

101 Variations on a Maximum Likelihood Procedure for a Structural Time Series Model

Javier López-de-Lacalle

October 20, 2014

Abstract

Several variations on the implementation of a maximum likelihood procedure to fit the basic structural time series model are discussed in this paper. The paper illustrates through an application that different choices in the parameterization of the model or in the design of the optimization procedure may have a significant effect on the results. Difficulties that practitioners sometimes find to replicate results reported by others are often due to the omission of these issues.

Key words: Kalman filter, seasonality, state-space model, unobserved components

1 Introduction

Structural time series models (STSM) consist of building blocks defining different unobserved or latent components such as a trend, a cycle or a seasonal component. As these models are usually arranged in state-space form and the Kalman filter is used to extract the components, they are sometimes referred to as state-space methods for time series analysis. This name involves a wider class of models, nonetheless. For example, autoregressive moving-average models (ARMA) can be cast into a state space form as well. Although an ARMA model can be decomposed into a trend, cycle and seasonal components, those components do not appear in the definition of the ARMA model. The possibility to explicitly select and define the components in a time series is one of the appeals of STSM.

The common procedure to fit a STSM is to maximize the time-domain likelihood function by means of numerical methods. The likelihood function is evaluated by means of the Kalman filter. Given the optimal parameter values, an optimal signal for the unobserved components can also be obtained via the Kalman smoother. Considering that most statistical packages provide an interface for the Kalman filter as well as a numerical optimization algorithm, the procedure to fit a STSM

may seem trivial. Given the apparent simplicity, the user is supposed to be able to implement by herself a procedure to fit a STSM. In this paper we will act as such user and will design and implement a procedure that fits the basic structural model to the classic Box & Jenkins (Box and Jenkins, 1976) airline passengers data. In doing so, we will see that it requires making several choices about the parameterization of the model and the design of the optimization procedure.

The combination of alternative and equally sensible options leads to a large number of variations for a seemingly simple procedure. The motivation to develop such an extensive variety of versions of the same procedure lies on the difficulties that are sometimes observed and experienced when replicating the results of a particular application. Sometimes we find different maximum likelihood estimates reported for the same model and data. For example, the documentation in the R statistical software (R Core Team, 2012) points out that the results obtained for the airline passengers data using the function `StructTS`¹ are more reliable than those reported on another source. In other cases, the results are not robust when the model is applied to relatively long subsamples of the original data. For example, Harvey (1989, §4.5.4) fits the same model and data but excludes the last two years in order to assess the performance of the forecasts based on the model. The results differ from those found in the previous references. Considering that the basic structural model is suited to capture the dynamics observed in the airline passengers series (Harvey, 1989, §2.7.5), large discrepancies between different implementations or subsamples were not expected. Contrary to the common view, it is sometimes easier to replicate results based on Monte Carlo methods such as the Gibbs sampler rather than results based on a maximum likelihood procedure. This is a puzzling situation that could be avoided if the particularities of each implementation were documented. This paper gives some guidelines about the issues that shouldn't be overlooked when describing an application or documenting a software package. At the same time, this paper gives an insight into the issues that may affect the performance of the optimization procedure in structural time series models.

The disparity of results is not intrinsic to the context of these models. It may also arise in other scenarios due to the presence of several local optima or other peculiarities of the objective function. Here, we do not pursue an analysis to detect all possible local optima. In that case, we could run a given algorithm for different starting values of the parameters. The solutions found in the present study are determined by the design of the procedure to fit the basic structural time series model rather than by the use of different starting points in the optimization algorithm. In the alternative framework of ARIMA models it is rare to find major differences in the results across

¹ This function is available in the `stats` core package.

implementations of the procedure to fit these models. It is therefore worth exploring the options and features that characterize alternative procedures to fit the basic structural model.

Despite these difficulties, the growing development of software packages for structural time series models reflects the usefulness and compelling performance of these models. Recently, the special issue ‘Statistical Software for State Space Methods’ published in the *Journal of Statistical Software* [Commandeur et al. \(2011\)](#) reviewed a large variety of software tools for time series analysis in the context of our application. Among that software, STAMP ([Koopman et al., 2009](#)) is the longest in existence and probably the most specialized software for STSM. The articles in that issue fit the local level model (random walk plus noise) to the series of annual flow in the rive Nile in order to illustrate the usage of the software. They also introduce some further functionalities available in each software tool. In this relatively simple application discrepancies in the results across software tools were not found.

Another interesting review relevant for our present discussion is given in [Tusell \(2011\)](#). The article compares the performance of different interfaces available in R to run the Kalman filter (the same interfaces that will be considered in this paper). The packages introduced in [Tusell \(2011\)](#) provide flexible interfaces to cast a broad range of time series models in state-space form as well as sound implementations of the Kalman filter. However, little attention is given to the procedure that optimizes the likelihood function. A general purpose algorithm –the L-BFGS-B algorithm– is generally used. The concern of this paper is the algorithm that *optimizes* the objective function, not the algorithm that *evaluates* the objective function.

