ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

Forecast comparison of SARIMA and Kalman-filter DLM models for seasonal data

By

Dimitrios Georgiou Rousis

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfillment of the requirements for

the degree of Master of Science in Statistics

Athens, Greece

July 2020


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

Σύγκριση προβλεπτικής ικανότητας SARIMA και Kalman-filter DLM μοντέλων σε εποχικά δεδομένα

Δημήτριος Γεωργίου Ρουσής

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής

του Οικονομικού Πανεπιστημίου Αθηνών

ως μέρος των απαιτήσεων για την απόκτηση

Διπλώματος Μεταπτυχιακών Σπουδών στη Στατιστική

Αθήνα

Ιούλιος 2020


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

DEDICATION

To my mother.


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Professor E. Ioannidis for his patient guidance, enthusiastic encouragement, useful critiques, and his valuable help in learning to overcome any difficulties that arose along the way until the completion of this master thesis. I would also like to thank Professor P.Besbeas, for the time he devoted to rekindle my interest in seeing the science of statistics from a different perspective. My grateful thanks are also extended to all the staff of the postgraduate secretariat that helped me in a timely and valid manner in every request that I submitted to them throughout my studies.

Finally, I wish to thank my family and my friends for their support and encouragement throughout my study.


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

VITA

I was born on October 5th, 1994 in Athens. In 2012 graduated from Evangeliki Scholi high school and I entered the Department of Applied Mathematics and Physical Sciences at the National and Technical University of Athens. In 2018, I got my bachelor degree and I enrolled the Master of Science in Statistics at Athens University of Economics and Business.


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ABSTRACT

In time series analysis, it is very common to encounter cases of data that show seasonality or some form of a trend. The general methodology developed by Box and Jenkins (1970) gave us the class of ARIMA models and the special subcategory of SARIMA models. This is still a very powerful tool for describing time series with such features and predicting future observations.

Another large class of models that are gaining ground in recent years is the so-called state space models. A very broad subcategory of state space models is the so-called Dynamic Linear Models (DLM's) which, despite their general simplicity, can be used to capture features that govern real-time time series such as trend and seasonality. The process of fitting the models and predicting future observations is done through the methodology known as the Kalman filter which was first introduced by Kalman in 1960.

This thesis analyzes all the key features of the ARIMA and SARIMA models and the DLM model class. Then, their predictive ability is tested based on two different evaluation methods which are the calculation of the usual error measures, namely MSE, MAD and MAPE, and the cross-validation methodology specifically adapted for time series data.

The application of the methods and the comparison of their predictive capacity of the were done in two different real data sets: the first concerns monthly birth numbers in New York City and the second is related to monthly retail automobile sales in the USA.


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΠΕΡΙΛΗΨΗ

Στην ανάλυση χρονοσειρών είναι πολύ σύνηθες να ερχόμαστε αντιμέτωποι με περιπτώσεις δεδομένων που παρουσιάζουν εποχικότητα ή και κάποιας μορφή τάση. Η γενική μεθοδολογία που αναπτύχθηκε από τους Box και Jenkins (1970) μας έδωσε την κλάση των ARIMA μοντέλων και της ειδικής υποκατηγορίας SARIMA μοντέλων. Αυτό μέχρι και σήμερα αποτελεί ένα πολύ ισχυρό εργαλείο για την περιγραφή χρονοσειρών με τέτοια χαρακτηριστικά και πρόβλεψη μελλοντικών παρατηρήσεων.

Μια άλλη μεγάλη γκάμα μοντέλων που κερδίζουν μεγάλο έδαφος τα τελευταία χρόνια είναι τα λεγόμενα state space μοντέλα. Μια πολύ ευρεία υποκατηγορία state space μοντέλων είναι τα λεγόμενα Dynamic Linear Models (DLM's) τα οποία παρά την γενικότερη απλότητα τους μπορούν να χρησιμοποιηθούν για την αποτύπωση χαρακτηριστικών που διέπουν χρονοσειρές αληθινών δεδομένων όπως είναι η τάση και η εποχικότητα. Η διαδικασία της προσαρμογής των μοντέλων και πρόβλεψης μελλοντικών παρατηρήσεων γίνεται μέσω της μεθοδολογίας που είναι γνωστή ως Kalman φίλτρο η οποία πρωτοπαρουσιάστηκε από τον Kalman το 1960.

Σε αυτήν την εργασία αναλύονται όλα τα βασικά χαρακτηριστικά της συνομοταξίας των ARIMA και SARIMA μοντέλων και της κλάσης των DLM μοντέλων. Στη συνέχεια διενεργείται έλεγχος της προβλεπτικής ικανότητας τους βασιζόμενοι σε δύο διαφορετικές μεθόδους αξιολόγησης οι οποίες είναι ο υπολογισμός των συνήθη μέτρων για σφάλματα, δηλαδή τα MSE, MAD και MAPE, και στη μεθοδολογία cross-validation ειδικά προσαρμοσμένη για δεδομένα χρονοσειρών.

Η εφαρμογή των μεθόδων και η σύγκριση της προβλεπτικής ικανότητας των μοντέλων έγινε σε δύο διαφορετικά πραγματικά σύνολα δεδομένων: το πρώτο αφορά μηνιαίους αριθμούς γεννήσεων στην πόλη της Νέας Υόρκης και το δεύτερο σχετίζεται με μηνιαίες πωλήσεις αυτοκινήτων λιανικού εμπορίου στις ΗΠΑ.


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Index

LIST OF TABLES

LIST OF FIGURES

Introduction

The element of seasonality is found in almost all areas of human activity. From financial measures such as the annual forecast of GDP or inflation that determines the policies of an entire country to meteorological data and the forecast of extreme weather events, the need for processing and modeling of chronological data with seasonality is becoming more and more imperative. However, a common element and key issue in any such study is the ability of predicting future data.

In 1970, Box and Jenkins introduced a new class of multiplicative linear autoregressive integrated moving average models, ARIMA(p,d,q), which are defined as:

$$\varphi(B)(1-B)^d x_t = \theta(B) w_t \, , \, w_t \sim WN(0,\sigma_w^2)$$

where $\varphi(z)$, $\theta(z)$ are the autoregressive and moving average polynomials and p,q are respectively the orders of these polynomials and d is the order of differencing.

This kind of models were considered path breaking. The seasonal ARIMA model, called SARIMA, is a special case of this model and is applied to seasonal data, for example weekly or monthly which is the case for this thesis. The general SARIMA(p,d,q)x$(P,D,Q)_s$ model is defined as:

$$\varphi(B)\Phi(B^s)x_t = \theta(B)\Theta(B^s)w_t, \; w_t \sim WN(0,\sigma_w^2)$$

where $\varphi(z)$, $\theta(z)$, $\Phi(z)$ and $\Theta(z)$ are the seasonal and non-seasonal AR and MA polynomials and P and Q are the orders of the non-seasonal polynomials, S is the period of the seasonal pattern, d is the order of differencing and D is the order of seasonal differencing.

The process of fitting a SARIMA (p, d, q) x (P, D, Q) model begins with the identification of the parameters d, D and then p, q, P, Q, which is done by using both graphic as well as analytical tools and tests. The usual procedure used by R is the use of the Hyndman-Khandakar (HK) algorithm developed by Hyndman and Khandakar (2008) which minimizes one of the well known information criteria for selecting parameters (AIC, BIC, AICc). However, a basic criticism of this algorithm of selecting the optimal model is that it poses very strict restrictions on the order of the parameters and in fact many models with higher orders of the AR and MA polynomials are excluded by definition from the investigation.

The Kalman filter was developed by Kalman in 1960 and is applied to models written on the so called state-space form. It was first proven as useful for applications in engineering. In 1977 Morrison and Pike stated that a general method for applying the Kalman filter for statistical forecasting has not yet been developed. However the procedure had at that time started to get its roots into the statistical field, for example Rosenberg (1973), Engle (1979) and Harvey and Phillips

(1979) found that the Kalman filter could be applied to econometric and statistical problems, for example for time series forecasting of economic properties. These were fields that ARIMA modeling at that point dominated (Harvey, 1989). One example of a clear benefit of the Kalman filter is that only the present state estimate and the next observation is required to update the whole system. This can be compared to models estimated with maximum likelihood or ordinary least squares where updating would imply use of the whole history of data.

One very important category of models which can be written in state-space form are the so called Dynamic Linear Models (DLM) which are specified by means of two equations

$$Y_t = F_t\theta_t + v_t,\ v_t \sim N_m(0, V_t)$$
$$\theta_t = G_t\theta_{t-1} + w_t,\ w_t \sim N_p(0, W_t)$$

where Gt and Ft are known matrices and the (vt) and (wt) are two independent white noise sequences, with mean zero and known covariance matrices Vt and Wt respectively. DLM's are very adaptable and can be used for time series which appear some kind of trend, or seasonality or both.

The purpose of this thesis is to compare the forecasting performance between the model specified by the HK algorithm, the model obtained by the minimization of the AIC criterion without posing the strict restrictions of the HK algorithm and a suitable DLM which captures the characteristics of the data in hand. The questions to be answered would be 1)"*Does the standard HK algorithm lead to a model adequate enough for forecasting future values or we can conclude to a more suitable model by applying less strict restrictions?"* and 2) *"Can the simple DLM's assimilate the time series characteristics and give forecasts equally good as the standard ARIMA class of models?"*

The comparisons will be made through two different methods a) use of the ordinary error measures (MSE, MAD and MAPE) along with the Diebold-Mariano test in a test set of observations and b) the extraction of forecast errors from a cross-validation methodology designed specifically for time series.

This thesis deals with two different real life datasets: 1) monthly rates of births in New York City and 2) monthly retail automobile sales in USA. Both of them appear to have a strong seasonal pattern with an underlying trend, the presence of which is obvious in the second one and more questionable in the first. The reason behind the choice of those so apparently different examples is to examine the comparison of the various models in a great variety of forecast examples, so as to obtain the big picture of how accurate they can be for future use in datasets which appear similar characteristics.

The structure of this thesis is organized as follows. Chapter 1 deals with some basic concepts and definitions which lead to the building of the general SARIMA model which is defined and discussed in Chapter 2. Also in Chapter 2 can be found all the various tools that are used for the estimation,

model fitting and forecast evaluation. Chapter 3 deals with state space models and particularly with the most commonly used forms of Dynamic Linear models and the Kalman filer methodology. Chapter 4 contains the application of all the methods discussed in the previous chapters and is divided in two parts; subsection 4.1 deals with the monthly rates of births dataset and subsection 4.2 with the monthly retail automobile sales dataset. Finally, chapter 5 concludes the thesis by summarizing the results and by discussing some details.

1. Basic concepts and definitions

In the present section we give the essential definitions and attributes of the models that we will go over in the following chapters. Besides, we additionally express the fundamental outcomes and hypotheses which are at the center of our investigation and a necessary piece of each ensuing area. In the present case we simply take into account discrete procedures, where time $t \in \mathbf{Z}$ and also the time horizon $h \in \mathbf{Z}$

1.1 Stationarity and ACF, PACF

We start with certain definitions so as to review the fundamental terms and furthermore concurring on documentation. We start with the concept of stationarity and autocovariance.

Definition 1.1.1 The mean function of a stochastic process x_t is

$$\mu_t = E(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx \qquad 1$$

provided it exists, where E denotes the usual expected value operator and $f_t(x)$ denotes process distribution density function.

Definition 1.1.2 The autocovariance function of a finite variance process x_t with mean value function μ_t is defined as

$$\gamma(s,t) = E\left[(x_s - \mu_s)(x_t - \mu_t)\right] \qquad 2$$

for all s and t

Definition 1.1.3 A weekly stationary time series, $\{x_t\}$, is a finite variance process such that

a) the mean value function, μ_t, is constant and does not depend on time t, and

b) the covariance function γ(s,t),depends on s and t only through their difference $|s-t|$.

From this time forward, when we utilize the term stationarity the term weak stationarity will be meant. The autocovariance function of a stationary time series will be written as a function of the lag h as:

$$\gamma(h)=E\left[\left(x_{t+h}-\mu\right)\left(x_t-\mu\right)\right],\ h\geq 1 \qquad 3$$

If we are interested in finding to what extent there is a numerical relationship between two variables of interest, using their correlation coefficient will give misleading results if there is another, confounding, variable that is numerically related to both variables of interest. This misleading information can be avoided by controlling for the confounding variable, which is done by computing the partial correlation coefficient

Definition 1.1.4 a) The partial correlation between X and Y given a set of n controlling variables $Z=\{Z_1,Z_2,...,Z_n\}$ is the correlation between the residuals e_X and e_Y resulting from the linear regression of X with Z and of Y with Z, respectively.

b)The partial autocorrelation function (PACF) gives the partial correlation of a stationary time series with its own lagged values, regressed the values of the time series at all shorter lags. It contrasts with the autocorrelation function, which does not control for other lags.Given a time series x_t the partial autocorrelation between x_t and x_{t-h} is defined as the conditional correlation between x_t and x_{t-h}, condtional on $x_{t-h+1},...,x_{t-1}$ the set of observations that come between the time points t and t-h.

- The 1st order partial autocorrelation will be defined to equal the 1st order autocorrelation.
- The 2nd order (lag) partial autocorrelation is $\dfrac{\mathrm{Cov}\left(x_t,x_{t-2}|x_{t-1}\right)}{\sqrt{\mathrm{Var}\left(x_t|x_{t-1}\right)\mathrm{Var}\left(x_{t-2}|x_{t-1}\right)}}$

This is the correlation between values two time periods apart conditional on knowledge of the value in between (the two variances in the denominator will equal each other in a stationary series.)

- The 3rd order (lag) partial autocorrelation is $\dfrac{\mathrm{Cov}\left(x_t,x_{t-3}|x_{t-1},x_{t-2}\right)}{\sqrt{\mathrm{Var}\left(x_t|x_{t-1},x_{t-2}\right)\mathrm{Var}\left(x_{t-3}|x_{t-1},x_{t-2}\right)}}$

And, so on, for any lag

Definition 1.1.5 If x_t is a stationary process with autocovariance function $\gamma(.)$ then its autocovariance generating function is defined by

$$G(z)=\sum_{k=-\infty}^{\infty}\gamma(k)z^k, z\in \square \quad (4)$$

provided the series converges for all z in some annulus $r^{-1}<|z|<r$ with $r>1$

At the point when we have characterized the autocovariance function, we can give the definition of the white noise process, which has a significant role in the examination of the considerable number of models we are going to utilize and furthermore in the inference of the prediction equations.

Definition 1.1.6 The process w_t (with mean 0 and variance σ^2) is said to be white noise process, and it will be symbolized as $w_t \sim WN(0,\sigma_w^2)$, if and only if w_t has zero mean and covariance function

$$\gamma(h)=\begin{cases}0, h\neq 0\\ \sigma^2, h=0\end{cases} \quad (5)$$

1.2 The ARMA models

A very large and widely used class of time series models is the ARMA models. In order to construct the ARMA models, first we have to give some definitions.

Definition 1.2.1 We say that stationary process x_t is an autoregressive process of order p, abbreviated AR(p), if it satisfies the equation

$$x_t=\varphi_1 x_{t-1}+...+\varphi_p x_{t-p}+w_t \quad (6)$$

where $\varphi_1,\varphi_2,...,\varphi_p$ are constants ($\varphi_p\neq 0$) and w_t is a white noise process with variance σ_w^2

Definition 1.2.2 The process x_t is said to be a moving average process of order q, or MA(q), if it satisfies the equation

$$x_t=w_t+\theta_1 w_{t-1}+...+\theta_q w_{t-q} \quad (7)$$

where $\theta_1,\theta_2,...,\theta_q$ are constants $(\theta_q \neq 0)$ and $w_t \sim WN(0,\sigma_w^2)$

A compact way of defining ARMA process is by defining the back-swift operator and introducing the autoregressive and moving average equations and operators.

Definition 1.2.3 We define the back-swift operator by

$$Bx_t = x_{t-1}, t \in \mathbf{Z} \qquad 8$$

and extend it to powers $B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$ and so on. Thus, $B^k x_t = x_{t-k}, t \in \mathbf{Z}$.

Definition 1.2.4A stationary process $\{x_t : t \in \mathbf{Z}\}$ is said to be an ARMA(p,q) process if $x_t = \varphi_1 x_{t-1} + ... + \varphi_p x_{t-p} + \theta_1 w_{t-1} + ... + \theta_q w_{t-q} + w_t$ or equivalently $\varphi(B)x_t = \theta(B)w_t$, where $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - ... - \varphi_p B^p$, $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 - ... - \theta_q B^q$ and $w_t \sim WN(0,\sigma_w^2)$, $\varphi_p, \theta_q \neq 0$ and the parameters p and q are called autoregressive and moving average orders respectively.

Definition 1.2.5 An ARMA(p,q) model $\varphi(B)x_t = \theta(B)w_t$ is said to causal if the time series $\{x_t : t \in \mathbf{Z}\}$ can be written as a one-sided linear filter of the white noise process $\{w_t : t \in \mathbf{Z}\}$:

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B)w_t \qquad 9$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and we set $\psi_0 = 1$

Definition 1.2.6 An ARMA(p,q) model $\varphi(B)x_t = \theta(B)w_t$ is said to be invertible if the innovations $\{w_t : t \in \mathbf{Z}\}$ can be written as a one-sided linear filter of the process $\{x_t : t \in \mathbf{Z}\}$:

$$w_t = \pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} \qquad 10$$

where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$ and $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and we set $\pi_0 = 1$

At the point when we have characterized the ideas of causality and invertibility, we now give criteria for an ARMA model to be causal and invertible.

Lemma 1.2.1 a) Let $\{x_t\}$ be a stationary ARMA(p,q) process for which the polynomials $\varphi(.)$, $\theta(.)$ have no common zeros. Then $\{x_t\}$ is causal if and only if $\varphi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$

b) Let $\{x_t\}$ be an ARMA(p,q) process for which the polynomials φ(.), θ(.) have no common zeros. Then $\{x_t\}$ is inverible if and only if $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$

From Brockwell and Davis (1991) we can locate a valuable tool called autocovariance generating function.

Lemma 1.2.2 The autocovariance generating function for a an ARMA(p,q) process $\varphi(B)x_t=\theta(B)w_t$ for which $\varphi(z) \neq 0$ when $|z|=1$ is of the following form

$$G(z)=\sigma^2 \frac{\theta(z)\theta(z^{-1})}{\varphi(z)\varphi(z^{-1})}, r^{-1}<|z|<r \text{ for some } r>1 \qquad 11$$

1.3 Integrated series and ARIMA models

A key point in our analysis is the definition of the unit root which can be found below.

Definition 1.3.1A time series x_t is integrated of order $d \neq 0$, abbreviated I(d), if

$$\nabla^d x_t := (1-B)^d x_t$$

is stationary but $\nabla x_t, \nabla^2 x_t, ..., \nabla^{d-1} x_t$ remain non-stationary.

Definition 1.3.2 Order of integration is a summary statistic that it indicates the minimum number of differences needed to get a stationary series.

A special mention should be made for the class of I(0) models. From definition 1.3.1 it is clear that x_t is I(0) if $y_t=(1-B)^0 x_t=x_t$ is stationary. From this definition $I(0) \Rightarrow$ stationary but not vice versa, i.eI(0) is a sufficient condition for stationarity, but stationarity is only a necessary condition for I(0).Therefore, all stationary processes are I(0), but not all I(0) processes are stationary.

Definition 1.3.3A random walk is a time series model x_t such that

$$x_t=x_{t-1}+w_t,\ w_t \sim WN(0,\sigma_w^2) \qquad 12$$

While the mean of a random walk is still zero, the covariance is actually time-dependent. Hence a random walk is non-stationary because

$$\gamma_x(h)=Cov(x_t,x_{t+h})=t\sigma_w^2$$

Random walk models are widely used for non-stationary data, particularly financial and economic data. Random walks typically have:

a) long periods of apparent trends up or down and b) sudden and unpredictable changes in direction

It is also easy to see that by differencing x_t,we find that $\nabla x_t:=(1-B)x_t=w_t$ is stationary, thus the random walk process is a I(1) process.

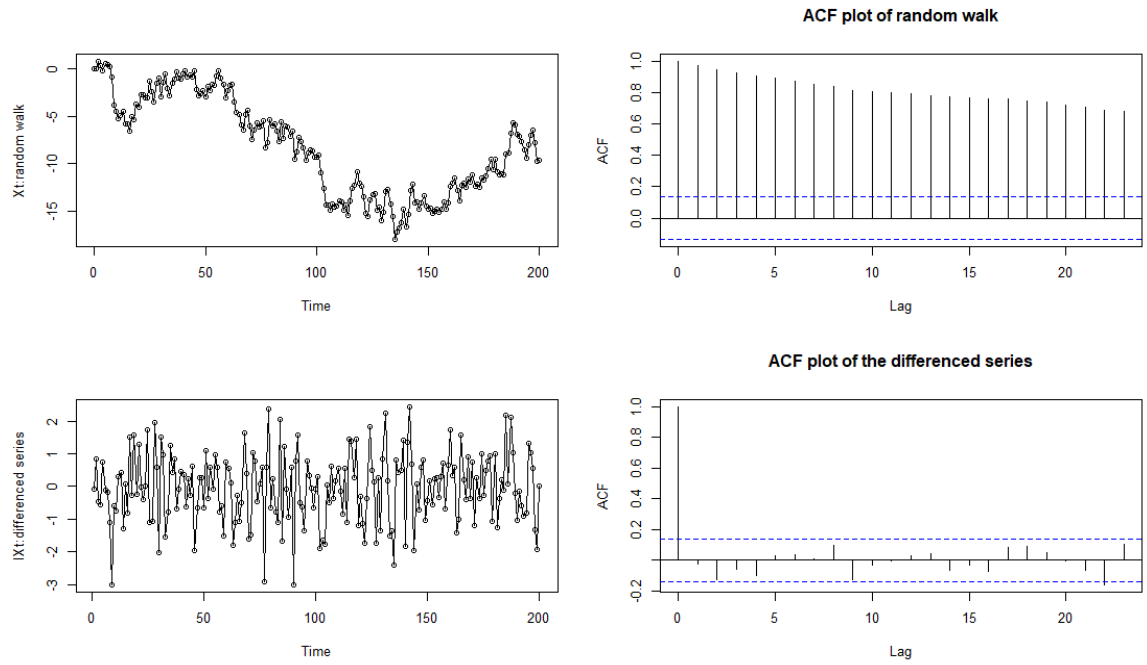


Figure 1-1:Simulated random walk Xt=Xt-1+Wt, Wt~WN(0,1), of sample size n=200 with its acf plot on the upper half and the differenced process with its acf plot on the lower half. The random walk procedure appears the charachteristics we expected with non-decreasing autocovariances when on the other hand the differenced series is a white noise procedure with statistically insignificant autocovariances.

The integrated ARMA, or ARIMA, model is a broadening of the class of ARMA models to include differencing.

Definition 1.3.4A process x_t is said to be ARIMA(p,d,q) if

$$\nabla^d x_t=(1-B)^d x_t \qquad 13$$

is ARMA(p,q). In general, we will write the model as

$$\varphi(B)(1-B)^d x_t=\theta(B)w_t \qquad 14$$

There are two different ways of modeling a linear trend. A *deterministic trend* is obtained using the regression model $x_t=\beta_0+\beta_1 t+\eta_t$ where η_t is an ARMA process.

A *stochastic trend* and a *deterministic trend* are obtained using the model $\nabla x_t := (1-B)x_t = \beta_0 + \eta_t$ where η_t is anARMA process. In the latter case, if $\beta_0 = 0$, then we have a stochastic trend only, otherwise, if $\beta_0 \neq 0$ we have both a deterministic and a stochastic trend. In other words,

$$x_t = x_{t-1} + \beta_0 + \eta_t$$

2. The SARIMA class of models

In time series data, seasonality is the presence of variations that occur at specific regular intervals such as weekly, monthly, quarterly or yearly. Seasonality may be caused by various factors, such as weather, vacation, and holidays and consists of periodic, repetitive, and generally regular and predictable patterns in the levels of a time series.

Seasonal series are characterized by a strong serial correlation at the seasonal lag (and possibly multiples thereof). The classical decomposition of the time series in its additive form: $x_t = \mu_t + s_t + y_t$ where μ_t is the trend component, s_t is the seasonal component and y_t is the white noise component. If deterministic seasonality of lag d is assumed, then it is implied that $s_t = s_{t-d}$, which means that the seasonal component repeats itself over every cycle of d data observations.

2.1 Definition of the model

In practice it may not be reasonable to assume that the seasonality component repeats itself precisely in the same way cycle by cycle. Seasonal ARIMA models allow for randomness in the seasonal pattern from one cycle to the next, describing thus a form of seasonality which is known as "stochastic seasonality".

