AN INVESTIGATION OF A GENERALIZED LEAST SQUARES ESTIMATOR FOR NON-LINEAR TIME SERIES MODEL

Xiaoling Dou

Received March 11, 2017 ; revised April 7, 2017

Abstract.

Ochi(1983) proposed an estimator for the autoregressive coefficient of the first-order autoregressive model (AR(1)) by using two constants for the end points of the process. Classical estimators for AR(1) , such as the least squares estimator, Burg's estimator, and Yule-Walker estimator are obtained as special cases by choice of the constants in Ochi's estimator. By writing the first-order autoregressive conditional heteroskedastic model, ARCH(1), in a form similar to that of AR(1), we extend Ochi 's estimator to ARCH(1) models. This allows introducing analogues of the least squares estimator, Burg's estimator and Yule-Walker estimator, and we compare the relations of these with Ochi's estimator for ARCH(1) models. We then provide a simulation for AR(1) models and examine the performance of Ochi 's estimator. Also, we simulate Ochi's estimator for ARCH(1) with different parameter values and sample sizes.

1 Introduction Let $\{x_1, \cdots, x_T\}$ be generated from the first order autoregressive process, AR(1),

$$x_t = \alpha x_{t-1} + \epsilon_t, \quad |\alpha| < 1, \quad \epsilon_t \sim N(0, \sigma^2), \quad t \in [2, \cdots, T] \tag{1}$$

with an unknown coefficient α, and independent and identically distributed (iid) errors ϵ_t. Ochi (1983) proposed an estimator of the autoregressive coefficient

$$\text{Ochi}(c_1, c_2) = \hat{\alpha}_{c_1,c_2} = \frac{\sum_{t=2}^{T} x_t x_{t-1}}{\sum_{t=2}^{T-1} x_t^2 + c_1 x_1^2 + c_2 x_T^2}, \tag{2}$$

where c_1 and c_2 are nonnegative constants, and can be considered as weights of the end points. It is also known that $\text{Ochi}(1, 0)$, $\text{Ochi}(0.5, 0.5)$, and $\text{Ochi}(1, 1)$ are the least squares estimator(LSE), Burg's estimator, and Yule-Walker estimator, respectively.

Recently, non-linear time series models have been increasing in popularity. Autoregressive conditional heteroskedastic (ARCH) models were proposed by Engle (1982). Chan and Tong (1986) and Tong (1990) introduced some threshold models, such as threshold

2010 *Mathematics Subject Classification.* subject classifications .

Key words and phrases. Burg's method, Least squares estimator, Mean Square Error, Ochi's estimator, Yule-Walker estimator.

autoregressive (TAR) models, self-exciting threshold (SETAR) models and smooth threshold autoregressive(STAR) models. Markov switching autoregressive(MAR) models were developed by Hamilton (1989). Davis et al. (2008) introduced some segmented time series. For more non-linear models, see Turkman et al. (2014). The ARCH process of order 1, ARCH(1) is one of the most famous and can be modeled as

$$y_t = \sqrt{\theta_0 + \theta_1 y_{t-1}^2}\, u_t, \quad u_t \sim \text{iid } N(0,1), \tag{3}$$

with parameters $\theta_0 > 0$ and $|\theta_1| < 1$. Rewriting (3), we have

$$\begin{aligned} y_t^2 &= (\theta_0 + \theta_1 y_{t-1}^2) u_t^2 \\ &= \theta_0 + \theta_1 y_{t-1}^2 + (\theta_0 + \theta_1 y_{t-1}^2)(u_t^2 - 1) \\ &= \theta_0 + \theta_1 y_{t-1}^2 + \xi_t, \end{aligned} \tag{4}$$

which has a form similar to that of AR(1) in (1). Here, $\xi_t := (\theta_0 + \theta_1 y_{t-1}^2)(u_t^2 - 1)$ is an uncorrelated process with mean 0 and variance

$$\begin{aligned} \text{Var}(\xi_t) &= E[(\theta_0 + \theta_1 y_{t-1}^2)^2 (u_t^2 - 1)^2] - \{E[(\theta_0 + \theta_1 y_{t-1}^2)(u_t^2 - 1)]\}^2 \\ &= E[(\theta_0 + \theta_1 y_{t-1}^2)^2] E[(u_t^2 - 1)^2] - \{\theta_0 + \theta_1 E[y_{t-1}^2]\}^2 \{E[u_t^2] - 1\}^2 \\ &= E[\theta_0^2 + \theta_1^2 y_{t-1}^4 + 2\theta_0\theta_1 y_{t-1}^2] E[u_t^4 + 1 - 2u_t^2] - 0 \\ &= 2\left(\theta_0^2 + \theta_1^2 E[y_{t-1}^4] + 2\theta_0\theta_1 E[y_{t-1}^2]\right). \end{aligned}$$

Since

$$E[y_t^2] = \theta_0 + \theta_1 E[y_{t-1}^2] + 0 = \frac{\theta_0}{1-\theta_1}, \quad E[y_t^2] = E[y_{t-1}^2],$$

$E[y_t^4] = E[y_{t-1}^4]$, $u_t^2 \sim \chi_1^2$, and $E[u_t^4] = 3$, we have

$$E[y_t^4] = E[(\theta_0 + \theta_1 y_{t-1}^2)^2]\, E[u_t^4] = 3(\theta_0^2 + \theta_1^2\, E[y_{t-1}^4] + 2\theta_0\theta_1 E[y_{t-1}^2]) = \frac{3\theta_0^2(1+\theta_1)}{(1-\theta_1)(1-3\theta_1^2)}.$$

Hence the expression for the variance of ξ_t can be simplified to

$$\text{Var}(\xi_t) = \frac{2\theta_0^2(1+\theta_1)}{(1-3\theta_1^2)(1-\theta_1)}, \tag{5}$$

and the variance of y_t^2 can be easily found

$$\text{Var}(y_t^2) = E[y_t^4] - (E[y_t^2])^2 = \frac{2\theta_0^2}{(1-3\theta_1^2)(1-\theta_1)^2}. \tag{6}$$

From variances (5) and (6), we see that $\theta_1 < \sqrt{1/3}$ is required. This is also discussed in Shumway and Stoffer (2011).

