Journal of Forecasting
J. Forecast. (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/for.1150

Bias-Corrected Bootstrap Prediction Intervals for Autoregressive Model: New Alternatives with Applications to Tourism Forecasting

JAE H. KIM,[1]* HAIYAN SONG[2] AND KEVIN K. F. WONG[2]
[1] *School of Economics and Finance, La Trobe University, Bundoora, Victoria, Australia*
[2] *School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hong Kong*

ABSTRACT
This paper proposes the use of the bias-corrected bootstrap for interval forecasting of an autoregressive time series with an arbitrary number of deterministic components. We use the bias-corrected bootstrap based on two alternative bias-correction methods: the bootstrap and an analytic formula based on asymptotic expansion. We also propose a new stationarity-correction method, based on stable spectral factorization, as an alternative to Kilian's method exclusively used in past studies. A Monte Carlo experiment is conducted to compare small-sample properties of prediction intervals. The results show that the bias-corrected bootstrap prediction intervals proposed in this paper exhibit desirable small-sample properties. It is also found that the bootstrap bias-corrected prediction intervals based on stable spectral factorization are tighter and more stable than those based on Kilian's stationarity-correction. The proposed methods are applied to interval forecasting for the number of tourist arrivals in Hong Kong. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS bias-correction; stationarity-correction; time series; tourist arrivals

INTRODUCTION

This paper is concerned with the bias-corrected bootstrap for interval forecasting of an autoregressive (AR) time series. Since the seminal work of Thombs and Schucany (1990), much attention has been paid to bootstrapping prediction intervals for AR models. In particular, application of the bias-corrected bootstrap of Kilian (1998a,b) has seen a great success, evidenced by the works of Clements and Taylor (2001), Kim (2001, 2004a,b), and Clements and Kim (2007). The bias-corrected bootstrap involves bias-correction in two stages of the bootstrap procedure. In the first stage, the bias-corrected estimates for AR parameters are obtained from the observed data. In the second stage, a bootstrap

*Correspondence to: Jae H. Kim, School of Economics and Finance, La Trobe University, Bundoora, VIC 3086, Australia. E-mail: J.Kim@latrobe.edu.au

Copyright © 2010 John Wiley & Sons, Ltd.

sample is generated using the bias-corrected estimates from the first stage; the bias-corrected AR parameter estimates are obtained from the bootstrap sample, and they in turn are used to generate AR forecasts. The second stage is repeated sufficiently many times to obtain a bootstrap distribution of AR forecasts, which approximates the sampling distribution of the AR forecasts.

The above-mentioned past studies, however, dealt with pure AR models,[1] neglecting the case where additional deterministic terms (such as the trend, seasonal or structural shift dummy variables) have to be included in the model. Since trend and dummy variables are widely used in practice, it is important to examine whether the bias-corrected bootstrap prediction intervals show desirable properties for an AR model with an arbitrary number of deterministic components. One may argue that the deterministic terms may be removed before the bias-corrected bootstrap is implemented to a pure AR model. However, in light of van Giersbergen's (2005) finding, additional small-sample biases due to the presence of extra deterministic terms should not be ignored. He showed that the presence of deterministic terms adds substantial small-sample bias to AR parameter estimation. In particular, he proved that the biases double (triple) when the linear (quadratic) trend is added to an AR model with an intercept. In practical applications, trend, trend breaks, seasonality, and deterministic shifts are often dominant features, and it is more desirable to include them explicitly in forecasting models. An example can be found in tourism forecasting, which will be presented later in this paper.

As a means of bias-correction, both bootstrapping and analytic bias formula are used in this paper, following Kim (2004a,b) and Clements and Kim (2007). We use analytic formula derived by Kiviet and Philips (1994), which has not previously been used in conjunction with the bias-corrected bootstrap. It is more flexible than the analytic bias formulae used in past studies, as it is applicable to an AR model with an arbitrary number of deterministic terms. In fact, it is applicable to a general autoregressive distributed lag (ARDL) model: we leave the application of the bootstrap prediction interval to an ARDL model as a future research study. The formula, however, is derived under the assumption of normal error term, which may be a restrictive condition in practice.

An important technical aspect of the past studies is the use of stationarity-correction. This is required when the parameter estimates are pushed to non-stationarity as a result of bias-correction. Past studies exclusively used Kilian's (1998a,b) method, which has been found to work well in their Monte Carlo evaluations and practical applications. However, it is an *ad hoc* adjustment which lacks a theoretical basis (see Clements and Taylor, 2001, p. 254). In this paper, we propose an alternative to Kilian's method that has a sound theoretical underpinning. It makes a correction keeping the auto-covariance generating function (or spectrum) associated with the initial bias-corrected estimates unchanged. The procedure is called stable spectral factorization, which is a special case of the method proposed by Poskitt and Salau (1993) for the vector ARMA model.[2]

As Chatfield (1993) and Christoffersen (1998) pointed out, interval forecasts are of higher value to decision makers than point forecasts. The former indicate a range of future outcomes with a prescribed level of probability, allowing for evaluation of future uncertainty and contingency planning. Although interval forecasting is widely neglected in the forecasting literature at large, we identify tourism forecasting as an area where interval forecasting can add high utility. Tourism forecasting has been an area of extensive research, and application of time series methods has been a key innovation in this area (see, for example, Witt and Witt, 1995; Song and Li, 2008). However, as we shall discuss later, all past studies have been preoccupied with point forecasting, neglecting the provision of prediction intervals.

[1] An exception is Clements and Kim (2007), who considered the AR model with a linear trend.
[2] We would like to thank Don Poskitt for bringing this method to our attention.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

An extensive Monte Carlo experiment is conducted to compare small-sample properties of alternative prediction intervals for the AR model with a number of deterministic terms. We found that both bias-corrected bootstrap prediction intervals based on bootstrap and analytic bias-corrections show desirable small-sample properties, although the former tends to slightly outperform the latter. It is also found that the new stationarity-correction based on stable spectral factorization proposed in this paper provides much tighter and more stable prediction intervals than those based on Kilian's (1998a,b) method. The application presented demonstrates that the bias-corrected bootstrap based on bootstrap bias-correction provides accurate and informative prediction intervals for tourist arrivals to Hong Kong.

The structure of the paper is as follows. The next section presents the model and the details of parameter estimation and forecasting. The third section presents alternative bias-correction and stationarity-correction methods, and the fourth section provides the details of the bias-corrected bootstrap. The fifth section presents the Monte Carlo simulation results and the sixth section empirical applications. The seventh section concludes the paper.

MODEL AND FORECASTING

We consider an AR(p) model with an arbitrary number of deterministic components, which can be written as

$$Y_t = \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{j=1}^{m} \beta_j D_{j,t} + u_t \tag{1}$$

where $D_{j,t}$ is a deterministic component that includes intercept, time trend and dummy variables, and u_t is an i.i.d. error term. The above model can be written in a matrix form as

$$Y = Z\alpha + u \tag{2}$$

where $Y = (Y_1, \ldots, Y_n)'$, $u = (u_1, \ldots, u_n)'$, and $Z = [W : D]$ is an $(n \times k)$ matrix with $k = p + m$, while W is an $(n \times p)$ matrix of lagged dependent variables and D is an $(n \times m)$ matrix of deterministic components. The vector of unknown coefficients is written as $\alpha = (\gamma : \beta)' = (\gamma_1, \ldots, \gamma_p : \beta_1, \ldots, \beta_m)'$. We assume that the AR component of the model is stationary, with all characteristic roots lying outside the unit circle. The least-squares (LS) estimators for α and σ^2 are $\hat{\alpha} \equiv (\hat{\gamma} : \hat{\beta}) \equiv (\hat{\gamma}_1, \ldots, \hat{\gamma}_p; \hat{\beta}_1, \ldots, \hat{\beta}_m) = (Z'Z)^{-1}Z'Y$ and $s^2 = e'e/(n-k)$, respectively, while $e = (e_1, \ldots, e_n)'$ denotes the vector of residuals.