Definition 2.1.1 If d and D are non-negative integers, then $\{x_t\}$ is said to be a seasonal $\mathrm{ARIMA}(p,d,q)\times(P,D,Q)_s$ with period s if the differenced process $y_t := (1-B)^d(1-B^s)^D x_t$ is a causal ARMA process,

$$\varphi(B)\Phi(B^s)y_t = \theta(B)\Theta(B^s)z_t,\ z_t \sim WN(0,\sigma_z^2) \qquad 15$$

where $\varphi(z) = 1-\varphi_1 z-\ldots-\varphi_p z^p$, $\Phi(z) = 1-\Phi_1 z-\ldots-\Phi_P z^P$, $\theta(z) = 1+\theta_1 z+\ldots+\theta_q z^q$ and $\Theta(z) = 1+\Theta_1 z+\ldots+\Theta_Q z^Q$.

Note that the process $\{y_t\}$ is causal if and only if $\varphi(z) \neq 0$ and $\Phi(z) \neq 0$ for all $|z| \leq 1$.

- p and seasonal P: indicate number of autoregressive terms (lags of the stationarized series)

- dand seasonal D: indicate differencing that must be done to stationarize series
- q and seasonal Q: indicate number of moving average terms (lags of the forecast errors)
- s: indicates seasonal length in the data

Box et al. (2008) propose that the SARIMA(p, d, q) × (P, D, Q)s for x_t can be seen as a special form of the equivalent representation of y_t as an ARMA(p + sP, q + sQ) written as

$$\varphi^*(B)y_t=\theta_0+\theta^*(B)z_t$$

where $\varphi^*(B)=\varphi(B)\Phi(B)$, $\theta^*(B)=\theta(B)\Theta(B)$ *16*

2.2Identifying d and D

The first (and most important) step in fitting aSARIMA model is the determination of the orders d and D of differencing needed to stationarize the series. The determination will be made by taking into consideration both the time series correlogram (ACF and PACF plots) and widely used tests such as the ADF test for unit roots and the HEGY test for seasonal unit roots.

2.2.1 Graphical tools

Normally, the correct amount of differencing is the lowest order of differencing that yields a time series which fluctuates around a well-defined mean value and whose autocorrelation function (ACF) plot decays fairly rapidly to zero, either from above or below. If the series still exhibits a long-term trend, or otherwise lacks a tendency to return to its mean value, or if its autocorrelations are positive out to a high number of lags (e.g., 10 or more), then it needs a higher order of differencing.

Some basic indications for determining the orders d and D of an SARIMA(p,d,q) model are the following:

Indication 1: If the series has positive autocorrelations out to a high number of lags, then it probably needs a higher order of differencing.

Differencing tends to introduce negative correlation: if the series initially shows strong positive autocorrelation, then a non-seasonal difference will reduce the autocorrelation and perhaps even drive the lag-1 autocorrelation to a negative value. If we apply a second non-seasonal difference (which is occasionally necessary), the lag-1 autocorrelation will be driven even further in the negative direction.

Indication 2: If the lag-1 autocorrelation is zero or negative, or the autocorrelations are all small and patternless, then the series does not need a higher order of differencing.

One of the most common errors in ARIMA modeling is to "overdifference" the series and end up adding extra AR or MA terms to undo the damage. If the lag-1 autocorrelation is more negative than -0.5 (and theoretically a negative lag-1 autocorrelation should never be greater than 0.5 in magnitude), this may mean the series has been overdifferenced. The time series plot of an overdifferenced series may look quite random at first glance, but from a closer look, a pattern of excessive changes in sign from one observation to the next will be observed.

Indication 3: The optimal order of differencing is often the order of differencing at which the standard deviation is lowest.

The first two rules do not always unambiguously determine the "correct" order of differencing. "Mild underdifferencing" can be compensated for by adding AR terms to the model, while "mild overdifferencing" can be compensated for by adding MA terms instead. In some cases, there may be two different models which fit the data almost equally well: a model that uses 0 or 1 order of differencing together with AR terms, versus a model that uses the next higher order of differencing together with MA terms. In trying to choose between two such models that use different orders of differencing, we may need to ask what assumption we are most comfortable making about the degree of nonstationarity in the original series

Indication 4: A model with no orders of differencing assumes that the original series is stationary. A model with one order of differencing assumes that the original series has a constant average trend (e.g. a random walk). A model with two orders of total differencing assumes that the original series has a time-varying trend (e.g. a random trend)

Another consideration in determining the order of differencing is the role played by the constant term in the model-if one is included. The presence of a constant allows for a non-zero mean in the series if no differencing is performed, it allows for a non-zero average trend in the series if one order of differencing is used, and it allows for a non-zero average trend-in-the-trend if there are two orders of differencing. We generally do not assume that there are trends-in-trends, so the constant is usually removed from models with two orders of differencing. In a model with one order of differencing, the constant may or may not be included, depending on whether we do or do not want to allow for an average trend.

Indication 5: If the series has a strong and consistent seasonal pattern, then you must use an order of seasonal differencing (otherwise the model assumes that the seasonal pattern will fade away over time). However, never use more than one order of seasonal differencing or more than 2 orders of total differencing (seasonal+nonseasonal)

2.2.2 Tests for unit roots and seasonal unit roots

Apart from the graphical tools and the rules base on the characteristics of the ACF and PACF plots, the order of intergration d can be found with the most common unit root test, the Augmented Dickey-Fuller (ADF) test.

The order of seasonal differencing D can be found with the Canova-Hansen (CH) test and the Hylleberg-Engle-Granger-Yoo (HEGY) test.These tests are used to evaluate the seasonal stationarity of each time series and have been shown to complement each other (Hylleberg, 1995). Further Ghysels et al. (1994) performed a Monte Carlo experiment to compare some seasonal unit root tests and concluded that the HEGY test performs best of all included tests.

2.2.2.1 The DF and ADF tests for unit roots

The Dickey-Fuller test is generally used for identifying the presence of a unit root in an AR model $y_t=py_{t-1}+w_t$, $w_t \sim WN(0,\sigma_w^2)$ which can be rewritten as a regression model as

$$\nabla y_t=(p-1)y_{t-1}+w_t=\delta y_{t-1}+w_t$$

It takes different forms depending on what is the type of the model under consideration:

a) Testing for a unit root: $\nabla y_t=(p-1)y_{t-1}+w_t$

b) Testing for a unit root with a drift: $\nabla y_t=a_0+(p-1)y_{t-1}+w_t$

c) Testing for a unit root with a drift and a deterministc trend: $\nabla y_t=a_0+a_1t+(p-1)y_{t-1}+w_t$

After the model parameters have been estimated we can test if any unit roots are present by checking the hypothesis: $\{H_0:p=1 \text{ vs } H_1:|p<1|\}$. It should be noted that although the test statistic used is similar to a t-test statistic, we cannot use the standard t-distribution to calculate critical values. For this purpose a special table of critical values, simply known as Dickey-Fuller table, is used.

The Dickey-Fuller, abbreviated DF, is generalized into the Augmented Dickey-Fuller, abbreviated ADF, test to accommodate the general ARIMA and ARMA models. The Augmented Dickey-Fuller test allows for higher order autoregressive processes by including ∇y_{t-p} in the model. But our test is still if p=1 in

$$\nabla y_t = a+bt+(p-1)y_{t-1}+\delta_1\nabla y_{t-1}+\delta_2\nabla y_{t-2}+...+\delta_{p-1}\nabla y_{t-p+1}+w_t,\ w_t \sim WN(0,\sigma_w^2)$$

It is not always easy to tell if a unit root exists because these tests have low power against near-unit-root alternatives. To some extent, that criticism is invalid, because all tests have low power when the actual parameter value is close to the null value. To some extent, it is valid because there are tests that apply under broader sets of assumptions, or are simply more powerful than the ADF test.

There are also size problems (false positives) because we cannot include an infinite number of augmentation lags as might be called for with MA processes.

2.2.2.2 The HEGY test for seasonal unit roots

There are 3 approaches for modeling seasonal time series: deterministic seasonal processes, stationary and non-stationary seasonal processes. The differences lie in how they react to the shocks to the seasonal patterns. In deterministic seasonal processes have no effect on the seasonal pattern, and in stationary seasonal processes they have temporary effect which would diminish with time passes by. But in non-stationary processes the shocks have non-diminishing effect, causing permanent changes to the seasonal pattern and increasing variance of the series. Therefore, the non-stationary seasonal process raises the most concern and testing seasonal unit roots has high priority in the modeling procedure. The miss-specification of the type of seasonality would cause severe bias in modeling and forecasting process.

Consider a basic autoregressive polynomial φ(B) having a form

$$\varphi(B)=1-B^S$$

where as usual B is the lag operator and S is the number of time periods in a seasonal pattern which repeats regularly. For example, S=12 for monthly data when the seasonal pattern repeats itself every year or S=7 for daily data. The equation φ(z)=0 has S roots on the unit circle

$$z_k = e^{\frac{2\pi i}{S}k} = \cos\left(\frac{2k\pi}{S}\right)+i\sin\left(\frac{2k\pi}{S}\right),\ k=0,1,...,S-1$$

Each root z_k in (17) is related to a specific frequency $\frac{2k\pi}{S}$. When k=0,the root z_k in (17) is called non-seasonal unit root. The other roots z_k in (17) are called seasonal unit roots.

Except for roots z_k in (17) at frequencies 0 and π, z_k in (17) are pairs of conjugation frequencies. We can reorder the S frequencies of z_k by putting their conjugations frequencies together. We distinguish two cases depending on whether S is an odd or an even number.

1) When S is an even number, the S frequencies are ordered as:

$$\theta_m=\begin{cases}0,\ m=1\\ \frac{m-1}{S}\pi,\ m=2,4,...,(S-2)\\ 2\pi-\frac{m-2}{S}\pi,\ m=3,5,...,(S-1)\\ \pi,\ m=S\end{cases} \qquad 18$$

In (18), θ_m and θ_{m+1} are conjugation frequencies if m=2,4,…,(S-2).

2) When S is an odd number, there is no unit root at frequency π, the S frequencies are ordered as:

$$\theta_m=\begin{cases}0,\ m=1\\ \frac{m-1}{S}\pi,\ m=2,4,...,(S-2)\\ 2\pi-\frac{m-2}{S}\pi,\ s=3,5,...,(S-1)\\ \pi,\ m=S\end{cases} \qquad 19$$

In (19), θ_m and θ_{m+1} are conjugation frequencies if m=2,4,…,(S-2).

For both cases (18) and (19),the unit roots corresponding to frequency θ_m are:

$$u_m=\cos\theta_m+i\sin\theta_m \qquad 20$$

The frequencies θ_m also indicate the number of cycles for u_m in the seasonal pattern, which are derived by $\frac{\theta_m S}{2\pi}$.For example, consider hourly data where S=24,setting m=2, $u_2=\cos\frac{\pi}{12}+i\sin\frac{\pi}{12}=$ $=\frac{\sqrt{6}+\sqrt{2}}{4}+i\frac{\sqrt{6}-\sqrt{2}}{4}$. Its frequency is $\theta_m=\frac{\pi}{12}$,and it corresponds to $\frac{\theta_m S}{2\pi}=1$ cycle every 24 hours.

As discussed above, the seasonal unit roots in time series would permanently change the seasonal patterns of the series and make the variance of the series increase linearly. Therefore testing seasonal unit roots proceeds modeling seasonality

Among, the seasonal unit root tests, the HEGY test which was proposed by Hylleberg, Engle, Granger and Yoo (1990) has the advantage of testing seasonal unit root at each frequency separately, thus it is widely adapted. The HEGY test was originally proposed for testing seasonal unit roots in quarterly data but later its use was extended also for monthly data (Beaulieu and Miron (1992)).

First, we define the seasonal difference operator of period s as

$$\nabla_s := 1-B^s$$

Note that: $\nabla_s \neq \nabla^s = (1-B)^s$

The HEGY test for seasonal integration is conducted by estimating the following regression (special case for quarterly data)

$$\nabla_4 y_t = a+\beta t+\sum_{j=2}^{4} b_j Q_{jt} + \sum_{i=1}^{4} \pi_i W_{i,t-1} + \sum_{m=1}^{p-1} \gamma_m \nabla_4 y_{t-m} + \varepsilon_t \qquad 21$$

where a and t are the drift and deterministic time trend terms, Q_{jt} is a seasonal dummy variable, and the W_{it} are given below:

$$W_{1t} = (1+B)(1+B^2)y_t$$
$$W_{2t} = -(1-B)(1+B^2)y_t$$
$$W_{3t} = -(1-B)(1+B)y_t$$
$$W_{4t} = -B(1-B)(1+B)y_t = W_{3t-1}$$

In addition, as in the Augmented Dickey-Fuller tests, some lagged values of the dependent variable are included to assure the appropriate behavior of the residuals. The literature suggests, among other criteria, to include all the lagged values until p-1; more details about how p is determined can be found in Perron (1995). The procedure starts from a very long lagged model and reduce the number of lags until the last lag included is statistically different from zero at some prespecified level of significance (in general, a 10% level of significance is used). The introduction of these lags (as in the Augmented Dickey-Fuller tests) is to assure that the residuals have the standard properties (no serial autocorrelation). Finally, ε_t is an error term with the standard white noise properties.

After OLS estimation, tests are conducted for $\pi_1=0$,for $\pi_2=0$ and a joint test of the hypothesis $\pi_3=\pi_4=0$ .If none of the π_i 's are equal to zero, then the series is stationary (both at seasonal and non-seasonal frequencies).

Interpretation of the different π_i 's is as follows:

1) If $H_0:\pi_1=0$ is rejected in favor of $\pi_1 < 0$, then there is no long-run (non-seasonal) unit root. π_1 is on $W_{1t}=S(B)y_t$ which has had all of the seasonal unit roots removed.

2) If $H_0:\pi_2=0$ is rejected in favor of $\pi_2 < 0$, then there is no semi-annual unit root.

3) If $H_0:\pi_3=\pi_4=0$ is rejected in favor of $\pi_3<0$ or $\pi_4<0$, then there is no unit root in the annual cycle.

The HEGY test has been expanded for monthly time series (12 roots) by Franses (1991) and Beaulieu and Miron (1993). For monthly data, the seasonal unit roots are

$$1,-1,\pm i,\left(-\frac{1}{2}\pm\frac{\sqrt{3}}{2}i\right),\left(\frac{1}{2}\pm\frac{\sqrt{3}}{2}i\right),\left(-\frac{\sqrt{3}}{2}\pm\frac{1}{2}i\right),\left(\frac{\sqrt{3}}{2}\pm\frac{1}{2}i\right)$$

These roots correspond to the frequencies: 0, π, ±π/2; ±2π/3; ±π/3; ±5π/6; and ±π/6. These frequencies correspond to the bi-monthly case, four-month case, quarterly case, six-month and 12-month cases[1].

In the monthly case, the procedure requires the estimation of the following expression

$$\nabla_{12}y_t = a+\beta t+\sum_{j=2}^{12}b_jQ_{jt}+\sum_{i=1}^{12}\pi_iW_{i,t-1}+\sum_{m=1}^{p-1}\gamma_m\nabla_{12}y_{t-m}+\varepsilon_t \qquad 22$$

Where Q_{jt} are deterministic seasonal dummies, and $W_{i,t-1}$ are transformations of the y_t that provide the basis for testing unit roots at zero and the rest of the frequencies.

Let assume that π_1 corresponds to the unit root at frequency 0 and π_2 corresponds to the unit root at frequency π. To verify the presence of unit roots, we need to test if the 2 parameters equal to 0. The null hypothesis is

[1]If the unit circle is 2π, a month has π of a cycle; therefore, every two months there is one cycle and there are six cycles in the year.

$$H_{0m}:\pi_m=0$$

against the alternative hypothesis $H_{1m}:\pi_m<0$,where m=1 or 2. The test statistics used here are t-statistics: $t_m=\pi_m/\sigma_m$,where π_m is the OLS estimator of π_m and σ_m is the sample standard error of π_m .

Due to the fact that the complex unit roots are conjugate pairs, the regressors appear in pairs and correspond with frequencies in pairs. Thus, only that both parameters are zero could proof the existence of unit roots. This leads to the joint test of each pair. The null hypothesis is:

$$H_{0m}:\pi_m=\pi_{m+1}=0$$

The alternative hypothesis $H_{1m}:\pi_m\neq 0$ or $\pi_{m+1}\neq 0$,m=3,5,7,…. The F-statistics

$$F_{m,m+1}=\frac{1}{2}\left(t_m^2+t_{m+1}^2\right) \qquad 23$$

are used, where the t-statistics t_m and t_{m+1} are derived in the same way as t_1 and t_2 . If the null hypothesis is not rejected, the test indicates unit roots exist at the corresponding 2 frequencies.

There is no known distribution for the HEGY test statistics but only asymptotic distributions which can be used in order for the tests to be carried out. In general, the asymptotic distributions of the HEGY test statistics typically depend on the specification chosen for the deterministic component.

2.3 Identifying the number of AR and MA terms

In this section the method of selecting the remaining orders will be described. The autoregressive orders p and P and the moving average orders q and Q will be decided partly by the correlogram (ACF,PACF) andand partly by minimizing information criterion's with the use of the HK-algorithm.

2.3.1 Graphical tools

The ACF and PACF plots arefirst used to make guesses for appropriate orders. More specifically for orders p,q:

Indication 1: If the PACF displays a sharp cutoff and ACF declines gradually $\Rightarrow$ AR signature. Also an AR series is usually positively autocorrelated at lag 1.

The lag at which the PACF cuts off is the indicated number of AR terms.

Indication 2: ACF that cuts off sharply after a few lags and PACF that declines gradually ⇒MA signature. Also, an MA series is usually negatively autocorrelated at lag 1.

The lag at which the ACF cuts off is the indicated number of MA terms.

Indication 3 It is possible for an AR term and an MA term to cancel each other's effects, so if a mixed ARMA model seems to fit the data, also try a model with one fewer AR term and one fewer MA term.

Now for the seasonal orders P, Q we have:

Indication 4: If the autocorrelation of the appropriately differenced series is positive at lag s, where s is the number of periods in a season, then consider adding an SAR term to the model. If the autocorrelation of the differenced series is negative at lag s, consider adding an SMA term to the model. The latter situation is likely to occur if a seasonal difference has been used, which should be done if the data has a stable and logical seasonal pattern. The former is likely to occur if a seasonal difference has not been used, which would only be appropriate if the seasonal pattern is not stable over time. You should try to avoid using more than one or two seasonal parameters (SAR+SMA) in the same model, as this is likely to lead to overfitting of the data and/or problems in estimation.

In order to sum up, assuming that the differencing implied by d and D have been done to the time series, meaning that the SARIMA(p,d,q)×(P,D,Q)s of the original series y_t has taken its equivalent representation as a SARMA(p,q)×(P,Q)s of the differenced series y_t^*, we can estimate graphically the appropriate orders for the SARIMA model by using the results in Table 2.1:

Table 2.2: ACF and PACF to identify the orders of SARMA(p,q)×(P,Q)s

	ACF	PACF
AR(p)	Exponentially decreasing or damped sine wave	Spikes to lag p then zero
MA(q)	Spikes to lag q then zero	Exponential decreasing or damped sine wave
ARMA(p,q)	Exponentially decreasing or damped sine wave after q-p lags	Exponentially decreasing or damped sine wave after p-q lags
SAR(P)s	Exponentially decreasing or damped sine wave for all lags times s	Spikes for lag Ps then zero

SMA(Q)s	Spikes for lag Qs then zero	Exponentially decreasing or damped sine wave for all lags times s
SARMA(P,Q)s	Exponentially decreasing or damped sine wave for all lags times s after lags (Q-P)s	Exponentially decreasing or damped sine wave for all lags times s after lags (P-Q)s

2.3.2 Selection with the HK-algorithm

However this procedure of finding patterns from ACF and PACF plots is considered subjective for mixed and seasonal processes. To make the model selection less subjective some frequently used likelihood based information criterions are applied. These are the Akaike information criterion (AIC), the AIC with correction for small samples (AICc) and the Bayes information criterion (BIC).

The Hyndman-Khandakar (HK) algorithm was developed by Hyndman and Khandakar (2008) and can be applied in R with the function auto.arima in the forecast package. They suggest an iterative time-saving procedure where the model with the smallest value of AIC, AICc or BIC will be found much faster, since it is now found without comparing every possible model.

To derive these information criterions the first thing that is needed is the likelihood function, $L(\Psi)$, where Ψ is the maximum likelihood estimates of the parameters for the SARIMA with n=p+q+P+Q+1 parameters and sample size T. The criteria are then derived by the following equations:

$$AIC=-2\log\left(L\left(\Psi\right)\right)+2n \qquad 24$$

$$AICc=AIC+\frac{2n\left(n+1\right)}{T-n+1} \qquad 25$$

$$BIC=-2\log\left(L\left(\Psi\right)\right)+n\log(T) \qquad 26$$

Note: It has been shown that AIC has a tendency to choose a model that is over-parametrized (Hurvich and Anderson, 1989). Further Burnham and Anderson (2004) suggest that AIC and AICc should be valued over BIC and Brockwell and Davis (1991) propose that AICc is most fit for selecting orders of SARIMA models. Thus, the model selected by AICc will be most valued.

The HK-algorithm then performs an iterative procedure to select the model that minimizes the value of each criterion. It begins with estimation of the following four models:

1)SARIMA(2,d,2)×(1,D,1)s

2)SARIMA(0,d,0)×(0,D,0)s

3)SARIMA(1,d,0)×(1,D,0)s

4)SARIMA(0,d,1)×(0,D,1)s

where d and D are assumed to have been found previously and a constant is included in the models if $d + D \leq 1$. The model which attains the smallest value for the chosen information criterion is then selected and the procedure continues with varying the parameters in the following ways:

• Let each of p, q, P and Q vary with ±1.

• Let both p and q vary with ±1 at the same time.

• Let both P and Q vary with ±1 at the same time.

• Include the intercept if previously not included otherwise do the opposite.

This step of the procedure will be repeated until none of these variations decreases the value of the criterion.

There are some constraints that follows with the use of this method. These are used to check that the model is reasonable and well-fitted and are the following

• The maximum orders of p and q are five.

• The maximum orders of P and Q are two.

• All non-invertible or non-causal models are rejected. These are found by computing the roots of the lag polynomials $\varphi(B)\Phi(B)$ and $\theta(B)\Theta(B)$, if any root is smaller than 1.001 then the model is rejected.

• If errors arise when fitting the model with the non-linear optimization routine then the model is rejected.

2.4 Estimating the parameters of the model selected

The second part of the Box-Jenkins methodology for SARIMA modeling is estimation. Once a model has been tentatively identified, its parameters have to be efficiently estimated, and the resulting fit assessed, mainly by an analysis of residuals, to see whether it can be accepted as a

plausible explanation of the series. If the model is found to be inadequate, then this assessment should indicate promising modifications to the identification, and the cycle is repeated; and so on until the analyst is satisfied (Box and Jenkins, 1976).

For the general ARMA(p,q) process $\varphi(B)x_t=\theta(B)w_t$, $w_t \sim WN(0,\sigma_w^2)$ in which, the order parameters, p and q, are known there are three techniques for estimating the p+q+1 unknown parameters:

a) Two-Step Regression Estimation b) Yule-Walker Estimation c) Maximum Likelihood Estimation

Although, for the class of SARIMA models the first two approaches prove to be inadequate for the estimation task and only the Maximum Likelihood method can be applied.

If the model contains only AR terms, the unknown coefficients can be estimated using ordinary least squares. However, if the model contains MA terms too, the task becomes more complicated, because the lagged values of the innovations are unobservable. Consequently, it is not possible to derive explicit expressions to estimate the unknown coefficients and therefore one has to use maximum likelihood for estimation purposes.

In order to convey an idea of what follows, we will shortly outline the procedure: first, one sets up the SARIMA model - in the following also called original model - with some initial values for the unknown vector of parametersψ. Restrictions can be imposed on the coefficients. The simplest restriction is that some of the coefficients are zero. Then the value of the likelihood function, given the initial parameters, is evaluated.

Unfortunately, in most cases the original SARIMA model cannot be estimated directly. If one looks at the general SARIMA representation one recognizes on the left hand side the product of two expressions. Both of them contain lag-operators. Such expressions have to be telescoped out first. The result we get is an ordinary ARMA(p,q) model which can be estimated.

Under the assumption that an ARMA model is stationary and invertible and that the observations are normally distributed it can be estimated using the maximum likelihood approach. Once the conditional sum of squares function is calculated numerical methods have to be applied to maximize the likelihood function with respect toψ. By making suitable assumptions about initial conditions, the maximum likelihood estimators can be obtained by minimizing the conditional sum of squares.

However, the conditional sum of squares is not always very satisfactory for seasonal series. In that case the calculation of the exact likelihood function becomes necessary (Box and Jenkins, 1976).

The approach of maximum likelihood (ML) requires the specification of a particular distribution for a sample of T observations y_t. Let

$$f_{y_T,y_{t-1},...,y_1}(y_T,y_{t-1},...,y_1|\psi)$$

denote the probability density of the sample given the unknown nx1 parametersψ. It can be interpreted as the probability of having observed the given sample (Hamilton, 1994).

With the sample $y=\{y_T,y_{t-1},...,y_1\}$ at hand, the above given probability can be rewritten as a function of the unknown parameters given the sample y. Following the notation of Box and Jenkins (1976), we use the notation $l(\psi|y)=\ln L(\psi|y)$.

