6.3735	4,655	0.3432	7.8830	813
12,000	-0.0048	0.5114	5,000	0.0183	6.1392	4,767	-0.3454	7.9886	710
13,000	-0.0041	0.4910	5,000	-0.1025	6.2142	4,871	-0.2629	8.3086	728

Note: R shows the number of replications which are stationary through ADF test.

Table 3.11: Estimation of α and β with γ fixed at 0.99 where $f(y) = y$

(i) $\gamma = 1.0, \alpha = 1.0$, and $\beta = 0.6$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.9902	0.0115	-2.3188	>1,000	3,696	
$\hat{\alpha}$	1.0232	0.0653	0.2411	83.09		
$\hat{\beta}$	0.6186	0.0361	0.5511	376.72		
$\hat{\alpha}^{\star}$	0.9971	0.0577	0.0517	4.51	3,696	$1.0002(3{,}696)^{\dagger}$
$\hat{\beta}^{\star}$	0.5985	0.0306	0.4186	278.36		
$\hat{\alpha}^{*}$	1.0238	0.0607	0.0785	7.42	3,696	$1.0019(3{,}696)^{\dagger}$
$\hat{\beta}^{*}$	0.6198	0.0329	0.4241	270.21		$1.0017(3{,}696)^{\ddagger}$

(ii) $\gamma = 1.0, \alpha = -3.0$, and $\beta = 0.1$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.9728	0.2892	-1.9698	>1,000	3,237	
$\hat{\alpha}$	-3.4221	2.0433	8.0786	>1,000		
$\hat{\beta}$	0.1750	0.2381	-4.4436	>1,000		
$\hat{\alpha}^{\star}$	-3.0018	0.1018	-0.2114	38.14	5,000	$0.9999(3{,}237)^{\dagger}$
$\hat{\beta}^{\star}$	0.1010	0.0243	0.0773	5.95		
$\hat{\alpha}^{*}$	-3.1169	0.1166	-0.2710	64.83	5,000	$0.9999(3{,}237)^{\dagger}$
$\hat{\beta}^{*}$	0.1118	0.0273	0.1193	16.01		$1.0000(5{,}000)^{\ddagger}$

Note: * and * indicate the estimation of α and β with γ fixed at 1 and with γ fixed at 0.99. † and † show MSE ratio over the estimation of γ unknown and of γ fixed at 1 respectively.

Table 3.12: Estimation of α and β with γ fixed at 0.99 where $f(y) = |y|$

(i) $\gamma = 1.0, \alpha = 1.0$, and $\beta = 0.6$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.9912	0.0104	-2.4185	0.93	3,171	
$\hat{\alpha}$	0.9708	0.1654	-0.0383	234.73		
$\hat{\beta}$	0.6418	0.0903	0.5444	>1,000		
$\hat{\alpha}^{\star}$	1.0209	0.1621	-0.0533	1.93	3,171	$1.0003(3,171)^{\dagger}$
$\hat{\beta}^{\star}$	0.5876	0.0761	0.3304	73.95		
$\hat{\alpha}^{*}$	0.9628	0.1663	-0.0471	1.92	3,171	$1.0021(3,171)^{\dagger}$
$\hat{\beta}^{*}$	0.6491	0.0836	0.2886	59.67		$1.0019(3,171)^{\ddagger}$

(ii) $\gamma = 1.0, \alpha = -3.0$, and $\beta = 0.1$, $n = 3,000$

	estimate	STD	skewness	JB	R	MSE ratio
$\hat{\gamma}$	0.8630	0.4914	-2.7935	>1,000	801	
$\hat{\alpha}$	-5.5763	10.3299	-2.6078	>1,000		
$\hat{\beta}$	0.5175	1.6735	4.1472	>1,000		
$\hat{\alpha}^{\star}$	-3.0251	0.2130	-0.1979	32.66	5,000	$0.9998(801)^{\dagger}$
$\hat{\beta}^{\star}$	0.1081	0.0580	0.1728	25.80		
$\hat{\alpha}^{*}$	-3.1287	0.2309	-0.2103	36.83	5,000	$0.9998(801)^{\dagger}$
$\hat{\beta}^{*}$	0.1164	0.0625	0.1711	25.28		$1.0000(5,000)^{\ddagger}$

Note: * and * indicate the estimation of α and β with γ fixed at 1 and with γ fixed at 0.99. † and † show MSE ratio over the estimation of γ unknown and of γ fixed at 1 respectively.

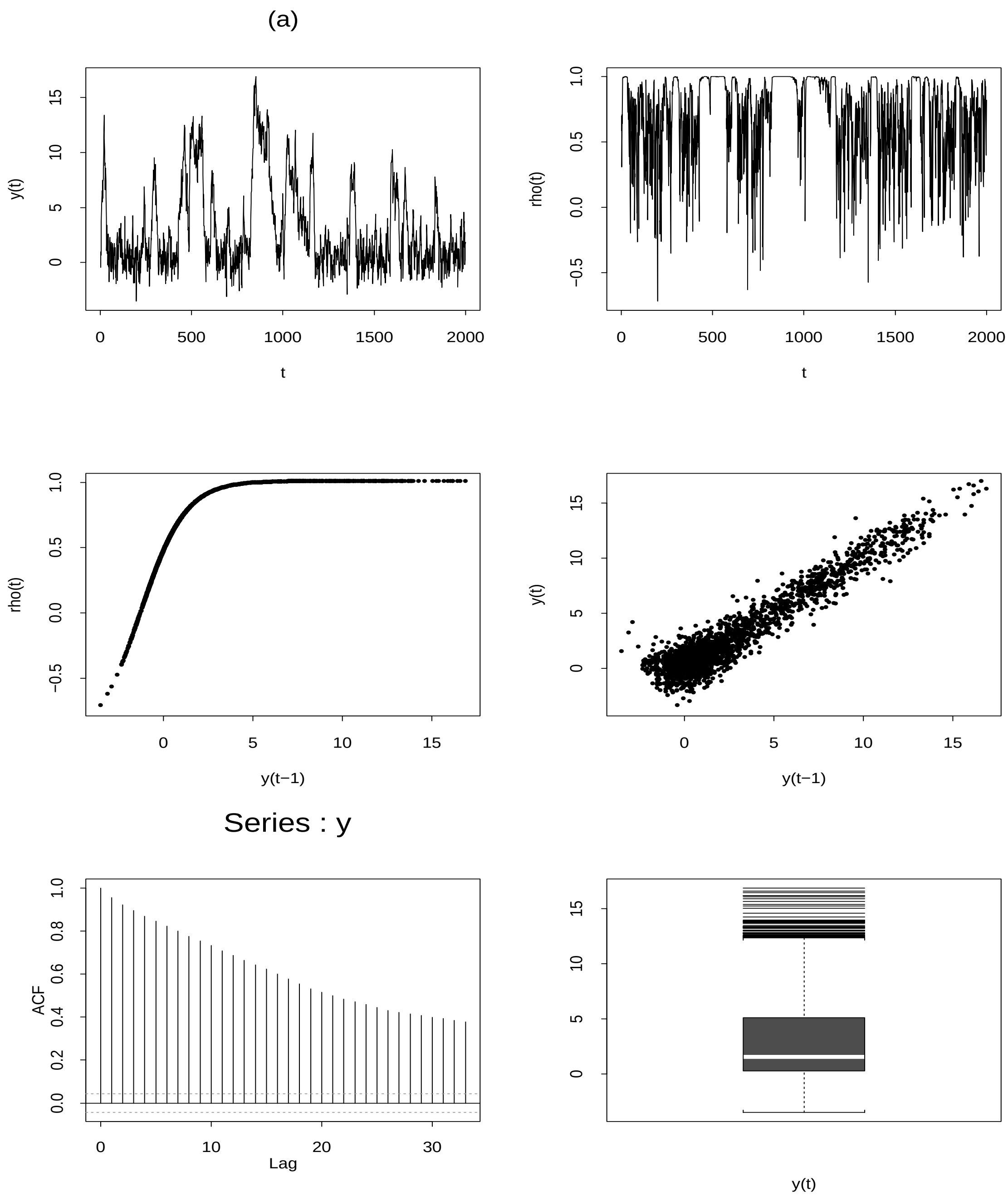


Figure 3.1: y_t and $\rho(y_{t-1})$ of series generated by (a) $(\gamma, \alpha, \beta) = (0.99, 1.0, 0.8)$ where $f(y) = y$

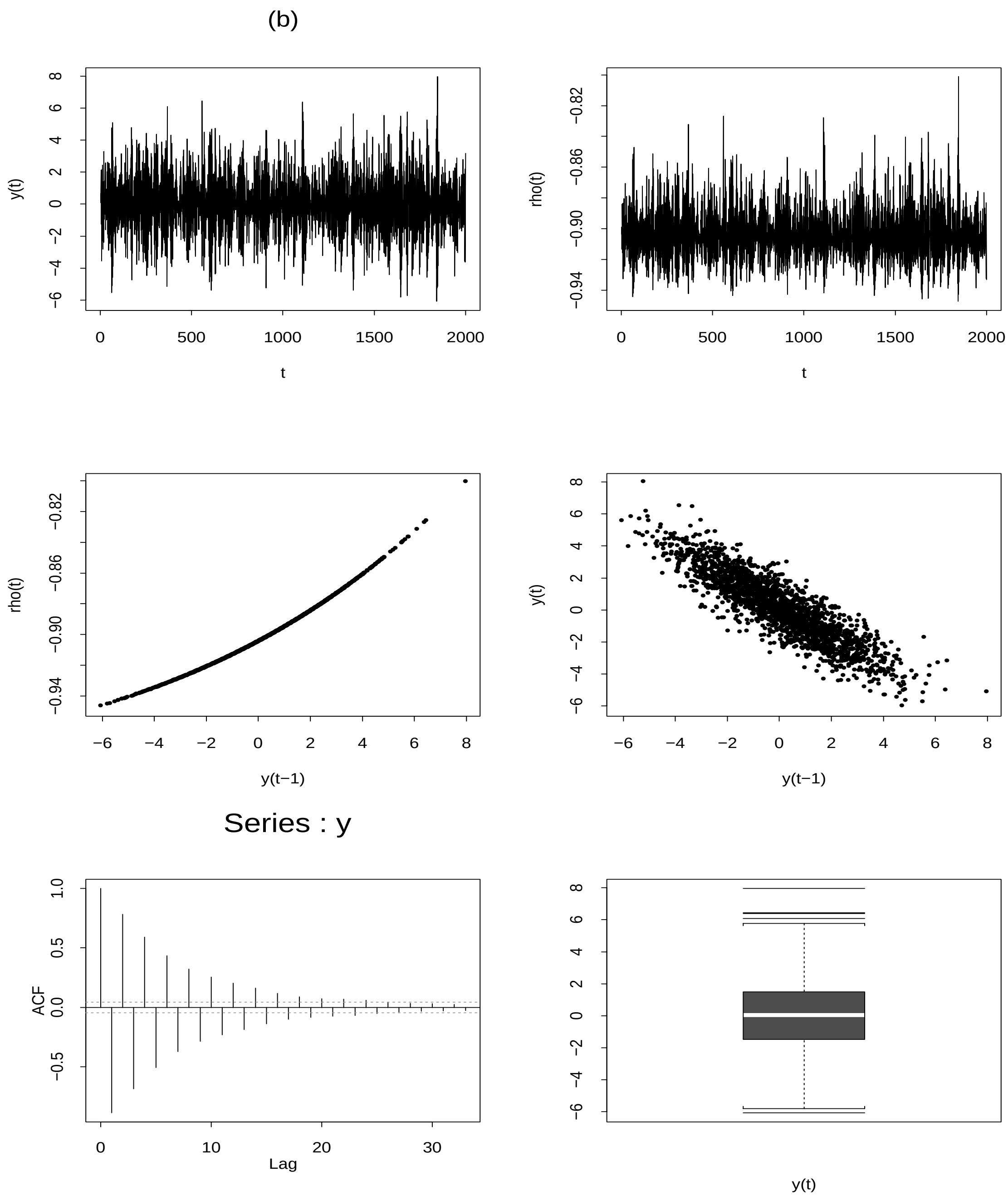


Figure 3.2: y_t and $\rho(y_{t-1})$ of series generated by (b) $(\gamma, \alpha, \beta) = (0.99, -3.0, 0.1)$ where $f(y) = y$

