Bootstrap for correcting the mean square error of prediction and smoothed estimates in structural models

Thiago R. dos Santos[a,1] **and Glaura C. Franco**[a]

[a]*Universidade Federal de Minas Gerais*

Abstract. It is well known that the uncertainty in the estimation of parameters produces the underestimation of the mean square error (MSE) both for in-sample and out-of-sample estimation. In the state space framework, this problem can affect confidence intervals for smoothed estimates and forecasts, which are generally built by state vector predictors that use estimated model parameters. In order to correct this problem, this paper proposes and compares parametric and nonparametric bootstrap methods based on procedures usually employed to calculate the MSE in the context of forecasting and smoothing in state space models. The comparisons are performed through an extensive Monte Carlo study which illustrates, empirically, the bias reduction in the estimation of MSE for prediction and smoothed estimates using the bootstrap approaches. The finite sample properties of the bootstrap procedures are analyzed for Gaussian and non-Gaussian assumptions of the error term. The procedures are also applied to real time series, leading to satisfactory results.

[1]Corresponding author. Thiago Rezende dos Santos. Address: Universidade Federal de Minas Gerais, Av. Antonio Carlos, 6627, Belo Horizonte, MG, 31270-901, Brazil. Tel.: +55 31 34095945

1

1 Introduction

One of the main objectives of modeling time series is to forecast future values. In classical inference, forecasting is usually performed using a plug-in approach, i.e., replacing the model parameters by their estimators. It is well known that this procedure leads to the underestimation of the prediction mean square error (PMSE), as it does not incorporate the uncertainty due to parameter estimation (see Pfeffermann and Tiller (2005), Ansley and Kohn (1986) and Yamamoto (1976)), resulting in prediction intervals with smaller widths.

In state space models, the underestimation of the PMSE can also influence inference for the state vector predictor, specially for short series. If model parameters are replaced by their estimators, the mean square error (MSE) and confidence intervals for the smoothed estimates of the state vector may be affected. The problem is worsened if the Gaussian distribution is assumed when there is no guarantee that this is the actual distribution of the error terms.

One way to incorporate the uncertainty in the parameter estimation is through asymptotic sampling of the maximum likelihood estimator (MLE) (Hamilton (1986); Quenneville and Singh (2000)). In this sense, for state space models, Quenneville and Singh (2000) proposed a modification in the methods of Hamilton (1986) and Ansley and Kohn (1986), which take into account the uncertainty of the hyperparameter estimation in a simple way, by incorporating the uncertainty in the estimation of the state vector MSE through the corrected hyperparameter estimator. A similar approach was adopted by Shephard (1993), who wrote the local level model as a normal mixed effects model in order to use the restricted maximum likelihood estimator (RMLE). Tsimikas and Ledolter (1994) presented an alternative way to build the restricted likelihood function, also using mixed effects models.

The bootstrap technique can also be employed to incorporate the uncertainty in the estimation of ψ, the hyperparameter vector. Wall and Stoffer (2002) proposed a method for the construction of the empirical distribution of the forecast error. Rodriguez and Ruiz (2009) presented a simpler nonparametric bootstrap procedure, compared to the proposal of Wall and Stoffer (2002), and the method is justified by the results of a Monte Carlo simulation in finite samples. They employ a different approach using a quasi-likelihood estimator of ψ, which has some interesting properties, such as consistency (Harvey (1989)). Rodríguez and Ruiz (2012) presented procedures for the prediction of the state vector based on bootstrap. Pfeffermann and Tiller (2005) proposed parametric and nonparametric bootstrap methods

for estimating the PMSE of the state vector.

The main objective of this paper is to propose and compare bootstrap methods to reduce the bias in the estimation of the MSE for in-sample and out-of-sample estimates, in state space models. The in-sample study is performed in smoothed estimates of the state vector, while the out-of-sample is accomplished in the PMSE for future observations. To this purpose, nonparametric bootstrap procedures are proposed by adapting some methods existing in the literature to evaluate the PMSE for estimates of the state vector. This first one, called here HaB, is a modification on the method of Hamilton (1986) and the second one, the AKB, is adapted from Ansley and Kohn (1986). Forecast intervals using the nonparametric bootstrap correction of the PMSE for future observations are also provided, along with a new proposal of a parametric bootstrap interval, called here PBFI. These procedures are compared, through Monte Carlo experiments, to the existing method of Pfeffermann and Tiller (2005) and to the standard procedures.

The paper is organized as follows. Section 2 presents the structural models and how to calculate forecasts and smoothed estimates of the state vector, as well as the standard method to build confidence intervals. Section 3 presents nonparametric bootstrap procedures for forecasts and smoothed state vectors while Section 4 presents the parametric bootstrap interval proposed here. Section 5 provides some simulation studies. Section 6 shows application to real data sets. Finally, Section 7 concludes the work.

2 Forecasting and smoothing in structural models

A univariate time series $\{y_t\}_{t \in Z}$ can be decomposed as the sum of its unobservable components, such as trend, seasonality and error. A procedure which has been widely used to model y_t is the state space model (SSM) (Harvey, 1989; West and Harrison, 1997), which can be written as

$$y_t = \mathbf{z}_t' \boldsymbol{\alpha}_t + \epsilon_t, \quad \epsilon_t \sim N(0, h_t) \tag{1}$$

$$\boldsymbol{\alpha}_t = \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \mathbf{R}_t \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \tag{2}$$

where $\boldsymbol{\alpha}_t$ is the state vector, $\mathbf{z}_t$, $\mathbf{T}_t$ and $\mathbf{R}_t$ are the system matrices, ϵ_t are uncorrelated errors with variance h_t, $\boldsymbol{\eta}_t$ is a vector of serially uncorrelated errors with covariance matrix given by $\mathbf{Q}_t$ and $\boldsymbol{\eta}_t$ and ϵ_t are independent. Covariates can be added to both equations to include information such as structural breaks, outliers or external variables.

For the Gaussian model, the one-step-ahead predictive distribution function is given by $(y_t|Y_{t-1}) \sim N(\widetilde{y}_{t|t-1}, F_t)$, where $Y_{t-1} = \{y_1, ..., y_{t-1}\}$ and $\widetilde{y}_{t|t-1}$ and F_t are obtained by the Kalman filter (KF) algorithm (Kalman (1960)). For more details refer to (Harvey, 1989).

Let $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_p)$ be the hyperparameter vector, which in this case is composed of the variances of the error terms in (1) and (2). Thus, for a univariate series of size n, the logarithm of the likelihood function is given by

$$lnL(Y_n; \boldsymbol{\psi}) = ln \prod_{t=1}^{n} p(y_t|Y_{t-1}) = -\frac{n}{2}ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n} ln|F_t| - \frac{1}{2}\sum_{t=1}^{n} \nu_t' F_t^{-1} \nu_t$$
(3)

where $\nu_t = y_t - \widetilde{y}_{t|t-1}$ is the one-step-ahead forecast error. Since this is a nonlinear function of the hyperparameters, estimation must be performed numerically. In this work, for using the optimization algorithm BFGS, suitable transformations in the hyperparameters are undertaken (details can be seen in Franco et al. (2008)).