Let $Y_{n+h} = Y_n(h) + u_{n+h}$, where $Y_n(h) = \sum_{i=1}^{p} \gamma_i Y_n(h-i) + \sum_{j=1}^{m} \beta_j D_{j,n+h}$ is the optimal forecast for Y_{n+h}. $Y_n(h)$ can be estimated by replacing unknowns with their LS estimates as

$$\hat{Y}_n(h) = \sum_{i=1}^{p} \hat{\gamma}_i Y_n(h-i) + \sum_{j=1}^{m} \hat{\beta}_j D_{j,n+h} \tag{3}$$

where $\hat{Y}_n(h) = Y_n(h) = Y_{n+h}$ for $h \le 0$. It can be shown that $\hat{Y}_n(h)$ is $\sqrt{n}$-consistent and asymptotically normal. Based on this, asymptotic prediction intervals using a normal approximation can be constructed. However, since asymptotic intervals are well known to perform poorly in small samples (see, for example, Thombs and Schucany, 1990), they are not considered in this paper.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

BIAS-CORRECTION AND STATIONARITY-CORRECTION

Analytic bias-correction based on asymptotic approximation

Kiviet and Phillips (1994) derived an analytic formula for the bias of $\hat{\alpha}$ to $O(n^{-1})$. That is,

$$E(\hat{\alpha}-\alpha)=B_{\hat{\alpha}}(\alpha,\sigma^2,Z)+o(n^{-1})$$

where $B_{\hat{\alpha}}(\alpha, \sigma^2, Z) \equiv (B_{\hat{\gamma}} : B_{\hat{\beta}})$ denotes the bias as a function of the unknown parameters and the data matrix, the explicit form of which is given in Theorem 1B of Kiviet and Phillips (1994). Note that this bias expression is derived under the normality assumption of u_t and fixed initial values ($Y_{1-p}, \ldots, Y_0$) of Y_t. The bias-corrected estimator for α can be obtained as

$$\hat{\alpha}_A^c=\left[\hat{\gamma}_A^c:\hat{\beta}_A^c\right]=\hat{\alpha}-B_{\hat{\alpha}}(\hat{\alpha},s^2,Z) \tag{4}$$

The estimator for the error variance σ^2, associated with $\hat{\alpha}_A^c$, is denoted as s^{2c}.

Bootstrap bias-correction

The non-parametric bootstrap based on residual resampling can be applied to model (1) to estimate the bias of $\hat{\alpha}$ to $O(n^{-1})$. First, a bootstrap sample $\{Y_t^*\}_{t=1}^n$ is generated as

$$Y_t^*=\sum_{i=1}^{p}\hat{\gamma}_i Y_{t-i}^*+\sum_{j=1}^{m}\hat{\beta}_j D_{j,t}+e_t^*$$

using $\{Y_t\}_{t=1}^p$ as starting values, where e_t^* is a random draw with replacement from $\{e_t\}_{t=1}^n$. Second, we obtain $\hat{\alpha}^* = (Z^{*\prime}Z^*)^{-1}Z^*Y^*$, the bootstrap estimator for α, where $Y^* = (Y_1^*, \ldots, Y_n^*)'$, $Z^* = [W^* : D]$, and W^* is a matrix of lagged dependent variables constructed from $\{Y_t^*\}_{t=1}^n$. The above steps are repeated B_1 times to obtain the bootstrap distribution $\{\hat{\alpha}^*(j)\}_{j=1}^{B_1}$. The bias of $\hat{\alpha}$ is estimated as $\text{Bias}(\hat{\alpha}) \equiv \text{Bias}(\hat{\gamma}, \hat{\beta}) = \bar{\alpha}^* - \hat{\alpha}$, where $\bar{\alpha}^*$ is the sample mean of $\{\hat{\alpha}^*(j)\}_{j=1}^{B_1}$. In this paper, we set $B_1 = 500$. The bootstrap bias-corrected estimator $\hat{\alpha}_B^c$ is obtained as

$$\hat{\alpha}_B^c=\hat{\alpha}-\text{Bias}(\hat{\alpha})\equiv\left[\hat{\gamma}_B^c:\hat{\beta}_B^c\right] \tag{5}$$

Stationarity-correction

Kilian's method

As mentioned earlier, it is possible that bias-correction pushes $\hat{\gamma}^c$ to the non-stationary part of the parameter space, especially when the sample size is small and the model is close to unit root non-stationarity. Kilian (1998a,b) proposed a correction which adjusts the bias-corrected estimates by shrinking the bias estimates until the condition of stationarity is satisfied. For simplicity, we use $\hat{\alpha}^c = [\hat{\gamma}^c : \hat{\beta}^c]$ as a generic notation for the bias-corrected estimator (based on either analytic or bootstrap bias-correction). In the case of bootstrap bias-correction, the procedure is conducted as follows: if $\hat{\gamma}^c$ implies non-stationarity, then let $\delta_1 = 1$, $\Delta_1 = \text{Bias}(\hat{\gamma})$(or $B_{\hat{\gamma}}$) and $\hat{\gamma}^c = \hat{\gamma} - \Delta_i$. Set $\Delta_{i+1} = \delta_i\Delta_i$, $\delta_{i+1} = \delta_i - 0.01$ for $i = 1, 2, 3, \ldots$. Iterate until $\hat{\gamma}^c$ satisfies the condition of stationarity. Once this correction is made, $\hat{\beta}^c$ is re-estimated restricting the AR coefficients to the adjusted value of $\hat{\gamma}^c$.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Stable spectral factorization

In this paper, an alternative stationarity-correction based on stable spectral factorization of Poskitt and Salau (1993) is presented. Consider the AR polynomial of the model (1):

$$\gamma(z) = 1 - \gamma_1 z - \ldots - \gamma_p z^p = \prod_{i=1}^{p}(1 - \delta_i z)$$

where δ_i are the reciprocals of the roots of the polynomial $\gamma(z)$. For stationarity, all δ_i should lie strictly inside the unit circle. The auto-covariance generating function (AGF) of model (1) is written as $\delta(\mathrm{z}) = \sigma^2/\gamma(z)\gamma(z^{-1})$. Suppose that bias-correction is given to the initial (stationary) LS estimates to obtain $\hat{\gamma}^c$.[3] The corresponding AGF is written as

$$\hat{\delta}^c(z) = \frac{s^{2c}}{\hat{\gamma}^c(z)\hat{\gamma}^c(z^{-1})} = \frac{s^{2c}}{\prod_{i=1}^{p}\left(1 - \hat{\delta}_i^c z\right)\prod_{i=1}^{p}\left(1 - \hat{\delta}_i^c z^{-1}\right)} \tag{6}$$

where $\hat{\gamma}^c(z) = 1 - \hat{\gamma}_1^c z - \ldots - \hat{\gamma}_p^c z^p = \prod_{i=1}^{p}\left(1 - \hat{\delta}_i^c z\right)$ and s^{2c} is the estimator for σ^2 associated with $\hat{\gamma}^c$. Suppose the model has become non-stationary as a result of bias-correction. Without loss of generality, we assume that the first p^* $\hat{\delta}_i^c$'s ($i = 1, \ldots, p^*$, where $p^* \le p$) lie outside the unit circle, while all others are strictly inside. The AGF is can be written as

$$\hat{\delta}^c(z) = \frac{s^{2c}}{\prod_{i=1}^{p^*}(1 - \hat{\delta}_i^c z^{-1})\prod_{i=p^*+1}^{p}(1 - \hat{\delta}_i^c z)\prod_{i=1}^{p^*}(1 - \hat{\delta}_i^c z)\prod_{i=p^*+1}^{p}(1 - \hat{\delta}_i^c z^{-1})} \tag{7}$$