Suppose the process $\{y_1, \cdots, y_T\}$ is generated from (3). To estimate the parameters $\boldsymbol{\theta} = (\theta_0, \theta_1)'$ in the ARCH(1) process, we apply Ochi's estimator to the squared process (4). The main purpose of this paper is to investigate the performance of Ochi's estimator for (4)

$$\text{(7)} \qquad \text{Ochi}^*(c_1, c_2) = \hat{\boldsymbol{\theta}}_{c_1,c_2} = \left(\sum_{t=2}^{T-1} \boldsymbol{y}_t \boldsymbol{y}_t' + c_1 \boldsymbol{y}_1 \boldsymbol{y}_1' + c_2 \boldsymbol{y}_T \boldsymbol{y}_T' \right)^{-1} \sum_{t=1}^{T-1} \boldsymbol{y}_t y_{1+t}^2,$$

by simulation. In this, $c_1, \ c_2 \geq 0$, and

$$\boldsymbol{y}_t = \begin{pmatrix} 1 \\ y_t^2 \end{pmatrix}.$$

To compare Ochi's estimator (7) with the LSE, Burg's estimator, and Yule-Walker estimator, we give the derivations of the three estimators in the ARCH(1) case.

The paper is organized as follows. In Section 2, we extend the LSE, Burg's estimator and Yule-Walker estimator to the ARCH(1) model. In Section 3, we evaluate and compare Ochi's estimator and the three estimators in AR(1) and ARCH(1) models by simulation. Finally, in Section 4, we discuss these results and conclude.

2 LSE, Burg's estimator, and Yule-Walker estimator for ARCH(1) model

2.1 The least squares estimator By minimizing the sum of squared errors

$$\sum_{t=2}^{T} \{y_t^2 - (\theta_0 + \theta_1 y_{t-1}^2)\}^2,$$

we can obtain the LSE (e.g., Taniguchi et al., 2008)

$$\hat{\boldsymbol{\theta}}_{LSE} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y},$$

where

$$\text{(8)} \qquad \boldsymbol{Y} = (y_2^2, \cdots, y_T^2)', \quad \boldsymbol{Z} = \begin{pmatrix} 1 & y_1^2 \\ \vdots & \vdots \\ 1 & y_{t-1}^2 \\ \vdots & \vdots \\ 1 & y_{T-1}^2 \end{pmatrix},$$

and

$$\boldsymbol{Z}' = \left(\begin{pmatrix} 1 \\ y_1^2 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ y_2^2 \end{pmatrix}, \cdots, \begin{pmatrix} 1 \\ y_{T-1}^2 \end{pmatrix} \right) = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_{T-1}).$$

Then the LSE can be rewritten as

$$\hat{\boldsymbol{\theta}}_{LSE} = \left(\sum_{t=1}^{T-1} \boldsymbol{y}_t \boldsymbol{y}_t' \right)^{-1} \sum_{t=1}^{T-1} \boldsymbol{y}_t y_{t+1}^2. \tag{9}$$

Recalling the form of Ochi's estimator (7) with constants c_1 and c_2, we see that when $c_1 = 1$ and $c_2 = 0$, Ochi's estimator becomes the LSE.

2.2 Burg's method Burg's idea (Burg, 1975) is simple. With a previous given value y_{t-1} and a next given value y_{t+1}, forward and backward linear prediction can be represented as

$$\hat{y}_t^2 = \hat{\theta}_0 + \hat{\theta}_1 y_{t-1}^2, \quad t \in \{2, 3, \cdots, T\} \tag{10}$$

and

$$\tilde{y}_t^2 = \hat{\theta}_0 + \hat{\theta}_1 y_{t+1}^2, \quad t \in \{1, 2, \cdots, T-1\}, \tag{11}$$

respectively. The sum of the squared errors for (10) is

$$S_f = \sum_{t=2}^{T} (y_t^2 - \theta_0 - \theta_1 y_{t-1}^2)^2 \tag{12}$$

and for (11) is

$$S_b = \sum_{t=1}^{T-1} (y_t^2 - \theta_0 - \theta_1 y_{t+1}^2)^2. \tag{13}$$

Minimizing the sum of (12) and (13),

$$\begin{aligned} S = S_f + S_b &= \sum_{t=2}^{T} (y_t^2 - \boldsymbol{\theta}' \boldsymbol{y}_{t-1})^2 + \sum_{t=1}^{T-1} (y_t^2 - \boldsymbol{\theta}' \boldsymbol{y}_{t+1})^2 \\ &= y_1^4 + 2 \sum_{t=2}^{T-1} y_t^4 + y_T^4 + (\boldsymbol{\theta}' \boldsymbol{y}_1)^2 + 2 \sum_{t=2}^{T-1} (\boldsymbol{\theta}' \boldsymbol{y}_t)^2 + (\boldsymbol{\theta}' \boldsymbol{y}_T)^2 - 2\boldsymbol{\theta}' \left(\sum_{t=2}^{T} y_t^2 \boldsymbol{y}_{t-1} + \sum_{t=1}^{T-1} y_t^2 \boldsymbol{y}_{t+1} \right), \end{aligned}$$

by setting the gradient with respect to $\boldsymbol{\theta}$ as $\mathbf{0}$

$$\frac{\partial S}{\partial \boldsymbol{\theta}} = 2\boldsymbol{y}_1 (\boldsymbol{\theta}' \boldsymbol{y}_1)' + 4 \sum_{t=2}^{T-1} \boldsymbol{y}_t (\boldsymbol{\theta}' \boldsymbol{y}_t)' + 2\boldsymbol{y}_T (\boldsymbol{\theta}' \boldsymbol{y}_T)' - 2 \left(\sum_{t=2}^{T} y_t^2 \boldsymbol{y}_{t-1} + \sum_{t=1}^{T-1} y_t^2 \boldsymbol{y}_{t+1} \right) = \mathbf{0},$$

we have

$$\hat{\boldsymbol{\theta}}_{Burg} = \left(\boldsymbol{y}_1 \boldsymbol{y}_1' + 2 \sum_{t=2}^{T-1} \boldsymbol{y}_t \boldsymbol{y}_t' + \boldsymbol{y}_T \boldsymbol{y}_T' \right)^{-1} \left(\sum_{t=2}^{T} y_t^2 \boldsymbol{y}_{t-1} + \sum_{t=1}^{T-1} y_t^2 \boldsymbol{y}_{t+1} \right)$$

$$= \left(\boldsymbol{y}_1 \boldsymbol{y}_1' + 2 \sum_{t=2}^{T-1} \boldsymbol{y}_t \boldsymbol{y}_t' + \boldsymbol{y}_T \boldsymbol{y}_T' \right)^{-1} \begin{pmatrix} y_1^2 + 2\sum_{t=2}^{T-1} y_t^2 + y_T^2 \\ 2\sum_{t=1}^{T-1} y_t^2 y_{t+1}^2 \end{pmatrix}.$$

We see that Burg's estimator gives $\hat{\theta}_0$ a different form from Ochi's estimator. However, θ_1 is estimated as a special case of Ochi's estimator (7) by setting $c_1 = c_2 = 0.5$.