2.1 Smoothing

The smoothed estimator of the state vector is the inference performed on any particular time t, using the whole sample information, $Y_n = \{y_1, ..., y_n\}$. In linear Gaussian models, the smoothed estimator of the state vector is denoted by

$$\mathsf{a}_{t|n}(\boldsymbol{\psi}) = E(\alpha_t|Y_n, \boldsymbol{\psi}).$$
(4)

The smoothed estimate can be calculated running the KF and storing the conditional means, $\mathsf{a}_t(\boldsymbol{\psi}) = E(\alpha_t|Y_t, \boldsymbol{\psi})$ and $\mathsf{a}_{t|t-1}(\boldsymbol{\psi}) = E(\alpha_t|Y_{t-1}, \psi)$, and variances, $P_t(\boldsymbol{\psi}) = \text{Var}(\alpha_t|Y_t, \boldsymbol{\psi})$ and $P_{t|t-1}(\boldsymbol{\psi}) = \text{Var}(\alpha_t|Y_{t-1}, \boldsymbol{\psi})$, of the state vector. The sequence of smoothed estimates $\mathsf{a}_{t|n}(\boldsymbol{\psi})$ is then calculated in reverse order, for $t = n, n-1, n-2, \ldots, 1$. The corresponding mean squared error is given by

$$\text{MSE}(\mathsf{a}_{t|n}(\boldsymbol{\psi})) = P_{t|n}(\boldsymbol{\psi}) = E\left[\left(\alpha_t - \mathsf{a}_{t|n}(\boldsymbol{\psi})\right)\left(\alpha_t - \mathsf{a}_{t|n}(\boldsymbol{\psi})\right)' |Y_n\right].$$
(5)

This work employs a moment smoothing algorithm, which is described in Koopman et al. (1999), using the KF output. More details about smoothing algorithms can be found in De Jong (1989) and Harvey (1989).

2.2 Forecasting

The forecast of a future value y_{n+k}, $k > 0$, based on all available data Y_n, can be obtained through the k-step-ahead forecast of the state vector, $\mathtt{a}_{n+k|n}(\boldsymbol{\psi})$, which is given by

$$\mathtt{a}_{n+k|n}(\boldsymbol{\psi}) = E(\alpha_{n+k}|Y_n, \boldsymbol{\psi}) = \left(\prod_{i=1}^{k} \boldsymbol{T}_{n+i}\right) \mathtt{a}_n(\boldsymbol{\psi}), \qquad (6)$$

where $\mathtt{a}_n(\boldsymbol{\psi}) = E(\alpha_n|Y_n, \boldsymbol{\psi})$.

By combining Equation (1) in time $n + k$ with Equation (6), the k-step-ahead forecast for $\{y_t\}$, defined by $\tilde{y}_{n+k|n}(\boldsymbol{\psi}) = E(y_{n+k}|Y_n, \boldsymbol{\psi})$, can be calculated as

$$\tilde{y}_{n+k|n}(\boldsymbol{\psi}) = \boldsymbol{z}'_{n+k}\left(\prod_{i=1}^{k} \boldsymbol{T}_{n+i}\right) \mathtt{a}_n(\boldsymbol{\psi}) = \boldsymbol{z}'_{n+k}\mathtt{a}_{n+k|n}(\boldsymbol{\psi}). \qquad (7)$$

The corresponding mean square error is given by

$$\mathrm{MSE}(\tilde{y}_{n+k}|Y_n, \boldsymbol{\psi}) = \mathbf{z}'_{n+k}\left(\prod_{i=1}^{k} \boldsymbol{T}_{n+i}\right) \mathbf{P}_n(\boldsymbol{\psi}) \left(\prod_{i=1}^{k} \boldsymbol{T}'_{n+i}\right) \mathbf{z}_{n+k} +$$

$$\boldsymbol{z}'_{n+k} \sum_{i=1}^{k} \left(\prod_{j=1}^{k-i} \boldsymbol{T}_{n+k-j+1}\right) \boldsymbol{R}_{n+i}\boldsymbol{Q}_{n+i}\boldsymbol{R}'_{n+i} \left(\prod_{j=1}^{k-i} \boldsymbol{T}'_{n+k-j+1}\right) \mathbf{z}_{n+k} + h_{n+k}, \quad (8)$$

where $P_n(\boldsymbol{\psi}) = \mathrm{Var}(\alpha_n|Y_n, \boldsymbol{\psi})$.

2.3 Confidence intervals for smoothed estimates of the state vector and prediction intervals

Let τ be a specific time (instant) taken for an in-sample or out-of-sample observation and let $\mathtt{a}_{\tau|n}(\boldsymbol{\psi})$ denote the optimal inference about α_τ conditional on the whole sample. Thus, for $\tau \leq n$, $\mathtt{a}_{\tau|n}(\boldsymbol{\psi})$ is the smoothed inference given in (4), while for $\tau > n$, $\mathtt{a}_{\tau|n}(\boldsymbol{\psi})$ and $\tilde{y}_{\tau|n}(\boldsymbol{\psi})$ are the forecasts for the state vector and for the future observations given in (6) and (7), respectively.

When the true value of $\boldsymbol{\psi}$ is known, the confidence interval of level $1 - \kappa$ for the smoothed state vector is given by

$$\left[\mathsf{a}_{\tau|n}(\boldsymbol{\psi}) \pm \left| z_{\kappa/2} \right| \sqrt{\mathrm{MSE}(\mathsf{a}_{\tau|n}(\boldsymbol{\psi}))} \right], \quad \text{for} \quad \tau \leq n \qquad (9)$$

while the prediction interval of level $1 - \kappa$ is given by

$$\left[\tilde{y}_{\tau|n}(\boldsymbol{\psi}) \pm \left| z_{\kappa/2} \right| \sqrt{\mathrm{MSE}(\tilde{y}_{\tau|n}(\boldsymbol{\psi}))} \right], \quad \text{for} \quad \tau > n \qquad (10)$$

where $\tilde{y}_{\tau|n}(\boldsymbol{\psi}) = \boldsymbol{z}_{\tau}' \mathsf{a}_{\tau|n}(\boldsymbol{\psi})$ and $z_{\kappa/2}$ is the quantile $\kappa/2$ of the standard normal distribution.

The value of $\boldsymbol{\psi}$, however, is frequently unknown. In this case, it should be replaced by its MLE, $\hat{\boldsymbol{\psi}}$, and the obtained interval is called Standard (S) (Brockwell and Davis, 1996).

S Procedure:

1. Calculate the MLE $\hat{\boldsymbol{\psi}}$ estimated from the original series;
2. Calculate $\tilde{y}_{\tau|n}(\hat{\boldsymbol{\psi}})$ and $\mathrm{MSE}(\tilde{y}_{\tau|n}(\hat{\boldsymbol{\psi}}))$;
3. The confidence interval for $\boldsymbol{\alpha}_{n+k}$ and the forecast interval of level $(1 - \kappa)$ for $y_{\tau|n}$ are given, respectively, by (9) and (10).

The problem is that this interval does not incorporate the uncertainty related to $\boldsymbol{\psi}$, which inevitably leads to underestimation of the prediction MSE in the classical inference (Harvey (1989)). Another question that arises is related to the normality assumption, which can be unrealistic in practice. For example, if the future observations assume an asymmetric distribution, the interval coverage rates in the tails may be affected (see Rodriguez and Ruiz (2009)).

In the next sections, some alternatives to correct these problems are proposed, using the bootstrap.

3 Nonparametric bootstrap for estimating the MSE

As stated in Section 2, the standard procedure to estimate the MSE replaces the unknown parameter vector $\boldsymbol{\psi}$ by its MLE, $\hat{\boldsymbol{\psi}}$. In order to correct the bias introduced in the MSE of the state vector estimates and the predictions by this practice, some alternative approaches using the bootstrap are described in this section. As the procedures are very similar in both forecasting and smoothing contexts, only the notation for the forecast of a future value is used here to present the proposed algorithms. The procedures are easily adapted to the smoothed inference by replacing y_{τ} by $\mathbf{a}_{\tau}$ with $\tau \leq n$.

Consider the model given in (1 - 2). Initially the hyperparameters, which are the unknown variances of the errors ϵ_t and $\boldsymbol{\eta}_t$, must be estimated. Then the KF is run to obtain the values of the estimated innovations, $\hat{\nu}_t$, and their variances, $\hat{F}_t$. It should be noted that these quantities are functions of the unknown parameters.

The standardized innovations, $\hat{e}_t = (\hat{\nu}_t - \bar{\nu})/\sqrt{\hat{F}_t}$, $t = 1, 2, \ldots, n$, where $\bar{\nu} = \sum_{j=1}^{n} \hat{\nu}_j/n$, are resampled to construct the bootstrap series (see Stoffer and Wall (1991)). Then samples can be taken with replacement, from $\hat{e}_t$, to obtain the bootstrap innovations, $\hat{e}_t^*$.