The AGFs given in (6) and (7) are identical, but only the components of $\hat{\gamma}^c(z)$ that yield non-stationary roots are exchanged with the corresponding components of $\hat{\gamma}^c(z^{-1})$ with stationary roots. The AGF in (7) can be rewritten as

$$\hat{\delta}^c(z) = \frac{s^{2c}\prod_{i=1}^{p^*}\hat{\delta}_i^{c^{-2}}}{\prod_{i=1}^{p^*}\left(1 - \hat{\delta}_i^{c^{-1}} z\right)\prod_{i=p^*+1}^{p}\left(1 - \hat{\delta}_i^c z\right)\prod_{i=1}^{p^*}\left(1 - \hat{\delta}_i^{c^{-1}} z^{-1}\right)\prod_{i=p^*+1}^{p}\left(1 - \hat{\delta}_i^c z^{-1}\right)}$$

The bias-corrected estimator is recalculated as $\hat{\gamma}^c$ corresponding to

$$\prod_{i=1}^{p^*}\left(1 - \hat{\delta}_i^{c-1} z\right)\prod_{i=p^*+1}^{p}\left(1 - \hat{\delta}_i^c z\right) \tag{8}$$

with the error variance scaled by $\prod_{i=1}^{p^*}\hat{\delta}_i^{c^{-2}}$. As before, $\hat{\beta}^c$ is re-estimated, restricting the AR coefficients to the adjusted values.

[3] If initial LS estimates are non-stationary, no bias-correction is made, following Kilian (1998a,b).

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

As an example, suppose that the initial (bias-unadjusted) LS estimates for an AR(2) model gives $\hat{\gamma}(z) = 1 - 1.45z + 0.475z^2 = (1 - 0.95z)(1 - 0.5z)$ with the characteristic roots 1.05 and 2. Assume that $B_{\hat{\alpha}}(\hat{\alpha}, s^2, Z)$(or Bias($\hat{\alpha}$)) = (–0.06, 0.03). Bias-correction yields $\hat{\gamma}^c(z) = 1 - 1.51z + 0.505z^2 = (1 - 1.01z)(1 - 0.5z)$, and the model becomes non-stationary. The correction based on stable spectral factorization given in (8) yields

$$\hat{\gamma}^c(z) = (1 - 1.01^{-1}z)(1 - 0.5z) = 1 - 1.49z + 0.495z^2$$

with the error variance adjusted to $s^{2c}/1.01^2$. As a result, the model has become stationary with the characteristic roots 1.01 and 2. Kilian's correction gives

$$\hat{\gamma}^c(z) = 1 - 1.498z - 0.499z^2 = (1 - 0.998z)(1 - 0.5z) \tag{9}$$

with the characteristic roots 1.001 and 2. This example illustrates that Kilian's correction may render the model unnecessarily close to non-stationarity.

Bias-corrected bootstrap
This section details the bias-corrected bootstrap procedure, again using $\hat{\alpha}^c = [\hat{\gamma}^c : \hat{\beta}^c]$ as a generic notation. We also note that a stationarity-correction method indicates either of the two methods presented in the previous section.

Using analytic bias-correction

Stage 1
Calculate the LS estimators $\hat{\alpha}$ and s^2 for model (1) and their bias-corrected versions $\hat{\alpha}^c$ following (4) and s^{2c}, which is the estimator for σ^2 calculated from $\hat{\alpha}^c$, implementing stationarity-correction if necessary.

Stage 2
Generate a bootstrap sample $\{Y_t^*\}_{t=1}^n$ recursively as

$$Y_t^* = \sum_{i=1}^{p} \hat{\gamma}_i^c Y_{t-i}^* + \sum_{j=1}^{m} \hat{\beta}_j^c D_{j,t} + e_t^* \tag{10}$$

using $\{Y_t\}_{t=1}^p$ as starting values, where e_t^* is a random draw from $N(0, s^{2c})$. Using (Y^*, Z^*), calculate $\hat{\alpha}^*$ and s^{2*}, the bootstrap version of $\hat{\alpha}$ and s^2; and

$$\hat{\alpha}^{c*} = \left[\hat{\gamma}^{c*} : \hat{\beta}^{c*}\right] = \hat{\alpha}^* - B_{\hat{\alpha}}\left(\hat{\alpha}^*, s^{2*}, Z^*\right)$$

following (4), again implementing stationarity-correction if necessary. Then, calculate

$$Y_n^{c*}(h) = \sum_{i=1}^{p} \gamma_i^{c*} Y_n^{c*}(h-i) + \sum_{j=1}^{m} \hat{\beta}_j^{c*} D_{j,n+h} + e_{n+h}^*$$

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

where $Y_n^{c*}(h) = Y_{n+h}$ for $h \le 0$ and e_{n+h}^* is a random draw from $N(0, s^{2c})$.

Stage 3
Repeat Stage 2 B times to obtain $\{Y_n^{c*}(h; j)\}_{j=1}^{B}$, which is the bootstrap distribution for the forecasts.

Using bootstrap bias-correction
Stage 1
This is the same as Stage 1 in the previous subsection, except that bias-correction is conducted following (5). Let $\{e_t^c\}_{t=1}^n$ be the residuals calculated from $\hat{\alpha}^c$.

Stage 2
This is the same as Stage 2 in the previous subsection, except that e_t^* and e_{n+h}^* are random draws from $\{e_t^c\}_{t=1}^n$ with replacement. Here the bias-correction is conducted as $\hat{\alpha}^{c*} = \hat{\alpha}^* - \text{Bias}(\hat{\alpha})$ using the bias estimate obtained in Stage 1 as an approximation for the bias in Stage 2, following Kilian (1998a).[4] Stationarity-correction is given if necessary.

Stage 3
This is the same as in the previous subsection.

The bias-corrected bootstrap prediction interval with nominal coverage rate of $100(1 - \theta)\%$ is given by $[Y_n^*(h, \tau), Y_n^*(h, 1 - \tau)]$, where $Y_n^*(h, \tau)$ is the 100τth percentile of the bootstrap distribution $\{Y_n^*(h; i)\}_{i=1}^{B}$, and $\tau = 0.5\theta$. This is based on what is called the percentile method. Although other alternatives such as Hall's (1992) percentile method and percentile-t method are available,[5] we prefer the percentile method as the past studies have found that it performs better than the others in small samples (for details see Kim, 1999; Clements and Taylor 2001).

