MODELING AND FORECASTING GAMBIA'S INFLATION RATES

SANNA MANJANG

MASTER OF SCIENCE IN MATHEMATICS (STATISTICS OPTION)

PAN AFRICAN UNIVERSITY

INSTITUTE FOR BASIC SCIENCES TECHNOLOGY AND INNOVATION

October, 2014

Modeling and Forecasting Gambia's Inflation Rates

Sanna Manjang

MS300-0003/12

A Thesis Submitted to Pan African University, Institute for Basic Sciences, Technology and Innovation in partial fulfillment of the requirements for the degree of Masters of Science in Mathematics (Statistics Option)

October, 2014

DECLARATION

Part(i)

This proposal is my original work and has not been presented for a degree in any other university:

Signature: .... Date: ....

Sanna Manjang

Part(ii)

This proposal has been submitted for examination with our approval as university supervisors:

Signature: .... Date: ....

Professor Abdou ka Diongue

Universit Gaston Berger, BP 234 Saint-Louis, Senegal,

URL: http://www.ugb.sn/

Signature: .... Date: ....

Professor Leo Odongo

Kenyatta University,

URL: http://www.ku.ac.ke

Acknowledgments

I wish to gratefully acknowledge financial support from the African Union Commission and from the Government of Kenya. I would also like to express my sincere gratitude to my supervisors Professor Abdou ka Diongue and professor Leo Odongo for their guidance and the helpful comments they provided me in accomplishing my objectives.

I am also grateful to all the lecturers and entire staff of Pan African University especially Professor Magoma (Director) for facilitating the provision of facilities and a conducive environment for my research.

I extend my sincere gratitude to officials of Central Bank of The Gambia for providing the data necessary for this study.

Dedication

This dissertation is lovingly dedicated to my wife and daughter, my mother and to the memory of my father who have all been my constant source of inspiration. They have given me the drive and discipline to tackle any task with enthusiasm and determination. Without their love and support this project would not have been made possible.

Abstract

In this thesis, we examine the most appropriate method for modeling and forecasting Gambia's inflation rates. We investigate the statistical properties of the inflation data and specify two models namely seasonal autoregressive integrated moving average (SARIMA) and k-factor Gegenbauer Autoregressive Moving Average (k-factor GARMA).The first model seasonal $ARIMA(1,1,1)(0,0,1)_{12}$ was selected using the H-K Algorithm developed by Hyndman and Khandakar (2008) and 3-factor *GARMA* from both the spectral density graph and further analysis of the residuals from the 3-factor Gegenbauer model. The in-sample characteristics such as the Akaike Criterion and Schwarz Criterion following estimation using the first data set show that the $ARIMA(1,1,1)(0,0,1)_{12}$ outperforms the 3-factor *GARMA* model. However,the second data set that was preserved and used for out-of-sample forecasting suggest that the 3-factor *GARMA* model outperforms the seasonal $ARIMA(1,1,1)(0,0,1)_{12}$ model in out-of -sample forecasting. Our results indicated that inflation in Gambia is stationary with long-memory behavior at three distinct frequencies. We also found that the k-factor GARMA outperforms the seasonal ARIMA in out-sample forecasting which may be ascribed to the forecast horizon been large and series strongly long-range dependent.

Contents

Nomenclature

ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller
AIC	Akaike Information Criterion
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroskedasticity
ARMA	Autoregressive Moving Average
BIC	Bayesian Information Criterion
CH	anova and Hansen
DESARIMA	Driftless Extended Seasonal AutoregressiveIintegrated Moving Average
DGP	Data Generating Process
FIGARCH	Fractionally Integrated Generalized Autoregressive Conditionally Heteroskedastic
GARMA	Gegenbauer Autoregressive Moving Average
HEGY	Hylleberg, Engle, Granger and Yoo
HK	Hyndman-Khandakar
KPSS	Kwiatkowski, Phillips, Schmidt and Shin

KPSS	Kwiatkowski, Phillips, Schmidt and Shin
KPSS	Kwiatkowski, Phillips, Schmidt and Shin
LM	Lagrange Multiplier
MDS	martingale difference sequence
MLE	Maximum Likelihood Estimation
OLS	Ordinary Least Squares
PACF	Partial Autocorrelation Function
SARIMA	Seasonal Autoregressive Integrated Moving Average
SMA	Seasonal Moving Average

List of Tables

List of Figures

Chapter 1

Introduction

Inflation as defined by Webster (2000) is the persistent increase in the level of consumer prices or a persistent decline in the purchasing power of money. Inflation causes global concern because it can distort economic patterns when not anticipated. Inflation as described by Aidoo (2010) can cause uncertainty about the future price, interest rate, and exchange rate etc which as a result might increase the risk among potential traders and partners of a country. Inflation in The Gambia exhibit evidence of cyclical persistence justifying the use of a seasonal model such SARIMA (seasonal autoregressive integrated moving average) which is an extension of the ARIMA model when the series contains both seasonal and non-seasonal behavior and a long memory model such as k-factor Gegenbauer autoregressive moving average (k-factor GARMA) in modeling it.

A stationary process is said to be long memory if it exhibits a slowly decaying autocorrelation function (ACF), denoted $\rho(n)$, approximated as follows : $\rho(n) \sim c_p(n)n^{2d-1}$, as n tends to infinity, where $c_p(n)$ is a slowly varying function at infinity and d is the long memory parameter such that $0 < d < 1/2$, and therefore the infinite sum $\sum|\rho(n)|$ diverges. In an equivalent way, in terms of spectral analysis, a stationary process is said to be long memory if for some frequency $\lambda \in [0, \pi]$ the spectral density becomes unbounded (Woodward *et al.*, 2011). Forecasters need to take into account long-range dependence in Modeling time series to obtain accurate results. In many economic and financial applications, time series often possess persistent cycles, especially when dealing with daily or monthly frequencies. To take into account this periodic and persistent behavior long memory models have been proposed which makes it possible to integrate a certain type of nonstationarity in modeling and therefore avoids a differentiation of the series which strongly reduces the available information set. Long-range dependence has been studied extensively and evidence of it can be found in many areas of applied statistics.For instance, the pioneer work of Hurst (1951) stems from problems dealing with hydrology, and later many authors analyzed river flows through long memory processes, see for instance Noakes *et al.* (1998) or Ooms and Franses (1998). More recently, researchers founded evidence of persistence in telecommunication data (Taqqu *et al.*, 1997) or in urban transport data (Ferrara and Guegan, 1999). However, most long memory applications concern financial time series such as stock market prices, exchange rates, inflation, etc.

In Gambia, the inflation rate measures a broad rise or fall in prices that consumers pay for a standard basket of goods. In general, Inflation is a change in the general price level of all goods and services. If inflation rises it means that a Gambian dalasi is worth less because it can buy less goods and services. For this reason it is important to model and forecast inflation rates in The Gambia to help policy makers in their decision making and also see the purchasing power of the dalasi. The central Bank of The Gambia's mandate is to keep inflation low, stable and predictable in order to preserve the value of dalasi. Targeting a low inflation rate ensures that prices will not rise so quickly as to seriously erode the purchasing power of the dalasi. Having a stable and predictable inflation rate is very important to the economy as well. It allows for greater certainty among households and firms regarding the purchasing power of their savings and income. For example, if inflation was very volatile, people would be uncertain about the value of their money from one month to the next. In addition, a stable inflation reduces the volatility of unemployment. While there is no long run tradeoff between inflation and unemployment, the negative slope of the short run Phillips curve suggests that higher inflation variability comes with larger fluctuations in unemployment. So keeping inflation stable reduces the volatility in unemployment. Furthermore, low inflation rates keeps both nominal and real interest rates low and Low real interest rates are good for the economy because they reduce the cost of funds thereby supporting capital growth which in turn increases future production possibilities. Price signals are also very important to both households and businesses. For efficient financial decision making, households and businesses must read price signals to determine if prices are changing due to changes in demand, supply or productivity. If inflation rates are volatile then it is difficult to read the real changes in price and allocate resources efficiently.Inflation targeting allows households and businesses to better read price changes and thus respond to price shocks more swiftly.

1.1 Statement Of The Problem

There is almost a consensus that macroeconomic stability, specifically defined as low inflation is positively related to economic growth. Over the years extensive research has been carried out in various countries on modeling and forecasting inflation. These researchers have been motivated by the fact that , in general, inflation has a negative effect on medium and long-term growth (Bruno and Easterly,1998). If inflation is detrimental to economic growth, it is important that policy makers be provided with means to predict future inflation and adjust their policies accordingly.In Gambia, no work has been done in modeling and forecasting inflation using the proposed models. This study therefore contributes to the existing literature by modeling and forecasting the inflation rates in Gambia using the proposed methods. Developing a model that can accurately predict inflation in The Gambia will go a long way in guiding policy makers in decision making.

1.2 Objectives Of The Study

1.2.1 Main Objective

The broad objective of this study is modeling and forecasting inflation.

1.2.2 Specific Objectives

The specific objectives are to:

- determine the trend of inflation in The Gambia over the years
- specify appropriate seasonal autoregressive integrated moving average and k-factor Gegenbauer autoregressive moving average models for The Gambian inflation rates
- determine the statistical properties of the proposed models
- use the proposed seasonal models to forecast inflation rates in The Gambia

1.3 Significance Of The Study

Economic agents base many current decisions among other things on their expectation of the future inflation pattern. The views held about future inflation may influence firms' price-setting behavior or workers' wage-demands, and thereby impact current purchasing power and labor costs; through its effect on the real interest rate and on inflation risk premia, the expected inflation rate furthermore influences savings and investment decisions. Inflation forecasts and projections are also often at the heart of economic policy decision-making, as is the case for monetary policy, which in most industrialized economies is mandated to maintain price stability over the medium term. Decision-makers hence need to have a view on the likely future path of inflation when taking measures that are necessary to reach their objective. This modeling and forecasting of inflation in The Gambia will enable the government and the consumers to make better decisions as they will be provided with forecasts for future inflation.