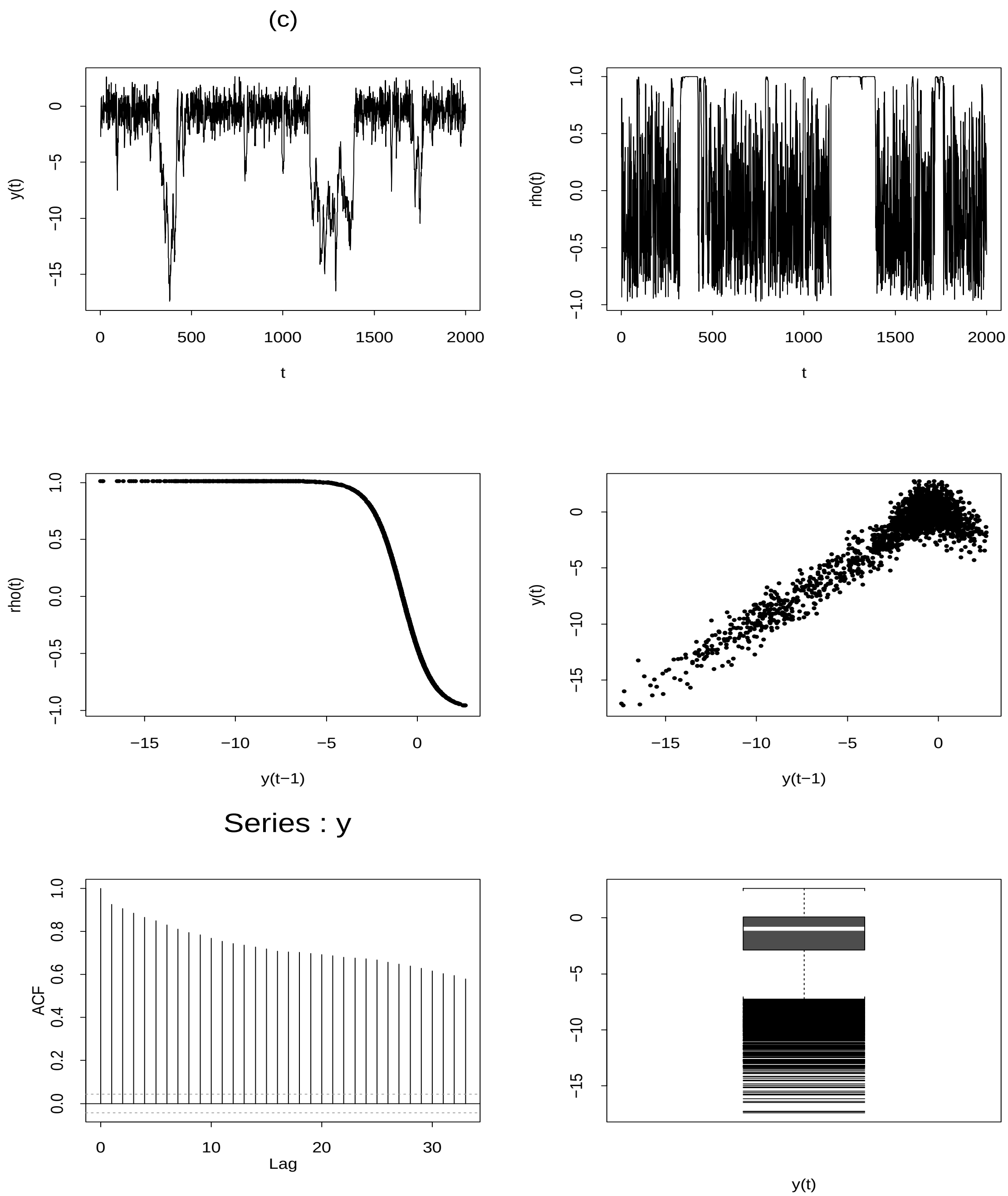


Figure 3.3: y_t and $\rho(y_{t-1})$ of series generated by (c) $(\gamma, \alpha, \beta) = (0.99, -1.0, -1.2)$ where $f(y) = y$

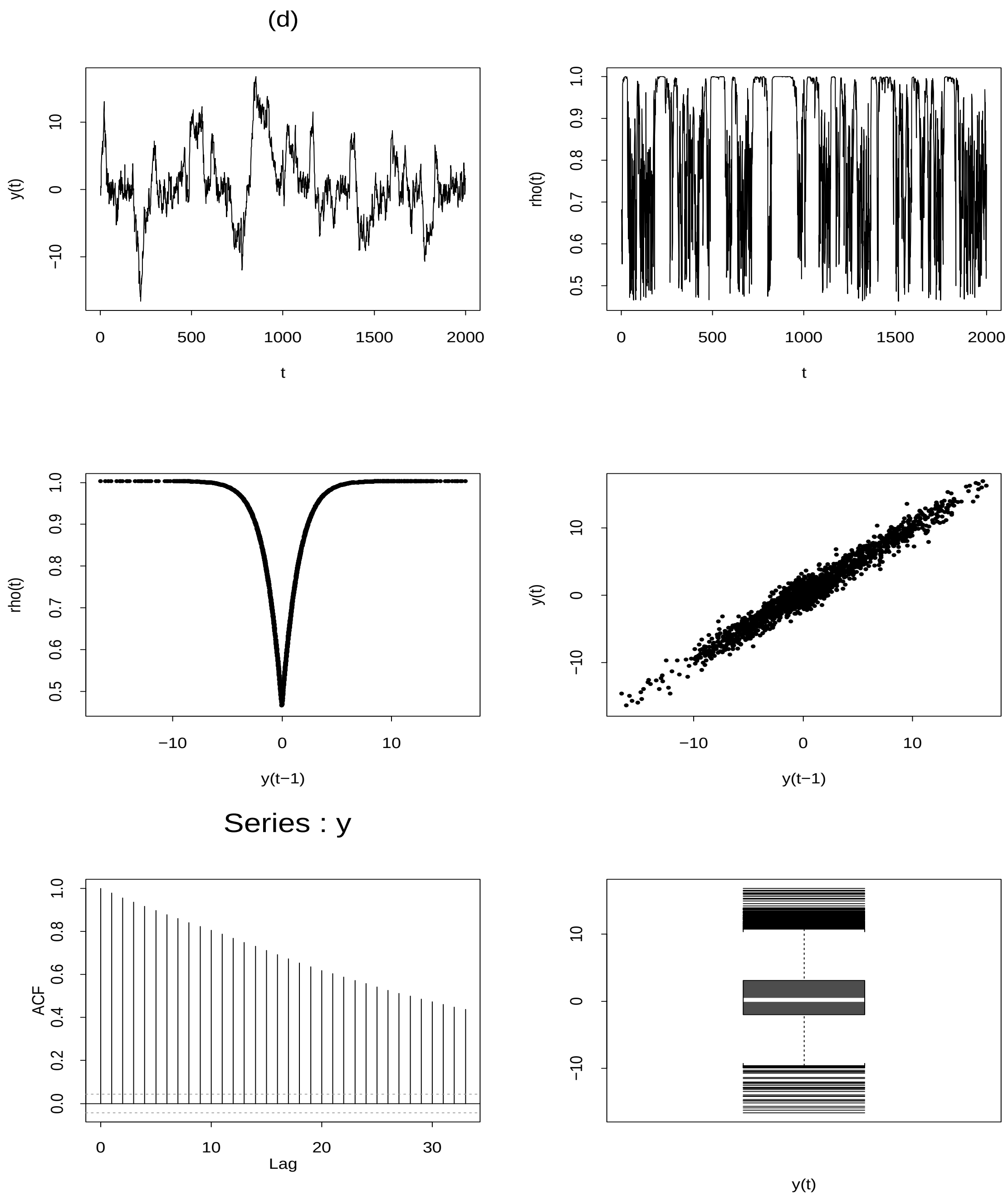


Figure 3.4: y_t and $\rho(y_{t-1})$ of series generated by (d) $(\gamma, \alpha, \beta) = (0.99, 1.0, 0.8)$ where $f(y) = |y|$

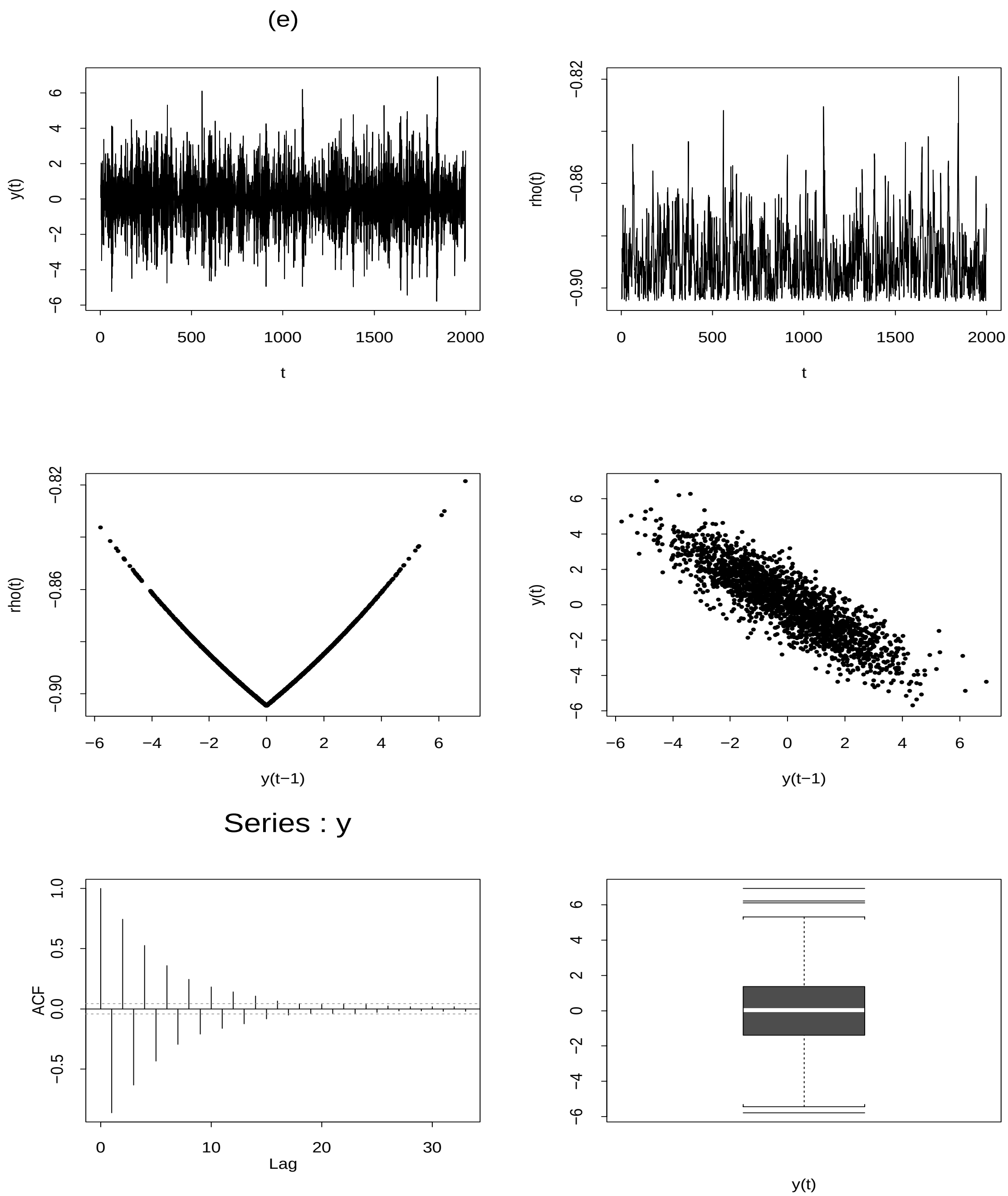


Figure 3.5: y_t and $\rho(y_{t-1})$ of series generated by (e) $(\gamma, \alpha, \beta) = (0.99, -3.0, 0.1)$ where $f(y) = |y|$

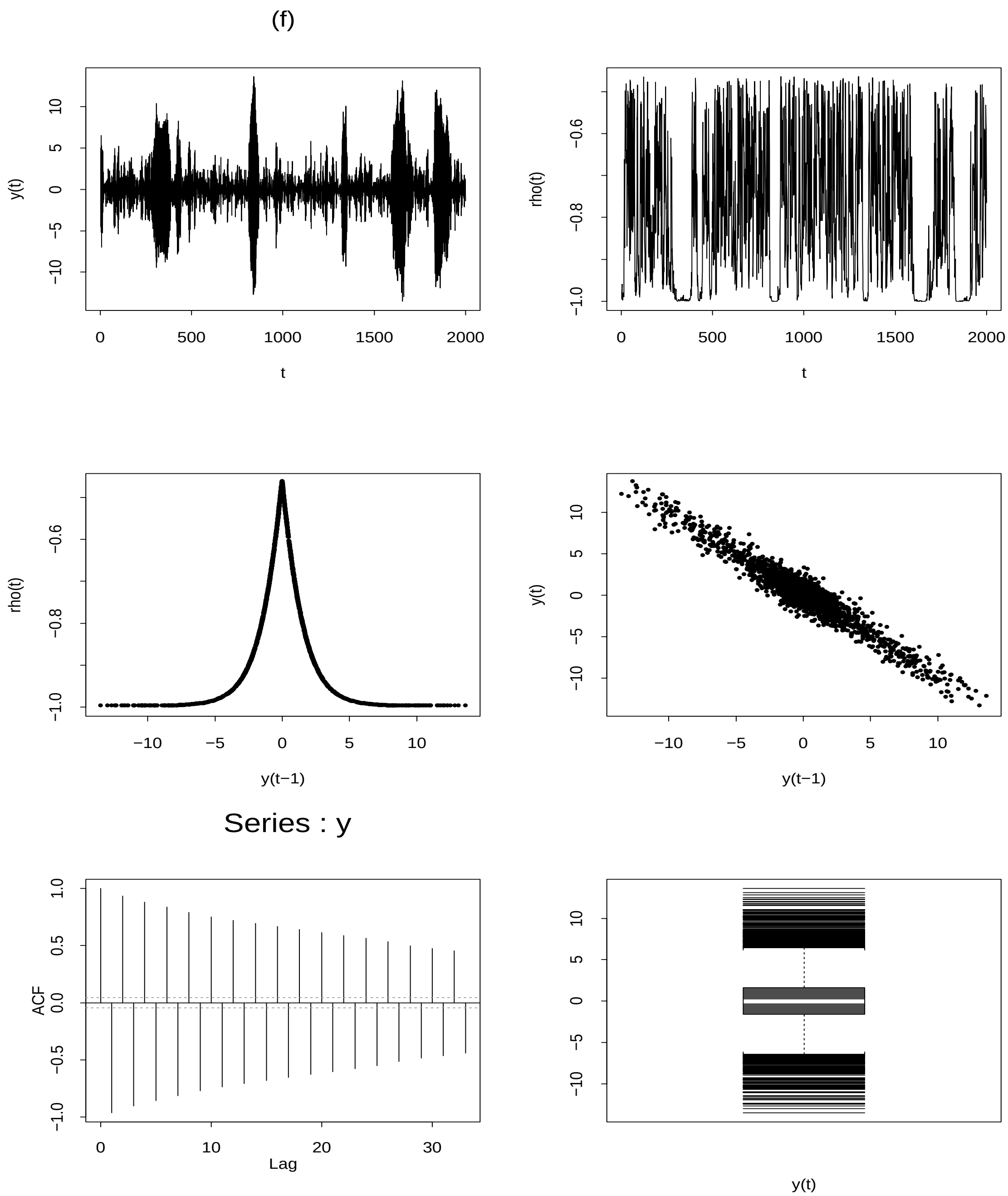


Figure 3.6: y_t and $\rho(y_{t-1})$ of series generated by (f) $(\gamma, \alpha, \beta) = (0.99, -1.0, -0.8)$ where $f(y) = |y|$