Further issues
In generating bootstrap sample $\{Y_t^*\}_{t=1}^n$ in Stage 2 of the bias-corrected bootstrap, the forward AR model is used as in (10). In some past studies such as Thombs and Schucany (1990) and Kim (2001, 2004a,b), the backward AR model is preferred. This is to generate bootstrap parameter estimates conditionally on the last p values of the observed series. While incorporating this conditionality is desirable in small samples, it does not affect the asymptotic validity of the bias-corrected bootstrap, since the contribution of the last p values of the time series becomes negligible in large samples. On this basis, other past studies such as Clements and Taylor (2001) and Clements and Kim (2007) preferred the use of the forward AR model, which is simpler to implement. Note that the bias-corrected bootstrap based on analytic bias-correction can be performed only with the forward AR model, because the Kiviet–Phillips (1994) formula is derived for the forward AR model and fixed pre-sample values. For the bias-corrected bootstrap based on bootstrap bias-correction, use of the backward AR model is possible although the procedure is more involved than those described in this section. In fact, simulations are also conducted for this case at an early stage of the study, but the overall results are better when the forward AR model is used. On this basis, the procedure based on the forward AR model is preferred in this paper.

It can be shown that the above bias-corrected bootstrap procedure is asymptotically valid, i.e., $Y_n^{c*}(h)$ converges to Y_{n+h} in distribution as the sample size increases, since

[4] According to Kilian (1998a), this approximation has an asymptotically negligible error rate of $O_p(n^{-3/2})$.
[5] See Efron and Tibshirani (1993) for the details of the percentile-t method.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

$$Y_n^{c*}(h) - Y_n(h) = \left(\sum_{i=1}^{p} \gamma_i^{c*} Y_n^{c*}(h-i) + \sum_{j=1}^{m} \hat{\beta}_j^{c*} D_{j,n+h} - Y_n(h) \right) + e_{n+h}^{*}$$
$$= O_p\left(n^{-1/2}\right) + O_p(1)$$

while e_{n+h}^{*} converges to u_{n+h} in distribution. A detailed proof, which is not presented for simplicity, can be given in a manner similar to that of Clements and Taylor (2001) if the forward AR model is used, and of Kim (2001, 2004a,b) if the backward AR model is used.

MONTE CARLO EXPERIMENT

Experimental design

We consider AR(1) and AR(2) models of the form

$$Y_t = \gamma_1 Y_{t-1} + \beta_1 + \beta_2 t + \beta_3 D_{1t} + \beta_4 D_{2t} + \beta_5 D_{3t} + u_t$$

and

$$Y_t = \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + \beta_1 + \beta_2 t + \beta_3 D_{1t} + \beta_4 D_{2t} + \beta_5 D_{3t} + u_t$$

where D_{1t} to D_{3t} are the quarterly seasonal dummy variables and $u_t \sim$ i.i.d. $N(0, 1)$. For the AR(1) case $\gamma_1 = a$, while for the AR(2) case $\gamma_1 = a + 0.5$, $\gamma_2 = 0.5a$, setting $a \in \{0.7, 0.9, 0.95, 0.975\}$. The coefficients of deterministic terms are fixed at $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = (1, 0.1, 2, 0, -2)$. This data generation process is typical of seasonal time series frequently encountered in practice. The sample sizes n considered are 50 and 100. The prediction intervals with nominal coverage of 80% and 95% are evaluated. The number of Monte Carlo trials is set to 1000, as is the number of bootstrap iterations B.

To calculate the coverage rates, 1000 future values are generated conditionally on the last p values of the series $(Y_{n-p+1}, \ldots, Y_n)$, following the data generation processes given above. The coverage rate refers to the proportion of these future values belonging to a prediction interval, following Thombs and Schucany (1990). We calculate the mean and median values of the coverage rates over Monte Carlo trials. We also calculate the length of prediction intervals in each Monte Carlo trial, using their mean and standard deviation values for comparison. We prefer a prediction interval, which provides the mean or median coverage rates closer to the nominal coverage level. When two alternative prediction intervals show similar coverage properties, we prefer the one with tighter and more stable length properties.

Monte Carlo results

We begin by comparing small sample properties of prediction intervals (PIs) based on alternative stationarity-correction methods, i.e., Kilian's method vs. the stable spectral factorization method. Figure 1 reports the mean coverage rates, mean and standard deviation of the length of PIs based on bootstrap bias-correction, for the AR(2) model with $a = 0.975$ and $n = 100$, with nominal coverage rates of 80% and 95%. The PIs based on stable spectral factorization has slightly lower coverage rates than those based on Kilian's correction. For 95% PIs, the differences in coverage rates are negligible, while 80% PIs based on stable spectral factorization show coverage rates to a degree

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

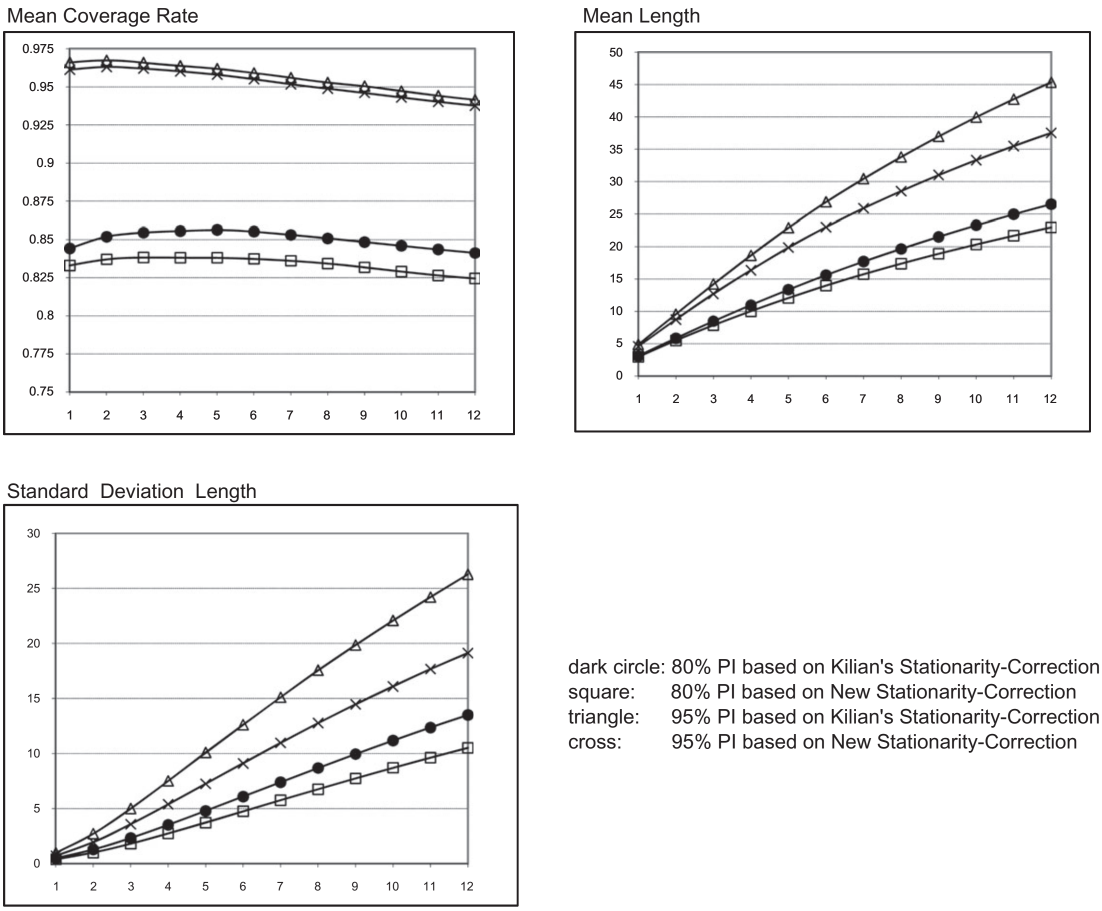

Figure 1. Comparison of bootstrap prediction intervals (AR(2) model with delta = 0.975; n = 100)

closer to the nominal level. For length properties, the PIs based on the stable spectral factorization method are much shorter and more stable, especially when the forecasting period h is long. The case with n = 50 provides similar results, although to conserve space the details are not reported here.

