Causality for CHARN models

Xiaoling Dou

Received April 26,2021 ; revised May 26,2021

Abstract. In this study, we consider Granger causality with a highly flexible nonlinear time series model, the conditional heteroscedastic autoregressive nonlinear (CHARN) model. We show that the causality of the CHARN models can be examined by a Portmanteau test based on a constrained maximum likelihood estimator of the parameters, and the test statistic has an approximate asymptotic Chi-square distribution. We describe the Chi-square asymptotics of the Portmanteau test for a CHARN model, provide calculations of the test statistic and investigate the performance of the Portmanteau test using a simulation. This idea is also illustrated using a real data set.

1 Introduction Causality is a relationship between a cause and an effect. The cause is considered to occur not later than the effect and it can help in predictions of the effect. Granger causality, defined by [8], is not necessarily a true causality, but a contributory factor in prediction. That is, for two random variables, X and Y, Granger causality does not clarify whether X causes Y, but focuses on whether X forecasts Y.

Granger causality was proposed in a vector autoregressive (VAR) processes, that is, a linear combination form of random vectors of stationary time series. A standard way to examine Granger causality is the Wald test for the coefficients of VAR model with a limiting χ^2-distribution ([14]). It tests whether the coefficients of the elements from distinct sequences in the VAR system are zero or not. Since the asymptotic χ^2 distribution is often a poor approximation when sample size is small, an F-version of the Wald test is often used instead. The test statistic is obtained by dividing the χ^2-statistic by its degrees of freedom, and is considered from an F-distribution. Likelihood ratio test, the Lagrange multiplier test ([16]) and the other test methods for Granger causality are discussed and compared in [7]. These classical tests give pairwise diagnoses for fixed time lag.

For multiple testing, Portmanteau test is popular. It can test overall significance of the serial correlations over various time lags. Portmanteau test was first proposed by Box and Pierce [2] for model diagnostics of autoregressive and moving average processes. For an autoregressive moving average model of order (p, q), ARMA(p, q), the Box and Pierce test statistic is defined as $n\sum_{k=1}^{h}\hat{r}_k^2$, where n is the sample size, $\hat{r}_k$ is the residual empirical autocorrelation at lag k. For moderately large n and h, the Box-Pierce test statistic is considered approximately χ^2 distributed with degrees of freedom $h-p-q$. A modified version of the Box-Pierce test, i.e. Ljung-Box test ([12]), substantially improves the approximation of the $\chi^2(h-p-q)$ distribution and is frequently applied in a variety of fields.

Many other modifications have been suggested. Among them, Taniguchi and Amano [17] pointed out that both the Box-Pierce statistic and the Ljung-Box statistic never converge to $\chi^2(h-p-q)$ for finite h. Instead, they proposed a modified Whittle likelihood ratio test which is asymptotically chi-square distributed for any finite h under ARMA(p, q) models and Bloomfield's exponential spectral density assumption. Recently, Chen and Lee [4] developed a Bayesian procedure for Granger causality test based on the generalized auto-regressive conditional heteroscedasticity (GARCH) type of integer-valued models and applied it to

2010 *Mathematics Subject Classification.* Primary 37M10; Secondary 62M10.
Key words and phrases. Asymptotic Chi-square distribution, CHARN model, Constrained maximum likelihood estimator, Nonlinear Granger causality, Portmanteau test.

testing causality relationships between temperature and crime data, as well as air pollution and human influenza data ([5]). Moreover, Akashi et al [1] proposed a new likelihood ratio based Portmanteau test which is applicable in more general situations. They also showed application examples for linear models.

Ideas for test of nonlinearity are also available. For example, Tsay [20] generalized Keenan's ([11]) Tukey nonadditivity-type test, improved its power, and proposed a test for concurrent nonlinearity as a diagnostic tool to examine the linearity assumption of time series models. Castle and Hendry [3] provided a Portmanteau test based on polynomial and exponential functions. It focuses on the nonlinearity in the conditional mean. Chen et al [6] proposed a hysteretic vector autoregressive (HVAR) model to test nonlinear Granger causality between two target time series and using posterior odds ratios for multiple testing.

To investigate Granger causality for nonlinear time series models, we consider a more flexible model, the conditional heteroscedastic autoregressive nonlinear (CHARN) model, where both the conditional mean and the residual are functions of the past. The CHARN model was introduced by [9] for financial data analysis. Because of its non-normality, nonlinearity and the blindingly general form, it has come into use in various fields of time series ([10], [19]).

We are interested in whether the Portmanteau test proposed by [1] can be used to detect the Granger causality for the CHARN model. In this paper, we examine nonlinear causality with this method. To show the feasibility and the performance of the method, we provide an example with the calculation of the test statistic for a specified CHARN model and conduct a simulation to confirm its capability for different sample sizes and different parameter settings of the CHARN model. We also demonstrate that the Portmanteau test can be used in practice if the the normality of the residuals of the CHARN model is satisfied.

The paper is organized as follows. Section 2 sets up the high dimensional stochastic process of the CHARN model, provides assumptions for stationarity of the process and requirements for the asymptotic optimal estimation theory of the parameters in the model, and formulates the nonlinear Portmanteau test for the CHARN model. Section 3 discusses the asymptotic distribution of the Portmanteau test and calculates the test statistic for a given CHARN model. In Section 4, we investigate the performance of the test by simulation. Finally, in Section 5, supposing that data follow CHARN models, we test whether the infection number of COVID-19 in Tokyo Granger causes the infection numbers in two of surrounding prefectures of Tokyo in Japan.

2 Assumptions and the Portmanteau Test

Let

$$\boldsymbol{X}(t) = \begin{pmatrix} \boldsymbol{X}_1(t) \\ \boldsymbol{X}_2(t) \end{pmatrix} = \begin{pmatrix} X_{1,1}(t) \\ \vdots \\ X_{1,m_1}(t) \\ X_{2,1}(t) \\ \vdots \\ X_{2,m_2}(t) \end{pmatrix} \tag{1}$$

be a $(m_1 + m_2 =)m$-dimensional stochastic process generated by

$$\boldsymbol{X}(t) = \boldsymbol{F_\theta}\{\boldsymbol{X}(t-1), \dots, \boldsymbol{X}(t-p)\} + \boldsymbol{H_\theta}\{\boldsymbol{X}(t-1), \dots, \boldsymbol{X}(t-q)\}\, \boldsymbol{U}(t), \tag{2}$$

where $\boldsymbol{F_\theta} : \mathbb{R}^{mp} \to \mathbb{R}^m$ is a vector-valued measurable function, $\boldsymbol{H_\theta} : \mathbb{R}^{mq} \to \mathbb{R}^{m\times m}$ is a positive definite matrix-valued measurable function, and $\boldsymbol{U}(t) = (\boldsymbol{U}_1(t), \boldsymbol{U}_2(t))'$ combining an m_1-vector $\boldsymbol{U}_1(t)$ and an m_2-vector $\boldsymbol{U}_2(t)$, is a sequence of m i.i.d. random variables with $E\{\boldsymbol{U}(t)\} = \boldsymbol{0}$, $E|\boldsymbol{U}(t)| < \infty$, and $\boldsymbol{U}(t)$ is independent of $\{\boldsymbol{X}(s),\ s < t\}$. Here, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)' \in \Theta \subset \mathbb{R}^r$ is a vector of unknown parameters.

We write $\boldsymbol{x} = (x_{11}, \dots, x_{1m}, x_{21}, \dots, x_{2m}, \dots, x_{p1}, \dots, x_{pm})'$ as an (mp)-vector, and $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2)' = (u_1, u_2, \dots, u_{m_1}, u_{m_1+1}, \dots, u_{m2})'$ an m-vector. From now on, without loss of generality, we assume $p = q$ and make the following assumptions.