2.3 Yule-Walker estimator The Yule-Walker method is derived by considering the following expectations

$$\begin{aligned} E[y_t^2 y_t^2] &= \theta_0 E[y_t^2] + \theta_1 E[y_t^2 y_{t-1}^2] + E[\xi_t y_t^2] \\ &= \theta_0 \mu + \theta_1 E[y_t^2 y_{t-1}^2] + V[\xi_t], \end{aligned} \tag{14}$$

$$\begin{aligned} E[y_t^2 y_{t-1}^2] &= \theta_0 E[y_{t-1}^2] + \theta_1 E[y_{t-1}^2 y_{t-1}^2] + E[\xi_t y_{t-1}^2] \\ &= \theta_0 \mu + \theta_1 E[y_{t-1}^2 y_{t-1}^2] \\ &= \theta_0 \mu + \theta_1 E[y_t^2 y_t^2], \end{aligned} \tag{15}$$

where

$$\mu = E[y_t^2] = \theta_0 + \theta_1 E[y_{t-1}^2] + E[\xi_t] = \theta_0 + \theta_1 \mu. \tag{16}$$

Hence,

$$\mu = \frac{\theta_0}{1-\theta_1}, \quad \text{and} \quad \theta_0 = \mu(1-\theta_1). \tag{17}$$

From (16) and(15), we can see that

$$\theta_1 = \frac{E[y_t^2 y_{t-1}^2] - \mu^2}{E[y_t^2 y_t^2] - \mu^2} = \frac{C_1}{C_0}.$$

That is, θ_1 can be estimated by the lag 1 autocorrelation function of the series. Then the Yule-Walker estimator of the parameters $\boldsymbol{\theta} = (\theta_0, \theta_1)'$ can be obtained by

$$\hat{\boldsymbol{\theta}}_{YW} = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} y_t^2 \left(1 - \frac{\hat{C}_1}{\hat{C}_0}\right) \\ \hat{C}_1/\hat{C}_0 \end{pmatrix}, \tag{18}$$

where $\hat{C}_0$ and $\hat{C}_1$ are the sample autocovariance functions of y_t^2 for lags 0 and 1, respectively. Comparing this with Ochi's estimator (7), we can see that the Yule-Walker estimator of θ_1 is a centered version of Ochi*(1,1).

3 Simulation and results In this section we provide a simulation study for Ochi's estimator, the LSE, Burg's estimator and Yule-Walker estimator for AR(1) models in (1) and ARCH(1) models in the form (3). The models and parameters used in the simulations are given in Table 1.

Table 1: Simulation setting

AR(1)	ARCH(1)
$x_t = \alpha x_{t-1} + \epsilon_t, \qquad t \in \{2, \cdots, T\}$	$y_t^2 = (\theta_0 + \theta_1 y_{t-1}^2) u_t^2, \qquad t \in \{2, \cdots, T\}$
	$\theta_0 = 1$
$\alpha = (0, 0.05, 0.1, \cdots, 0.95)'$	$\theta_1 \in \{0, 0.05, 0.1, \cdots, 0.5, 0.55 < \sqrt{1/3}\}$
$\epsilon_t \sim \text{iid } N(0,1)$	$u_t \sim \text{iid } N(0,1)$
$\{x_1, x_2, \cdots, x_T\}$	$\{y_1^2,\ y_2^2, \ldots, y_T^2\}$
$T \in \{100,\ 200,\ 300,\ 500,\ 1000\}$	
Replications of time series sequences $N = 1000$	
$c_1,\ c_2 \in \{0, 0.2, 0.5, 0.7, 1\}$	

3.1 Simulation of Ochi's estimator for AR(1) For model (1) with $\sigma^2 = 1$ and α as each of $(0, 0.05, 0.1, \cdots, 0.95)'$, we generate $N = 1000$ sequences as time series with length 1000. That is, 20×1000 sequences of length 1000 are generated. For each sequence, we consider five different sample sizes T,

$$x_1,\ x_2, \cdots, x_T, \quad T \in \{100,\ 200,\ 300,\ 500,\ 1000\}$$

and estimate the parameter α by Ochi's estimator (2) for each T. In the calculation, we try all 25 pairs of (c_1, c_2), $c_1,\ c_2 \in \{0, 0.2, 0.5, 0.7, 1\}$ for the constants of Ochi's estimator. Then, we evaluate the performance of the different pairs (c_1, c_2) by comparing the resulting mean square errors (MSEs) for different sample sizes as α changes.

Figure 1 shows a part of the simulation results for $T = 100$ and $T = 200$. As expected, a bigger sample size T gives a smaller MSE, and hence provides a better estimate of α. For each sample size, MSE is calculated for all 25 pairs of (c_1, c_2). The MSE curves are plotted as dashed lines. MSE curves for four special pairs are plotted as solid lines, in red for $(0, 0)$, green for $(0.5, 0.5)$, blue for $(1, 0)$, and light blue for $(1, 1)$. We see that as α grows, the MSE curves decrease.This means that better estimation is obtained when α is bigger. At around $\alpha = 0.5$, the MSEs for different pairs of (c_1, c_2) do not show big differences. When $\alpha < 0.4$, the two extreme cases, $(c_1, c_2) = (0, 0)$ and $(c_1, c_2) = (1, 1)$, give the biggest and smallest MSEs, respectively. However, when $\alpha > 0.6$, these two curves exchange their positions. In contrast, Burg's method(Ochi(0.5, 0.5)) and the LSE (Ochi(1,0)) give intermediate MSEs. In particular, when α is close to 1, Burg's method is slightly better than the LSE. When the sample size is large, different pairs of (c_1, c_2) give only small differences in MSE.

The figures for the variance and squared bias show that as α increases the variance becomes smaller but the squared bias becomes bigger. In particular, Ochi$(1, 1)$ (the Yule-Walker estimator) shows the largest squared bias among these methods. We can also see from the last panel of the figure that the mean of the estimated α is slightly less than the real α (gray line).

3.2 Simulation of Ochi's estimator for ARCH(1) We set the true parameters in model (4) as $\theta_0 = 1$ and $\theta_1 \in \{0, 0.05, 0.1, \cdots, 0.5, 0.55 < \sqrt{1/3}\}$, and then we use Ochi's estimator (7) to estimate the parameters by simulation with different constants $c_1,\ c_2 \in \{0, 0.2, 0.5, 0.7, 1\}$. For every value of θ_1, $N = 1000$ sequences of length 1000 are generated. For each sequence, we consider different values of T;

$$y_1^2,\ y_2^2, \ldots, y_T^2, \qquad T \in \{100, 200, 300, 500, 1000\}.$$

We estimate θ_0 and θ_1 and then compute the MSEs, variances, and squared biases of the estimates for each case.