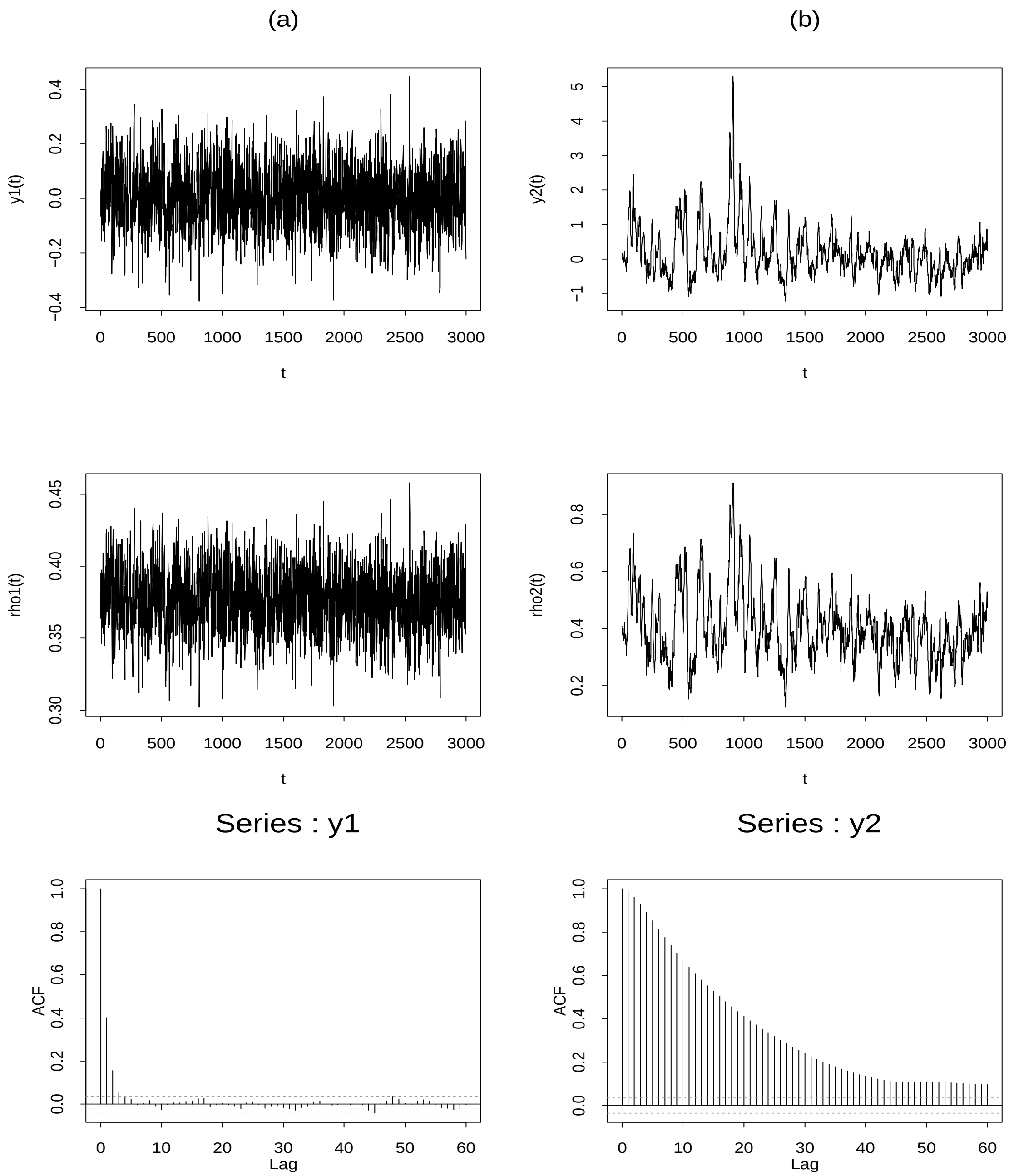


Figure 3.7: Examples of series generated by $(\gamma, \alpha, \beta, \sigma) = (0.99, 0.8, 0.45, 0.1)$ under (a) $\eta_t = e_t$ (b) $\eta_t = 0.95\eta_{t-1} + e_t$ where $f(y) = y$

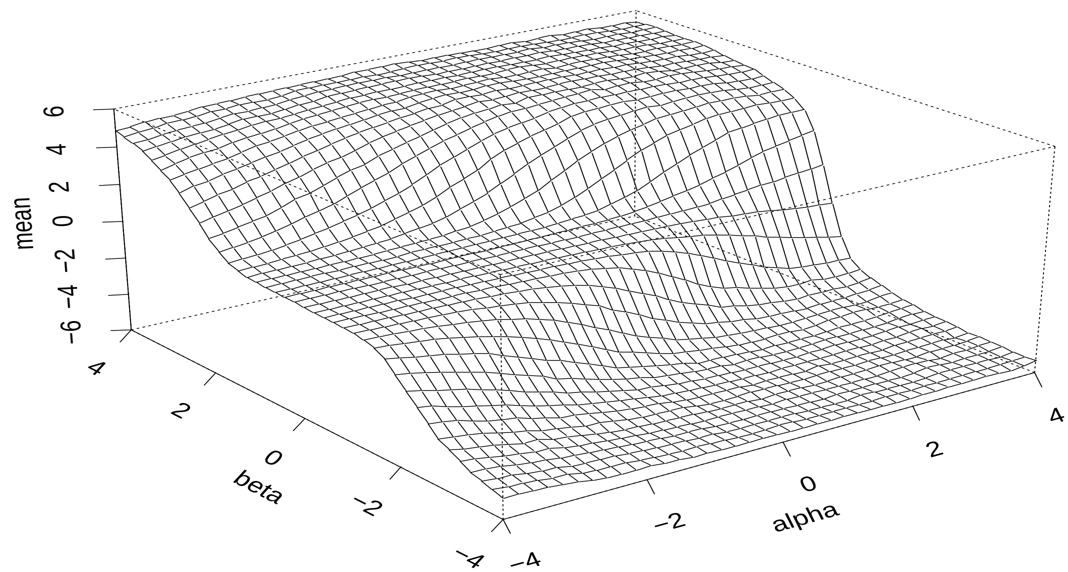


Figure 3.8: The trend of mean with $\sigma^2 = 1.0$ where $f(y) = y$

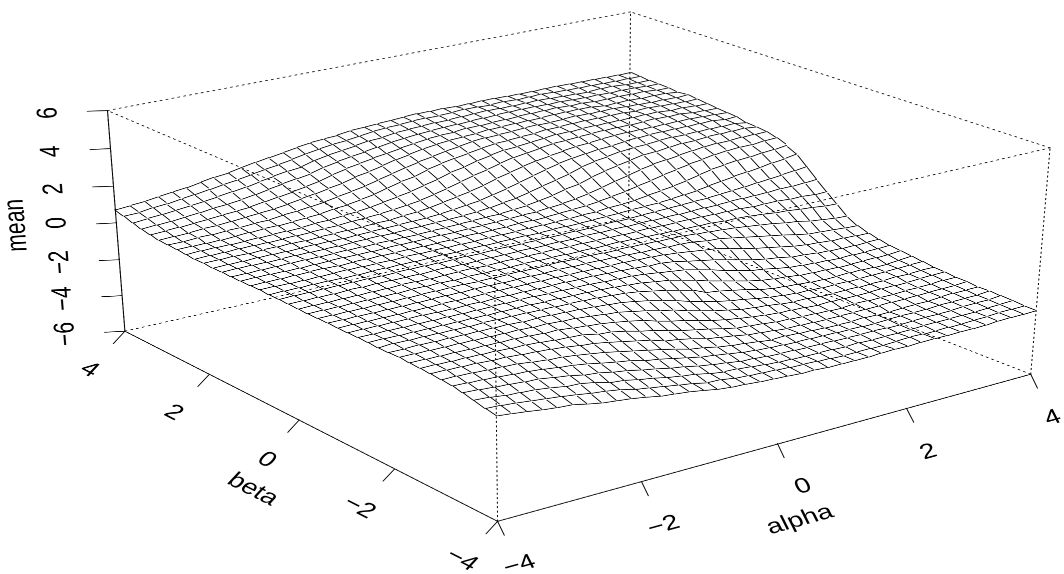


Figure 3.9: The trend of mean with $\sigma^2 = 0.25$ where $f(y) = y$

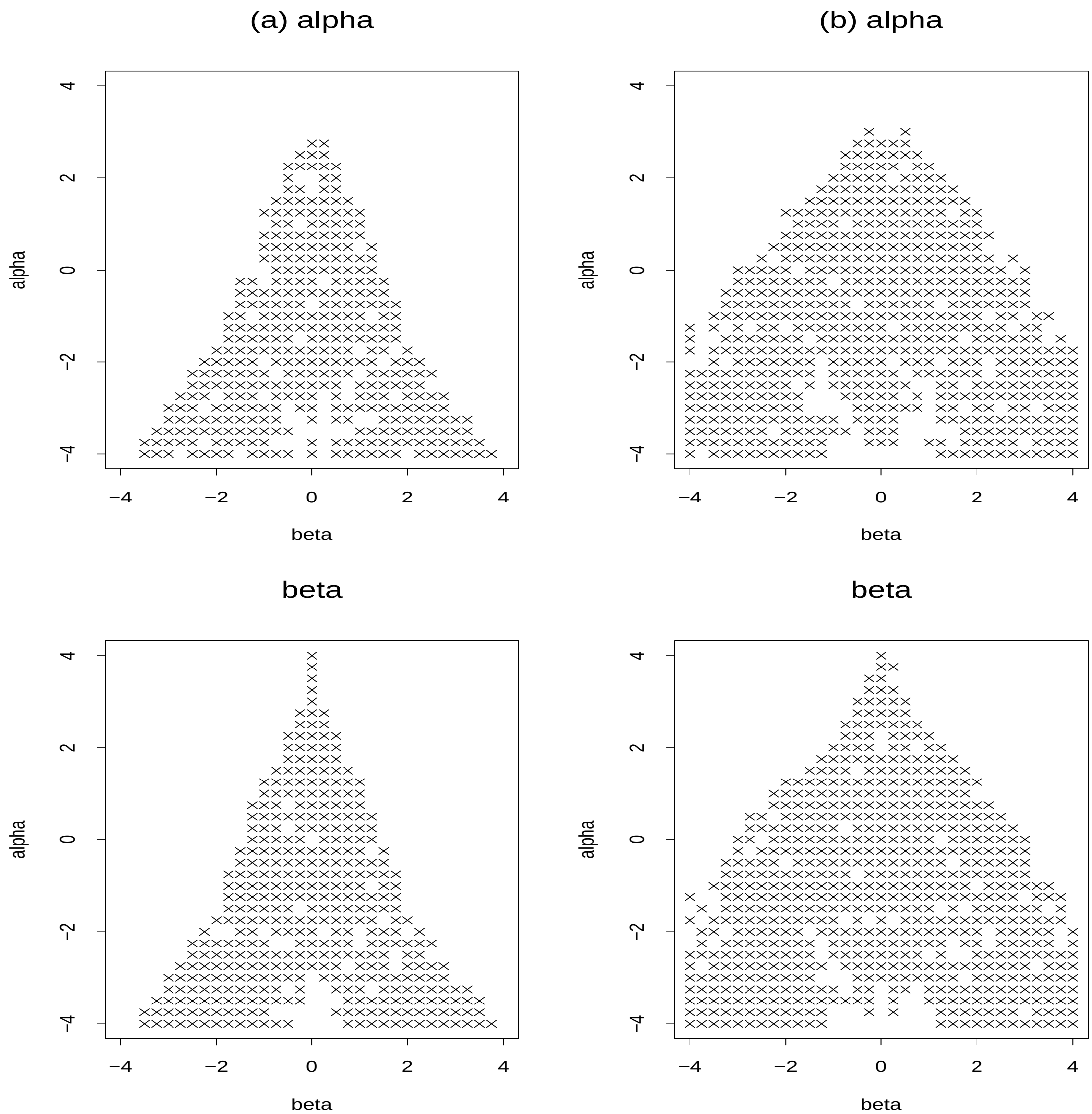


Figure 3.10: The areas where t tests of $\alpha = \alpha_0$ and $\beta = \beta_0$ have empirical rejection rates not significantly different from 0.05 based on binomial test. $f(y) = y$ and nonlinear estimation convergence rates bigger than 99.5%; (a) $\sigma = 1.0$ and (b) $\sigma = 0.5$