Assumption 1 ([13]) (A.1) *$\boldsymbol{U}(t)$ has a probability density $p(\boldsymbol{u}) > 0$ on $\mathbb{R}^m$.*

(A.2) *There exist constants $a_{ij} \geq 0,\ b_{ij} \geq 0,\ 1 \leq i \leq p,\ 1 \leq j \leq m$, such that as $|\boldsymbol{x}| \to \infty$,*

$$|\boldsymbol{F}_{\boldsymbol{\theta}}(\boldsymbol{x})| \leq \sum_{i=1}^{p} \sum_{j=1}^{m} a_{ij} |x_{ij}| + o(|\boldsymbol{x}|),$$
$$|\boldsymbol{H}_{\boldsymbol{\theta}}(\boldsymbol{x})| \leq \sum_{i=1}^{p} \sum_{j=1}^{m} b_{ij} |x_{ij}| + o(|\boldsymbol{x}|),$$

where $|\boldsymbol{A}|$ denotes the sum of the absolute values of all entries of $\boldsymbol{A}$.

(A.3) *$\boldsymbol{H}_{\boldsymbol{\theta}}(\boldsymbol{x})$ is continuous and symmetric on $\mathbb{R}^{mp}$, and there exists a positive constant λ such that*

$$\lambda_m\{\boldsymbol{H}_{\boldsymbol{\theta}}(\boldsymbol{x})\} \geq \lambda \qquad \text{for all } \boldsymbol{x} \in \mathbb{R}^{mp},$$

where $\lambda_m\{(\cdot)\}$ is the minimum eigenvalue of $(\cdot)$.

(A.4)

$$\max_{1 \leq j \leq m} \left\{ \sum_{i=1}^{p} a_{ij} + E|\boldsymbol{U}(t)| \sum_{i=1}^{p} b_{ij} \right\} < 1.$$

Assumption 1 guarantees that $\{X_t\}$ is strictly stationary.

Assumption 2 ([18]) (B.1)

$$E_{\boldsymbol{\theta}} ||\boldsymbol{F}_{\boldsymbol{\theta}}(\boldsymbol{X}(t-1), \dots, \boldsymbol{X}(t-p))||^2 < \infty,$$
$$E_{\boldsymbol{\theta}} ||\boldsymbol{H}_{\boldsymbol{\theta}}(\boldsymbol{X}(t-1), \dots, \boldsymbol{X}(t-p))||^2 < \infty, \quad \text{for all } \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

where $||\boldsymbol{A}||$ indicates the Euclidian norm of a vector $\boldsymbol{A}$ or a matrix $\boldsymbol{A}$.

(B.2) *There exists $c > 0$ such that*

$$c \leq ||\boldsymbol{H}_{\boldsymbol{\theta}'}^{-1/2}(\boldsymbol{x}) \boldsymbol{H}_{\boldsymbol{\theta}}(\boldsymbol{x}) \boldsymbol{H}_{\boldsymbol{\theta}'}^{-1/2}(\boldsymbol{x})|| < \infty,$$

for all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$, and for all $\boldsymbol{x} \in \mathbb{R}^{mp}$.

(B.3) *$\boldsymbol{H}_{\boldsymbol{\theta}}$ and $\boldsymbol{F}_{\boldsymbol{\theta}}$ are continuously differentiable with respect to $\boldsymbol{\theta}$, and their derivatives $\partial_j \boldsymbol{H}_{\boldsymbol{\theta}}$ and $\partial_j \boldsymbol{F}_{\boldsymbol{\theta}}$ ($\partial_j = \partial / \partial \theta_j$), $j = 1, \dots, r$, satisfy the condition that there exist square-integrable functions A_j and B_j such that $||\partial_j \boldsymbol{H}_{\boldsymbol{\theta}}|| \leq A_j$, and $||\partial_j \boldsymbol{F}_{\boldsymbol{\theta}}|| \leq B_j$ ($j = 1, \dots, r$), for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.*

(B.4) *Density $p(\cdot)$ satisfies*

$$\lim_{||\boldsymbol{u}|| \to \infty} ||\boldsymbol{u}|| p(\boldsymbol{u}) = 0 \quad \text{and} \quad \int \boldsymbol{u}\boldsymbol{u}' p(\boldsymbol{u}) d\boldsymbol{u} = \boldsymbol{I}_m,$$

where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix.

(B.5) *The continuous derivative* $D_p = D_p(\boldsymbol{u}) \equiv \left(\frac{d}{du_1}p(\boldsymbol{u}), \ldots, \frac{d}{du_m}p(\boldsymbol{u})\right)'$ *exists on* $\mathbb{R}^m$ *and*

$$\int ||p^{-1}D_p||^4 p(\boldsymbol{u})d\boldsymbol{u} < \infty,$$
$$\int ||\boldsymbol{u}||^2 ||p^{-1}D_p||^2 p(\boldsymbol{u})d\boldsymbol{u} < \infty.$$

Assumption 2 is necessary in construction of the asymptotic optimal estimation theory for $\boldsymbol{\theta}$.

For given time series data in (1), to consider the Granger causality from $\boldsymbol{X}_2(t)$ to $\boldsymbol{X}_1(t)$, we focus on the prediction of $\boldsymbol{X}_1(t)$, $t = 1, \ldots, n$. Similar to (2), assume that $\boldsymbol{X}_1(t)$ is observed from the following CHARN model

$$\boldsymbol{X}_1(t) = \boldsymbol{F}_{1,\boldsymbol{\theta}}\{\boldsymbol{X}(t-1), \ldots, \boldsymbol{X}(t-p)\} + \boldsymbol{H}_{1,\boldsymbol{\theta}}\{\boldsymbol{X}(t-1), \ldots, \boldsymbol{X}(t-q)\}\, \boldsymbol{U}_1(t), \tag{3}$$

where $\boldsymbol{\theta}$ is a vector of unknown parameters, $\boldsymbol{F}_{1,\boldsymbol{\theta}} : \mathbb{R}^{mp} \to \mathbb{R}^{m_1}$ is a vector-valued measurable function, $\boldsymbol{H}_{1,\boldsymbol{\theta}} : \mathbb{R}^{mq} \to \mathbb{R}^{m_1 \times m_1}$ is a positive definite matrix-valued measurable function, and the model satisfies Assumptions 1 and 2 for $\boldsymbol{U}_1(t)$, $\boldsymbol{F}_{1,\boldsymbol{\theta}}$ and $\boldsymbol{H}_{1,\boldsymbol{\theta}}$ instead of $\boldsymbol{U}(t)$, $\boldsymbol{F}_{\boldsymbol{\theta}}$ and $\boldsymbol{H}_{\boldsymbol{\theta}}$, respectively.

Suppose that the dynamic part of (3) can be expressed as

$$\begin{aligned}
\boldsymbol{F}_{1,\boldsymbol{\theta}} &= \boldsymbol{F}_{1,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}\{\boldsymbol{X}(t-1), \ldots, \boldsymbol{X}(t-p)\} \\
&= \boldsymbol{F}_{1,\boldsymbol{\theta}_1}\{\boldsymbol{X}_1(t-1), \ldots, \boldsymbol{X}_1(t-p)\} + \boldsymbol{\theta}^*_{r_1+1}\boldsymbol{X}_2(t-1) + \ldots + \boldsymbol{\theta}^*_{r_1+r_1'}\boldsymbol{X}_2(t-p) \\
&\quad + \sum_{\ell=1}^{p} \exp\left\{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{X}_1(t-\ell)\boldsymbol{X}_1(t-\ell)'\}\right\}\boldsymbol{\theta}^*_{r_1+r_1'+\ell}\boldsymbol{X}_2(t-\ell),
\end{aligned} \tag{4}$$

where $\boldsymbol{F}_{1,\boldsymbol{\theta}}$ is the prediction of $\boldsymbol{X}_1(t)$ using information of both $\boldsymbol{X}_1(t-\ell)$ and $\boldsymbol{X}_2(t-\ell)$, $\boldsymbol{F}_{1,\boldsymbol{\theta}_1}$ is the prediction of $\boldsymbol{X}_1(t)$ with only information of $\boldsymbol{X}_1(t-\ell)$, for $\ell = 1, \ldots, p$, and the subscript of $\boldsymbol{\theta}^*_{r_1+r_1'+\ell}$ changes from $(r_1 + r_1' + 1)$ to $(r_1 + r_1' + p) = (r_1 + r_2)$.