Figure 2 shows the MSE, variance, and squared biase for Ochi's estimator for estimating θ_0 and θ_1. Different colors indicate different lengths of the time series (or sample sizes), with these sizes $T \in \{100, 200, 300, 500, 1000\}$. In each panel, for each T, 25 curves obtained from different pairs of (c_1, c_2), $c_1,\ c_2 \in \{0, 0.2, 0.5, 0.7, 1\}$, are plotted with respect to θ_1.

The first panel shows MSE curves obtained in estimating θ_0. The MSE curve obtained from $(c_1 = 1, c_2 = 1)$ is plotted as a solid line, and the other 24 MSE curves are plotted as dashed lines. The graph shows that as the sample size T increases, the corresponding MSE becomes smaller. For big sample sizes, such as $T = 1000$, the choice of (c_1, c_2) makes almost no difference. In contrast, with smaller sample sizes, there are bigger differences

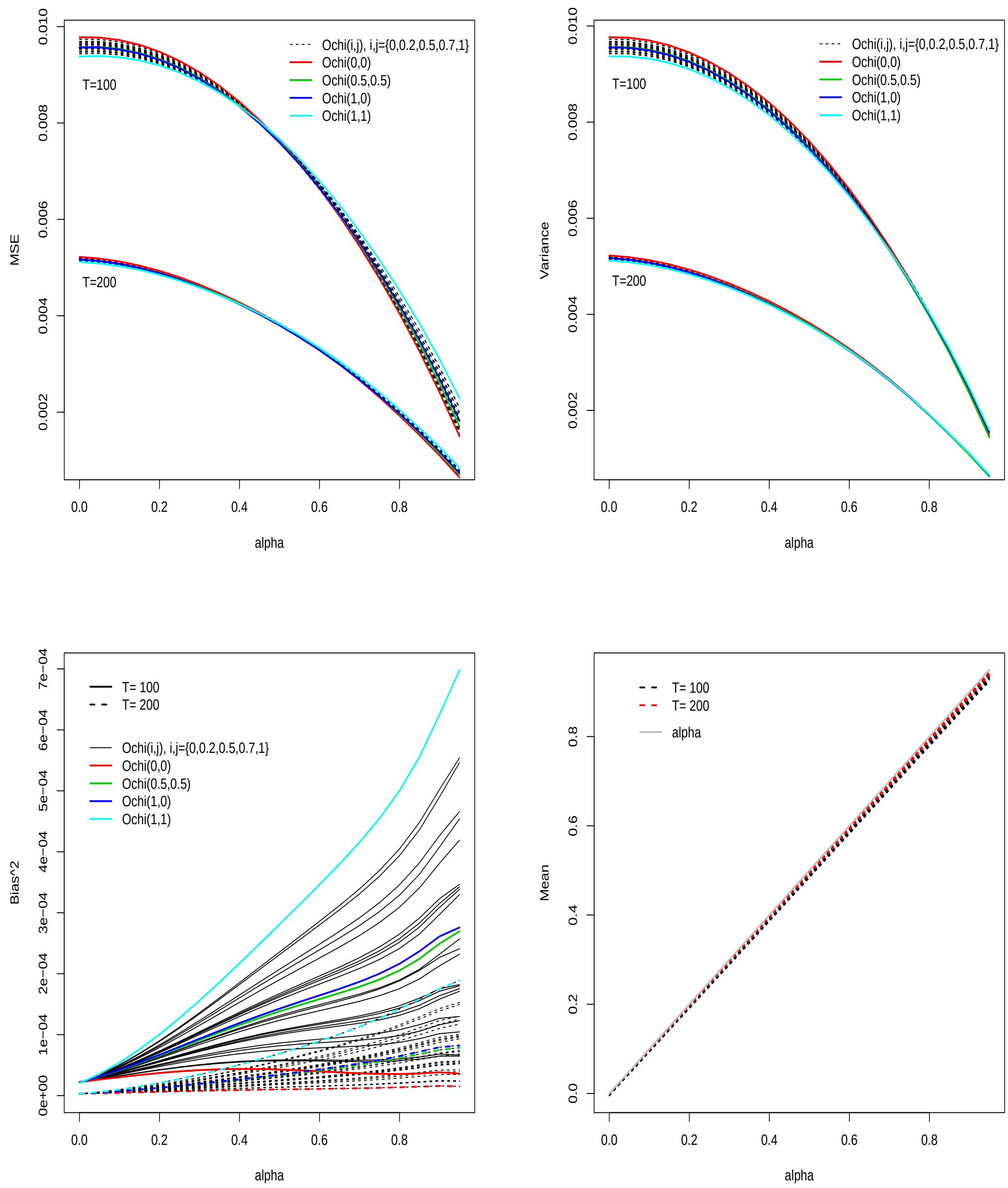

Figure 1: Comparison of Ochi(c_1, c_2) estimators for AR(1) model.

among different pairs of (c_1, c_2). We also see that $(c_1, c_2) = (1, 1)$ gives a better estimation for θ_0 than the other pairs do. However, increasing θ_1 enlarges MSE and gives worse estimation of θ_0.

The second panel shows MSE curves in estimating θ_1. Comparing this with the first panel, we see that a similar trend is obtained, except that Ochi$^*(c_1 = 1, c_2 = 1)$ works better than the others when $\theta_1 \leq 0.3$; around a θ_1 of $[0.3, 0.4]$, the difference of attributable to (c_1, c_2) is very small; after that, Ochi$^*(c_1 = 1, c_2 = 1)$ becomes worse than the others, and Ochi$^*(c_1 = 0, c_2 = 0)$ works better, instead. As special cases of Ochi's estimator, the LSE (Ochi$^*(1, 0)$) and Burg's estimator (Ochi$^*(0.5, 0.5)$) perform similarly in between Ochi$^*(1, 1)$ and Ochi$^*(0, 0)$.

Plots of variance and squared biases for both θ_0 and θ_1 are given in the second and third rows, respectively in Figure 2. We can see that $(c_1 = 1, c_2 = 1)$ has better performance than the other pairs of (c_1, c_2) for estimating θ_0. For θ_1, Ochi$^*(c_1 = 1, c_2 = 1)$ has smaller variance than with other constants, but its squared bias is bigger.