The bootstrap series, y_t^*, is built recursively using the bootstrap innovations, $\hat{e}_t^*$, and the quantities $\hat{F}_t$ and $\hat{\mathbf{K}}_t$ obtained from the KF, where $\hat{\mathbf{K}}_t = \mathbf{T}_{t+1}\hat{\boldsymbol{P}}_{t|t-1}\mathbf{z}_t'\hat{F}_t^{-1}$ and $\hat{\mathbf{P}}_{t|t-1} = Var(\boldsymbol{\alpha}_t|Y_{t-1})$. Initially, the state vector estimate $\mathbf{a}_{t+1|t} = E(\boldsymbol{\alpha}_{t+1}|Y_t)$ and y_t are written in function of the innovations,

$$\mathbf{a}_{t+1|t} = \mathbf{T}_t\mathbf{a}_{t|t-1} + \mathbf{K}_t\nu_t$$

$$y_t = \mathbf{z}_t'\mathbf{a}_{t|t-1} + \nu_t.$$

Stoffer and Wall (1991) have proposed a procedure to build the bootstrap series using the following recursion

$$\mathbf{S}_t = \mathbf{A}_t\mathbf{S}_{t-1} + \mathbf{B}_t\nu_t, \quad t = 1, 2, ..., n, \tag{11}$$

where $\mathbf{S}_t = \begin{bmatrix} \mathbf{a}_{t+1|t} \\ y_t \end{bmatrix}$, $\mathbf{A}_t = \begin{bmatrix} \mathbf{T}_t & 0 \\ \mathbf{z}_t & 0 \end{bmatrix}$ and $\mathbf{B}_t = \begin{bmatrix} \mathbf{K}_t\sqrt{F_t} \\ \sqrt{F_t} \end{bmatrix}$.

The bootstrap series, y_t^*, $t = 1, 2, ..., n$, is obtained by solving Equation (11), replacing ν_t by $\hat{e}_t^*$, $\mathbf{K}_t$ by $\hat{\mathbf{K}}_t$ and F_t by $\hat{F}_t$.

The procedures described in the next subsections are the proposals of this work to obtain confidence intervals for the state vector and prediction intervals, using the (non-)parametric bootstrap. They are based on the methods of Hamilton (1986) and Ansley and Kohn (1986), which calculate the prediction MSE of the state vector for the state space model with estimated parameters. For comparison purposes, the procedure proposed by Pfeffermann and Tiller (2005) is also presented.

3.1 Hamilton procedure with bootstrap resampling

Following the proposal of Hamilton (1986), which incorporates the uncertainty of the parameter estimation in estimating the state vector, the effect

of the estimation of ψ in the forecasts can be eliminated using Monte Carlo integration. The predictive distribution of future observations is given by

$$
\begin{aligned}
p(y_\tau, \psi|Y_n) &= \int p(y_\tau, \psi|Y_n)d\psi \\
&= \int p(y_\tau|Y_n, \psi)p(\psi|Y_n)d\psi \\
&\cong \frac{1}{MC}\sum_{i=1}^{MC} p(y_\tau|Y_n, \psi^i),
\end{aligned}
$$

where ψ^i are samples from $p(\hat{\psi}|Y_n) \cong N(\hat{\psi}, I^{-1}(\hat{\psi}))$, the asymptotic distribution of the MLE of ψ, where $I(\cdot)$ is the Fisher information matrix and MC is the required number of Monte Carlo replications.

The mean of the predictive distribution is given by:

$$
\begin{aligned}
\bar{y}_{\tau|n} &= \int \tilde{y}_{\tau|n}(\psi)p(\psi|Y_n)d\psi \\
&\cong \frac{1}{MC}\sum_{i=1}^{MC} \tilde{y}_{\tau|n}(\psi^{(i)}). \tag{12}
\end{aligned}
$$

Following the results in Hamilton (1986), with some modifications, the prediction MSE can be calculated as

$$
\begin{aligned}
\mathrm{MSE}_{\tau|n}^{Ha} &= E\left[(y_\tau - \bar{y}_{\tau|n})^2\right] \\
&= \mathrm{MSE}_{p(\psi|Y_n)}(\tilde{y}_{\tau|n}(\psi)) + E_{p(\psi|Y_n)}[(\tilde{y}_{\tau|n}(\psi) - \bar{y}_{\tau|n})^2].
\end{aligned}
$$

Then,

$$
\widehat{\mathrm{MSE}}_{\tau|n}^{Ha} \cong \frac{1}{MC}\sum_{i=1}^{MC}\widehat{\mathrm{MSE}}(\tilde{y}_{\tau|n}(\psi^i)) + \frac{1}{MC}\sum_{i=1}^{MC}(\tilde{y}_{\tau|n}(\psi^i) - \bar{y}_{\tau|n})^2, \tag{13}
$$

where $\bar{y}_{\tau|n}$ is given in equation (12), $\tilde{y}_{\tau|n}(\psi^i)$ in (7) and $\widehat{\mathrm{MSE}}(\tilde{y}_{\tau|n}(\psi^i))$ in (8).

Pfeffermann and Tiller (2005) argue that sampling from $N(\hat{\psi}, I^{-1}(\hat{\psi}))$ may result in several problems, such as parameters being close to their boundary values, the distribution of $\hat{\psi}$ can be asymmetric in small samples and the calculation of the Fisher information matrix may become unstable for complex models. To avoid this, a simple and efficient procedure is to use bootstrap resamples. Thus, the first procedure proposed in this work is a variation of the Hamilton approach discussed above, using sampling bootstrap, called HaB procedure.

Procedure 1 (HaB):

1. Generate B nonparametric bootstrap series, using $\hat{\boldsymbol{\psi}}$ estimated from the original series;
2. Calculate $\hat{\boldsymbol{\psi}}^{*(b)}$ for each bootstrap series, $b = 1, \ldots, B$;
3. Using $\hat{\boldsymbol{\psi}}^{*(1)}, \ldots, \hat{\boldsymbol{\psi}}^{*(B)}$, compute:
   - $\bar{y}^*_{\tau|n}$ based on equation (12), with the number of Monte Carlo replications now equal to B;
   - the $\widehat{\mathrm{MSE}}_{\tau|n}^{HaB}$ for $\bar{y}^*_{\tau|n}$, using Equation (13).
4. The forecast interval of level $(1 - \kappa)$ for $y_{\tau|n}$ is given by,

$$\left[\bar{y}^*_{\tau|n} \pm |z_{\kappa/2}| \sqrt{\widehat{\mathrm{MSE}}_{\tau|n}^{HaB}} \right],$$

where $z_{\kappa/2}$ is the $(\frac{\kappa}{2})^{th}$ quantile of the standard normal distribution.

Assuming the state space form with Gaussian errors given in (1-2), Stoffer and Wall (1991) ensure that $p(\hat{\boldsymbol{\psi}}^*|Y_n) \cong p(\hat{\boldsymbol{\psi}}|Y_n)$ when B (the number of bootstrap replicates) is sufficiently large. That is, the distribution of $\hat{\boldsymbol{\psi}}$ is approximated by the bootstrap distribution of the MLE and they are equivalent when the sample size is sufficiently large. Therefore, the distribution of $p(\hat{\boldsymbol{\psi}}|Y_n)$ can be replaced by the distribution of $p(\hat{\boldsymbol{\psi}}^*|Y_n)$ in the above procedure. It should be noted that the number of replications necessary to calculate $\widehat{\mathrm{MSE}}^{HaB}$ is equal to $2 \times B$, as it can be perceived from (13), with $MC = B$.