In practice, most of the practitioners confine the analysis to the capabilities and options provided by the software that they are more familiar with. A comparison of software tools is beyond the scope of this paper. The focus of the present study is the design and comparison of variations on the baseline procedure to fit a STSM by maximum likelihood. This paper reveals some issues that are often neglected by practitioners when describing or conducting an application or when documenting a software tool. The discussion is illustrated with an application to the airline passengers data, which is often taken as an example to apply the basic structural model and other time series models.

The remaining of the paper is organized as follows. Section [2](#) introduces the basic structural model. The variations on the maximum likelihood procedure are described in Section [3](#). A discussion based on the application to the airline passengers data is presented in Section [4](#). Section [5](#) gives some concluding remarks. The R scripts that replicate the results are provided along with this paper.

2 The basic structural model

The basic structural model (BSM) is a pure variance model commonly used in applications. It plays a central role in the approach advocated in [Harvey \(1989\)](#) for time series analysis. The model is defined as follows:

$$\begin{aligned}
 \text{observed series: } y_t &= \mu_t + \gamma_t + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \sigma_\epsilon^2); \\
 \text{latent level: } \mu_t &= \mu_{t-1} + \beta_{t-1} + \xi_t, & \xi_t &\sim \text{NID}(0, \sigma_\xi^2); \\
 \text{latent drift: } \beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim \text{NID}(0, \sigma_\zeta^2); \\
 \text{latent seasonal: } \gamma_t &= \sum_{j=1}^{s-1} -\gamma_{t-j} + \omega_t, & \omega_t &\sim \text{NID}(0, \sigma_\omega^2),
 \end{aligned}$$

for $t = s, \dots, n$; s is the periodicity of the data (e.g. $s = 4$ for quarterly data).

A detailed insight into the theoretical properties of this model can be found in the latter reference. The model provides a flexible framework suitable to analyze many of the observed macroeconomic time series. Setting $\sigma_\omega^2 = 0$ yields a model with deterministic seasonality. Setting also $\gamma_1 = \dots = \gamma_{s-1} = 0$ and $\sigma_\zeta^2 = 0$ removes the seasonal component and gives the local level model (random walk plus noise model with drift β_0). If $\sigma_\zeta^2 > 0$ we obtain the local trend model, where the drift follows a random walk.

The state space form of the basic structural model is given by the following representation:

$$\begin{aligned}
 y_t &= Z\alpha_t + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \sigma_\epsilon^2), \\
 \alpha_t &= T\alpha_{t-1} + R\eta_t, & \eta_t &\sim \text{NID}(0, Q), & \text{with } Q &= \begin{bmatrix} \sigma_\xi^2 & 0 & 0 \\ 0 & \sigma_\zeta^2 & 0 \\ 0 & 0 & \sigma_\omega^2 \end{bmatrix}, \\
 \alpha_0 &\sim \text{N}(a_0, P_0),
 \end{aligned}$$

for $t = 1, \dots, n$. For $s = 4$, the matrices of this representation are defined as follows:

$$\begin{aligned}
 y_t &= \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \end{bmatrix} \alpha_t + \epsilon_t, \\
 \alpha_t \equiv \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \\ \gamma_{t-2} \\ \gamma_{t-3} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \xi_t \\ \zeta_t \\ \omega_t \end{bmatrix}.
 \end{aligned}$$

The initial state vector a_0 and its covariance matrix P_0 are assumed known or chosen beforehand.

Given the variance parameters, the Kalman filter and smoother can be applied to extract an estimate of the unobserved components. The traditional approach to obtain maximum likelihood parameter estimates is by maximization of the log-likelihood through a quasi-Newton algorithm. At each step of the optimization algorithm, the likelihood function is evaluated by means of the

Kalman filter. The derivation of the Kalman filter and the resulting iterative equations can be found, for instance, in (Harvey, 1989, Chapter 3), (Durbin and Koopman, 2001, Chapter 4). The computations carried out at each iteration can be used to evaluate the log-likelihood function of a linear Gaussian state space model, which is given by:

$$LL = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \log f_t - \frac{1}{2} \sum_{t=1}^n \frac{v_t^2}{f_t}, \quad (1)$$

where v_t is the prediction error and f_t its covariance matrix at time t .

3 Variations on the maximum likelihood procedure

In this section we introduce the issues and choices that lead to different variations on the maximum likelihood procedure in the context of structural time series models.