Chapter 2

Literature Review

Empirical researches have been carried out in the area of forecasting using Autoregressive Integrated Moving Average (ARIMA) models popularised by Box and Jenkins (1976). Junttila (2001), applied the Box and Jenkins (1976) approach to model and forecast Finnish inflation. Pufnik and Kunovac (2006) applied a similar approach to forecast short term inflation in Croatia. Alnaa and Ferdinand (2011) used ARIMA approach to predict inflation in Ghana. In their study, they used monthly data from June, 2000 to December, 2010 and found that ARIMA (6, 1, 6) is best for forecasting inflation in Ghana. An extended version of the Seasonal ARIMA, known as the Driftless Extended Seasonal ARIMA (DESARIMA) was introduced in a study by Pincheira and Medel (2012) to forecast inflation across 12 countries. Also, Barros and Gil-Alana (2012) employed a fractional approach (Autoregressive Fractionally Integrated Moving Average) to forecast inflation in Angola. Adjepong *et al.* (2013) considered the most appropriate short-term forecasting method for Ghana's inflation: Seasonal ARIMA vs. Holt-Winters. They concluded by proposing the Seasonal-ARIMA process as the most appropriate short-term forecasting method for Ghana's inflation. Recently Long-memory models have then become increasingly popular (Hassler and Wolters., 1995). (see, e.g., Chung and Baillie, 1993, and Franses and Ooms, 1997). Much of the evidence supports the view that inflation is fractionally integrated with a differencing parameter that is significantly different from zero or unity (Sutcliffe, 1994).For instance, using US monthly data, Backus and Zin (1993) found a fractional degree of integration. They argued that aggregation across agents with heterogeneous beliefs results in long memory in the inflation process. Hasser (1993) and Delgado and Robinson (1994) provided strong evidence of long memory in the Swiss and Spanish inflation rates respectively. A new class of Fractionally Integrated Generalized Autoregressive Conditionally Heteroskedastic (FIGARCH) processes was introduced by **?**. Empirical evidence of long memory has also been found in monthly river flows Ooms and Franses (1998), stock market prices (Cheung and Lai, 1995), (Barkoulas and Baum, 1996), Willinger *et al.* (1999)or exchange rates (Cheung, 1993), Bisaglia and Guégan (1998), Velasco (1999). Caporale and Gil-Alana (2011) modeled European inflation rates using Multi-Factor Gegenbauer Processes. Their findings show that inflation in France and Italy is nonstationary, but in the former country this applies to both the long-run and the seasonal frequencies, whilst for the latter the nonstationarity concerns exclusively the long-run or zero frequency, and the contribution of the long-range dependence in the seasonal structure is relatively small. For the UK, inflation seems to be stationary, though with a large component of long-memory behavior, especially at the zero frequency. In the present study inflation rates in The Gambia is modeled which to the best of my knowledge has not been done before. A seasonal autoregressive integrated moving average (SARIMA) and k-factor Gegenbauer Autoregressive Moving Average (k-factor GARMA) models are specified for the Gambia inflation rates and their in-sample characteristics and out-sample forecasting performances compared. The results indicated that inflation in Gambia is stationary with long-memory behavior at three distinct frequencies. We also found that the k-factor GARMA outperforms the seasonal ARIMA in out-sample forecasting which may be ascribed to the forecast horizon been large and series strongly long-range dependent.

The outline of this thesis is as follows: section III describes the inflation data from the Central Bank of The Gambia. In section IV we present the two models namely seasonal autoregressive integrated moving average (SARIMA) model and k-factor Gegenbauer Autoregressive Moving Average (k-factor GARMA) model, section V deals with applications, and section VI gives a summary of our findings.

Chapter 3

Data and Methodology

The data was obtained from the Central Bank of the Gambia spanning the period January 1987 to June 2013 and is measured in percentages based on the consumer price index (CPI) of The Gambia. Figure 3.1 shows the graphical representations of realizations, distribution function, periodogram and the autocorrelation function.The spectral density is unbounded at the low frequencies and the autocorrelation function decay very slowly which suggest that the inflation series seems to be long memory process with fractional integration behavior. The periodogram exhibits three distinct spikes at the low frequencies and smaller peaks at other seasonal frequencies which are all indications that the inflation process in Gambia is possibly fractionally integrated with $0 < d_i < 1$, $i = 1, ..., k$, where k is a finite integer indicating the number of seasonal structures. Patterns of seasonality are evident in the plot of the realizations and the autocorrelation function.

Description	statistic
Number of observations	318
Mean	6.8533
Median	5.4000
Maximum	-2.9100
Minimum	58.330
Std	6.4450
C.V.	0.94042
Skewness	2.8965
Kurtosis	15.839

Table 3.1: Summary statistics, using the observations 1987:01 - 2013:06 for the inflation rates

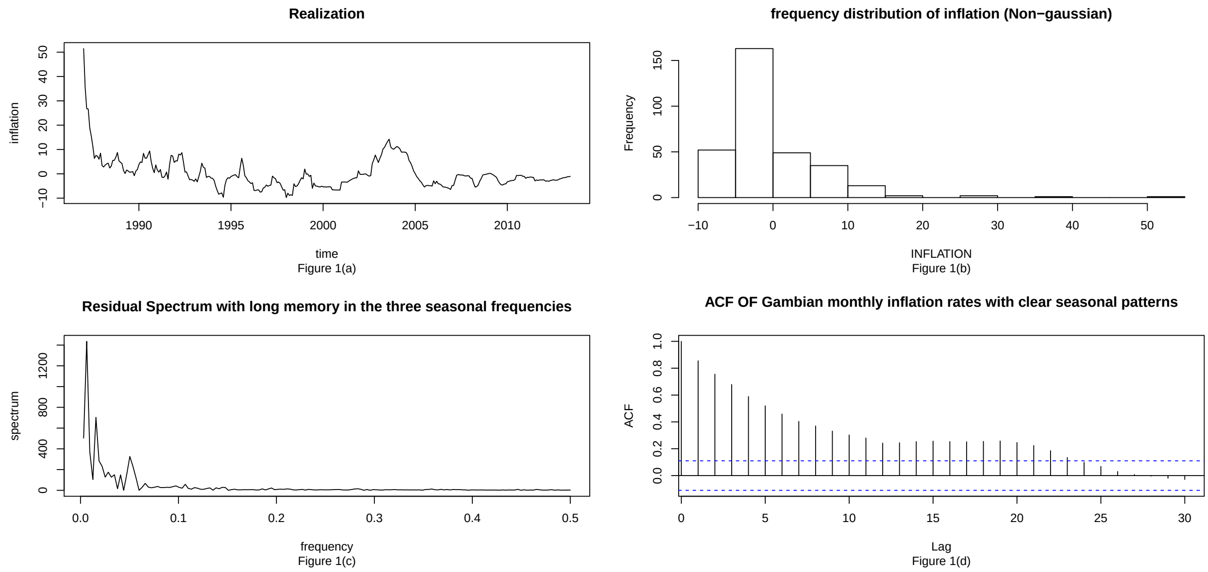


Figure 3.1: realization,frequency distribution, periodogram and autocorrelations of inflation rates

3.1 Test for Stationarity

Because the order of integration of a time series is of great importance for analysis, a number of statistical tests have been developed for investigating it. In this study we check the stationarity of the inflation series using Augmented Dickey-Fuller (ADF) Test and Kwiatkowski, Phillips, Schmidt and Shin (KPSS) Test. A description of the two tests will be given in this section.

3.1.1 Augmented Dickey-Fuller (ADF) Test

Dickey and Fuller (1979) tests are based on models of the form

$$\triangle x_t = \Psi x_{t-1} + \sum_{i=1}^{p-1} \beta_j^* \triangle x_{t-i} + \varepsilon_t. \tag{3.1.1}$$

The following pair of hypotheses

$$H_0 : \Psi = 0 \ versus \ H_1 : \Psi < 0$$

is tested based on the t-statistic of the coefficient Ψ from an OLS estimation of equation 3.1.1 (Fuller (1976), Dickey and Fuller (1979)). The null hypothesis H_0 is rejected if the t-statistic is smaller than the relevant critical value. If $\Psi = 0$ (that is, under H_0) the series x_t has a unit root and is nonstationary, whereas it is regarded as stationary if the null hypothesis is rejected.

The test statistic has a nonstandard limiting distribution. Critical values have been obtained by simulation and they are available, for instance, in Fuller (1976) and Davidson and MacKinnon (1993). The limiting distribution depends on the deterministic terms which have to be included. Therefore, different critical values are used when a constant or linear trend term is added in equation 3.1.1. Also seasonal dummies may be included.In these tests a decision on the AR order or, equivalently, on the number of lagged differences of x_t has to be made.

3.1.2 KPSS Test

The integration properties of a series x_t may also be investigated by testing

$$H_0 : x_t \sim I(0) \ versus \ H_1 : x_t \sim I(1),$$

that is, the null hypothesis that the DGP is stationary is tested against a unit root. Kwiatkowski *et al.* (1992) have derived a test for these pair of hypotheses. If there is no linear trend term, they start from a DGP

$$x_t = y_t + z_t \tag{3.1.2}$$

where x_t is a random walk, $y_t = y_{t-1} + v_t$, $v_t \sim iid(0, \sigma_v^2)$ and z_t is a stationary process.In this framework the foregoing pair of hypotheses is equivalent to the pair H_0 : $\sigma_v^2 = 0$ versus $H_0 : \sigma_v^2 > 0$. Kwiatkowski et al. (1992) propose the following test statistic

$$KPSS = \frac{1}{T^2} \sum_{t=1}^{T} \frac{S_t^2}{\sigma_\infty^2} \tag{3.1.3}$$

where $S_t = \sum_{i=}^{t} \hat{w}_i$ with $\hat{w}_t = x_t - \bar{x}$ and σ_∞^2 is an estimator of the long-run variance of the process z_t. Kwiatkowski et al. (1992) propose a nonparametric estimator for this quantity based on a Bartlett window with a lag truncation parameter $l_q = q(T/100)^{\frac{1}{4}}$:

$$\sigma_\infty^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{w}_t^2 + 2\sum_{i=1}^{l_q} w_i(\frac{1}{T}\sum_{t=i+1}^{T}\hat{w}_t\hat{w}_{t-i}) \tag{3.1.4}$$

where $w_i = 1 - \frac{i}{l_q+1}$. Critical values may be found, e.g., in Kwiatkowski et al. (1992). The null hypothesis of stationarity is rejected for large values of KPSS.

