ARIMA FORECASTS OF THE NUMBER OF BENEFICIARIES OF SOCIAL SECURITY GRANTS IN SOUTH AFRICA

by

FULULEDZANI LUCY LURULI

submitted in accordance with requirements for the degree of
MASTER OF SCIENCE
in the subject
STATISTICS
at the
UNIVERSITY OF SOUTH AFRICA

Supervisor: PROFESSOR P NDLOVU

DECEMBER 2011

Abstract

The main objective of the thesis was to investigate the feasibility of accurately and precisely forecasting the number of both national and provincial beneficiaries of social security grants in South Africa, using simple autoregressive integrated moving average (ARIMA) models. The series of the monthly number of beneficiaries of the old age, child support, foster care and disability grants from April 2004 to March 2010 were used to achieve the objectives of the thesis. The conclusions from analysing the series were that: (1) ARIMA models for forecasting are province and grant-type specific; (2) for some grants, national forecasts obtained by aggregating provincial ARIMA forecasts are more accurate and precise than those obtained by ARIMA modelling national series; and (3) for some grants, forecasts obtained by modelling the latest half of the series were more accurate and precise than those obtained from modelling the full series.

Key terms

Social security grants; Autoregressive integrated moving average models; Forecasts; Social Pension System; Portmanteau test; Model identification; Standard error; Mean square error; Autocorrelation function; Akaike's information criterion.

Contents

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to:

The almighty **God**, for inspiration and guidance throughout this study.

The following individuals, who took the time and effort to assist in the development of this document and without whom I would not have succeeded:

My supervisor, **Prof P Ndlovu**, whose advice provided me with inspiration when most needed and whose assistance made a great impact on my study.

I would like to acknowledge the South African Social Security Agency (SASSA) and National Department of Social Development (NDSD) for allowing me to use their data.

Alexa Barnby for professional language editing.

My mother, Avhazwifuni Annah Khethani, my late father, Mphatheleni Nelson Khethani and my husband, Moleko Abbey Luruli, for their constant emotional support, advice and availability, as well as for their interest in my work and their belief in me. My son, Mpho Luruli, your existence encouraged me to work harder.

Finally, I thank all those whose direct and indirect advice has helped me to complete this dissertation.

DEDICATION

To my father, Mphatheleni Nelson Khethani (1938-2002), my mother, Avhazwifuni Annah Khethani, my husband, Moleko Abbey Luruli, and my son, Mpho Luruli

DECLARATION BY STUDENT

The research work illustrated in this thesis was carried out at the University of South Africa, in the Department of Statistics, under the supervion of Prof. Principal Ndlovu.

I, Fululedzani Lucy Luruli, declare that this thesis is mine and it has not been submitted to any other university.

Declared on the (date):

Signed:

Name: **Mrs Fululedzani Lucy Luruli**

Declared on the (date):

Signed:

Name: **Prof. Principal Ndlovu** (Supervisor)

Chapter 1

Introduction

1.1 Background

The South African Social Security Agency (SASSA) of the National Department of Social Development (NDSD) was established in April 2006 for the purpose of implementing the norms and standards set by the NDSD for improving the delivery of social assistance grants to deserving South Africans. Deserving South Africans are those individuals and families who are unable to avoid poverty because of various disabilities and circumstances. The purpose of giving social security grants to these South Africans is to prevent and alleviate poverty for social compensation and income distribution.

In 1995, there were 2 848 344 beneficiaries of all types of social security grants. This means that about 7 out of every 100 South Africans were the recipients of government social assistance grants. At that time, the amounts of social assistance grants were very low in relation to the cost of living. Previously, the types and amounts of social assistance grants, as well as the range of social services available, were determined on racial grounds. Furthermore, many people were unaware of their rights to social security. Parity in the amounts of all grants given to people was achieved in 1993. Subsequently, in 1997, regulations were formulated for the uniformity and integration of grants (White paper for social welfare; 1997).