Since Ochi$^*(c_1 = 1, c_2 = 1)$ works well in all the cases for estimating θ_0, we compare it with the methods of LSE, Burg and Yule-Walker in Figure 3 by evaluating their MSE, variance and squared bias curves. In each panel of Figures 3, the results for different sample sizes are indicated by different colors. For each sample size, four curves with respect to θ_1 are plotted, one for each of four different methods. Figure 3 shows that Ochi$^*(c_1 = 1, c_2 = 1)$ works well in all the cases for θ_0.

The last two graphs in Figure 3 show the means of the estimates, $\hat{\theta}_0$ and $\hat{\theta}_1$, obtained with different methods and different sample sizes. In the simulation, the true value of θ_0 is fixed to 1, and the true θ_1 takes values from $\{0, 0.05, 0.1, \cdots, 0.55\}$. We see that when θ_1 becomes bigger, the means of the estimates spread from above (with $\hat{\theta}_0$) and below (with $\hat{\theta}_1$) the gray lines in the two graphs. That is, θ_0 is over estimated and θ_1 is under estimated in the ARCH(1) model. The simulation also shows that, in estimating θ_1, Ochi$^*(1, 1)$ and the Yule-Walker estimator are not exactly the same but their results are particularly close.

3.3 Data with heavy-tailed distributions In time series analysis, data with heavy-tailed distributions are often of interest. Here, we also evaluate Ochi's estimator, the LSE, Burg's estimator, and the Yule-Walker estimator by simulation when errors ϵ_t and u_t have t distributions with $4, \cdots, 10$ degrees of freedom. Since the simulations give similar results, we show only here the cases of $\epsilon_t \sim$ iid $t(5)$ for the AR(1) model, and $u_t \sim$ iid $t(5)$ for the ARCH(1) model.

For the AR(1) model, Ochi(c_1, c_2) performs similar to its performance in the case of $\epsilon_t \sim N(0, 1)$. This can be seen by comparing Figures 1 and 4.

When $u_t \sim$ iid $t(5)$, among Ochi$^*(c_1, c_2)$, $c_1, c_2 \in \{0, 0.2, 0.5, 0.7, 1\}$, Ochi$^*(1, 1)$ is better at estimating θ_0. For θ_1, Ochi$^*(1, 1)$ maintains a small MSE and shows stability, as seen in the first row of Figure 5. Figure 5 also shows the variance and squared bias for $\hat{\theta}_0$ and $\hat{\theta}_1$ compared for instances of Ochi$^*(c_1, c_2)$.

Comparing Ochi$^*(1, 1)$ with the LSE, Burg's estimator, and the Yule-Walker estimator in Figure 6, we see that Ochi$^*(1, 1)$ works well in most cases for estimating θ_0. In estimating θ_1, from the last panel of Figure 6, we can see that Ochi$^*(1, 1)$ and the Yule-Walker estimator give similar performance, and this performance is close to that of Burg's estimator and is more stable than the LSE.

From the ranges of MSE and mean in Figures 5 and 6, we also see that, for the ARCH(1) model, estimation of θ_0 is difficult when u_t has heavy-tailed distributions.

4 Conclusions Ochi's estimator is examined for estimating the parameters in both AR(1) and ARCH(1) models. The simulation for AR(1) models shows that Ochi(1,1) (equivalently Yule-Walker estimator) works well when $\alpha < 0.4$, and Ochi(0,0) gives a smaller MSE when $\alpha > 0.6$. Around $\alpha = 0.5$, there is not much difference in MSEs of Ochi(c_1, c_2). Ochi(1,1) also has bigger squared biases than the other methods. Ochi(1,0) and Ochi(0.5, 0.5) are the LSE and Burg's estimator, respectively. They give intermediate MSE values.

Since ARCH(1) models can be written as a form of AR(1), we introduced Ochi's estimator to ARCH(1). With different pairs of Ochi parameters (c_1, c_2), we investigated its performance by simulation and found that Ochi$^*(c_1 = 1, c_2 = 1)$ works well for estimating θ_0. Ochi$^*(c_1 = 1, c_2 = 1)$ performs similarly to the Yule-Walker estimator, having relatively smaller MSEs than given by the LSE and Burg's estimator for $\theta_1 < 0.3$; when $\theta_1 > 0.4$, the LSE and Burg's estimator work better.

When the data are heavy-tailed, such as when $\epsilon_t \sim t(5)$, for AR(1), Ochi's estimator robustly estimates α. However, for ARCH(1), from the large MSEs, we see that Ochi's estimator gives poor estimation of θ_0. The MSEs for θ_1 are also big, but much smaller than those for θ_0. Moreover, Ochi$^*(1, 1)$ shows performance similar to the performance of the

Yule-Walker estimator and Burg's estimator for θ_1.

Acknowledgments The author thanks the editors and the referee for their constructive comments that led to an improved article. The author also expresses her sincere thanks and appreciation to Professor Masanobu Taniguchi for his great help and valuable advices on the paper.

References

[1] Burg, J. P. (1975). Maximum entropy spectral analysis, *PhD thesis, Stanford University* , Available from `http://sepwww.stanford.edu/theses/sep06`.

[2] Chan, K. S., Tong, H. (1986). On estimating thresholds in autoregressive models, *Journal of Time Series Analysis*, **7**, 179–190.

[3] Davis, R.A., Lee, T., and Rodriguez-Yam, G. (2008). Break Detection for a Class of Nonlinear Time Series Models, Journal of Time Series Analysis, **29**, 834–867.

[4] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, **50** (4), 987–1007.

[5] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57** 357–384.

[6] Ochi, Y. (1983). Asymptotic expansions for the distribution of an estimator in the first-order autoregressive process, *Journal of time series analysis*, **4** (1), 57–67.

[7] Shumway, R. H. and Stoffer, D. S. (2011). *Time Series Analysis and Its Applications: With R Examples*, Third Edition. Springer, New York.

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

[9] Tong, H. (1990). *Non-linear time series*, Oxford Science Publications, Oxford.

[10] Turkman, K., Scotto, M. G., de Zea Bermudez, P. (2014). *Non-Linear Time Series: Extreme Events and Integer Value Problems*, Springer, New York.

[11] Vos, K. (2013). A fast implementation of Burg's method, Available from `http://creativecommons.org/licenses/by/3.0`.