The above results are also evident for other cases with lower values of a, although the differences of the PIs based on two different stationarity-correction methods become smaller with decreasing values of a. This is because the stationarity-correction is not often required when the value of a is far from 1. The above results also hold for the PIs based on analytic bias-correction, although the details are not reported for simplicity. These results indicate that PIs based on Kilian's stationarity-correction are wider and more unstable than those based on stable spectral factorization, especially when the model is close to unit root non-stationarity. From this point on, only the results associated with stationarity-correction based on stable spectral factorization are reported.

Figure 2 reports the properties of the PIs based on alternative bias-correction for the AR(1) model with $a = 0.7$. When the sample is 50, the mean and median coverage rates between the two PIs are

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

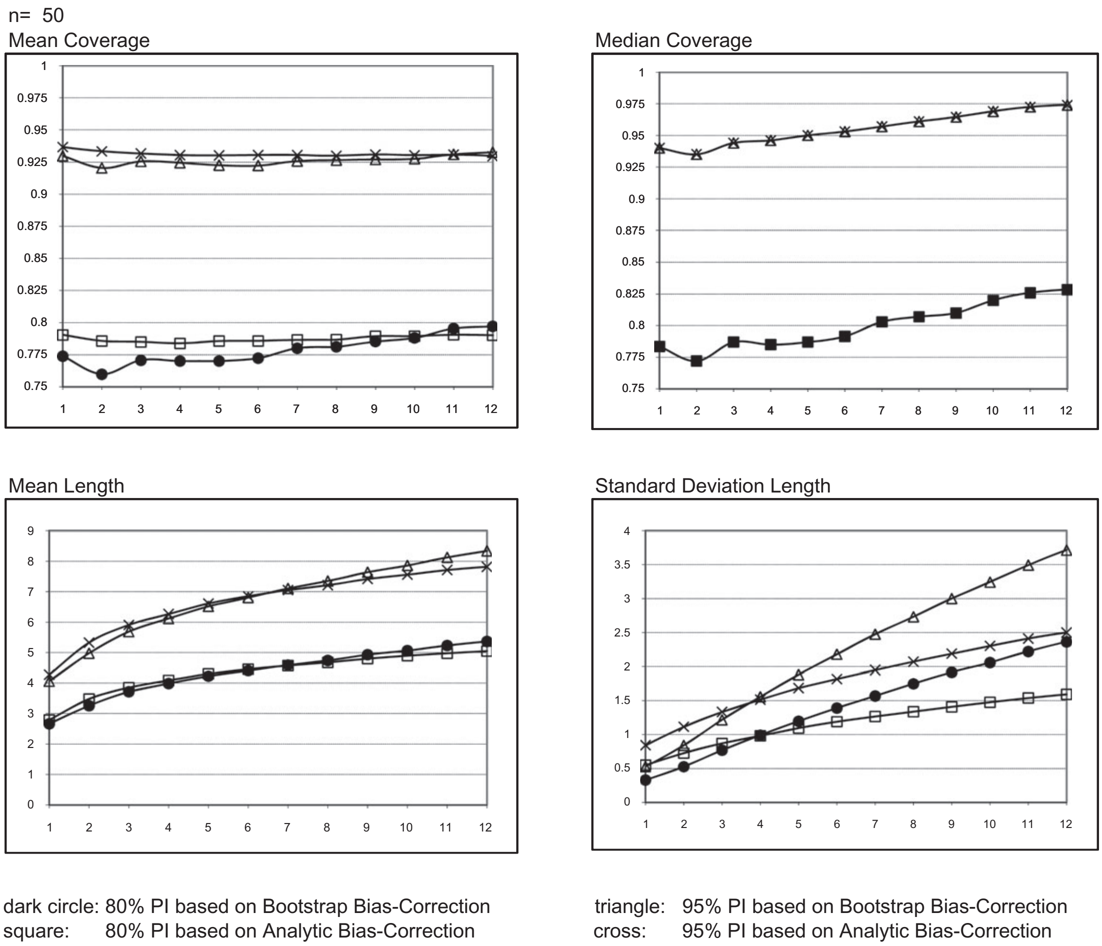

Figure 2. Comparison of alternative prediction intervals (AR(1) model with delta = 0.7)

virtually identical, although PIs based on bootstrap bias-correction slightly underestimate the nominal level. They again show virtually identical properties in terms of mean length. For the standard deviation of length, the PIs based on analytic bias-correction show lower values, especially when the value of h is long. When n = 100, the two PIs show nearly identical coverage and length properties. The values of the standard deviation of length are slightly lower when bootstrap bias-correction is used, although the differences are not substantial.

Figure 3 reports the properties of the PIs based on alternative bias-correction for the AR(1) model with $a = 0.95$. When $n = 50$, the PIs based on bootstrap bias-correction show much better performance in terms of mean coverage rates, while those based on analytic bias-correction underestimate the nominal coverage rate. If the median coverage rate is used for comparison, the reverse is the case. That is, the PIs based on bootstrap bias-correction overestimate the nominal coverage, while those based on analytic bias-correction slightly underestimate. The PIs based on analytic bias-correction are shorter on average, with less variability. Similar results are evident when n = 100, although, for simplicity, the details are not reported.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