If a deterministic trend is suspected, the point of departure is a DGP

$$x_t = \delta_1 t + y_t + z_t, \tag{3.1.5}$$

and the $\hat{w}_t$ are residuals from a regression

$$x_t = \delta_0 + \delta_1 t + w_t \tag{3.1.6}$$

With these quantities the test statistic is computed in the same way as before. Its limiting distribution under H_o is different from the case without trend term, however. Critical values for the case with trend are available from Kwiatkowski et al. (1992).

3.2 Testing for Seasonal Unit Roots

3.2.1 The HEGY test

Tests have been proposed by Hylleberg *et al.* (1990) to check for seasonal unit roots in quarterly time series. They are based on the model

$$\triangle_4 x_t = \pi_1 z_{1,t-1} + \pi_2 z_{2,t-1} + \pi_3 z_{3,t-1} + \pi_4 z_{3,t-2} + \sum_{i=1}^{p}\beta_i^*\triangle_4 x_{t-i} + \varepsilon_t \tag{3.2.1}$$

where $z_{1t} = (1+L+L^2+L^3)x_t$, $z_{2t} = -(1-L+L^2-L^3)x_t$, and $z_{3t} = -(1-L^2)x_t$

with L being the lag operator.The null hypotheses $H_0 : \pi_1 = 0$, $H_0 : \pi_2 = 0$ and $H_0 : \pi_3 = \pi_4 = 0$ correspond to tests for regular, semiannual and annual unit roots, respectively. These hypotheses can be tested by estimating the model in 3.2.1 by OLS and considering the relevant $'t-'$ and $'F-tests'$. These tests are known as HEGY tests.

For monthly series the corresponding tests for seasonal unit roots were discussed by Franses (1990) based on the model

$$\begin{aligned}\triangle_{12}x_t &= \pi_1 z_{1,t-1} + \pi_2 z_{2,t-1} + \pi_3 z_{3,t-1} + \pi_4 z_{4,t-2} + \pi_5 z_{4,t-1} + \pi_6 z_{4,t-2} \\ &\quad + \pi_7 z_{5,t-1} + \pi_8 z_{5,t-2} + \pi_9 z_{6,t-1} + \pi_{10} z_{6,t-2} + \pi_{11} z_{7,t-1} + \pi_{12} z_{7,t-2} \\ &\quad + \sum_{i=1}^{p} \beta_i^* \triangle_{12} x_{t-i} + \varepsilon_t \qquad (3.2.2)\end{aligned}$$

where

$$\begin{aligned}z_{1,t} &= (1+L)(1+L^2)(1+L^4+L^8)x_t && (3.2.3)\\ z_{2,t} &= -(1-L)(1+L^2)(1+L^4+L^8)x_t && (3.2.4)\\ z_{3,t} &= -(1-L^2)(1+L^4+L^8)x_t && (3.2.5)\\ z_{4,t} &= -(1-L^4)(1-\sqrt{3}L+L^2)(1+L^4+L^8)x_t && (3.2.6)\\ z_{5,t} &= -(1-L^4)(1+\sqrt{3}L+L^2)(1+L^4+L^8)x_t && (3.2.7)\\ z_{6,t} &= -(1-L^4)(1-L^2+L^4)(1-L+L^2) && (3.2.8)\\ z_{7,t} &= -(1-L^4)(1-L^2+L^4)(1+L+L^2) && (3.2.9)\end{aligned}$$

The process x_t has a regular (zero frequency) unit root if $\pi_1 = 0$ and it has seasonal unit roots if any one of the other $\pi_i = 0, (i = 2, ..., 12)$ is zero. For the conjugate complex roots, $\pi_i = \pi_{i+1,}(i = 3, 5, 7, 9, 11)$ is required. The corresponding statistical hypotheses can again be checked by t- and F-statistics, critical values for which are given by Franses and Hobijn (1997). If all the $\pi_i(i = 1, ..., 12)$ are zero, then a stationary model for the monthly seasonal differences of the series is suitable.

3.2.2 The CH test

The Canova and Hansen (1995) (CH) test was developed in 1992 and is used to test the null hypothesis that the time series process is stationary with deterministic seasonality against the alternative that it has a seasonal unit root.

Let x_t be a real valued variable observed S times per year. The Canova-Hansen tests for stability are based on the residuals of the following auxiliary regression:

$$x_t = \mu + y_t'\beta + S_t + \varepsilon_t \quad t = 1, 2, ..., T \tag{3.2.10}$$

where T is the number of observations or sample size, x_t is a $k \times 1$ vector of explanatory variables, S_t is a deterministic seasonal component and $\varepsilon_t = (0, \sigma^2)$ is an error uncorrelated with y_t and S_t. The dependent variable x_t must bee free of unit roots at the zero frequency, so Canova and Hansen suggest to difference the observed series in order to eliminate the zero frequency unit root. f no explanatory variables are included then the error ε_t will be the difference between the modeled process x_t and its seasonal component S_t .

The deterministic seasonal component may be specified into two different ways: seasonal dummies or trigonometric terms. In the first case the auxiliary regression 3.2.10 is:

$$x_t = y_t'\beta + d_t'\alpha + \varepsilon_t \tag{3.2.11}$$

where d_t is an $S \times 1$ vector of seasonal dummy indicators and α is an $S \times 1$ parameter vector, S being the seasonal periodicity ($S = 4$ for quarterly data, $S = 12$ for monthly data,...). To study whether the seasonal intercepts change over time, Canova and Hansen consider stochastic variation of a martingale form:$A'\alpha_t = A'\alpha_{t-1} + u_t$ where α_0 is fixed and μ_t is a martingale difference sequence (MDS) with covariance matrix $E(\mu_t\mu_t') = \tau^2 G$. A is an $S \times a$ selection matrix which serve to select the elements of α that we allow to stochastically change under the alternative. Testing stability of the jth intercept can be achieved by choosing A to be the unit vector with a 1 in the jth element and zeros elsewhere.

Equation 3.2.10 may be also specified in a form equivalent to 3.2.11 defining the seasonal component by means of trigonometric terms. In this case, the auxiliary regression may be written as:

$$x_t = \mu + y_t'\beta + \sum_{j=1}^{q} f_{jt}'\gamma_j + \varepsilon_t, \tag{3.2.12}$$

where $q = [\frac{S}{2}]$ is the integer part of $\frac{S}{2}$ and f_{jt} is the deterministic cyclical process at the seasonal frequency $\theta_j = \frac{2\pi j}{S}; j = 1, ..., q$, which, defining S^* as $\frac{S}{2} - 1$ if S is even and

$(\frac{S-1}{2})$ if S is odd, may be defined as:

$$f'_{jt} = \begin{cases} (cos(\theta_j t), sin(\theta_j t)) & j = 1, ..., S* \\ (cos(\theta_j t) & j = \frac{S}{2}\ (only\, for\, S\, even) \end{cases} \tag{3.2.13}$$

Stacking the q elements of the previous sum in a vector $\gamma = (\gamma_1, ..., \gamma_q)$,$f_t = (f_{1t}, ..., f_{qt})'$ the equation 3.2.12 may be expressed as:

$$x_t = \mu + y'_t \beta + f'_t \gamma + \varepsilon_t \tag{3.2.14}$$

Under 3.2.14, the seasonal pattern is stable and γ is a vector of $S-1$ seasonal coefficients which are constant over time. To setup the tests of seasonal stability against seasonal unit roots Canova and Hansen propose for γ_t the process $A'\gamma_t = A'\gamma_{t-1} + u_t$ with γ_0 fixed and μ_t a MDS, where A is a $(S-1) \times a$ selection matrix such that, for example, $A = I_{s-1}$ may be used to test whether the entire vector is stable, $A = (\tilde{0}, I_2, \tilde{0})'$ (commensurate with γ_i) to test for a unit root only at a specific frequency j $(\neq \pi)$ and $A = (\tilde{0}, 1)'$ serve for testing a unit root at frequency π .

The variance-covariance matrix of μ_t is $E(\mu_t \mu'_t) = \tau^2 G$ where G is a full rank $a \times a$ matrix and $\gamma^2 \geq 0$ is real valued. When $\gamma^2 = 0$,$\gamma_t = \gamma_0$ the model has no seasonal unit roots. When $\gamma^2 > 0$, x_t has a unit root at the seasonal frequencies determined by A.

Considering the hypothesis test of $H_0 : \gamma^2 = 0$ against $H_1 : \gamma^2 > 0$ they propose a LM test statistic which takes the form:

$$L = \frac{1}{T^2} tr((A'\hat{\Omega}^f A)^{-1} A' \sum_{t=1}^{T} \hat{F}_t \hat{F'}_t A), \tag{3.2.15}$$

where $\hat{F}_t = \sum_{j=1}^{t} f_j \varepsilon_j$, $\hat{\Omega}^f$is a consistent estimate of the long run covariance matrix of $f_t \varepsilon_t$,

$$\hat{\Omega}^f = \sum_{k=-m}^{k=m} w(\frac{k}{m}) \frac{1}{T} \sum_{t=1}^{T} f_{t+k} \hat{\varepsilon}_{t+k} f'_t \hat{\varepsilon}_t \tag{3.2.16}$$

and $w()$ is a kernel which gives a positive semidefinite matrix, such as the Bartlett kernel.

The large sample distribution was studied by Nyblom(1989) and Hansen(1990,1992), establishing that under $H_{0,}L \underset{\rightarrow}{d} VM(a)$, the generalized Von Mises distribution with a degrees of freedom which is tabulated in the Canova-Hansen paper.