Communicated by *Masanobu Taniguchi*

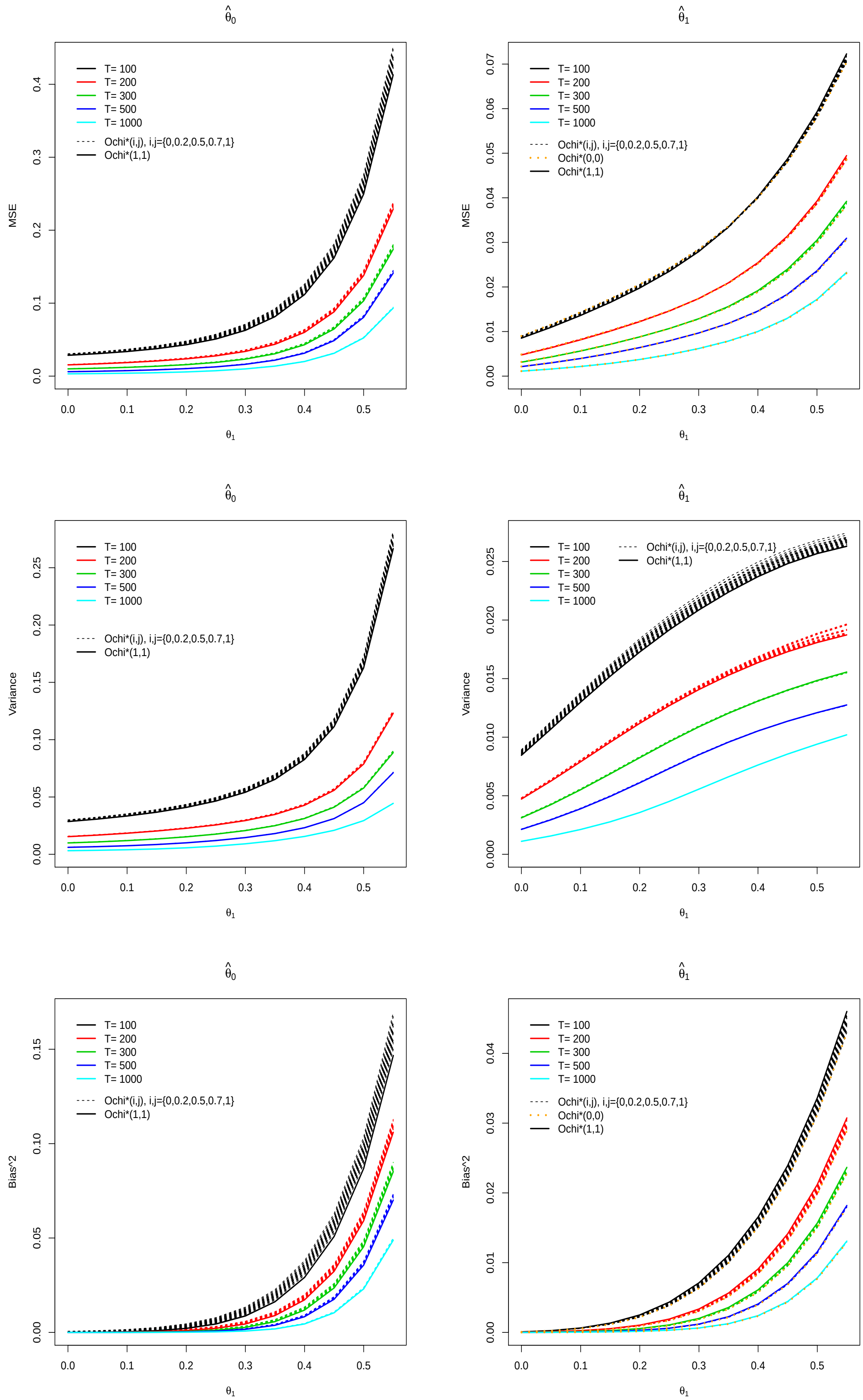

Figure 2: Comparison of Ochi's estimators with different pairs of (c_1, c_2). Left panels show θ_0, right panels show θ_1.

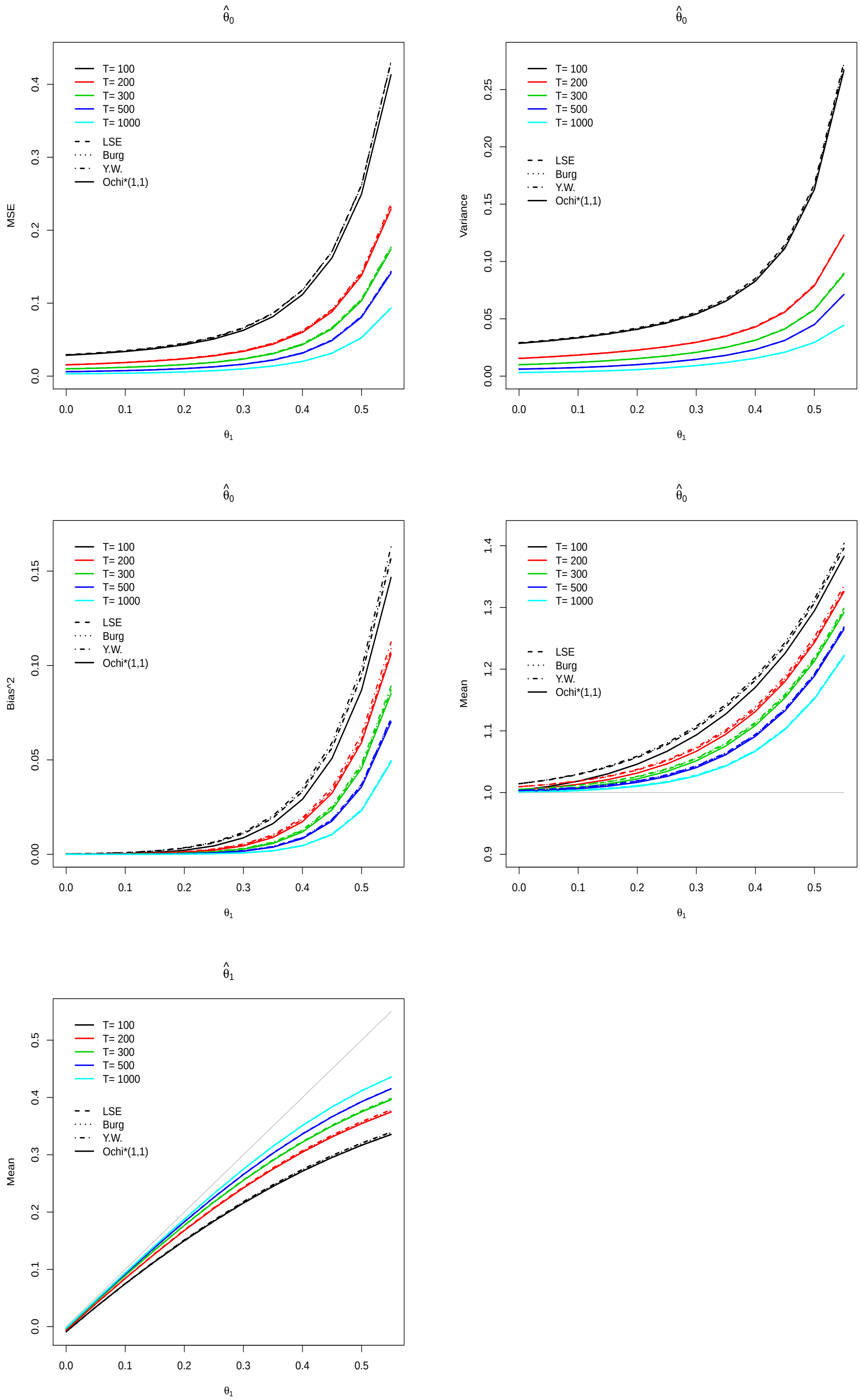

Figure 3: Comparison of Ochi$(1,1)$ estimator with methods of LSE, Burg and Yule-Walker for θ_0. The last two panels show means of the estimates of parameters θ_0 and θ_1.

