Bootstrap prediction intervals for threshold autoregressive models

Jing, Li

South Dakota State University

January 2009

Bootstrap Prediction Intervals For Threshold Autoregressive Models

Jing Li*

Department of Economics
South Dakota State University

Abstract

This paper examines the performance of prediction intervals based on bootstrap for threshold autoregressive models. We consider four bootstrap methods to account for the variability of estimates, correct the small-sample bias of autoregressive coefficients and allow for heterogeneous errors. Simulation shows that (1) accounting for the sampling variability of estimated threshold values is necessary despite super-consistency, (2) bias-correction leads to better prediction intervals under certain circumstances, and (3) two-sample bootstrap can improve long term forecast when errors are regime-dependent.

Keywords: Bootstrap; Interval Forecasting; Threshold Autoregressive Models; Time Series; Simulation

*Department of Economics, Scobey Hall, Box 504, South Dakota State University, Brookings, SD 57007. Phone: 605.688.4848, Fax: 605.688.6386, Email: Jing.Li@sdstate.edu.

Introduction

Constructing prediction intervals is an important topic in forecasting time series. One can construct classical prediction intervals (CPI) for an autoregression (AR) using the Box-Jenkins method, see Granger and Newbold (1986) for instance. In practice two factors may worsen the finite-sample performance of CPI. First, the distribution of prediction errors may be non-normal. Second, CPI may fail to take into account the sampling variability of estimated autoregressive coefficients. The second factor directly causes the under-nominal coverage of CPI found in previous studies such as Thombs and Schucany (1990) and Kim (2001).

The non-normal prediction error plays a more important role for the threshold autoregressive (TAR) model developed by Tong (1983). All estimated autoregressive coefficients in TAR models are functions of the estimated threshold value, which is well known to follow a nonstandard distribution. Therefore prediction errors of TAR models (calculated as the difference between true future values and predicted values) follow nonstandard distributions by construction.

Given these limitations of CPI, bootstrap prediction intervals (BPI) for TAR models are widely used by practitioners, though the performance of BPI has not been systematically investigated. This paper is intended to fill the gap. In particular, four methods of constructing BPI are considered. The first issue is unique to TAR models, and is concerned with the estimated threshold value. Chan (1993) proves that the threshold value estimated by grid-searching is super-consistent. That means for some problems of statistical inference we can treat the estimated threshold value as the true value and ignores its sampling variability. In this paper we investigate whether it is worthwhile to account for that sampling variability when constructing BPI. The second issue is related to the small-sample bias of estimated autoregressive coefficients. Following Kilian (1998) and Kim (2001) we adopt the bootstrap-after-bootstrap method and compute BPI based on bias-corrected autoregressive coefficients.

The last issue is unique to TAR models with regime-dependent errors, for which constructing BPI belongs to the "two-sample problem" in the terminology of Efron and Tibshirani (1993). Existing literature typically downplays this problem. But we propose a method of bootstrapping separately the residuals in each regime (called two-sample bootstrap hereafter), instead of bootstrapping the pooled residuals. In this paper we do not consider the percentile-t method since it is theoretically difficult to compute the asymptotic standard error and its bootstrap counterpart ($\hat{\sigma}_k^c(h)$ and $\hat{\sigma}_k^*(h)$ in Kim (2001)) for TAR models.

BPI for linear autoregressive models is studied by several authors. Among others, Thombs and Schucany (1990) generate bootstrap replicates based on backward AR models. Kim (2001) and Kim (2002) apply the bootstrap-after-bootstrap method of Kilian (1998) to correct the bias of autoregressive coefficients. Masarotto (1990) and Grigoletto (1998) build BPI based on forward AR models. Kim (1999) and Kim (2004) consider BPI for vector autoregressions.

In this paper, the performance of four BPIs for the nonlinear TAR model is compared by extensive simulation. Special attention is paid to scrutinizing the effects of (I) magnitude of varying coefficients across regimes, (II) the number of observations subject to regime-switching, (III) the degree of heterogeneity in errors and (IV) the possible non-stationarity. (I) (II) and (III) have not been discussed in above papers since they are irrelevant to linear models.

The remainder of the paper is organized as follows. Section 2 specifies the TAR model. A simple simulation is used to highlight the nonstandard distribution of predicted values of TAR models. Three methods of constructing BPI for TAR models with regime-invariant errors are provided in Section 3. Section 4 constructs BPI for TAR models with regime-varying errors. Simulation is conducted in Section 5, and Section 6 concludes. Without further explanation the nominal coverage rate is 0.95 throughout the paper.

TAR Models

The observed data are $(y_1, \ldots, y_n)$, with initial conditions $(y_0, y_{-1}, \ldots, y_{-p+1})$. A two-regime self-exciting threshold autoregressive (SETAR) model of order p is specified as

$$y_t = \left(\beta_{10} + \sum_{j=1}^{p} \beta_{1j} y_{t-j}\right) 1(y_{t-1} > \gamma) + \left(\beta_{20} + \sum_{j=1}^{p} \beta_{2j} y_{t-j}\right) 1(y_{t-1} \leq \gamma) + e_t, \qquad (1)$$

where e_t is the error with a common distribution function F_e. We do not assume a specific parametric form of F_e. However, we do assume $e_t \sim \texttt{iid}(0, \sigma^2)$ to facilitate the residual bootstrap. The lag order p can be chosen based on information criteria such as AIC and BIC. After fitting (1) one may check whether residuals are serially uncorrelated to ensure bootstrap works properly.

The threshold value is denoted by γ, and $1(\,.)$ denotes the indicator function that equals one if the event inside the parentheses is true and zero otherwise. In regime one y_{t-1} is greater than γ, and is less than or equal to γ in regime two. At a given lag j, model (1) allows for different autoregressive coefficients across regimes. The threshold effect exists if $\beta_{1j} \neq \beta_{2j}$ for some j. A formal test for the threshold effect is developed in Chan (1990).

The unknown threshold value can be estimated by grid-searching as follows. For given γ we define the indicator and fit model (1) by OLS. We do this for a range of $\gamma \in [\gamma^l, \gamma^u]$, where the lower and upper searching bounds are the τth and $(100 - \tau)$th percentiles of the empirical distribution of y_{t-1}. In this paper $\tau = 15$ is used throughout. The value of γ minimizing the residual sum of squares (RSS) is the estimated threshold value:

$$\hat{\gamma} = \texttt{argmin}_{\gamma \in [\gamma^l, \gamma^u]} \texttt{RSS}(\gamma). \qquad (2)$$

Chan (1993) shows that the estimated threshold value (2) is super-consistent, thereby converging to the true value very quickly. Model (1) can be generalized in several ways. For

example, an exogenous variable can serve as the threshold variable entering the indicator. A three-regime or band TAR model can be specified by defining $1(y_{t-1} > \gamma_2)$ for regime 1, $1(\gamma_1 \leq y_{t-1} \leq \gamma_2)$ for regime 2 and $1(y_{t-1} < \gamma_1)$ for regime 3. In addition, one can specify the indicator as $1(y_{t-d} > \gamma)$ with unknown d. Then grid-searching γ can be nested into the search for the number of regimes and d. For expositional purpose, the following discussions are based on the basic model (1) (so that TAR is synonymous with SETAR in this paper).