In the vector of the unknown parameters

$$\boldsymbol{\theta} = (\mathrm{vec}(\boldsymbol{\theta}_1), \boldsymbol{\theta}_2)' = \left(\mathrm{vec}(\boldsymbol{\theta}_1), \mathrm{vec}(\boldsymbol{\theta}^*_{r_1+1}), \cdots, \mathrm{vec}(\boldsymbol{\theta}^*_{r_1+r_1'}), \cdots, \mathrm{vec}(\boldsymbol{\theta}^*_{r_1+r_2})\right)',$$

$\boldsymbol{\theta}_1$ is an $m_1 \times r_1$ matrix in function $\boldsymbol{F}_{1,\boldsymbol{\theta}_1}$ of $\boldsymbol{X}_1(t-\ell)$; $\boldsymbol{\theta}^*_{r_1+1}, \ldots, \boldsymbol{\theta}^*_{r_1+r_1'}, \ldots,$ and $\boldsymbol{\theta}^*_{r_1+r_2}$ are $m_1 \times m_2$ matrices for terms containing $\boldsymbol{X}_2(t-\ell)$, $\ell = 1, \ldots, p$.

Then the prediction error of $\boldsymbol{X}_1(t)$ by $\boldsymbol{F}_{1,\boldsymbol{\theta}_1}$ becomes

$$P_1 = E[\mathrm{tr}\{(\boldsymbol{X}_1(t) - \boldsymbol{F}_{1,\boldsymbol{\theta}_1})(\boldsymbol{X}_1(t) - \boldsymbol{F}_{1,\boldsymbol{\theta}_1})'\}]$$

and that by $\boldsymbol{F}_{1,\boldsymbol{\theta}}$ is

$$P_2 = E[\mathrm{tr}\{(\boldsymbol{X}_1(t) - \boldsymbol{F}_{1,\boldsymbol{\theta}})(\boldsymbol{X}_1(t) - \boldsymbol{F}_{1,\boldsymbol{\theta}})'\}].$$

Letting $V \equiv P_1 - P_2$, we can introduce a nonlinear causality from $\{\boldsymbol{X}_2(t)\}$ to $\{\boldsymbol{X}_1(t)\}$ by V, i.e., if $V = 0$, then we say that $\{\boldsymbol{X}_2(t)\}$ does not cause $\{\boldsymbol{X}_1(t)\}$ in our CHARN setting (for short, $\boldsymbol{X}_2(t) \nrightarrow \boldsymbol{X}_1(t)$). We can understand that $\boldsymbol{X}_2(t) \nrightarrow \boldsymbol{X}_1(t)$ is grasped by the testing problem:

$$H : \boldsymbol{\theta}_2 = \boldsymbol{0}, \quad v.s. \quad A : \boldsymbol{\theta}_2 \neq \boldsymbol{0}. \tag{5}$$

Let $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ be an observed stretch from (2), and let

$$\begin{aligned}\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= \log\{\text{likelihood function based on } \boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)\} \\ &= \sum_{t=p}^{n} \log \left[p\{\boldsymbol{H}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{X}(t) - \boldsymbol{F}_{\boldsymbol{\theta}})\}\{\det \boldsymbol{H}_{\boldsymbol{\theta}}\}^{-1} \right].\end{aligned}$$

In what follows we deal with the following marginal log-likelihood function:

$$\ell_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \sum_{t=p}^{n} \log \left[p_1 \left\{ \boldsymbol{H}_{1,\boldsymbol{\theta}}^{-1} \left(\boldsymbol{X}_1(t) - \boldsymbol{F}_{1,\boldsymbol{\theta}} \right) \right\} \left(\det \boldsymbol{H}_{1,\boldsymbol{\theta}} \right)^{-1} \right], \tag{6}$$

where $p_1(\cdot)$ is the marginal pdf of $\boldsymbol{U}_1(t)$.

Define

$$\widehat{\boldsymbol{\theta}}_1 = \arg \max_{\boldsymbol{\theta}_1} \ell_1(\boldsymbol{\theta}_1, \boldsymbol{0}), \quad \widehat{\boldsymbol{\theta}}_2 = \arg \max_{\boldsymbol{\theta}_2} \ell_1(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2).$$

For the problem of testing (5), we introduce the following test of Portmanteau type:

$$PT = 2[\ell_1(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2) - \ell_1(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{0})]. \tag{7}$$

The Fisher information matrix for the general model is given by

$$\begin{aligned}\mathcal{F}(\boldsymbol{\theta}) &\equiv \lim_{n\to\infty} \frac{1}{n} E \left[\frac{\partial}{\partial \boldsymbol{\theta}} \ell_1(\boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}'} \ell_1(\boldsymbol{\theta}) \right] \\ &= \begin{bmatrix} \boldsymbol{F}_{11} & \boldsymbol{F}_{12} \\ \boldsymbol{F}_{21} & \boldsymbol{F}_{22} \end{bmatrix}, \quad \text{(say)}.\end{aligned}$$

The following two lemmas follow from [1].

Lemma 1 *Under H,*

$$PT = \boldsymbol{N}_{R_2}' \boldsymbol{F}_{22\cdot 1}^{1/2} \boldsymbol{F}_{22}^{-1} \boldsymbol{F}_{22\cdot 1}^{1/2} \boldsymbol{N}_{R_2} + o_p(1),$$

where $\boldsymbol{N}_{R_2}$ is the $R_2(= m_1 m_2 r_2)$-dimensional standard normal random vector, and $\boldsymbol{F}_{22\cdot 1} = \boldsymbol{F}_{22} - \boldsymbol{F}_{21}\boldsymbol{F}_{11}^{-1}\boldsymbol{F}_{12}$.

Lemma 2 **(i)** *Let $R_1 = m_1 r_1$. If $R_1 < R_2$, $\boldsymbol{F}_{22} = \boldsymbol{I}_{R_2}$ and $\boldsymbol{F}_{21}\boldsymbol{F}_{11}^{-1}\boldsymbol{F}_{12}$ is idempotent with rank $\bar{r}$, then*

$$PT \xrightarrow{d} \chi^2_{R_2 - \bar{r}} \qquad \textit{under } H.$$

(ii) *If $\boldsymbol{F}_{22} \neq \boldsymbol{I}_{R_2}$ and $\boldsymbol{F}_{12} = \boldsymbol{0}$, then*

$$PT \xrightarrow{d} \chi^2_{R_2} \qquad \textit{under } H.$$

3 Asymptotic Distribution of Portmanteau Test In this section, we describe the χ^2-asymptotics of the Portmanteau test PT for the simplest case of (1), where $m_1 = m_2 = 1$, that is, $\boldsymbol{X}(t) = (X_1(t), X_2(t))'$.

Let

$$\begin{aligned}\boldsymbol{Z}(t) = \Big(& X_2(t-1), \ldots, X_2(t-p), \exp\left(-\frac{1}{2}X_1(t-1)^2\right) X_2(t-1), \ldots, \\ & \exp\left(-\frac{1}{2}X_1(t-p)^2\right) X_2(t-p) \Big)',\end{aligned}$$

whose dimension is r_2. Then we can write $\boldsymbol{F}_{1,\boldsymbol{\theta}}$ of (4) as

$$\begin{aligned}\boldsymbol{F}_{1,\boldsymbol{\theta}} &= \boldsymbol{F}_{1,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}\{\boldsymbol{X}(t-1),\dots,\boldsymbol{X}(t-p)\}\\ &= \boldsymbol{F}_{1,\boldsymbol{\theta}_1}\{X_1(t-1),\dots,X_1(t-p)\} + \boldsymbol{\theta}_2'\boldsymbol{Z}(t),\end{aligned}$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are r_1 and r_2-vectors, respectively.