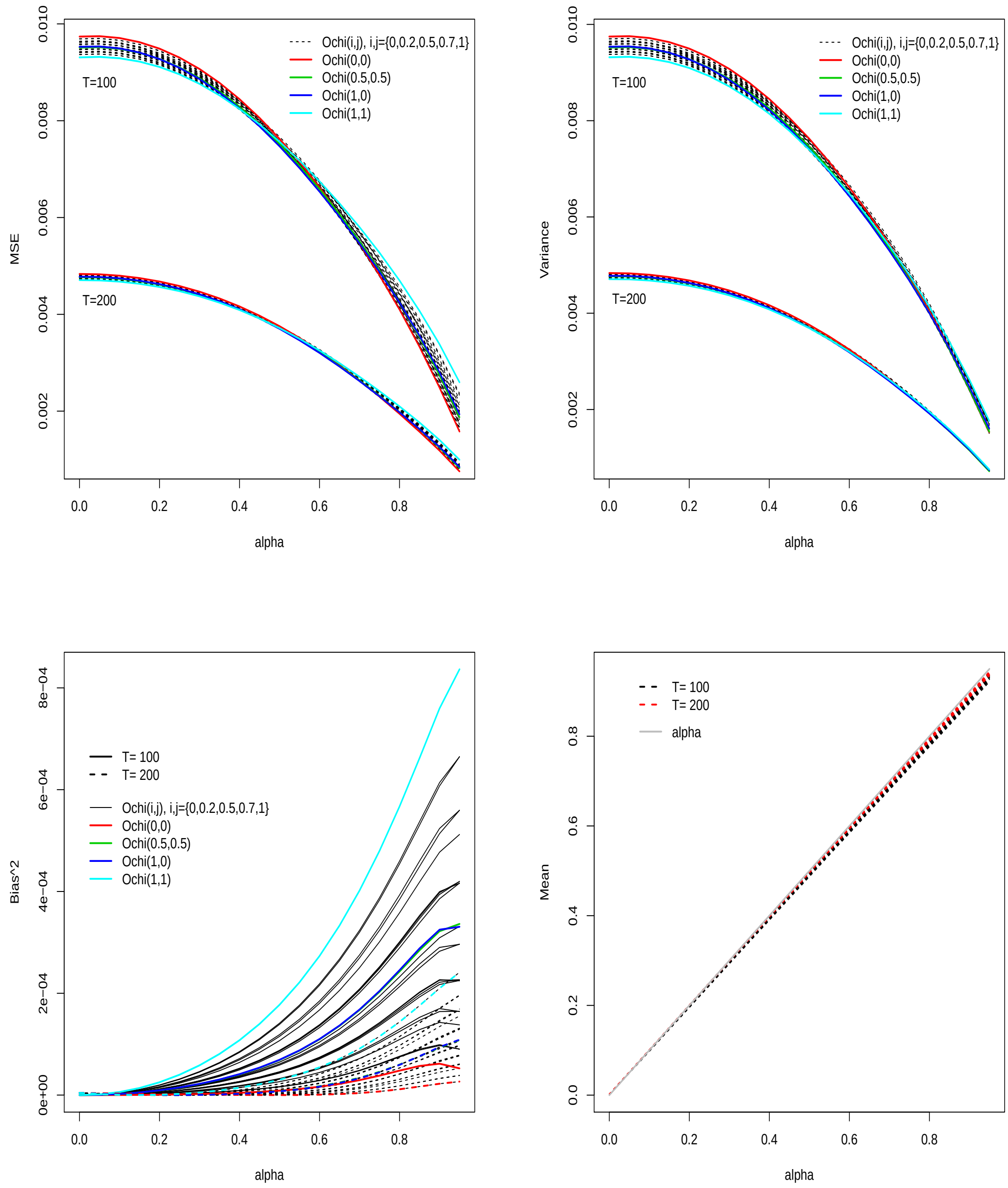

Figure 4: For AR(1), when $\epsilon_t \sim$ iid $t(5)$, Ochi's estimators perform similar to case with $\epsilon_t \sim$ iid $N(0,1)$.

