# Confidence Intervals for the Hyperparameters in Structural Models

Glaura C. Franco†, Thiago R. Santos,
Juliana A. Ribeiro, and F. R. B. Cruz

*Department of Statistics,*
*Federal University of Minas Gerais,*
*31270-901 - Belo Horizonte - MG, Brazil*
*E-mail: {glauraf,thiagors,jujujar,fcruz}@ufmg.br*

November 13, 2006

**Abstract**

This paper deals with the bootstrap as an alternative method to construct confidence intervals for the hyperparameters of structural models. The bootstrap procedure considered is the classical nonparametric bootstrap in the residuals of the fitted model using a well-known approach. The performance of this procedure is empirically obtained through Monte Carlo simulations implemented in Ox. Asymptotic and percentile bootstrap confidence intervals for the hyperparameters are built and compared by means of the coverage percentages. The results are similar but the bootstrap procedure is better for small sample sizes. The methods are applied to a real time series and confidence intervals are built for the hyperparameters.

**Keywords:** state space models, asymptotic confidence intervals, nonparametric bootstrap.

## 1 Introduction

One way of modeling a time series is through its decomposition in non-observable components, using structural models (Harvey, 1989). This technique, in spite of its simplicity

†Corresponding author: Prof. Glaura C. Franco. E-mail: glauraf@ufmg.br. Phone: (+55 31) 3499 5949. Fax: (+55 31) 3499 5929.

due to the direct interpretation of the components, was initially undertaken due to computational difficulties involved in the implementation of the Kalman filter. Nowadays, the models can be adjusted either using new softwares developed specially to fit structural models, such as STAMP (Koopman et al., 1995) or by means of some subroutines available in R (R Development Core Team, 2005) or Ox (Doornik, 1999).

Parameter estimation, in many of the structural models, can be performed by simply estimating the variance of the errors of the non-observable components, named hyperparameters. Inference about these quantities can be done using the asymptotic distribution of the hyperparameters, in which the calculation of the asymptotic variance can be performed in an analytic or computational way (Harvey, 1989). More details about the use of structural models can be found in Harvey (2001), Harvey et al. (2004) and Durbin & Koopman (2001).

Some attention has also been given in applying the nonparametric bootstrap methodology in time series data. The most common alternative to perform the bootstrap in discrete time series is to resample the residuals of the fitted model, which are generally uncorrelated if there is not order misspecification (Efron & Tibshirani, 1993). Recently, different bootstrap procedures in time series data have been one of the main research focus in the area (for instance, see Silva et al., 2006, and references therein). Some references to the use of bootstrap in structural models include Stoffer & Wall (1991), Franco et al. (1999), Franco & Souza (2002), Davis & Yam (2003), Pfeffermann & Tiller (2004) and Olsson & Rydén (2005). However, there is still not many works on the use of bootstrap procedures in the estimation and test of the hyperparameters of structural models, and this is the main motivation of this paper.

The main objective of this work is to test, empirically, the efficiency of the bootstrap proposed by Stoffer & Wall (1991) in order to make inferences about the hyperparameters of structural models, extending the work of Franco & Souza (2002) and considering, besides the local level model (LLM) and the local linear trend model (LLT), a wide class of structural models, named the basic structural model (BSM). In addition, percentile bootstrap confidence intervals are built and compared to the asymptotic intervals by means of the coverage percentage and width of the intervals. Additionally, this work will enable future developments of inference procedures, such as the hypothesis tests to verify the significance of the hyperparameters. An application to a real time series is also presented, as an illustration of the methodology.

This paper is organized as follows. Section 2 presents the structural models considered in this work, as well as the bootstrap technique. Section 3 presents some Monte Carlo simulations and the confidence intervals for the hyperparameters. Section 4 applies the methodology to a real time series and Section 5 concludes the work.

# 2 Structural Models and Bootstrap

In this section, the structural models considered in this work and the construction of the bootstrap series are described in some details.

## 2.1 Structural Models

Decomposition models, also known as structural models, are often used to model and predict a time series. This kind of modeling assumes that the characteristic movements of a time series can be decomposed in four components:

1. trend ($\mu_t$): refers to the general direction a series takes over a large time interval;
2. seasonal component ($\gamma_t$): refers to a similar pattern a series seems to follow during successive period of times, resulting from periodic events that occur annually;
3. cyclical component ($\psi_t$): refers to the oscillations occurring on a long turn;
4. random component or error ($\epsilon_t$): refers to the sporadic movings of a series, due to casual events.

Thus, a very general model to a univariate time series can be written as:

$$y_t = \mu_t + \gamma_t + \psi_t + \epsilon_t, \tag{1}$$

in which $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$, independent.

The structural models considered in this work are the local level model (LLM), the local linear trend model (LLT), and the basic structural model (BSM). For example, the

BSM is defined by dropping $\psi_t$ from Eq. (1) and using the following equations to model the non-observable components

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2), \tag{2}$$

$$\beta_t = \beta_{t-1} + \xi_t, \qquad \xi_t \sim \mathcal{N}(0, \sigma_\xi^2), \tag{3}$$

$$\gamma_t = -\gamma_{t-1} - \ldots - \gamma_{t-s+1} + \omega_t, \qquad \omega_t \sim \mathcal{N}(0, \sigma_\omega^2), \tag{4}$$

in which $\epsilon_t$, $\eta_t$, $\xi_t$, and $\omega_t$ are white noise disturbances mutually non-correlated and $s$ is the size of the seasonal component. The seasonal component $\gamma_t$ is calculated here performing a stochastic modeling using the dummy approach $\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t$. The LLT is obtained by dropping Eq. (4) and the components $\gamma_t$ and $\psi_t$ from Eq. (1). The LLM is obtained by dropping Eq. (3) and Eq. (4), the component $\beta_{t-1}$ from Eq. (2), and the components $\gamma_t$ and $\psi_t$ from Eq. (1). More details about structural models can be found in Harvey (1989).