The constitution of the Republic of South Africa (Act No. 108; 1996) affirms that everyone has a right to have access to social security if they are economically unable to support themselves. The constitution is the highest law in South Africa. Therefore in attempting to fulfill the obligations of the constitution the NDSD has actively engaged in providing grants to eligible beneficiaries. A grant is a cash transfer from the state to people who are unable to provide for themselves. It is paid to poor people to provide for basic necessities that may have a basic quality of life in sustaining a minimum livelihood. Grants are provided in support of the mission of the Department of Social Development, which is to empower the poor and to secure a better life for those in need. A person is entitled to the appropriate social assistance if he or she, in terms of section 17, is resident in the republic, and is a South African. These are requirements which have to be met in order to qualify for social assistance grant. Means testing officers check if the applicants meet the requirements to qualify for the social assistance grants.

Currently in South Africa there are eight social security grants, namely, old age, child support, disability, foster child, care dependency, war veteran, grant-in-aid and social relief of distress grants. Three of these grants have been developed especially for children, namely, the foster care grant, the care dependency grant and the child support grant. According to Social Assistance Act No. 13; 2004, a person is eligible for a disability grant, if he or she has attained the prescribed age, and is owing to a physical or mental disability, unfit to obtain by virtue to any service, employment or profession the means needed to enable him or her to provide for his or her maintenance. There are two types of disability grants, permanent disability grant, this grant paid to the people who are permanently disabled where the disability will continue for more than 12 months. Temporary disability grant, this grant also paid to people whose disability is temporary where they are disabled for not less than 6 months and no longer than 12 months. The temporary disability grant lapses anytime if the Doctor confirms that this grant beneficiary is no longer disabled.

Child support grant means a person is eligible for a child support grant if he or she is a parent, primary care giver of that child. The child support grant lapses when the child turns 18 years of age. A person is eligible to older persons grant if in the case of a woman, she has attained the age of 60 years and in the case of a man, he has attained the age of 65 years. Social assistance amendment (Act No. 6; 2008) confirms that a man qualified for older person's grant after 1 April 2008 if he attained the age of 63 years, 1 April 2009 if he attained the age of 61 years and 1 April 2010 if he attained the age of 60 years.

Foster care grant means any child who is in need of care and has been placed in the custody of any foster parent. Foster parent, means any person, except a parent or guardian, in whose custody a child has been placed in terms of (Criminal Procedure Act No. 51; 1977). The foster parents and children need not be necessarily be South African citizens, the child must be under 18 years, the foster parents and the foster child must live in South Africa during the time of the application. The foster care grant lapses when the child turns 18 years of age, the eligible of this grant can be extended to 21 years if the child provide the proof of school registration.

The NDSD has now decided to address the problem of the huge unexpected increase in the number of social security grant beneficiaries, from 6.4 million in 2004 to 14 million in 2010. One of the problems to be addressed is that of obtaining accurate and precise forecasts of the number of beneficiaries of the various types of social security grant. Such forecasts would help the NDSD and the National Department of Treasury (NDT) in making future policy decisions, drawing up management plans and doing financial planning.

One of the proposed methods of forecasting the number of beneficiaries of the various types of social security grant is here termed autoregressive integrated moving average (ARIMA) forecasting. This is not a new forecasting method. Reikard (2009), for example, used ARIMA forecasting to obtain short-term forecasts of daily solar power radiation for operational planning, switching sources, programming back up, short-term power purchases, as well as for planning for reserve usage and peak load matching. This author reported that the ARIMA model with time varying coefficients obtained accurate forecasts mainly due to the model's ability to capture the diurnal cycle more effectively than other methods. In this thesis, the performance of the ARIMA forecasting method in obtaining accurate and precise forecasts of the number of beneficiaries of social security grants in South Africa is investigated.

1.2 Objectives of the thesis

The major objective of this thesis is to investigate the feasibility of accurately and precisely forecasting the number of both national and provincial beneficiaries of various types of social grant using autoregressive integrated moving average (ARIMA) models. In particular, the objectives of the thesis are the following:

1. Find simple ARIMA models for accurately and precisely forecasting the number of both national and provincial beneficiaries of various types of social grant. These models, if found, will be used by the NDSD and the NDT to obtain forecasts for planning and making policy decisions.

2. Confirm the hypothesis that the simple ARIMA models for accurately and precisely forecasting the number of provincial beneficiaries of various types of social grant are province and type of social grant specific. If confirmed, then for each type of social grant, separate models for forecasting the number of national beneficiaries and for forecasting the number of beneficiaries in each province would have to be developed for use by the NDSD and the NDT.