If $p_1(\cdot)$ is Gaussian, it is not difficult to show

$$\boldsymbol{F}_{12} = \lim_{n\to\infty}\frac{-1}{n}E\left[\frac{\partial^2}{\partial\boldsymbol{\theta}_1\partial\boldsymbol{\theta}_2'}\ell_1(\boldsymbol{\theta})\right] = \boldsymbol{0}.$$

Then we have

Proposition 1 *If $p_1(\cdot)$ is a Gaussian probability density, under H,*

$$PT \xrightarrow{d} \chi^2_{(r_2)}.$$

Example 1. In (4), let

$$\begin{aligned}\boldsymbol{F}_{1,\boldsymbol{\theta}} &= \theta_1 X_1(t-1) + \theta_2\exp\left(-\frac{1}{2}X_1(t-1)^2\right) + \theta_3 X_2(t-1) + \theta_4\exp\left(-\frac{1}{2}X_1(t-1)^2\right)X_2(t-1)\\ &= \boldsymbol{Y}'(t-1)\boldsymbol{\theta},\end{aligned}$$

where

(8)

$$\begin{aligned}\boldsymbol{Y}(t-1) &:= (Y_1(t-1), Y_2(t-1), Y_3(t-1), Y_4(t-1))'\\ &:= \left(X_1(t-1),\ \exp\left(-\frac{1}{2}X_1(t-1)^2\right),\ X_2(t-1),\ \exp\left(-\frac{1}{2}X_1(t-1)^2\right)X_2(t-1)\right)',\end{aligned}$$

and $\boldsymbol{\theta} = (\theta_1,\theta_2,\theta_3,\theta_4)'$. In (6), let $\boldsymbol{H}_{1,\boldsymbol{\theta}} = \sqrt{0.1+\varepsilon X_1(t-1)^2+\delta X_2(t-1)^2} =: \sqrt{W(t-1)}$, where ε and δ are small positive values providing minor effects on the residual part of the CHARN model. Assume that $p_1(\cdot)$ is the pdf of $N(0,1)$. We see that $\boldsymbol{\theta}_1 = (\theta_1,\theta_2)'$, $\boldsymbol{\theta}_2 = (\theta_3,\theta_4)'$, $r_1 = 2$, $r_2 = 2$. Suppose that $\sum_{j=1}^4|\theta_j| < 1$. Then we can find PT in the following way.

Since

$$p_1\left\{\frac{X_1(t)-\boldsymbol{F}_{1,\boldsymbol{\theta}}}{\boldsymbol{H}_{1,\boldsymbol{\theta}}}\right\}\times\frac{1}{\boldsymbol{H}_{1,\boldsymbol{\theta}}} = \frac{1}{\sqrt{2\pi\times\{0.1+\varepsilon X_1(t-1)^2+\delta X_2(t-1)^2\}}}\exp\left\{-\frac{\left(X_1(t)-\boldsymbol{F}_{1,\boldsymbol{\theta}}\right)^2}{2\left(\boldsymbol{H}_{1,\boldsymbol{\theta}}\right)^2}\right\}$$

$$\begin{aligned}&=\frac{1}{\sqrt{2\pi\times\{0.1+\varepsilon X_1(t-1)^2+\delta X_2(t-1)^2\}}}\\ &\quad\exp\left(-\frac{[X_1(t)-\theta_1X_1(t-1)-\theta_2\exp\{-\frac{1}{2}X_1(t-1)^2\}-\theta_3X_2(t-1)-\theta_4\exp\{-\frac{1}{2}X_1(t-1)^2\}X_2(t-1)]^2}{2\times\{0.1+\varepsilon X_1(t-1)^2+\delta X_2(t-1)^2\}}\right),\end{aligned}$$

and

$$\begin{aligned}&\log\left[p_1\left\{\frac{X_1(t)-\boldsymbol{F}_{1,\boldsymbol{\theta}}}{\boldsymbol{H}_{1,\boldsymbol{\theta}}}\right\}\times\frac{1}{\boldsymbol{H}_{1,\boldsymbol{\theta}}}\right]\\ &= -\frac{1}{2}\log(2\pi\times\{0.1+\varepsilon X_1(t-1)^2+\delta X_2(t-1)^2\})\\ &\quad-\frac{[X_1(t)-\theta_1X_1(t-1)-\theta_2\exp\{-\frac{1}{2}X_1(t-1)^2\}-\theta_3X_2(t-1)-\theta_4\exp\{-\frac{1}{2}X_1(t-1)^2\}X_2(t-1)]^2}{2\times\{0.1+\varepsilon X_1(t-1)^2+\delta X_2(t-1)^2\}},\end{aligned}$$

Equ. (6) becomes

(9)

$$\ell_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \sum_{t=2}^{n} \log \left[p_1 \left\{ \frac{X_1(t) - \boldsymbol{F}_{1,\boldsymbol{\theta}}}{\boldsymbol{H}_{1,\boldsymbol{\theta}}} \right\} \times \frac{1}{\boldsymbol{H}_{1,\boldsymbol{\theta}}} \right]$$

$$= -\frac{1}{2} \sum_{t=2}^{n} \log(2\pi \times \{0.1 + \varepsilon X_1(t-1)^2 + \delta X_2(t-1)^2\})$$

$$- \sum_{t=2}^{n} \frac{[X_1(t) - \theta_1 X_1(t-1) - \theta_2 \exp\{-\frac{1}{2} X_1(t-1)^2\} - \theta_3 X_2(t-1) - \theta_4 \exp\{-\frac{1}{2} X_1(t-1)^2\} X_2(t-1)]^2}{2 \times \{0.1 + \varepsilon X_1(t-1)^2 + \delta X_2(t-1)^2\}}.$$

Under the null hypothesis H, the log-likelihood function becomes

$$\ell_1(\boldsymbol{\theta}_1, \mathbf{0}) = -\frac{1}{2} \sum_{t=2}^{n} \log(2\pi \times \{0.1 + \varepsilon X_1(t-1)^2 + \delta X_2(t-1)^2\}) - \sum_{t=2}^{n} \frac{[X_1(t) - \theta_1 X_1(t-1) - \theta_2 \exp\{-\frac{1}{2} X_1(t-1)^2\}]^2}{2 \times \{0.1 + \varepsilon X_1(t-1)^2 + \delta X_2(t-1)^2\}}.$$

Setting $\partial \ell_1 / \partial \theta_i = 0, \ i \in \{1, 2\}$, with notations $Y_1(t-1)$, $Y_2(t-1)$ and $W(t-1)$, we see that

$$\begin{pmatrix} \sum_{t=2}^{n} \{(Y_1(t-1), Y_2(t-1)) Y_1(t-1)\} / W(t-1) \\ \sum_{t=2}^{n} \{(Y_1(t-1), Y_2(t-1)) Y_2(t-1)\} / W(t-1) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$
$$= \begin{pmatrix} \sum_{t=2}^{n} \{(Y_1(t-1), Y_2(t-1)) Y_1(t-1)\} / W(t-1) \\ \sum_{t=2}^{n} \{(Y_1(t-1), Y_2(t-1)) Y_2(t-1)\} / W(t-1) \end{pmatrix} \boldsymbol{\theta}_1$$
$$= \begin{pmatrix} \sum_{t=2}^{n} \{X_1(t) Y_1(t-1)\} / W(t-1) \\ \sum_{t=2}^{n} \{X_1(t) Y_2(t-1)\} / W(t-1) \end{pmatrix}$$

and obtain

$$\widehat{\boldsymbol{\theta}}_1 = \begin{pmatrix} \sum_{t=2}^{n} \{(Y_1(t-1), Y_2(t-1)) Y_1(t-1)\} / W(t-1) \\ \sum_{t=2}^{n} \{(Y_1(t-1), Y_2(t-1)) Y_2(t-1)\} / W(t-1) \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=2}^{n} \{X_1(t) Y_1(t-1)\} / W(t-1) \\ \sum_{t=2}^{n} \{X_1(t) Y_2(t-1)\} / W(t-1) \end{pmatrix}.$$