After estimating $\hat{\gamma}$ by (2) and fitting model (1) with $\hat{\gamma}$, compute residuals as

$$\hat{e}_t = y_t - \left(\hat{\beta}_{10} + \sum_{j=1}^{p} \hat{\beta}_{1j} y_{t-j}\right) 1(y_{t-1} > \hat{\gamma}) - \left(\hat{\beta}_{20} + \sum_{j=1}^{p} \hat{\beta}_{2j} y_{t-j}\right) 1(y_{t-1} \leq \hat{\gamma}), \qquad (3)$$

where $\hat{\beta}_{ij}$ denotes the least squares estimate of the coefficient. By construction the residual is centered at zero. We do not re-scale the residual since preliminary simulation shows little effect of re-scaling. Let $h \geq 1$ denote the forecast horizon. Then the h-step-ahead predicted value conditional on the last p observations of y_t can be computed as

$$\hat{y}_{t+h} = \left(\hat{\beta}_{10} + \sum_{j=1}^{p} \hat{\beta}_{1j} \hat{y}_{t+h-j}\right) 1(\hat{y}_{t+h-1} > \hat{\gamma}) + \left(\hat{\beta}_{20} + \sum_{j=1}^{p} \hat{\beta}_{2j} \hat{y}_{t+h-j}\right) 1(\hat{y}_{t+h-1} \leq \hat{\gamma}), \quad (4)$$

where $\hat{y}_t = y_t, (t = n, n-1, \ldots, n-p+1)$.

The nonstandard distribution of estimated threshold value is established in Hansen (2000). The predicted values of TAR models by construction are linear combinations of estimated autoregressive coefficients, which are themselves functions of the estimated threshold value. Thus, it is not surprising that the predicted values (and prediction errors) follow nonstandard distributions. To see this, we generate 5000 1-step-ahead predicted values from the TAR model: $y_t = 0.2y_{t-1}1(y_{t-1} > 0) + 0.8y_{t-1}1(y_{t-1} \leq 0) + e_t, (t = 1, \ldots, n)$, where $n = 60, e_t \sim \texttt{iidn}(0, 1), y_0 = 0$. The standard normal error e_t is used to highlight the role of nonstandard distribution of $\hat{\gamma}$. After obtaining $\hat{\gamma}$ (2), we compute $\hat{y}_{t+1}$ using (4). Figure 1

displays the histogram of standardized $\hat{y}_{t+1}$ and corresponding statistics. First the significant skewness and kurtosis cast doubt on normality. Then the Jarque-Bera test of Jarque and Bera (1987) formally rejects normality for standardized $\hat{y}_{t+1}$ at 0.01 level. This simulation demonstrates that the assumption of normal prediction errors is invalid for TAR models.

One-Sample Bootstrap Prediction Intervals for TAR Models

Given the nonstandard distribution of prediction errors, we consider bootstrap prediction intervals for TAR models. The point is to use bootstrap to "automatically" account for the variability of estimated parameters and non-normality of prediction errors. Four bootstrap methods are proposed depending on whether accounting for the variability of estimated threshold values, whether correcting for the bias of autoregressive coefficient and whether generalizing model (1) by heterogeneous errors.

Chan (1993) shows that $\hat{\gamma}$ in (2) is n-consistent. This super-consistency implies that it is plausible to construct prediction intervals by ignoring the variability of $\hat{\gamma}$. Method 1 uses this idea and consists of the following steps.

Method 1

Step 1-1. Let $F_{\hat{e}}$ denote the empirical cdf of $\hat{e}_t$ computed by (3). Then use $\hat{\gamma}$, $\hat{\beta}_{ij}$, and generate recursively the bootstrap replicate of y_t (denoted by y_t^*) as

$$y_t^* = y_t, t = 1, \ldots, p,$$
$$y_t^* = \left(\hat{\beta}_{10} + \sum_{j=1}^{p} \hat{\beta}_{1j} y_{t-j}^*\right) 1(y_{t-1}^* > \hat{\gamma}) + \left(\hat{\beta}_{20} + \sum_{j=1}^{p} \hat{\beta}_{2j} y_{t-j}^*\right) 1(y_{t-1}^* \leq \hat{\gamma}) + e_t^*, (t > p)$$

where e_t^* is a random draw from $F_{\hat{e}}$ (ie., a draw from $\{\hat{e}_t\}_{t=p+1}^n$ with replacement).

Step 1-2. Re-estimate Model (1) using y_t^* and $\hat{\gamma}$, and obtain bootstrap coefficients $\hat{\beta}_{ij}^*$. Then

compute a bootstrap h-step-ahead future value (denoted by $\hat{y}^*_{t+h}$):

$$\hat{y}^*_t = y_t, t = n, n-1, \ldots, n-p+1,$$
$$\hat{y}^*_{t+h} = \left(\hat{\beta}^*_{10} + \sum_{j=1}^{p} \hat{\beta}^*_{1j}\hat{y}^*_{t+h-j}\right) 1(\hat{y}^*_{t+h-1} > \hat{\gamma}) + \left(\hat{\beta}^*_{20} + \sum_{j=1}^{p} \hat{\beta}^*_{2j}\hat{y}^*_{t+h-j}\right) 1(\hat{y}^*_{t+h-1} \leq \hat{\gamma}) + e^*_{t+h},$$

where e^*_{t+h} is a random draw from $F_{\hat{e}}$.

Step 1-3. Repeat Step 1-1 and 1-2 B times and obtain a series of bootstrap future values $\left\{\hat{y}^*_{t+h}(i)\right\}_{i=1}^{B}$, where i indexes resampling. The Method-1 h-step-ahead BPI at 0.95 nominal level is given by

$$\texttt{BPI1} = [\hat{y}^{.025}_{t+h}, \hat{y}^{.975}_{t+h}] \tag{5}$$

where $\hat{y}^{.025}_{t+h}$ and $\hat{y}^{.975}_{t+h}$ are the 2.5th and 97.5th percentiles of the empirical cdf of $\left\{\hat{y}^*_{t+h}(i)\right\}_{i=1}^{B}$.

First notice that in Step 1-1 we generate the bootstrap replicate in the forward model. The backward representation used by Thombs and Schucany (1990), Kim (2001) and others is unavailable for TAR models since it is impossible to invert the lag polynomial augmented with indicators. Second, following Efron and Tibshirani (1986) we use the first p observations of observed series as initial values for bootstrap replicates. Alternatively one may use any block of p observations of y_t. In Step 1-2 we compute bootstrap forecasts always using the last p observations of observed series. The importance of this conditionality on last p observations is stressed in Maekawa (1987), Chatfield (1993) and Kim (2001). Note Method 1 accounts for the sampling variability of $\hat{\beta}_{ij}$ but ignores its small-sample bias shown by Shaman and Stine (1988). That bias is corrected by Method 2 as follows.

Method 2

Step 2-1. The same as Step 1-1.