3.3 Models

In this section the SARIMA and k-factor GARMA models will be presented and identification, estimation, diagnostic checking, and forecasting discussed.

3.3.1 The SARIMA Model

The extension of ARIMA model to the SARIMA model comes in when the series contains both seasonal and non-seasonal behavior. This behavior of the series makes the ARIMA model inefficient to be applied to the series. This is because it may not be able to capture the behavior along the seasonal part of the series and therefore mislead to a wrong order selection for non-seasonal component.

The $SARIMA(p,d,q)(P,D,Q)_S$ model is specified as follows:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D x_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t \qquad (3.3.1)$$

with $\phi_p(B), \Phi_P(B^S), \theta_q(B), \Theta_Q(B^S)$ defined in the following equations 3.3.2, 3.3.3, 3.3.4, and 3.3.5:

$$\begin{aligned}
\phi_p(B) &= 1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p && (3.3.2)\\
\Phi_P(B^S) &= 1-\Phi_1 B^S-\Phi_2 B^{2S}-\ldots-\Phi_P B^{PS} && (3.3.3)\\
\theta_q(B) &= 1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q && (3.3.4)\\
\Theta_Q(B^S) &= 1-\Theta_1 B^S-\Theta_2 B^{2S}-\ldots-\Theta_Q B^{QS} && (3.3.5)
\end{aligned}$$

where,

p, *d* and *q* are the order of non-seasonal AR, differencing and MA respectively.

P, *D* and *Q* are the order of seasonal AR, differencing and MA respectively.

x_t represents the time series data at period t.

ε_t represents white noise error (random shock) at period t.

B represents the backward shift operator ($B^k x_t = x_{t-k}$).

$(1-B^S)^D$ represents the seasonal difference.

$(1-B)^d$represents the non-seasonal difference.

S represents seasonal order ($S = 12$ in this case).

The $SARIMA(p,d,q) \times (P,D,Q)$ model defined in equation (3.3.1) is referred to as a *multiplicative model*.

This is because the nonseasonal and seasonal *AR* operators are multiplied together on the left hand side while the *MA* operators are multiplied together on the right hand side (Brockwell and Davis, 2002).

In addition, the nonseasonal and seasonal differencing operators are multiplied together with the *AR* operators.

To obtain a SARIMA model that does not have separate differencing factors, we write the model in generalized form as

$$\phi'(B)x_t^{(\lambda)} = \theta'(B)\varepsilon_t \tag{3.3.6}$$

where $\phi'(B) = \phi(B)\Phi(B^s)\nabla^d\nabla_s^D$ is the generalized AR parameter, and $\theta'(B) = \theta(B)\Theta(B^s) = \tilde{\Theta}(B)$ is the generalized *MA* operator for which θ_i' is the ith generalized *MA* operator and $x_t^{(\lambda)}$ represents the series. The generalized form is useful for calculating MMSE forecasts.

3.3.2 Stationarity and Invertibility condition of the SARIMA

For the nonseasonal part the roots of $\phi(B) = 0$ must lie outside the unit circle for stationarity of the series and for the invertibility condition the roots of $\theta(B) = 0$ must fall outside the unit circle. For the seasonal stationarity the roots of the equation $\Phi(B^s) = 0$ must lie outside the unit circle and for seasonal invertibility the roots of the characteristics equation $\Theta(B^s) = 0$ must fall outside the unit circle (Hipel and McLeod, 1994).

3.3.3 Identification step

The identification procedure comprises the following steps: plotting the data, possibly transforming the data, identifying the dependence order of the model to determine the values of p,P,q,Q,D,and d which are obtained using the Autocorrelation function ACF and the Partial Autocorrelation Function PACF . The ACF plays a major role in modeling the dependence among observations, because, it characterizes the process describing the evolution of Y_t over time. From the ACF we can infer the extent to which one value of the process is correlated with the previous values and thus the length and strength of the memory of the process. It indicates how long (how strongly) a shock in the process affects the values of Y_t . The partial autocorrelation function measures correlation between (time series) observations that are k time periods apart after controlling for correlation at intermediate lags i.e. Correlation between Y_t and Y_{t-k} after removing the effect of the intermediate Ys.In general, the seasonal and nonseasonal autoregressive components have their PACF and IACF cutting off at the seasonal and nonseasonal lags. On the other hand, the seasonal and nonseasonal moving average components produce PACF and IACF that show exponential decay or damped sine waves at the seasonal and nonseasonal lags (Johnston and Dinardo).

3.3.4 Selection with the HK-algorithm

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 some information criterions 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, $L(\tilde{\Psi})$, where $\tilde{\Psi}$ is the maximum likelihood estimates of the parameters for the SARIMA with $n = p+q+P+Q+1$ parameters and sample size T . The criterions are

$$AIC = -2log[L(\tilde{\Psi})] + 2n \tag{3.3.7}$$

$$AIC_c = AIC + \frac{2n(n+1)}{T-n+1} \tag{3.3.8}$$

$$BIC = -2log[L(\tilde{\Psi})] + nlog(T) \tag{3.3.9}$$

The HK-algorithm then performs an iterative procedure to select the model that minimizes the value of each criterion.

3.3.5 Estimation of the parameters

The next step is the estimation of the model parameters for the chosen model using maximum likelihood estimation and the final model selection based on the model with the maximum log – likelihood and minimum values of Akaike Information Criterion (AIC),and normalized Bayesiam Information Criterion.

3.3.6 Maximum likelihood estimation of the SARMA Model

McLeod and Salas (1983) provide an algorithm for calculating an approximation to the likelihood function of the multiplicative SARMA model given as:

$$\phi(B)\Phi(B^s)x_t = \theta(B)\Theta(B^s)a_t \tag{3.3.10}$$

where

$$\begin{aligned}
\phi(B) &= 1-\phi_1 B - \dots - \phi_p B^p, \\
\theta(B) &= 1-\theta_1 B - \dots - \theta_q B^q \\
\Theta(B^s) &= 1-\Theta_1 B^s - \dots - \Theta_{qs} B \\
\Phi(B^s) &= 1-\Phi_1 B^s - \dots - \Phi_{p_s} B^{sp_s}
\end{aligned}$$

B is the backshift operator, s the seasonal period and a_t a sequence of independent normal variables with mean 0 and variance σ^2. The $a_t's$, called the innovations represent the one step forecast errors when the vector of the model parameters,

$$\beta = (\phi_1, \dots, \phi_1, \theta_1, \dots, \theta_q, \Phi_1, \dots, \Phi_{ps}\Theta_1, \dots, \Theta_{qs})$$

are known. The $ARMA(p,q)$ model is obtained by taking $p_s = q_s = 0$. The stationarity and invertibiliy conditions stated above applied to the equation 3.3.10 respectively. Although the *SARMA* model maybe consider as a special case of the $ARMA(p*,q*)$ model by taking $p* = p + sp_s$, $q* = q + sq_s$, $\phi^*(B) = \Phi(B^s)\phi(B)$ and

$\theta^*(B) = \Theta(B^s)\theta(B)$, it will be shown how a more efficient estimation algorithm can be developed utilizing the multiplicative structure of the SARMA model.

Given the observations x_t $(t = 1, 2, ..., n)$ the exact log-likelihood function maximized over σ^2 may be written, apart from an arbitrary constant, as

$$logL(\beta) = -nlog(S_m/n)/2 \tag{3.3.11}$$

where S_m, the modified sum of squares, is

$$S_m = S[M_n(p, q, p_s, q_s, s)]^{\frac{-1}{n}} \tag{3.3.12}$$

S represents the unconditional sum of squares of Box and Jenkins (1976) define by

$$S = \sum_{t=-\infty}^{n} [a_t]^2 \tag{3.3.13}$$

where [.] denotes the expectation given $x_1, ..., x_n$

The evaluation of S by the iterative unconditional sum of squares method may involve two types of truncation error. First, the infinite sum in 3.3.13 is replaced by

$$S = \sum_{t=1-Q}^{n} [a_t]^2 \tag{3.3.14}$$

for suitably large Q. Theoretically, Q should be chosen so that

$$\gamma_0/\sigma_a^2 - \sum_{i=0}^{Q} \psi_i^2 < e_{to1} \tag{3.3.15}$$

where $\gamma_0 = var(x_t)$, ψ_i is the coefficient of a_{t-i} in the infinite moving average representation of 3.3.10 and $e_{to1}$is an error tolerance. Thus if the model contains an autoregressive factor with roots near the unit cycle, a fairly large Q might be necessary. In practice,

$$Q = q + sq_s + 20(p + sp_s) \tag{3.3.16}$$

is often sufficient. The other truncation error involves terminating the iterative procedure used to calculate $[a_t]$. Several iterations maybe required to obtain convergence when the model contains a moving average factor with roots near the unit cycle. However, sufficient accuracy is usually obtained on the first evaluation.

McLeod (1977,1982) suggested that the term $M_n(p,q,p_s,q_s,s)$ be replaced by $m(p,q,p_s,q_s,s)$, given by:

$$m(p,q,p_s,q_s,s) = M(p,q)[M(p_s,q_s]s \tag{3.3.17}$$

where $M(p,q)$is defined for any ARMA(p,q) model as

$$M(p,q) = M_p^2 M_q^2 / M_{p+q} \tag{3.3.18}$$

where the terms M_p, M_q *and* M_{p+q} are defined in terms of the auxiliary autoregressions, $\phi(B)v_t = a_t$ and $\theta(B)u_t = a_t$ and the left-adjoint autoregression $\phi(B)\theta(B)y_t = a_t$. For the autoregression, $\phi(B)v_t = a_t$, M_p is the determinant of the $p \times p$ matrix with (i,j) entry

$$\sum_{k=1}^{min(i,j)} \phi_{i-k}\phi_{j-k} - \phi_{p+k-i}\phi_{p+k-j} \tag{3.3.19}$$

and similarly for the autoregressions. The $p \times p$ matrix define by equation 3.3.19 is called the Schur matrix of $\phi(B)$**.** Pagano (1973) has shown that a necessary and sufficient condition for stationarity of an autoregression is that its Schur matrix be positive-definite. Thus calculation of $m(p,q,p_s,q_s,s)$ also provides a check on the stationary and invertibility conditions and so during estimation the parameters maybe constrained to the admissible region. Modified Cholesky decomposition is used to evaluate $M(p,q)$. To obtain the MLE for the model parameters, the modified sum of squares must be minimized using standard optimization algorithm .