Kalman filter interfaces The Kalman filter is required to evaluate the time-domain log-likelihood function given in equation (1) as well as to extract an estimate of the unobserved components given the parameters of the model. As there are computational issues that differ between implementations, we will consider the choice of different interfaces available in R: `d1m` version 1.1-4 (Petris and Petrone, 2011; Petris, 2010; Petris et al., 2009).² `FKF` version 0.1.3 (Luethi et al., 2009); `KFAS` 1.0.4-1 (Helske, 2010); `KFKSDS` version 1.6; `sspir` version 0.2.10 (Dethlefsen et al., 2009) and `StructTS` in the `stats` package of R version 3.1.0 (R Core Team, 2012); All of these interfaces evaluate the same expression (up to a constant) of the log-likelihood function given in equation (1). Some of them differ in the approach to implement the filter. For example, `KFKSDS` is a direct transcription of the recursive equations of the Kalman filter. It was developed as a reference for the results shown in this paper and was useful to detect some relevant aspects in the design of the procedure. In addition, it provides analytical derivatives of the likelihood function that can be passed to the optimization algorithm. It is nevertheless a naive implementation subject to potential rounding errors that may render non-positive definite covariance matrices. Some packages circumvent this potential difficulty by means of a square root filter. For a review of computational issues involved in the Kalman filter see for instance Durbin and Koopman (2001, §6.3) and Tusell (2011). As far as we are concerned here, we observe that the practitioner should choose an implementation of the Kalman filter. Regardless of the potential effect on the final result, this issue involves a variation in the design of the maximum likelihood procedure.

²A reader familiar with older and active R packages may notice that the package `dse` (Gilbert, 2006; Gilbert and P.D., 1993) is missing. This package computes a large sample approximation of the log-likelihood function. Since this version is not common or comparable in the context of our application, this package is not considered here.

Initialization of the Kalman filter In real applications, the initial state vector a_0 is unknown and the filter needs to be somehow initialized. This topic has been discussed extensively in the literature. A common approach is to set a_0 equal to an arbitrary set of values along with a large value for its variance P_0 , which will account for the uncertainty on a_0 . Since the first observations act as a warming period, the filtered state vector may look somewhat erratic at the beginning of the sample. Another possibility is to include a_0 in the set of parameters to be estimated by maximum likelihood. The drawback of this approach is that the dimension of the vector of parameters in the objective function may become overly large. Another approach is to use the diffuse filter. It is a more sophisticated solution which requires elaborate modifications in the Kalman filter. For a detailed introduction see for instance [de Jong \(1991\)](#), [Koopman \(1997\)](#), [Durbin and Koopman \(2001, Chapter 5\)](#). Loosely speaking, a model for the initial state vector is defined as a random vector with an arbitrary large variance: $a_0 = \delta + \eta_0$, where δ is a diffuse vector, i.e., a random vector of the same length as a_0 and with distribution $\delta \sim N(0, \kappa I)$ with $\kappa \rightarrow \infty$; η_0 is a random variable distributed as $\eta_0 \sim N(0, Q)$. Under this framework, [Koopman \(1997\)](#) derives the expression of the diffuse likelihood function. The diffuse Kalman filter is available in package **KFAS**.

Covariance matrix of the initial state vector Another issue related to the initialization of the filter is the definition of the covariance matrix of the initial state vector, P_0 . In principle, it is assumed to be a diagonal matrix containing the same value for all the elements in the diagonal. The function `StructTS` in the `stats` package of R defines P_0 as a non-diagonal matrix containing the same value for all the elements (inside and outside the diagonal). In particular, all the elements are set equal to 10 000 times the variance of the series. We will consider both options (diagonal and non-diagonal matrices).

Parameterization of the model In order to ensure that the variance parameters returned by the optimization procedure are non-negative values, the model can be reparameterized. For example, the variance can be defined as $\sigma^2 = \theta^2$ or $\sigma^2 = \exp(\theta)$. Then, the likelihood is maximized for θ , which will lead to a non-negative σ^2 . As we shall see, another alternative is to use box constrained optimization. The implementation of `StructTS` uses a kind of regularization on the parameters: $\sigma^2 = \theta \times \kappa$, where κ is the variance of the data divided by 100. Although it does not transform the parameters to non-negative values, it may be a useful alternative, especially when the parameters are close to zero. We will consider variations on the parameterization of the model combined with constrained or unconstrained optimization.

Concentration of a parameter Another useful parameterization is in terms of relative variances. The interest is that one of the parameters can be concentrated out of the likelihood function. The variance parameters can be defined proportional to a positive scalar κ : $\sigma_\epsilon^2 = \kappa h$, $Q = \kappa Q^*$. Setting $h = 1$, the parameters in Q^* can be interpreted as variances relative to σ_ϵ^2 (alternatively, one of the elements in the diagonal of Q^* could be normalized to unity). It can be checked that the optimal value for κ in the model with parameters h and the elements in the diagonal of Q^* is given by $\hat{\kappa} = (1/n) \sum_{t=1}^n v_t^2 / f_t$. Thus, the dimension of the vector of parameters to be estimated by numerical methods is reduced in one parameter. The concentrated likelihood function is evaluated by the **stsm** package as follows:

$$LL^{(c)} = -\frac{n}{2} \log(2\pi + 1) - \frac{1}{2} \sum_{t=1}^n \log(f_t) - \frac{n}{2} \log(\hat{\kappa}).$$

Optimization algorithm The common approach to maximize the likelihood function is to use a quasi-Newton optimization algorithm. The BFGS algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) is widely used in this and other contexts. We will use this algorithm for those parameterizations that do not require restrictions on the parameters. For the other parameterizations, the L-BFGS-B algorithm (Byrd et al., 1995), which enforces box constraints (i.e., lower bounds equal to zero for the variance parameters), will be used. An adaptive barrier algorithm combined with the Nelder-Mead optimization algorithm (Nelder and Mead, 1965) (Rardin, 1998, §6.4, §13.8, §14.5) is also considered as a way to enforce non-negative lower bounds.