The maximum likelihood estimate ψ is the parameter vector that maximizes the probability for the observed sample y. Thus, the MLE satisfies the so-called likelihood equations, which are obtained by differentiating $l(\psi|y)$ with respect to each of the unknown parameters of the vectorψ and setting the derivatives to zero. Using vector notation and suppressing y, we obtain

$$\frac{\partial l(\psi)}{\partial \psi}=0 \qquad 27$$

As a rule, the likelihood equations are non-linear. Therefore, the ML estimates must be found in the course of an iterative procedure. This is true for the exact likelihood function of every Gaussian ARMA(p,q) process (Hamilton, 1994).

In many applications of ARMA models the conditional likelihood function is an alternative to the exact likelihood function. In that case, one assumes that the first p observations of a Gaussian ARMA(p,q) process are deterministic and are equal to its observed values $\{y_p,...,y_1\}$. The initial residuals a_t for $t\in\{p,...,p-q+1\}$ are set to its expected values 0. The log likelihood function for this setting is of the following form with $\psi=(\psi',\sigma^2)$

$$l(\psi)=-\frac{1}{2}(T-p)\ln(2\pi)-\frac{1}{2}(T-p)\ln(\sigma^2)-\frac{S(\psi')}{2\sigma^2} \qquad 28$$

and $S(\psi')$ denoting the sum of squares function

$$S(\psi')=\sum_{t=p+1}^{T}\left(a_t(\psi')\right)^2 \qquad 29$$

The notation $a_t(\psi')$ emphasizes that a_t is no longer a disturbance, but a residual which depends on the value taken by the variables in ψ'.

Note, that the parameter σ^2 is an additional one, that is not included in vector ψ'. It is easy to see that (28) is maximized with respect to ψ' if the sum of squares $S(\psi')$ is minimized. Using the condition (27), this leads to

$$\sum_{t=1}^{T} \frac{\partial a_t(\psi')}{\partial \psi'} a_t = 0 \qquad 30$$

Thus, the ML estimate for ψ' can be obtained by minimizing (29). Furthermore, we obtain from (28) and (27) that

$$\sigma^2 = \frac{S\left(\psi^{'}\right)}{T\text{-}p} \qquad 31$$

Thus, given the parameter vector $\psi^{'}$ that maximizes the sum of squares, and thus the conditional log likelihood function (28), one divides the sum of squares by T-p to obtain the ML estimate σ^2.

Once the ML estimate ψ is calculated, one wants to have standard errors for testing purposes. If T is sufficiently large, the ML estimate $l(\psi)$ is approximately normally distributed with the inverse of the information matrix divided by T as covariance matrix. The inverse of the Hessian for $l(\psi)$ is one way to estimate the covariance matrix (Hamilton, 1994). One can calculate the Hessian applying numerical methods.

2.5 Diagnostics

The fit of the model is evaluated by diagnostic checks of the residuals. The residuals should behave like Gaussian white noise that is appear random, homoscedastic and normal (Box and Jenkins, 1976). The first part is a graphical check of the standardized residuals, meaning the residuals divided with their standard deviation. These should look random and homoscedastic. The number of outliers are also important where a good indication would be that about 95 percent of the residuals lie inside their 95 percent confidence interval±1.96 (Brockwell and Davis, 1991). The next step is to evaluate the assumption of randomness by using the sample autocorrelation function of the residuals. The autocorrelations of interest are those that are significantly different from zero, that is those who lie outside the sample size dependent approximately 95 percent confidence interval $\pm 2/\sqrt{T}$ (Hamilton, 1994). Those significant lags suggest some kind of inconsistency in the residuals, but there is no reason to worry if only about five percent of the autocorrelations are significant. (Brockwell and Davis, 1991) In this thesis the most important part of the diagnostic checking is the use of tests to possibly acquire statistically significant results which would imply a rejection of the fitted model.

The chosen tests are the Box-Ljung test which is used to test the serial independence and the Jarque-Bera test which tests the normality of the residuals.

2.5.1 The Box-Ljung test

The Ljung and Box (BL) test was developed in 1978 and is used to test the randomness of the residuals. For this test the first step is to extract the residuals $\hat{e}_t$ for the fitted model. The T residuals are then used to derive the sample autocorrelations of the residuals with the following equation

$$\hat{r}_k = \frac{\sum_{t=k+1}^{T} \hat{e}_t \hat{e}_{t-k}}{\sum_{t=1}^{T} \hat{e}_t^2}, \quad k=1,2,... \qquad 32$$

This equation is used until a set of autocorrelations $\hat{r}_1, \hat{r}_2, ..., \hat{r}_m$ have been obtained. These are then used to test the null hypothesis of serially independent residuals versus the alternative hypothesis that they are not serially independent with the following test statistic (Ljung and Box, 1978)

$$Q(r) = T(T+2)\sum_{k=1}^{m}(T-k)^{-1}\hat{r}_k^2 \qquad 33$$

which for an appropriate model was shown to be asymptotically distributed as a $\chi^2_{1-\alpha}(m)$ where m is the number of lagged autocorrelations included and α is the selected significance level. Harvey, 1989 suggest that the number of lags should be a function of T for example the truncated value of $m = \sqrt{T}$ and that the degrees of freedom should be corrected for SARIMA models to df = m−p−q−P−Q. The critical value is included in (Ljung and Box, 1978) and then compared to the value of the test statistic. The null hypothesis of randomness is rejected for large values of the test statistic.

2.5.2 The Jarque-Bera test

The Jarque-Bera test is used to test the normality of the residuals. The null hypothesis of the test is normality and it is tested against the alternative hypothesis of non-normality. The statistic that is used is the first part of equation 4 in Jarque and Bera (1980), written in the following way

$$JB = T\left(\frac{S^2}{6} + \frac{K^2-3}{24}\right) \qquad 34$$

where T is the number of observations, S is the skewness and K is the kurtosis. The statistic is assumed to be distributed as a $\chi^2_{1-\alpha}(2)$ variable and the null hypothesis of normality is rejected for large values of the statistic (Schwert, 2009).

2.6 Forecasting

If the selected SARIMA model has already been estimated and has passed all the diagnostic tests the final step remaining is forecasting. In forecasting, the goal is to predict future values of a time series x_{n+m} ,m=1,2,…, based on the data collected to the present $\underset{\rightarrow}{\mathbf{x}}=\{x_n,x_{n-1},...,x_1\}$.

It has been shown that the SARIMA(p,d,q)×(P,D,Q)s for the variable x_t , t = 1,...,T can be written equivalently as an ARMA(p + sP,q + sQ) for y_t from Equation (16). The forecast function for the assumed stationary variable y_t is then written as

$$\left(y_{t+1|t}-\mu\right)=\varphi(B)^*\left(y_t-\mu\right)+\theta(B)^*\hat{e}_t \qquad 35$$

where $\hat{e}_t=x_t-y_{t|t-1}$ (Hamilton, 1994). The forecast for lead time τ, meaning the time that follows after the last observed information, is then derived by

$$\left(y_{t+s|t}-\mu\right)=\varphi(B)^*\left(y_{t+s-1|t}-\mu\right)+\theta(B)^*\hat{e}_{t+s-1} \qquad 36$$

That is the forecast for lead time τ will be derived by the previously observed values of y_t , previous forecasts of y and the residuals $\hat{e}_t$ which have been derived for all time points up to the last observed observation but are equal to zero for the ones where the real values have not yet been observed. (Hamilton, 1994)

2.7 Goodness of fit

2.7.1 Error measures and tests

The accuracy of the forecasts obtained will be evaluated by some error measures. Error measures are easily derived and give an indication for which model that is most suitable for forecasting. These will be derived by the use of y_t which is the real value, y_t which is the forecast value, n which is the length of the forecast horizon and t which goes from one to n. The first error measure that is presented is the mean squared error (MSE) which is estimated by

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - y_t\right)^2 \qquad 37$$

A value of zero would imply a perfect forecast and a negative value should not be possible to obtain. The next error measure is the mean absolute deviation (MAD) is derived by

$$MAD = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - y_t\right| \qquad 38$$

For this measure the value of zero would imply a perfect forecast and a negative value should not be possible. The difference between MAD and MSE is that MSE places relatively greater penalty on large forecast errors. Further the mean absolute percentage error (MAPE) is calculated by

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - y_t}{y_t}\right| \qquad 39$$

A value of zero would imply a perfect forecast and a negative value cannot possibly be obtained. It is also important to state that this error measure is not defined if $y_t = 0$ for any t. The MAPE scales the errors, that is this measure puts relatively more penalty to the forecast error if the true value of the observation is small. The last of these measures is a statistic called Theil's U. This statistic compares the ratio between the MSE implied by the model forecast and the MSE of forecasts obtained by a naive model which sets the upcoming observation at time t+1 as equal to the value observed at time t. It is derived by (Petris et al., 2009)

$$U = \sqrt{\frac{\sum_{t=1}^{n}\left(y_t - y_t\right)^2}{\sum_{t=1}^{n}\left(y_t - y_{t-1}\right)^2}} \qquad 40$$

The value of Theil's U will be smaller than one if the model forecast has higher accuracy than the naive model forecast. The estimate will be zero if the forecasts are perfect and it cannot possibly be derived as negative. It is not defined if the naive model perform perfect forecasts (Petris et al., 2009).

The Diebold and Mariano (DM) test was developed in 1995 and it's first assumed that there are two forecasts $w_1,...,w_h$ and $\hat{z}_1,...,\hat{z}_h$ of the true time series $y_1,...,y_h$ where h is the forecast horizon. These are then used to derive the forecast errors of the first $e_{1i} = y_i - w_i$ and the second model $e_{2i} = y_i - \hat{z}_i$. (Diebold and Mariano, 1995).

This test works even if the forecast errors have a non-zero mean are non-Gaussian and are correlated to each other. For the test in R with function dm.test the forecast package with default specifications

implies that the errors are used to derive the loss functions $g(e_{1i}) = |e_{1i}|^2$ and $g(e_{2i}) = |e_{2i}|^2$ which lead to the loss-differential series $d_i = \{g(e_{1i}) - g(e_{2i})\}_{i=1}^{h}$ which is assumed normally distributed. The sample mean is then derived by $\bar{d} = \frac{\sum_{i=1}^{h} d_i}{h}$

The next step is to derive the autocovariance at lag 0 for di, that is $\hat{\gamma}(0)$ which is assumed to be a consistent estimate of the variance of $h\bar{d}$. The test statistic then follows and is derived by

$$DM = \frac{\bar{d}}{\sqrt{\hat{\gamma}(0)/h}} \sim N(0,1) \qquad 41$$

Thus, the DM statistic is assumed to be standard normal and the rejection values then follows from that. The null hypothesis of the test is that there is no difference between the accuracy of the two forecasts. The alternative hypothesis is either two-sided to test if either model perform better than the other or one-sided to test if one specific model is more accurate than the other. (Diebold and Mariano, 1995)

2.7.2 Cross-validation for time series

Because of the nature of time series data there are strong temporal dependencies and so it is not possible to perform the usual validation techniques. A very useful way to overcome the difficulty of biased evaluations is to use the procedure called time series cross-validation.

Most cross validation schemes appear to rely on having i.i.d. data because the training-test splits do not take time indices into account. For example, 5-fold cross validation applied naively over 5 time periods would ignore the sequential nature of time, mixing up past, present and future as in figure 2.1

	Time 1	Time 2	Time 3	Time 4	Time 5
Split 1:	Test set	Training set	Training set	Training set	Training set
Split 2:	Training set	Test set	Training set	Training set	Training set
Split 3:	Training set	Training set	Test set	Training set	Training set
Split 4:	Training set	Training set	Training set	Test set	Training set
Split 5:	Training set	Training set	Training set	Training set	Test set

Figure 2-1.: Usual 5-fold cross validation scheme

This pattern would lead to wrong conclusions since autoregressive models require a contiguous block of data, because they rely on the presence of autocorrelations at predefined lags (instead of having training sets split into 2 parts).

Instead we can use a *rolling* or *sliding window* approach to cross validation to avoid this issue. For 5 time periods, you would split the sets as follows:

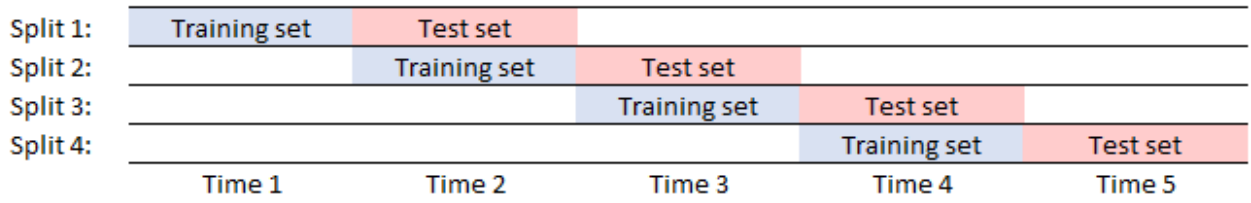


Figure 2-2:Rolling or sliding window cross-validation scheme for time series data

Assume we want to perform a 1-step ahead forecast for a time series of n-values starting with a training set of the first k-values, then the time series cross validation follows the next steps:

Step1: Select the observations at, and prior to, time k

Step2: Select the observation at k+1 for the test data

Step3: Compute the forecasted value for k+1 based on the values 1,2,…,k and the forecast error

Step4: Slide the training set window by one and use the values 2,3,…,k, k+1 (where the (k+1)th observation is the real value of the time series)

Step4: Slide the training set window by one and use the values 2,3,…,k, k+1 (where the (k+1)th observation is the real value of the time series)to forecast the value for time k+2.

3.State space models and the Kalman filter

In this chapter we discuss the basic notions about state-space models and their use in time series analysis. Dynamic linear models (DLM) are presented as a special case of general state space models, being linear and Gaussian. For DLM, estimation and forecasting can be obtained recursively by the well known Kalman filter.

State-space models appeared in the time series literature in the seventies (Akaike (1974), Harrison and Stevens (1976)) and became established during the eighties (West and Harrison (1997)). In the last decades they have become a focus of interest. This is due on one hand to the development of models well suited to time series analysis, but also to an even wider range of applications, including for instance molecular biology or genetics, and on the other hand to the development of computational tools, such as modern Monte Carlo methods, for dealing with more complex nonlinear and non-Gaussian situations.

State-space models consider a time series as the output of a dynamic system perturbed by random disturbances. As we shall see, they allow a natural interpretation of a time series as the result of several components, such as trend, seasonal or regressive components. At the same time, they have an elegant and powerful probabilistic structure, offering a flexible framework for a very wide range of applications. Computations can be implemented by recursive algorithms. The problems of estimation and forecasting are solved by recursively computing the conditional distribution of the quantities of interest, given the available information. In this sense, they are quite naturally treated from a Bayesian approach.

State-space models are based on the idea that the time series Y_t is an incomplete and noisy function of some underlying unobservable process $\theta_t, t = 1, 2, \ldots,$ called the state process.. More generally, we might think of as an auxiliary random process which facilitates the task of specifying the probability law of the time series: the observable process Y_t depends on the latent state process θ_t, which has a simpler, Markovian dynamics, and we can reasonably assume that the observation Y_t only depends on the state of the system at the time the measurement is taken, θ_t. Figure 3.1 represents graphically the dependences among variables that we are assuming.

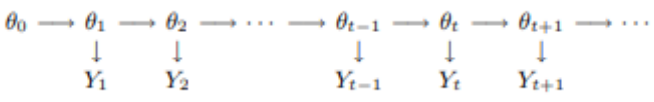


Figure 3-1: Dependence structure for a state space model

More formally, the assumptions of a state space model are

A1) $(\theta_t, t=0,1,2,...)$ is a Markov chain; that is, θt depends on the past values $(\theta_0,\theta_1,...,\theta_{t-1})$ only through θ_{t-1}. Thus, the probability law of the process $(\theta_t, t=0,1,2,...)$ is specified by assigning the initial density $p_0(\theta_0)$ of θ_0 and the transition densities $p_0(\theta_t|\theta_{t-1})$ of θ_t conditionally on θ_{t-1}.

A2) Conditionally on $(\theta_t, t=0,1,...)$, the Yt's are independent and Y_t depends on θ_t only. It follows that, for any $n \geq 1$, $(Y_1,...,Y_n)|\theta_1,...,\theta_n$ have joint conditional density $\prod_{i=1}^{n} f(y_t|\theta_t)$.

With the help of the graph for understanding the conditional independence relations implied by the model, we find that, for any $n \geq 1$,

$$(\theta_0,\theta_1,...,\theta_n,Y_1,...,Y_n) \sim p_0(\theta_0)\prod_{t=1}^{n} p(\theta_t,Y_t|\theta_0,\theta_1,...,\theta_{t-1},Y_1,...,Y_{t-1}) =$$

$$p_0(\theta_0)\prod_{t=1}^{n} f(Y_t|\theta_0,...,\theta_t,Y_1,...,Y_{t-1})\, p(\theta_t|\theta_0,...,\theta_{t-1},Y_1,...,Y_{t-1}) = p_0(\theta_0)\prod_{t=1}^{n} f(Y_t|\theta_t)\, p(\theta_t|\theta_{t-1})$$

In particular, we see that the process $((\theta_t,Y_t), t=1,2,...)$ is Markovian. The density of $(Y_1,...,Y_n)$ can be obtained by integrating out all the θ-variables from the joint density of the above expression. As we shall see, computations are fairly simple in Gaussian linear state space models; however, in general the density of $(Y_1,...,Y_n)$ is not available in close form and the observable process Y_t is not Markovian.

3.1 DLM for time series analysis

The first, important class of state-space models is given by Gaussian linear state-space models, also called dynamic linear models (DLM). These models are specified by means of two equations

$$Y_t = F_t\theta_t + v_t,\ v_t \sim N_m(0,V_t)$$
$$\theta_t = G_t\theta_{t-1} + w_t,\ w_t \sim N_p(0,W_t)$$

42

where Gt and Ft are known matrices and the (vt) and (wt) are two independent white noise sequences (i.e., they are independent, both between them and within each of them), with mean zero and known covariance matrices Vt and Wt respectively. The first equation above is called the observation equation, the second state equation or system equation. Furthermore, it is assumed that θ_0 has a Gaussian distribution,

$$\theta_0 \sim N_p(m_0, C_0) \qquad 43$$

for some non-random vector m_0 and matrix C_0, and it is independent on (vt) and (wt). One can show that the DLM satisfies the assumptions (A.1) and (A.2) of the previous section, with $Y_t|\theta_t \sim N(F_t\theta_t, V_t)$ and $\theta_t|\theta_{t-1} \sim N(G_t\theta_{t-1}, W_t)$.

DLM is a specific case of the general state space models specified by the set of equations

$$Y_t = h_t(\theta_t, v_t)$$
$$\theta_t = g_t(\theta_{t-1}, w_t)$$

with arbitrary functions g_t and h_t. For the class of DLM we add also the assumption that g_t and h_t are linear functions and the assumption of Normal distributions.

3.1.1 Trend models

Polynomial DLM are the models most commonly used for describing the trend of a time series, where the trend is viewed as a smooth development of the series over time. At time t, the expected trend of the time series can be thought of as the expected behavior of Y_{t+k} for $k \geq 1$, given the information up to time t.

Let us denote with D_t the information provided by the first t observations $Y_1,...,Y_t$, the expected trend is the forecast function $f_t(k) = E(Y_{t+k}|D_t)$. A polynomial model of order n is a DLM with constant matrices Ft = F and Gt = G, known covariance matrices Vt and Wt, and a forecast function of the form

$$f_t(k) = E(Y_{t+k}|D_t) = a_{t,0} + a_{t,1}k + ... + a_{t,n-1}k^{n-1}, k \geq 0 \qquad 44$$

where $a_{t,0},...,a_{t,n-1}$ are linear functions of $m_t = E(\theta_t|D_t)$ and are independent of k. Thus, the forecast function is a polynomial of order $(n - 1)$ in k. Roughly speaking, any reasonable shape of the forecast function can be described or closely approximated by a polynomial, by choosing n sufficiently large. However, one usually thinks of the trend as a fairly smooth function of time, so that in practice small values of n are used. The most popular polynomial models are the random walk plus noise model, also known as the local level model, which is a polynomial model of order n = 1, and the linear growth model, that is a polynomial model of order n = 2.

3.1.1.1 The local level model

The random walk plus noise, or local level model, is defined by the following two equations

$$\begin{aligned} Y_t &= \mu_t + v_t,\ v_t \sim N(0,V) \\ \mu_t &= \mu_{t-1} + w_t,\ w_t \sim N(0,W) \end{aligned} \qquad 45$$

where the error sequences (vt) and (wt) are independent, both within them and between them. This is a DLM with m = p = 1, $\theta_t = \mu_t$ and Ft = Gt = 1. Intuitively, it is appropriate for time series showing no clear trend or seasonal variation: the observations (Yt) are modeled as random fluctuations around a level (μ_t); in turn, the level can evolve randomly over time (described by a random walk). This is why the model is also called local level model; if W = 0, we are back to the constant mean model.

3.1.1.2 The linear growth model

A slightly more elaborated model is the linear growth model, which has the same observation equation as the local level model, but includes a time-varying slope in the dynamics for μ_t.

$$\begin{aligned} Y_t &= \mu_t + v_t,\ v_t \sim N(0,V) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + w_{1t},\ w_{1t} \sim N(0,\sigma^2_{w_1}) \\ \beta_t &= \beta_{t-1} + w_{2,t},\ w_{2,t} \sim N\left(0,\sigma^2_{w_2}\right) \end{aligned} \qquad 46$$

with uncorrelated errors. This is a DLM with

$$\theta_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix},\ G = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},\ W = \begin{bmatrix} \sigma^2_{w_1} & 0 \\ 0 & \sigma^2_{w_2} \end{bmatrix},\ F = \begin{pmatrix} 1 & 0 \end{pmatrix} \qquad 47$$

The system variances $\sigma^2_{w_i}$, are allowed to be zero.

3.1.2 Seasonal factor models

Suppose that we have seasonal data from quarters $Y_t, t=1,2,...$ Assume also for brevity that the series has zero mean: a non-zero mean, or a trend component, can be modelled separately, so for the moment we consider the series as purely seasonal. We might describe the series by introducing seasonal deviations from the zero mean, expressed by different coefficients a_i for the different quarters, i=1,2,3,4. So, if Y_{t-1} refers to the first quarter of the year and Y_t to the second quarter, we assume

$Y_{t-1}=a_1+v_{t-1}$
$Y_t=a_2+v_t$ *48*

and so on. This model can be written as a DLM as follows. Let $\theta_{t-1}=(a_1,a_4,a_3,a_2)^T$ and $F_t=F=(1,0,0,0)$. Then the observation equation of the DLM is given by

$$Y_t=F\theta_t+v_t$$

which corresponds to (48). The state equation must 'rotate' θ_{t-1} into a vector $\theta_t=(a_2,a_1,a_4,a_3)^T$, so that $Y_t=F\theta_t+v_t=a_2+v_t$. The required permutation of the state vector can be obtained by a permutation matrix G so defined

$$G=\begin{bmatrix}0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\end{bmatrix}$$

Then the state equation can be written as

$$\theta_t=G\theta_{t-1}+w_t=(\alpha_2,\alpha_1,\alpha_4,\alpha_3)^T+w_t$$

In the static seasonal model, w_t is degenerate on a vector of zeros (i.e., $W_t=0$). More generally, the seasonal effects might change in time, so that W_t is non-zero and has to be carefully specified.

In general, a seasonal time series with period s can be modeled through an s-dimensional state vector θ_t of seasonal deviations, by specifying a DLM with F = (1, 0, . . . , 0) and G given by a s by s permutation matrix. Identifiability constraints have to be imposed on the seasonal factors $a_1,...,a_s$. A common choice is to impose that they sum to zero, $\sum_{j=1}^{s}a_j=0$. The linear constraint on the s seasonal factors implies that there are effectively only s−1 free seasonal factors, and this suggests an alternative, more parsimonious representation that uses an (s − 1)-dimensional state vector. For the example given by (48), one can consider $\theta_{t-1}=(a_1,a_4,a_3)^T$ and $\theta_t=(a_2,a_1,a_4)^T$, with F = (1, 0, 0). To go from θ_{t-1} to θ_t, assuming for the moment a static model without system evolution errors and using the constraint $\sum_{i=1}^{4}a_i=0$, one has to apply the linear transformation given by the matrix

$$G=\begin{bmatrix}-1&-1&-1\\1&0&0\\0&1&0\end{bmatrix}$$

More generally, for a seasonal model with period s, one can consider an (s-1) dimensional state space, with F=(1,0,…,0) and

$$G=\begin{bmatrix} -1 & -1 & \ldots\ldots & -1 & -1 \\ 1 & 0 & \ldots\ldots & 0 & 0 \\ 0 & 0 & \ldots\ldots & 1 & 0 \end{bmatrix}$$

A dynamic variation in the seasonal components may be introduced via a system evolution error with variance $W=\mathrm{diag}\left(\sigma_w^2,0,....,0\right)$.