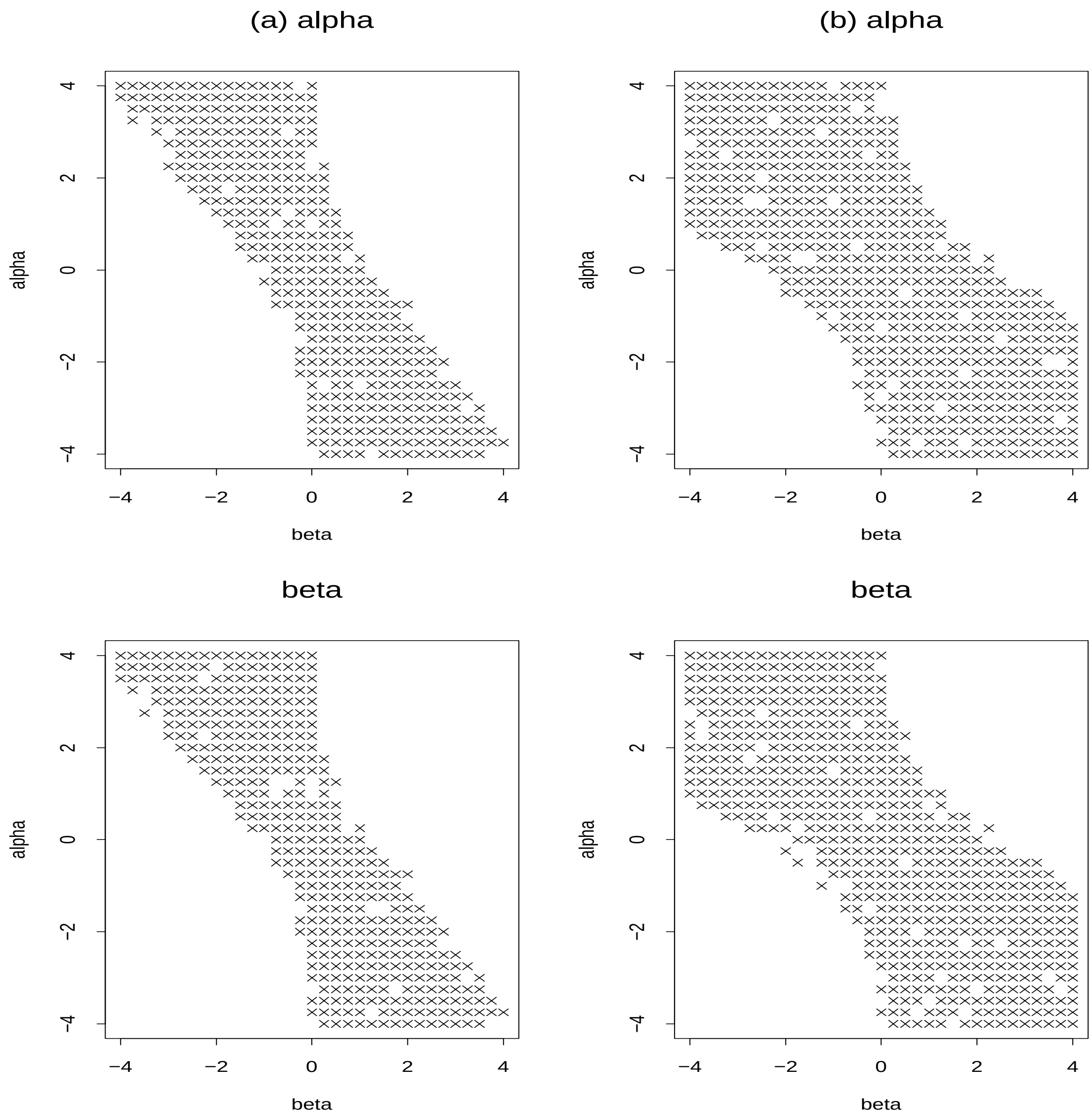


Figure 3.11: The areas where t tests of $\alpha = \alpha_0$ and $\beta = \beta_0$ have empirical rejection rates not significantly different from 0.05 based on binomial test. $f(y) = |y|$ and nonlinear estimation convergence rates bigger than 99.5%; (a) $\sigma = 1.0$ and (b) $\sigma = 0.5$

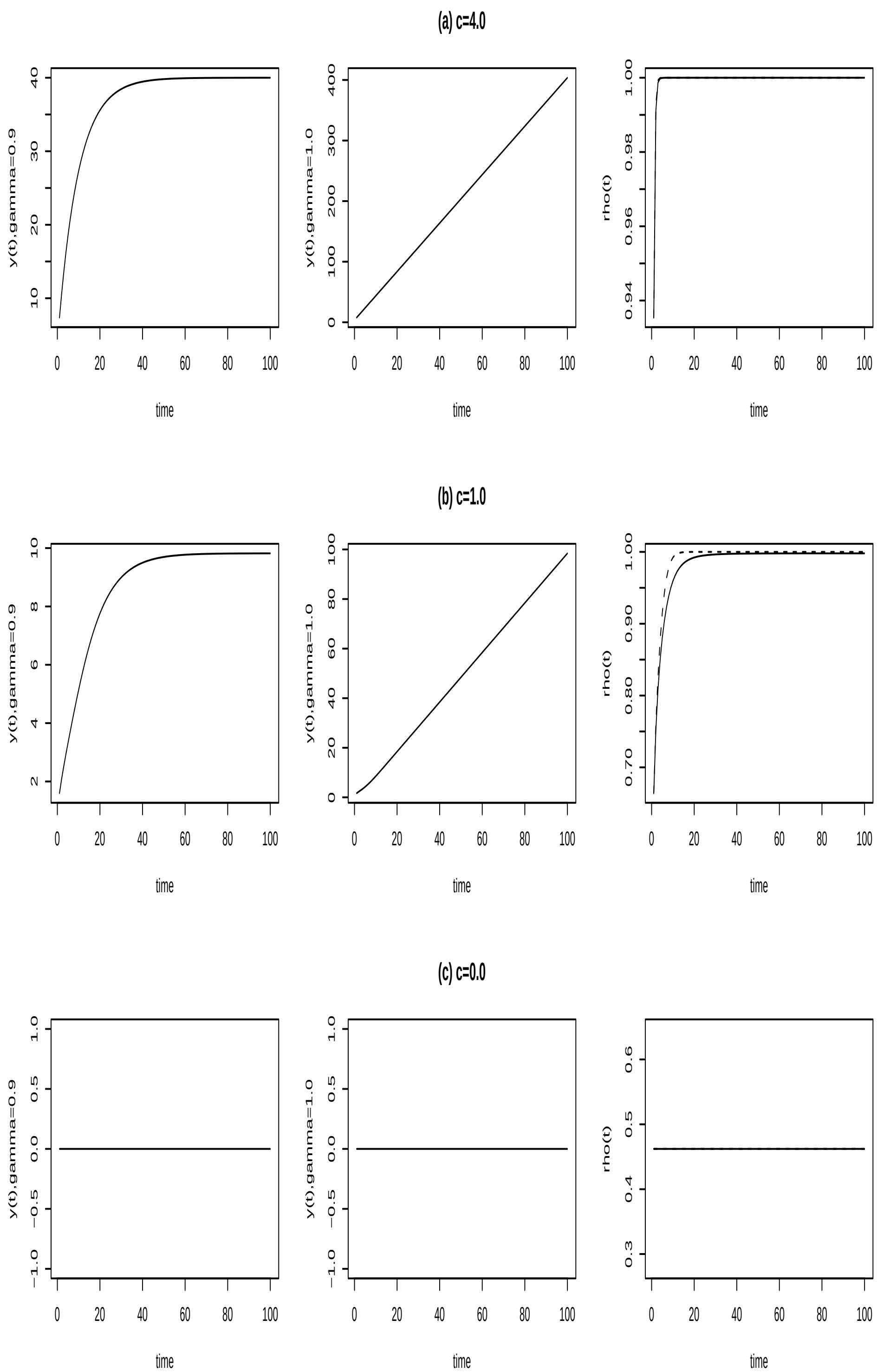


Figure 3.12: The trend of y_t and $\rho(y_{t-1})$; (a) $c = 4.0$, (b) $c = 1.0$, (c) $c = 0.0$. From the left, $\gamma = 0.9$, $\gamma = 1.0$. For $\rho(y_{t-1})$, the solid line is at $\gamma = 0.9$ and the dotted line is at $\gamma = 1.0$.

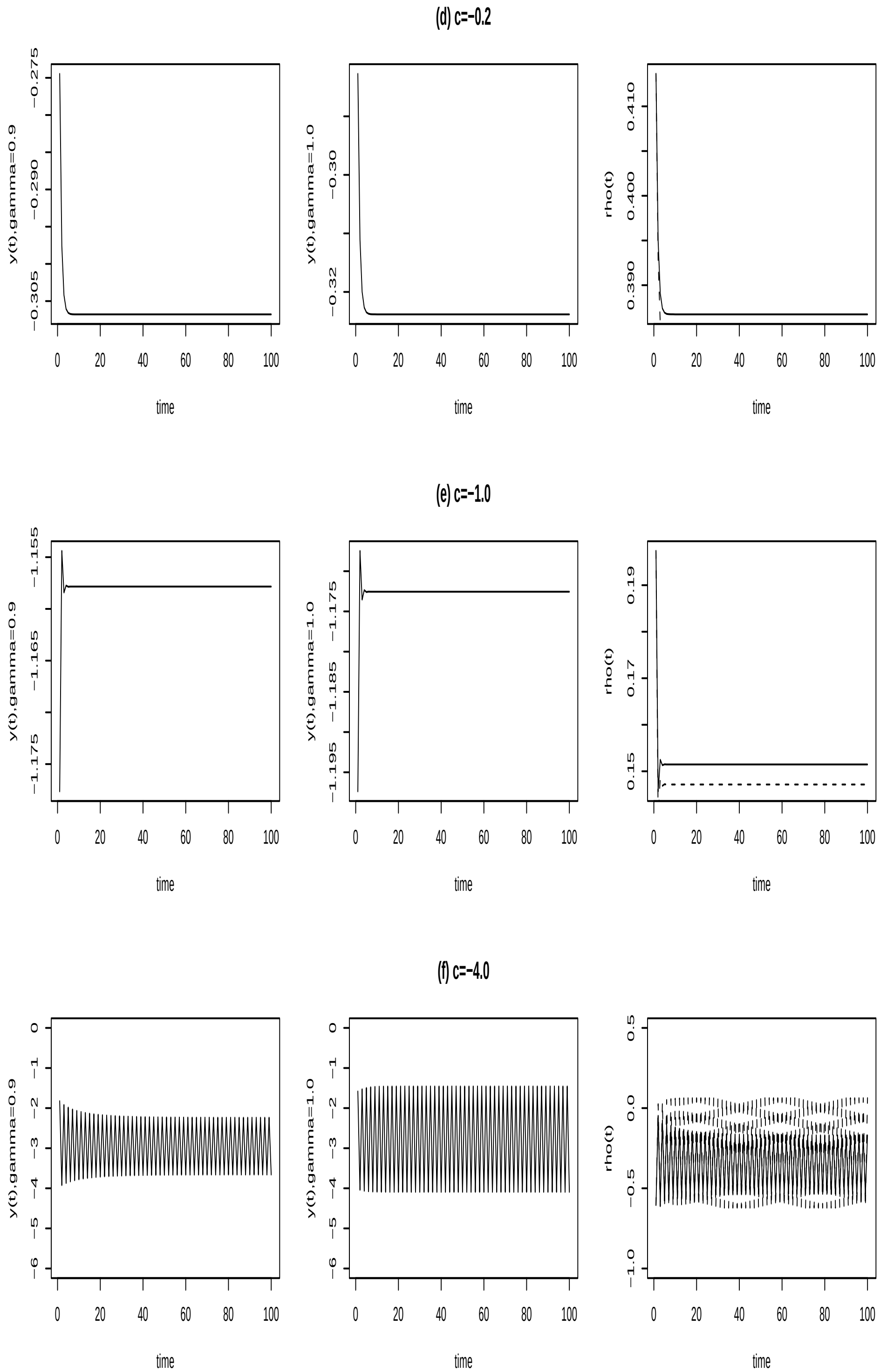


Figure 3.13: The trend of y_t and $\rho(y_{t-1})$; (d) $c = -0.2$, (e) $c = -1.0$, (f) $c = -4.0$. From the left, $\gamma = 0.9$, $\gamma = 1.0$. For $\rho(y_{t-1})$, the solid line is at $\gamma = 0.9$ and the dotted line is at $\gamma = 1.0$.