Numerical derivatives are employed except when the Kalman filter interface from package **KFKSDS** is used. This package evaluates the analytical derivatives and can be passed to the optimization algorithm.

Split the optimization procedure in three steps As mentioned above, the initialization of the filter requires choosing the values of the initial state vector, a_0 . The common approach is to set the first value in the vector a_0 equal to the value of the first observation and set the remaining elements equal to zero. Alternatively, the elements in a_0 can be taken as parameters to be estimated by maximum likelihood. In the latter case, we can consider splitting the procedure in three steps as follows (a similar idea is used in Brockwell and Davis (1996, §8.5)). In the first step a_0 is fixed as described before and the variance parameters are estimated by maximum likelihood. In the second step, the variances are fixed to their estimates and the vector a_0 is estimated. In the third step, the variance parameters are estimated again using the estimate of the vector a_0 obtained in the previous step. Although in a relatively long time series the effect of a_0 on the estimate of the variances may be small, the quality of the estimated components may improve, in particular at the

beginning of the sample. The standard approach is not to estimate a_0 . Two-steps or three-steps procedure involve new variations in the design of the procedure.

Restricted models Some of the variances may be found to be equal to zero. In that case, it is advisable to estimate the model imposing those restrictions. Two restricted models will be considered: $\sigma_\epsilon^2 = \sigma_\zeta^2 = 0$ and $\sigma_\zeta^2 = 0$. The easiest and most common way to enforce these restrictions is to fix the corresponding parameters to zero. Another possibility to enforce $\sigma_\zeta^2 = 0$ is to remove the drift term from the trend component and redefine the matrices of the state-space representation given in Section 2 accordingly. The model is exactly the same in both cases, what changes is the dimension of the matrix in the measurement equation. Both the restricted basic structural model and the unrestricted local level plus seasonal component model will be considered.

4 Application to the airline passengers data

The combination of all options discussed above leads to a large number of variations on the maximum likelihood procedure. The **bsm.airp** R package that is provided along with this paper contains some functions to aid in the process of generating and inspecting the results based on these variations. The functions available in the package can also be used to filter the results according to a set of options selected by the user and to cluster into groups of variations. For instance, given a Kalman filter interface, it is possible to compare the results across different parameterizations of the model. This section summarizes the main results that were obtained for the popular Box & Jenkins airline data (monthly totals of international airline passengers, 1949 to 1960). Following a common practice for this series, the analysis is conducted for the logarithms of the data. The Tables show the results for the one step procedure (i.e., the vector a_0 is fixed beforehand as indicated in Section 3). Variance parameter estimates are reported multiplied by 10 000.

Table 1 compares the results for different choices of the initial covariance matrix, P_0 . The table reports the number of variations where a particular solution was found, the parameter estimates and the corresponding value of the log-likelihood function. A non-diagonal covariance matrix for the initial state vector (labelled as $P_0 = 1$) leads to a larger variety of solutions. Although the values of the log-likelihood function at each solution are relatively close to each other, they have different implications. For example, with $P_0 = 0$ (P_0 is diagonal) the variance of the seasonal component σ_ω^2 is close to zero and, hence, it involves an almost deterministic pattern. With $P_0 = 1$, σ_ω^2 turns out to be a positive value and, hence, the seasonal pattern is stochastic.

[Table 1 about here.]

Table 2 reports parameter estimates for those variations that use unconstrained optimization. In order to ensure non-negative variance parameters, auxiliary parameters θ are estimated according to the following parameterizations: $\sigma^2 = \theta^2$ (labelled pt = 2) and $\sigma^2 = \exp(\theta)$ (labelled pt = 3). The results shown in Table 2 are sorted by the choice of the structure of P_0 (diagonal or non-diagonal). Given a value for P0, the choice of the parameterization leads to different results. When P_0 is diagonal, P0 = 0, there are minor differences across the choice of the parameterization, except for the fact that the seasonal component turns out to be fully deterministic when the exponential transformation of θ is used (labelled pt = 3). When P0 = 1 differences are more relevant. For example, σ_ω^2 is well above zero when the parameterization pt = 3 is used, while the relative weight of the seasonal component is lower with pt = 2). In addition, σ_ζ^2 takes on a positive value when the latter parameterization is used and it is zero for pt = 3). As we will explain below, the two sets of solutions that are found for P0 = 1 are distinguished by the choice of the Kalman filter interface.

[Table 2 about here.]