3.1.3 Combining component models

In the previous sections we have presented some common models for the different components (trend, seasonality) of a time series. These models can be used as "building blocks" for constructing a DLM for a time series with a more complex behavior. The additive structure of the DLM allows to easily combine the different component models, as discussed at the beginning of this section.

Suppose, for example, that a series Y_t is the sum of a trend component $Y_{L,t}$ and a seasonal component $Y_{S,t}$:

$$Y_t = Y_{L,t} + Y_{S,t} + v_t$$

We can construct a DLM for each component, so that

$$\begin{aligned} Y_{L,t} &= F_{L,t}\theta_{L,t} \\ \theta_{L,t} &= G_{L,t}\theta_{L,t-1} + w_{L,t}, \quad w_{L,t} \sim N\left(0, W_{L,t}\right) \end{aligned}$$

and

$$\begin{aligned} Y_{S,t} &= F_{S,t}\theta_{S,t} \\ \theta_{S,t} &= G_{S,t}\theta_{S,t-1} + w_{S,t}, \quad w_{S,t} \sim N\left(0, W_{S,t}\right) \end{aligned}$$

Define F_t and θ_t as partitioned matrices

$$F_t = \left(F_{L,t}, F_{S,t}\right), \ \theta_t = \begin{pmatrix} \theta_{L,t} \\ \theta_{S,t} \end{pmatrix}$$

and G and W_t as block-diagonal matrices

$$G_t = \begin{pmatrix} G_{L,t} & \\ & G_{S,t} \end{pmatrix}, \ W_t = \begin{pmatrix} W_{L,t} & \\ & W_{S,t} \end{pmatrix}$$

Then Y_t is described by a DLM with observation equation

$$Y_t = F_t\theta_t + v_t = F_{L,t}\theta_{L,t} + F_{S,t}\theta_{S,t} + v_t, \quad v_t \sim N(0, V_t) \qquad 49$$

and state equation

$$\theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N(0, W_t) \qquad 50$$

3.2 Kalman filtering

Here, to get started, we consider the model as given, that is we assume that the densities $f(y_t|\theta_t)$ and $p(\theta_t|\theta_{t-1})$ have been specified. For a given state-space model, the main tasks are to make inference on the unobserved states or predict future observations based on a part of the observation sequence. Estimation and forecasting are solved by computing the conditional distributions of the quantities of interest, given the available information.

For estimating the state vector we compute the conditional densities $p(\theta_s|y_1,...,y_t)$. We distinguish between problems of filtering (when s = t), state prediction (s > t) and smoothing (s < t). It is worth to underline the difference between filtering and smoothing. In the filtering problem, the data are supposed to arrive sequentially in time. This is the case in many applied problems: think for example of the problem of tracking a moving object, or of financial applications where one has to estimate, day by day, the term structure of interest rates, updating the current estimates as new data are observed on the markets the following day, etc. In these cases, we want a procedure for estimating the current value of the state vector, based on the observations up to time t ("now"), and for updating our estimates and forecasts as new data become available at time t + 1. For solving the filtering problem, we compute the conditional density $p(\theta_t|y_1,...,y_t)$.In DLM, the Kalman filter provides the formulas for updating our current inference on the state vector as new data become available, that is for passing from the filtering density $p(\theta_t|y_1,...,y_t)$ to $p(\theta_{t+1}|y_1,...,y_{t+1})$.

However in general the actual computation of the relevant conditional densities is not at all an easy task. Dynamic linear models are one relevant case where the general recursions simplify considerably. In this case, using standard results about the multivariate Gaussian distribution, it is easily proved that the random vector $(\theta_0,\theta_1,...,\theta_t,Y_1,...,Y_t)$ has a Gaussian distribution for any

$t \geq 1$. It follows that the marginal and conditional distributions are also Gaussian. Since all the relevant distributions are Gaussian, it suffices to compute their means and covariances.

Theorem 3.1 (Kalman filter): For the DLM specified by (42), if $\theta_{t-1}|D_{t-1} \sim N(m_{t-1}, C_{t-1}), t \geq 1$ then

a) The one step ahead state predictive density of θ_t, given D_{t-1} is Gaussian with parameters

$$a_t = E(\theta_t|D_{t-1}) = G_t m_{t-1}$$
$$R_t = Var(\theta_t|D_{t-1}) = G_t C_{t-1} G_t^T + W_t$$

b) The one step ahead predictive density of Y_t given D_{t-1} is Gaussian, with parameters

$$f_t = E(Y_t|D_{t-1}) = F_t a_t$$
$$Q_t = Var(Y_t|D_{t-1}) = F_t R_t F_t^T + V_t$$

c) The filtering density of θ_t given D_t is Gaussian, with

$$m_t = E(\theta_t|D_t) = a_t + R_t F_t^T Q_t^{-1} e_t$$
$$C_t = Var(\theta_t|D_t) = R_t - R_t F_t^T Q_t^{-1} F_t R_t$$

where $e_t = Y_t - f_t$ is the forecast error.

The Kalman filter allows to compute the predictive and filtering densities recursively, starting from $\theta_0|D_0 \sim N(m_0, C_0)$ then computing $p(\theta_1|D_1)$ and proceeding recursively as new data information becomes available.

3.3 Kalman smoothing

The problem of smoothing, or retrospective analysis, consists instead in estimating the state sequence at times 1,…,t, given the data $y_1,...,y_t$. In many applications, one has observations on a time series for a certain period, and wants to retrospectively study the behavior of the system underlying the observations; for example one can think the problem of an economic researcher trying to understand the socio-economic behavior that explains the time series GDP of a country for the same period of time. The smoothing problem is solved by computing the conditional distribution of $\theta_1,...,\theta_t$ given $Y_1,...,Y_t$.

Again, one has a backward-recursive algorithm for computing the conditional densities of $\theta_t|D_T$, for t< T, starting from the filtering density $p(\theta_T|D_T)$ and estimating backward all the states' history.

Theorem 3.2 (Kalman smoothing): For the DLM specified by (42), if $\theta_{t+1}|D_T \sim N(s_{t+1},S_{t+1})$, then $\theta_t|D_T \sim N(s_t,S_t)$ where

$$s_t = m_t + C_t G_{t+1}^T R_{t+1}^{-1}(s_{t+1}-a_{t+1})$$
$$S_t = C_t + C_t G_{t+1}^T R_{t+1}^{-1}(S_{t+1}-R_{t+1})R_{t+1}^{-1}G_{t+1}C_t$$

The Kalman-smoother allows to compute the densities of $\theta_t|D_T$, starting from t=T-1, in which case $\theta_t|D_T \sim N(s_T=m_T,S_T=C_T)$, and then proceeding backward for computing the densities of $\theta_t|D_T$ for t=T-2, t=T-3 and so on.

3.4 DLM innovations and model checking

As we have seen, for DLM the Kalman filter provides the filtering estimate m_t, given the information D_t, as the previous estimate m_{t-1} corrected by a term which depends on the forecast error

$$e_t = Y_t - E(Y_t|D_{t-1}) = Y_t - f_t \qquad 51$$

The forecast errors can alternatively be written in terms of the estimation errors as follows

$$e_t = Y_t - F_t a_t = F_t\theta_t + v_t - F_t a_t \qquad 52$$

For the forecast errors e_t, $t \geq 1$ some interesting properties hold:

a) The expected value of e_t is zero, since $E(e_t)=E(E(e_t|D_{t-1}))=E(0)=0$

b) For $s \neq t$, e_s and e_t are uncorrelated.

c) The random vector e_t is uncorrelated with any function of $Y_1,...,Y_{t-1}$

d) $(e_t, t \geq 1)$ is a Gaussian process with $e_t \sim N_m(0,Q_t)$, t=1,2,...

The forecasts errors e_t are also called innovations. The representation $Y_t = f_t + e_t$ justifies the terminology, since one can think of Y_t as the sum of a component which is predictable from past observations, f_t, and another component, e_t, which is independent of the past and therefore contains the new information carried by the observation by the observation Y_t.

When the observations are univariate, the standardized innovations, defined by

$$\tilde{e}_t = \frac{e_t}{\sqrt{Q_t}},$$

are a Gaussian white noise, i.e. a sequence of independent identically distributed zero-mean normal random variables. This property can be exploited to check the model assumptions: if the model is correct, the sequence $\tilde{e}_1,...,\tilde{e}_t$ computed from the data should look like a sample of size t from a standard normal distribution. Many statistical tests, several of them readily available in R, can be carried out on the standardized innovations. They fall into two broad categories: those aimed at checking if the distribution of the $\tilde{e}_t$'s is standard normal, and those aimed at checking whether the $\tilde{e}_t$'s are uncorrelated. However, most of the time we take a more informal approach to model checking, based on the subjective assessment of selected diagnostic plots. The most illuminating are, in the authors' opinion, a QQ-plot and a plot of the empirical autocorrelation function of the standardized innovations. The former is used to assess normality, while the latter reveals departures from the property of uncorrelated innovations. A time series plot of the standardized innovations may prove useful in detecting outliers, change points and other unexpected patterns.

Finally, the same goodness of fit measures as described in section 2.7 can be used for the innovations e_t to evaluate the forecasting performance of the DLM model.

4. Application

4.1 New York monthly number of births data

For our first example, a dataset from monthly number of births from the city of New York is used. The dataset includes 168 entries of number of births from January 1946 to December 1959.

Firstly, restricting our investigation in the SARIMA class of models, the most suitable SARIMA(p,d,q)x(P,D,Q)s will be extracted and it will be evaluated by various graphical tools and tests. The model estimation will be performed on a training set of number of births from January 1946 to December 1956. A test set of 3 years from January 1957 to December 1959 will be used to evaluate its forecast ability.

For the second model we use the same training set for estimation and Kalman filtering, Kalman smoothing and forecasting of a local level state space model with a monthly seasonal component. It should be noted that the training set should consist of at least 50 observations since that is needed for efficient estimation (Box and Jenkins, 1976). Forecasts are found on the same test set as before from January 1957 to December 1959.

Finally, for the last part of the analysis assuming that both models have passed the diagnostic checks their forecast performance will be compared with some frequently used error measures and tests.

4.1.1 Description of the data

Firstly, we give a graphical representation of the whole dataset in figure 4.1 and calculate some basic descriptive measures for our data in table 4.1:

Table 4.1: Descriptive statistics for monthly number of births in thousands.

N	Mean	Median	Maximum	Minimum	Std	Skewness	Kurtosis
168	25.06	24.96	30	20	2.32	-0.02	-0.88

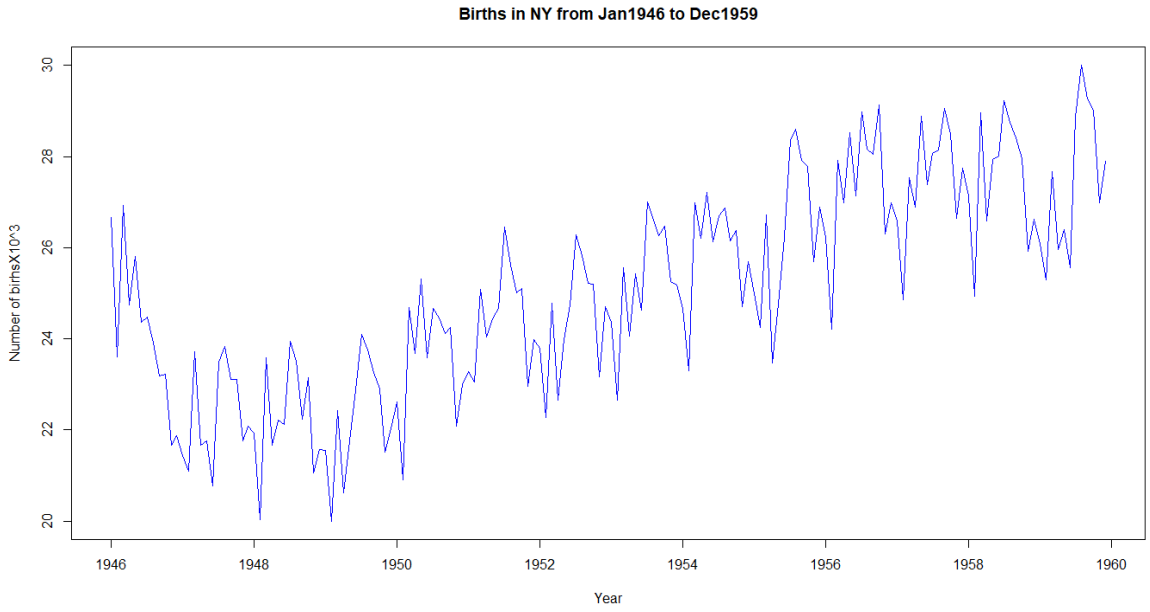


Figure 4-1: Number of births in New York City from January 1946 to December 1959

From figure 4.1, we can see that there is certainly some seasonal variation in the number of births per month; there is a peak every summer, and a trough every winter. It seems like this could be described using an additive model, as the seasonal fluctuations are roughly constant in size over time and do not seem to depend on the level of the time series, and the random fluctuations seem constant over time. A slight upward trend may also be present in the time series.

A short note here about time series decomposition how is done in practice. If we assume an additive decomposition $y_t=S_t+R_t+T_t$ where $y_t$is the data, S_t is the seasonal component, T_t is the trend-cycle component and R_t is the residuals component for period t.

The first step in a classical decomposition is to use a moving average method to estimate the trend-cycle component. A moving average of order m can be written as

$$T_t = \frac{1}{m}\sum_{j=-k}^{k} y_{t+j}$$

where m=2k+1. That is, the estimate of the trend-cycle at time t is obtained by averaging values of the time series within k periods of t. Observations that are nearby in time are also likely to be close in value. Therefore, the average eliminates some of the randomness in the data, leaving a smooth trend-cycle component.

The additive decomposition has the following steps:

Step1: If m is an even number, compute the trend-cycle component T_t using a 2×m-MA. If m is an odd number, compute the trend-cycle component T_t using an m-MA.

Step2: Calculate the detrended series y_t-T_t

Step3: To estimate the seasonal component for each season, simply average the detrended values for that season. For example, with monthly data, the seasonal component for March is the average of all the detrended March values in the data. These seasonal component values are then adjusted to ensure that they add to zero. The seasonal component is obtained by stringing together these monthly values, and then replicating the sequence for each year of data. This gives $\hat{S}_t$

Step4: The remainder component is calculated by subtracting the estimated seasonal and trend-cycle components: $R_t = y_t - \hat{S}_t - T_t$

Command decompose() in R calculates the time series components which are shown in figure 4.2.

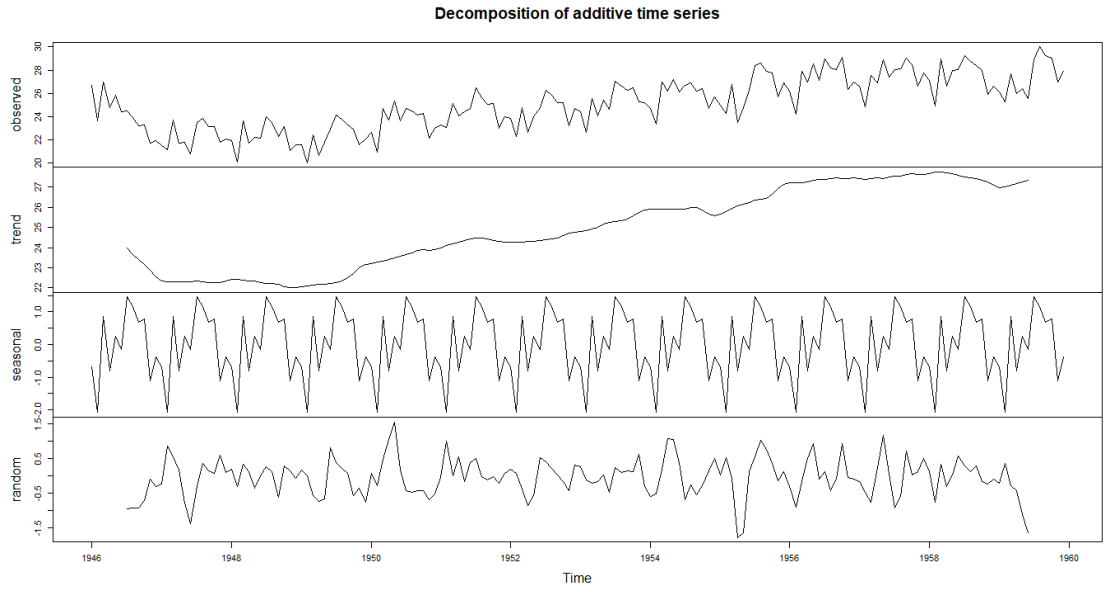


Figure 4-2: From top to bottom: a) The original births series b) The trend component c)The seasonal component and d) The random component

These are some first indications that we should take seasonal and non-seasonal differences, thus take d=1 and D=1, to stationarize the process. Our hypotheses are confirmed by figure 4.3.

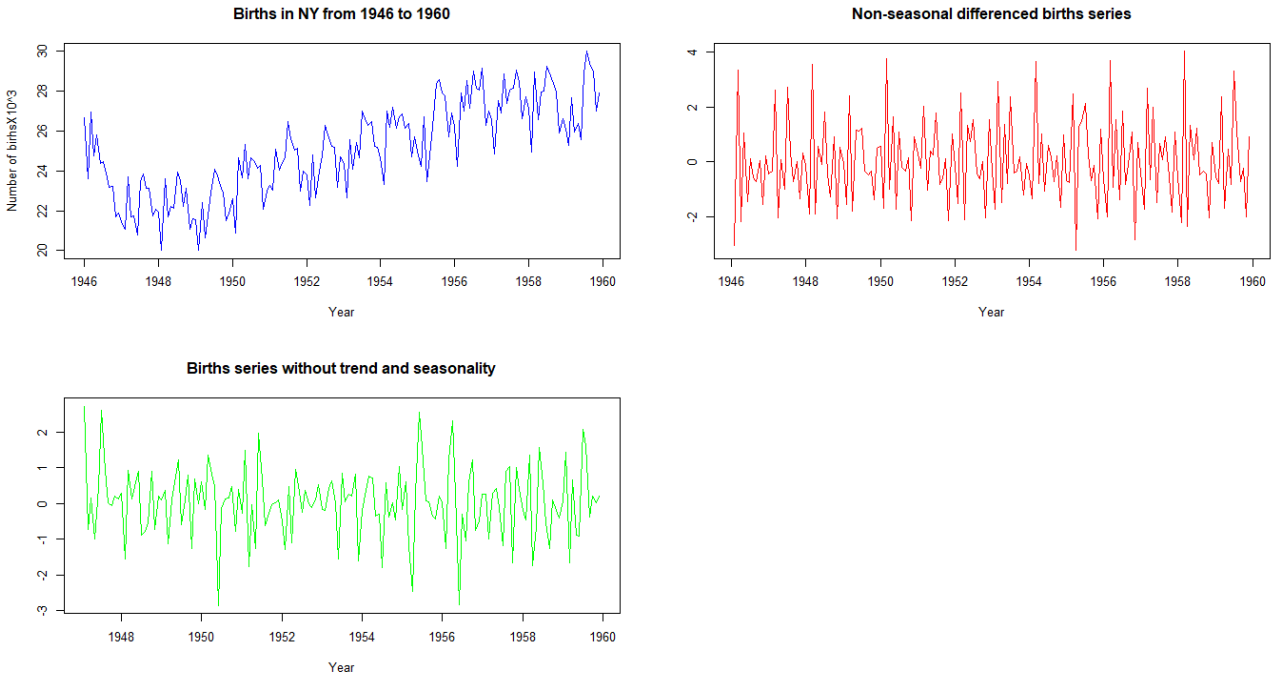


Figure 4-3: Time series plots of: a) The original births series (top left) b) The non-seasonal differenced births series (top right) and c) The seasonal and non-seasonal births series (bottom left)

From Figure 4.3, it is clear that once we have taken non-seasonal differences the trend has disappeared but there is still a seasonal pattern that is eliminated when we take seasonal and non seasonal differences. As we can see in the last panel of figure 4.3 the transformed series seems to have lost any trend and seasonal pattern and can be considered as a stationary process.

A strong graphical tool for the identification of the parameters of a SARIMA model is the study of the characteristics of the ACF and PACF plots of the transformed process (after taking seasonal and non-seasonal differences) which are presented in figure 4.4.

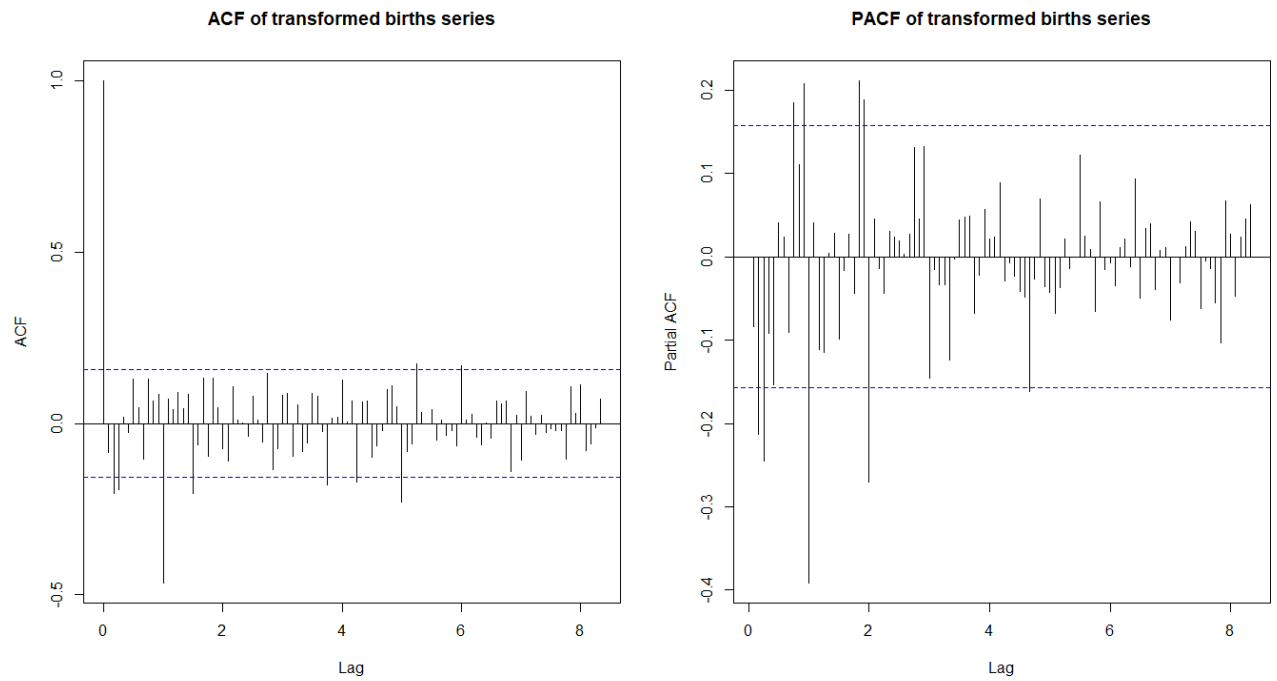


Figure 4-4: ACF plot (left) and PACF (right) of the transformed births series

From figure 4.4, if we look on seasonal lags there is only one major spike on seasonal lag 1, so that indicates the existence of an SMA(1) term in the model and thus the order of Q is probably 1. In the same manner, from the PACF plot of the transformed series if we look on seasonal lags there are two clear spikes on seasonal lag 1 and 2 that cross the 95% CI, that indicate the existence of SAR(1) and SAR(2) terms in the model and thus the order of P is probably 2. For the orders p,q it is not clear from the ACF and PACF plots if there are any AR and/or MA terms.

4.1.2 ARIMA and SARIMA models

After the preliminary analysis we are now ready to fit the most suitable SARIMA model for the births series. The first 11 years from January 1946 to December 1956 are used as a training set on which the model fit, parameters estimation and diagnostic tests will be carried out and the last 3 years of the births dataset from January 1957 to December 1959 will be utilized as a test set for the forecasting evaluation of the model.

The first step is to estimate the model parameters of integration and seasonal integration d, D respectively. From section we have already graphical evidence that d=D=1 but in order to substantiate our assumptions we conduct the ADF test for unit roots and the HEGY test for seasonal unit roots the results of which are shown in tables 4.2 and 4.3 respectively.

Table 4.2: ADF test statistics and p-values for the births series and first differenced births series. The null hypothesis is that the series has a unit root and the lag length used is 15.