The hyperparameter vector $\boldsymbol{\psi} = \left(\sigma_\epsilon^2, \sigma_\eta^2, \sigma_\xi^2, \sigma_\omega^2\right)$ can be estimated maximizing the likelihood function

$$L(\boldsymbol{\psi}) = (y_1, \ldots, y_n \mid \ \boldsymbol{\psi}) = \prod_{t=1}^{n} f\left(y_t \mid \boldsymbol{Y}_{t-1}, \boldsymbol{\psi}\right),$$

in which $\boldsymbol{Y}_{t-1} = (y_1, y_2, \ldots, y_{t-1})$ and $n$ is the size of the series.

One way of calculating the predictive distribution $f\left(y_t \mid \boldsymbol{Y}_{t-1}, \boldsymbol{\psi}\right)$ is by means of the Kalman filter (Kalman, 1960). The Kalman filter is a recursive procedure which decomposes the series in its non-observable components through recursive equations that update sequentially the state vector, once the model is in the state space form. The state of the process condenses all the necessary information from the past to predict the future.

A general representation for the state space form is given by

$$y_t = \boldsymbol{z}_t' \boldsymbol{\alpha}_t + \epsilon_t \tag{5}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{T}_t \boldsymbol{\alpha}_{t-1} + \boldsymbol{R}_t \boldsymbol{\eta}_t, \quad t = 1, ..., n \tag{6}$$

in which $\boldsymbol{\alpha}_t$ is a $p \times 1$ unobservable state vector, the disturbances $\epsilon_t$ and $\boldsymbol{\eta}_t$ are serially uncorrelated, with zero means and variances given by $h_t$ and $\boldsymbol{Q}_t$, respectively, and $\boldsymbol{z}_t$,

$\boldsymbol{T}_t$ and $\boldsymbol{R}_t$ are the system matrices. It is straightforward to write the LLM, LLT and SBM models in the state space form and to implement the Kalman filter. For details see Harvey (1989).

Given that the model is in the state space form, the likelihood function can be written in the following way

$$f(y_t|\boldsymbol{Y}_{t-1}, \boldsymbol{\psi}) = (2\pi)^{-1/2}|F_t|^{-1/2} \exp\left\{(-1/2)(y_t - \tilde{y}_{t|t-1})' F_t^{-1}(y_t - \tilde{y}_{t|t-1})\right\},$$

in which $\tilde{y}_{t|t-1}$ is the one-step-ahead prediction and $F_t$ is the variance of the one-step-ahead prediction error, given by $\nu_t = y_t - \tilde{y}_{t|t-1}$. The values $\tilde{y}_{t|t-1}$ , $F_t$ and $\nu_t$ can be obtained through the Kalman filter.

As the likelihood function is nonlinear, the estimation should be done via numerical methods. The maximization method used here is the BFGS, taking advantage of Ox (Doornik, 1999), which has a fair built-in implementation of this well-known optimization algorithm.

Harvey (1989) states that, under some regularity conditions, the maximum likelihood estimator, $\hat{\boldsymbol{\psi}}$, is asymptotically normal with mean $\boldsymbol{\psi}$ and covariance matrix $aVar(\boldsymbol{\psi}) = n^{-1}\mathbf{IA}^{-1}(\boldsymbol{\psi})$, in which

$$\mathbf{IA}(\boldsymbol{\psi}) = \lim_{n \to \infty} n^{-1}\mathbf{I}(\boldsymbol{\psi}),$$

and $\mathbf{I}(\boldsymbol{\psi})$ is the information matrix.

**Asymptotic Confidence Intervals**

In order to calculate the variances of the hyperparameters, Harvey (1989) proposes a numerical approximation to calculate the derivatives of $\nu_t$ and $F_t$, which are extremely useful in calculating the Fisher information matrix, whose form is given by

$$I_{ij}(\psi) = \frac{1}{2}\sum_t \left\{ \mathrm{tr}\left[ F_t^{-1}\frac{\partial F_t}{\partial \psi_i} F_t^{-1}\frac{\partial F_t}{\partial \psi_j}\right]\right\} + \mathrm{E}\left\{\sum_t \left(\frac{\partial \nu_t}{\partial \psi_i}\right)' F_t^{-1}\frac{\partial \nu_t}{\partial \psi_j}\right\},$$

in which $i, j = 1, \ldots, k$, $t = 1, \ldots, n$, and $k$ is the number of hyperparameters to be estimated.

The derivatives of $\nu_t$ and $F_t$ can be calculated through the following procedure:

1. for $i = 1, \ldots, k$, a small quantity, $\delta_i$, is added to $\psi_i$;

2. the Kalman filter is run with the new value of the hyperparameter, but keeping fixed the values of the other hyperparameters;

3. a new prediction error vector, $\nu_t^{(i)}$, and its covariance matrix, $F_t^{(i)}$, are obtained;

4. the expressions $\delta_i^{-1}[\nu_t^{(i)} - \nu_t]$ and $\delta_i^{-1}[F_t^{(i)} - F_t]$ are then numerical approximations to the required derivatives.