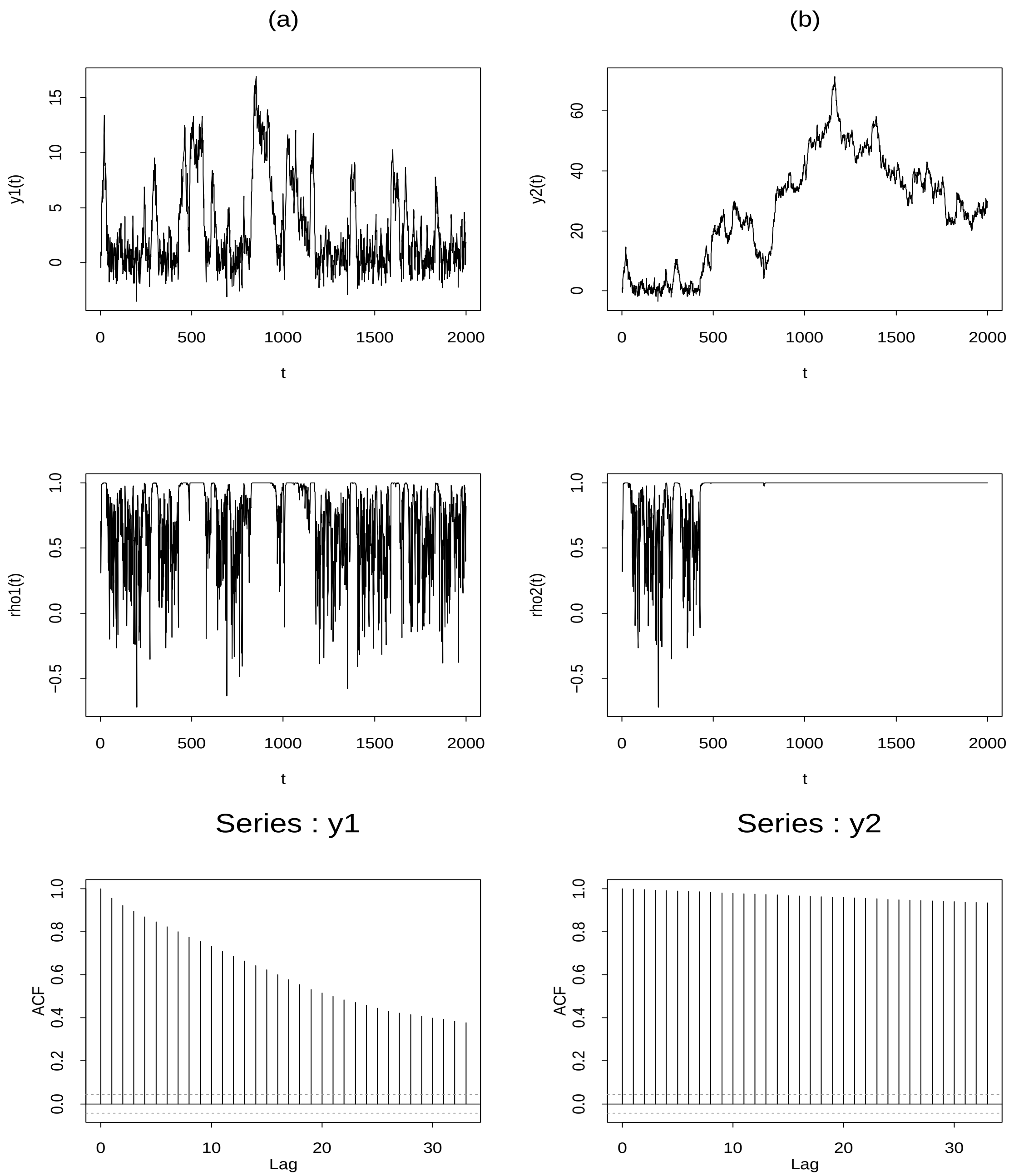


Figure 3.14: Examples of series generated by (a) $(\gamma, \alpha, \beta) = (0.99, 1.0, 0.8)$, (b) $(\gamma, \alpha, \beta) = (1.0, 1.0, 0.8)$ where $f(y) = y$

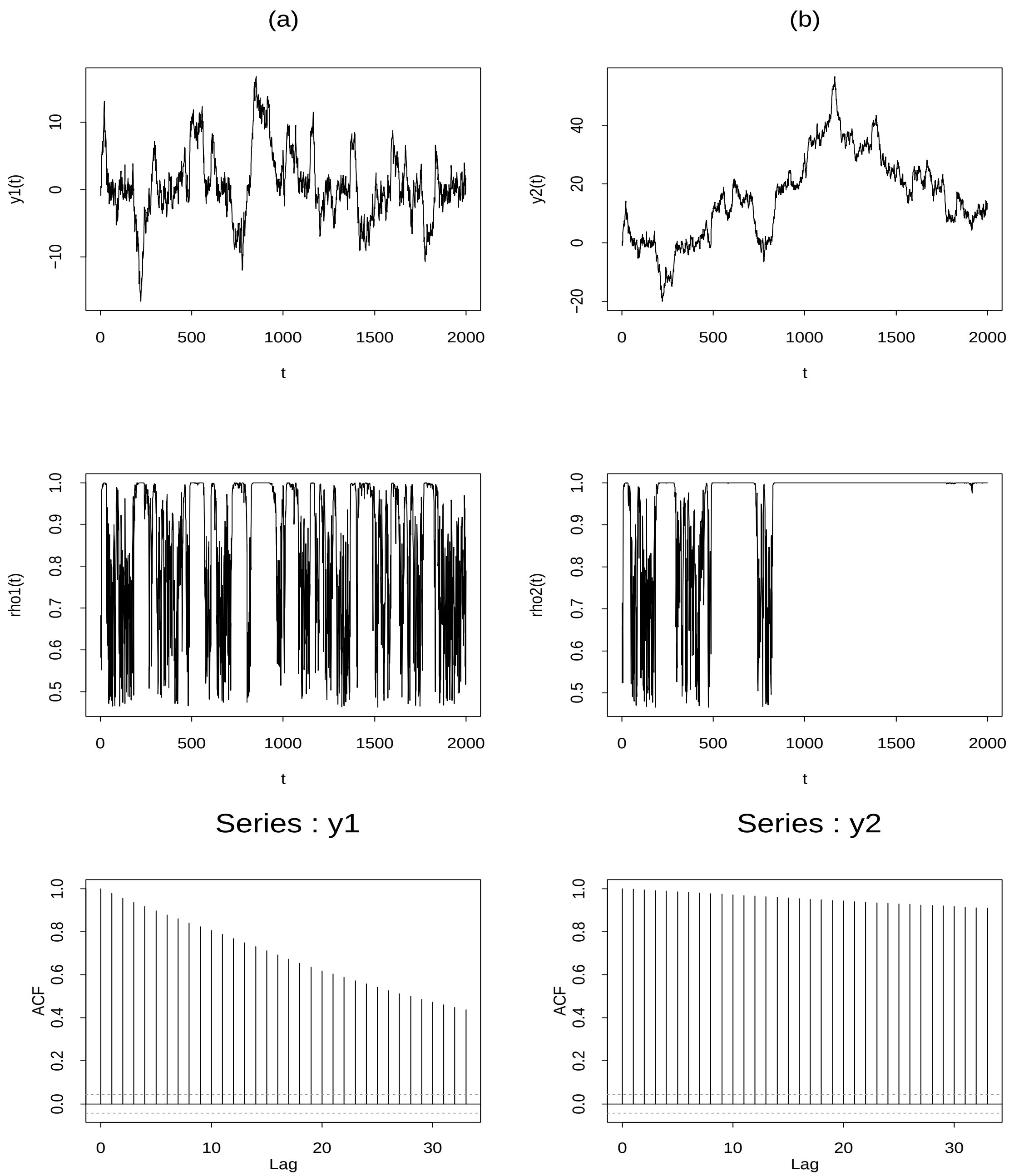


Figure 3.15: Examples of series generated by (a) $(\gamma, \alpha, \beta) = (0.99, 1.0, 0.8)$, (b) $(\gamma, \alpha, \beta) = (1.0, 1.0, 0.8)$ where $f(y) = |y|$

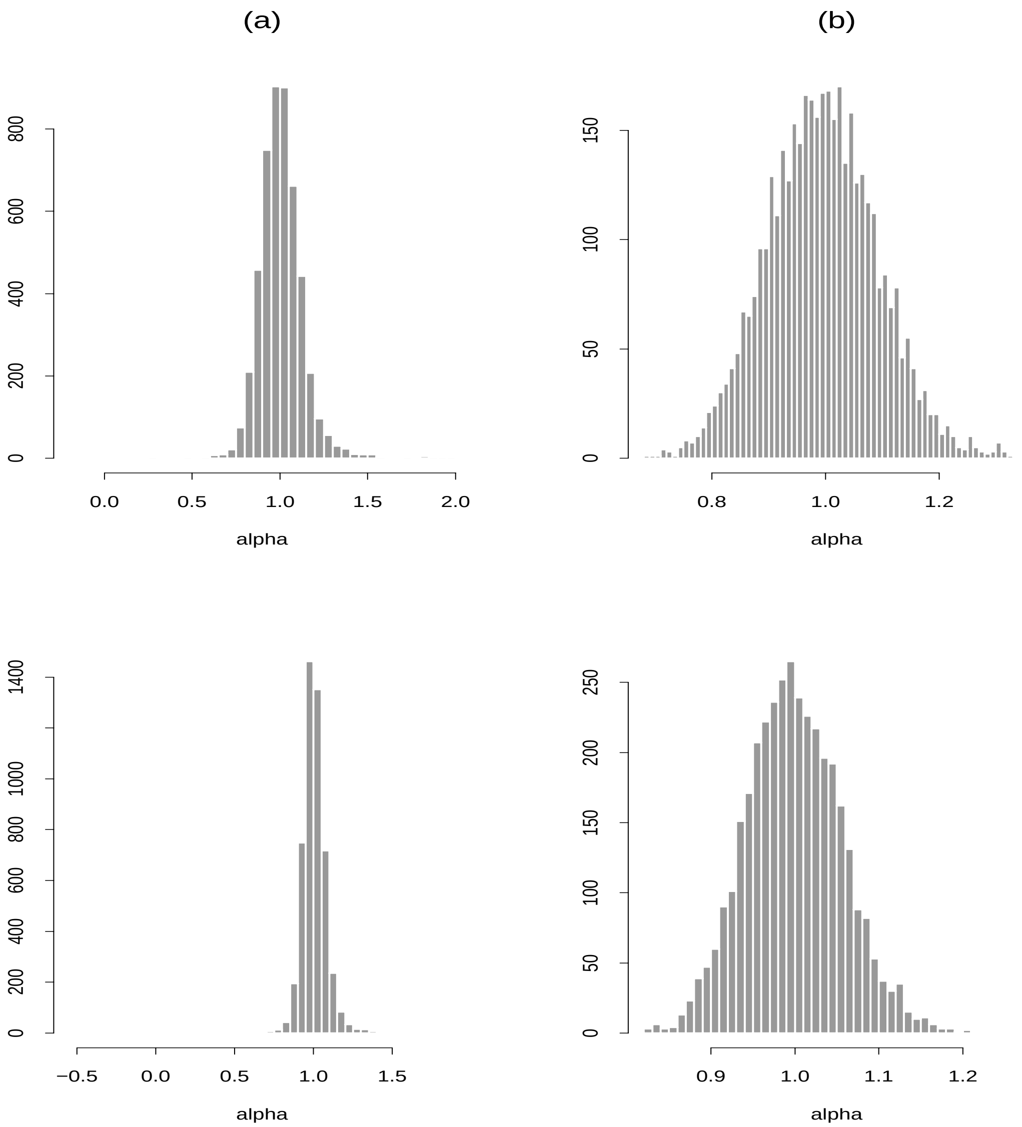


Figure 3.16: The distribution of $\hat{\alpha}$; (a) before unit root test, extreme observations eliminated, (b) after unit root test. From the top, $n = 1,000$ and $n = 3,000$ repectively