Series	ADF statistic	p-value
Births	-2.9977	0.1616
Diff(Births)	-4.249	<0.01

Table 4.3:HEGY test with test-statistics for the first differenced births series and the first and seasonal differenced births series .The p-value is found in the parentheses

Frequency	Diff(Births)	Diff(Diff(Births),12)
0	-5.0503 (<0.001)	-4.001 (0.0011)
π/6	0.3075 (0.7363)	28.0599 (<0.001)
π/3	4.2288 (0.0161)	29.4525 (<0.001)
π/2	2.4559 (0.0889)	24.7756 (<0.001)
2π/3	4.5867(0.0114)	27.2751 (<0.001)
5π/6	1.5437 (0.2174)	16.4363 (<0.001)
π	-0.9243 (0.2487)	-6.0974 (<0.001)

The results from table 4.2 indicate that for the births series the ADF test with the null hypothesis of a unit root is not rejected on the five percent significance level but for the differenced births series we have strong evidence in order to reject the null hypothesis and thus we can conclude that the level of integration d=1.

From table 4.3 the seasonal unit root hypothesis is not rejected for all frequencies for the differenced births series (d=1, D=0) and the test is then performed on both first and seasonally differenced births series(d=1, D=1) with the result that a seasonal unit root is rejected on all frequencies on the five percent significance level. Thus, seasonally differencing first differenced births series seem to have made the time series stationary enough. Hence a seasonal difference operator of order D=1 seems appropriate.

The orders of integration that seem to be needed to make each series stationary enough have now been decided. The procedure of identifying the SARIMA can continue to the next step which is to determine the autoregressive and moving average orders.

The auto.arima function of the forecast package in R is running the HK algorithm for the specification of the ordersp,P,q,Q and the selection of the most suitable SARIMA model based on the information criteria AIC, AICc, and BIC. All the information criteria conclude to the same SARIMA(0,1,0)x(2,1,1)12 model. The estimated parameters with their standard errors are shown in table 4.4.

Table 4.4: Estimated model parameters of the SARIMA(0,1,0)x(2,1,1)12 model with their standard errors in parentheses

Parameter	Estimate	Standard Error
Φ_1	-0.4608	0.1399
Φ_2	-0.2548	0.1340
Θ_1	-0.7333	0.1736
σ^2	0.3784	-

The next step is to examine the usual model diagnostics in order to determine the model suitability. The model residuals are tested for their normality and independence via the Jarque-Bera test and the Ljung-Box test, the results of which are shown in table 4.5.

Table 4.5: Box-Ljung and Jarque-Bera tests for the assumptions of normality and independence of the residuals of the estimated SARIMA(0,1,0)x(2,1,1)12 model.

Test	Statistic	p-value
Box-Ljung	0.76909	0.3805
Jarque-Bera	7.7601	0.05

From table 4.5 we can see that forthe Box-Ljung test the null hypothesis of uncorrelated residuals cannot be rejected. On the other hand, the p-value of the Jarque-Bera test does not lead us to any conclusion about the assumption of normal distributed residuals since it is exactly equal to the default level of significance a=0.05. For this reason, it is advisable to check the QQplot of the residuals. In figure 4.5 are all the graphical results extracted from the function sarima() of the astsa package in R in order to check the model assumptions.

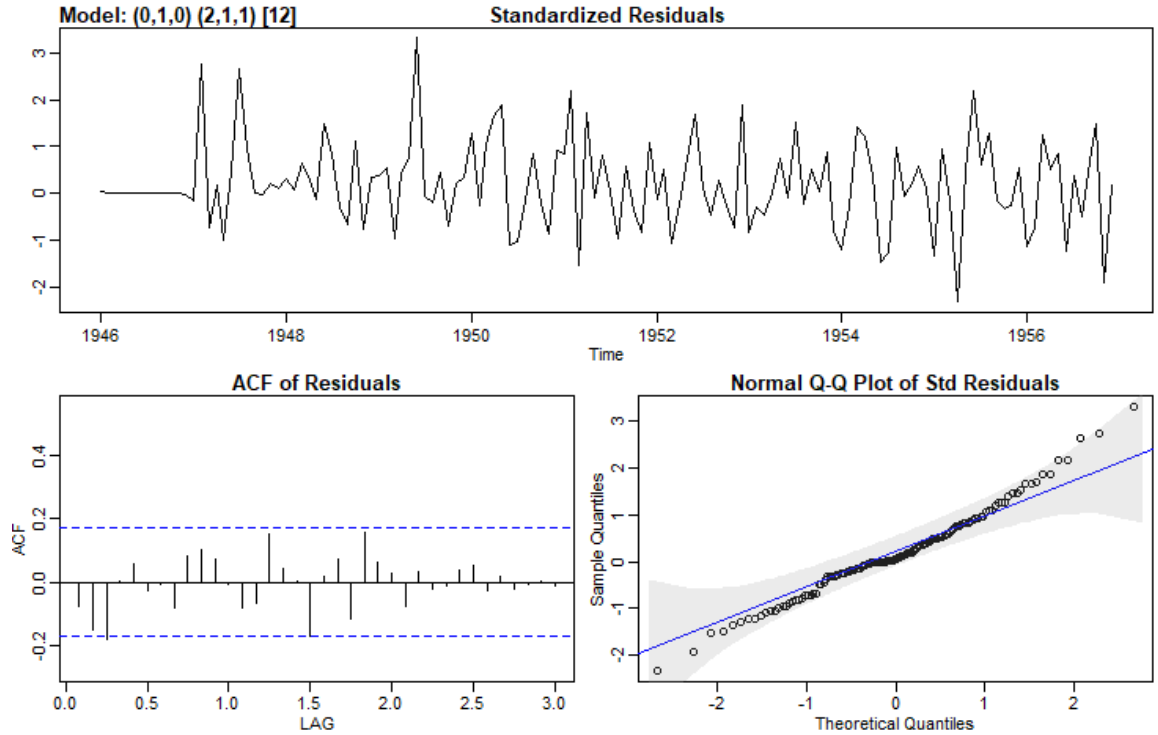


Figure 4-5: Plot of the standardized residuals (top), ACF plot of the residuals (bottom left) and QQplot of the standardized residuals (bottom right).

The standardized residuals in the first plot of figure 4.5 seem quite homoscedastic and to have a mean close to zero. Moreover, the ACF plot in figure 4.5 tells us that there is no significant autocorrelation for any lag and thus the residuals can be considered independent. Finally, the QQplotin figure 4.5 show that most of the residuals are close to the theoretical line and almost are inside the 95% CI bounds. So, the normality assumption of the residuals cannot be rejected.

An alternative approach of modeling the problem would be to consider also the case of a usual non-seasonal ARIMA model but without the strict parameters restrains of the HK algorithm. More specifically, we will consider an ARIMA model with d=1 where the parameters search would be performed in the grid $p \in [0,12], q \in [0,14]$. The most suitable model will be selected according to the combination of parameters that minimizes the AIC.

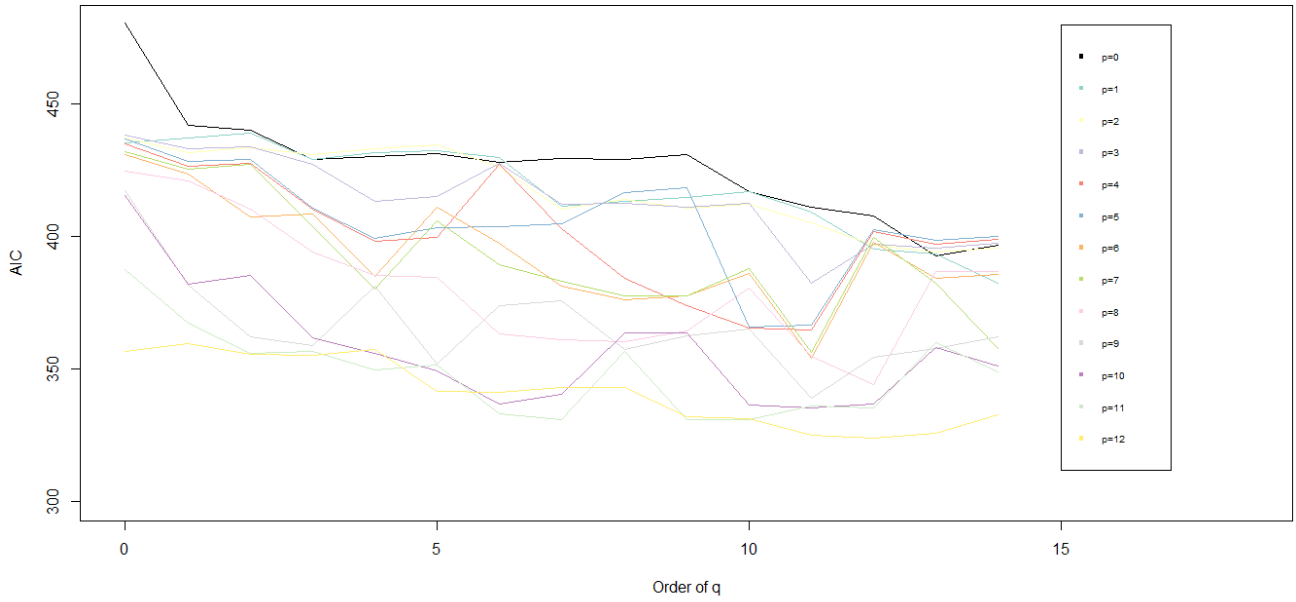


Figure 4-6: AIC as a function of the orders p and q

From figure 4.6 it is clear that the AIC is almost always smaller in value when the order of p=12. Moreover, for p=12 the AIC as a function of q appears to be a straight line after a local minimum for q=5. So, it is reasonable to assume that the best ARIMA model that minimizes the AIC is the one with parameters p=12 and q=5.

The estimated parameters of the ARIMA(12,1,5) model with their standard errors are presented in table 4.6.

Table 4.6: Parameters estimates with their standard errors for the ARIMA(12,1,5) model.

Parameter	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7	φ_8	φ_9
Estimate	0.1620	0.7025	-0.1022	-0.0959	0.0668	-0.0644	-0.1339	0.1194	0.2113
Std.Error	0.4096	0.1587	0.2937	0.1396	0.2657	0.1795	0.1179	0.1227	0.1511

Parameter	φ_{10}	φ_{11}	φ_{12}	θ_1	θ_2	θ_3	θ_4	θ_5	σ^2
Estimate	0.2113	-0.3227	0.2504	-0.4476	-1.000	0.3590	0.9793	-0.3612	0.5479
Std.Error	0.1117	0.1636	0.1472	0.4311	0.2081	0.4469	0.2240	0.4339	-

The model diagnostics for this model are shown in table 4.7 and figure 4.7.

Table 4.7:Box-Ljung and Jarque-Bera tests for the assumptions of normality and independence of the residuals of the estimated ARIMA(12,1,5) model.

Test	Statistic	p-value
Box-Ljung	0.10013	0.7517
Jarque-Bera	0.46371	0.7655

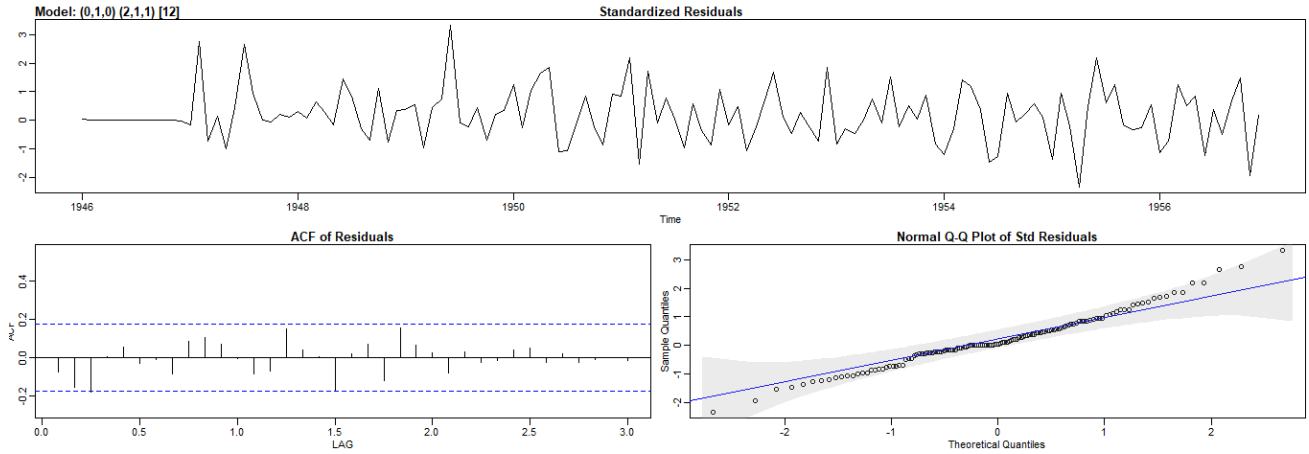


Figure 4-7: Plot of the standardized residuals (top), ACF plot of the residuals (bottom left) and QQplot of the standardized residuals (bottom right).

From the results of table 4.7 and the figure 4.7 it is logical to assume that the assumptions of normally distributed, homoscedastic and uncorrelated residuals cannot be rejected for the ARIMA(12,1,5) model.

Finally, since the results of table 4.3 revealed that not all the hypotheses for seasonal unit roots have been rejected for the differenced births time series by the HEGY test it would be reasonable to consider a SARIMA model with not very strictly restricted orders for p and q in order to adjust for stochastic seasonality in certain frequencies. Moreover, for the differenced and seasonally differenced births series the HEGY test rejects the null hypotheses for seasonal unit roots at all frequencies and thus we do not need any further seasonal differences to achieve stationarity. So for our search we will use a seasonal AR polynomial $\Phi(B)=1-B^S$ which means P=1 and for simplicity reasons we will consider that the seasonal MA polynomial is the simplest $\Theta(B)=1$ which mean that

Q=0. As we have already mentioned the most suitable model will be selected according to the combination of parameters that minimizes the AIC.

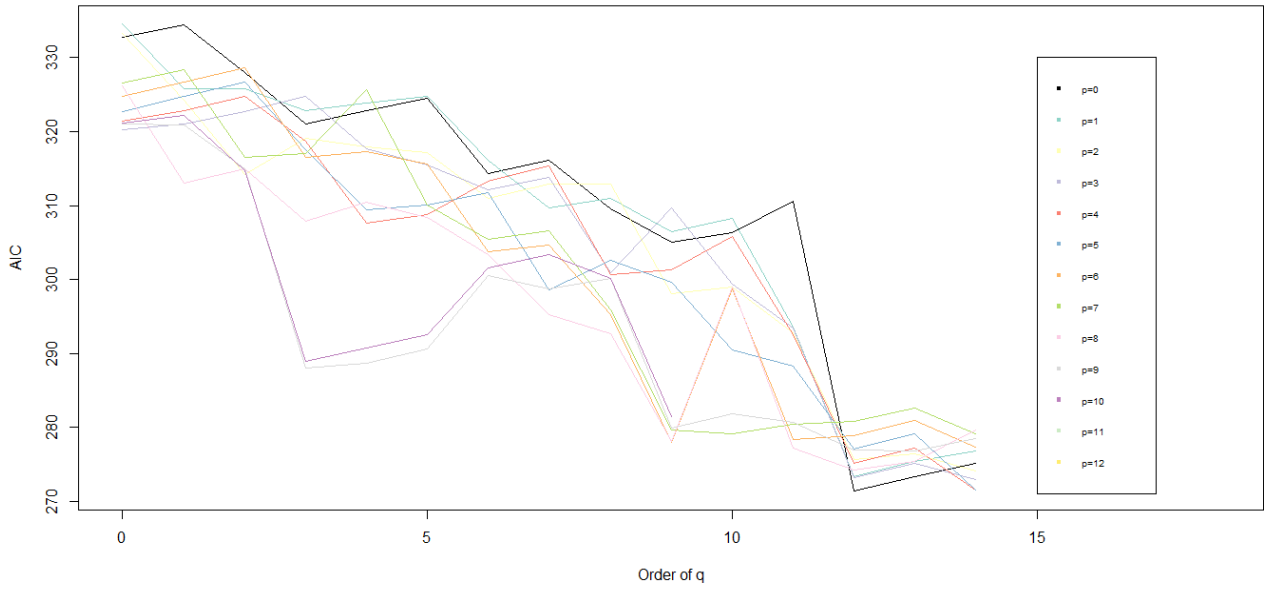


Figure 4-8: AIC as a function of the orders p and q

Figure 4.8 shows that the minimization of the AIC is achieved for p=0 and q=12, and therefore a SARIMA(0,1,12)x(1,1,0)12 will be fitted. The estimated parameters for this model are presented in table 4.8.

Table 4.8:Parameters estimates with their standard errors for the SARIMA(0,1,12)x(1,1,0)12 model

Parameter	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
Estimate	-0.0132	-0.1127	-0.0040	-0.0232	0.0097	-0.0080	-0.0207
Std.Error	0.3112	0.3160	0.2921	0.2842	0.2794	0.2901	0.2742

Parameter	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	Φ_1	σ^2
Estimate	0.0430	0.0660	0.2401	0.1995	-0.8937	-0.3633	0.322
Std. Error	0.2802	0.2991	0.2748	0.3457	0.2934	0.0970	-

The model diagnostics for this model are shown in table 4.9 and figure 4.9.

Table 4.9:Box-Ljung and Jarque-Bera tests for the assumptions of normality and independence of the residuals of the estimated SARIMA(0,1,12)x(1,1,0)12 model.

Test	Statistic	p-value
Box-Ljung	0.33056	0.5653
Jarque-Bera	7.439	0.0305

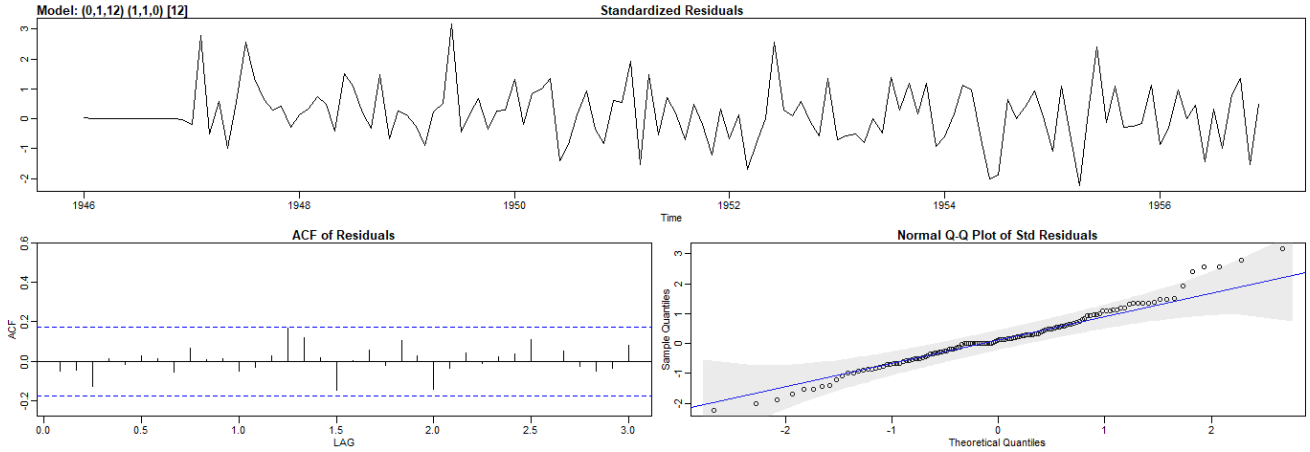


Figure 4-9: Plot of the standardized residuals (top), ACF plot of the residuals (bottom left) and QQplot of the standardized residuals (bottom right).

From the results of table 4.9 and the figure 4.9 it is logical to assume that the assumptions of homoscedastic and uncorrelated residuals cannot be rejected for the SARIMA(0,1,12)x(1,1,0)12 model but on the other hand the standardized residuals do not seem to satisfy the normality assumption.

The final step is to make forecasts of future values based on the various fitted models. The forecast() function from the forecast package in R calculates the predicted values of monthly number of births in New York city from January 1957 to December 1959. The graphical results are shown in figure 4.10.

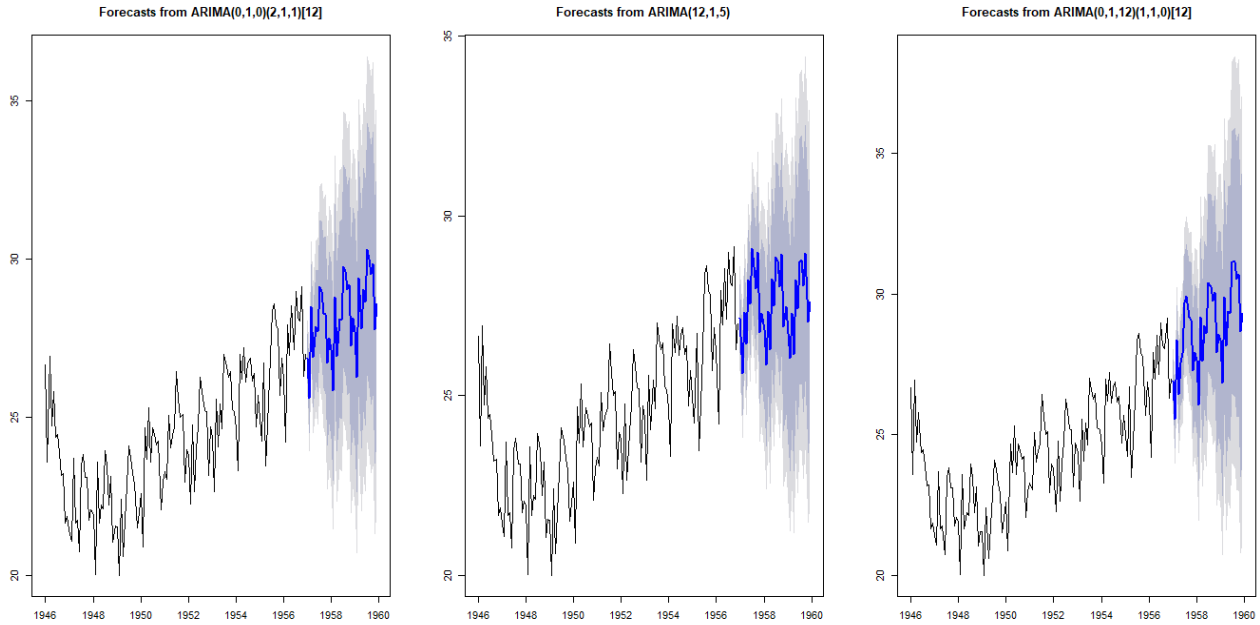


Figure 4-10: Forecasts based on the SARIMA(0,1,0)x(2,1,1)12 model (left), the ARIMA(12,1,5) model (middle) and the SARIMA(0,1,12)x(1,1,0)12 model (right). The light shaded area is the 95%CI of the forecasts while the narrowest dark shaded area is the 80%CI of the forecasts.

4.1.3 The Kalman filtered DLM

Suppose that we want to describe the seriesby a local level DLM. The DLM package in R provides all the necessary tools for parameter estimation, Kalman filtering, Kalman smoothing and forecasting.

Firstly, after giving some initial values the dlmMLE function returns the MLE estimates of the initial state estimates V, σ_w^2 of the local level component and V_s of the seasonal component. The next steps is to apply the Kalman filter for the estimated model, via the dlmFilter function, for the training set of observations from Jan1946 to Dec1956 and then the Kalman smoothing using the dlmSmooth function. Finally, we calculate forecasts for the next 3 years from Jan1957 to Dec1959 with the dlmForecast function. The graphical results are shown in figure 4.11.

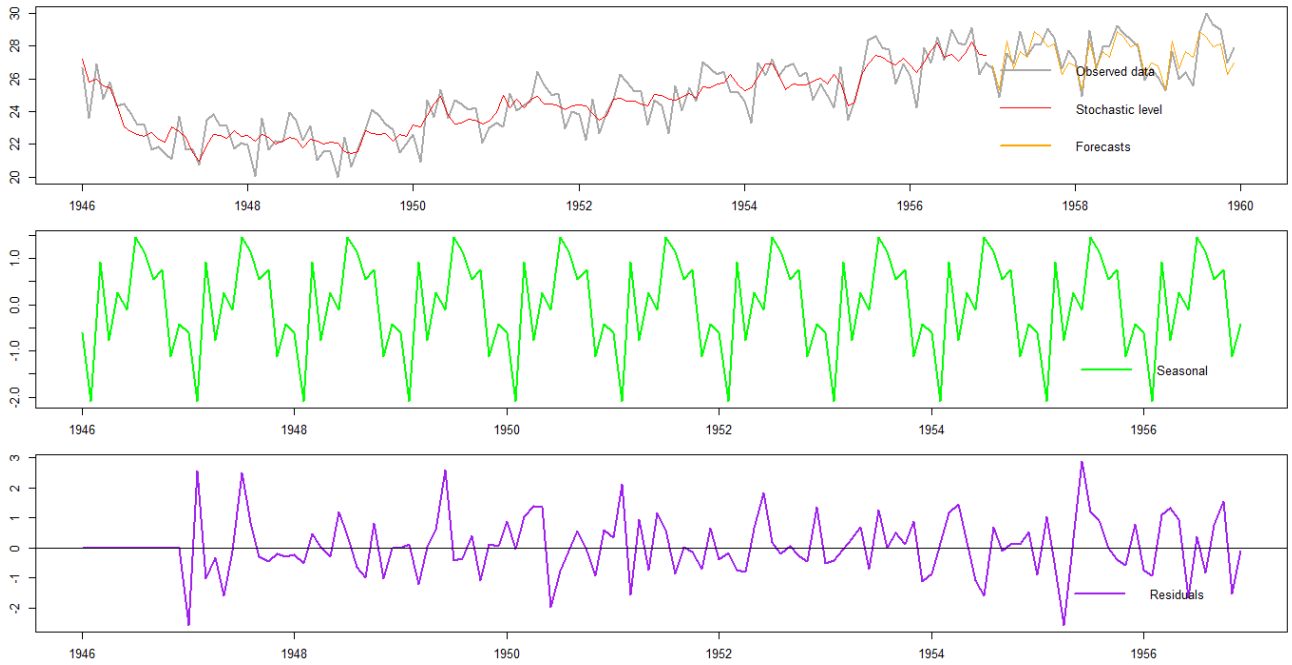


Figure 4-11: Stochastic level (red line), training set (grey line) and forecasted values (orange line) in the top panel, the seasonal component (green line) in the middle panel and the residuals(purple line) in the bottom panel.