3. Compare the accuracy and the precision of the forecasts of the number of national beneficiaries of social grants obtained using simple ARIMA models with those obtained by aggregating forecasts of the number of provincial beneficiaries of social grants. If the latter forecasts of the number of national beneficiaries of social grants turn out to be more accurate and precise, then the NDSD and the NDT will be advised to use these forecasts rather than those obtained from ARIMA modelling the number of national beneficiaries of social grants.

4. Investigate the effect of the length of social grant series on the accuracy and precision of the forecasted numbers of national and provincial beneficiaries. The null hypothesis here is that there is no difference between national beneficiaries and provincial beneficiaries of social grant series in terms of the accuracy and precision of forecasts. The alternative hypothesis is that there is a difference between the national beneficiaries and the provincial beneficiaries of social grant series on the accuracy and precision of forecasts.

1.3 Data

The data used in this thesis to achieve the objectives mentioned in Section 1.2 were obtained from SASSA. The data consist of records of the monthly and yearly numbers of national beneficiaries (Republic of South Africa)[RSA] and provincial beneficiaries (Eastern Cape [EC], Free State [FS], Gauteng [GP], KwaZulu-Natal [KZN], Limpopo [LP], Mpumalanga [MP], North West [NW], Northern Cape [NC] and Western Cape [WC]) of various types of social grant from April 2004 to March 2010. The social grant types in the data are the care dependency grant, the child support grant, the foster care grant, the grant-in-aid, the old age grant, the disability grant and the war veteran grant. The data exclude the grant for social relief of distress, because this grant is paid out to beneficiaries only in the event of a disaster. Therefore, in this thesis only the old age, the child support, the foster care, and the disability grants were investigated.

The data used in this thesis refer to the social security grant and are available in the Social Pension System (SOCPEN) of SASSA. These data were collected from the provinces of South Africa (SA), which in turn were collected from provincial district and regional offices. The reliability of the SOCPEN data depends on the fact that the applicants for social security grants must produce their national identity books when submitting the application forms and that SASSA officers conduct a means tests on the applicants. Furthermore, all new applicants as well lapsed beneficiaries owing to death, children being over the age limit, and the recovery of the previously disabled, are recorded by SASSA officers in the SOCPEN (Social Pension System Database; 2010).

1.4 Organisation of the thesis

This thesis consists of four chapters. This chapter gave some background about the social security grants in South Africa and discussed the objectives of the thesis. The chapter also briefly reviewed related studies. Chapter 2 describes the theories of ARIMA modelling and forecasting. In particular, the theories of ARIMA model identification, model estimation methods, model diagnostics and forecasting are discussed. Chapter 3 examines ARIMA modelling and forecasting of the number of national and provincial beneficiaries of the selected social security grants. This is done with the help of the Proc ARIMA procedure in SAS Version 9.2 (2004). The results of the analysis are used to answer the questions implied by the objectives of this thesis. Chapter 4 contains the conclusions reached and the recommendations made by this thesis.

Chapter 2

Theory of ARIMA Modelling and Forecasting

2.1 Introduction

The major objective of this thesis is to investigate the feasibility of accurately and precisely predicting (forecasting) the number of both national and provincial beneficiaries of various types of social grant using autoregressive integrated moving average (ARIMA) models. The theories relevant to ARIMA modelling and forecasting are discussed in this chapter. Definitions and notation in this regard are given in Section 2.2, the theory of ARIMA modelling is discussed in Section 2.3, model identification, estimation and diagnostics are discussed in Section 2.4 and forecasting is discussed in Section 2.5.

2.2 Definitions and notations

In discussing the theory of ARIMA modeling, the following definitions and notations will be used:

1. $\{X_t\}_{t=1}^{n}$ or simply X_t where t is time and represents a time series such as the number of beneficiaries of a social grant in year t.

2. X_{t-k}, where $k = 0, \pm1, \pm2, ...$ is called the lag period and is the observation/measurement at k units of time before time t if $k > 0$ or at k units of time after time t if $k < 0$.

For example, number of beneficiaries in k years before year t if $k > 0$ or in k years after year t if $k > 0$.

3. The **backshift operator** is $\mathbf{B}$ such that $\mathbf{B}^k\ X_t = X_{t-k}$ where $k = 0, \pm1, \pm2, ...$ is the lag period.