Step 2-2. Re-estimate Model (1) using y_t^* and $\hat{\gamma}$, and obtain bootstrap coefficients $\hat{\beta}_{ij}^*$. Repeat this process C times, and get a series of bootstrap coefficients $\left\{\hat{\beta}_{ij}^*(k)\right\}_{k=1}^{C}$, where k indexes resampling. Compute the bias-corrected autoregressive coefficients as

$$\hat{\beta}_{ij}^c = 2\hat{\beta}_{ij} - C^{-1}\sum_{k=1}^{C}\hat{\beta}_{ij}^*(k), (i = 1, 2, j = 0, 1, \ldots, p). \tag{6}$$

Step 2-3. Use $\hat{\beta}_{ij}^c$ to generate the bias-corrected bootstrap replicate as $y_t^{c*} = y_t, t = 1, \ldots, p$, and

$$y_t^{c*} = \left(\hat{\beta}_{10}^c + \sum_{j=1}^{p}\hat{\beta}_{1j}^c y_{t-j}^{c*}\right) 1(y_{t-1}^{c*} > \hat{\gamma}) + \left(\hat{\beta}_{20}^c + \sum_{j=1}^{p}\hat{\beta}_{2j}^c y_{t-j}^{c*}\right) 1(y_{t-1}^{c*} \leq \hat{\gamma}) + e_t^*.(t > p),$$

where e_t^* is a random draw from $F_{\hat{e}}$.

Step 2-4. Re-estimate Model (1) using y_t^{c*} and $\hat{\gamma}$, and obtain coefficients $\hat{\beta}_{ij}^{c*}$. Then compute a bias-corrected bootstrap future value as $\hat{y}_t^{c*} = y_t, t = n, n-1, \ldots, n-p+1$, and

$$\hat{y}_{t+h}^{c*} = \left(\hat{\beta}_{10}^{c*} + \sum_{j=1}^{p}\hat{\beta}_{1j}^{c*}\hat{y}_{t+h-j}^*\right) 1(\hat{y}_{t+h-1}^{c*} > \hat{\gamma}) + \left(\hat{\beta}_{20}^{c*} + \sum_{j=1}^{p}\hat{\beta}_{2j}^{c*}\hat{y}_{t+h-j}^{c*}\right) 1(\hat{y}_{t+h-1}^{c*} \leq \hat{\gamma}) + e_{t+h}^*,$$

where e_{t+h}^* is a random draw from $F_{\hat{e}}$.

Step 2-5. Repeat Step 2-1, 2-2, 2-3 and 2-4 B times and obtain a series of bias-corrected bootstrap future values $\left\{\hat{y}_{t+h}^{c*}(i)\right\}_{i=1}^{B}$. The Method-2 h-step-ahead BPI is given by

$$\texttt{BPI2} = [\hat{y}_{t+h}^{.025c}, \hat{y}_{t+h}^{.975c}] \tag{7}$$

where $\hat{y}_{t+h}^{.025c}$ and $\hat{y}_{t+h}^{.975c}$ are the 2.5th and 97.5th percentiles of the empirical cdf of $\left\{\hat{y}_{t+h}^{c*}(i)\right\}_{i=1}^{B}$.

Basically, Step 2-2 calculates the bias-corrected autoregressive coefficients following Kilian (1998) and Kim (2001). Then Step 2-3 generates bootstrap replicates using the bias-corrected coefficients. Notice that Steps 2-3 and 2-4 resample original residuals. Alternatively, one may compute new residuals by (3) using bias-corrected coefficients (6), and redo Step 2-3 and 2-4 using random draws from new residuals.

Method 1 and Method 2 both ignore the variability of the estimated threshold value since $\hat{\gamma}$ is not re-estimated using bootstrap replicates. Method 3, on the other hand, explicitly takes into account the variability of $\hat{\gamma}$. The algorithm of Method 3 is as follows.

Method 3

Step 3-1. The same as Step 1-1.

Step 3-2. use y_t^* and estimate the *bootstrap* threshold value $\hat{\gamma}^*$ from (2). Then re-estimate Model (1) using y_t^* and $\hat{\gamma}^*$, and obtain bootstrap coefficients $\hat{\beta}_{ij}^{\gamma^*}$. Next compute a bootstrap h-step-ahead future value $\hat{y}_t^{\gamma^*} = y_t, t = n, n-1, \ldots, n-p+1$, and

$$\hat{y}_{t+h}^{\gamma^*} = \left(\hat{\beta}_{10}^{\gamma^*} + \sum_{j=1}^{p} \hat{\beta}_{1j}^{\gamma^*} \hat{y}_{t+h-j}^{\gamma^*}\right) 1(\hat{y}_{t+h-1}^{\gamma^*} > \hat{\gamma}^*) + \left(\hat{\beta}_{20}^{\gamma^*} + \sum_{j=1}^{p} \hat{\beta}_{2j}^{\gamma^*} \hat{y}_{t+h-j}^{\gamma^*}\right) 1(\hat{y}_{t+h-1}^{\gamma^*} \leq \hat{\gamma}^*) + e_{t+h}^*,$$

where e_{t+h}^* is a random draw from $F_{\hat{e}}$.

Step 3-3. Repeat Step 3-1 and 3-2 B times and obtain a series of bootstrap future values $\left\{\hat{y}_{t+h}^{\gamma^*}(i)\right\}_{i=1}^{B}$. The Method-3 h-step-ahead BPI is given by

$$\texttt{BPI3} = [\hat{y}_{t+h}^{.025\gamma}, \hat{y}_{t+h}^{.975\gamma}] \tag{8}$$

where $\hat{y}_{t+h}^{.025\gamma}$ and $\hat{y}_{t+h}^{.975\gamma}$ are the 2.5th and 97.5th percentiles of the empirical cdf of $\left\{\hat{y}_{t+h}^{\gamma^*}(i)\right\}_{i=1}^{B}$.

Note that the threshold value is re-estimated in Step 3-2. To ease computation Method 3 does not correct the bias of autoregressive coefficients, though it is straightforward to add the bias-correcting procedure. Bias-correcting $\hat{\gamma}$ is unnecessary because of its super consistency.

Two-Sample Bootstrap Prediction Intervals of TAR Models

Model (1) assumes a regime-invariant distribution function for e_t. Then bootstrapping model (1) belongs to what is called "one-sample problems" in Efron and Tibshirani (1993). More generally, we can allow for regime-dependent errors and write the generalized model as

$$y_t = \left(\beta_{10} + \sum_{j=1}^{p} \beta_{1j} y_{t-j} + e_{1t}\right) 1(y_{t-1} > \gamma) + \left(\beta_{20} + \sum_{j=1}^{p} \beta_{2j} y_{t-j} + e_{2t}\right) 1(y_{t-1} \leq \gamma), \quad (9)$$

where $e_{1t} \sim \texttt{iid}(0, \sigma_1^2)$ and $e_{2t} \sim \texttt{iid}(0, \sigma_2^2)$. Let F_{e1} and F_{e2} be the distribution functions for e_{1t} and e_{2t}. Then Model (9) makes it possible $F_{e1} \neq F_{e2}$. Method 4 explicitly takes into account this possible heterogeneity by bootstrapping separately two samples of residuals.

Method 4

Step 4-1. Estimate $\hat{\gamma}$ by (2) and define regime 1 and 2 accordingly. Then fit Model (1) using $\hat{\gamma}$ and compute the residual $\hat{e}_t$ by (3). Collect observations of $\{\hat{e}_t\}_{t=p+1}^{n}$ in regime 1 as the series of $\{\hat{e}_{1t}\}_{t=1}^{n1}$, and in regime 2 $\{\hat{e}_{2t}\}_{t=1}^{n2}$, where $n1 + n2 = n - p$.