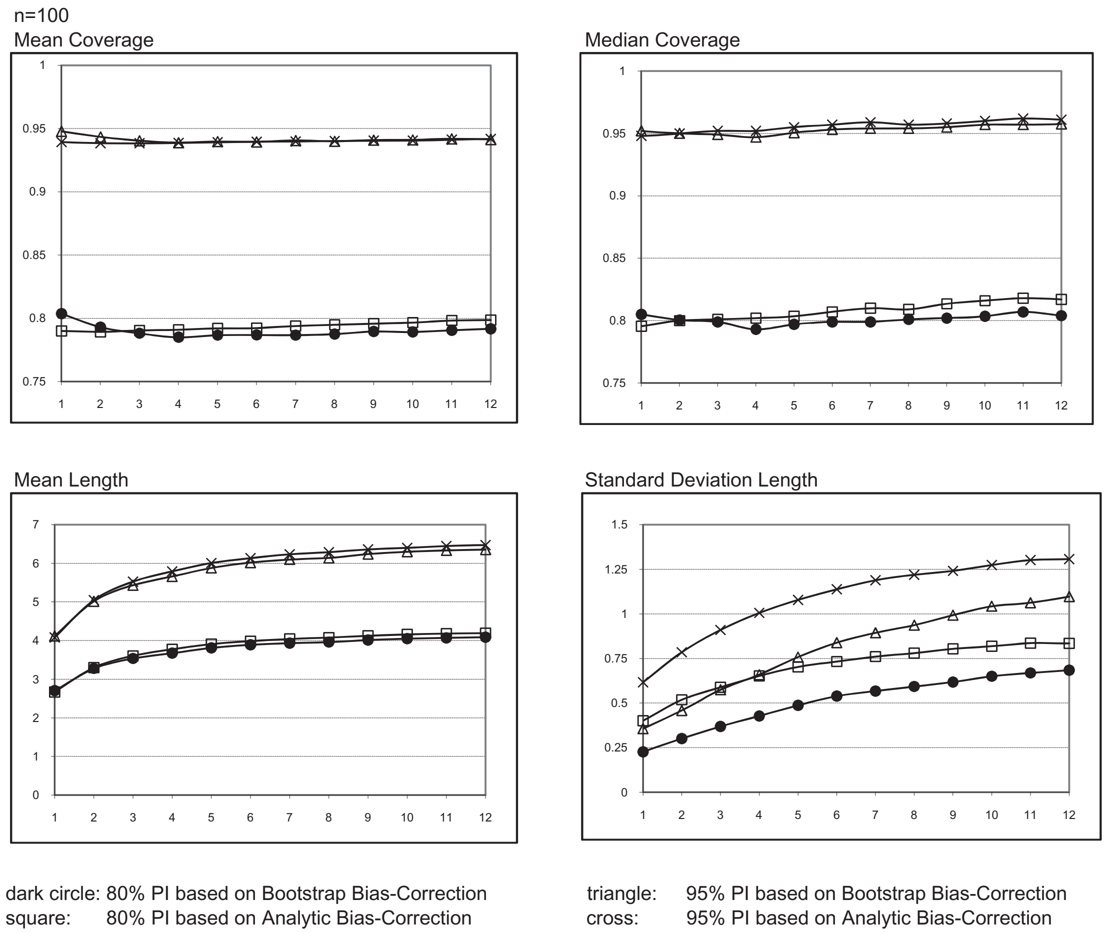

Figure 2. *Continued*

Overall, both PIs exhibit desirable properties in small samples. However, it should be noted that the PIs based on analytic bias-correction depend heavily on the assumption of a normal error term, while those based on bootstrap bias-correction do not require it. Past studies that examined small-sample properties of the latter reported that they are robust to non-normality (see, for example, Kim, 2004a,b). Hence, when the error term is non-normal, the use of bootstrap bias-correction should be preferred.

EMPIRICAL APPLICATION

The bias-corrected bootstrap discussed above is applied to interval forecasting for tourist arrivals to Hong Kong. Before we present our empirical applications, brief reviews of the literature on tourism forecasting and of the Hong Kong tourism market are presented.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

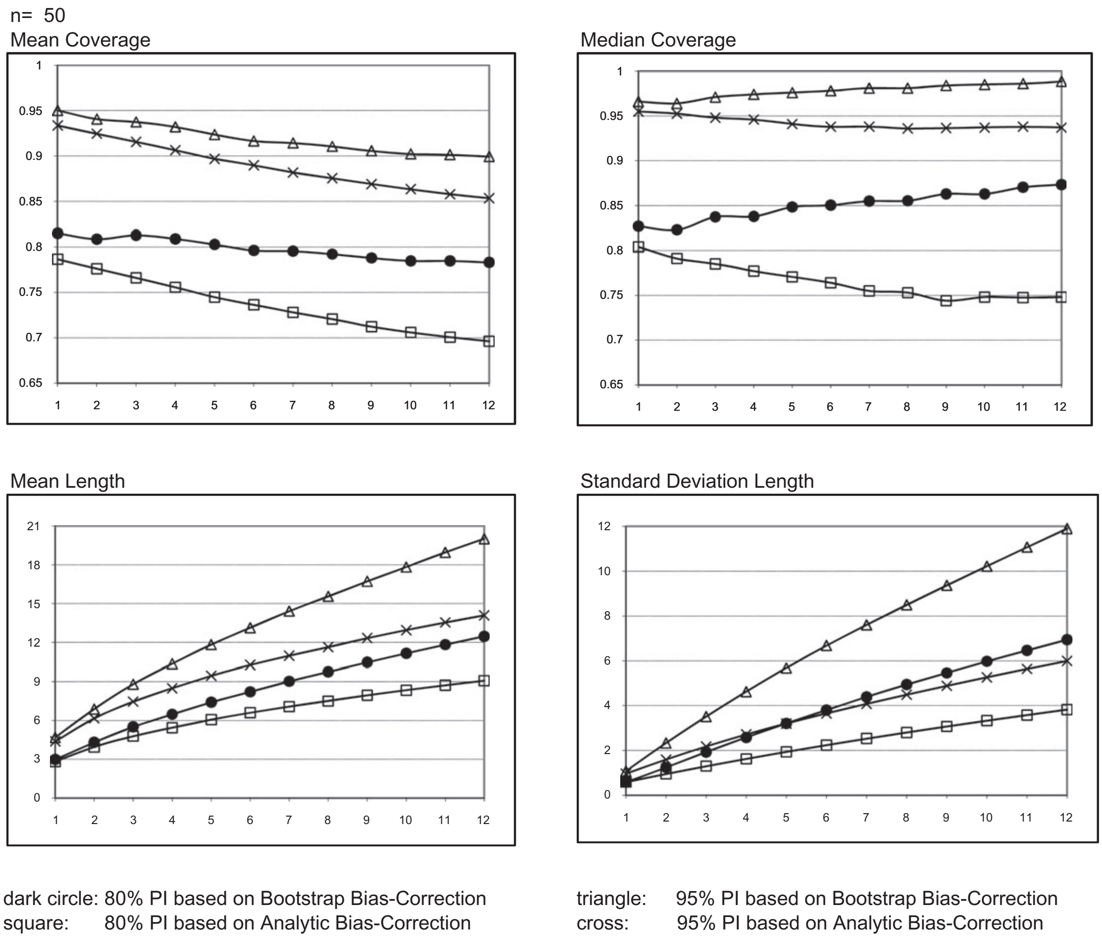

Figure 3. Comparison of alternative prediction intervals (AR(1) model with delta = 0.95)

A brief review of tourism demand forecasting literature

A large number of studies on tourism demand modeling and forecasting have been published over the past two decades. In their review of the tourism demand forecasting literature, Song and Li (2008) found that, out of the 121 published studies since 2000, 72 used time series techniques to model the demand for tourism. Among those papers that use time series approaches, the integrated autoregressive moving-average models (ARIMAs) proposed by Box and Jenkins (1970) dominate the research, as about two-third of these published studies use different versions of the ARIMA models. Depending on the frequency of the time series, either simple ARIMA or seasonal ARIMA (i.e., SARIMA) models could be used, with the latter gaining popularity over the last few years: seasonality is such a dominant feature of the tourism industry that decision makers are very much interested in the seasonal variation in tourism demand. With regard to the forecasting performance of ARIMA and SARIMA models, empirical studies present contradictory evidence. For example, Cho (2003) showed that the ARIMA model outperforms two other time series models in all cases. Goh and Law (2002)

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

suggested that the SARIMA models outperform eight other time series methods, while the non-seasonal (simple) ARIMA model's performance was above average of all forecasting models considered. However, Smeral and Wüger (2005) found that the ARIMA or SARIMA model cannot even outperform the Naïve 1 (no-change) model.

Given the inconsistency in forecasting performance of the ARIMA/SARIMA models, researchers have recently tried to improve the forecasting performance of ARIMA/SARIMA by using alternative time series approaches. One of the efforts has been to extend the univariate time series models to a multivariate dimension, and to examine whether the additional information involved in 'parallel' time series (e.g., tourism demand for a destination by a number of origin countries/regions) may contribute to the improvement of forecast accuracy. For example, Goh and Law (2002) introduced a multivariate SARIMA (i.e., MARIMA) model which includes an intervention function to capture the potential spill-over effects of the 'parallel' demand series on a particular tourism demand series. Their study showed that the multivariate SARIMA model significantly improved the forecasting performance of the simple SARIMA as well as other univariate time series models. However, in a similar attempt, Gustavsson and Nordström (2001) found that their multivariate ARIMA model could not beat its univariate counterpart. Moreover, Du Preez and Witt (2003) investigated the intervention effects of the time series models on forecasting performance within a state space framework. It was found that the multivariate state space time series model was outperformed by the simple ARIMA model. The authors argued that the unsatisfactory forecasting performance of the multivariate state space time series model was attributed mainly to the absence of a 'rich' cross-correlation structure amongst 'parallel' demand series.