Substitute $\widehat{\boldsymbol{\theta}}_1 = (\hat{\theta}_1, \hat{\theta}_2)'$ into Equ. (9) and let $Y_0(t) := X_1(t) - \hat{\theta}_1 X_1(t-1) - \hat{\theta}_2 \exp\{-\frac{1}{2} X_1(t-1)^2\}$, $t = 2, 3, \ldots$, we maximize

$$\ell_1(\widehat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2) = -\frac{1}{2} \sum_{t=2}^{n} \log(2\pi \times \{0.1 + \varepsilon X_1(t-1)^2 + \delta X_2(t-1)^2\})$$

$$- \sum_{t=2}^{n} \frac{[X_1(t) - \hat{\theta}_1 X_1(t-1) - \hat{\theta}_2 \exp\{-\frac{1}{2} X_1(t-1)^2\} - \theta_3 X_2(t-1) - \theta_4 \exp\{-\frac{1}{2} X_1(t-1)^2\} X_2(t-1)]^2}{2 \times \{0.1 + \varepsilon X_1(t-1)^2 + \delta X_2(t-1)^2\}}$$

$$= -\frac{1}{2} \sum_{t=2}^{n} \log(2\pi \times \{0.1 + \varepsilon X_1(t-1)^2 + \delta X_2(t-1)^2\}) - \sum_{t=2}^{n} \frac{[Y_0(t) - \theta_3 X_2(t-1) - \theta_4 \exp\{-\frac{1}{2} X_1(t-1)^2\} X_2(t-1)]^2}{2 \times \{0.1 + \varepsilon X_1(t-1)^2 + \delta X_2(t-1)^2\}}.$$

Set $\partial \ell_1 / \partial \theta_i = 0, \ i \in \{3, 4\}$, we have

$$\begin{pmatrix} \sum_{t=2}^{n} [\{(Y_3(t-1), Y_4(t-1)) Y_3(t-1)\} / W(t-1)] \\ \sum_{t=2}^{n} [\{(Y_3(t-1), Y_4(t-1)) Y_4(t-1)\} / W(t-1)] \end{pmatrix} \begin{pmatrix} \theta_3 \\ \theta_4 \end{pmatrix} = \begin{pmatrix} \sum_{t=2}^{n} [Y_0(t) Y_3(t-1) / W(t-1)] \\ \sum_{t=2}^{n} [Y_0(t) Y_4(t-1) / W(t-1)] \end{pmatrix},$$

then $\boldsymbol{\theta}_2$ can be estimated as follows

$$\widehat{\boldsymbol{\theta}}_2 = \begin{pmatrix} \hat{\theta}_3 \\ \hat{\theta}_4 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{t=2}^{n} \left[\{(Y_3(t-1), Y_4(t-1))Y_3(t-1)\}/W(t-1)\right] \\ \sum_{t=2}^{n} \left[\{(Y_3(t-1), Y_4(t-1))Y_4(t-1)\}/W(t-1)\right] \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=2}^{n} \left[Y_0(t)Y_3(t-1)/W(t-1)\right] \\ \sum_{t=2}^{n} \left[Y_0(t)Y_4(t-1)/W(t-1)\right] \end{pmatrix}.$$

By substituting the obtained estimates $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$ into (7), we then can calculate the Portmanteau test statistic PT. That is, with $Y_0(t) = X_1(t) - (Y_1(t-1), Y_2(t-1))\,\widehat{\boldsymbol{\theta}}_1$,

$$\begin{aligned} PT = & -\sum_{t=2}^{n} \log(2\pi W(t-1)) - \frac{2}{2}\sum_{t=2}^{n} \frac{\left\{Y_0(t) - (Y_3(t-1), Y_4(t-1))\ \widehat{\boldsymbol{\theta}}_2\right\}^2}{W(t-1)} \\ & + \sum_{t=2}^{n} \log(2\pi W(t-1)) + \frac{2}{2}\sum_{t=2}^{n} \frac{Y_0^2(t)}{W(t-1)} \\ = & \sum_{t=2}^{n} \frac{Y_0^2(t)}{W(t-1)} - \sum_{t=2}^{n} \frac{\left\{Y_0(t) - (Y_3(t-1), Y_4(t-1))\ \widehat{\boldsymbol{\theta}}_2\right\}^2}{W(t-1)}. \end{aligned}$$

4 Simulation Study To evaluate the availability of the Portmanteau test for Granger causality, we carry out the following simulation. We generate data from the two dimensional stochastic process below in which $X_2(t)$ is AR(1) and $X_1(t)$ is a CHARN model:

$$\boldsymbol{X}(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}'(t-1)\boldsymbol{\theta} + \sqrt{0.1 + \varepsilon X_1(t-1)^2 + \delta X_2(t-1)^2}\, U_1(t) \\ \theta_{21} X_2(t-1) + U_2(t) \end{pmatrix},$$

where $\boldsymbol{Y}(t-1)$ is defined in (8), $\boldsymbol{\theta} = (\theta_{11}, \theta_{12}, \theta_{13}, \theta_{14})'$, and $U_i,\ i = 1, 2$ are i.i.d. $N(0, 1)$. For eight models, the parameters are set as in the table below. We generate data for each model with three different lengths, $n = 50, 300, 1000$. For each model and each length, 3000 replications are made. We then test the non-linear causality $H_0:\ \theta_{13} = \theta_{14} = 0$, and calculate the empirical rejection ratios for each situation. The last three columns in Table 1 present the empirical rejection ratios of non-causality $X_2(t) \nrightarrow X_1(t)$. Nominal significance level is 0.05.

Table 1: Empirical rejection ratio of the null hypothesis of non-causality for eight models

Model	ε	δ	$\boldsymbol{\theta}$	θ_{21}	$n = 50$	$n = 300$	$n = 1000$
i	0	0	$(0.1, 0, 0, 0)'$	0.2	0.035	0.049	0.048
ii	0.01	0	$(0.1, 0, 0, 0)'$	0.2	0.036	0.043	0.049
iii	0.01	0.01	$(0.1, 0, 0, 0)'$	0.2	0.036	0.045	0.055
iv	0.01	0.05	$(0.1, 0, 0, 0)'$	0.2	0.041	0.051	0.055
v	0.01	0	$(0.1, 0, 0.1, 0)'$	0.2	0.462	0.999	1.000
vi	0.01	0.01	$(0.1, 0, 0.1, 0)'$	0.2	0.362	0.995	1.000
vii	0.01	0.01	$(0.1, 0.1, 0.1, 0)'$	0.2	0.376	0.996	1.000
viii	0.01	0.01	$(0.1, 0.1, 0.1, 0.1)'$	0.2	0.916	1.000	1.000

From this result, we see that when the sample size is large or moderately large, the Portmanteau test works well, although there is a need to improve the power when the sample size is small. When there is no Granger causality (Models i – iv), the empirical rejection ratio is close to the significance level; when the Granger causality exists (Models v – viii), the empirical rejection ratio is close to one.

5 Data Analysis We examine Granger causality between the numbers of infected people with COVID-19 in Tokyo and its two neighboring prefectures. The data are taken from the website of NHK `https://www3.nhk.or.jp/news/special/coronavirus/data-widget/#mokuji1`. We focus on the data of Kanagawa Prefecture, Yamanashi Prefecture and the Tokyo metropolitan area, from January 16 to December 17, 2020. The three time series as well as their cross-autocorrelation functions (CCFs) are plotted in Figure 1. A clear seven day period can be seen from the original data and the CCFs.

Since the variance increases substantially when the number of infections grows and there are clear trends in the sequences, we set the zero values in the data set as 0.5, take logarithm for all the data and take the first difference to remove the trends. The detrended data and their CCFs are given in Figure 2. We also plot the autocorrelation functions (ACFs) and the partial autocorrelation functions (PACFs) for Kanagawa and Yamanashi Prefectures in Figure 3.

For the detrended data, we test whether the number of infections in Tokyo causes the numbers of infections in Kanagawa and Yamanashi in the Granger sense. We denote the detrended time series of Kanagawa, Yamanashi and Tokyo in Figure 2 as $\{X_1(t)\}, \{X_2(t)\}$ and $\{X_3(t)\}$, respectively. According to the CCFs of Kanagawa and Tokyo in Figure 2 and the ACFs and PACFs of Kanagawa in Figure 3, we try time lags $P = 7, 14, \ldots, 20$. We also prepare the nine possible terms in Table 2 for model selection for $p = 1, \ldots, P$. The use of $\exp\{-0.5(X_i(t-p))^2\}$ and $\exp\{-0.5(X_i(t-p))^2\}X_j(t),\ \ i, j \in \{1, 2, 3\}$, is due to Assumption (B.1), and the exponential function can moderate sharp fluctuations in the time series.