3.2 Ansley and Kohn procedure

Ansley and Kohn (1986) proposed a procedure for incorporating the uncertainty in the estimation of $\boldsymbol{\psi}$ through a conditional mean square error for the estimate of the state vector. According to Harvey (1989, page 151), the idea can be used in the forecasting context through the approximation of $\mathrm{MSE}(\tilde{y}_\tau|Y_n, \boldsymbol{\psi})$ by $\mathrm{MSE}(\tilde{y}_\tau|Y_n, \hat{\boldsymbol{\psi}})$ and the expansion of $\tilde{y}_{\tau|n}(\hat{\boldsymbol{\psi}})$ around $\tilde{y}_{\tau|n}(\boldsymbol{\psi})$ up to the second term. The Prediction MSE estimator is given by

$$\widehat{\mathrm{MSE}}_{\tau|n}^{\mathrm{AK}} = \widehat{\mathrm{MSE}}(\tilde{y}_{\tau|n}(\hat{\boldsymbol{\psi}})) + \left[\frac{\partial \tilde{y}_{\tau|n}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right]'_{\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}} \left[I^{-1}(\hat{\boldsymbol{\psi}}) \right] \left[\frac{\partial \tilde{y}_{\tau|n}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right]_{\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}}, (14)$$

where $\hat{\boldsymbol{\psi}}$ is the MLE of $\boldsymbol{\psi}$ and $I^{-1}(\hat{\boldsymbol{\psi}})$ is the Fisher information matrix evaluated at $\hat{\boldsymbol{\psi}}$. $I^{-1}(\hat{\boldsymbol{\psi}})$ may be calculated using a numerical approximation, as in Franco et al. (2008) and Santos and Franco (2010).

The procedure of Ansley and Kohn (1986) incorporates the uncertainty in the estimation of hyperparameters in an elegant way, although the calculation of the Fisher information matrix and the derivation of the function $\tilde{y}_{\tau|n}(\boldsymbol{\psi})$ with respect to $\boldsymbol{\psi}$ can be a difficult task. To compute the Fisher information matrix, we propose in this work to use a bootstrap approximation $I_n^{-1}(\hat{\boldsymbol{\psi}}) \cong \widehat{Cov}(\hat{\boldsymbol{\psi}}^*)$, which is supported by the same argument of the asymptotic validity of Procedure 1. With respect to the calculation of $\frac{\partial \tilde{y}_{\tau|n}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}$, with $\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}$, it can be performed using numerical derivatives.

Thus, using the empirical bootstrap distribution of the MLE of $\boldsymbol{\psi}$, the following Ansley and Kohn procedure with bootstrap sampling (AKB) is proposed. This proposal is a variation of the Ansley and Kohn procedure, where the Fisher information matrix is approximated by the covariance matrix of the empirical bootstrap distribution for the MLE of $\boldsymbol{\psi}$.

Procedure 2 (AKB):

1. The first steps are Steps (1) and (2) in Procedure 1.

2. With $\hat{\boldsymbol{\psi}}^{*(1)}, \ldots, \hat{\boldsymbol{\psi}}^{*(B)}$, compute:

   - $\bar{y}_{\tau|n}^*$ using equation (12);

   - the $\widehat{\mathrm{MSE}}_{\tau|n}^{AKB}$ for $\bar{y}_{\tau|n}^*$, using Equation (14) .

3. The forecast interval of level $(1 - \kappa)$ for $y_{\tau|n}$ is given by,

$$\left[\bar{y}_{\tau|n}^* \pm |z_{\kappa/2}| \sqrt{\widehat{\mathrm{MSE}}_{\tau|n}^{AK}} \right],$$

where $z_{\kappa/2}$ is the $\left(\frac{\kappa}{2}\right)^{th}$ quantile of the standard normal distribution.

To obtain $\widehat{\mathrm{MSE}}_{\tau|n}^{AKB}$ it is necessary to calculate p numerical derivatives with respect to the hyperparameter vector and the Fisher information matrix of $p \times p$ dimension, which is similar to the $2 * B$ replications of the HaB procedure.

3.3 Pfeffermann and Tiller procedure

Pfeffermann and Tiller (2005) proposed a procedure that estimates the prediction MSE by incorporating the uncertainty in the hyperparameter estimation. An advantage of this procedure is that its prediction MSE estimator is $O(1/n^2)$ under certain conditions. Therefore, it is expected that the Pfeffermann and Tiller method produces better estimates for the prediction MSE. The procedure, called here PT, is described below.

Procedure 3 (PT):

1. The first steps are Steps (1) and (2) in Procedure 1.
2. Compute

$$\widehat{\text{MSE}}_{\tau|n}^{PT} = \widehat{\text{MSE}}^{bs}(\tilde{y}_{\tau|n}(\boldsymbol{\psi})) + 2\widehat{\text{MSE}}(\tilde{y}_{\tau|n}(\hat{\boldsymbol{\psi}})) - \overline{\text{MSE}}^{bs}(\tilde{y}_{\tau|n}(\boldsymbol{\psi})),$$

where $\widehat{\text{MSE}}^{bs}(\tilde{y}_{\tau|n}(\boldsymbol{\psi})) = \frac{1}{B}\sum\limits_{b=1}^{B}\left[y_{\tau|n}^{\star(b)}(\hat{\boldsymbol{\psi}}^{*(b)}) - y_{\tau|n}^{\star(b)}(\hat{\boldsymbol{\psi}})\right]^2$

and

$\overline{\text{MSE}}^{bs}(\tilde{y}_{\tau|n}(\boldsymbol{\psi})) = \frac{1}{B}\sum\limits_{b=1}^{B}\widehat{\text{MSE}}(\tilde{y}_{\tau|n}(\hat{\boldsymbol{\psi}}^{*(b)})).$

It is important to observe that $y_{\tau|n}^{\star(b)}(\hat{\boldsymbol{\psi}})$ and $y_{\tau|n}^{\star(b)}(\hat{\boldsymbol{\psi}}^{*(b)})$ are the bootstrap estimates of the future observations with the estimates $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\psi}}^{*(b)}$ of $\boldsymbol{\psi}$, respectively. $\widehat{\text{MSE}}(\tilde{y}_{\tau|n}(\hat{\boldsymbol{\psi}}^{\star(b)}))$ is the natural estimator in Equation (8) with the original series, but using the estimate $\hat{\boldsymbol{\psi}}^{*(b)}$.

3. The forecast interval of level $(1 - \kappa)$ for $y_{\tau|n}$ is given by,

$$\left[\bar{y}_{\tau|n}^{*} \pm |z_{\kappa/2}|\sqrt{\widehat{\text{MSE}}_{\tau|n}^{PT}}\right],$$

where $\bar{y}_{\tau|n}^{*}$ is given in (12).

The PT procedure is computationally most costly compared to the HaB and AKB, as it needs $3.B$ replications in order to calculate the MSE, that is, B replications for each of the following terms $\widehat{\text{MSE}}(\tilde{y}_{\tau|n}(\hat{\boldsymbol{\psi}}^{\star(b)}))$, $y_{\tau|n}^{\star(b)}(\hat{\boldsymbol{\psi}}^{*(b)})$ and $y_{\tau|n}^{\star(b)}(\hat{\boldsymbol{\psi}})$.

4 Parametric bootstrap forecast interval

Unlike the nonparametric bootstrap, in which the residuals of the fitted model are resampled with replacement, the parametric bootstrap uses only the parameter estimates from the original series. The procedure is performed as follows. The bootstrap errors ϵ_t^* and η_t^* are sampled from the Gaussian distribution with zero mean and variance replaced by the estimates obtained from the original series.

Then the bootstrap series y_t^* is constructed as:

$$\begin{aligned}
y_t^* &= \mathbf{z}_t^{'}\alpha_t^* + \hat{\epsilon}_t^*, \\
\alpha_t^* &= \mathbf{T}_t\alpha_{t-1}^* + \mathbf{R}_t\hat{\eta}_t^*, \quad t = 1, 2, \ldots, n.
\end{aligned}$$

In the above equations, α_0^* can be initialized using the same values specified when running the KF for the original series.

A new parametric bootstrap forecast interval, called here PBFI, which incorporates the uncertainty due to parameter estimation is presented in Procedure 4. Although this method can be performed for observations not necessarily Gaussian, the Gaussian distribution is assumed here for the future observations.