The choice of this small quantity, $\delta$, to be presented in Section 4, will be performed by a Monte Carlo study.

A $100(1-\kappa)\%$ asymptotic confidence interval for $\boldsymbol{\psi}$ is then given by

$$\hat{\boldsymbol{\psi}} \pm z_{\kappa/2}\sqrt{a\mathrm{Var}(\boldsymbol{\psi})},$$

in which $z_{\kappa/2}$ is the $\kappa/2$ percentile of the Normal distribution.

## 2.2 Bootstrap

The bootstrap is a resample method (Efron, 1979), used mainly to make inferences about the parameters of a given model. This is done by taking resamples (with replacement) of the original data and trying to approximate the distribution of a function of the observations by the empirical distribution of the data.

In this work, the bootstrap is done in the residuals of the fitted model, following the work of Stoffer & Wall (1991). Using the Kalman filter, the innovations, $\nu_t$, and their variances, $F_t$, are obtained recursively. As they are a function of the hyperparameters $\boldsymbol{\psi}$, the notation $\nu_t(\boldsymbol{\psi})$ and $F_t(\boldsymbol{\psi})$ will be used. First the hyperparameters should be estimated and the innovations, $\nu_t(\widehat{\boldsymbol{\psi}})$, calculated. Next, the innovations should be re-scaled as

$$e_t(\widehat{\boldsymbol{\psi}}) = \frac{\nu_t(\widehat{\boldsymbol{\psi}}) - \overline{\nu}_t(\widehat{\boldsymbol{\psi}})}{\sqrt{F_t(\widehat{\boldsymbol{\psi}})}},$$

in which $\overline{\nu}_t(\widehat{\boldsymbol{\psi}}) = \frac{\sum_{j=1}^{n} \nu_j(\widehat{\boldsymbol{\psi}})}{n}$.

The bootstrap innovations, $e_t^*(\widehat{\boldsymbol{\psi}})$, are obtained by resampling $e_t(\widehat{\boldsymbol{\psi}})$, with replacement.

Let the vector $\boldsymbol{S}_t$ be defined as

$$\boldsymbol{S}_t = \begin{bmatrix} \boldsymbol{a}_t \\ y_{t-1} \end{bmatrix},$$

in which $\boldsymbol{a}_t$ is a linear estimator for the state vector $\boldsymbol{\alpha}_t$, with variance $\boldsymbol{V}_t$. Then

$$\boldsymbol{S}_{t+1} = \begin{bmatrix} \boldsymbol{T}_t & \boldsymbol{0} \\ \boldsymbol{z}_t & \boldsymbol{0} \end{bmatrix} \boldsymbol{S}_t + \begin{bmatrix} \boldsymbol{T}_t \boldsymbol{V}_t \boldsymbol{z}_t' F_t^{-1} \sqrt{F_t} \\ \sqrt{F_t} \end{bmatrix} e_t. \tag{7}$$

The bootstrap series $y_t^*$ can be calculated solving Eq. (7) by substituting $e_t$ by $e_t^*$ and using the estimated values $F_t(\widehat{\boldsymbol{\psi}})$, $\boldsymbol{a}_t(\widehat{\boldsymbol{\psi}})$ and $\boldsymbol{V}_t(\widehat{\boldsymbol{\psi}})$ obtained from the Kalman filter.

**Bootstrap Confidence Intervals**

The percentile bootstrap confidence interval (Efron & Tibshirani, 1986) will be used in this work to build intervals for the hyperparameters. In practice, after estimating the hyperparameter values for each one of the $B$ bootstrap series, the values are ordered and the 100$\kappa$-th value is taken as the inferior limit and the 100(1-$\kappa$)-th as the superior limit of the interval. Thus, the percentile bootstrap interval is given by

$$\left[\hat{\boldsymbol{\psi}}^{(\kappa)};\ \hat{\boldsymbol{\psi}}^{(1-\kappa)}\right].$$

# 3 Simulation Results

All the algorithms described were implemented in Ox and are available from the authors upon request. The estimates were compared through their means and mean square errors (MSE) and the confidence intervals were compared by means of the coverage percentage. The nominal level was fixed at 95%. The performance of the bootstrap procedure was investigated by carrying out a series of simulations in the LLM, LLT and SBM. Series sizes of $n = \{50, 100, 200\}$, bootstrap resamples of $B = 1000$ , and Monte Carlo replications of MC $= 500$ were considered. The hyperparameter values were fixed at $\sigma_\varepsilon^2 = 1.0$, $\sigma_\eta^2 = 0.5$, for the LLM, $\sigma_\varepsilon^2 = 1.0$, $\sigma_\eta^2 = 0.5$ and $\sigma_\xi^2 = 0.1$, for the LLT, and $\sigma_\varepsilon^2 = 1.0$, $\sigma_\eta^2 = 0.5$, $\sigma_\xi^2 = 0.03$, and $\sigma_\omega^2 = 0.1$, for the SBM. These choices were based on earlier studies (from Ribeiro, 2006, not shown). The burn-in used was equal to 100 and the number of iterations in the BFGS algorithm equal to 50.