Table 3 shows the results obtained through constrained optimization. Instead of transforming the parameters as in the previous variations shown in Table 2, the L-BFGS-B optimization algorithm and an adaptive barrier alongside the Nelder-Mead optimization algorithm (AB-NM) are employed. The variance parameters are directly estimated, i.e. none reparameterization is considered. Lower bounds for the parameters are set equal to zero in the optimization algorithms. In this application, the AB-NM algorithm is more robust to changes in the design of the procedure and converges in all cases to a reliable solution. The L-BFGS-B algorithm failed to converge or led to an unsatisfactory solution (variances going to infinite or the same value for all parameters) in more than half of the cases. The kind of regularization described in Section 3 requires also a constrained optimization algorithm since the regularization does not ensure non-negative variances. Although not included in Table 3, it was observed that the regularization improves the convergence and reliability of the L-BFGS-B algorithm, since the number of non-plausible solutions was reduced. The interpretation of results for the AB-NM optimization algorithm remained similar to those reported in Table 3.

[Table 3 about here.]

Although the Kalman filter interfaces considered here implement the filter in different ways, in this application no major differences are found in the results based on different implementations of the filter. The main observation that is worth mentioning here is that, compared to the other packages, the KFAS and FKF interfaces led to a different set of solutions when P_0 is non-diagonal. The arguments passed to these interfaces accept a non-diagonal covariance matrix. However, in

a preliminary tests of the packages it was observed that the value returned for the log-likelihood differs slightly from the other interfaces when a non-diagonal covariance matrix P_0 is used. The difference increases as the variance parameter goes to zero. In the application, the results based on these interfaces often led to a solution where σ_ζ^2 is different from zero. When the parameterization $\text{tp} = 3, \sigma^2 = \exp(\theta)$, is used, the solution is close to that found with the other implementations of the filter (even when P_0 is non-diagonal).

Table 4 shows the results obtained when the parameter σ_ϵ^2 was concentrated out of the likelihood function. This procedure was robust across different variations. Given a value for P_0 , all the variations led essentially to the same result regardless of the parameterization and the optimization algorithm. Contrary to the results discussed before for the standard likelihood function, here, σ_ζ^2 is estimated to be zero regardless of the structure of P_0 . As some of the parameters were estimated to be close to zero, two restricted models were fit. In the first model (labelled r1) the restrictions $\sigma_\epsilon^2 = \sigma_\zeta^2 = 0$ are imposed and in the second model (labelled r2) $\sigma_\zeta^2 = 0$. Given a structure of P_0 (diagonal or not), the results are close to each other regardless of the restrictions imposed to the model, except for the fact that σ_ϵ^2 takes a more prominent role relative to the other parameters in the model restricted with $\sigma_\zeta^2 = 0$. Although not reported, the restricted models were also fitted using the non-concentrated likelihood function. In that case, the results were not robust across variations and varied in a similar way as described above for the unrestricted model. In addition, in those variations where a non-plausible solution was obtained for the unrestricted model, the restricted model did not improve the quality of results. In those cases where the unrestricted model yielded estimates for σ_ζ^2 and/or σ_ϵ^2 close to zero, the estimates for the other parameters were similar in the unrestricted and restricted models.

[Table 4 about here.]

As an alternative to the restriction $\sigma_\zeta^2 = 0$, we fit the model redefining the matrices of the state space representation as a local level plus a seasonal component (LLS). The results for this set of variations summarized in Table 5. We can see that a new set of solutions characterized by a smaller relative value of σ_ω^2 is found. As before, two solutions are obtained depending on the structure of P_0 . Given a structure of P_0 , the results are robust across variations. Parameter estimates remain almost identical for different choices of the parameterization, the optimization algorithm and the Kalman filter interface. The results for the LLS model restricted with $\sigma_\epsilon^2 = 0$ and those based on the concentrated likelihood function for the same model remained very similar to those reported in Table 5.

[Table 5 about here.]

So far we have discussed the results based on the one-step procedure. In the three-steps procedure the vector a_0 is estimated instead of being fixed. The patterns observed in the three-steps procedure were in line with those summarized above. In general, the solutions that were found in the one-step procedure remained very similar in the three-steps procedure, both in the unconstrained and constrained models. A smoother pattern for the estimated components may be expected when a polished initial vector a_0 is used, especially at the beginning of the sample. However, in this application, changes in the pattern of the estimated components were not noticeable graphically, possibly because the estimate of the vector a_0 turned out to be close to the fixed vector chosen in the one step procedure.

Table 6 assesses the quality of some of the solutions. A diagnostic is done based on the Akaike information criterion (AIC), the sample autocorrelations to detect the presence of seasonality in the estimated trend and in the standardized residuals and a range-mean plot to assess the homoscedasticity of the estimated seasonal component. The trend and seasonal components are estimated by means of the Kalman smoother, which gives the best estimate of the state at each period conditioned on the whole observed series. The solutions are shown in the order in which they have been shown in the Tables above. Although the values of the AIC are close to each other, the quality of results is better for some of the solutions. When seasonality is found in the estimated trend component, it is also found in the standardized residuals. When seasonality is detected in the trend, the estimated seasonal component is not satisfactory either, since a heteroscedastic is observed in the seasonal component. A deterministic seasonal pattern (fourth row, $\sigma_\omega^2 = 0$) is not appropriate according to these diagnostics. The solutions with a lower ratio between the variance parameters in the seasonal and trend components show a better performance in terms of AIC and the diagnostic of the estimated components.