Procedure 4 (PBFI):

1. Generate B parametric bootstrap series, using $\hat{\psi}$ estimated from the original series;
2. Calculate $\hat{\psi}^{*(b)}$ for each bootstrap series, $b = 1, \ldots, B$;
3. Using $\hat{\psi}^{*(1)}, \ldots, \hat{\psi}^{*(B)}$, run the KF for the original series and obtain $\mathsf{a}_n^{(b)}(\boldsymbol{\psi})$ and $\mathbf{P}_n^{(b)}(\boldsymbol{\psi})$, for $b = 1, \ldots, B$;
4. Generate the k-steps-ahead bootstrap forecasts using the following equations:

   (a) For $b = 1, \ldots, B$, calculate
   $$\alpha_{n+k}^{*(b)} = \left(\prod_{i=1}^{k} \mathbf{T}_{n+i}^{*(b)}\right) \alpha_n^{*(b)} + \sum_{i=1}^{k} \left(\prod_{j=1}^{k-i} \mathbf{T}_{n+k-j+1}^{'*(b)}\right) \mathbf{R}_{n+i}^{'*(b)} \eta_{n+i}^{*(b)},$$
   where $\eta_{n+i}^{*(b)} \sim F(\mathbf{0}, \hat{\mathbf{Q}}_t^{*(b)})$ and $\alpha_n^{*(b)}$ is generated from the Normal distribution with mean $\mathsf{a}_n^{(b)}$ and covariance matrix $\mathbf{P}_n^{(b)} + (\mathsf{a}_n^{(b)} - \mathsf{a}_n) \times (\mathsf{a}_n^{(b)} - \mathsf{a}_n)'$.

   (b) $y_{n+k}^{*(b)} = \mathbf{z}_{n+k}^{'*(b)} \alpha_{n+k}^{*(b)} + \epsilon_{n+k}^{*(b)}$, where $\epsilon_{n+k}^{*(b)}$ is generated from a generic continuous distribution (G) with zero mean and variance $\hat{h}_t^{*(b)}$.

5. The bootstrap forecast interval of level $(1 - \kappa)$ for y_{n+k} is given by,
$$\left[y_{n+k}^{*(\kappa/2)} ; y_{n+k}^{*(1-\kappa/2)} \right],$$
where $y_{n+k}^{*(\kappa)}$ is the quantile of order κ.

The G distribution is not necessarily Gaussian, despite being the most frequent case. The term $(\mathsf{a}_n^{(b)} - \mathsf{a}_n) \times (\mathsf{a}_n^{(b)} - \mathsf{a}_n)'$ was introduced in the covariance matrix to incorporate the uncertainty in the estimation of $\boldsymbol{\psi}$ when generating values of the state vector. It can be viewed as an approximation to the second term of the variance decomposition of the state vector, $Var(\alpha_n | Y_n) = E_{\psi|Y_n}(Var(\alpha_n | Y_n, \psi)) + Var_{\psi|Y_n}(E(\alpha_n | Y_n, \psi))$, which is the same argument given by Hamilton (1986) to calculate the MSE in the Ha procedure of Eq. (13).

The asymptotic validity of Procedure 4 under Gaussian errors can be easily obtained, although it is not difficult to prove the non-Gaussian case. If the distribution of the observation equation is not Gaussian, the MLE of $\boldsymbol{\psi}$ becomes the maximum quasi-likelihood estimator of $\boldsymbol{\psi}$, which is consistent and preserves good properties (Harvey, 1989).

5 Simulation results

The bootstrap procedures described in Sections 3 and 4 are compared through a Monte Carlo study and the results are presented separately for smoothing and forecasting.

The experiments are performed for the simplest structural model, known as the Local Level Model (LLM) (Harvey (1989)), defined as

$$
\begin{array}{rcl}
y_t &=& \alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \\
\alpha_t &=& \alpha_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2).
\end{array}
\tag{15}
$$

Different values for the signal-to-noise ratio, $q = \sigma_\eta^2 / \sigma_\epsilon^2$, are used in the simulations, with $\sigma_\epsilon^2 = 1$. To check the robustness of the methods with respect to the Gaussian assumption, a Gamma distribution for the error terms, re-centered and scaled so that the errors have zero mean and variance 1, was also assumed. All models were fitted using the KF and the Gaussian likelihood function and the simulations were implemented in the Ox Software (Doornik, 2002).

5.1 Smoothing

The MSE of the smoothed state vector is estimated using the nonparametric bootstrap procedures proposed in Section 3. To this purpose, MC=1000 time series are generated for the LLM, with Gaussian and non-Gaussian distributions, for three sample sizes $n = 40$, $n = 100$ and $n = 500$ and with signal-to-noise ratio $q = 0.25$. These values and distributions were chosen according to a simulation study in Quenneville and Singh (2000). The number of bootstrap replications is 2000.

The estimators of the MSE compared here are: Standard (S), obtained by substituting the hyperparameters in equation (15) by their MLE to calculate the MSE of $\mathtt{a}_{t|n}(\boldsymbol{\psi})$; Hamilton with bootstrap resample (HaB) described in Procedure 1; Ansley and Kohn with bootstrap resample (AKB), defined in Procedure 2; and Pfeffermann & Tiller estimator (PT) described in Procedure 3. For the first estimator, a numerical approximation for the Fisher information matrix was used.

Following the Pfeffermann and Tiller (2005) and Quenneville and Singh (2000) experiments, the procedures are compared using the relative bias (Rel-Bias) and relative square root of the mean square error (Rel-SMSE), given respectively by,

$$\text{Rel-Bias} = \frac{100}{n} \sum_{t=1}^{n} \frac{\bar{d}_t(\boldsymbol{\psi})}{MSE_t(\boldsymbol{\psi})}$$

and

$$\text{Rel-SMSE} = \frac{100}{n} \sum_{t=1}^{n} \frac{(\bar{d}_t^{(2)}(\boldsymbol{\psi}))^{1/2}}{MSE_t(\boldsymbol{\psi})},$$

where $\bar{d}_t(\boldsymbol{\psi}) = \sum_{i=1}^{MC} \frac{d_{i,t}}{MC}$, $\bar{d}_t^{(2)}(\boldsymbol{\psi}) = \sum_{i=1}^{MC} \frac{d_{i,t}^2}{MC}$, $d_{i,t} = [\hat{MSE}_t^{(i)}(\boldsymbol{\psi}) - MSE_t(\boldsymbol{\psi})]$ and MSE_t is the true PMSE of $\mathbf{a}_{t|n}(\boldsymbol{\psi})$ calculated for each time $t = 1, \ldots, n$ by simulating 5000 series for each lenght ($n = 40, 100, 500$) according to

$$MSE_t(\boldsymbol{\psi}) \cong \sum_{i=1}^{5000} \frac{(\mathbf{a}_{t|n}(\hat{\boldsymbol{\psi}})^{(i)} - \alpha_t^{(i)})^2}{5000}.$$

Table 1 shows the Rel-Bias and Rel-SMSE for the PMSE of the smoothed estimator in the LLM with Gaussian and non Gaussian errors. In the case of Gaussian errors, it can be seen that the bootstrap procedures greatly reduce the underestimation of the PMSE compared to the standard (S) method with respect to the bias, especially for small sample sizes. However, in spite of possessing higher Rel-Bias, the S estimator has a similar or inferior Rel-SMSE compared to the bootstrap estimators. This is not an unexpected result, once the bias correction can increase the variance (see Pfeffermann and Tiller (2005)). With respect to the nonparametric bootstraps proposed here, the AKB is superior to the HaB procedure. In the non-Gaussian case, Gamma(16/9, 3/4) and Gamma(25/16, 4/5) distributions for ϵ_t and η_t, respectively, were assumed. The Rel-Bias and Rel-SMSE are larger for non-Gaussian errors, although the behavior of the PMSE estimators under the smoothing procedure is similar to the Gaussian case. Although the PT presents a better performance with respect to the Rel-Bias, the nonparametric bootstrap procedures proposed here (HaB and AKB) are simpler, require less Monte Carlo replications and their computational time is about one third less than the PT procedure. This can bring a significant gain when the model complexity increases. Comparing HaB and AKB, the last one is slightly faster than the former.