## 3.1 Determining $\delta$

In order to calculate the Fisher information matrix and to build the asymptotic intervals, a Monte Carlo study was performed with the aim of defining $\delta$, which was varied from $10^{-6}$ to $10^{2}$, in powers of 10, for a sample size of 100. For each value of $\delta$, 1000 Monte Carlo simulations were performed. The asymptotic intervals for the hyperparameter of each series were calculated and the inferior and superior limits were averaged for the intervals obtained in the simulations. Figures 1, 2 and 3 show the results of the Monte Carlo simulation for the asymptotic intervals in the LLM, LLT, and BSM, respectively. It can be observed that the range of the intervals tends to increase as $\delta$ increases. For values of $\delta$ up to 0,0001 there is not a substantial alteration in the range of the intervals. Therefore this value was selected for the calculation of the asymptotic interval.

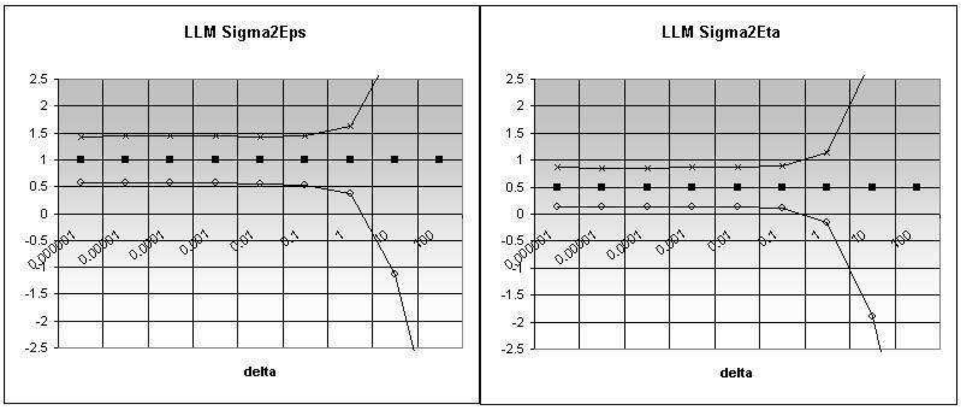


Figure 1: Monte Carlo study to determine $\delta$ in the LLM.

## 3.2 Comparing the Asymptotic and Bootstrap Intervals

Table 1 presents the maximum likelihood estimates (MLE) and the MSE for the Monte Carlo and bootstrap replications, as well as the coverage percentages of the asymptotic and bootstrap intervals for the LLM. It can be observed that the MLE's are always very close to the fixed value of the hyperparameter (the distance is in order of $10^{-3}$). The bootstrap procedure seems to approximate satisfactorily the values of the estimates obtained from the Monte Carlo replications, and they become closer as $n$ increases. The coverage of the intervals also tends to the nominal value of 95%, as the sample size increases. It should be noted that the coverage of the bootstrap intervals is better than

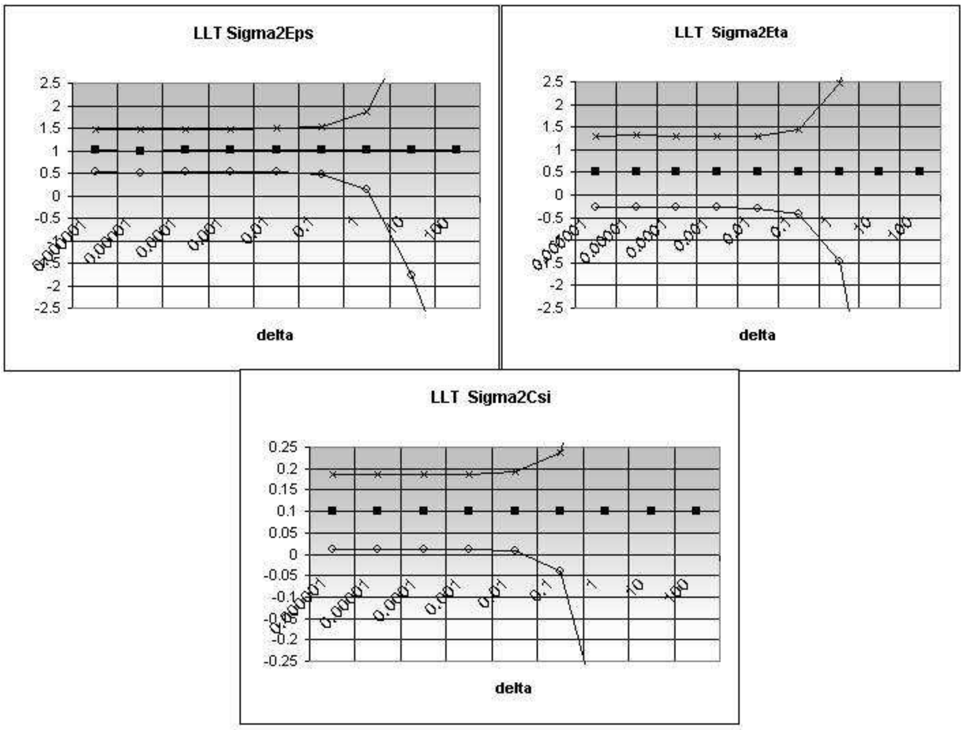


Figure 2: Monte Carlo study to determine $\delta$ in the LLT.