Step 4-2. Let $F_{\hat{e}1}$ and $F_{\hat{e}2}$ denote the empirical cdfs of $\hat{e}_{1t}$ and $\hat{e}_{2t}$ computed in Step 4-1. Then generate the bootstrap replicate of y_t as $y_t^{t*} = y_t, t = 1, \ldots, p$, and

$$y_t^{t*} = \left(\hat{\beta}_{10} + \sum_{j=1}^{p} \hat{\beta}_{1j} y_{t-j}^{t*} + e_{1t}^{*}\right) 1(y_{t-1}^{t*} > \hat{\gamma}) + \left(\hat{\beta}_{20} + \sum_{j=1}^{p} \hat{\beta}_{2j} y_{t-j}^{t*} + e_{2t}^{*}\right) 1(y_{t-1}^{t*} \leq \hat{\gamma}), (t > p)$$

where e_{1t}^{*} and e_{2t}^{*} are random draws from $F_{\hat{e}1}$ and $F_{\hat{e}2}$ respectively.

Step 4-3. Re-estimate Model (1) using y_t^{t*} and $\hat{\gamma}$, and obtain bootstrap coefficients $\hat{\beta}_{ij}^{t*}$. Then

compute a bootstrap h-step-ahead future value: $\hat{y}_t^{t*} = y_t, t = n, n-1, \ldots, n-p+1$,

$$\hat{y}_{t+h}^{t*} = \left(\hat{\beta}_{10}^{t*} + \sum_{j=1}^{p} \hat{\beta}_{1j}^{t*} \hat{y}_{t+h-j}^{t*} + e_{1t+h}^{*} \right) 1(\hat{y}_{t+h-1}^{t*} > \hat{\gamma}) + \left(\hat{\beta}_{20}^{t*} + \sum_{j=1}^{p} \hat{\beta}_{2j}^{t*} \hat{y}_{t+h-j}^{t*} + e_{2t+h}^{*} \right) 1(\hat{y}_{t+h-1}^{t*} \leq \hat{\gamma}),$$

where e_{1t+h}^{*} and e_{2t+h}^{*} are random draws from $F_{\hat{e}1}$ and $F_{\hat{e}2}$ respectively.

Step 4-4. Repeat Step 4-2 and 4-3 B times and obtain a series of bootstrap future values $\left\{\hat{y}_{t+h}^{t*}(i)\right\}_{i=1}^{B}$, where i indexes the number of resampling. The Method-4 h-step-ahead BPI is given by

$$\texttt{BPI4} = [\hat{y}_{t+h}^{.025\texttt{two}}, \hat{y}_{t+h}^{.975\texttt{two}}] \tag{10}$$

where $\hat{y}_{t+h}^{.025\texttt{two}}$ and $\hat{y}_{t+h}^{.975\texttt{two}}$ are the 2.5th and 97.5th percentiles of the empirical cdf of $\left\{\hat{y}_{t+h}^{t*}(i)\right\}_{i=1}^{B}$.

It is easy to modify Method 4 to account for the variability of $\hat{\gamma}$ and correct the bias of autoregressive coefficients.

Simulation

This section compares the performance of BPIs by Monte Carlo simulation. Following Thombs and Schucany (1990), the criterion of comparison is the average coverage rate computed as

$$m^{-1} \sum_{i=1}^{m} 1\left(y_{t+h} \in \texttt{PI}\right), \tag{11}$$

where $m = 100$ for each replication and PI denotes BPI1 (5), BPI2 (7), BPI3 (8) and BPI4 (10). The forecast horizon h ranges from 1 to 8 (no qualitative changes in simulation results are found when h ranges from 1 to 12, or larger value). The number of Monte Carlo simulations (replications) is set to 500. The number of resampling is $B = 999$ for Method 1,

Method 3 and Method 4. For Method 2, the number of resampling is $C = 200$ for Step 2-2, and $B = 999$ for Step 2-5. The nominal coverage rate is 0.95. The method that produces prediction intervals with the average coverage rate closest to 0.95 is deemed the best method.

We first consider the data generating process with homogeneous (regime-invariant) errors:

$$y_t = c_1 y_{t-1} 1(y_{t-1} > \gamma) + c_2 y_{t-1} 1(y_{t-1} \leq \gamma) + e_t, (t = 1, \ldots, n), \tag{12}$$

where the threshold effect (or magnitude of regime-switching) is measured by coefficients c_1 and c_2. The number of observations subject to regime switching is controlled by γ. Figures 2, 3, 4, 5 plot the average coverage rates of BPI1, BPI2 and BPI3 against the forecast horizon h (this is a "one-sample" problem so we ignore BPI4). A note about the key in those figures: the number after the underlining sign indexes graphs. For example, in Figure 2 BPI1_1 denotes the coverage rate of BPI1 in the first graph, and BPI3_2 the coverage rate of BPI3 in the second graph, and etc.

First we let $c1 = 0.2, c2 = 0.8, \gamma = 0.0, n = 60$. Figure 2 shows the effect of various distributions of e_t on the coverage rate. We simulate errors using the standard normal distribution, the Student-T distribution with 5 degrees of freedom, and the chi-squared distribution with 4 degree of freedom. The Student-T and chi-squared distributions are fat tailed and skewed, respectively. Following Thombs and Schucany (1990) the non-normal errors are standardized prior to simulation.

We have four findings from Figure 2. First we see no severe distortion in coverage rates: the coverage rates of BPIs are bounded between 0.94 and 0.96 in most cases. This means bootstrap prediction intervals work generally well with various error distributions. Second, the coverage rate of BPI1 is always the lowest one among the three methods. This fact comes as no surprise since BPI1 ignores the variability of estimated threshold value, $\hat{\gamma}$. On the other hand, the bias-corrected BPI2 has the highest coverage rate except when $h = 1$.

The wide BPI2 is consistent with Efron and Tibshirani (1993), which point out that more variability is introduced by bias-corrected statistics. In this case BPI2 is wider than BPI1 and BPI3 by using bias-corrected autoregressive coefficients. By accounting for variability of $\hat{\gamma}$ but without correcting the autoregressive bias, BPI3 yields the coverage in the middle. For practitioners the lesson is that BPI2 and BPI3 may be more conservative than BPI1 in terms of coverage rates. The third finding is, the skewed chi-squared distribution causes more coverage-distortion (by shifting coverage lines up further) than the fat-tailed T distribution, while the latter seems not to worsen BPIs very much. Finally, we find that as forecast horizon h increases, the coverage rates of all BPIs increase as well. Again this result is intuitive since long-term forecast intrinsically involves more uncertainty than short-term forecast.