The above accumulated evidence suggests that the comparison of point forecasts does not provide any decisive conclusion as to which model should be preferred for tourism demand forecasting. It may be the case that these competition forecasting models in fact generate point forecasts which are statistically no different from each other in their accuracy. In view of this, researchers need to turn their attention to provision of accurate and reliable PIs, based on a reasonably specified forecasting model, instead of attempting to find the model that generates the most accurate point forecasts.

Hong Kong tourism market

Although the tourism industry has played an ever-increasing role in the generation of wealth and employment in Hong Kong, it also faces the critical problem of creating and maintaining a sustainable competitive advantage in an environment of increased globalization and economic integration in Southeast Asia. In particular, Hong Kong will have to compete for international tourists with such destinations as mainland China, Singapore, Taiwan, Thailand and Indonesia. In order to respond effectively to this challenge, policy makers and businesses in the service sector need to know the comparative position of the industry in the region. To achieve this, the future trends of international tourism demand in Hong Kong should be properly assessed.

International tourism arrivals in Hong Kong increased from 0.93 million in 1975 to 25.3 million in 2006, representing an average annual growth rate of around 11.2%. The growth of tourist receipts has been even more spectacular, rising from HK $2975 million in 1975 to HK $117.3 billion in 2006, with an annual growth rate of about 12.5%. Japan was the biggest origin country of Hong Kong tourism in the 1970s, occupying more than 30% of the total market, followed by Southeast Asia (21.3%) and the Americas (17.8%). Europe (11.0%) and Australia and New Zealand (9.1%) ranked fourth and fifth, respectively. In the 1980s, Southeast Asia and the Americas emerged as the leading markets, replacing Japan. Since the mid 1990s, however, mainland China and Taiwan became the largest source markets for Hong Kong tourism, followed by Japan, Southeast Asia and

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Europe. According to the most recent statistics published by the Hong Kong Tourism Board (HKTB) in 2006, the following are the top tourism-generating countries/regions for Hong Kong in 2006: Australia (2.13%), China (55.36%), Japan (5.05%), Korea (2.75%), Macao (2.26%), Philippines (1.78%), Singapore (2.11%), Taiwan (8.79%), UK (1.89%) and USA (4.49%), and tourist arrivals from these 10 countries/regions accounted for about 86.66% of total tourist arrivals in Hong Kong in the same period.

International demand for Hong Kong tourism has been increasing over the past 20 years; there were, however, a few setbacks. For example, in 1997, when Hong Kong was handed over to mainland China, there was some uncertainty about Hong Kong's future. As a result, international tourist arrivals in Hong Kong, especially those from long haul-markets such as Europe and north Americas declined. However, this trend did not last for long, as the figures for tourist arrivals and tourist receipts started to pick up growth momentum from 1999. Owing to the SARS epidemic in the second quarter of 2003, there were significant downturns in total tourist arrivals and tourist receipts in Hong Kong, but the declines were short lived as tourists returned to Hong Kong shortly after the SARS epidemic was over.

Interval forecasting exercises

The forecasting exercise in this study looks at the number of monthly tourist arrivals to Hong Kong from its three major long haul markets—Canada, UK and USA—from 1990:01 to 2006:12. Figure 4 presents the time plot for the US series; those for Canada and UK are not reported since they show similar features. The data show that the time series contain the trend, seasonality and deterministic shifts. In order to take account of the effects of events such as the 1997 handover and 2003 SARS epidemic, we introduce the following dummy variables: $D_1 = 0$ before 1997:07 and 1 there afterwards; $D_2 = 1$ for 2003:04 and 0 elsewhere; $D_3 = 1$ for 2003:05 and 0 elsewhere; $D_4 = 1$ for 2003:06 and 0 elsewhere; and $D_5 = 0$ before 2003:06 and 1 afterwards. D_1 and D_5 are used to capture the permanent effects after the respective events, while D_2 to D_4 to capture the temporary effects during the period of SARS outbreak.

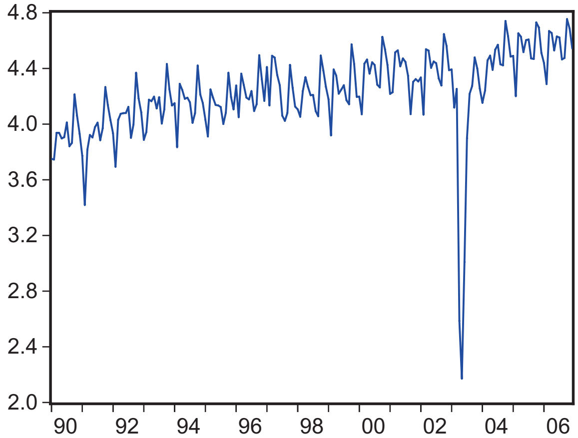

Figure 4. Time plot of tourist arrivals from the USA (in natural logs)

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

From their time plots, we find the following model appropriate for Canada and the USA:

$$Y_t = \sum_{i=1}^{p} \gamma_i Y_{t-i} + \beta_1 t + \beta_2 D_{1t} + \beta_3 t D_{1t} + \beta_4 D_{2t} + \beta_5 D_{3t} + \beta_6 D_{4t} + u_t$$

since we observe a trend shift after the handover in 1997. Monthly dummy variables are included in the model, but not written in the above equation for simplicity. For the UK, the model proposed takes the form

$$Y_t = \sum_{i=1}^{p} \gamma_i Y_{t-i} + \beta_1 t + \beta_2 D_{1t} + \beta_3 t D_{1t} + \beta_4 D_{2t} + \beta_5 D_{3t} + \beta_6 D_{4t} + \beta_7 t D_{5t} + u_t$$

as we observe an additional trend shift after the SARS outbreak. Again, monthly dummy variables are included in the model but not written in the above equation for simplicity.

We estimate the above models using the rolling window of 120 monthly observations, generating 12-step-ahead interval forecasts from each window. The first window covers the period from 1990:01 to 1999:12, with PIs generated from 2000:01 to 2000:12. The next window covers the period from 1990:02 to 2000:01, with PIs from 2000:02 to 2001:01. This continues until we reach the last window from 2005:01 to 2005:12, with PIs for the next 12 months. We do not include the period from 2003:04 to 2004:05 for forecast evaluation since the model estimation and forecasting for this period were adversely affected by the extreme observations due to the SARS outbreak, which is an event totally unpredictable by the model. This gives 70 out-of-sample PIs to compare with the actual values. For each rolling sample, the unknown AR order is estimated using AIC, with the number of bootstrap iterations set to 1000.

We present the results associated with the stationarity-correction based on stable spectral factorization, as it was found to provide more stable interval estimates in the Monte Carlo experiment. We also do not report the results associated with the PIs based on analytic bias-correction, since they showed very poor performance, with actual coverage rates much lower than the nominal level. Only those based on bootstrap bias-correction showed desirable coverage properties from this exercise. We believe the reason for this is strong non-normality of the error term for all estimated models. The Jarque–Bera test rejects the normality of error term for all three markets and for nearly the entire sample period, particularly with high excess kurtosis values. This means that the PIs based on asymptotic bias-correction are misspecified. Table I presents the actual coverage rates and mean length of PIs based on bootstrap bias-correction for 70 PIs. For both 80% and 95% PIs, the actual coverage rates are close to the nominal level. For all cases, the coverage rates are inside the 90% confidence interval for the true coverage value for 80% and 95% nominal coverage, except for the UK case, when $h = 6$. The mean length properties indicate that the intervals are tight and informative. The results indicate that the bias-corrected bootstrap PIs provide statistically adequate interval forecasts for tourist arrivals to Hong Kong from its major source markets.