3.3.7 Diagnostic checking

As a last step in the model-building cycle some checks on the model adequacy are required. Possibilities are doing a residual analysis and over fitting the specified model. For example for $ARIMA(p,d,q)(P,D,Q)_S$,we could over fit this model and test the significance of the additional parameters. A residual analysis is usually based on the fact that the residuals of an adequate model Should be approximately while noise. A plot of the residual can be a useful tool in checking for outliers. The estimated residual autocorrelations are usually examined and for a while noise series these autocorrelations are zero. An ARCH LM-test is performed on the residuals of the fitted model. This test is performed to check for homoscedasticity of the residuals. To check the overall acceptability of the residual autocorrelations, the L Jung –Box test statistics or the Q statistic developed by Box-Pierce are often used (Johnston and Dinardo).

3.3.8 ARCH TEST

This is the conditional heteroscedasticity test introduced by Engel (1982) and is also know as the Lagrange Multiplier test. The test is similar to the F test for $\alpha_i = 0$ $(i = 1, ..., m)$ in the linear regression

$$\varepsilon_t^2 = \alpha_o + \sum_{i=1}^{m} \alpha_i \varepsilon_{t-i} + a_t,\ t = i+1, ..., n,$$

where a_t denote an error term and m is a prespecified integer. Under the null hypothesis of no ARCh effect, the test is $\chi^2_{1-\alpha}$ distributed with m degrees of freedom and α is the significance level.

3.3.9 LJUNG-BOX TEST

Ljung and Box (1978) test is used to test that several autocorrelation of a time series are zero. It has the test statistics

$$Q(m) = n(n+2) \sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{n-l},$$

where $\hat{\rho}_l = \frac{\sum_{t=l+1}^{n} (\varepsilon_t - \bar{\varepsilon}_t)(\varepsilon_{t-l} - \bar{\varepsilon}_t)}{\sum_{t=1}^{n} (\varepsilon_t - \bar{\varepsilon}_t)^2}$ and m is the number of lag autocorrelations to be considered. Under the null hypothesis of no serial dependence in the first m autocorrelations, the decision rule is to reject the null hypothesis if the p-value is less than or equal to α, the significance level.

3.3.10 Forecasting with SARIMA Models

To calculate the conditional expectation of x_{t+1} at time t, and hence, the minimum mean square error forecast $\hat{x}_t(l)$, one takes conditional expectation of the generalized model (3.3.6) to obtain

$$\begin{aligned}[x_{t+1}] &= \phi_1'[x_{t+l-1}] + \phi_2'[x_{t+l-2}] + \ldots + \phi_{p+sP+sD}'[x_{t+l-p-d-sD}] - \theta_1'[\varepsilon_{t+l-1}] \\ &\quad -\theta_2'[\varepsilon_{t+l-2}] - \ldots - \theta_{q+sQ}'[\varepsilon_{t+l-q-sQ}] \end{aligned} \tag{3.3.20}$$

where $l = 1,2,...$, is the lead time for the forecast, $[x_{t+1}]$ denotes the conditional expectation

$$E_t[x_{t+1}|x_t, x_{t-1}, \ldots];$$

θ_i' is the generalized AR parameter defined by

$$\phi_1'(B) = \phi(B)\Phi(B^s)\nabla^d\nabla_s^D \tag{3.3.21}$$

and θ_i' is the generalized MA parameter defined by

$$\theta'(B) = \theta(B)\Theta(B^s) \tag{3.3.22}$$

3.4 The k-factor Gegenbauer process

The Gegenbauer polynomials $(C_j(d,v))_{j\in Z}$ are defined by:

$$(1-2vz+z^2)^{-d} = \sum_{j\geq 0} C_j(d,v)z^j \tag{3.4.1}$$

where $|z| \leq 1$ and $|v| \leq 1$and can be obtained by using the following recursive formula:

$$
\begin{aligned}
C_0(d,v) &= 1 && (3.4.2)\\
C_1(d,v) &= 2dv && (3.4.3)\\
C_j(d,v) &= 2v(\frac{d-1}{j}+1)C_{j-1}(d,v)-(2\frac{d-1}{j}+1)C_{j-2}(d,v) && (3.4.4)
\end{aligned}
$$

$\forall j > 1.$

Using the definition above, the k-factor Gegenbauer process $(X_t)_{t\in Z}$ is defined by the following equation:

$$
\prod_{j=1}^{k}(I-2v_jB+B^2)^{d_j}(X_t-\mu)=\varepsilon_t \tag{3.4.5}
$$

where k is a finite integer indicating the maximum number of seasonal structures, where $|v_j| \leq 1$ for $j = 1,...,k$, where $d_j's$ are fractional values for $j = 1,...,k$, where μ is the mean of the process and where $(\varepsilon_t)_{t\in Z}$ is covariance stationary with mean zero and variance σ_ε^2 (Gray *et al.*, 1989).

see Gray et al. (1998), Robinson (1994), Giraiti and Leipus (1995) and Woodward et al. (1998)

3.4.1 Properties of the k-factor Gegenbauer process defined in equation(2)

Let $d_j \neq 0$, for $j = 1,...,k$, be such that

$$
d_i < \begin{cases} \frac{1}{2} \; if \; |v_j| < 1 \\ \frac{1}{4} \; if \; |v_j| = 1 \end{cases}
$$

hence, the k-factor Gegenbauer process $(X_t)_{t\in Z}$ defined in equation 3.4.5

has the following properties:

1. $(X_t)_{t\in Z}$ is stationary, casual and invertible
2. If $d_j > 0$, for $i = 1,...,k$,then $(X_t)_{t\in Z}$ is long memory
3. The spectral density $f_X(\lambda)$ of $(X_t)_{t\in Z}$ equals

$$f_X(\lambda) = \frac{\sigma_{\in}^2}{2\pi}\prod_{j=1}^{k}|2cos(\lambda)-v_i|^{-2d_j} \quad (3.4.6)$$

$$= \frac{\sigma_{\in}^2}{2\pi}\prod_{j=1}^{k}|4sin(\frac{\lambda+\lambda_j}{2})sin(\frac{\lambda-\lambda_j}{2}|^{-2dj} \quad (3.4.7)$$

where $0 \leq \lambda \leq \pi$ and for $j = 1,...,k$, $\lambda_j = cos^{-1}(v_j)$ are called Gegenbauer frequencies.

3.4.2 The k-factor GARMA process for modeling this inflation series is given by the following equation:

$$\phi(B)\prod_{j=1}^{k}(I-2v_jB+B^2)^{d_j}(X_t-\mu) = \theta(B)\varepsilon_t \quad (3.4.8)$$

where $0 < d_j < \frac{1}{2}$ if $|v_j| < 1$ or $0 < d_j < \frac{1}{4}$ if $|v_j| = 1 \ \forall \ j = 1,...,k$. (For stationarity and invertibility conditions refer to Gray *et al.* (1989))

3.4.3 Properties of the k-factor GARMA

It has been shown that (see Gray *et al.* (1989)) that:

1. the k-factor GARMA process is stationary if:

- $|v_j| < 1$ and $d_j < \frac{1}{2}$ or
- $|v_j| = 1$ and $d_j < \frac{1}{4}$, $\forall j = 1,...,k$

2. the k-factor GARMA process is invertible if:

- $|v_j| < 1$ and $d_j > -\frac{1}{2}$ or
- $|v_j| = 1$ and $d_j > -\frac{1}{4}$, $\forall j = 1,...,k$

3. the stationary k-factor GARMA will have long memory property if:

- $|v_j| < 1$ and $0 < d_j < \frac{1}{2}$ or
- $|v_j| = 1$ and $0 < d_j < \frac{1}{4}$, $\forall j = 1,...,k$

3.4.4 WHITTLE ESTIMATION

This methodology is used to obtain approximate maximum-likelihood estimates and is based on the calculation of the periodogram by means of the fast Fourier transform (FFT) and the use of the so-called Whittle approximation of the Gaussian log-likelihood function. Since the calculation of the FFT has a numerical complexity of order $O[nlog_2(n)]$ this approach produces very fast algorithms for computing parameter estimates (Palma, 2007).

Suppose that the sample vector $y = (y_1, ..., y_n)'$ is normally distributed with zero mean and variance Γ_θ. Then, the log-likelihood function divided by the sample is given by

$$\mathscr{L}(\theta) = -\frac{1}{2n} log\, det\, \Gamma_\theta - \frac{1}{2n} y \Gamma_\theta^{-1} y \tag{3.4.9}$$

Notice that the variance-covariance matrix re may be expressed in terms of the spectral density of the process $f_\theta(.)$ as follows:

$$(\Gamma_\theta)_{ij} = \gamma_\theta(i-j), \tag{3.4.10}$$

where

$$\gamma_\theta(k) = \int_{-\pi}^{\pi} f_\theta(\lambda) exp(i\lambda k) d\lambda \tag{3.4.11}$$

In order to obtain the Whittle method, two approximations are made. Since

$$\frac{1}{n} log\, det\, \Gamma_\theta \to \frac{1}{2\pi} \int_{-\pi}^{\pi} log[2\pi f_\theta(\lambda)] d\lambda$$

as $n \to \infty$, the first term in equation 3.4.9 is approximated by

$$\frac{1}{2n} log\, det\, \Gamma_\theta \approx \frac{1}{4\pi} \int_{-\pi}^{\pi} log[2\pi f_\theta(\lambda)] d\lambda$$

On the other hand, the second term in equation 3.4.9 is approximated by

$$
\begin{aligned}
\frac{1}{2n} y \Gamma_\theta^{-1} y &\approx \sum_{l=1}^{n}\sum_{j=1}^{n} y_l \left\{ \frac{1}{8\pi^2 n} \int_{-\pi}^{\pi} f_\theta^{-1}(\lambda) exp[(i\lambda(l-j)] d\lambda \right\} y_j \\
&= \frac{1}{8\pi^2 n} \int_{-\pi}^{\pi} f_\theta^{-1}(\lambda) \sum_{l=1}^{n}\sum_{j=1}^{n} y_l y_j) exp[(i\lambda(l-j)] d\lambda \\
&= \frac{1}{8\pi^2 n} \int_{-\pi}^{\pi} f_\theta^{-1}(\lambda) |\sum_{j=1}^{n} y_l exp(i\lambda j)|^2 d\lambda \\
&= \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\theta(\lambda)} d\lambda,
\end{aligned}
$$

where

$$
I(\lambda) = \frac{1}{2\pi n} |\sum_{j=1}^{n} y_j e^{i\lambda j}|^2
$$

is the periodogram of the series $\{y_t\}$.