Table 1 *Relative bias (%) and relative MSE square root (%) of the smoothed estimators for a local level model with Gaussian (G) and Non-Gaussian (NG) errors and signal-to-noise ratio q = 0.25.*

		n = 40		n = 100		n = 500	
		Rel-Bias	Rel-SMSE	Rel-Bias	Rel-SMSE	Rel-Bias	Rel-SMSE
G	S	-23.422	19.413	-8.677	10.315	-2.235	4.021
	HaB	-8.224	24.204	-5.726	10.407	-2.158	3.917
	AKB	-4.460	28.131	-4.003	11.224	-1.396	3.940
	PTB	0.110	19.828	0.742	10.108	0.598	3.903
NG	S	-24.498	28.389	-10.294	15.093	-1.545	6.007
	HaB	-13.536	34.921	-10.102	16.990	-1.592	6.269
	AKB	-10.934	31.445	-4.906	15.621	-0.547	6.084
	PT	2.567	31.125	1.147	15.012	0.519	5.987

5.2 Forecasting

The performance of the bootstrap methods presented in Sections 3 and 4 are studied here in the out of sample forecasting context. The following procedures: Standard (S) shown in Equation (10), Procedure 1 (HaB), Procedure 2 (AKB), Procedure 3 (PT) and the parametric bootstrap from Procedure 4 (PBFI) are compared through the average width and coverage rates of the forecast intervals. For this study, series of size $n = 50, 100$ and 500 were generated with a burn-in equal to 100. The values of the hyperparameters were chosen to equal the following signal-to-noise ratio: $q = 0.1$ and $q = 1.0$, according to the simulation study in Rodriguez and Ruiz (2009). The forecasts were calculated k-steps-ahead for $k = 1, 5$ and 15. The number of Monte Carlo replications was $MC = 1000$, the number of bootstrap resamples was $B = 2000$ and the nominal level of the prediction intervals was fixed at 95%. For the non-Gaussian case, a Gamma$(1/9, 1/3)$ distribution for ϵ_t was assumed.

Tables 2 and 3 contain the results for Gaussian errors with $q = 0.1$ and $q = 1.0$, respectively. It can be seen that the proposed methods are consistent, as for all forecast lags, the coverage rates get close to the fixed 95% level assumed, as the sample size increases. It can be also noted that, even for small samples, all methods present very satisfactory results, with the AKB, PT and PBFI intervals showing coverage slightly better than the other procedures. Note that the bootstrap approaches present width greater than the S method, particularly for small samples, once they incorporate the uncertainty inherent in the estimation of ψ. The parametric bootstrap (PBFI) shows a better performance when $k = 1$, but its width is slightly larger than the other methods. The coverage on the tails appears to be symmetric for

Table 2 *95% prediction intervals for y_{n+k} with $q = 0.1$ and Gaussian errors.*

		S Width Coverage	PBFI Width Coverage	HaB Width Coverage	AKB Width Coverage	PT Width Coverage
n	k	Tails	Tails	Tails	Tails	Tails
50	1	4.462	4.707	4.536	4.590	4.601
		0.913	0.933	0.919	0.920	0.922
		0.041/0.046	0.035/0.032	0.038/0.043	0.037/0.043	0.037/0.041
100	1	4.533	4.631	4.574	4.561	4.603
		0.955	0.953	0.954	0.954	0.956
		0.021/0.024	0.019/0.028	0.020/0.026	0.020/0.026	0.020/0.024
500	1	4.582	4.566	4.579	4.586	4.592
		0.949	0.952	0.955	0.954	0.955
		0.022/0.029	0.020/0.028	0.017/0.028	0.016/0.030	0.016/0.029
50	5	5.099	5.340	5.174	5.229	5.215
		0.938	0.942	0.939	0.941	0.941
		0.031/0.031	0.028/0.030	0.030/0.031	0.029/0.032	0.029/0.030
100	5	5.147	5.240	5.180	5.174	5.213
		0.936	0.944	0.938	0.939	0.941
		0.031/0.033	0.025/0.031	0.03/0.032	0.030/0.031	0.030/0.029
500	5	5.214	5.198	5.210	5.218	5.225
		0.950	0.949	0.951	0.951	0.951
		0.023/0.027	0.026/0.025	0.023/0.026	0.023/0.026	0.023/0.026
50	15	6.331	6.606	6.410	6.448	6.406
		0.918	0.926	0.923	0.924	0.919
		0.044/0.033	0.045/0.029	0.047/0.030	0.044/0.032	0.049/0.032
100	15	6.409	6.504	6.429	6.440	6.472
		0.926	0.930	0.928	0.929	0.930
		0.037/0.037	0.037/0.033	0.038/0.034	0.037/0.034	0.037/0.033
500	15	6.485	6.464	6.479	6.489	6.498
		0.946	0.938	0.944	0.946	0.947
		0.028/0.026	0.033/0.029	0.029/0.027	0.028/0.026	0.028/0.025

all methods, as expected. It can be also noted that the average width of the intervals increases as q increases.

Tables 4 and 5 contain the results for Gamma errors with $q = 0.1$ and $q = 1.0$, respectively. Once again, it can be seen that the coverage rates get close to the fixed 95% level as the sample size increases, for all forecast lags. In this case, the coverage rate of the S method is slightly smaller than the bootstrap methods when the sample size is small ($n = 50$). It can be also seen that the bootstrap methods, which incorporate the uncertainty in the estimation of $\boldsymbol{\psi}$, have larger width, as expected, especially for small sample sizes. Due to the asymmetry of the Gamma distribution, the coverage rates in the tails are not symmetric, but the asymmetry is more pronounced when $q < 1$.

Table 3 *95% prediction intervals for y_{n+k} with $q = 1.0$ and Gaussian errors.*

		S	PBFI	HaB	AKB	PT
		Width	Width	Width	Width	Width
		Coverage	Coverage	Coverage	Coverage	Coverage
n	k	Tails	Tails	Tails	Tails	Tails
50	1	6.212	6.398	6.282	6.292	6.381
		0.924	0.933	0.927	0.929	0.932
		0.041/0.035	0.036/0.031	0.038/0.035	0.036/0.035	0.038/0.030
100	1	6.287	6.330	6.292	6.326	6.366
		0.952	0.952	0.952	0.955	0.956
		0.022/0.026	0.023/0.025	0.022/0.026	0.021/0.024	0.021/0.023
500	1	6.326	6.315	6.326	6.333	6.334
		0.949	0.946	0.949	0.949	0.949
		0.020/0.031	0.020/0.034	0.020/0.031	0.020/0.031	0.020/0.031
50	5	9.772	10.015	9.861	9.829	9.845
		0.925	0.933	0.927	0.927	0.929
		0.036/0.039	0.032/0.035	0.035/0.038	0.032/0.041	0.033/0.038
100	5	9.915	9.996	9.956	9.940	9.928
		0.945	0.938	0.944	0.944	0.944
		0.024/0.031	0.026/0.036	0.024/0.032	0.024/0.032	0.024/0.032
500	5	10.035	10.008	10.046	10.040	10.033
		0.949	0.943	0.948	0.948	0.950
		0.026/0.025	0.031/0.026	0.026/0.026	0.025/0.027	0.025/0.025
50	15	15.504	15.836	15.521	15.504	15.517
		0.911	0.913	0.910	0.911	0.913
		0.048/0.041	0.047/0.040	0.048/0.042	0.047/0.042	0.047/0.040
100	15	15.793	15.943	15.830	15.808	15.788
		0.928	0.927	0.925	0.925	0.925
		0.034/0.038	0.036/0.037	0.035/0.040	0.036/0.039	0.037/0.038
500	15	15.881	15.842	15.886	15.884	15.882
		0.955	0.952	0.955	0.955	0.955
		0.021/0.024	0.021/0.027	0.021/0.024	0.021/0.024	0.021/0.024