CONCLUDING REMARKS

This paper makes three main contributions to the interval forecasting literature. First, in generating prediction intervals based on the bias-corrected bootstrap, we have extended the model structure so that the AR model contains an arbitrary number of deterministic components. In addition

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Table I. Actual coverage rates and mean length of the bootstrap bias-corrected prediction intervals based on bootstrap bias-correction

	80% Prediction intervals		95% Prediction intervals	
	Coverage	Mean length	Coverage	Mean length
USA				
$h = 1$	0.77	0.18	0.95	0.31
$h = 6$	0.84	0.21	0.93	0.34
$h = 12$	0.86	0.21	0.93	0.35
Canada				
$h = 1$	0.79	0.20	0.96	0.33
$h = 6$	0.83	0.27	0.96	0.44
$h = 12$	0.84	0.29	0.96	0.45
UK				
$h = 1$	0.79	0.22	0.91	0.35
$h = 6$	0.89	0.27	0.99	0.42
$h = 12$	0.80	0.27	0.96	0.42

Note: 90% confidence interval for the coverage for 80% interval is (0.72, 0.88); 90% confidence interval for the coverage of 95% interval is (0.91, 0.99). The lengths of the intervals are calculated using the natural logged scale.

to bootstrap bias-correction, we have used the analytic formula derived by Kiviet and Phillips (1994), which was not used in previous studies. Second, we have proposed a new stationarity-correction based on stable spectral factorization, which has better theoretical basis than Kilian's (1998a,b) method exclusively used in past studies. Third, we have evaluated the performance of the bias-corrected bootstrap prediction intervals for the number of tourist arrivals in Hong Kong from a set of selected markets. This kind of empirical assessment has never been done in the tourism forecasting literature.

We have conducted an extensive Monte Carlo experiment to evaluate the performance of these new alternatives. We have found that the stationarity-correction based on stable spectral factorization provides more stable and tighter prediction intervals than Kilian's correction. We also found that that both bootstrap bias-corrected prediction intervals based on analytic and bootstrap bias-correction show desirable small-sample properties. The proposed methods are applied to forecasting international tourist arrivals to Hong Kong from Canada, the UK and USA. It is found that prediction intervals based on bootstrap bias-correction provide statistically adequate interval forecasts. Those based on analytic bias-correction performed poorly, possibly due to strong non-normality in the data.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support of the Hong Kong Polytechnic University (funding no. GYF35). The GAUSS or R code for the computation of bias-corrected bootstrap prediction intervals is available from the corresponding author on request. An earlier version of this paper was presented at the International Symposium of Forecasting held in Nice, France, in 2008.

Copyright © 2010 John Wiley & Sons, Ltd.

REFERENCES

Box GEP, Jenkins GM. 1970. *Time Series Analysis, Forecasting and Control.* Holden Day: San Francisco, CA.

Chatfield C. 1993. Calculating interval forecasts. *Journal of Business and Economic Statistics* **11**: 121–135.

Cho V. 2003. A comparison of three different approaches to tourist arrival forecasting. *Tourism Management* **24**: 323–330.

Christoffersen PF. 1998. Evaluating interval forecasts. *International Economic Review* **39**: 841–862.

Clements MP, Kim JH. 2007. Bootstrap prediction intervals for autoregressive time series. *Computational Statistics and Data Analysis* **51**: 3580–3594.

Clements MP, Taylor N. 2001. Bootstrapping prediction intervals for autoregressive models. *International Journal of Forecasting* **17**: 247–267.

Du Preez J, Witt SF. 2003. Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting* **19**: 435–451.

Efron B, Tibshirani RJ. 1993. *An Introduction to the Bootstrap*. Chapman & Hall: New York.

Goh C, Law R. 2002. Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management* **23**: 499–510.

Gustavsson P, Nordström J. 2001. The impact of seasonal unit roots and vector ARMA modelling on forecasting monthly tourism flows. *Tourism Economics* **7**: 117–133.

Hall P. 1992. *The Bootstrap and Edgeworth Expansion*. Springer: New York.

Kilian L. 1998a. Small sample confidence intervals for impulse response functions. *Review of Economics and Statistics* **80**: 218–230.

Kilian L. 1998b. Confidence intervals for impulse responses under departure from normality. *Econometric Reviews* **17**: 1–29.

Kim JH. 1999. Asymptotic and bootstrap prediction regions for vector autoregression. *International Journal of Forecasting* **15**: 393–403.

Kim JH. 2001. Bootstrap-after-bootstrap prediction intervals for autoregressive models. *Journal of Business and Economic Statistics* **19**: 117–128.

Kim JH. 2004a. Bootstrap prediction intervals for autoregression using asymptotically mean-unbiased parameter estimators. *International Journal of Forecasting* **20**: 85–97.

Kim JH. 2004b. Bias-corrected bootstrap prediction regions for vector autoregression. *Journal of Forecasting* **23**: 141–154.

Kiviet JF, Phillips GDA. 1994. Bias assessment and reduction in linear error-correction models. *Journal of Econometrics* **63**: 215–243.

Poskitt DS, Salau MO. 1993. Stable spectral factorization with applications to the estimation of time series models. *Communications in Statistics, Theory and Methods* **22**: 347–367.

Smeral E, Wüger M. 2005. Does complexity matter? Methods for improving forecasting accuracy in tourism: the case of Australia. *Journal of Travel Research* **44**: 100–110.

Song H, Li G. 2008. Tourism demand modelling and forecasting: a review of literature. *Tourism Management* **29**: 203–220.

Thombs LA, Schucany WR. 1990. Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association* **85**: 486–492.

van Giersbergen NPA. 2005. On the effect of deterministic terms on the bias in stable AR models. *Economics Letters* **89**: 75–82.

Witt SF, Witt CA. 1995. Forecasting tourism demand: a review of empirical approach. *International Journal of Forecasting* **11**: 447–475.

Authors' biographies:

Jae H. Kim is a Professor of Finance in the School of Economics and Finance at La Trobe University. He has published widely in the areas of empirical finance, econometrics, and time series forecasting. He is an author of two software packages written in R: BootPR for bias-corrected time series forecasting and vrtest for testing for asset market efficiency and return predictability. After completing his Ph.D. at the University of Sydney in 1997, he has worked at James Cook University, La Trobe University, and Monash University, before returning to La Trobe in June 2009.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Haiyan Song is the Chair Professor of Tourism in the School of Hotel and Tourism Management, The Hong Kong Polytechnic University. His main research interests are in the areas of applied econometrics and forecasting with a particular focus on tourism demand modeling and forecasting. He has published in such journals as *Journal of Applied Econometrics, Journal of Development Economics, International Journal of Forecasting, Journal of Travel Research and Tourism Management*.

Kevin K. F. Wong is an Associate Professor of Tourism at The Hong Kong Polytechnic University. His main research interests are in the areas of tourism economics and tourism demand modeling and forecasting.

Authors' addresses:
Jae H. Kim, School of Economics and Finance, La Trobe University, Bundoora, Victoria, 3086, Australia.

Haiyan Song and **Kevin K. F. Wong**, School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hong Kong.

Copyright © 2010 John Wiley & Sons, Ltd.

DOI: 10.1002/for