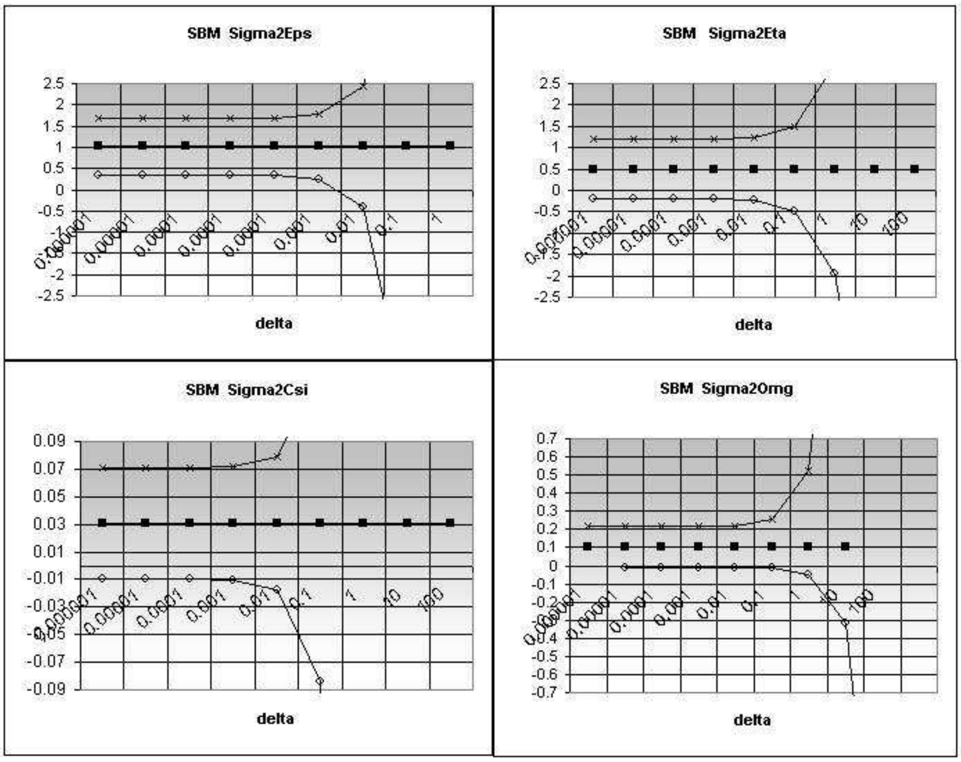


Figure 3: Monte Carlo study to determine $\delta$ in the BSM.

the asymptotic interval for $\sigma_\eta^2$, for any size of the series, while for $\sigma_\varepsilon^2$, they are practically the same. One of the disadvantages of the asymptotic interval is that it can lead to negative values for the inferior limit (for example, see $\sigma_\eta^2$ and $n = 50$).

Tables 2 and 3 present the maximum likelihood estimates (MLE) and the MSE for the Monte Carlo and bootstrap replications, and also the coverage percentages of the asymptotic and bootstrap intervals for the LLT and SBM, respectively. The results are quite the same as the ones obtained for the LLM. For the SBM (see Table 3), the superiority of the bootstrap confidence interval is more evident, as it is seen that, even for sizes as small as 50, the coverage percentages are closer to the 95% nominal level, varying in the range 93% to 97% (except for $\sigma_\omega^2$).

# 4 Application to the APCI Series

In this section, the methodology described in the previous sections is applied to a series of inflation index in Belo Horizonte, Brazil. The series is the Ample Price to Consumer Index (APCI), calculated by IPEAD*. This index measures the evolution of the incomes in families spending from 1 to 40 minimum salaries per month. The APCI series is composed of 105 monthly observations in the period January, 1997 to October, 2005.

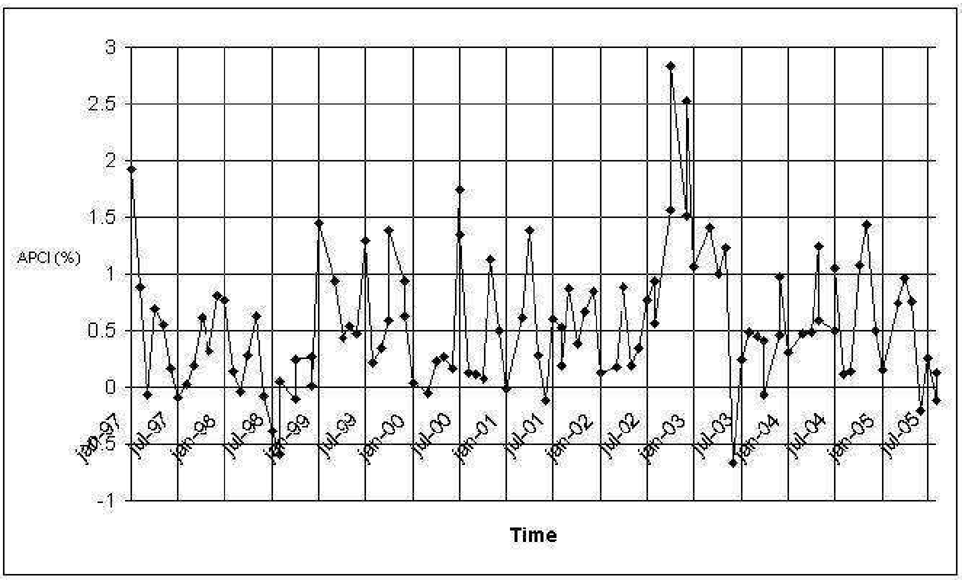


Figure 4: Plot of the APCI series.