Table 2: Possible terms for model selection

$X_1(t-p)$	$X_2(t-p)$	$X_3(t-p)$
$\exp\{-0.5(X_1(t-p))^2\}$	$\exp\{-0.5(X_2(t-p))^2\}$	$\exp\{-0.5(X_3(t-p))^2\}$
$\exp\{-0.5(X_1(t-p))^2\}X_2(t)$	$\exp\{-0.5(X_1(t-p))^2\}X_3(t-p)$	$\exp\{-0.5(X_2(t-p))^2\}X_3(t-p)$

The data analysis below is carried out in a two-step procedure. We first use AIC to select models for $X_1(t)$ and $X_2(t)$, respectively; then for the models containing effects from $X_3(t)$, we do the Portmanteau test to examine the Granger causality. Since the Portmanteau test requires that the distribution of the residuals under H_0 be normal, we also take this into account in the model selection. That is, the selected model should minimize AIC, and if there are several models having similar small values of AIC, we choose the model whose residuals under H_0 are closest to the normal distribution.

For Kanagawa Prefecture, the following model with $P = 18$ is selected.

$$X_1(t) = \sum_{p=1}^{P} \alpha_p X_1(t-p) + \sum_{p=1}^{P} \beta_p \exp\{-0.5(X_1(t-p))^2\} + \sum_{p=1}^{P} \theta_p \exp\{-0.5(X_3(t-p))^2\} + \varepsilon_1(t), \tag{10}$$

where $\varepsilon_1(t)$ is assumed normal distributed $N(0, \sigma_1^2)$.

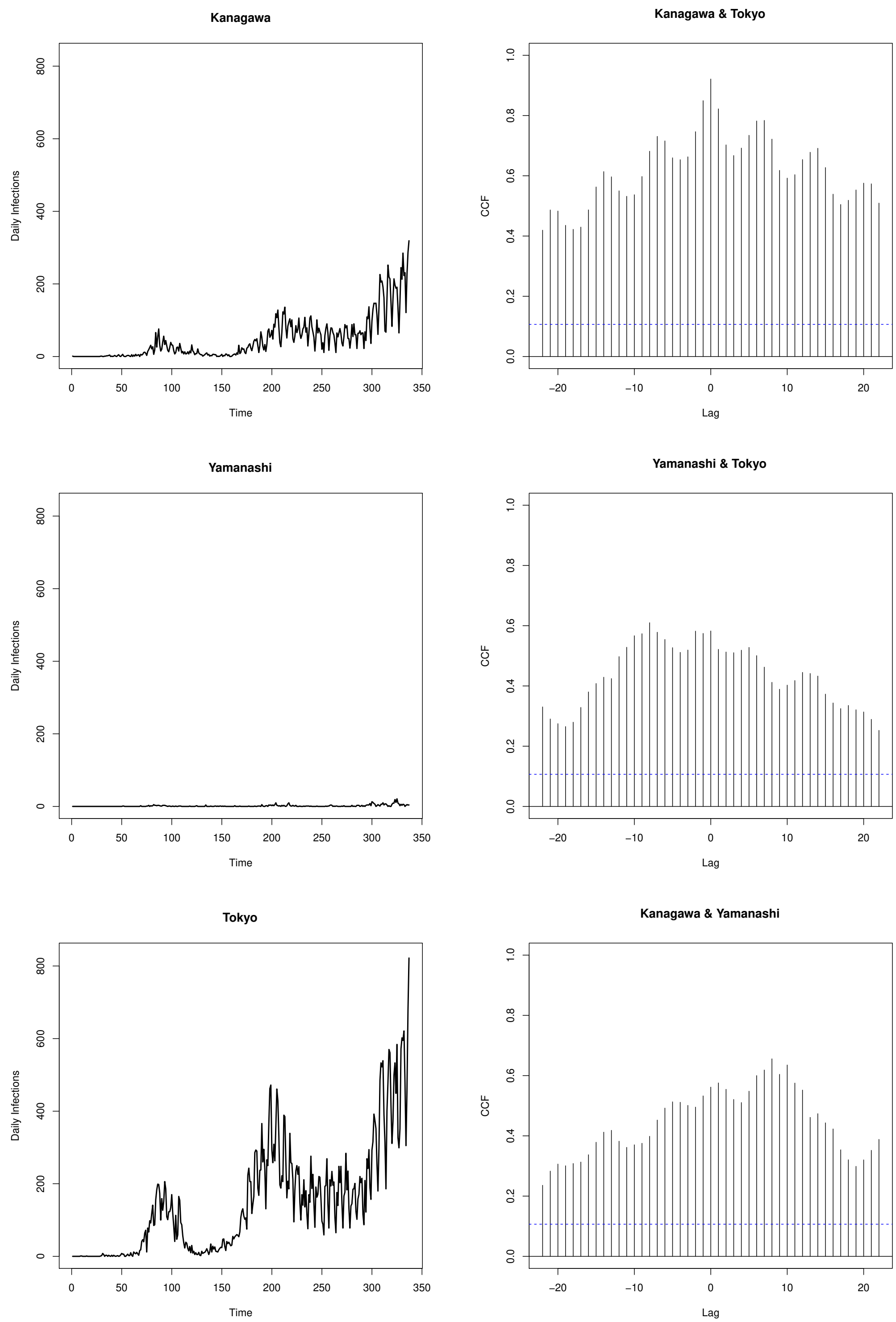

Figure 1: Left: Infection numbers of COVID-19 in Kanagawa, Yamanashi and Tokyo; Right: CCFs of Kanagawa, Yamanashi and Tokyo.

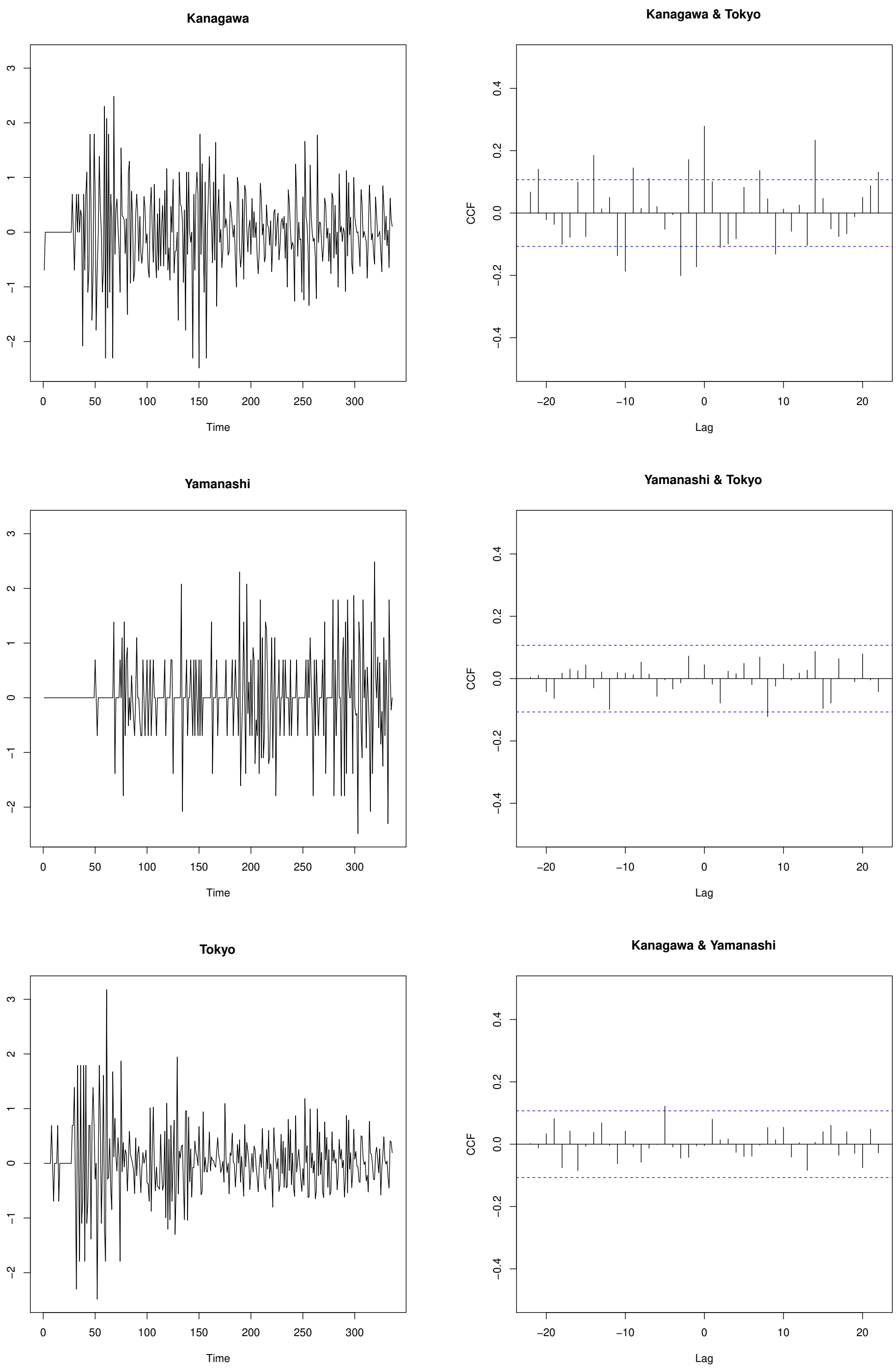

Figure 2: The detrended time series and their CCFs.