Thus, the log-likelihood function is approximated, up to a constant, by

$$
\mathscr{L}_3(\theta) = -\frac{1}{4\pi}[\int_{-\pi}^{\pi} log\, f_\theta(\lambda) d\lambda + \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\theta(\lambda)} d\lambda] \tag{3.4.12}
$$

The evaluation of the log-likelihood function in equation 3.4.12 requires the calculation of integrals. To simplify this computation, the integrals can be substituted by Riemann sums as follows:

$$
\int_{-\pi}^{\pi} log\, f_\theta(\lambda) d\lambda \approx \frac{2\pi}{n} \sum_{j=1}^{n} log\, f_\theta(\lambda_j)
$$

and

$$
\int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\theta(\lambda)} d\lambda] \approx \frac{2\pi}{n} \sum_{j=1}^{n} \frac{I(\lambda_j)}{f_\theta(\lambda_j)},
$$

where $\lambda_j = \frac{2\pi j}{n}$ are the Fourier frequencies. Thus, a discrete version of the log-likelihood function in equation 3.4.12 is

$$\mathscr{L}_4(\theta) = -\frac{1}{2\pi}[\sum_{j=1}^{n} log\, f_\theta(\lambda) + \sum_{j=1}^{n} \frac{I(\lambda_j)}{f_\theta(\lambda_j)} \tag{3.4.13}$$

Other versions of the Whittle likelihood function are obtained by making additional assumptions. For instance, if the spectral density is normalized as

$$\int_{-\pi}^{\pi} log\, f_\theta(\lambda) d\lambda = 0, \tag{3.4.14}$$

then the Whittle log-likelihood function is reduced to

$$\mathscr{L}_5(\theta) = -\frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\theta(\lambda)} d\lambda] \tag{3.4.15}$$

with the corresponding discrete version

$$\mathscr{L}_6(\theta) = \frac{1}{2n}\sum_{j=1}^{n} \frac{I(\lambda_j)}{f_\theta(\lambda_j)} \tag{3.4.16}$$

Theorem 1: Let $\hat{\theta_n^{(i)}}$ be the value that maximizes the log-likelihood function $\mathscr{L}_i(\theta)$ for $i = 3, ..., 6$ for a Gaussian process $\{y_t\}$. Then, under some regularity conditions, m$\hat{\theta_n^{(i)}}$ is consistent and $\sqrt{n}[\hat{\theta_n^{(i)}} - \theta_0] \to N[0, \Gamma(\theta_0)^{-1}]$ as $n \to \infty$, where $\Gamma(\theta)$ is the matrix defined in equation 3.4.17 below.

$$\Gamma_{ij}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} [\frac{\partial log\, f_\theta(\lambda)}{\partial \theta_i}][\frac{\partial log\, f_\theta(\lambda)}{\partial \theta_j}]d\lambda \tag{3.4.17}$$

where $f_\theta$is the spectral density of the process refer to [Taqqu (1986) and Dahlhaus (1989)].

For non-gaussian processes we can still find well-behaved Whittle estimates (see Giraitis and Surgailis (1990).

3.4.5 Forecast Performance

We will assess the predictive performance of each model by using the root mean square error (RMSE) of prediction, the mean squared error (MSE), the mean absolute deviation MAD and Theil's U, .These performance measures are defined in the following equations:

$$RMSE = \sqrt{\frac{1}{h}\sum_{l=1}^{h}(X_{t+1} - \hat{X}_t(l))^2} \quad (3.4.18)$$

$$MSE = \frac{\sum(X_t - \hat{X}_t(l))^2}{n} \quad (3.4.19)$$

$$MAD = \frac{|X_{t+l} - \hat{X}_t(l)|}{h} \quad (3.4.20)$$

$$U = \sqrt{\frac{\sum_{t=1}^{n-1}(\frac{\hat{X}_{t+1} - X_{t+1}}{X_t})^2}{\sum_{t=1}^{n-1}(\frac{X_{t+1} - X_t}{X_t})^2}} \quad (3.4.21)$$

where X_t is the actual value of a point for a given time period t, n is the number of data points, and $\hat{X}_t$ is the forecast value.

Chapter 4

Applications

In this section, the Gambia inflation rate data is used for empirical modeling and tests conducted on the two models defined in the previous section.

The graphical inspection of the inflation series from Figure 3.1 suggests that this series might not be stationary as it exhibits certain amount of volatility with a trend cycle. In the unit root tests that follow both constant and trend are included. Using the HEGY test for seasonal roots (Hylleberg *et al.*, 1990), with results presented in Table 4.1 it is concluded that the null hypothesis of a unit root at zero frequency cannot be rejected. In addition, both the ADF (Dickey and Fuller, 1979) and KPSS (Kwiatkowski *et al.*, 1992) tests (see Table 4.2) give the same result as the HEGY. However, the presence of seasonal unit roots is rejected at all frequencies at the 5% significance level.

Table 4.1: HEGY Seasonal Unit Root plus trend for inflation rates

			Constant+Trend
Auxillary Regression	Seasonal Frequency	Critical Values	Test statistic
t-test:$\pi_1 = 0$	0	-3.37	-0.7762
t-test;$\pi_2 = 0$	π	-1.94	-6.1885*
F-test:$\pi_3 = \pi_4 = 0$	$\frac{\pi}{2}$	3.05	53.6823*
F-test:$\pi_5 = \pi_6 = 0$	$\frac{2\pi}{3}$	3.05	59.1069*
F-test:$\pi_7 = \pi_8 = 0$	$\frac{\pi}{3}$	3.08	45.1287*
F-test:$\pi_9 = \pi_{10} = 0$	$\frac{5\pi}{6}$	3.08	67.8777*
F-test:$\pi_{11} = \pi_{12} = 0$	$\frac{\pi}{6}$	3.09	33.2156*
F-test:$\pi_1 = ...\pi_{12} = 0$		1.88	61.5118*
F-test:$\pi_2 = ...\pi_{12} = 0$		2.30	64.1506*

*Seasonal unit root is rejected at the 5% significance level

Table 4.2: Unit root test for inflation in level form

	Constant		Constant+Trend	
Test type	Critical Value	Test Statistic	Critical Value	Test Statistic
ADF	-2.878	-2.601	-3.428	-2.555
KPSS	0.463	0.9014	0.148	0.3727

Test of the null hypothesis of a unit root with test statistic and critical values

4.1 Model Selection

The proper order of differencing (both seasonal and nonseasonal) has been decided from the previous tests .i.e. $d = 1, D = 0$. We now determine the appropriate autoregressive orders p and P and moving average orders q and Q. We will use the autocorrelation function (ACF) and partial autocorrelation (PACF) plots of the difference series to identify the numbers of AR and/or MA terms that are needed tentatively and conclude the order selection with the Hyndman-Khandakar (HK) algorithm. In Figure 2(b) the partial autocorrelation function (PACF) has a positive significant spike at lag 1 with several insignificant spikes before the seasonal lag suggesting that p is at least equal to 1. In Figure 2(a) the first 3 autocorrelations are relatively significant, that is, outside the confidence band suggesting that q is at least 3. For the seasonal orders, since autocorrelation at the seasonal period is negative,we consider adding an SMA term to the model, in this case SMA(1). To complete the model selection we use the Hyndman-Khandakar (HK) algorithm to consider the various possibilities. Using the HK-algorithm, the best model in terms of AIC, AIC_c, and BIC is the $ARIMA(1,1,1)(0,0,1)_{12}$ as shown in Table 4.3.