A common feature of the variations in which the solutions in Table 6 are based is that those solutions that do not pass the diagnostic tests use a non-diagonal covariance matrix for the initial state vector, $P_0 = 1$ (except for the fourth solution where $\sigma_\omega^2 = 0$). The same diagnostic tests were carried out for these solutions setting $P_0 = 0$, i.e., off-diagonal elements are set equal to zero. The estimated components turned then acceptable, i.e., seasonal fluctuations did not leak into the estimated trend or residuals and the estimated seasonal component was homoscedastic. However, the AIC revealed that those solutions were inferior to the other estimates in terms of the AIC: the AIC increased from -317.418 to -251.572 and from -280.142 to -208.369 respectively for the second and third solutions reported in Table 6 and from -306.190 to -256.499 in the ninth solution which was also obtained with $P_0 = 0$. Thus, in this application the option $P_0 = 1$ (non-

diagonal covariance matrix for the initial state vector) does not seem appropriate. However, the documentation of the function `StructTS` which originally uses this option claims that the solution found by this function for these data is superior to the solution reported in other source. For other time series not reported in this paper, it was observed that the optimization algorithm failed to convergence when P_0 was defined as a diagonal matrix, while setting $P_0 = 1$ the algorithm converged to a feasible solution. Therefore, this variation is an option that can be helpful in some cases but, unfortunately, as it happens with the other choices, we don't know beforehand which one is more appropriate for the data at hand.

[Table 6 about here.]

Figure 1 displays the estimated trend and seasonal components based on the second and eighth solutions reported in Table 6. These solutions involve a different allocation of the dynamics of the data among the components. A lower relative value of the variance in the seasonal component (right hand side plot) yields a compelling seasonal pattern. The trend is dominated by a smooth long-term pattern and, hence, the seasonal fluctuations detected in the solution of the left hand side plot cannot be graphically observed.

[Figure 1 about here.]

5 Conclusion

We have seen several variations on the common maximum likelihood procedure to fit a structural time series model. The application to a popular time series shows that the choice of particular issues that are often overlooked in the description of applications may have an effect on the results, making it difficult to replicate the results obtained by others. Moreover, the design of the procedure may also affect the quality of results.

Unfortunately, it is not possible to know beforehand which combination of options will be more suitable for a particular application. Thus, flexible interfaces to fit these models are desirable so that the user does not need to implement the procedure by herself. Comprehensive software documentation is also necessary to compare designs implemented in different statistical packages.

Despite we cannot anticipate the performance of each variation in a given application, the results obtained in this paper suggest the following recommendations. 1) The concentration of a parameter –a technique often consigned to textbooks and some specialized software– makes the procedure more robust and more likely to converge to a reliable solution. 2) The optimization

algorithm performs better with the unrestricted model. It is a common practice to choose a broad model, e.g. the basic structural model, and restrict the model by fixing some of the parameters to zero. The application shown in this paper suggests that it is more appropriate to redefine the matrices of the state space representation of the model according to the desired restrictions and fit the equivalent unrestricted model. 3) An adaptive barrier algorithm along with the Nelder-Mead optimization procedure is a useful alternative to the L-BFGS-B algorithm when the latter fails to converge to a reliable result.

References

- Box, G. and Jenkins, G. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.
- Brockwell, P. J. and Davis, R. A. (1996), *Introduction to Time Series and Forecasting*, Springer Texts in Statistics, Springer-Verlag.
- Broyden, C. G. (1970), ‘The convergence of a class of double rank minimization algorithms’, *Journal of the Institute of Mathematics and its Applications* **6**, 76–90.
- Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995), ‘A limited memory algorithm for bound constrained optimization’, *SIAM Journal on Scientific Computing* **16**(5), 1190–1208.
- Commandeur, J. J. F., Koopman, S. J. and Ooms, M. (2011), ‘Statistical software for state space methods’, *Journal of Statistical Software* **41**(1), 1–18.
URL: <http://www.jstatsoft.org/v41/i01/>
- de Jong, P. (1991), ‘The diffuse Kalman filter’, *The Annals of Statistics* **19**(2), 1073–1083.
- Dethlefsen, C., Lundbye-Christensen, S. and Christensen, A. L. (2009), *sspir: State Space Models in R*. R package version 0.2.8.
URL: <http://CRAN.R-project.org/src/contrib/Archive/sspir/>
- Durbin, J. and Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford Statistical Science Series, Oxford University Press.
- Fletcher, R. (1970), ‘A new approach to variable metric algorithms’, *The Computer Journal* **13**, 317–322.
- Gilbert, P. D. (2006), *Brief User’s Guide: Dynamic Systems Estimation*.