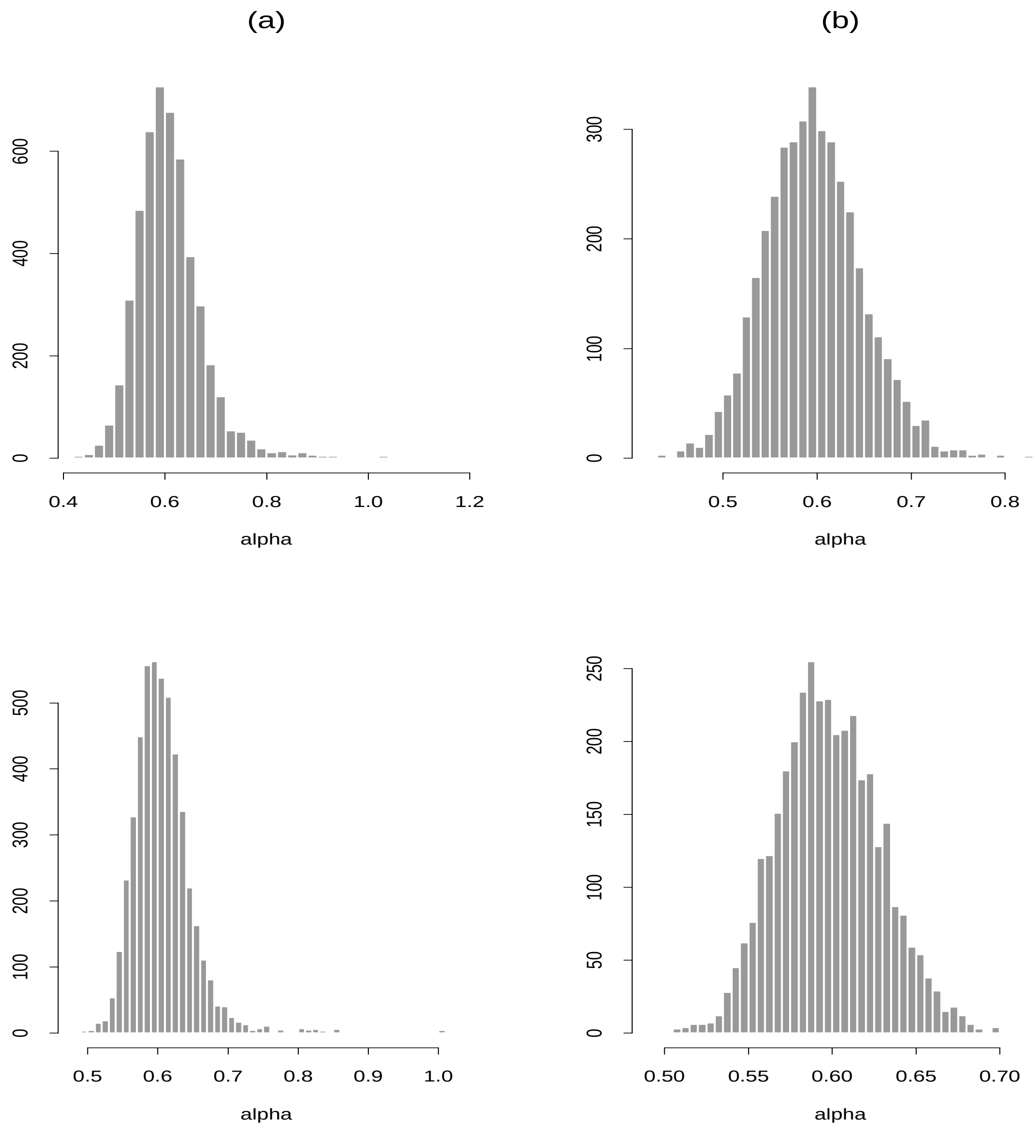


Figure 3.17: The distribution of $\hat{\beta}$; (a) before unit root test, extreme observations eliminated, (b) after unit root test. From the top, $n = 1,000$ and $n = 3,000$ repectively

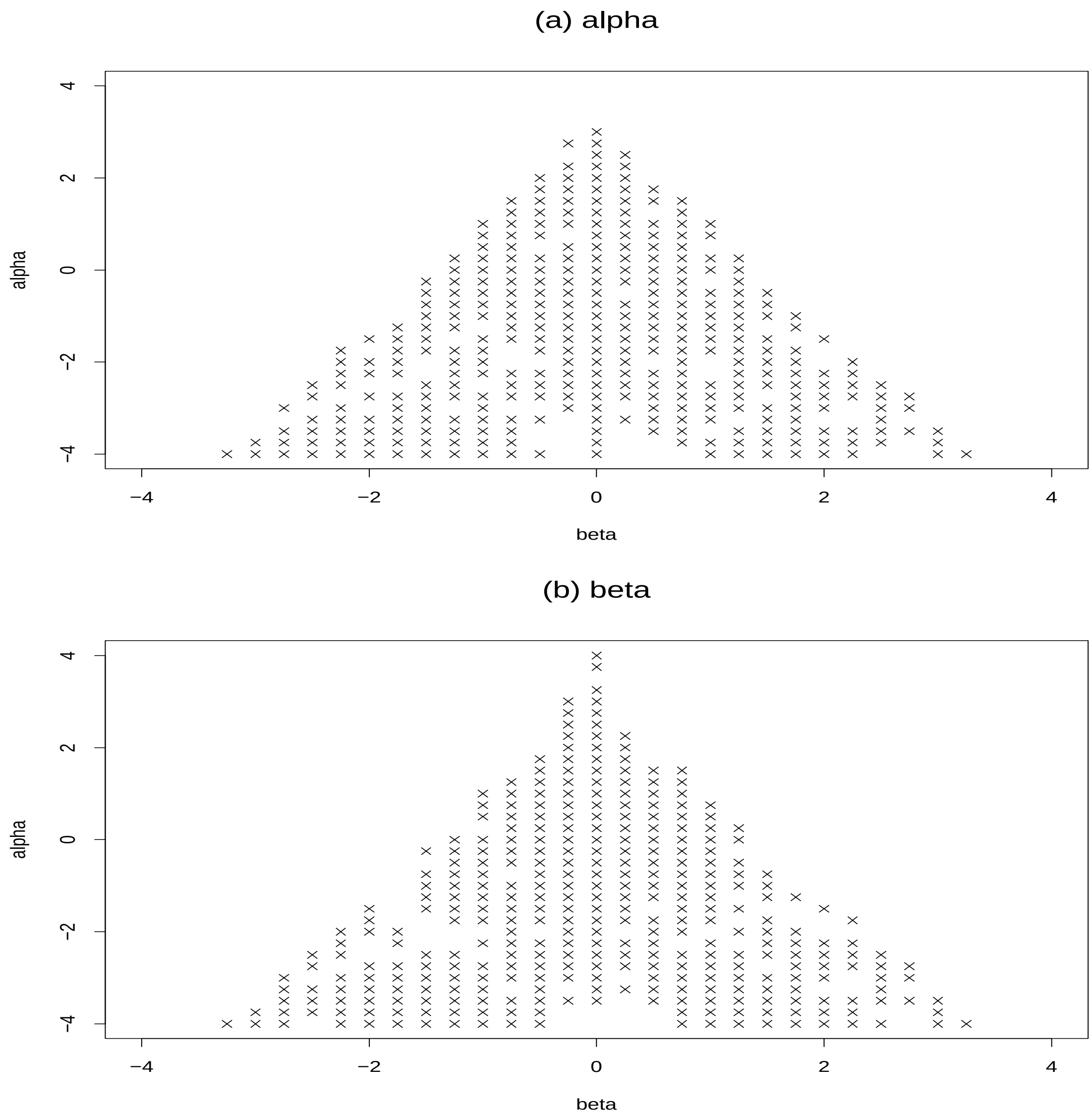
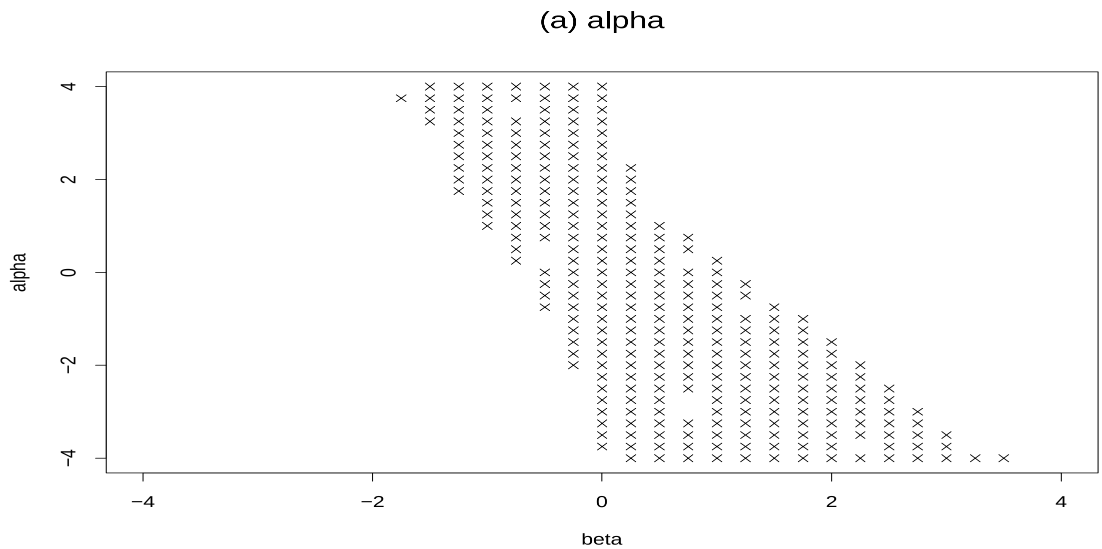


Figure 3.18: The areas where t tests of $\alpha = \alpha_0$ and $\beta = \beta_0$ have empirical rejection rates not significantly different from 0.05 based on binomial test. Nonlinear estimation convergence rates bigger than 99.5%. $\sigma^2 = 1.0$ and $f(y) = y$; (a) $\alpha = \alpha_0$, (b) $\beta = \beta_0$



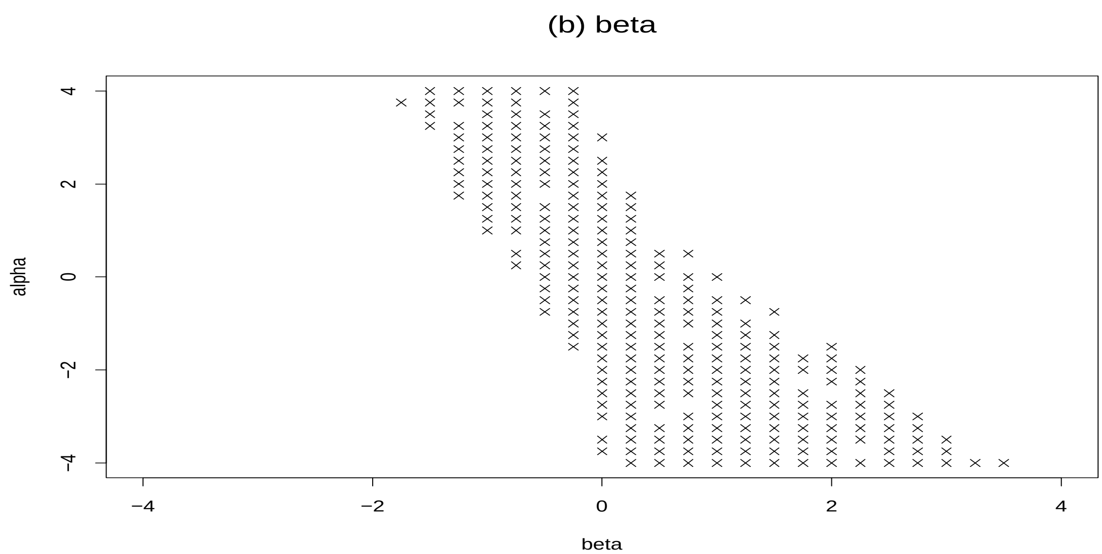


Figure 3.19: The areas where t tests of $\alpha = \alpha_0$ and $\beta = \beta_0$ have empirical rejection rates not significantly different from 0.05 based on binomial test. Nonlinear estimation convergence rates bigger than 99.5%. $\sigma^2 = 1.0$ and $f(y) = |y|$; (a) $\alpha = \alpha_0$, (b) $\beta = \beta_0$

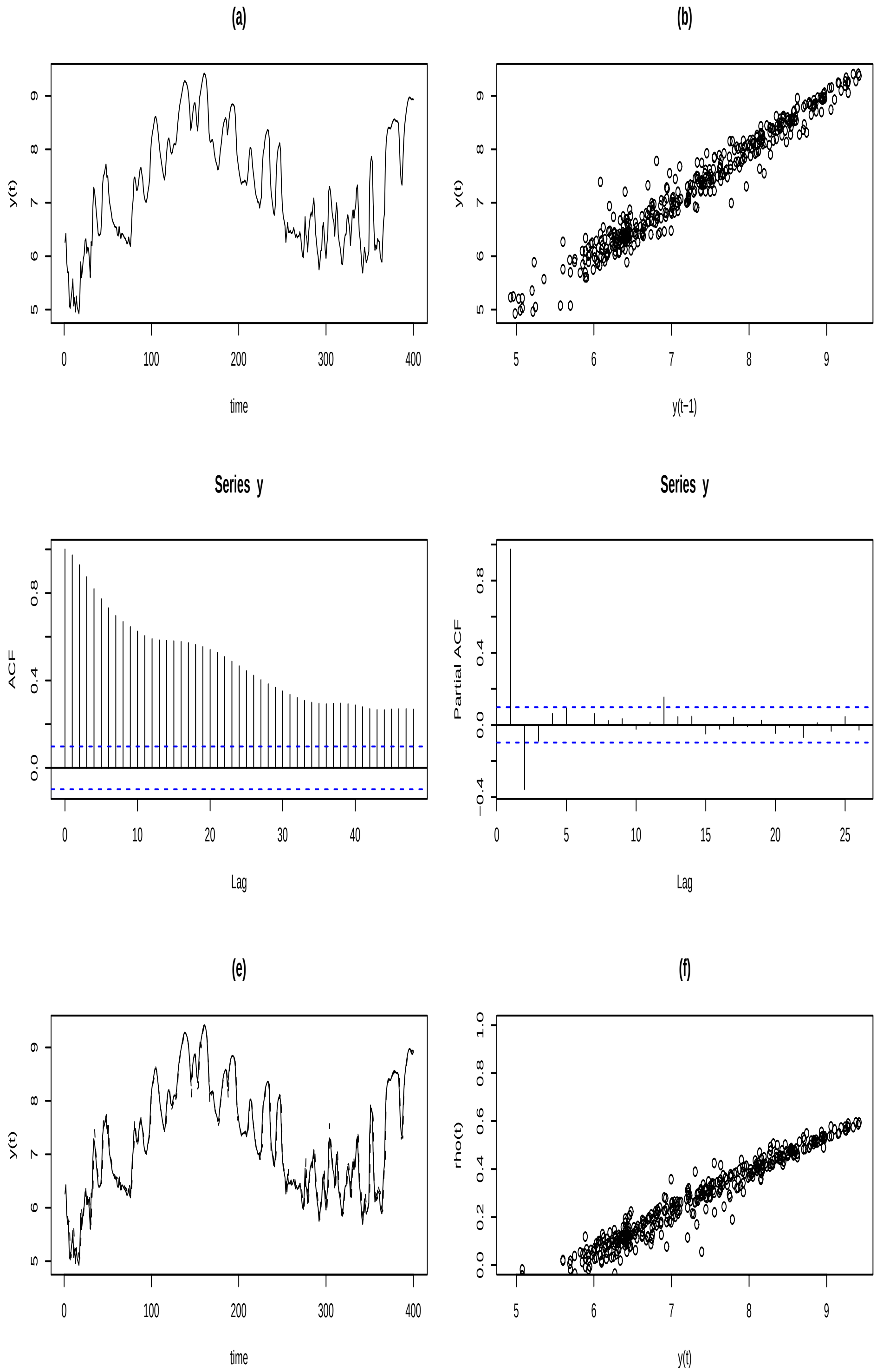


Figure 3.20: Stream flows in Goldsboro; (a) original series(y_t), (b) scatter plot of y_t and y_{t-1}, (c) autocorrelation of y_t, (d) partial autocorrelation of y_t, (e) prediction(y_t: solid line, prediction: dotted line), (f) $\rho(y_{t-1})$ vs y_{t-1}