*Fundação Instituto de Pesquisas Econômicas, Administrativas e Contábeis de Minas Gerais - Brazil*

Table 1: MLE, confidence intervals, and coverage for the LLM.

| | | MLE | | Confidence Intervals | | Coverage | |
|---|---|---|---|---|---|---|---|
| | | Monte Carlo | Bootstrap | Asymptotic | Bootstrap | Asymptotic | Bootstrap |
| $n$ | $\psi$ | (MSE) | (MSE) | (range) | (range) | | |
| 50 | $\sigma^2_\eta = 0.5$ | 0.509 | 0.566 | [-0.012;1.030] | [0.107;1.276] | 0.84 | **0.90** |
| | | (0.090) | (0.109) | (1.042) | (1.169) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 1.008 | 0.976 | [0.389;1.628] | [0.359;1.653] | **0.91** | 0.90 |
| | | (0.118) | (0.127) | (1.239) | (1.294) | | |
| 100 | $\sigma^2_\eta = 0.5$ | 0.495 | 0.525 | [0.132;0.859] | [0.197;0.975] | 0.88 | 0.90 |
| | | (0.042) | (0.048) | (0.727) | (0.778) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 1.010 | 0.990 | [0.571;1.448] | [0.556;1.461] | **0.93** | 0.93 |
| | | (0.056) | (0.059) | (0.877) | (0.905) | | |
| 200 | $\sigma^2_\eta = 0.5$ | 0.504 | 0.515 | [0.243;0.765] | [0.277;0.808] | 0.93 | **0.94** |
| | | (0.016) | (0.017) | (0.522) | (0.531) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 1.000 | 0.992 | [0.690;1.310] | [0.689;1.319] | 0.93 | 0.94 |
| | | (0.026) | (0.027) | (0.620) | (0.630) | | |
| 500 | $\sigma^2_\eta = 0.5$ | 0.496 | 0.502 | [0.333;0.659] | [0.348;0.674] | **0.95** | **0.95** |
| | | (0.007) | (0.007) | (0.326) | (0.326) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 0.999 | 0.995 | [0.803;1.194] | [0.804;1.195] | 0.94 | **0.95** |
| | | (0.010) | (0.011) | (0.391) | (0.391) | | |

Obs.: See in bold the closest percentages to the nominal level of 95%.

Table 2: MLE, confidence intervals, and coverage for the LLT.

| | | MLE | | Confidence Intervals | | Coverage | |
|---|---|---|---|---|---|---|---|
| | | Monte Carlo | Bootstrap | Asymptotic | Bootstrap | Asymptotic | Bootstrap |
| $n$ | $\psi$ | (MSE) | (MSE) | (range) | (range) | | |
| 50 | $\sigma^2_\eta = 0.5$ | 0.606 | 0.633 | [-0.732;1.948] | [0.001;2.098] | **0.99** | **0.99** |
| | | (0.450) | (0.273) | (2.680) | (2.097) | | |
| | $\sigma^2_\xi = 0.1$ | 0.095 | 0.104 | [-0.055;0.246] | [0.002;0.287] | 0.77 | **0.88** |
| | | (0.006) | (0.006) | (0.301) | (0.285) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 0.951 | 0.943 | [0.140;1.760] | [0.177;1.757] | 0.91 | **0.95** |
| | | (0.185) | (0.156) | (1.620) | (1.580) | | |
| 100 | $\sigma^2_\eta = 0.50$ | 0.555 | 0.582 | [-0.379;1.489] | [0.032;1.655] | **0.96** | 0.98 |
| | | (0.240) | (0.213) | (1.868) | (1.623) | | |
| | $\sigma^2_\xi = 0.1$ | 0.088 | 0.093 | [-0.009;0.186] | [0.015;0.201] | 0.80 | **0.87** |
| | | (0.003) | (0.002) | (0.195) | (0.186) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 0.980 | 0.967 | [0.401;1.558] | [0.337;1.536] | 0.94 | **0.95** |
| | | (0.090) | (0.088) | (1.157) | (1.199) | | |
| 200 | $\sigma^2_\eta = 0.5$ | 0.523 | 0.527 | [-0.159;1.204] | [0.046;1.294] | **0.97** | 0.98 |
| | | (0.120) | (0.111) | (1.363) | (1.248) | | |
| | $\sigma^2_\xi = 0.1$ | 0.100 | 0.103 | [0.024;0.175] | [0.037;0.185] | 0.90 | **0.94** |
| | | (0.001) | (0.001) | (0.151) | (0.148) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 0.982 | 0.980 | [0.566;1.398] | [0.553;1.390] | **0.96** | **0.96** |
| | | (0.045) | (0.044) | (0.832) | (0.837) | | |
| 500 | $\sigma^2_\eta = 0.5$ | 0.497 | 0.485 | [0.065;0.929] | [0.110;0.945] | **0.93** | **0.93** |
| | | (0.055) | (0.054) | (0.864) | (0.835) | | |
| | $\sigma^2_\xi = 0.1$ | 0.103 | 0.101 | [0.053;0.149] | [0.059;0.155] | **0.95** | 0.94 |
| | | (0.001) | (0.001) | (0.096) | (0.096) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 0.999 | 1.005 | [0.734;1.265] | [0.735;1.271] | **0.94** | **0.94** |
| | | (0.020) | (0.020) | (0.531) | (0.536) | | |

Obs.: See in bold the closest percentages to the nominal level of 95%.