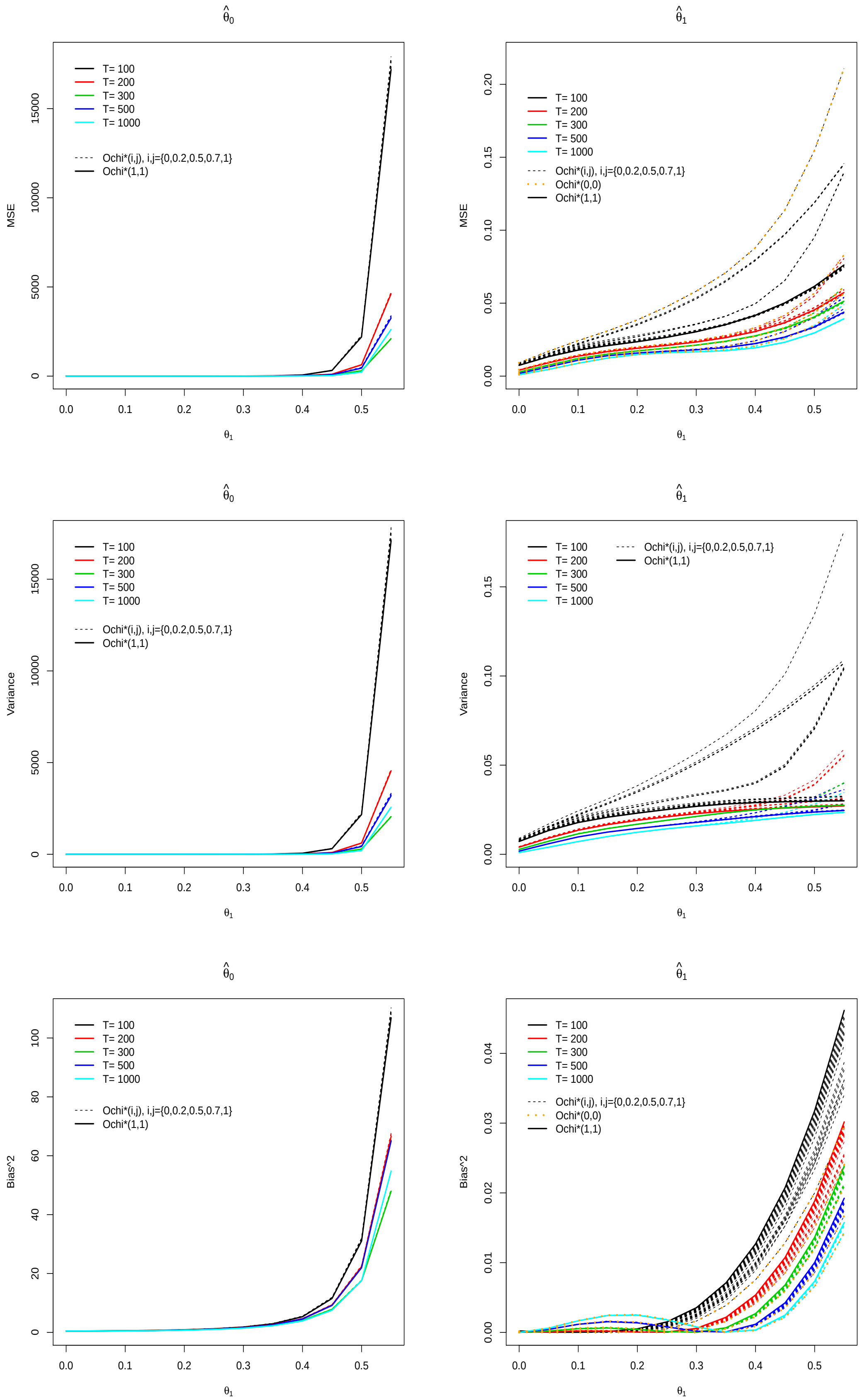

Figure 5: Comparison of Ochi's estimators with different pairs of (c_1, c_2) in case of $u_t \sim$iid $t(5)$. Left panels show θ_0, right panels show θ_1.

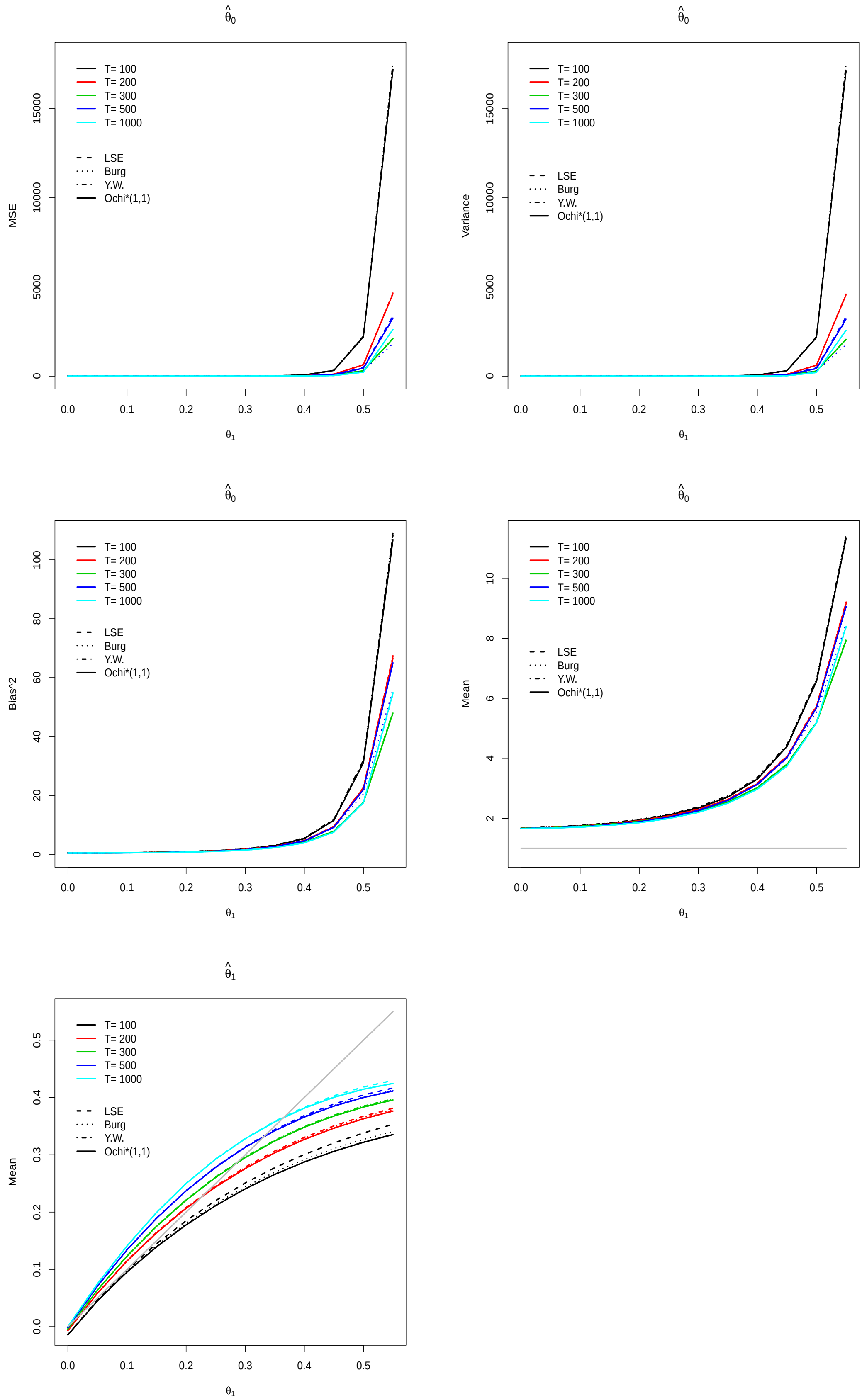

Figure 6: Comparison of Ochi$(1, 1)$ estimator and methods of LSE, Burg and Yule-Walker for θ_0 in the case of u_t ∼iid $t(5)$. The last two panels show means of the estimates of parameters θ_0 and θ_1 in the case of u_t ∼iid $t(5)$.