- Gilbert and P.D. (1993), State space and arma models: An overview of the equivalence, Working Paper 93-4, Bank of Canada, Ottawa, Canada. ISSN 1192-5434. ISBN 0-662-20494-8.
- Goldfarb, D. (1970), ‘A family of variable metric methods derived by variational means’, *Mathematics of Computation* **24**, 23–26.
- Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.
- Helske, J. (2010), *KFAS: Kalman filter and smoothers for exponential family state space models*. R package version 0.6.0.
URL: <http://CRAN.R-project.org/package=KFAS/>
- Koopman, S. J. (1997), ‘Exact initial Kalman filtering and smoothing for nonstationary time series models’, *JASA. Journal of the American Statistical Association* **92**(440), 1630–1638.
- Koopman, S. J., Harvey, A. C., Doornik, J. A. and Shephard, N. (2009), *STAMP 8.2: Structural Time Series Analyser, Modeler, and Predictor*, Timberlake Consultants Ltd, London. ISBN 978-0-9557076-2-9.
- Luethi, D., Erb, P. and Otziger, S. (2009), *FKF: Fast Kalman Filter*. R package version 0.1.0.
URL: <http://CRAN.R-project.org/package=FKF/>
- Nelder, J. A. and Mead, R. (1965), ‘A simplex algorithm for function minimization’, *Computer Journal* **7**, 308–313.
- Petris, G. (2010), *dlm: Bayesian and Likelihood Analysis of Dynamic Linear Models*. R package version 1.1-1.
URL: <http://CRAN.R-project.org/package=dlm/>
- Petris, G. and Petrone, S. (2011), ‘State space models in R’, *Journal of Statistical Software* **41**(4), 1–25.
URL: <http://www.jstatsoft.org/v41/i04/>
- Petris, G., Petrone, S. and Campagnoli, P. (2009), *Dynamic Linear Models with R*, Use R, Springer.
- R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: <http://www.R-project.org/>
- Rardin, R. L. (1998), *Optimization in Operations Research*, Prentice Hall.

Shanno, D. F. (1970), ‘Conditioning of quasi-Newton methods for function minimisation’, *Mathematics of Computation* **24**, 647–657.

Tusell, F. (2011), ‘Kalman filtering in R’, *Journal of Statistical Software* **39**(2), 1–27.

URL: <http://www.jstatsoft.org/v39/i02/>

Figure 1: Estimated components for the logarithms of the airline data. The estimated trend and seasonal components are based on the second and eighth solutions reported in Table 6, respectively for the left and right hand side plots. The line in gray is the logarithm of the original data.

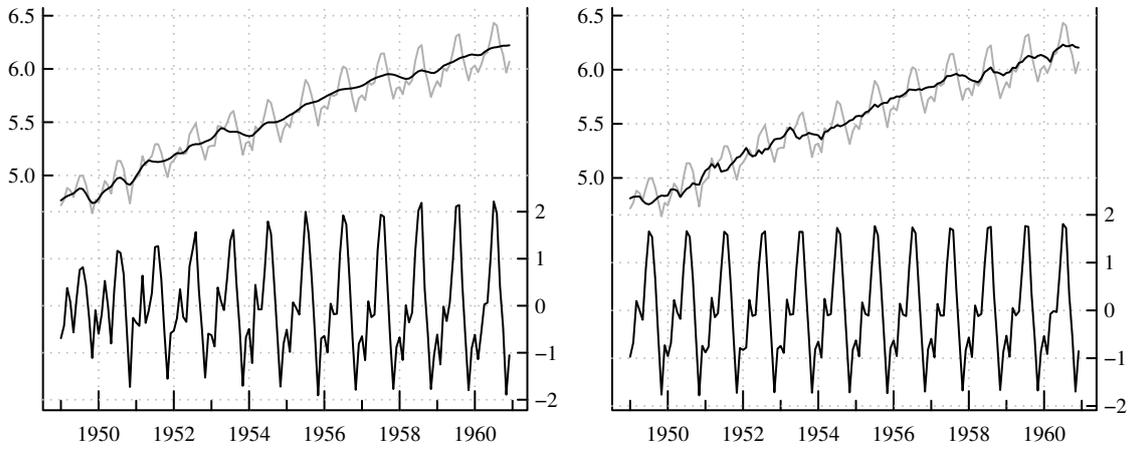


Table 1: Initial covariance matrix of the state vector (P0)

P0	#	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_\zeta^2$	$\hat{\sigma}_\omega^2$	LL
0	41	1.147	7.070	0.000	0.687	168.175
0	7	3.676	7.668	0.000	0.000	163.223
0	8	Non-plausible solution				
1	17	0.000	7.718	0.000	13.969	162.709
1	1	0.000	12.476	0.000	14.142	161.021
1	6	0.118	11.638	0.006	13.658	160.620
1	1	0.000	6.603	0.000	24.596	158.009
1	3	0.000	0.000	22.679	3.521	146.154
1	8	0.000	6.640	19.455	3.317	144.067
1	4	Non-plausible solution				

Cells report parameter estimates (multiplied by 10 000) that were obtained when the initial covariance matrix of the state vector is diagonal (P0 = 0) and when it contains non-zero values outside the diagonal (P0 = 1).