Figure 3 shows how regime-varying autoregressive coefficients affect coverage rates with $c_1 = 0.2, 0.8, 1.0$ and $c2 = 0.8, \gamma = 0.0, n = 60, e_t \sim \texttt{iidn}(0, 1)$. The threshold effect is present when $c_1 = 0.2 \neq c_2$, and the results are basically the same as the left panel of Figure 2. Threshold effect disappears (and a TAR model reduces to a linear AR model) when $c_1 = c_2 = 0.8$. In this case, BPI1 and BPI3 suffer more under-nominal distortion than BPI2. Hence it pays to apply the bias-correction procedure for the linear model, a result in line with Kim (2001) and Kilian (1998). Something interesting happens when the data are nonstationary in the regime with $c_1 = 1.0$. Now BPI1 suffers severe under-nominal distortion, with coverage rate declining monotonically with forecast horizon. BPI3 also has decreasing coverage rate, though less severe than BPI1. The most stable (though slightly above nominal level) coverage rate is produced by BPI2. Based on these findings, BPI2 is recommended when the threshold effect is marginal or when data are possibly nonstationary.

Figure 4 illustrates how the frequency of regime-switching affects coverage rates with varying γ and $c_1 = 0.2, c2 = 0.8, n = 60, e_t \sim \texttt{iidn}(0, 1)$. As γ increases there is less and less likelihood for regime-switching, and so more and more observations stay in one regime. In the limit as $\gamma \to \infty$ the threshold autoregression is reduced to a linear autoregression. The

performance of BPIs reflects this fact. When $\gamma = 0.0$, regime-switching occurs frequently and the graph looks almost the same as the left panel of Figure 2. When $\gamma = 1.0$ regime-switching becomes less likely (and the model is more like a linear model) and so the performance of BPI2 is the best (without obvious under-nominal distortion) among the three methods. The key message is that using BPI2 is a good idea where there is a small number of observations subject to regime-switching.

Figure 5 investigates the effect of sample sizes on coverage rates of BPIs, with $n = 50, 100, 150$ and $c_1 = 0.2, c2 = 0.8, \gamma = 0.0, e_t \sim \texttt{iidn}(0, 1)$. The increasing sample size is seen to improve the coverage rate of all BPIs in two ways. First, as n rises the coverage rate gets closer to the nominal level 0.95. Meanwhile, in large sample ($n = 150$) as h increases the coverage rate increases more slowly than in the small sample ($n = 50$). For instance, as h rises, BPI2 increases from 0.940 to above 0.955 when $n = 50$, while only increases from 0.945 to 0.955 when $n = 150$.

Next we consider the following TAR model with regime-dependent errors to investigate the performance of BPI4:

$$y_t = (c_1 y_{t-1} + e_{1t}) 1(y_{t-1} > \gamma) + (c_2 y_{t-1} + e_{2t}) 1(y_{t-1} \leq \gamma), (t = 1, \ldots, n), \tag{13}$$

where $e_{1t} \sim \texttt{iidn}(0, 1)$ and $e_{2t} \sim \texttt{iidn}(0, s^2), (s = 0.5, 1.0, 2.0)$ are (possibly) heteroskedastic normal errors. Figure 6 only compares BPI1 and BPI4 with $c_1 = 0.2, c2 = 0.8, n = 60, \gamma = 0.0$. We do not consider BPI2 and BPI3 here in order to focus on the difference between "one-sample bootstrap" and "two-sample bootstrap." The readers are reminded again that it is straightforward to modify BPI4 so that variability of $\hat{\gamma}$ and autoregressive bias can be taken care of.

First of all, Figure 6 shows that both coverage rates of BPI1 and BPI4 are below the nominal level, which makes sense because both methods ignore the variability of $\hat{\gamma}$. Second,

it is shown that BPI4 has less under-nominal coverage distortion than BPI1 in most cases. The exception is when $s = 0.5$, and when h is small. Nevertheless, it is constructive to emphasize that BPI4 outperforms BPI1 when errors are homoskedastic ($s = 1.0$). So using two-sample bootstrap loses nothing related to one-sample bootstrap when heteroskedasticity is uncertain. In addition, by comparing the three panels of Figure 6, we see that the position of the coverage line for BPI4 is relatively fixed, whereas the coverage line for BPI1 keeps shifting down as s rises. In light of this, loosely speaking, BPI4 is "heteroskedasticity-robust" while BPI1 is not.

Figure 6 simulates e_{2t} using standard normal distributions. Figure 7 instead simulates e_{2t} using the standard normal distribution, the Student-T distribution with 5 degrees of freedom, and the chi-squared distribution with 4 degree of freedom. Now e_{1t} and e_{2t} are heterogeneous, not just heteroskedastic. The findings from Figure 7 are more favorable to BPI4 than Figure 6. With normal errors, the graph looks the same as the middle panel of Figure 6. The chi-squared distribution tends to shift BPI4 up further (and cause less under-nominal distortion) than BPI1, so does the T distribution. Overall, BPI4 outperforms BPI1 with various distributions, and the gain of using two-sample bootstrap increases with the forecast horizon. For short term forecast, BPI1 may do better than BPI4 thanks to its relatively simple algorithm.

To summarize, the key findings of simulation are following: (1) BPIs perform generally well. (2) It is necessary to account for the sampling variability of estimated threshold value in finite sample even if it is asymptotically super-consistent. (3) The bias-correction in the bootstrap-after-bootstrap procedure can generate better prediction intervals when the threshold effect is minimal, when data are possibly nonstationary, and when the number of observations subject to regime-switching is small. (4) Two-sample bootstrap, which separately resamples residuals in two regimes, are necessary especially when errors are regime-dependent and when the focus is on long term forecast.

Conclusion

This paper considers four methods of constructing bootstrap prediction intervals for TAR models. The Method 1 is the simplest because it only accounts for the variability of estimated autoregressive coefficients. Method 2 corrects the finite-sample bias of autoregressive coefficients. Method 3 takes into account the variability of estimated threshold values. Method 4 resamples residuals in separate regimes. The main finding of simulation is that bootstrap prediction intervals perform generally well. Method 2 yields better prediction intervals under certain circumstances. The two-sample bootstrap prediction intervals outperform one-sample bootstrap prediction intervals when errors are regime-dependent and when forecast horizon is long.

References

Chan, K. S. (1990). Testing for threshold autoregression. *The Annals of Statistics*, 18, 1886–1894.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, 21, 520–533.

Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business & Economic Statistics*, 11, 121–135.

Efron, B. and Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54–75.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. London: Chapman and Hall.

Granger, C. and Newbold, P. (1986). *Forecasting Economic Time Series*. San Diego: Academic Press.

Grigoletto, M. (1998). Bootstrap prediction intervals for autoregressions: some alternatives. *International Journal of Forecasting*, 14, 447–456.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68, 575–603.

Jarque, C. M. and Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55, 163–172.

Kilian, L. (1998). Small sample confidence intervals for impulse response functions. *The Review of Economics and Statistics*, 80, 218–230.

Kim, J. (1999). Asymptotic and bootstrap prediction regions for vector autoregression. *International Journal of Forecasting*, 15, 393–403.

Kim, J. (2001). bootstrap-after-bootstrap prediction intervals for autoregressive models. *Journal of Business & Economic Statistics*, 19, 117–128.

Kim, J. (2002). Bootstrap prediction intervals for autoregressive models of unknown or infinite lag order. *Journal of Forecasting*, 21, 265–280.

Kim, J. (2004). Bias-corrected bootstrap prediction regions for vector autoregression. *Journal of Forecasting*, 23, 141–154.

Maekawa, K. (1987). Finite sample properties of several predictors from an autoregressive model. *Econometric Theory*, 3, 359–370.