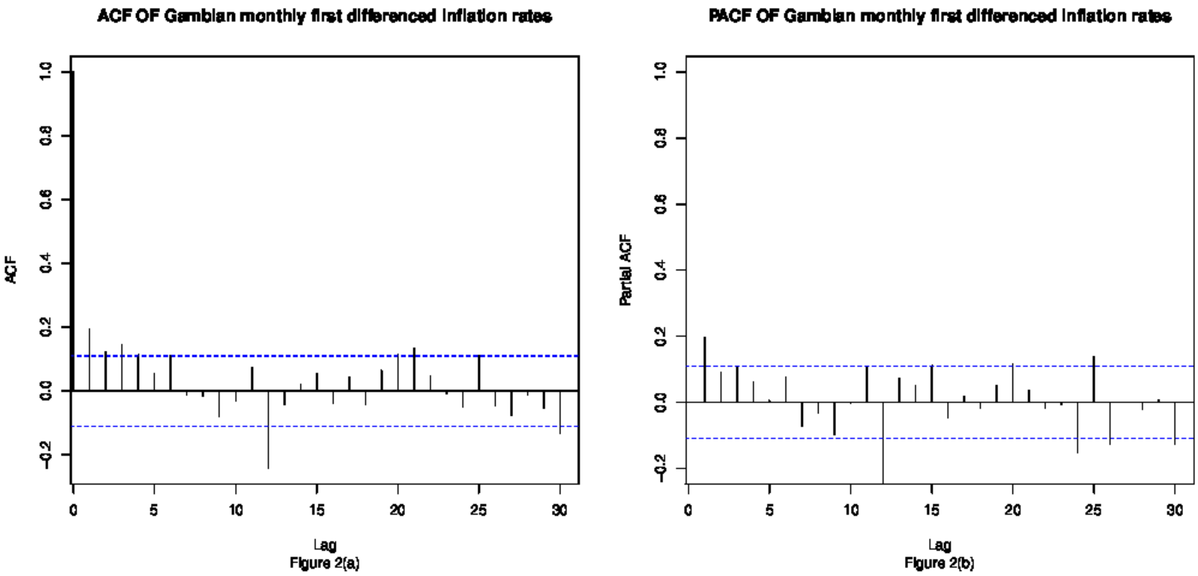


Figure 4.1: ACF and PACF of Gambian monthly first differenced inflation rates

Table 4.3: $SARIMA(p,d,q) \times (P,D,Q)_{12}$ chosen by each criterion(H-K Algorithm)

Criterion	AIC	BIC	AIC_c	(AIC,BIC,AIC_c)
Model chosen	$ARIMA(1,1,1)(0,0,1)_{12}$	$ARIMA(0,1,0)(0,0,1)_{12}$	$ARIMA(1,1,1)(0,0,1)_{12}$	$ARIMA(1,1,1)(0,0,1)_{12}$

Table 4.4: AIC and BIC for the Suggested SARIMA Models

Model	AIC	AIC_c	BIC
$ARIMA(1,1,1)(0,0,1)_{12}$	1488	1488	1507
$ARIMA(0,1,0)(0,0,1)_{12}$	1495	1495	1503

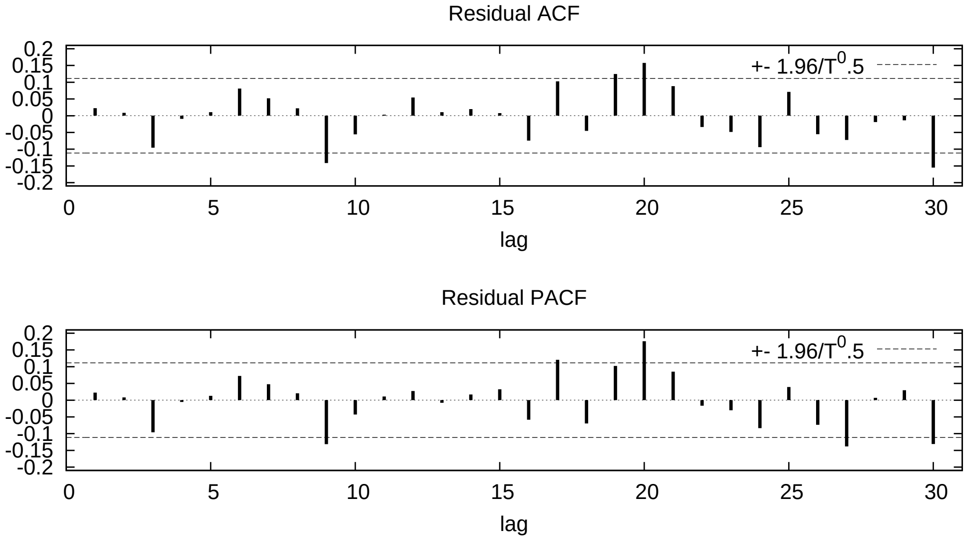
The estimated parameters of the model are presented in Table 4.5.

Table 4.5: Estimates of $SARIMA(1,1,1) \times (0,0,1)$

Parameter	Estimate	Standard Error	P-value
Constant	-0.014	0.0175	
$\hat{\phi}_1$	0.6128	0.0143	0.0000^{***}
$\hat{\theta}_1$	-0.5941	0.0386	0.0000^{***}
$\hat{\Theta}_1$	-0.5606	0.0383	0.0000^{***}
$\hat{\sigma}^2_{\varepsilon}$	2.0512		

***parameters are statistically significant

Figure 4.2: Residual ACF and PACF



The ACF of the residuals displayed in Figure 4.2 depicts that the autocorrelation of the residuals are approximately all zero, that is to say they are uncorrelated and the p-value for the Ljung-Box statistic in Table 4.6 clearly exceed 5% at lag 12, indicating that there is no significant departure from white noise for the residuals. The ARCH–LM test is used to test for constant variance. From the ARCH–LM test results shown in Table 4.7, we fail to reject the null hypothesis of no ARCH effect (homoscedasticity) in the residuals of the selected model.

Table 4.6: Box-Ljung test of the null hypothesis of randomness for the residuals

lag	Q'	P-value
12	14.61	0.1024

Table 4.7: ARCH–LM Test for Homoscedasticity

Model	P-value
$ARIMA(1,1,1)(0,0,1)_{12}$	0.0659*

*Null hypothesis: no arch effect is present

The spectral density in Figure 3.1 is unbounded at the low frequencies which suggest that the inflation series seems to be a long memory process with fractional integration behavior (Woodward *et al.*, 2011). The empirical ACF of the series in Figure 3.1 clearly shows a strong dependence between distant observations, as well as a cyclical behavior with a pretty long period. Moreover, the spectral density of the series clearly possesses three distinct peaks with the first peak having frequency located very close, but not necessarily equal, to zero which are all distinctive features of a weakly stationary Gegenbauer process. The properties of the data suggest using a 3-factor Gegenbauer process to model inflation in Gambia.

We assume the inflation series has the following specification:

$$\prod_{j=1}^{3}(I-2v_jB+B^2)^{d_j}(X_t-\mu) \quad = \quad \varepsilon_t \tag{4.1.1}$$

where $\lambda_j = cos^{-1}(v_j)$, $0 \leq \lambda \leq \pi$ are called the Gegenbauer frequencies.

The following results were obtained from the estimation of the long memory model, the 3-factor Gegenbauer model.The parameter estimation is done using the parametric Whittles method. The three orders of integration are $d_1 = 0.0117$ $d_2 = 0.2744$ $d_3 = 0.0831$, with $LLF = 1.9098$. (see Figure 4.3 for the residual correlogram).

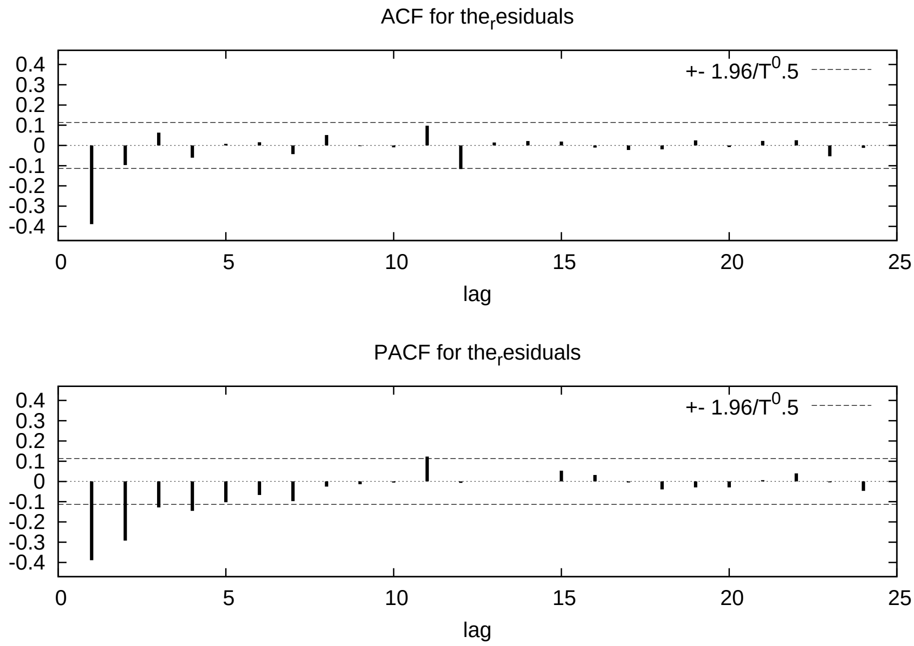


Figure 4.3: residual ACF and PACF

From the autocorrelation function in Figure 4.3 it is obvious that the residuals are not white noise. An ARCH-LM test performed on the residuals fail to reject evidence of

ARCH effect. We adjust various $ARMA(p,q)$ specifications on the residuals and found $ARMA(0,1)$ to be most adequate with the parameter $\theta_1 = -0.654$.

The resulting k-factor $GARMA$ model is given below.

$$
\begin{aligned}
&(I-2(0.9992)B+B^2)^{0.0117}(I-2(0.9951)B+B^2)^{0.2744}) \\
&\qquad \times(I-2(0.9982)B+B^2)^{0.0831}(X_t-\mu) \;=\; (1-0.654B)\varepsilon_t
\end{aligned}
\tag{4.12}
$$

We finally performed Box-Pierce and Ljung-Box- Pierce tests on the estimated residuals and found no evidence of autocorrelation in the model.