As we see from figure 4.11 the stochastic level of the dlm follows the pattern of the training set values. The forecasts seem to mimic the series behavior but their efficiency will be evaluated later.

As we have discussed the model fit will be examined through the assumptions about the model standardized innovations. Figure 4.12 sums up all the graphical inference from the standardized innovations.

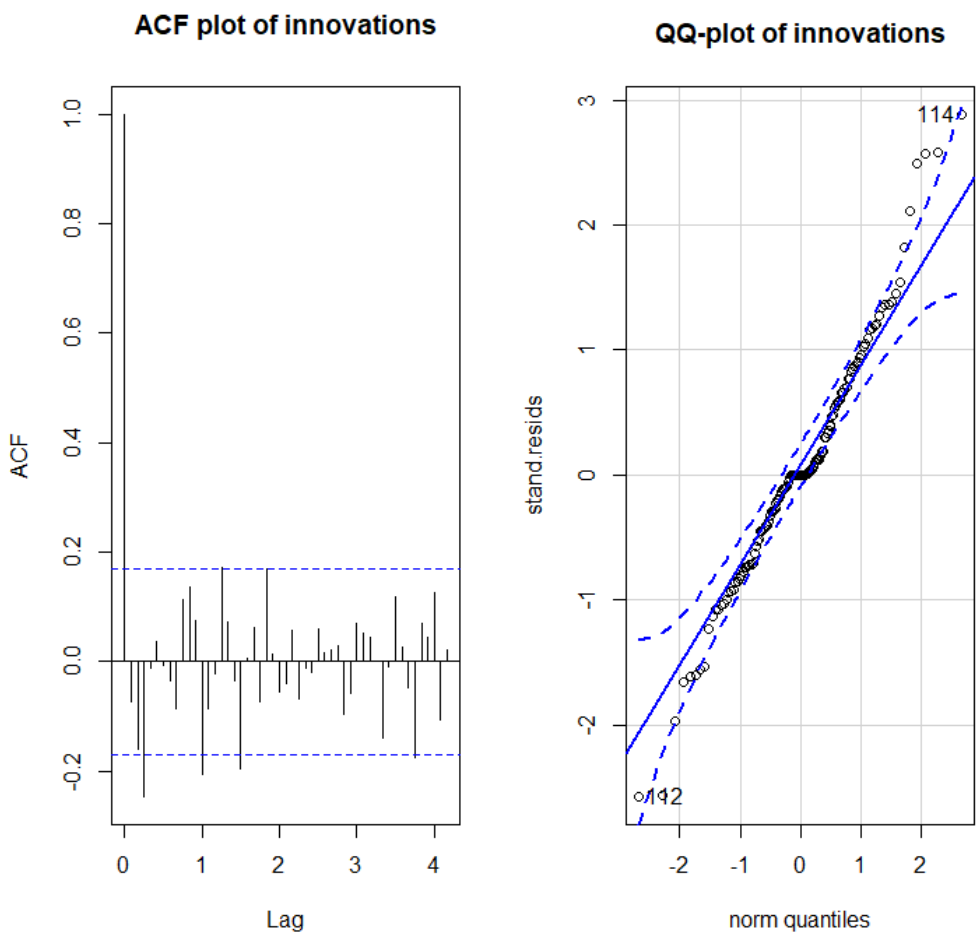


Figure 4-12: ACF plot (left) and QQ-plot (right) of the Kalman filter innovations

From figure 4.12, it is clear from the ACF plot that more than one lags are statistically significant since they surpass the 95% CI and from the QQ-plot of the innovations we observe that although most of the standardized innovations are inside the 95% CI there are at least two that break the normality assumption. Our final decision about the model assumptions are extracted by the Jarque-Bera and Box-Ljung tests the results of which are shown in table 4.10.

Table 4.10: Jarque-Bera and Box-Ljung test statistics and their correspondent p-values for the Kalman filter innovations.

Test	Statistic	p-value
Box-Ljung	0.71051	0.3993
Jarque-Bera	4.9454	0.0555

The results from table 4.10 indicate that for the Box-Ljung test on 5% level of significance we cannot reject the null hypothesis that the model innovations are uncorrelated and the Jarque-Bera test

gives a p-value slightly larger than the a=5% level of significance and thus the normality assumption cannot be rejected either.

If we are convinced that the model has passed the diagnostic tests, the final step is to evaluate its forecasting ability with respect to the real values from the test set and then it can be compared with the SARIMA model.

4.1.4 Error measures and forecasts comparison

The final part is to compare the two models forecasts with the actual values from the test sets and perform the Diebold-Mariano test to compare the two models forecasting ability. Figure 4.13 gives us a first notion of how adequately the two models can forecast new values.

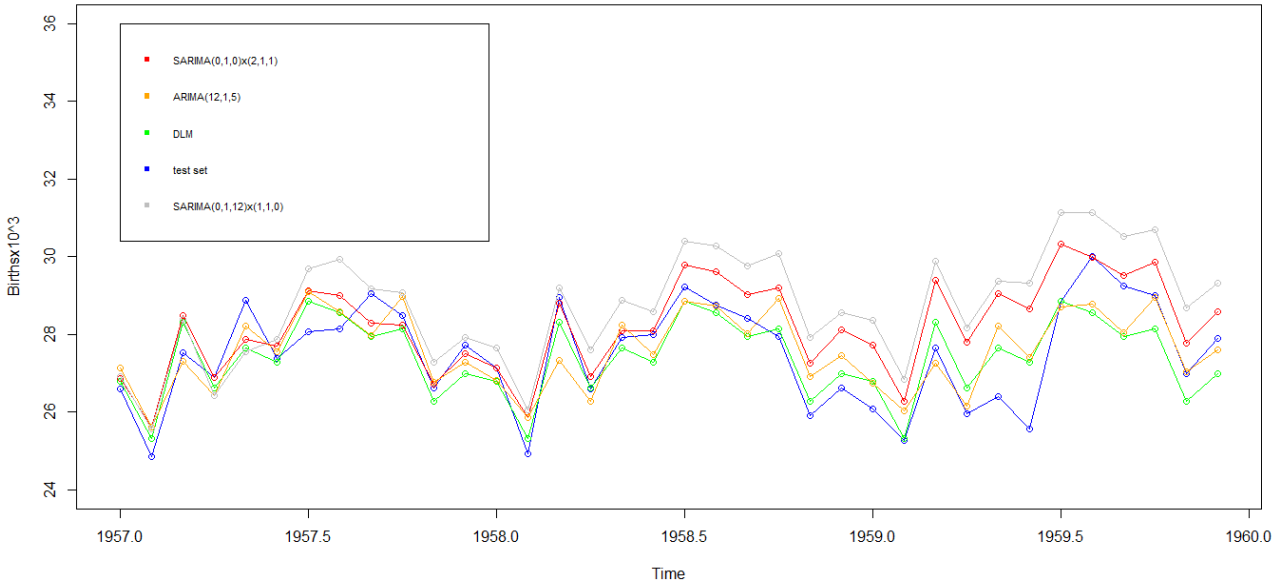


Figure 4-13: Test set values of the births series (blue line), SARIMA(0,1,0)x(2,1,1) forecasts (red line), Kalman filtered DLM, ARIMA(12,1,5) forecasts (green line) and SARIMA(0,1,12)x(1,1,0) forecasts (grey line)

Figure 4.13 shows us that both models follow the pattern of the test set but as the forecast horizon grows the distance between the model forecasts and the real time series value grows. It should be noted here that the flattening of the trend throughout the time period of the test set we observe from figure 4.2 is basically the reason why the two SARIMA models (SARIMA(0,1,0)x(2,1,1) and SARIMA(0,1,12)x(1,1,0)) always over-estimate the number of births. This means that these particular models predict new values under the assumption that there is a slight upward trend which evolves through time while the ARIMA(12,1,5) and the Kalman filtered DLM do not suppose any kind of trend.

The usual error measures and the DM test between the two models are calculated in tables 4.11, and 4.12 respectively.

Table 4.11: MSE, MAD and MAPE error measures for the various models

	MSE	MAD	MAPE
SARIMA(0,1,0)x(2,1,1)	1.2212	0.8427	0.03122
Kalman filtered DLM	0.5278	0.5945	0.02153
ARIMA(12,1,5)	0.6276	0.6317	0.02314
SARIMA(0,1,12)x(1,1,0)	2.4559	1.3409	0.04925

Table 4.12.: Diebold-Mariano test for the various models

Null hypothesis	p-value
The Kalman filtered DLM performs equally well as the SARIMA(0,1,0)x(2,1,1) model	0.01895
The Kalman filtered DLM performs better than the SARIMA(0,1,0)x(2,1,1) model	0.9905
The Kalman filtered DLM performs equally well as the SARIMA(0,1,12)x(1,1,0) model	<0.001
The Kalman filtered DLM performs better than the SARIMA(0,1,12)x(1,1,0) model	1
The Kalman filtered DLM performs equally well as the ARIMA(12,1,5) model	1

In table 4.11 all three error measures MSE, MAD and MAPE give their smaller value for the forecasts derived from the Kalman filtered DLM. Although it should be noted that the ARIMA(12,1,5) model gives almost equally good results as the Kalman filtered DLM for both three measures. Furthermore, the results from table 4.12 indicate that the Diebold-Mariano suggests that the Kalman filtered model performs better than the SARIMA(0,1,0)x(2,1,1) and SARIMA(0,1,12)x(1,1,0) models but it produces equally good forecasts as the ARIMA(12,1,5) model. So, we have strong evidence that the most suitable models to make forecasts for future values are the Kalman filtered DLM and the ARIMA(12,1,5) which perform almost equally well although the Kalman filtered DLM gives slightly better error measures.

As an alternative version of evaluating the 4 models predictive ability we can rerun our analysis as follows:1)we keep the same 4 model types, SARIMA(0,1,0)x(2,1,1), ARIMA(12,1,5), SARIMA(0,1,12)x(1,1,0) and Kalman filtered DLM 2) we run time series cross-validation with a sliding window of 132 values (11 years), which has the same length as the previous training set used, and we calculate forecasts for various h-steps ahead, for the four different competitive models 3) we construct the boxplots of the squared forecasts errors for h=6,12,24 steps ahead.

The results are shown in figure 4.14.

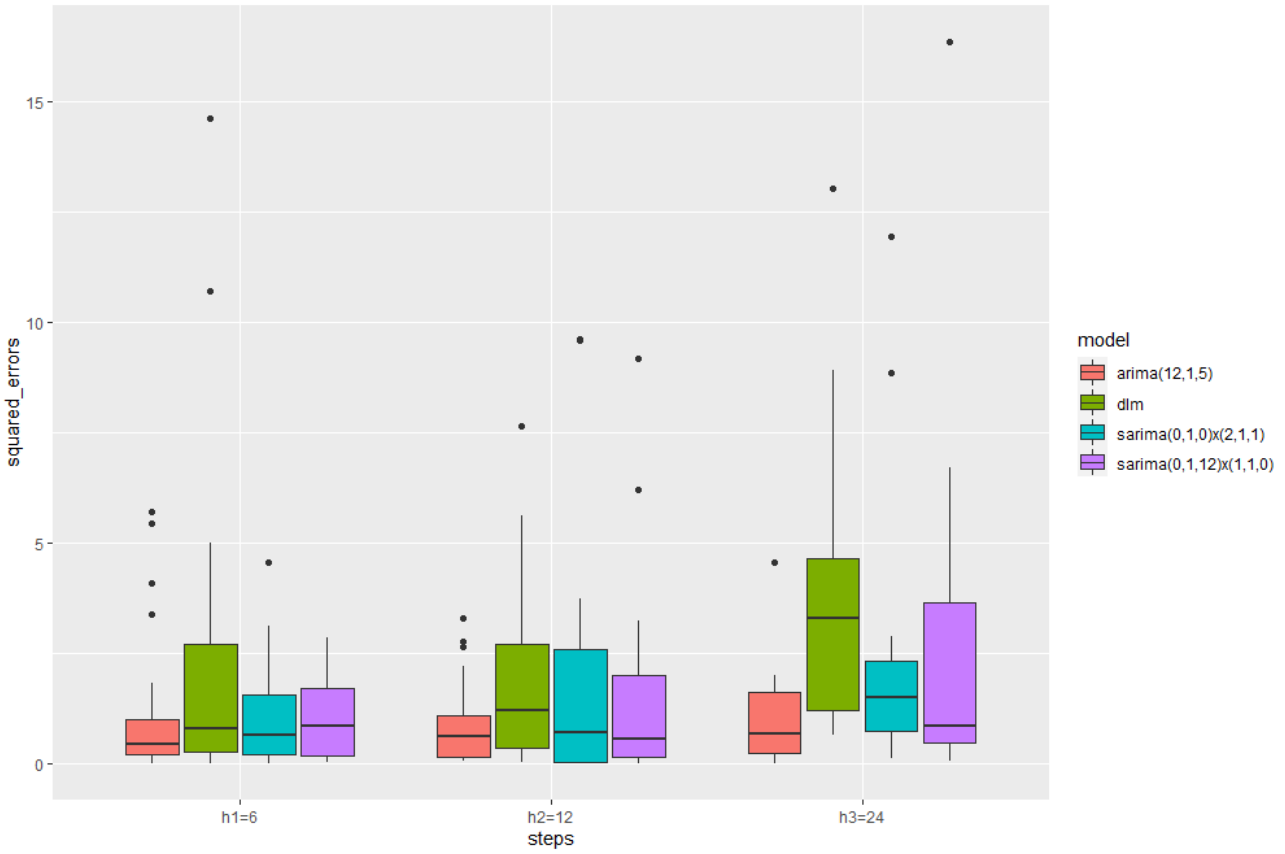


Figure 4-14: Boxplots of squared forecasts errors for h=6,12,24 steps ahead for the four different models: SARIMA(0,1,0)x(2,1,1)12 (blue), ARIMA(12,1,5) (red), Kalman filtered dlm model (green), SARIMA(0,1,12)x(1,1,0) model (purple)

Figure 4.14 reveals that, although the boxplots of the squared forecasts errors overlap for the different forecast horizons, the ARIMA(12,1,5) model has the lowest medium line of squared errors and the narrowest boxplot for every h=6,12,24. This is a strong evidence of a better forecasting ability regardless of the forecast horizon.

4.2 US monthly retail automobiles sales

For our second example, a dataset from monthly retail automobiles sales in United States is used. The dataset includes 341 entries of retail automobiles sales in billions of dollars from January 1970 to May 1998.

Firstly, the most suitable SARIMA(p,d,q)x(P,D,Q)s will be extracted and it will be evaluated by various graphical tools and tests. The model estimation will be performed on a training set of automobiles sales from January 1970 to May 1993. A test set of 5 years from June 1993 to May 1998 will be used to evaluate its forecast ability.

For the second model, the same training set for estimation and Kalman filtering, Kalman smoothing and forecasting of a Linear Growth state space model with a monthly seasonal component will be used. Forecasts are found on the same test set as before from June 1993 to May 1998.

Finally, for the last part of the analysis assuming that both models have passed the diagnostic checks their forecast performance will be compared with some frequently used error measures and tests.

4.2.1 Description of the data

Firstly, we give a graphical representation of the whole dataset in figure 4.15 and calculate some basic descriptive measures for our data in table 4.13:

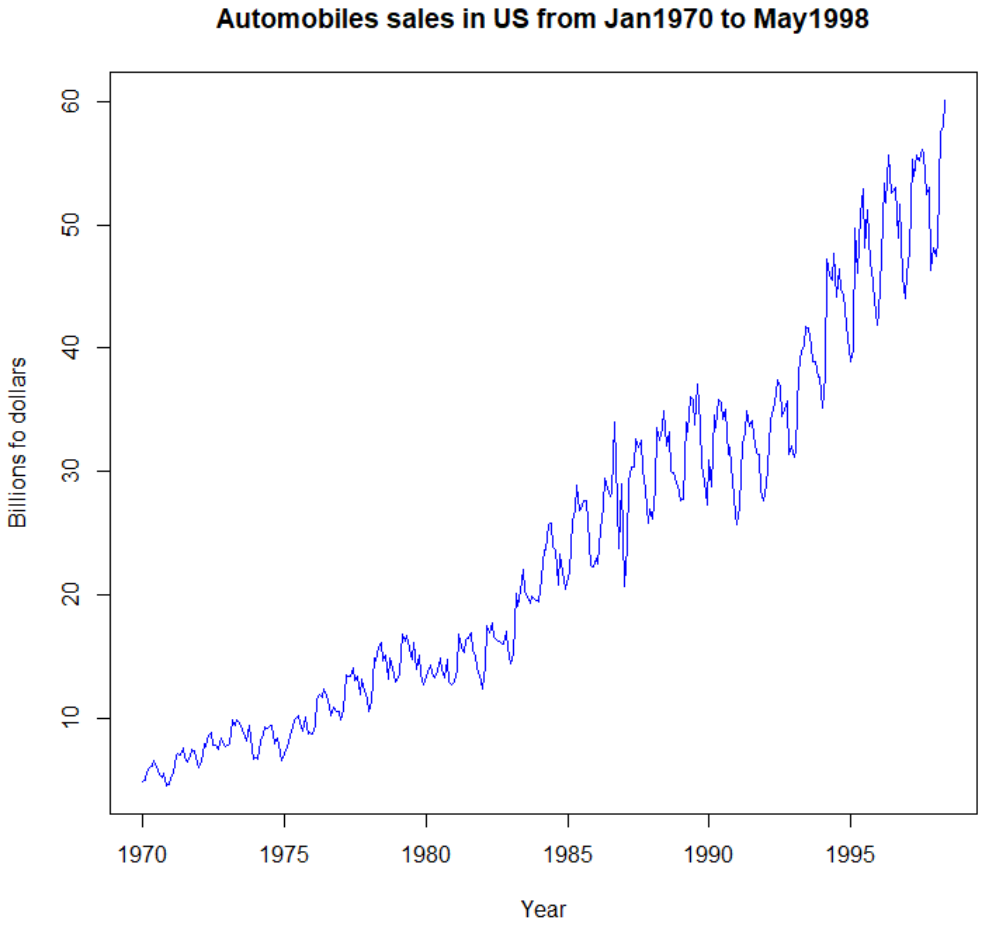


Figure 4-15: Monthly retail automobiles sales in US from January 1970 to May 1998

Table 4.13.: Descriptive statistics for US retail automobiles sales in billions of dollars

N	Mean	Median	Maximum	Minimum	Std	Skewness	Kurtosis
341	23.89	20.73	60.12	4.49	14.37	0.58	-0.73

From figure 4.15, we can see that there is certainly some seasonal variation in the automobiles sales per month; there is a peak every summer, and a trough every winter. A possible problem that we may face is that the series doesn't seem to have constant volatily over time. Thus a transformation of the data maybe needed e.g taking the natural logarithm of the series.Finally a clear upward trend is present in the series. The time series components are shown in figure 4.16.

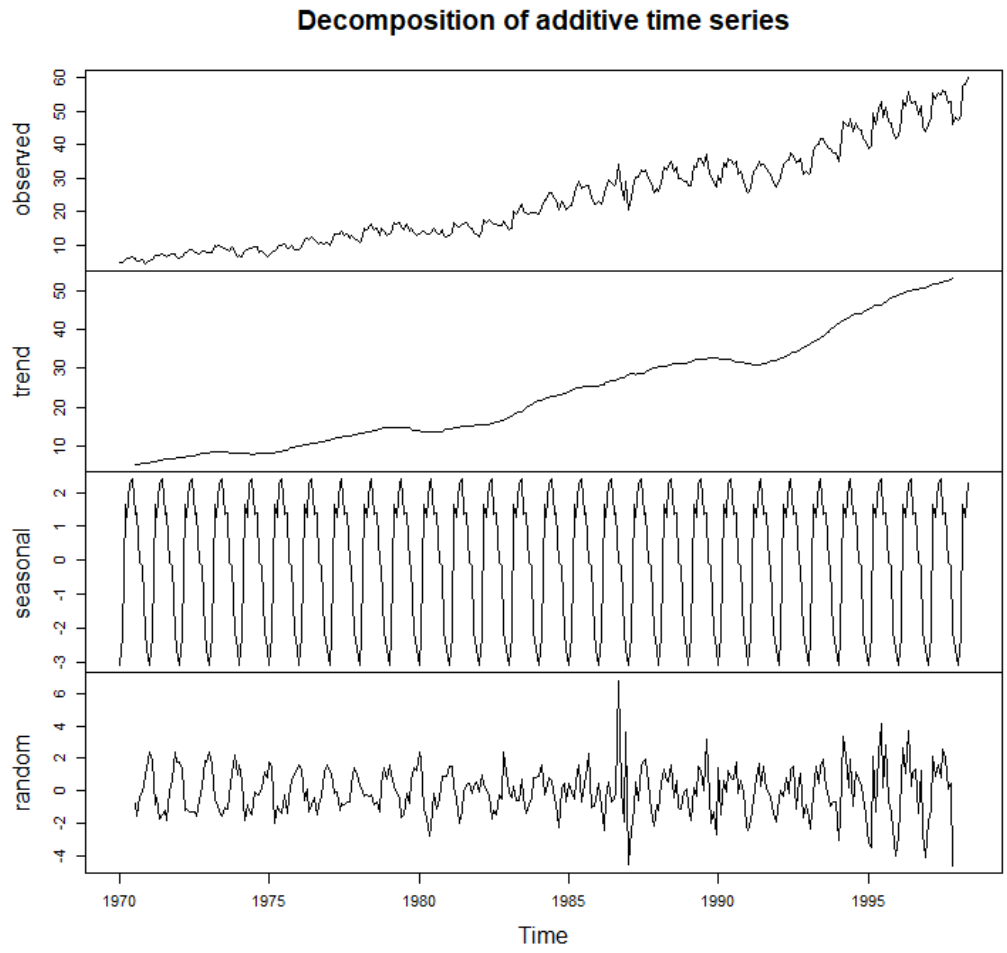


Figure 4-16: From top to bottom: a) The original sales series b) The trend component c) The seasonal component and d) The random component

As we have already mentioned, figure 4.16 indicates a clear linear upward trend, a strong seasonal component and a random component that presents an increment in its variance approximately after 1987.