Table 3: MLE, confidence intervals, and coverage for the SBM.

| | | MLE | | Confidence Intervals | | Coverage | |
|---|---|---|---|---|---|---|---|
| | | Monte Carlo | Bootstrap | Asymptotic | Bootstrap | Asymptotic | Bootstrap |
| $n$ | $\psi$ | (MSE) | (MSE) | (range) | (range) | | |
| 50 | $\sigma^2_\eta = 0.5$ | 0.547 | 0.506 | [-0.428;1.520] | [0.000;1.793] | 0.96 | **0.95** |
| | | (0.241) | (0.165) | (1.948) | (1.793) | | |
| | $\sigma^2_\xi = 0.03$ | 0.030 | 0.038 | [-0.031;0.090] | [0.000;0.119] | 0.74 | **0.93** |
| | | (0.001) | (0.001) | (0.121) | (0.119) | | |
| | $\sigma^2_\omega = 0.1$ | 0.108 | 0.095 | [-0.073;0.289] | [0.001;0.298] | 0.83 | **0.88** |
| | | (0.010) | (0.007) | (0.362) | (0.297) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 0.932 | 1.083 | [0.029;1.834] | [0.122;2.157] | 0.89 | **0.97** |
| | | (0.283) | (0.291) | (1.805) | (2.035) | | |
| 100 | $\sigma^2_\eta = 0.5$ | 0.505 | 0.485 | [-0.179;1.190] | [0.023;1.385] | **0.94** | **0.94** |
| | | (0.138) | (0.127) | (1.369) | (1.362) | | |
| | $\sigma^2_\xi = 0.03$ | 0.0029 | 0.033 | [-0.010;0.068] | [0.002;0.081] | 0.79 | **0.89** |
| | | (0.000) | (0.000) | (0.078) | (0.079) | | |
| | $\sigma^2_\omega = 0.1$ | 0.103 | 0.097 | [-0.007;0.214] | [0.013;0.226] | 0.86 | **0.90** |
| | | (0.004) | (0.003) | (0.221) | (0.213) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 0.982 | 1.064 | [0.336;1.627] | [0.304;1.810] | 0.91 | **0.94** |
| | | (0.136) | (0.162) | (1.291) | (1.506) | | |
| 200 | $\sigma^2_\eta = 0.5$ | 0.499 | 0.466 | [-0.004;1.002] | [0.059;1.105] | **0.92** | **0.92** |
| | | (0.071) | (0.077) | (1.006) | (1.046) | | |
| | $\sigma^2_\xi = 0.03$ | 0.031 | 0.035 | [0.002;0.061] | [0.009;0.070] | 0.87 | **0.93** |
| | | (0.000) | (0.000) | (0.059) | (0.061) | | |
| | $\sigma^2_\omega = 0.1$ | 0.102 | 0.099 | [0.027;0.177] | [0.035;0.184] | 0.93 | **0.94** |
| | | (0.002) | (0.001) | (0.150) | (0.149) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 1.001 | 1.058 | [0.531;1.471] | [0.505;1.582] | 0.93 | **0.95** |
| | | (0.059) | (0.077) | (0.940) | (1.077) | | |
| 500 | $\sigma^2_\eta = 0.5$ | 0.501 | 0.476 | [0.183;0.818] | [0.168;0.833] | **0.93** | **0.93** |
| | | (0.029) | (0.032) | (0.635) | (0.665) | | |
| | $\sigma^2_\xi = 0.03$ | 0.030 | 0.032 | [0.012;0.048] | [0.015;0.052] | 0.91 | **0.92** |
| | | (0.000) | (0.000) | (0.036) | (0.037) | | |
| | $\sigma^2_\omega = 0.1$ | 0.100 | 0.099 | [0.055;0.146] | [0.057;0.148] | 0.91 | **0.92** |
| | | (0.001) | (0.001) | (0.091) | (0.091) | | |
| | $\sigma^2_\varepsilon = 1.0$ | 0.993 | 1.022 | [0.694;1.292] | [0.696;1.348] | **0.95** | 0.94 |
| | | (0.024) | (0.029) | (0.598) | (0.652) | | |

Obs.: See in bold the closest percentages to the nominal level of 95%.

It can be seen from Figure 4 that the series does not seem to have seasonal nor trend components. Therefore, the most probable model to be fit to this series is the LLM. Even thus, the LLT will be fit, in order to check if the confidence intervals built can confirm that the variance of the trend component is close to zero.

Table 4 shows the fit of the LLT. It can be noticed that the hyperparameter $\sigma^2_\xi$ related to the trend component and the respective bootstrap and asymptotic confidence intervals are approximately zero, as it was expected. This can indicate that this component should not be considered in the model.

Table 5 presents the fit of the LLM and the respective bootstrap and asymptotic 95%

confidence intervals. In this case, the maximum likelihood estimates are not that close to zero, as well as the confidence intervals. This can indicate the presence of the stochastic level component $\sigma_\eta^2$, suggesting that the series can follow a LLM. It can be also observed that the bootstrap and asymptotic intervals are very similar.