Non-plausible solutions are those where the optimization algorithm did not converge or returned an unreliable solution (e.g. estimates go to infinity or all the variance parameters take the same value).

Results are based on a one step procedure with no restrictions on the parameters. The number of variations that led to a particular solution is shown in the column labelled as #. Each variation is characterized by the choice of the Kalman filter interface, the parameterization of the model and the optimization algorithm. LL is the value of the log-likelihood function at a given solution.

Table 2: Unconstrained optimization (BFGS algorithm)

pt	P0	#	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_\zeta^2$	$\hat{\sigma}_\omega^2$	LL
2	0	7	1.274	7.000	0.000	0.647	168.183
2	1	3	0.000	0.000	22.679	3.521	146.154
2	1	2	0.000	6.631	19.452	3.317	144.071
3	0	7	3.676	7.668	0.000	0.000	163.223
3	1	2	0.000	7.718	0.000	13.969	162.708
3	1	2	0.118	11.638	0.006	13.658	160.620
3	1	1	Non-plausible solution				

Cells report parameter estimates for those variations that use unconstrained optimization (BFGS algorithm). In order to ensure non-negative variance parameters, the model is parameterized in terms of an auxiliary vector of parameters θ as follows: $\sigma^2 = \theta^2$ (labelled pt = 2) and $\sigma^2 = \exp(\theta)$ (labelled pt = 3). Results are split by the choice of P0. See footnote to Table 1.

Table 3: Constrained optimization

Optim. alg.	P0	#	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_\zeta^2$	$\hat{\sigma}_\omega^2$	LL
L-BFGS-B	0	7	Non-plausible solution				
L-BFGS-B	1	1	0.802	6.630	0.000	14.148	162.033
L-BFGS-B	1	1	0.000	12.476	0.000	14.142	161.021
L-BFGS-B	1	1	0.000	6.603	0.000	24.596	158.009
L-BFGS-B	1	2	Non-plausible solution				
AB-NM	0	7	1.147	7.070	0.000	0.687	168.175
AB-NM	1	3	0.000	7.941	0.000	13.840	162.704
AB-NM	1	2	0.121	11.636	0.006	13.659	160.624

Cells report parameter estimates for those variations that use constrained optimization either via the L-BFGS-B algorithm or an adaptive barrier algorithm along with the Nelder-Mead algorithm (AB-NM). Optim. alg.: optimization algorithm. See second and third footnotes to Table 1.

Table 4: Results based on the concentrated likelihood function

	P0	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_\zeta^2$	$\hat{\sigma}_\omega^2$	LL
r0	0	0.071	7.848	0.000	0.953	299.502
	1	0.000	7.692	0.000	13.996	231.955
r1	0	—	5.437	—	1.572	276.509
	1	—	7.384	—	14.159	227.670
r2	0	2.593	3.935	—	0.863	278.405
	1	0.000	7.384	—	14.159	227.670

Cells report parameter estimates for those variations that use maximize the the concentrated likelihood function. Unrestricted (labelled r0) and restricted models are considered (r1 enforces the restrictions $\sigma_\epsilon^2 = \sigma_\zeta^2 = 0$ and in r2 $\sigma_\zeta^2 = 0$).

Table 5: Local level plus seasonal model

P0	#	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_\omega^2$	LL
0	47	0.282	10.280	0.537	170.765
0	9	Non-plausible solution			
1	36	0.000	13.101	11.675	156.095
1	4	Non-plausible solution			

Cells report parameter estimates for the local level plus seasonal model. It is equivalent to the basic structural model with $\sigma_\zeta^2 = 0$. Here, instead of estimating a restricted BSM model, an unrestricted LLS is estimated. See second and third footnotes to Table 1.

Table 6: Diagnostic of a set of solutions

$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_\zeta^2$	$\hat{\sigma}_\omega^2$	r^a	AIC ^b	D1 ^c	D2 ^d	D3 ^e
1.295	6.994	0.000	0.642	0.092	-328.366	No	No	Yes
0.000	7.719	0.000	13.969	1.810	-317.418	Yes	Yes	No
0.000	6.631	19.452	3.317	0.500	-280.142	Yes	Yes	No
3.677	7.664	0.000	0.000	0.000	-318.455	Yes	Yes	Yes
0.071	7.848	0.000	0.953	0.121	-327.425	No	No	Yes
0.000	5.437	0.000	1.572	0.289	-324.737	No	No	Yes
2.593	3.935	0.000	0.863	0.219	-324.392	No	No	Yes
0.282	10.280	—	0.537	0.052	-335.530	No	No	Yes
0.000	13.101	—	11.674	0.891	-306.190	Yes	Yes	No

^a Ratio $\hat{\sigma}_\omega^2/\hat{\sigma}_\xi^2$.

^b Akaike information criterion.

^c Is seasonality present in the first differences of the estimated trend? Based on the autocorrelation function.

^d Is seasonality present in the standardized residuals? Assessed as D1.

^e Is the estimated seasonal component homoscedastic? Based on a range-mean plot and graphical view of the estimated component.