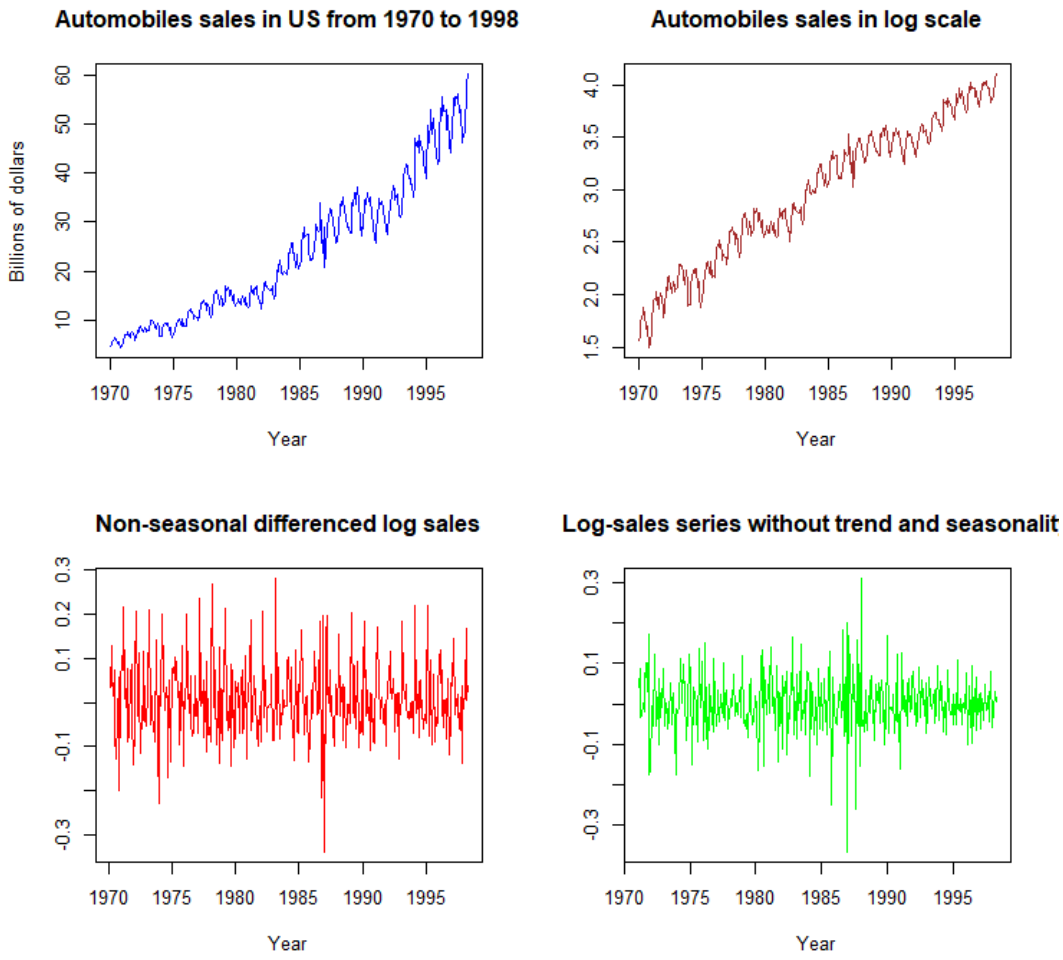


Figure 4-17: Time series plots of: a) The original sales series (top left) b) The log transformed sales series (top right) c) The non-seasonal differenced sales series (bottom left) and d) The seasonal and non-seasonal log sales series (bottom right)

From Figure 4.17, it is clear that once we have taken the log transform on the data the variance seems to be stabilized. After taking non-seasonal differences the trend has disappeared but there is still a seasonal pattern that is eliminated when we take seasonal and non-seasonal differences. As we can see in the last panel of figure 4.17 the transformed series seems to have lost any trend and seasonal pattern and can be considered as a stationary process.

4.2.2 ARIMA and SARIMA models

After the preliminary analysis we are now ready to fit the most suitable SARIMA model for the sales series. The time period from January 1970 to May 1993 are used as a training set on which the model fit, parameters estimation and diagnostic tests will be carried out and the last 5 years of the sales dataset from June 1993 to May 1998 will be utilized as a test set for the forecasting evaluation of the model.

The first step is to estimate the model parameters of integration and seasonal integration d, D respectively. From section we have already graphical evidence that d=D=1 but in order to substantiate our assumptions we conduct the ADF test for unit roots and the HEGY test for seasonal unit roots the results of which are shown in tables 4.14 and 4.15 respectively.

*Table 4.14.**:** ADF test statistics and p-values for the log-sales series and first differenced log-sales series. The null hypothesis is that the series has a unit root and the lag length used is 15.*

Series	ADF statistic	p-value
Log-sales	-2.6049	0.3212
Diff(log-sales)	-3.8216	0.01836

*Table 4.15.**:**HEGY test with test-statistics for the log-sales series and the seasonal differenced log-sales series. The p-value is found in the parentheses*

Statistic	Diff(log-sales)	Diff(Diff(log-sales),12)
0	-6.2927 (<0.001)	-5.6634 (0.0011)
π/6	4.8938 (0.0083)	43.4935 (<0.001)
π/3	6.6688 (0.0015)	40.5491 (<0.001)
π/2	3.0797 (0.0496)	40.3158 (<0.001)
2π/3	5.3585 (0.0053)	47.563 (<0.001)
5π/6	11.8332 (<0.001)	31.9685 (<0.001)
π	-3.7795 (<0.001)	-6.1858 (<0.001)

The results from table 4.14 indicate that for the log-sales series the ADF test with the null hypothesis of a unit root is not rejected on the five percent significance level but for the differenced log-sales series we have strong evidence in order to reject the null hypothesis and thus we can conclude that the level of integration d=1.

From table 4.15 the seasonal unit root hypothesis is rejected for all frequencies for the differenced log-sales series (d=1, D=0) but as we can see the p-value for the F3,4 statistic is very close to the 5% level of significance and so it is recommended that the test be performed again on both first and seasonally differenced log-sales series(d=1, D=1). The result is that the seasonal unit root hypothesis is rejected on all frequencies on the five percent significance level. Thus, seasonally differencing first differenced log-sales series seem to have made the time series stationary enough. Hence a seasonal difference operator of order D=1 seems appropriate.

The orders of integration that seem to be needed to make each series stationary enough have now been decided. The procedure of identifying the SARIMA can continue to the next step which is to determine the autoregressive and moving average orders.

The auto.arima function of the forecast package will be used for the specification of the orders p,P,q,Q and the selection of the most suitable SARIMA model based on the information criteria AIC, AICc, and BIC. The model selected from each of the information criteria is shown in table 4.16.

Table 4.16.:The most appropriate SARIMA model based on different information criteria

Informaton criterion	AIC	BIC	AICc
Model selected	SARIMA(1,1,2)x(2,1,1)12	SARIMA(0,1,1)x(2,1,1)12	SARIMA(1,1,2)x(2,1,1)12

Table 4.16 indicates that both AIC and AICc suggest that a SARIMA(1,1,2)x(2,1,1)12 should be fitted to the data while BIC uses a different more parsimonious SARIMA(0,1,1)x(2,1,1)12 model. The selection now is based on what we want to achieve; generally the estimation of complicated model is time consuming compared to a simpler model but on the other hand it may give us better forecasting results. Thus we conclude that a SARIMA(1,1,2)x(2,1,1)12 will be fitted to the log-sales series. The estimated parameters with their standard errors are shown in table 4.17.

Table 4.17: Estimated model parameters of the SARIMA(1,1,2)x(2,1,1)12 model with their standard errors in parentheses

Parameter	Estimate	Standard error
φ_1	-0.4468	0.2495
θ_1	0.0435	0.2373
θ_2	-0.3850	0.1016
Φ_1	0.1671	0.0744
Φ_2	-0.1410	0.0705
Θ_1	-0.8809	0.0565
σ^2	0.003488	-

The next step is to examine the usual model diagnostics in order to determine the model suitability. The model residuals are tested for their normality and independence via the Jarque-Bera test and the Ljung-Box test, the results of which are shown in table 4.18.

Table 4.18.: Ljung-Box and Jarque-Bera tests for the assumptions of normality and independence of the residuals of the estimated SARIMA(1,1,2)x(2,1,1)12 model.

Test	Statistic	p-value
Box-Ljung	0.00025468	0.9873
Jarque-Bera	89.9	<0.001

From table 4.18 we can see that for the Box-Ljung test the null hypothesis of uncorrelated residuals cannot be rejected. On the other hand, the p-value of the Jarque-Bera leads to the conclusion that the residuals cannot be normally distributed.

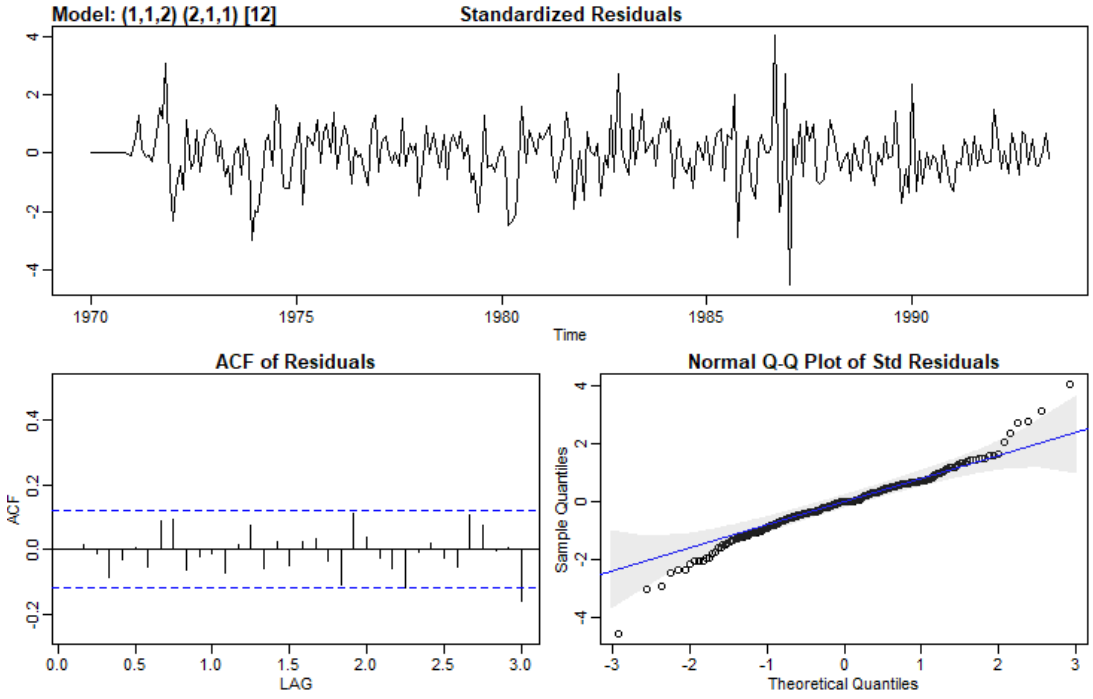


Figure 4-18: Plot of the standardized residuals (top), ACF plot of the residuals (bottom left) and QQplot of the standardized residuals (bottom right).

Figure 4.18 shows that indeed the model residuals do not appear any significant serial autocorrelation but on the other hand from the QQ-plot of the standardized residuals there are plenty of points that are not inside the 95% CI bounds. So the assumption of the normally distributed residuals cannot be accepted.

As we did in the first application it would be interesting to find the best non-seasonal ARIMA model based on the minimization of the AIC. The parameter search for p and q will be performed in the grid $p \in [0,12], q \in [0,14]$.

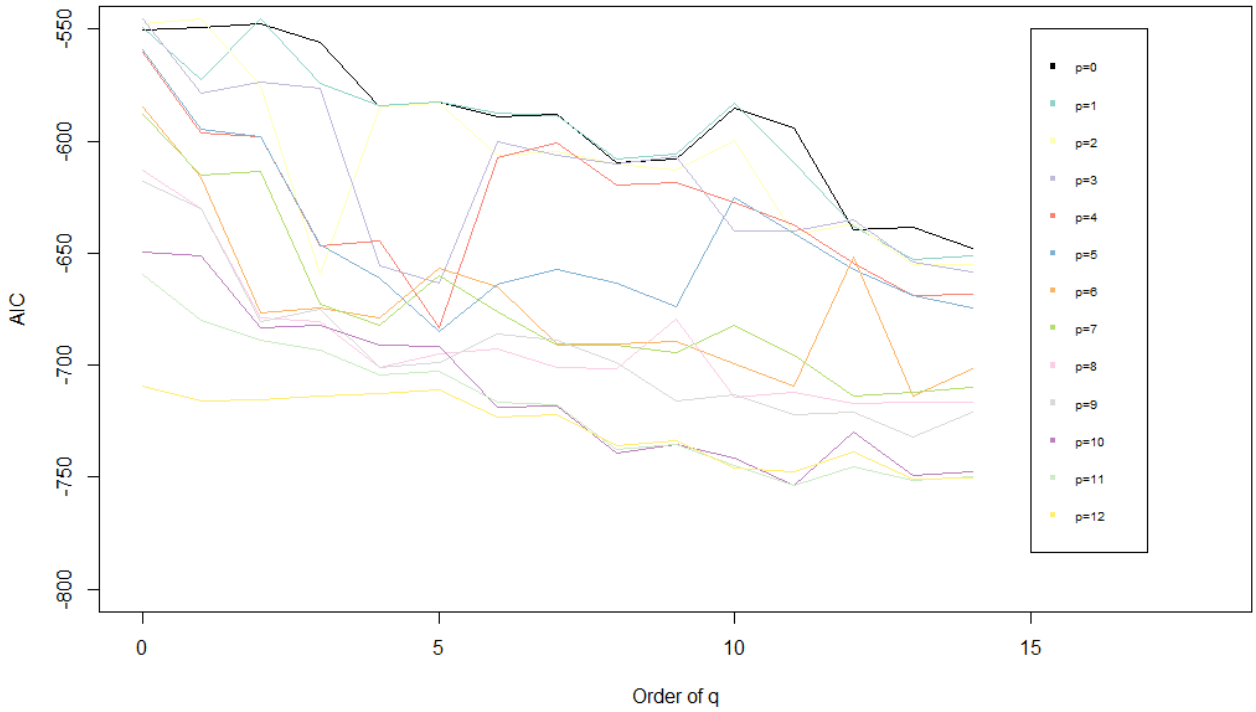


Figure 4-19: AIC as a function of orders p and q

From figure 4.19 it is clear that the AIC is almost always smaller in value when the order of p=12. Moreover, for p=12 the AIC as a function of q does not appear any significant changes after a local minimum for q=6. So, it is reasonable to assume that the best ARIMA model that minimizes the AIC is the one with parameters p=12 and q=6.

The estimated parameters of the ARIMA(12,1,5) model with their standard errors are presented in table 4.19.

Table 4.19.: Parameters estimates with their standard errors for the ARIMA(12,1,6) model

Parameter	**φ_1**	**φ_2**	**φ_3**	**φ_4**	**φ_5**	**φ_6**	**φ_7**
Estimate	-0.0677	-0.1024	-0.0603	-0.0249	-0.4899	-0.1240	-0.1479
Std.Error	0.1007	0.1230	0.0933	0.0867	0.0858	0.1361	0.0766

Parameter	**φ_8**	**φ_9**	**φ_{10}**	**φ_{11}**	**φ_{12}**	**θ_1**	**θ_2**
Estimate	-0.1479	-0.0590	-0.2065	0.0110	0.3774	-0.2640	-0.1001
Std.Error	0.0784	0.0697	0.0688	0.0666	0.0792	0.0981	0.1356

Parameter	θ_3	θ_4	θ_5	θ_6	σ^2
Estimate	0.0379	0.0778	0.1003	0.5076	0.003897
Std.Error	0.0962	0.1239	0.0872	0.1643	-

The model diagnostics for this model are shown in table 4.20 and figure 4.20.

Table 4.20.:Box-Ljung and Jarque-Bera tests for the assumptions of normality and independence of the residuals of the estimated ARIMA(12,1,6) model.

Test	Statistic	p-value
Box-Ljung	55.615	<0.001
Jarque-Bera	0.22192	0.6376

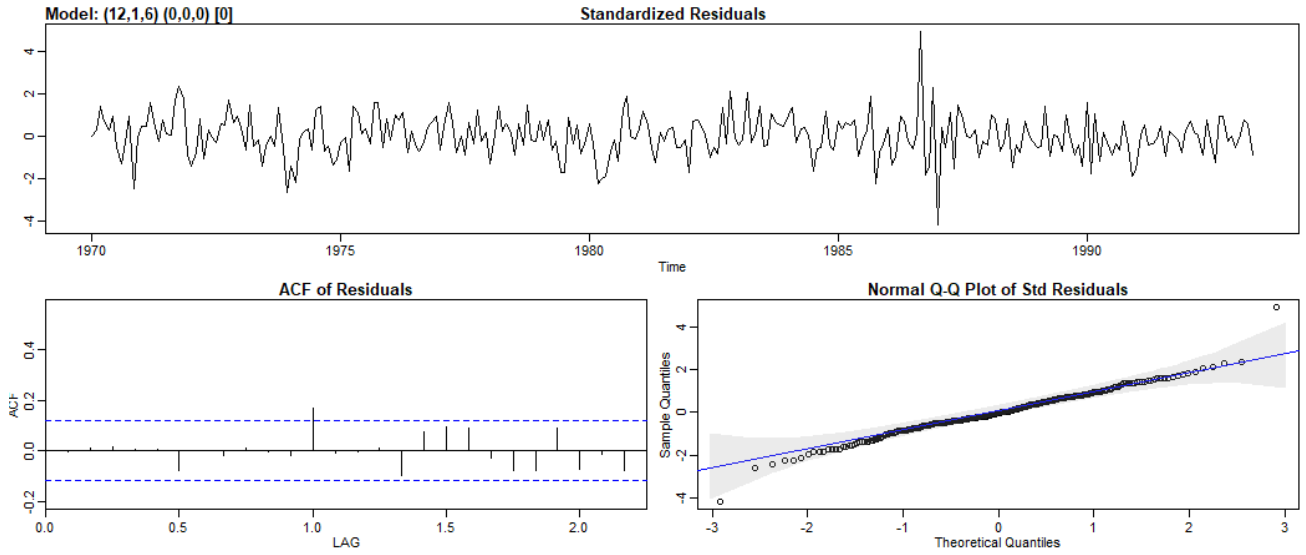


Figure 4-20: Plot of the standardized residuals (top), ACF plot of the residuals (bottom left) and QQplot of the standardized residuals (bottom right).

From the results of table 4.20 and the figure 4.20 it is logical to assume that the assumptions of homoscedastic and uncorrelated residuals can be accepted for the ARIMA(12,1,6) model but the Jarque-Bera test and the QQplot of the standardized residuals indicate that we cannot accept the assumption of normally distributed residuals.

Finally, following the same reasoning as in the previous application we are going to fit a SARIMA model with lighter restrictions for the p and q parameters and the orders P=1 and Q=0. The most suitable model will be extracted by the minimization of the AIC criterion as a function of p and q.

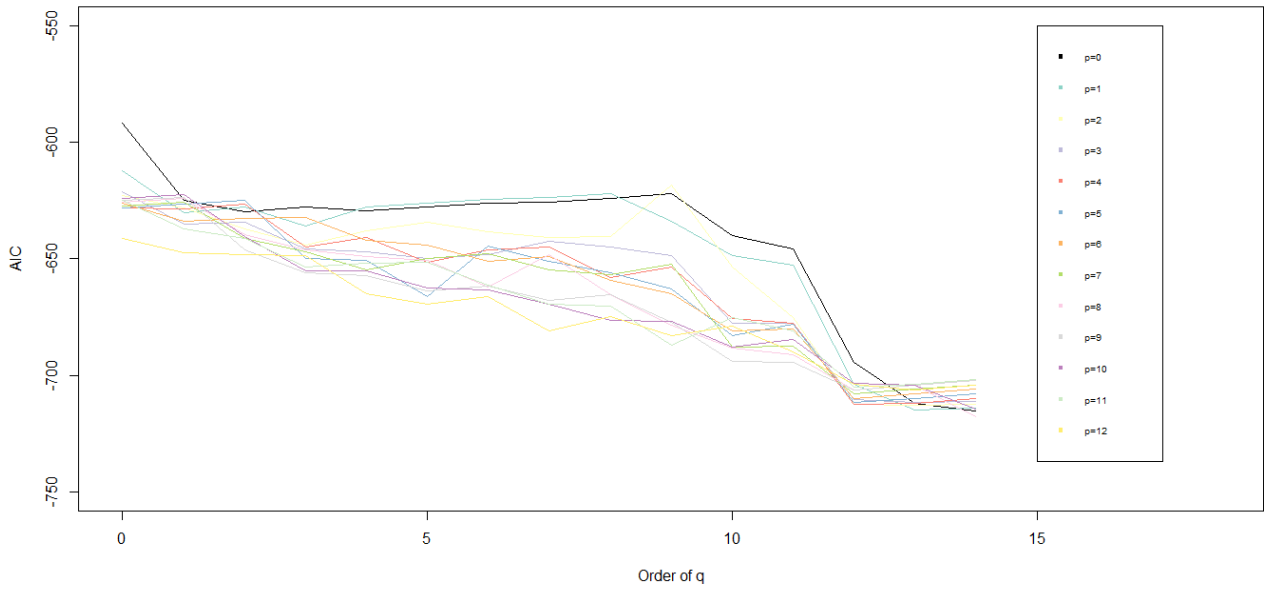


Figure 4-21: AIC as a function of orders p and q

Figure 4.21 does not show a clear "winner" among all the possible models since all of the models appear to have a sudden drop of the AIC value for q=12 and for larger q's they almost coincide. For this reason, the simplest model will be chosen i.e the model with p=0, q=12. Therefore a SARIMA(0,1,12)x(1,1,0)12 will be fitted. The estimated parameters for this model are presented in table 4.21.

Table 4.21.:Parameters estimates with their standard errors for the SARIMA(0,1,12)x(1,1,0)12 model

Parameter	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
Estimate	-0.1988	-0.0779	0.1147	-0.0560	-0.0330	0.0467	-0.0503
Std.Error	0.2131	0.1613	0.0704	0.1692	0.2613	0.2780	0.2130

Parameter	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	Φ_1	σ^2
Estimate	0.0682	0.0595	-0.1450	0.1079	-0.8359	0.0835	0.003617
Std. Error	0.0993	0.0949	0.1682	0.2075	0.2215	0.0680	-

The model diagnostics for this model are shown in table 4.22 and figure 4.22.

Table 4.22.Box-Ljung and Jarque-Bera tests for the assumptions of normality and independence of the residuals of the estimated SARIMA(0,1,12)x(1,1,0) model.

Test	Statistic	p-value
Box-Ljung	7.8794	0.005
Jarque-Bera	67.758	<0.001

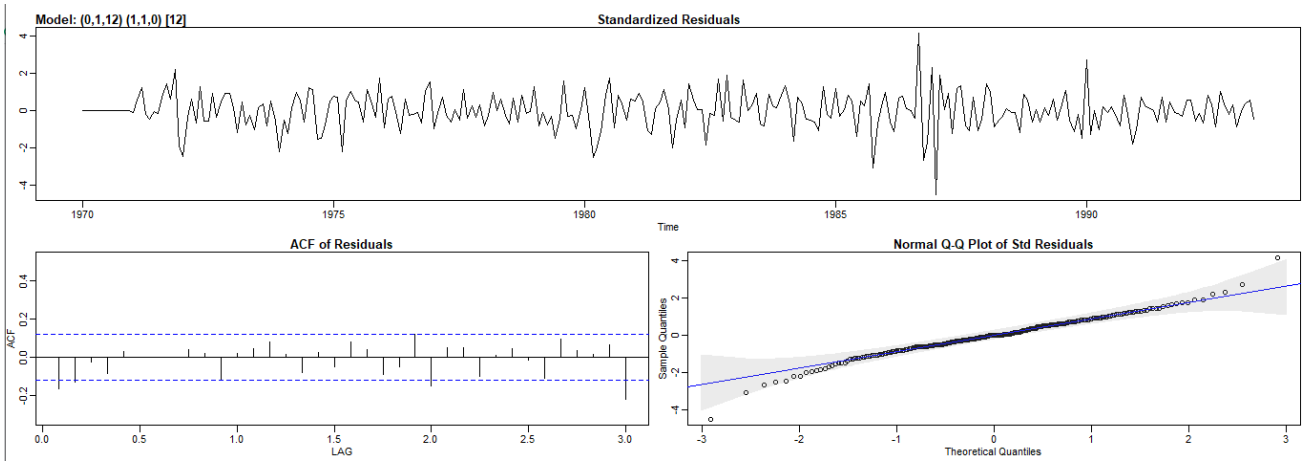


Figure 4-22: Plot of the standardized residuals (top), ACF plot of the residuals (bottom left) and QQplot of the standardized residuals (bottom right).

From the results of table 4.22 we cannot accept the assumptions of uncorrelated and normally distributed residuals although the graphical patterns of figure 4.22 do not indicate any clear violation of these assumptions

Although the residuals do not obey the necessary assumption we can use the estimated models to extract forecasts of future values. The forecast() function from the forecast package in R calculates the predicted values of monthly retail automobiles sales in logarithmic scale from June 1993 to May 1998. The graphical results are shown in figure 4.23.