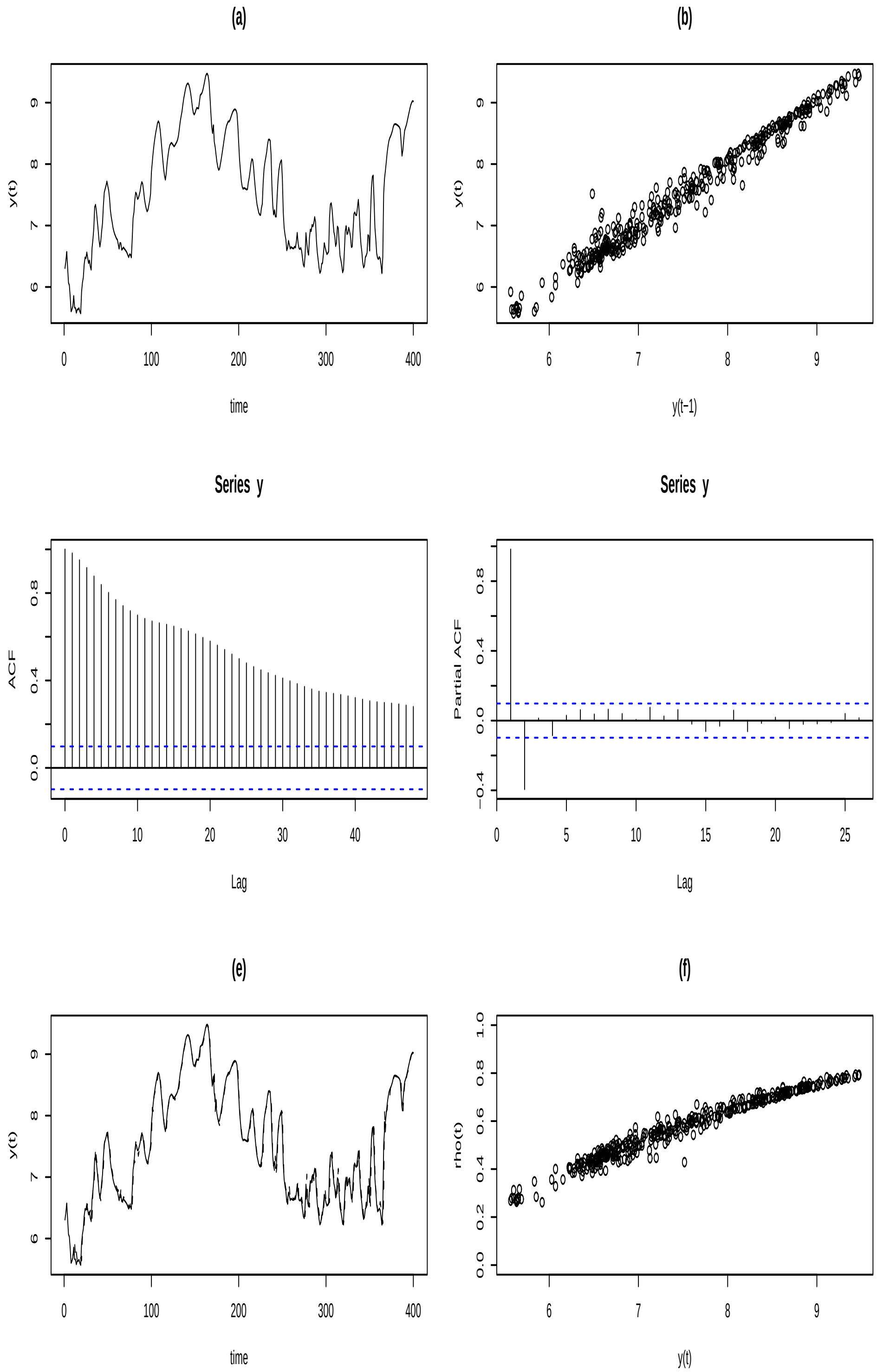


Figure 3.21: Stream flows in Kinston; (a) original series(y_t), (b) scatter plot of y_t and y_{t-1}, (c) autocorrelation of y_t, (d) partial autocorrelation of y_t, (e) prediction(y_t: solid line, prediction: dotted line), (f) $\rho(y_{t-1})$ vs y_{t-1}

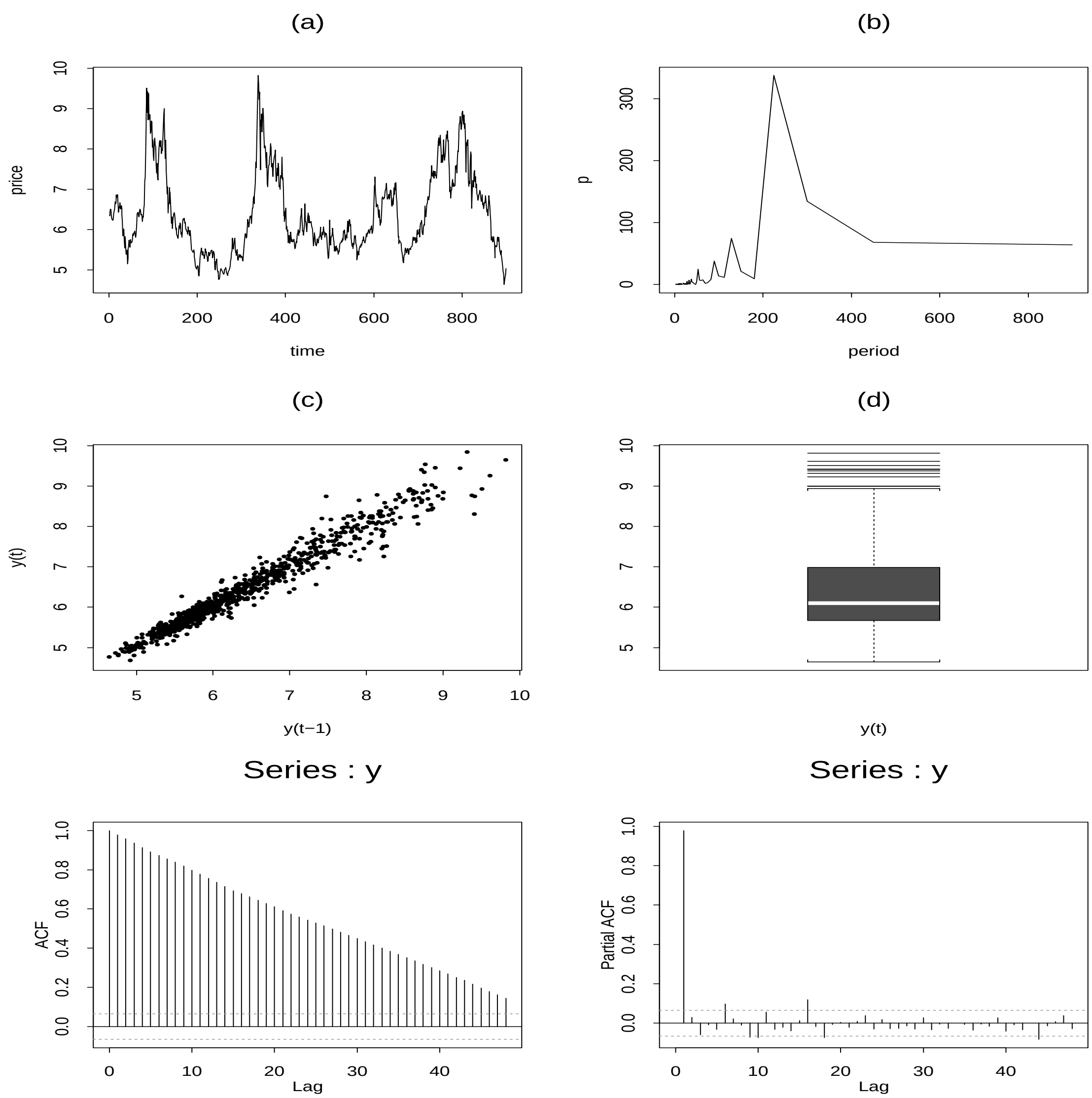


Figure 3.22: Weekly soybean prices in North Carolina; (a) soybean prices, (b) periodgram of soybean prices, (c) scatter plot of y_t and y_{t-1}, (d) boxplot of y_t, (e) autocorrelation of y_t, (f) partial autocorrelation of y_t

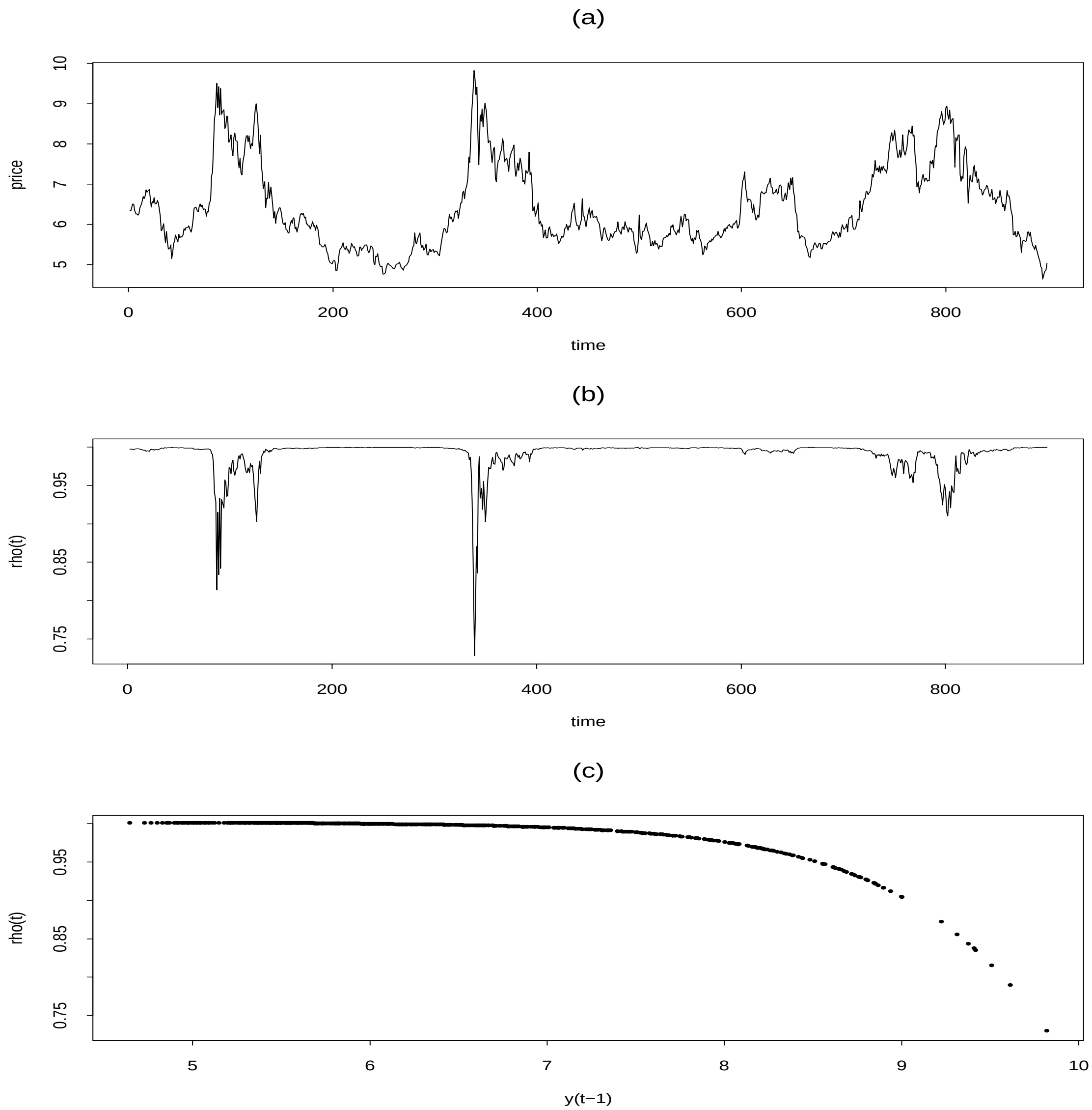


Figure 3.23: Prediction of Soybean prices; (a) prediction of prices(dotted line), (b) $\rho(y_{t-1})$ vs time, (c) $\rho(y_{t-1})$ vs y_{t-1}