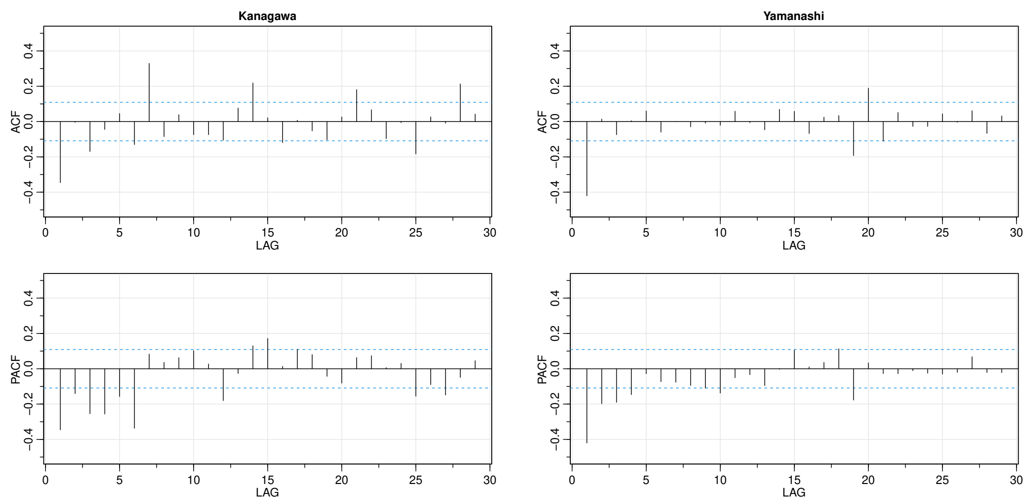

Figure 3: The ACFs and PACFs of Kanagawa and Yamanashi Prefectures.

Testing the Granger causality for this model is equivalent to test

$$H_0: \quad \theta_1 = \cdots = \theta_P = 0.$$

In the calculation of PT, we first compute $\widehat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_P)'$ and $\widehat{\boldsymbol{\beta}} = (\hat{\beta}_1, \cdots, \hat{\beta}_P)'$ under H_0, then substitute them into model (10) and find $\widehat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_P)'$ by maximizing the log-likelihood function.

The residuals from the model under the hypothesis have mean 0.012, standard deviation 0.546; the residuals obtained from the model under the alternative have mean 0.003, and standard deviation 0.502. The two sequences of residuals are shown in the first row of Figure 4. The ACFs and PACFs of the sequences of residuals in the second row of the figure show that there is almost no correlation in the sequences. Corresponding Q-Q plots of the standardized residuals are shown in the last row of Figure 4. We see that their distributions are close to the standard normal distribution.

As the value of the test statistic, $PT = 14.514$, is smaller than the critical point $\chi^2_{0.95}(18) = 28.869$, we cannot reject the hypothesis H_0 at an $\alpha = 0.05$ significance level, and cannot conclude that the number of infections in Kanagawa is Granger caused by the number of infections in Tokyo.

For Yamanashi Prefecture, we try time lags $P \in \{7, \ldots, 20\}$ for the model selection. From the possible terms given in Table 2 and their linear combinations, AIC selects the simple AR(P) model

$$X_2(t) = \sum_{p=1}^{P} \gamma_p X_2(t-p) + \varepsilon_2(t), \tag{11}$$

with $P = 19$. According to the ACFs and PACFs of Yamanashi in Figure 3, we see this model is suitable, because the ACF tails off and the PACF cuts off after lag 19.

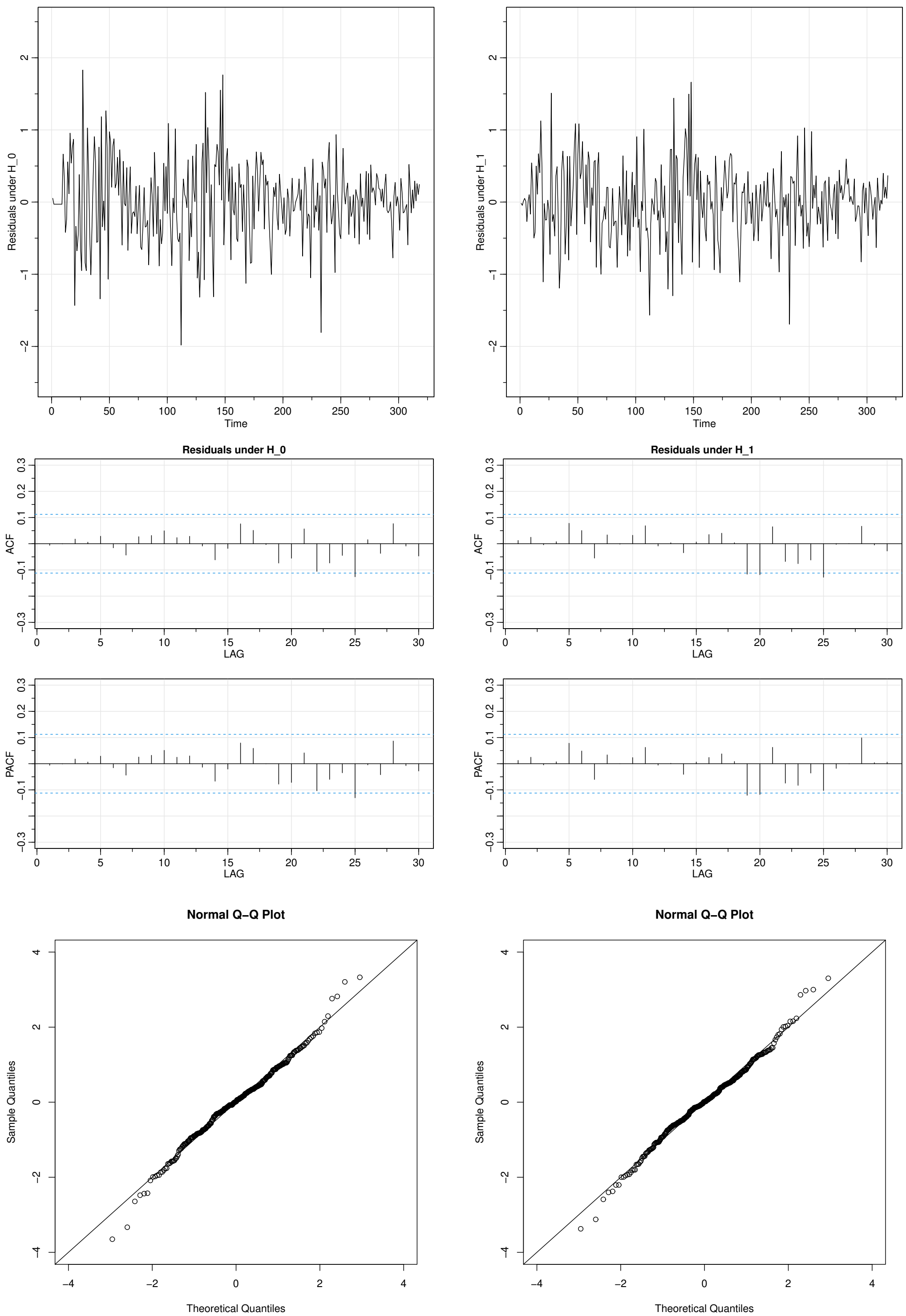

Figure 4: The left and right panels are the sequences of residuals, their ACFs, PACFs, and their Q-Q plots under H_0 and H_1, respectively.