Table 4 *95% prediction intervals for y_{n+k} with $q = 0.1$ and Gamma errors.*

		S	PBFI	HaB	AKB	PT
		Width	Width	Width	Width	Width
		Coverage	Coverage	Coverage	Coverage	Coverage
n	k	Tails	Tails	Tails	Tails	Tails
50	1	4.131	4.386	4.183	4.748	4.265
		0.921	0.929	0.925	0.929	0.927
		0.017/0.062	0.011/0.060	0.014/0.061	0.011/0.060	0.013/0.060
100	1	4.345	4.502	4.382	4.433	4.413
		0.962	0.968	0.963	0.963	0.962
		0.010/0.028	0.006/0.026	0.009/0.028	0.009/0.028	0.010/0.028
500	1	4.531	4.531	4.600	4.536	4.543
		0.956	0.956	0.956	0.956	0.956
		0.006/0.038	0.008/0.036	0.006/0.038	0.006/0.038	0.021/0.025
50	5	4.893	5.170	4.993	5.488	4.993
		0.934	0.941	0.934	0.945	0.937
		0.024/0.042	0.023/0.036	0.024/0.042	0.018/0.037	0.025/0.038
100	5	5.020	5.168	5.047	5.166	5.078
		0.960	0.959	0.959	0.959	0.960
		0.004/0.036	0.004/0.037	0.004/0.037	0.004/0.037	0.004/0.036
500	5	5.156	5.151	5.154	5.160	5.165
		0.952	0.952	0.952	0.952	0.952
		0.009/0.039	0.009/0.039	0.009/0.039	0.009/0.039	0.009/0.039
50	15	6.139	6.435	6.220	6.580	6.191
		0.917	0.924	0.923	0.927	0.923
		0.044/0.039	0.040/0.036	0.041/0.036	0.037/0.036	0.042/0.035
100	15	6.383	6.562	6.404	6.523	6.434
		0.942	0.940	0.943	0.946	0.945
		0.021/0.037	0.022/0.038	0.020/0.037	0.017/0.037	0.018/0.037
500	15	6.462	6.468	6.457	6.466	6.475
		0.957	0.960	0.957	0.958	0.959
		0.019/0.024	0.016/0.024	0.019/0.024	0.018/0.024	0.018/0.023

As a general conclusion, it can be seen that, although the forecasting methods proposed here were built in order to correct the PMSE, this correction does not seem to affect the coverage rates. Similar results were reached by Rodriguez and Ruiz (2009), in a study with a nonparametric bootstrap proposal.

6 Application to real data sets

6.1 Electric energy consumption

To illustrate the smoothing procedures, this section presents a study on a time series of electric energy consumption in the Northeast region of Brazil.

Table 5 *95% prediction intervals for y_{n+k} with $q = 1.0$ and Gamma errors.*

		S	PBFI	HaB	AKB	PT
		Width	Width	Width	Width	Width
		Coverage	Coverage	Coverage	Coverage	Coverage
n	k	Tails	Tails	Tails	Tails	Tails
50	1	5.978	6.155	6.026	6.068	6.136
		0.936	0.939	0.935	0.938	0.940
		0.027/0.037	0.025/0.036	0.026/0.039	0.025/0.037	0.025/0.035
100	1	6.151	6.216	6.157	6.191	6.228
		0.943	0.951	0.945	0.947	0.945
		0.030/0.027	0.025/0.024	0.027/0.028	0.026/0.027	0.029/0.026
500	1	6.299	6.267	6.298	6.306	6.314
		0.948	0.954	0.948	0.948	0.949
		0.021/0.031	0.018/0.028	0.021/0.031	0.020/0.032	0.020/0.031
50	5	9.804	9.974	9.790	9.874	9.961
		0.927	0.929	0.927	0.926	0.929
		0.032/0.041	0.031/0.040	0.032/0.041	0.029/0.045	0.029/0.042
100	5	9.031	9.994	9.934	9.953	9.976
		0.942	0.949	0.944	0.944	0.944
		0.016/0.042	0.013/0.038	0.015/0.041	0.014/0.042	0.016/0.040
500	5	10.031	10.011	10.035	10.035	10.035
		0.942	0.940	0.942	0.942	0.940
		0.023/0.035	0.025/0.035	0.025/0.035	0.023/0.035	0.023/0.037
50	15	15.443	15.655	15.341	15.476	15.628
		0.918	0.915	0.914	0.919	0.920
		0.043/0.039	0.043/0.042	0.043/0.043	0.042/0.039	0.042/0.038
100	15	15.816	15.981	15.817	15.831	15.846
		0.925	0.923	0.927	0.926	0.927
		0.033/0.042	0.036/0.041	0.031/0.042	0.031/0.043	0.032/0.041
500	15	15.926	15.917	15.934	15.929	15.924
		0.941	0.943	0.942	0.941	0.943
		0.027/0.032	0.025/0.032	0.027/0.031	0.026/0.033	0.026/0.031

Table 6 *MLE and confidence intervals for the hyperparameters of CHESF series.*

ψ	MLE	Conf. Int. 95%
σ_η^2	0.14	[0.03; 0.31]
σ_ϵ^2	0.25	[0.11; 0.40]

Table 7 *Percentual increasing of MSE compared to the Standard procedure for the CHESF series.*

Methods	Percent
PT	9.56
HaB	7.71
AKB	6.40

These data were obtained from a large study concerning the quantity of energy necessary to answer the maximum demand in the peak interval (from 6:00 pm to 9:00 pm). The series are monthly observations of electric consumption from CHESF (São Francisco Hydroelectric Company), in the period from May 1991 to December 1996 ($n = 68$). The data are shown in Figure 1 and the LLM was fitted to the series.

Point and interval estimates for the hyperparameters are shown in Table 6. The signal-to-noise ratio is less than 1 (0.56). A residual analysis was carried out and no evidence of correlation across time in the error term was found. As the zero value is not included in the intervals, the LLM can be an adequate model for this series.

Table 7 provides the percentage increase in the MSE compared to the standard procedure. The PT method presents the highest increasing, while the AKB and HaB procedures have a similar performance. Figure 1 provides smoothed estimates (point and 95% confidence intervals) for the level component using the PT method, which follows well the series behavior.

6.2 Income of a small Brazilian city

This application deals with the net income series from a small Brazilian city, whose name is omitted here for confidentiality. It is important to forecast future values of income series for planning and control of the annual budget and costs.

The income series, shown in Figure 2, consists of 72 monthly observations in the period 2006/01 to 2011/12. The last **six** observations of the series were omitted to compare the forecast intervals to the future values. The data do not seem to follow a Normal distribution, although they are not too asymmetric.

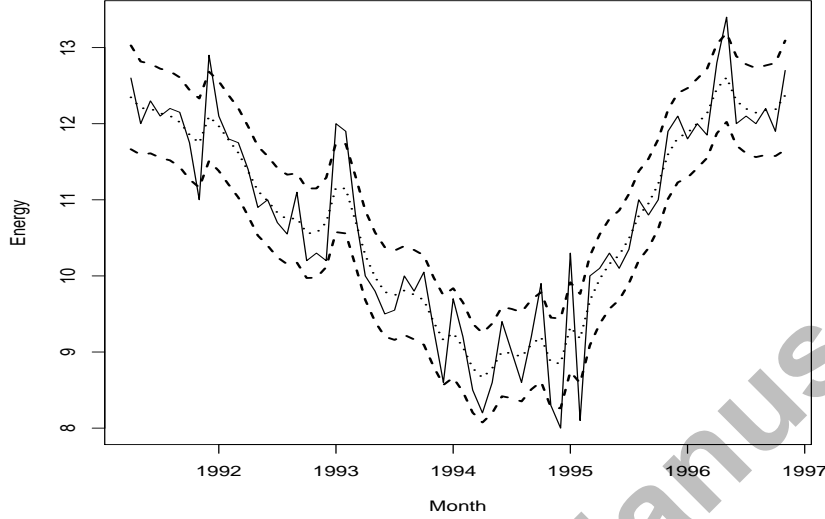The series is non-stationary and has a seasonal component, thus a struc-