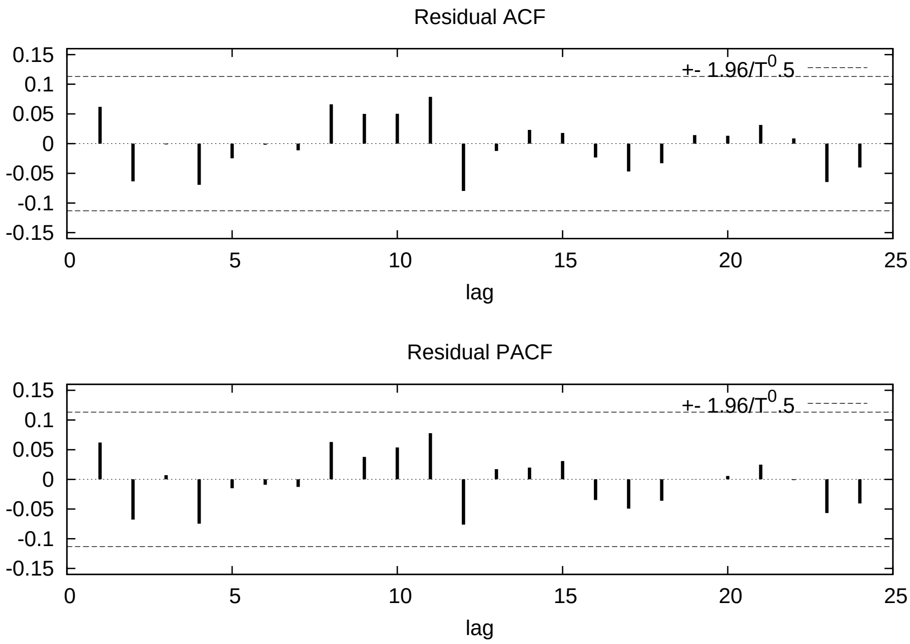


Figure 4.4: Behavior of ACF and PACF the distribution of the estimated residuals

4.2 Model Comparison

When comparing the two models using the AIC, Schwarz Criterion, Hannan-Quinn Criterion and the standard deviation of innovation it is obvious that the SARIMA model is better than the k-factor GARMA. See Table 4.8 for the in-sample characteristics of the two models.

Table 4.8: In-sample comparison

	SARIMA model	3-factor GARMA model
Akaike Criterion	1081.3	1651.0
Schwarz Criterion	1099.8	1662.1
Hannan-Quinn Criterion	1088.7	1655.4
S.D. of innovations	1.4601	3.7533

To complete the study, the forecasting performance of the two models are compared using standard measures of forecast accuracy such as the following: the root-mean-squared error (RMSE), the mean-squared error (MSE), the mean absolute deviation (MAD) and Theil's U (Brockwell and Davis, 1996).

The data set was divided into two sets, the first set for model estimation and the preserved observations for out-of-sample forecasting. Table 4.9 shows the forecast evaluation statistics for both the $ARIMA(1,1,1)(0,0,1)_{12}$ and 3-factor *GARMA* model. The $ARIMA(1,1,1)(0,0,1)_{12}$ recorded MSE, RMSE, MAE and Theil's U of 1.3781, 1.1739, 0.8674, and 4.4847 respectively while the 3-factor *GARMA* recorded recorded MSE, RMSE, MAE and Theil's U of 1.2320, 1.1100, 0.9870 and 0.0952 respectively. The 3-factor *GARMA* has the minimum values of MSE, RMSE, and Theil's U compared to the seasonal $ARIMA(1,1,1)(0,0,1)_{12}$ model , hence the conclusion is that the 3-factor *GARMA* performs better than the seasonal ARIMA model in out-of-sample forecasting.

Table 4.9: Forecast results for monthly Gambia inflation time series

Model	MSE	RMSE	MAE	Theil's U
$ARIMA(1,1,1)(0,0,1)_{12}$	1.3781	1.1739	0.8674	4.4847
3-factor *GARMA*	**1.2320**	**1.1100**	0.9870	**0.0952**

Chapter 5

Summary and Conclusion

In this study the statistical properties of the Gambia inflation rates is investigated and two models specified namely seasonal autoregressive integrated moving average (SARIMA) and k-factor Gegenbauer Autoregressive Moving Average (k-factor GARMA). The data set was divided into two and $ARIMA(1,1,1)(0,0,1)_{12}$ and 3-factor *GARMA* fit to the first data set of the inflation series. The first model seasonal $ARIMA(1,1,1)(0,0,1)_{12}$ was selected using the H-K Algorithm developed by Hyndman and Khandakar (2008) and 3-factor *GARMA* from both spectral density graph and further analysis of the residuals from the 3-factor Gegenbauer model. The in-sample characteristics such as log-likelihood, Akaike Criterion, Schwarz, Hannan-Quinn, and S.D. of innovations following estimation have show that the $ARIMA(1,1,1)(0,0,1)_{12}$ outperforms the 3-factor *GARMA* model. The second data set (preserved) was used for out-of-sample forecasting and the forecast evaluation statistics such as the MSE, RMSE, and Theil's U suggest that the 3-factor *GARMA* model outperforms the seasonal $ARIMA(1,1,1)(0,0,1)_{12}$ model in out-of-sample forecasting. The results indicated that inflation in Gambia is stationary with long-memory behavior at three distinct frequencies. It is also found that the k-factor GARMA outperforms the seasonal ARIMA in out-sample forecasting which may be ascribed to the forecast horizon been large and series strongly long-range dependent.

It is important to note that the SARIMA methodology has certain limitations. For instance, it requires large number of observations for model identification and sometimes estimation and selection involves some form of art. Also, differencing the series in the case of nonstationarity may reduce the available information set. However, the model is parsimonious with respect to the coefficients and good in providing unconditional forecasts. Note also that the number of peaks to be chosen in the spectral density for the Gegenbauer model remains unclear and must be investigated.

Bibliography

Adjepong, M. O., Oduro, F. T., and Oduro, S. D. (2013). Determing the better approach for short-term forecasting of ghana's inflation: Seasonal-arima vs. holt-winters. *International Journal of Business, Humanities and Technology*, 3(1).

Aidoo, E. (2010). Modelling and forecasting inflation rates in ghana: An application of sarima models. Master's thesis, Dalarna University, School of Technology and Business Studies, Statistics.

Alnaa and Ferdinand (2011). Arima approach to predicting inflation in ghana. *Journal of Economics and International Finance*, 3(5):328–336.

Backus, D. and Zin, S. (1993). Long memory inflation uncertainty. evidence from the term structure of interest rates. *Money, Credit and Banking*, 25:681–700.

Barkoulas, J. T. and Baum, C. F. (1996). Long term dependence in stock returns. *Economics Letters*, 53(3):253–59.

Barros and Gil-Alana (2012). Inflation forecasting in angola: A fractional approach. Working Paper 103, Instituto Superior de Economia e Gestã.

Bisaglia, L. and Guégan, D. (1998). A comparison of techniques of estimation in long-memory processes: Application to intra-day. *Computational Statistics and Data Analysis*, 27:61–81.

Box, G. and Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

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

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting.* Springer, New York, second edition.

Canova, F. and Hansen, B. (1995). Are seasonal patterns constant over time? a test for seasonal stability. *Journal of Business and Economic Statistics*, 13:237–252.

Caporale, G. M. and Gil-Alana, L. A. (2011). Multi-factor gegenbauer processes and european inflation rates. *Journal of Economic Integration*, 26(2):386–409.

Cheung, Y. W. (1993). Long memory in foreign-exchange rates. *Journal of Business and Economic Statistics*, 11(1):93–101.

Cheung, Y. W. and Lai, K. (1995). A search of long memory in international stock market returns. *Journal of International Money and Finance*, 14:597–615.

Delgado, M. and Robinson, P. (1994). New methods for the analysis of long memory time series: Application to spanish inflation. *Journal of Forecasting*, 13:97–107.

Dickey, D. and Fuller, W. (1979). Distribution of the estimators for autoregressive time series with unit root. *Journal of the American Statistical Association*, 74:427–431.

Engel, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of uk inflations. *Econometrica*, 50:987–1007.

Ferrara, L. and Guegan, D. (1999). Estimation and applications of gegen-bauer processes. Technical Report 9927, CREST-INSEE.

Franses and Hobijn (1997). Critical values for unit root tests in seasonal time series. *Journal of Applied Statistics*, 24(1):25–47.

Franses, P. (1990). Testing for seasonal unit roots in monthly data. Technical report 9032, Econometric Institute.

Gray, H. L., F, N., and Woodward, W. A. (1989). On generalized fractional processes. *Journal of Time Series Analysis*, 10:233–257.

Hasser (1993). Regression of spectral estimators with fractionally integrated time series. *Journal of Time Series Analysis*, 14:369–380.

Hassler, U. and Wolters., J. (1995). Long memory in inflation rates. international evidence. *Journal of Business and Economic Statistics*, 13:37–45.

Hipel, K. W. and McLeod, A. I. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–799.

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

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

Johnston, J. and Dinardo, J. *Econometric Methods*. McGraw-Hill, fourth edition.

Junttila, J. (2001). Structural breaks, arima model and finnish inflation forecasts. *International Journal of Forecasting*, 17:203–230.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54:159–178.

Ljung and Box (1978). On a measure of lack of fit in time-series models. *Biometrika*, 66:297–303.

Noakes, D. J., Hipel, K. W., McLeod, A. I., Jimenez, C., and Yakowitz, S. (1998). Forecasting annual geophysical time series. *International Journal of Forecasting*, 4.

Ooms, M. and Franses, P. H. (1998). A seasonal periodic long memory model for monthly river flows. Technical report, Econometric Institute, Erasmus University Rotterdam.

Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. A John Wiley and Sons , INC., Publication.

Pincheira, P. and Medel, C. A. (2012). Forecasting inflation with a simple and accurate benchmark: A cross-country analysis. Working Papers 677, Central Bank of Chile.

Pufnik and Kunovac (2006). Short-term forecasting of inflation in croatia with seasonal arima processes. Working Paper W-16, Croatia National Bank.

Sutcliffe, A. (1994). Time series forecasting using fractional differencing. *Journal of Forecasting*, 13:383–393.

Taqqu, M. S., Willinger, W., and Sherman, R. (1997). Proof of a fundamental result in self-similar traffic modeling. *Computer Communication Review*, 27:5–23.

Velasco, C. (1999). Gaussian semi-parametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20(1):87–127.

Willinger, W., Taqqu, M. S., and Teverovsky, V. (1999). Stock market prices and long range dependence. *Finance and Stochastics*, 3:1–13.

Woodward, W. A., Gray, H. L., and Elliott, A. C. (2011). *Applied Time Series Analysis*. Statistics: A series of Textbooks and Monographs. CRC Press.