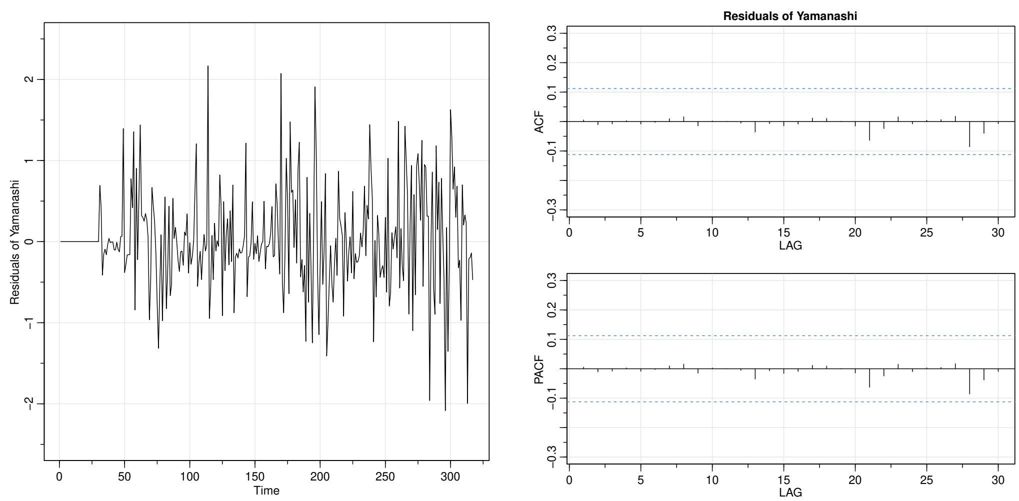

Figure 5: The sequence of residuals of Yamanashi Prefecture and the ACFs and PACFs.

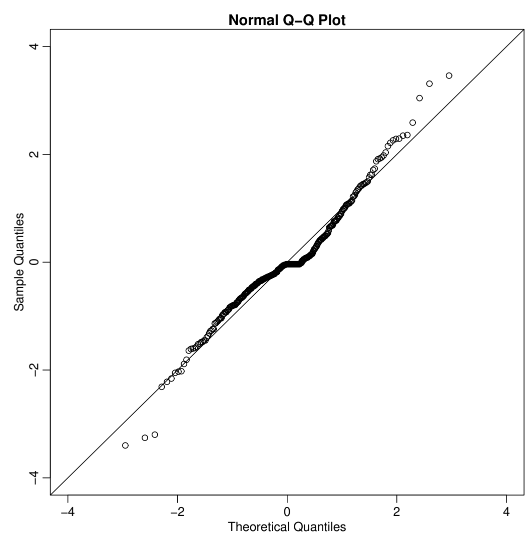

Figure 6: Q-Q plot of the standardized residuals of Yamanashi Prefecture.

The residuals of Yamanashi Prefecture obtained from model (11) have mean 0.022 and standard deviation 0.619. The sequence of the residuals, their ACFs and PACFs are plotted in Figure 5. This model does not contain any information of $X_3(t)$ and we cannot perform the Portmanteau test. However, since AIC can be used in Granger causality detection by the selection of the orders of bivariate autoregressive models when the sample size is large ([15]), we can conclude that there is no Granger causality of Tokyo to Yamanashi Prefecture.

For a simultaneous test of the Granger causality from Tokyo to both of the prefectures of Kanagawa and Yamanashi, a possible solution is to take $\boldsymbol{X}_1(t) = (X_1(t), X_2(t))'$ together as Kanagawa and Yamanashi, and investigate the effect of Tokyo $X_3(t)$ to $\boldsymbol{X}_1(t)$ using the

Portmanteau test. In the multivariate case, we also need a precondition that the distribution of the residuals under H_0 should be a two-dimensional normal distribution. However, the Q-Q plot of the standardized residuals of $X_2(t)$ obtained from (11) in Figure 6 shows a significant departure from the standard normal distribution. This means that the selected models (10) and (11) are not able to make the precondition satisfied and the Portmanteau test cannot be applied directly to this data set. Besides, in our models (10) and (11), different time lags P are used: $P = 18$ for $X_1(t)$, and $P = 19$ for $X_2(t)$. This results in different lengths of residuals of the two time series and makes it difficult to construct a two-dimensional normal distribution. Additionally, because of the different patterns of the ACFs and PACFs for Kanagawa and Yamanashi Prefectures in Figure 3, it is unlikely to obtain the same or similar models for $X_1(t)$ and $X_2(t)$ even if other model selection methods are used. For these reasons, we decide to cease the simultaneous Granger causality test for this data set.

Acknowledgements The author thanks the Editors and the editorial reviewer for their valuable time, careful comments and helpful suggestions, which have helped improve this paper. The author is grateful to Professor Masanobu Taniguchi of Waseda University for his kind help and valuable advice on this work.

References

[1] Akashi, F., Odashima, H., Taniguchi, M. and Monti, A.C. (2018) A new look at Portmanteau tests. *Sankhya* **80**, 121–137.

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

[3] Castle, J. L. and Hendry, D. F. (2010) A low-dimension portmanteau test for non-linearity. *Journal of Econometrics* **158**, 231–245.

[4] Chen, C. W. S. and Lee, S. (2017) Bayesian causality test for integer-valued time series models with applications to climate and crime data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **66**, 797–814.

[5] Chen, C. W. S., Hsieh, Y. H., Su, H. C. and Wu, J. J. (2018) Causality test of ambient fine particles and human influenza in Taiwan: Age group-specific disparity and geographic heterogeneity. *Environment International* **111**, 354–361.

[6] Chen, C. W. S., Than-Thi, H. and So, M. K. P. (2019) On hysteretic vector autoregressive model with applications. *Journal of Statistical Computation and Simulation* **89**, 191–210.

[7] Geweke, J., Meese, R. and Dent, W. (1983) Comparing alternative tests of causality in temporal systems: Analytic results and experimental evidence. *Journal of Econometrics* **21**, 161–194.

[8] Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438.

[9] Härdle, W., Tsybakov, A. and Yang, L. (1998) Nonparametric vector autoregression. *Journal of Statistical Planning and Inference* **68**, 221–245.

[10] Kato, H., Taniguchi, M. and Honda, M. (2006) Statistical analysis for multiplicatively modulated nonlinear autoregressive model and its applications to electrophysiological signal analysis in humans. *IEEE Transactions on Signal Processing* **54**, 3414–3425.

[11] Keenan, D. M. (1985) A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika* **72**, 39–44.

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

[13] Lu, Z. and Jiang, Z. (2001) L_1 geometric ergodicity of a nultivariate nonlinear AR model with an ARCH term. *Statistics & Probability Letters* **51**, 121–130.

[14] Lütkepohl, H. (2007) *New Introduction to Multiple Time Series Analysis*. Springer, New York.

[15] Odaki, M. (1987) Tests of Granger causality by selection of the orders of a bivariate autoregressive model. *Economics Letters* **22**, 223–227.

[16] Silvey, S. D. (1959) The lagrangian multiplier test. *Annals of Mathematical Statistics* **30**, 389–407.

[17] Taniguchi, M. and Amano, T. (2009) Systematic approach for portmanteau tests in view of whittle likelihood ratio. *Journal of the Japan Statistical Society* **39**, 177–192.

[18] Taniguchi, M., Hirukawa, J. and Tamaki, K. (2008) *Optimal Statistical Inference in Financial Engineering*. Chapman & Hall/CRC, New York.

[19] Taniguchi, M., Shiraishi, H., Hirukawa, J., Kato, H. S. and Yamashita, T. (2018) *Statistical Portfolio Estimation*. Chapman & Hall/CRC, New York.

[20] Tsay, R. S. (1986) Nonlinearity tests for time series. *Biometrika* **73**, 461–466.

Communicated by *Masanobu Taniguchi*