Masarotto, G. (1990). bootstrap prediction intervals for autoregressions. *International Journal of Forecasting*, 6, 229–239.

Shaman, P. and Stine, R. A. (1988). The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association*, 83, 842–848.

Thombs, L. A. and Schucany, W. R. (1990). bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, 85, 486–492.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. New York: Springer-Verlag.

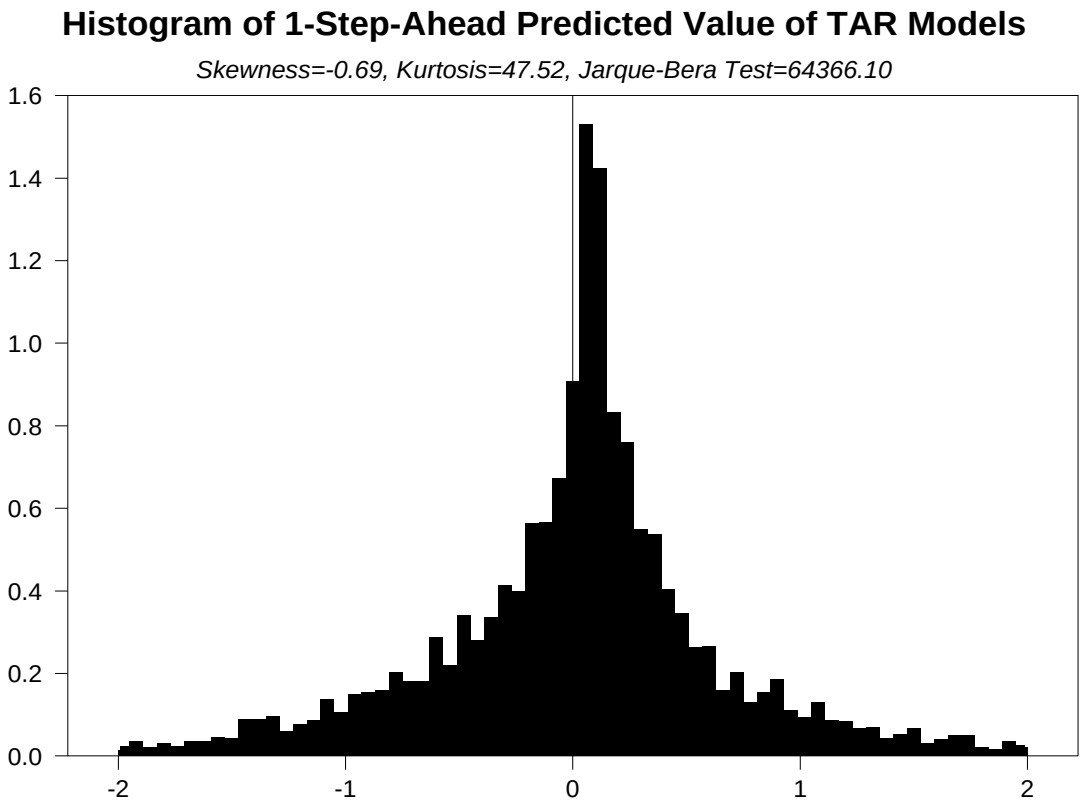

Figure 1: Histogram of Standardized 1-Step-Ahead Predicted Value of TAR Models

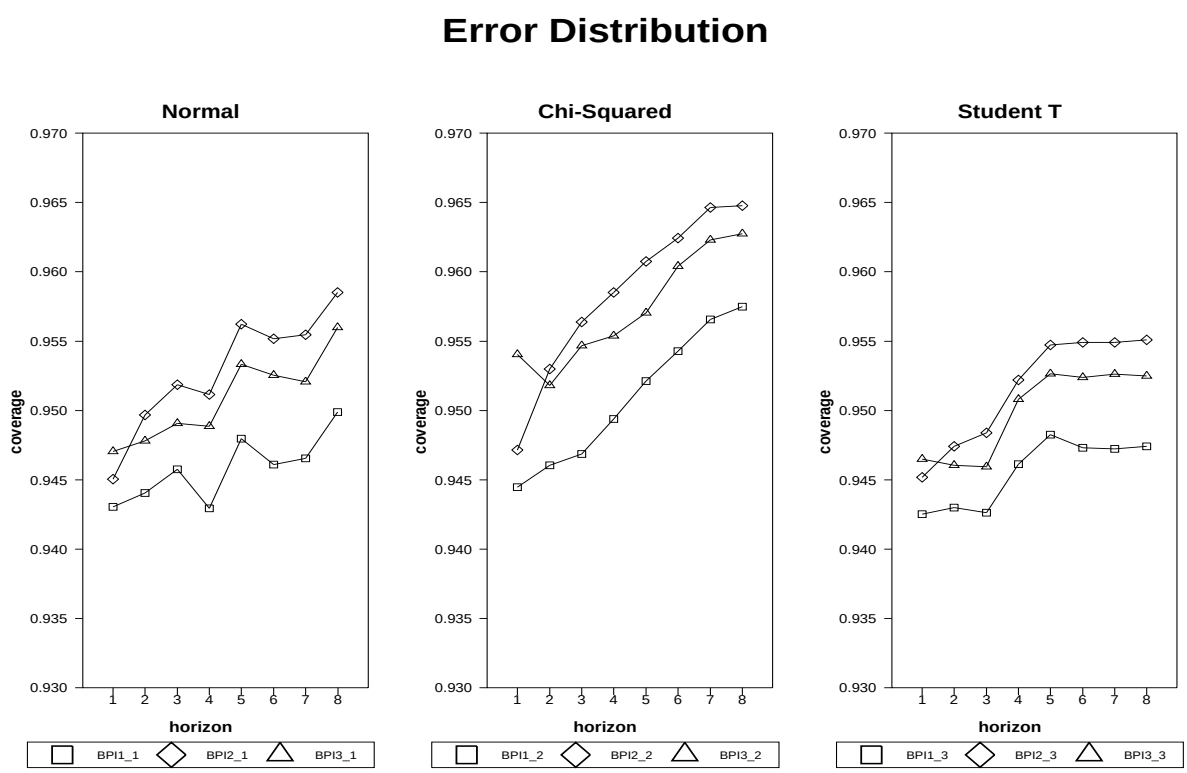

Figure 2: Average Coverage Rate of BPI for Different Error Distributions

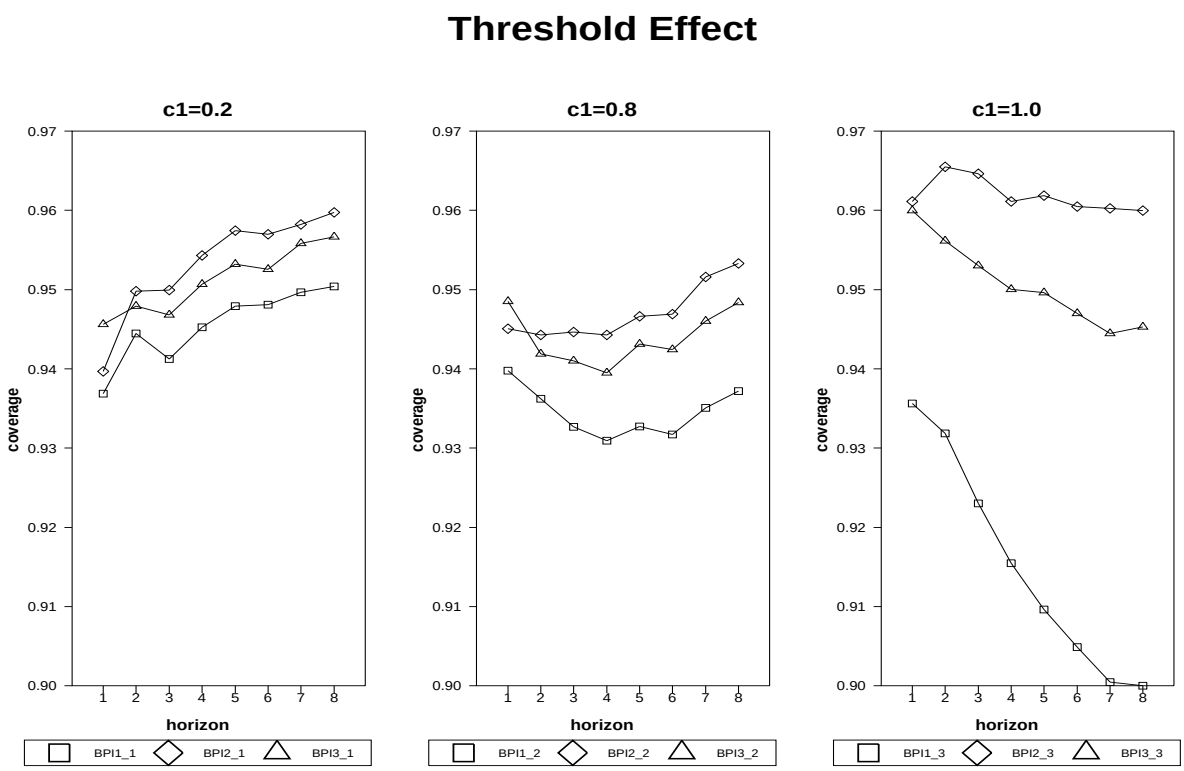

Figure 3: Average Coverage Rate of BPI for Different Autoregressive Coefficients

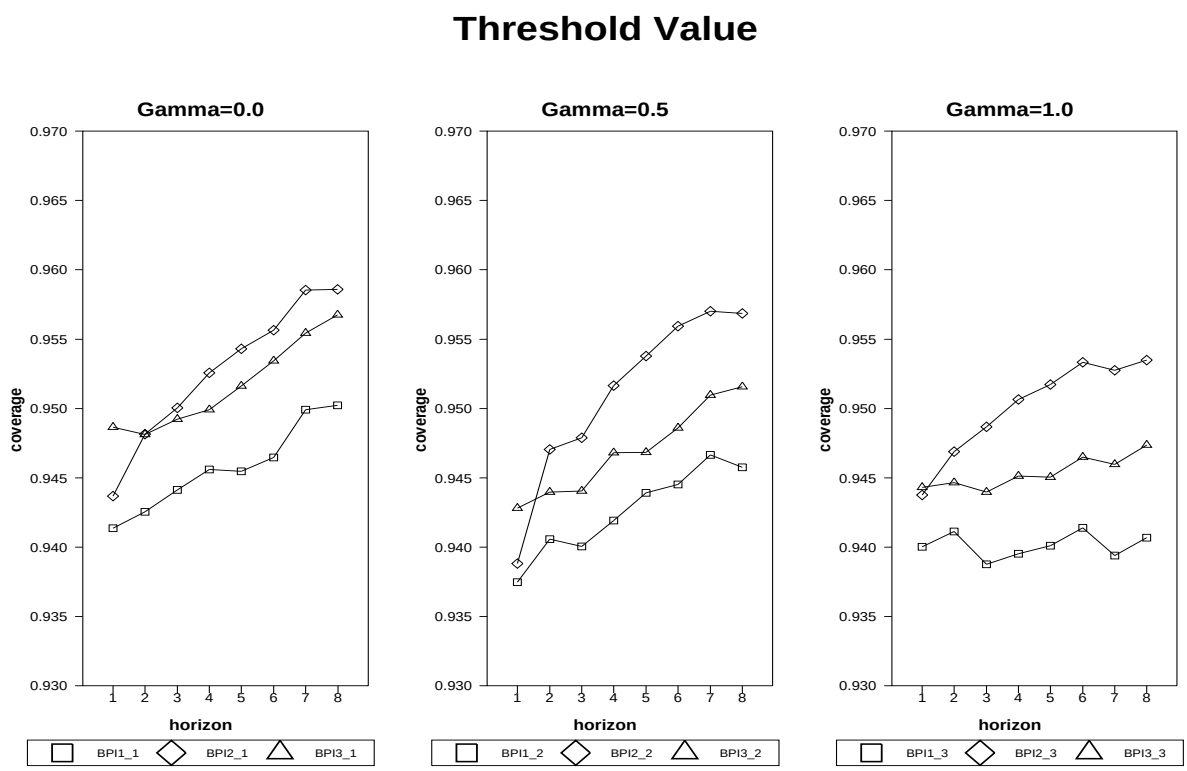