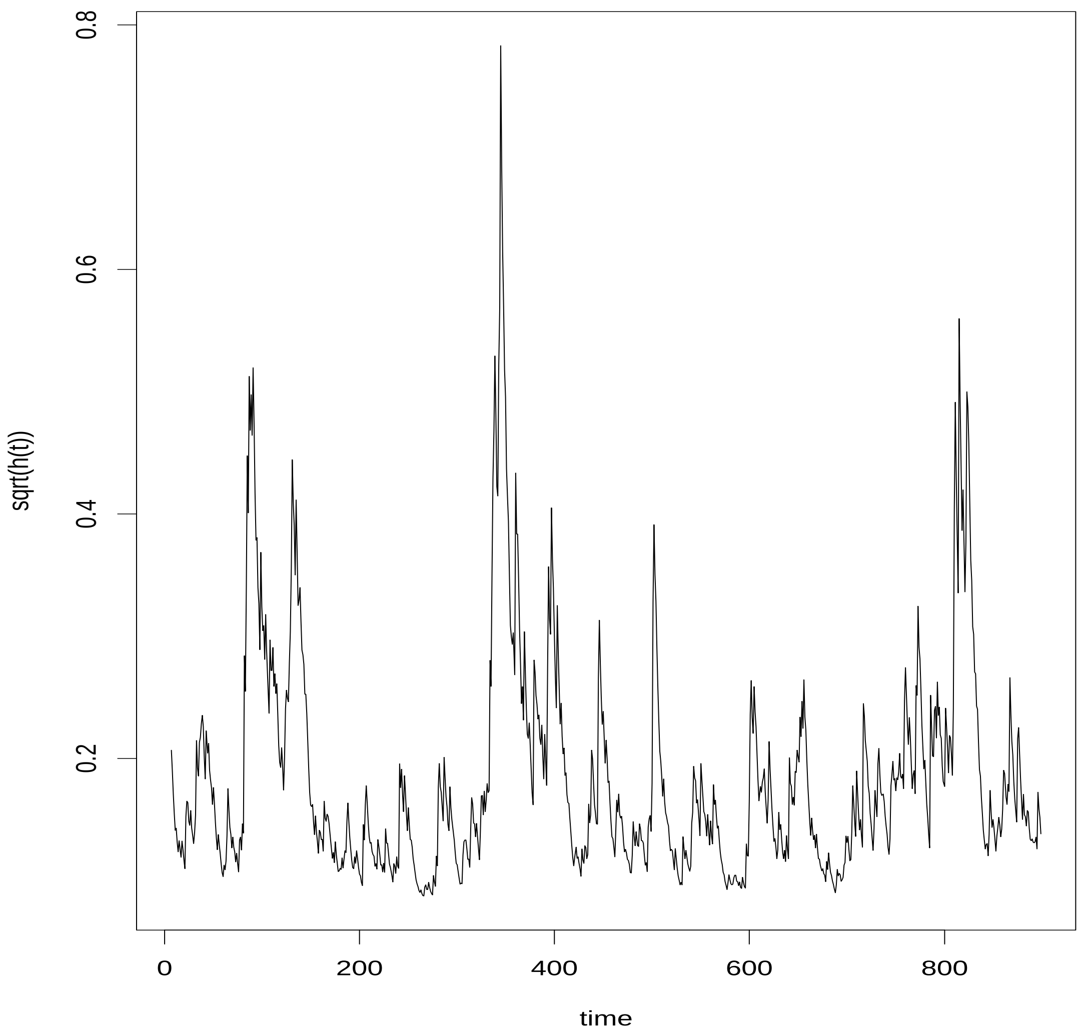


Figure 3.24: The estimated conditional standard errors($\sqrt{\hat{h}_t}$)

Chapter 4

Conclusion

We have investigated a set of autoregressive time series models whose coefficients have the form of a logistic function. The transfer function type models give additional flexibility over the fixed coefficients models and include them as a special case.

We have analyzed two stream flow series using the nonlinear model where the weight of the coefficients are determined by the logistic function. Our models include variants of

$$y_t = \gamma_1 \rho(x_{t-d}) x_{t-1} + \gamma_2 \left(1 - \rho(x_{t-d})\right) x_{t-2} + e_t$$

where $\rho(x_{t-d}) = \frac{1}{\exp(\alpha+\beta x_{t-d})+1}$ and $d = 1, 2 \cdots$, and

$$\begin{aligned} y_t &= \gamma_1[1 - \rho_1(x_{t-d}) - \rho_2(x_{t-d})]x_t + \gamma_2 \rho_1(x_{t-d}) x_{t-1} \\ &+ \gamma_3 \rho_2(x_{t-d}) x_{t-2} + e_t \end{aligned}$$

where $\rho_1(x_{t-d}) = \frac{\exp(\alpha_1+\beta_1 x_{t-d})}{\sum_{i=1}^{2} \exp(\alpha_i+\beta_i x_{t-d})+1}$ and $\rho_2(x_{t-d}) = \frac{\exp(\alpha_2+\beta_2 x_{t-d})}{\sum_{i=1}^{2} \exp(\alpha_i+\beta_i x_{t-d})+1}$. $d = 0, 1, 2, \cdots$.

The stream flow between Goldsoboro and Kinston North Carolina has been well fitted by the two weights model and that between Kinston and Fort Barnwell North

Carolina has been explained well by giving three weights to the coefficients of lagged variables. These nonlinear models add insights unavailable with the fixed coefficient model.

NLAR model with the AR(1) coefficient being a hyperbolic tangent function has been introduced.

$$y_t = \gamma\rho(y_{t-1})y_{t-1} + e_t$$

where $\rho(y_{t-1}) = \frac{\exp(\alpha+\beta f(y_{t-1}))-1}{\exp(\alpha+\beta f(y_{t-1}))+1}$. $f(y) = |y|$ or $f(y) = y$.

They work well for modeling series which have asymmetric stochastic volatility or changing amplitude around 0 with a persistent autocorrelation and local nonstationary behavior.

Geometric ergodicity of the series, and the consistency and the asymptotic normality of the parameter estimates have been established for the series generated with $|\gamma| < 1$. Where $\gamma = 1$, it appears that no single distribution applies, even for large samples, across the full range of possible (α, β) values. we have shown that similar distributional results to those of parameter estimates with $|\gamma|$ near but less than 1 are obtained using only those series which reject a unit root using a standard test, and we have found a region in which the normal approximation works reasonably well by conducting a Monte Carlo study.

In addition, we have found that estimating the other parameters with γ set to near 1 gives good prediction mean square errors, even if the true γ is not near 1, implying that the fitting is fairly robust with respect to the assumed γ.

We have applied the model to the analysis of the stream flow series in Goldsoboro and Kinston North Carolina and the soybean price data in North Carolina. It has been found that their performance is better compared to the standard ARMA model and the slowly decaying autocorrelation is well modeled.

In particular, in the case of the soybean price series, we fit a model in which a change of the conditional innovation variance(GARCH) as well as the dynamically changing difference equation coefficients(NLAR) are incorporated. We call this the NLAR-GARCH model.

A previously introduced model, the STAR model, could be misspecified and over-parameterized in its estimation stage using the usual STAR fitting process in the analysis of a series which has a rather persistent autocorrelation or ARCH type data structure. Our suggested NLAR could be one alternative.

Finally, a unit root test such as the ADF test still can be applied to check stationarity of the nonlinearly generated time series. For series that reject unit roots, we get well-behaved asymptotic distributions of parameter estimates.

References

An, H. Z. and Huang, F. C. (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica*, **6**, 943-56.

Bacon, D. W. and Watts, D. G. (1971). Estimating the Transition between two intersecting straight lines. *Biometrika*, **58**, 525-34.

Box, G. E., Jenkins, G. M. and Reinsel, G. C. (1994). *Time series analysis.* 3rd edition. Prentice Hall.

Brocklebank, J. C. and Dickey, D. A. (2003). *SAS system for forecasting time series.* SAS institute inc.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods.* Springer.

Brown, B. and Mariano, R. (1989). Predictors in dynamic-nonlinear models: large sample behaviour. *Economic Theory*, **5**, 430-52.

Chan, F. and McAleer, M. (2002). Maximum likelihood estimation of STAR and STAR-GARCH models: theory and Monte Carlo evidence. *Journal of Applied Econommetrics*, **17**, 509-34.

Chan, F. and McAleer, M. (2003). On the structure, asymptotic theory and application of STAR-GARCH models. CIRJE-F-216.

Chan, K. S., Petruccelli, J. D., Tong, H. and Woolford, S. W. (1985). A multiple-threshold AR(1) model. *J. Appl. Probab.*, **22**, 267-79.

Chan, K. S. and Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Prob.*, **17**, 666-78.

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, **7**, 178-90.

Chan, K. S. and Tong, H. (1994). A note on noisy chaos. *Journal of the Royal Statistical Society. Series B(Methodological)*, **56**, 301-11.

Chen, R. and Tsay, R. S. (1991). On the ergodicity of TAR(1) process. *Annals of Applied Probabability*, **1**, 613-634.

Cline, D. B. H. and Pu, H. H. (1999). Geometric ergodicity of nonlinear time series. *Statistica Sinica*, **9**, 1103-118.

Davidson, J. (1994). *An introduction for econometricians.* Oxford University Press.

Eitrheim, Φ. and Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, **74**, 59-75.

Enders, W. and Granger, C. W. J. (1998). Unit-root tests and asymptotic adjustment with an example using the term structure of interest rates. *Journal of Business Economic Statistics*, **16**, 304-11.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987-1007.

Escribano, A. and Jordá, Φ. (1999). Improved testing and specification of smooth transition regression models. In *Nonlinear Time Series Analysis of Economic and Financial Data*. Rothman, P., Ed.;Kluwer: Boston.

Franses, P. H. and van Dijk, D. (2000). *Non-linear time series models in empirical finance*. Cambridge univeristy press.

Fuller, W. A. (1996). *Introduction to statistical time series*. 2nd edition. Wiley-interscience.

Gallant, A. R. (1986). *Nonlinear statistical models*. Wiley-interscience.

Gallant, A. R. and Goebel, J. J. (1976). Nonlinear regression with autocorrelated errors. *Journal of the American Statistical Association*, **71**, 961-7.

Granger, C. W. J. and Teräsvirta, T. (1993). *Modeling nonlinear economic relationships*. Oxford university press.

Guo, M. and Petruccelli, J. D. (1991). On the null recurrence and transience of a first-order SETAR model. *J. Appl. Probab*, **28**, 584-92.

Hall, A. D., Skalin, J. and Teräsvirta, T. (2001). A nonlinear time series model of El Niño. *Environmental Modelling & Software*, **16**, 139-46.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, **40**, 633-43.

Jones , D. A. (1976). Non-linear autoregressive processes. Unpublished PhD Thesis, University of London.

Medeiros, M. C. and Veiga, A. (2000). *A flexible coefficient smooth transition time series model.* Working Paper Series in Economics and Finance No. 360, Stockholm School of Economics, 2000.

Meyn, S. P. and Tweedie, R. L. (1994). State-dependent criteria for convergence of Markov chains. *The Annals of Applied Probability*, **4**, 149-68.

Lin, J. L. and Granger, C. W. J. (1994). Forecasting from non-linear models in practice. *Journal of Forecasting*, **13**, 1-9.

Lundbergh, S. and Teräsvirta, T. (1998). Modelling economic high-frequency time series with STAR-STGARCH models. Stockholm School of Economics.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, **75**, 491-9.

Ozaki, T. and Oda, H. (1978). Non-linear time series model identification by Akaike's information criterion. *Proc. IFAC workshop on Information and Systems, Compiegn, France, Ocotber 1977.*

Petruccelli, J. D. and Woolford, S. W. (1984). A threshold AR(1) model. *Journal of Applied Probability*, **21**, 270-85.

Seber, G. A. F. and Wild, C. J. (1988). *Nonlinear regression.* John Wiley & Sons.

Shumway, R. H. and Stoffer, D. S. (1999). *Time series analysis and its applications.*

Springer.

Stout, W. F. (1974). *Almost sure convergence.* Academic Press.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series.* Springer.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**, 208-18.

Tjϕstheim, D. (1986). Estimation in nonlinear time series model. *Stochastic Processes and their Applications*, **21**, 251-73.

Tjϕstheim, D. (1990). Non-linear time series and Markov chains. *Journal of Applied Probability*, **22**, 587-611.

Tong, H. (1990). *Non-linear time series: A dynamical system approach.* Oxford science publications.

Tweedie, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Processes and their Applications*, **3**, 385-403.

van Dijk, D. and Franses, P. H. (1999). Modeling multiple regimes in the business cycle. *Macroeconomic Dynamics*, **3**, 311-40.

van Dijk, D. and Franses, P. H. (2000). A nonlinear long memory model for US unemployment. *Economic institute research report*.

van Dijk, D., Teräsvirta, T. and Franses, P. H. (2002). Smooth transition autoregressive models-a survey of recent developments. *Econometric Reviews*, **21**, 1-47.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817-38.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1-25.