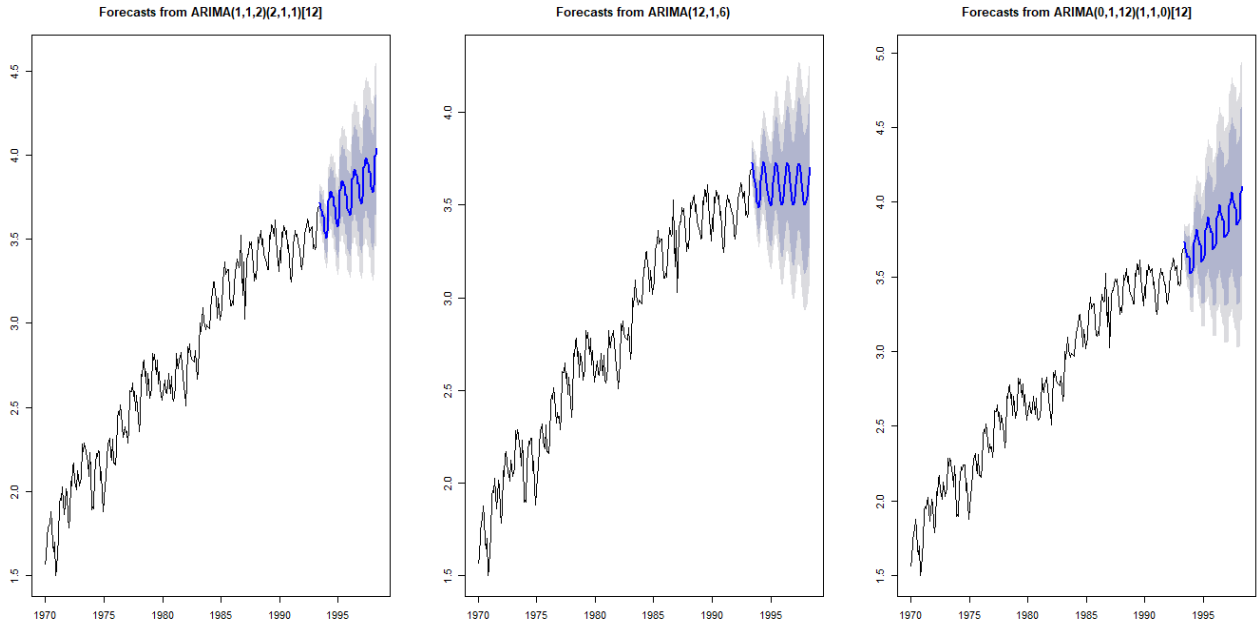


Figure 4-23:Forecasted values for the log sales series for the time period from June 1993 to May 1998based on the SARIMA(1,1,2)x(2,1,1) model (left), the ARIMA(12,1,6) model (middle) and SARIMA(0,1,12)x(1,1,0). The light shaded area is the 95%CI of the forecasts while the narrowest dark shaded area is the 80%CI of the forecasts.

4.2.3 The Kalman filtered DLM

Suppose that we want to describe the series by a linear growth DLM. The DLM package in R provides all the necessary tools for parameter estimation, Kalman filtering, Kalman smoothing and forecasting.

After following the same procedure of estimating, Kalman filtering and Kalman smoothing finally we calculate forecasts for the time period from June1993 to May1959. The graphical results are shown in figure 4.24.

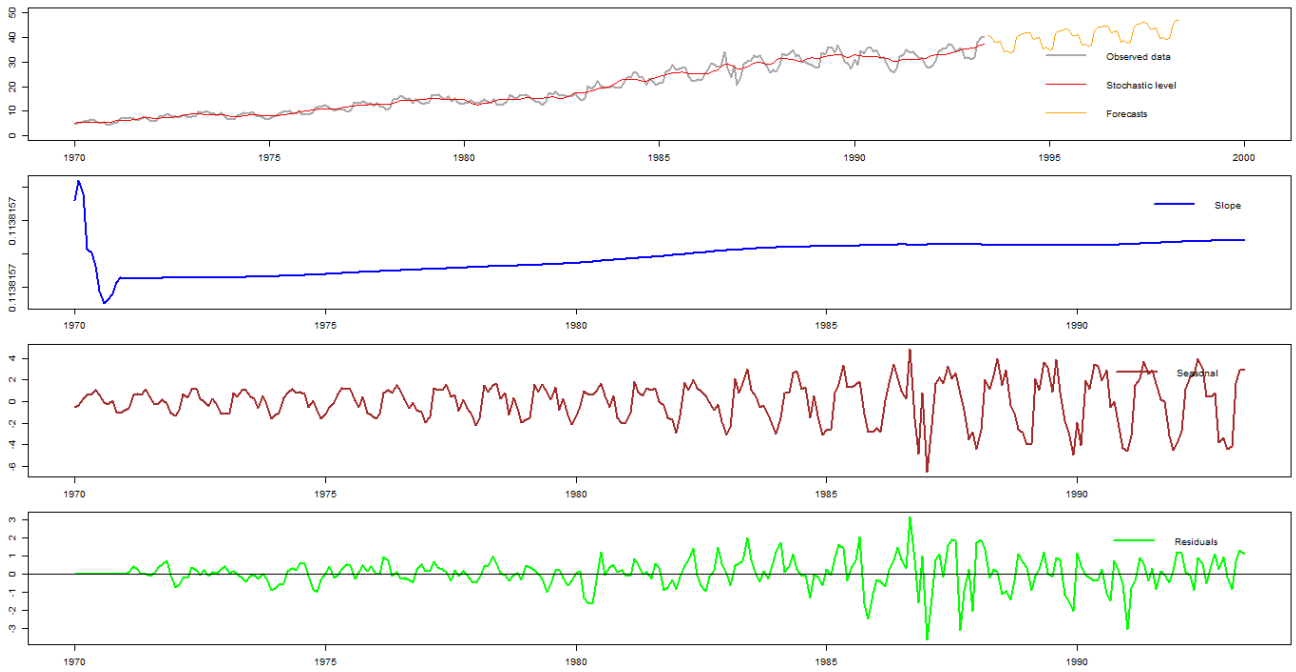


Figure 4-24: Stochastic level (red line), training set (grey line) and forecasted values (orange line) in the top panel, the trend component (blue line) in the second panel, the seasonal component (brown line) in the third panel and the residuals (green line) in the bottom panel.

As we see from figure 4.24 the forecasts seem to mimic the series behavior but their efficiency will be evaluated later.

The model fit will be examined through the assumptions about the model standardized innovations. Figure 4.25 sums up all the graphical inference from the standardized innovations.

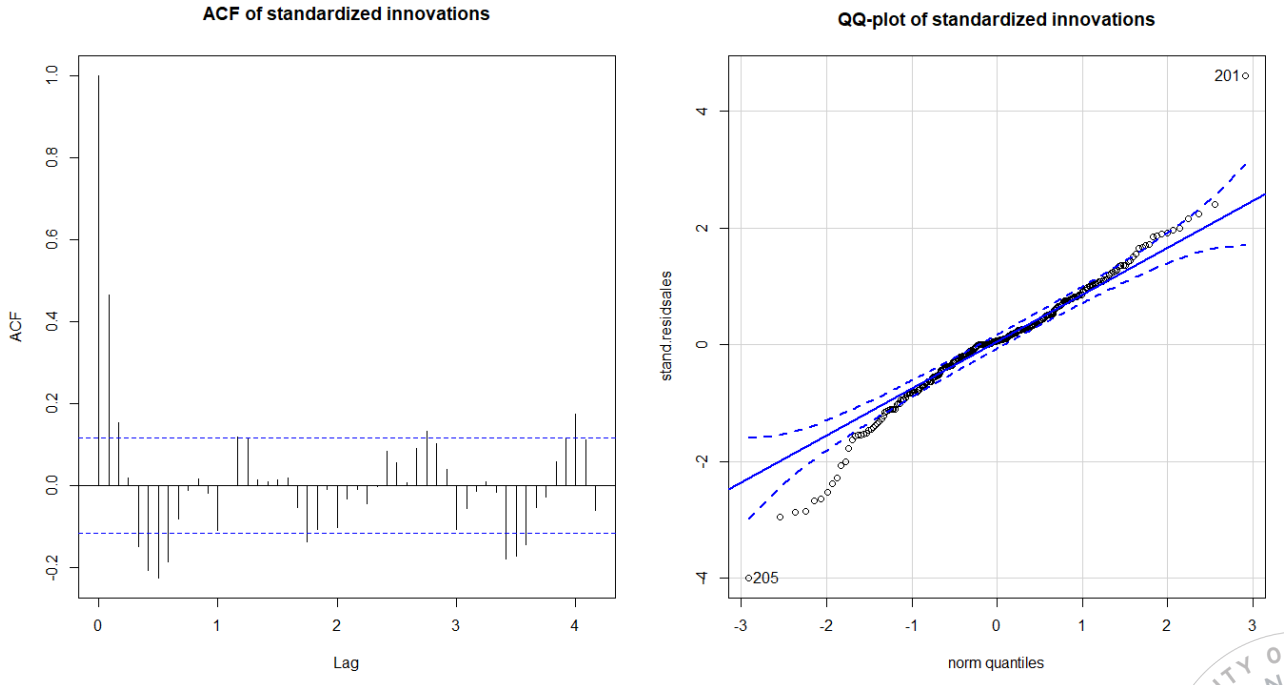


Figure 4-25: ACF of standardized innovations (left) and QQ-plot of standardized innovations (right)

Figure 4.25 indicates that none of the model assumptions is verified since there exists statistical significant serial autocorrelation in the standardized innovations and the QQ-plot shows that many points are outside the 95% CI bounds.

4.2.4 Error measures and forecasts comparison

The final part is to compare the three models forecasts with the actual values from the test sets and perform the Diebold-Mariano test to compare the two models forecasting ability. Figure 4.26 gives us a first notion of how adequately the two models can forecast new values.

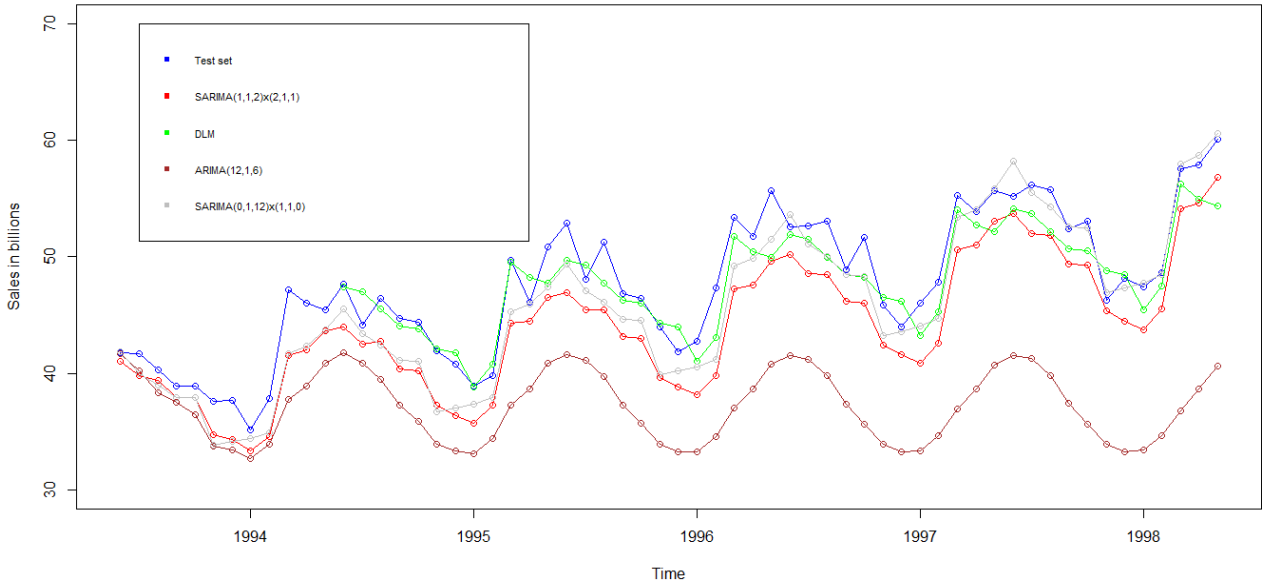


Figure 4-26: Test set values of the births series (blue line), SARIMA(1,1,2)x(2,1,1) forecasts (red line), Kalman filtered DLM forecasts (green line), ARIMA(12,1,6) forecasts (brown line) and SARIMA(0,1,12)x(1,1,0) forecasts (grey line)

Figure 4.26 shows us that the SARIMA models and the Kalman filtered DLM follow the pattern of the test set but not the ARIMA model. Moreover, although the distance between all the model forecasts and the real time series value grows as the forecast horizon grows but the most 'efficient' model seems to be the SARIMA(0,1,12)x(1,1,0). The worst model for predicting new values from figure 4.26 seems to be the ARIMA(12,1,6) model (brown line in figure 4.26) and this is mostly because this is the only model which does not take into account any trend in the time series which evolves over time.

The usual error measures and the DM test between the two models are calculated in tables 4.23 and 4.24 respectively.

Table 4.23.: MSE, MAD and MAPE error measures for the SARIMA and the Kalman filtered DLM

	MSE	MAD	MAPE
SARIMA(1,1,2)x(2,1,1)12	14.71742	3.543356	0.07422295
ARIMA(12,1,6)	131.2522	10.30852	0.2085977
Kalman filtered DLM	57.72231	6.902226	0.1399843
SARIMA(0,1,12)x(1,1,0)12	7.034042	2.13822	0.04609026

Table 4.24.: Diebold-Mariano test for the various models

Null hypothesis	p-value
The SARIMA(0,1,12)x(1,1,0) performs equally well as the Kalman filtered DLM	<0.001
The SARIMA(0,1,12)x(1,1,0) performs better than the Kalman filtered DLM	1
The SARIMA(0,1,12)x(1,1,0) performs equally well as the SARIMA(1,1,2)x(2,1,1)	<0.001
The SARIMA(0,1,12)x(1,1,0) performs better than the SARIMA(1,1,2)x(2,1,1)	1
The SARIMA(0,1,12)x(1,1,0) performs equally well as the ARIMA(12,1,6)	<0.001
The SARIMA(0,1,12)x(1,1,0) performs better than the ARIMA(12,1,6)	1

In table 4.23 all three error measures MSE, MAD and MAPE give their smaller value for the forecasts derived from the SARIMA(0,1,12)x(1,1,0) model. Also the results from table 4.24 indicate that the Diebold-Mariano suggests that the SARIMA(0,1,12)x(1,1,0) model performs better from all the other models. So, we have strong evidence that the best among the two models to make forecasts for future values is the SARIMA(0,1,12)x(1,1,0) model.

As an alternative version of evaluating the 4 models predictive ability we can rerun our analysis as follows:1)we keep the same 4 model types, SARIMA(1,1,2)x(2,1,1), ARIMA(12,1,6), SARIMA(0,1,12)x(1,1,0) and Kalman filtered DLM 2) we run time series cross-validation with a sliding window of 281 values, which has the same length as the previous training set used, and we calculate forecasts for various h-steps ahead, for the four different competitive models 3) we construct the boxplots of the squared forecasts errors for h=12,24,36,48 steps ahead.

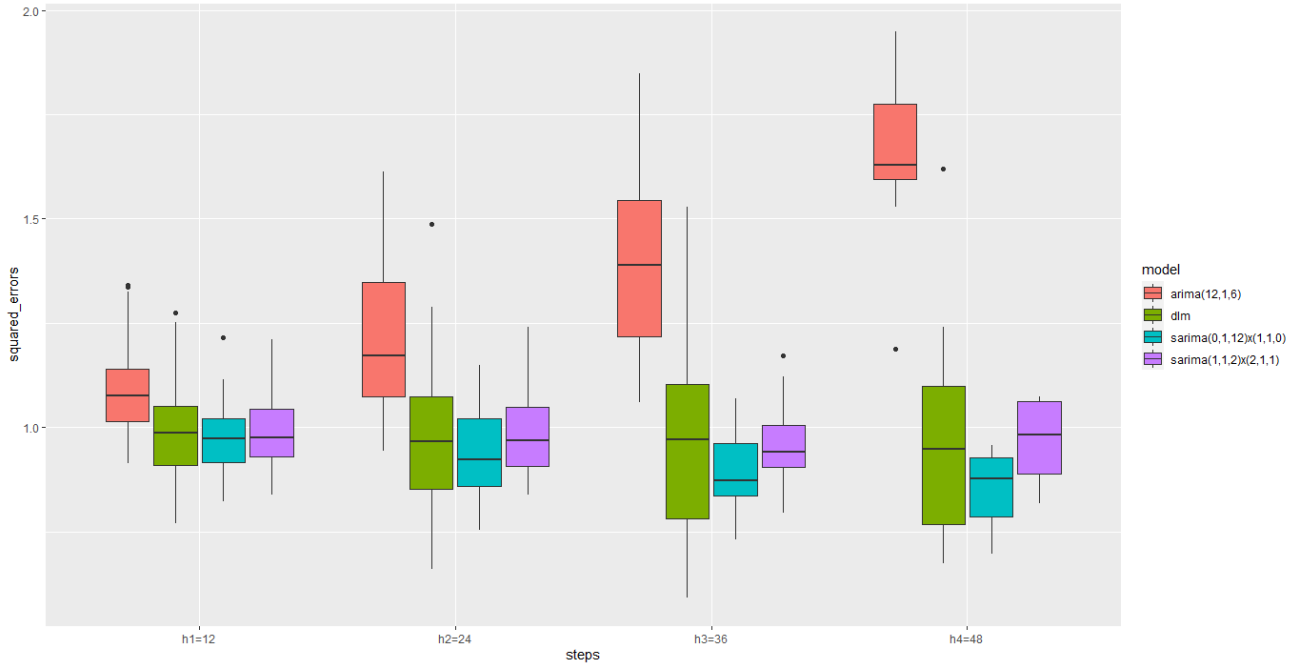


Figure 4-27: Boxplots of squared forecasts errors for h=12,24,36,48 steps ahead for the four different models: SARIMA(0,1,12)x(1,1,0)(blue), ARIMA(12,1,6) (red), Kalman filtered dlm model (green) and SARIMA(1,1,2)x(2,1,1)(purple)

Figure 4.27 reveals that apart from the ARIMA(12,1,6) model the 3 other competitive models have overlapping boxplots. Moreover, the SARIMA(0,1,12)x(1,1,0) model, especially for forecast horizons greater than 12 steps ahead, seems to have the lowest medium line of squared forecast errors among all models and the narrowest errors boxplots along with the SARIMA(1,1,2)x(2,1,1) model. This is an indicator of a better forecasting ability.

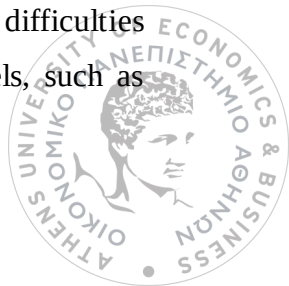
5. Conclusion

In this study ARIMA-SAIRIMA models and Kalman filtered Dynamic Linear Models have been fitted to two different real life datasets.

More specifically, for the first dataset with the monthly number of births in the city of New York, only two models, the ARIMA(12,1,5) and the SARIMA(0,1,0)x(2,1,1), pass all the model diagnostics among the four competitive models. Moreover, from the error measures calculated in the test set, the models that seem to produce the best forecasts are the ARIMA(12,1,5) and the Kalman filtered local level model. This is also confirmed by the Diebold-Mariano test for the comparison of the predictive ability of all models where the two aforementioned models seem to perform equally well and give more accurate forecasts than the other two. Furthermore, the time series cross-validation procedure for various h-steps ahead forecast horizons showed that the model which constantly gives the best forecasts with respect to the forecast errors is the ARIMA(12,1,5) model.

For the second dataset with the monthly retail automobiles sales in USA, none of the 4 candidate models passed the model diagnostics about the assumptions for the model residuals. The out of sample error measures showed that the best forecasts are produced by the SARIMA(0,1,12)x(1,1,0) model and this is also confirmed by the results of the Diebold-Mariano test which revealed that this model gives more accurate forecasts for the given test set than all the other competitive models. The time series cross-validation also confirmed that there is strong evidence to consider the SARIMA(0,1,12)x(1,1,0) model as the most suitable for predicting future values for various h-steps ahead horizons.

Overall, we can say that a) the ARIMA or SARIMA model which emerge from the minimization of a standard information criterion without posing very strict restrictions on the orders of p and q seems to produce better forecasts than the model evaluated by the default HK algorithm. This is the case in both datasets examples we studied and it is confirmed with both estimation methods for evaluating model forecasts, meaning standard out of sample error measures and time series cross-validation with a sliding window. Of course this comes with the price of estimating a much more complex model than the rather simplistic model from the HK algorithm. We should bear in mind that for much bigger datasets the problem of minimizing an information criterion over many candidate models may be very time-consuming or even impossible in practice b) Simple polynomial DLM models with a seasonal component can give adequate forecasts for time series that appear seasonality and linear trend. However, their ability to produce forecasts may depend a lot on the first initial values of the state estimates and error variances and as it is shown in the two examples difficulties may arise in confirming the model assumptions. More sophisticated state space models, such as SARIMA models expressed in a state space form, may produce far better results.



Appendix

Critical values for seasonal unit roots[a]

		Fractiles								
		't'π_1			't'π_2			'F':$\pi_3 n \pi_4$		
Model	T	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
No Inter	48	-2.72	-1.95	-1.59	-2.67	-1.95	-1.60	5.02	3.26	2.45
No S dum.	100	-2.60	-1.97	-1.61	-2.61	-1.92	-1.57	4.89	3.12	2.39
No trend	200	-2.62	-1.94	-1.62	-2.60	-1.95	-1.61	4.76	3.12	2.37
Inter.	48	-3.66	-2.96	-2.62	-2.68	-1.95	-1.60	4.78	3.04	2.32
No S. dum.	100	-3.47	-2.88	-2.58	-2.68	-1.95	-1.60	4.77	3.08	2.35
No Trend	200	-3.48	-2.87	-2.57	-2.58	-1.92	-1.59	4.76	3.12	2.37
Inter.	48	-3.77	-3.08	-2.72	-3.75	-3.04	-2.69	9.22	6.60	5.50
S. dum.	100	-3.55	-2.95	-2.63	-3.60	-2.94	-2.63	8.74	6.57	5.56
No. Trend	200	-3.51	-2.91	-2.59	-3.50	-2.89	-2.60	8.93	6.61	5.56
Inter.	48	-4.23	-3.56	-3.21	-2.65	-1.91	-1.57	4.64	2.95	2.23
No. S. dum.	100	-4.07	-3.47	-3.16	-2.58	-1.94	-1.60	4.70	2.98	2.31
No. Trend	200	-4.05	-3.44	-3.15	-2.59	-1.95	-1.62	4.66	3.07	2.34
Inter.	48	-4.46	-3.71	-3.37	-3.80	-3.08	-2.73	9.27	6.55	5.37
S. dum.	100	-4.09	-3.71	-3.37	-3.80	-3.08	-2.73	8.79	6.60	5.52
Trend	200	-4.05	-3.49	-3.18	-3.52	-2.91	-2.60	8.96	6.57	5.56

Table 1: Critical values for HEGY one-sided t1 and t2 tests and for the F1,2 test in quarterly time series. Source: Hylleberg et. al. (1990 pp 226-227)

References

Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control, 19,* 716-723.

Beaulieu, J.J., MacKie-Mason, J.K., and Miron, J.A. (1992), Why Do Countries and Industries with Large Seasonal Cycles Also Have Large Business Cycles?, *The Quarterly Journal of Economics, 107,* 621–656.

Beaulieu, J., and Miron, J.A. (1993), Seasonal unit roots in aggregate U.S. data, *Journal of Econometrics, 55,* 305-328.

Box, G., and Jenkins, G. (1976), Time Series Analysis: Forecasting and Control (Revised Edition), Holden-Day, Oakland, California, rev. edition.

Box, G., Jenkins, G., and Reinsel, G. (2008), *Time Series Analysis: Forecasting and Control*, Wiley, Oxford, 4thedition.

Brockwell, P.J., and Davis, R.A. (1991), *Time Series: Theory and Methods,* Springer: New York.

Burnham, K., and Anderson, D. (2004), Multimodel inference: Understanding aic and bic in model selection, *Sociological Methods and Research, 33,* 261-304.

Diebold, F., and Mariano, R. (1995), Comparing predictive accuracy, *Journal of Business and Economic Statistics, 13,* 134-144.

Franses, P. H. (1991), Seasonality, non-stationarity and the forescasting of monthly time series, *International Journal of Forecasting, 7,* 199-208.

Ghysels, E., Lee, H., and Noh, J. (1994), Testing for unit roots in seasonal time series,*Journal of Econometrics, 62,* 415-442.

Hamilton, James D. (1994), *Time Series Analysis,* Princeton University Press, Princeton: New Jersey.

Harrison P. J., and Stevens C. F. (1976), Bayesian Forecasting, *Journal of the Royal Statistical Society, 38,* 205-247.

Harvey, A. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter,* Cambridge University Press, Cambridge, 1 edition.

Hurvich, C., and Anderson, T. C. (1989), Regression and time series model selection in small samples, *Biometrika, 76,* 297-307.

Hylleberg, S. (1995), Tests for seasonal unit roots. General to specific or specific to general? *Journal of Econometrics*, *69*, 5-25.

Hylleberg, S., Engle, R.F., Granger, C.W.J., and Yoo, B.S (1990), Seasonal integration and cointegration, *Journal of Econometrics, 44*, 215-238.

Hyndman, R., and Khandakar, Y. (2008), Automatic time series forecasting: The forecast package for r, *Journal of Statistical Software, 27*,1-22.

Jarque, C., and Bera, A. (1980), Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters, 6*, 255-259.

Kalman, R. (1960), A new approach to linear filtering and prediction problems, *Transactions of the ASME-Journal of Basic Engineering, 82*, 35-45.

Ljung, G., and Box, G. (1978), On a measure of a lack of fit in time series models, *Biometrika, 65*,297-303.

Perron, P. (1995), Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag, *Journal of the American Statistical Association,* 90, 268-281.

Petris, G., Petrone, S., and Campagnoli, P. (2009), *Dynamic Linear Models with R*, Springer, New York.

Schwert,G. (2009), *EViews 7 User's Guide II*, Quantitative Micro Software, LLC, Irvine, California.

West, M., and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models,* Springer, 2nd Edition.


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