Figure 4: Average Coverage Rate of BPI for Different Threshold Values

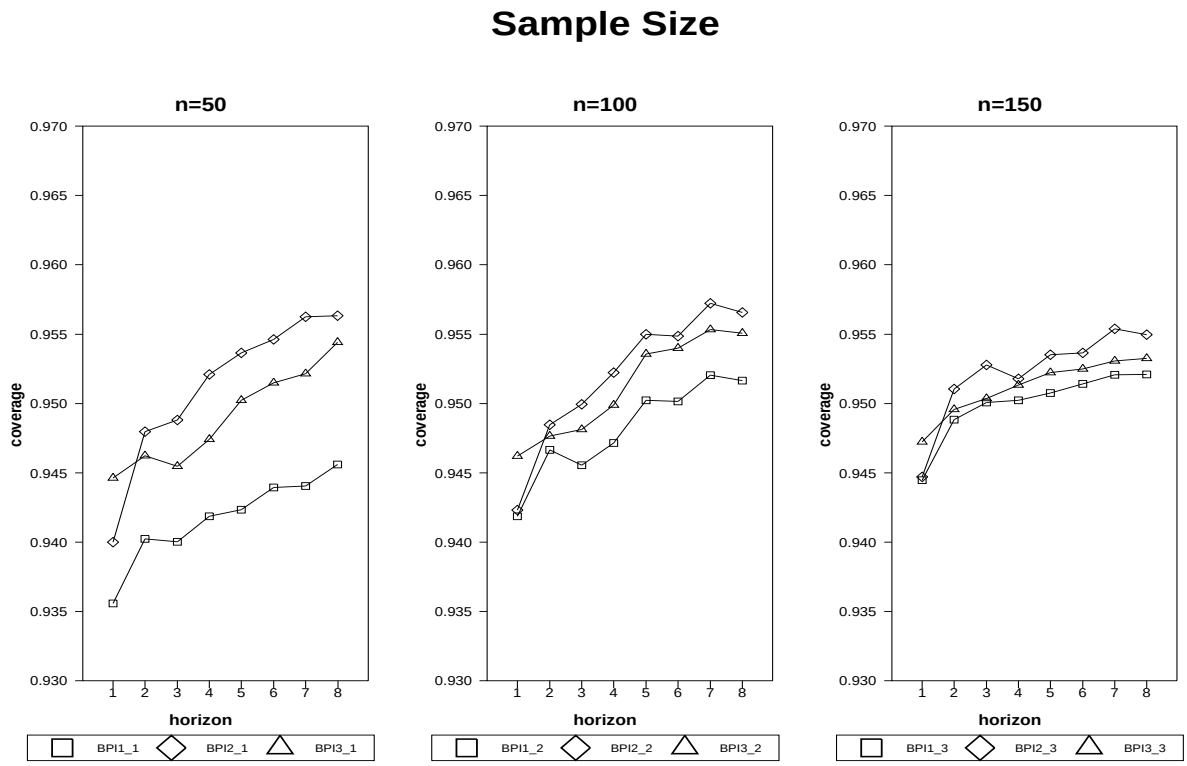

Figure 5: Average Coverage Rate of BPI for Different Sample Sizes

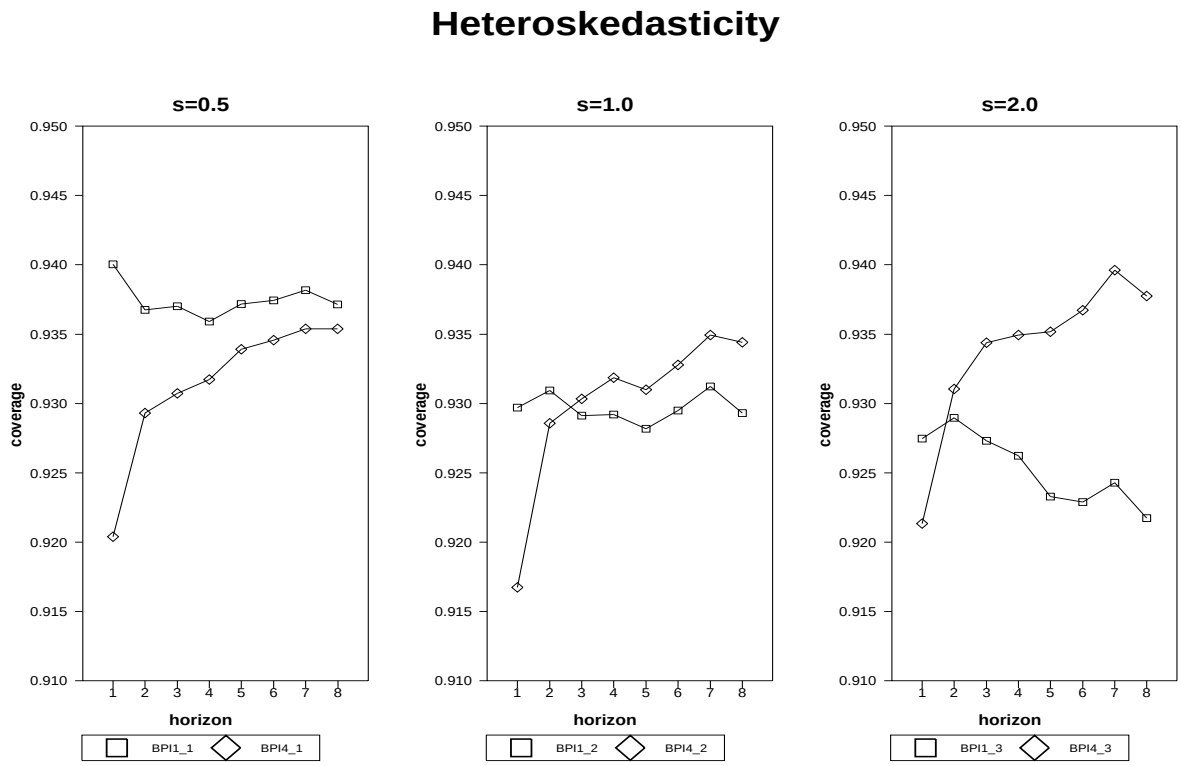

Figure 6: Average Coverage Rate of BPI for Heteroskedastic Errors

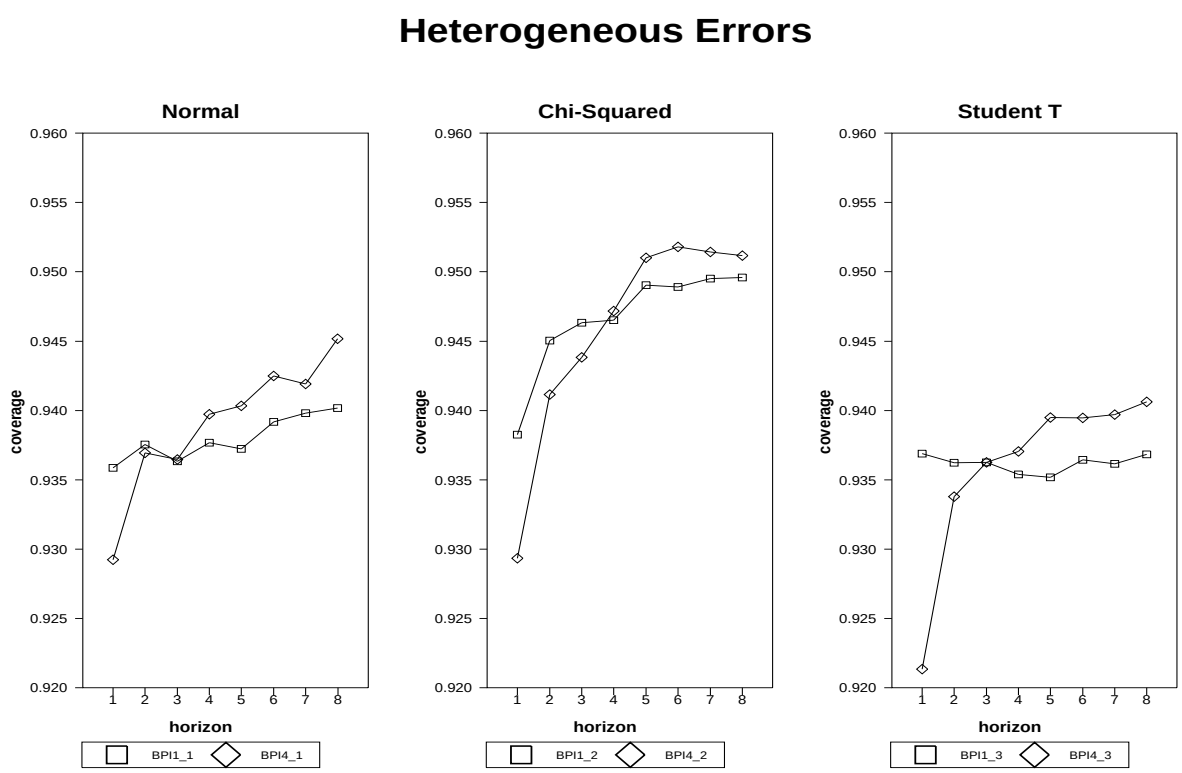

Figure 7: Average Coverage Rate of BPI for Heterogeneous Errors