Figure 1 *Electric energy consumption of CHESF. Values of the series were divided by 100. The dotted and dashed lines represent, respectively, the smoothing point estimates and the confidence intervals of level 95% for the level component using PT metohd.*

tural basic model (SBM) is fitted to the series (Harvey (1989)). For the SBM, the hyperparameter vector is given by $\psi_t = (\sigma_\epsilon^2, \sigma_\eta^2, \sigma_\xi^2, \sigma_\omega^2)$, where the components represent, respectively, the variance of the error, level, slope and seasonality.
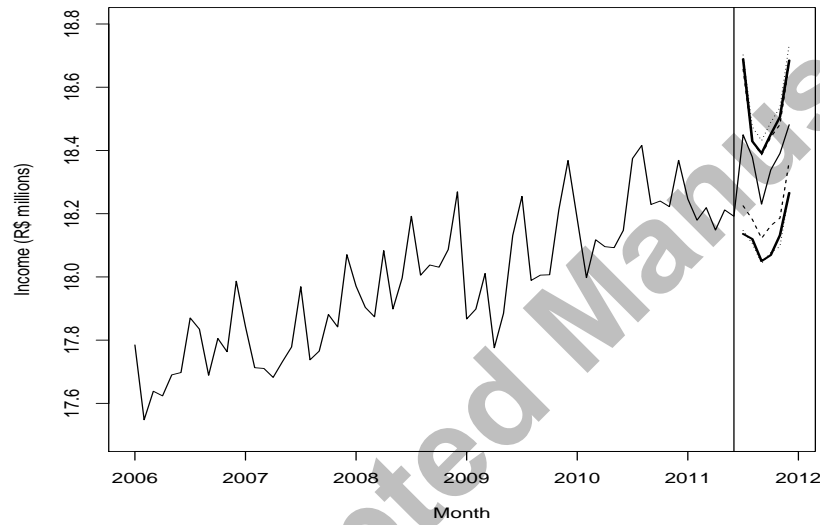
The residuals of the fitted model do not seem to follow a normal distribution, but they are not autocorrelated. Table 8 presents the point and interval estimates for the hyperparameters. The point estimates for the slope (σ_ξ^2) and seasonal (σ_ω^2) hyperparameters are very close to zero, implying that these components may have a deterministic behavior. For this application, the forecast intervals S, HaB and PBFI were calculated and they are shown in Figure 2. It can be seen that the real future values are all of them inside the prediction intervals. Note that the HaB and PBFI intervals present slightly larger widths compared to S, as it was shown in the simulation studies.

7 Conclusion and final remarks

In this work, different bootstrap procedures to estimate the PMSE, which take into account the uncertainty associated to the hyperparameters estima-

Table 8 *MLE and bootstrap confidence intervals of nominal level 95% for the hyperparameters of the SBM fitted to the income series.*

ψ	Estimate	Bootstrap confidence interval
σ_η^2	0.001	$(5.61 \times 10^{-12} \; ; \; 0.015)$
σ_ϵ^2	0.005	$(2.75 \times 10^{-9} \; ; \; 0.001)$
σ_ξ^2	1.14×10^{-10}	$(7.15 \times 10^{-16} \; ; \; 1.157 \times 10^{-5})$
σ_ω^2	2.66×10^{-8}	$(5.15 \times 10^{-21} \; ; \; 0.0005)$

Figure 2 *The full, dashed, dotted and full in bold lines indicate, respectively, the income series and the forecast intervals S, HaB and $PBFI$. The vertical line separates the data and the future values. The confidence level is 0.95.*

tion, have been proposed and compared for the SSM. These proposals are variations of the procedures developed in Hamilton (1986) and Ansley and Kohn (1986), along with a new parametric bootstrap procedure (PBFI). The methods were empirically compared to the standard procedure and to the proposal of Pfeffermann and Tiller (2005), under Gaussian and non-Gaussian assumptions for the errors, in the forecasting and smoothing contexts. The results confirmed that the performance of the bootstrap prediction intervals is slightly better than the Standard intervals with respect to the coverage rate, in the case of Gaussian distribution. The PMSE for the future observations was corrected using the bootstrap technique and had a larger value than the S procedure, but this correction does not seem to drastically affect

the coverage rate of the intervals. The results also showed that the presence of non-Gaussian errors with a very strong asymmetry directly interferes in the coverage rate in the tails.

The major advantage of the bootstrap methods (HaB and AKB) addressed here is their computational simplicity, that is, they require less replications compared to the PT method and they are simple to code. With respect to the computational time, the HaB and AKB spend approximately one third less time than the PT procedure.

Smoothing results were very satisfactory, especially for the simple AKB and HaB procedures, taking into account the quality of the estimates.

Future research includes the development of forecasting methods that do not employ the normality assumption for future values. This includes the use of asymmetric distributions for correcting the coverage rate of the intervals. New possibilities for the G continuous distribution, other than the Gaussian, can also be studied in the PBFI method.

References

Ansley, C.F. and Kohn, R. (1986). Prediction mean square error for state space models with estimated parameters, *Biometrika* **73**, 467-473.

Brockwell, P.J. and Davis, R.A. (1996). Introduction to Time Series and Forecasting, Springer text in Statistics, New York: Springer.

De Jong, P. (1989). Smoothing and interpolation with the state space model, *Journal of American Statistical Association* **84**, 1085-1088.

Doornik, J. A. (2002). Ox: An Object-Oriented Matrix Language, 3. ed. London: Timberlake Consultants Press.

Franco, G.C., Santos, T.R., Ribeiro, J.A. and Cruz, F.R.B. (2008). Confidence intervals for hyperparameters in structural models, *Communications in Statistics: Simulation and Computation* **37**, 486-497.

Hamilton, J.D. (1986). A standard error for the estimated state vector of a state space model, *Journal of Econometrics* **33**, 387-397.

Harvey, A.C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge: University Press.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems, Trans. ASME J. Basic Eng. **82**, 35-45.

Koopman, S.J., Shephard, N. and Doornik, J.A. (1999). Statistical algorithms for inference in state space using SsfPack 2.2, *The Econometrics Journal* **47**, 107-160.

Pfeffermann, D. and Tiller, R. (2005). Bootstrap approximation to prediction MSE for state-space models with estimated parameters. *Journal of Time Series Analysis* **26**, 893-916.

Quenneville, B. and Singh, A.C. (2000). Bayesian prediction mean squared error for state space models with estimated parameters, *Journal of Time Series Analysis* **21**, 219-236.

Rodriguez, A. and Ruiz, E. (2009). Bootstrap prediction intervals in state space models, *Journal of Time Series Analysis* **30**, 167-178.

Rodríguez, A. and Ruiz, E. (2012). Bootstrap prediction mean squared errors of unob-

served states based on the Kalman filter with estimated parameters, *Computational Statistics & Data Analysis* **56**, 62-74.

Santos, T.R. and Franco, G.C. (2010). Inference for the hyperparameters of structural models under classical and Bayesian perspectives: a comparison study, *Communications in Statistics: Simulation and Computation* **39**, 1671-1689.

Shephard, N. (1993). Maximum likelihood estimation of regression models with stochastic trend components, *Journal of American Statistical Association* **88**, 590-595.

Stoffer, D.S. and Wall, K.D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter, *Journal of American Statistical Association* **86**, 1024-1033.

Tsimikas, J. and Ledolter, J. (1994). REML and best linear unbiased prediction in state space models, Communications in Statistics: Theory and Methods **23**, 2253-2268.

Wall, K.D. and Stoffer, D.S. (2002). A state space approach to bootstrapping conditional forecasts in ARMA models, *Journal of Time Series Analysis* **23**, 733-751.

West, M. and Harrison, J. (1997). Bayesian Forecasting and Dynamic Models, New York: Springer.

Yamamoto, T. (1976). Asymptotic mean squared error prediction for an autoregressive model with estimated coefficients, *Applied Statistics* **25**, 123-7.