Table 4: Fit of the LLT to the APCI series.

| | | 95% confidence intervals | |
|---|---|---|---|
| $\psi$ | MLE | Asymptotic | Bootstrap |
| $\sigma_\eta^2$ | 0.050 | [0.000;0.107] | [0.003;0.105] |
| $\sigma_\xi^2$ | 0.000 | [0.000;0.000] | [0.000;0.001] |
| $\sigma_\varepsilon^2$ | 0.198 | [0.123;0.274] | [0.113;0.287] |

Table 5: Fit of the LLM to the APCI series.

| | | 95% confidence intervals | |
|---|---|---|---|
| $\psi$ | MLE | Asymptotic | Bootstrap |
| $\sigma_\eta^2$ | 0.042 | [0.005;0.088] | [0.010;0.090] |
| $\sigma_\varepsilon^2$ | 0.206 | [0.132;0.280] | [0.123;0.305] |

# 5 Conclusions and Final Remarks

This work presents an empirical study of the performance of the bootstrap applied to some structural models, in particular, the local level model (LLM), the local linear trend model (LLT) and the basic structural model (BSM). The approached used was the non-parametric bootstrap in the residuals of the fitted model, which was showed to mimic well the behavior of the hyperparameters of the structural models analyzed, possessing mean and mean square error very close to the Monte Carlo results.

Besides, bootstrap and asymptotic confidence intervals for the hyperparameters were built and compared with respect to the width of the intervals and coverage percentages. It was noticed that the bootstrap intervals were closer to the 95% nominal level than the asymptotic intervals, regardless the size of the series. It should be stressed that the asymptotic intervals can present boundary problems, that is, when the hyperparameter lies on the boundary of the parameter space, as this is one of the regularity conditions needed to calculate the asymptotic distribution. Empirically, it can be seen that the lower limit can be negative, and this should not be expected, as in this case the hyperparameters are the variances of the errors. On the other hand, the computational time for the bootstrap is approximately five times the computational time for the asymptotic interval.

The methodology was also applied to a real series of inflation index in Belo Horizonte, Brazil. The bootstrap and asymptotic intervals were constructed and they led to the conclusion that the series follows a LLM.

Future research includes the construction of hypothesis tests to verify the significance of the components in the structural model.

## Acknowledgments

The authors gratefully acknowledge the CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil, grants 201046/1994-6, 301809/1996-8, 307702/2004-9, 472066/2004-8, and 472877/2006-2, the FAPEMIG (*Fundação de Amparo à Pesquisa do Estado de Minas Gerais*), grants CEX-289/98 and CEX-855/98, and PRPq-UFMG, grant 4081-UFMG/RTR/FUNDO/PRPq/99, for partial financial support.

## References

Davis, R. A. & Yam, G. R. (2003). *Approximate Likelihood Approach*, Colorado State University, Colorado.

Doornik, J. A. (1999). *Ox: An Object-Oriented Matrix Language*, 3rd edn, Timberlake Consultants Press, London.

Durbin, J. & Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*, Oxford Statistical Science Series: Oxford University Press, Oxford.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *The Annals of Statistics* **7**: 1–26.

Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy, *Statistical Science* **1**(1): 54–77.

Efron, B. & Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, London, UK.

Franco, G. C., Koopman, S. J. & Souza, R. C. (1999). Bootstrap tests when parameters of nonstationary time series models lie on the boundary of the parameter space, *Brazilian Journal of Probability and Statistics* **13**(1): 41–54.

Franco, G. C. & Souza, R. C. (2002). A comparison of methods for bootstrapping in the local level model, *Journal of Forecasting* **21**: 27–38.

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

Harvey, A. C. (2001). Testing unobserved component models, *Journal of Forecasting* **20**: 1–19.

Harvey, A. C., Koopman, S. J. & Shephard, N. (2004). *State Space and Unobserved Component Models: Theory and Applications*, University of Cambridge, Cambridge.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Journal Basic Engineering, Transactions ASME, Series D* **82**: 35–45.

Koopman, S. J., Harvey, A. C., Doornik, J. A. & Shephard, N. (1995). *STAMP 5.0: Structural Time Series Analyser, Modeller and Predictor*, Chapman & Hall, London.

Olsson, J. & Rydén, T. (2005). Asymptotic properties of the bootstrap particle filter maximum likelihood estimator for state space models, *Technical report*, Center for Mathematical Sciences. Lund University.

Pfeffermann, D. & Tiller, R. (2004). Bootstrap approximation to prediction MSE for state-space models with estimated parameters, *S3RI Methodology Working Papers M03/05*, Southampton Statistical Sciences Research Institute, Southampton, UK.

R Development Core Team (2005). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Ribeiro, J. A. (2006). *Inferência sobre os hiperparâmetros dos modelos estruturais usando bootstrap (bootstrap inferences on the hyperparameters of structural models)*, Master's thesis, Statistics Department - UFMG, Belo Horizonte, Brazil (in Portuguese).

Silva, E. M., Franco, G. C., Reisen, V. A. & Cruz, F. R. B. (2006). Local bootstrap approaches for fractional differential parameter estimation in ARFIMA models, *Computational Statistics & Data Analysis* **51**(2): 1002–1011.

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