4. The autocorrelation function (ACF), $\rho_k = \text{Correlation}(X_t, X_{t+k})$ is given by

$$\rho_k = \frac{\gamma(t+k,t)}{\sqrt{\gamma(t+k,t+k)\gamma(t,t)}},\ \ k = 0, 1, 2, ... \tag{2.2.1}$$

where $\gamma(t+k,t) = \text{Covariance}(X_t, X_{t+k})$ and $\gamma(t+k,t+k) = \text{Variance}(X_{t+k})$, and $\gamma(t,t) = \text{Variance}(X_t)$ (Montgomery et al., 2008, pp28−30).

ρ_k measures the correlation between X_t and X_{t+k} and has the usual properties of correlations (Box, 1976, p30). The estimator of ρ_k is the sample ACF given by

$$\hat{\rho}_k = \frac{\hat{\gamma}(t+k,t)}{\sqrt{\hat{\gamma}(t+k,t+k)\hat{\gamma}(t,t)}}, \quad k = 0,1,2,... \tag{2.2.2}$$

where $\hat{\gamma}(t+k,t) = \frac{1}{n-k}\sum_{t=1}^{n-k}(X_t - \bar{X}_t)(X_{t+k} - \bar{X}_t)$, $\hat{\gamma}(t+k,t+k) = \frac{1}{n-k}\sum_{t=1}^{n-k}(X_{t+k} - \bar{X}_t)^2$ and $\hat{\gamma}(t,t) = \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X}_t)^2$.

5. A series X_t is said to be second-order stationary or covariance stationary if the ACF $\rho_k = \text{Correlation}(X_t, X_{t+k})$ does not depend on t but depends only on k. In this case (2.2.2) becomes

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X}_t)(X_{t+k} - \bar{X}_t)}{\sum_{t=1}^{n}(X_t - \bar{X}_t)^2}, \quad k = 1,2,... \tag{2.2.3}$$

Possible causes of the nonstationarity of X_t are the presence of trend (long-term change in $E[X_t]$), seasonality (constant changes in $E[X_t]$ at regular periods), and nonconstant variance of the random errors in X_t. Possible remedial measures to stationarise X_t are differencing X_t to remove the trend and/or seasonality (Chatfield, 2004, p19) and/or transforming X_t to achieve constant variance of the random component of the series (Diggle, 1990, p31).

6. Differencing is a special type of filtering, which is particularly useful for removing trend and/or seasonality in a series. For d a positive integer, the differencing filter $\nabla^d X_t = (1 - \mathbf{B})^d X_t$ removes the trend of the polynomial of degree d in the series X_t if present, and the differencing filter $(1 - \mathbf{B}^d)X_t = X_t - X_{t-d}$ removes seasonality of period d in the series X_t if present.

For example, if $X_t = \beta_0 + \beta t + \epsilon_t$, then $(1-\mathbf{B})\mathbf{X_t} = \mathbf{X_t} - \mathbf{X_{t-1}} = \beta_\mathbf{0} + \beta\mathbf{t} + \epsilon_\mathbf{t} - \beta_\mathbf{0} - \beta(\mathbf{t-1}) - \epsilon_\mathbf{t-1} = \beta + \epsilon_\mathbf{t} - \epsilon_\mathbf{t-1}$, is free of the linear trend.

2.3 Theory of ARIMA modelling

In this section, the theories of ARIMA modeling are briefly discussed. In the discussion, it is assumed that the series X_t is second-order stationary (see definition in Section 2.2, item 5). In this case, there are three basic types of ARIMA model, namely

- the autoregressive model of order p a positive integer, denoted by AR(p)
- the moving average model of order q a positive integer, denoted by MA(q)
- the autoregressive moving average of order (p,q), denoted by ARMA(p,q)

2.3.1 Autoregressive models of order p (AR(p))

The AR(p) model for a time series X_t is given by

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + \epsilon_t \tag{2.3.1}$$

where p is a nonnegative integer, $\alpha_1, \alpha_2, ..., \alpha_p$ are model parameters to be estimated, and ϵ_t is a series of random errors each with zero mean and constant variance σ^2. Using the backshift operator $\mathbf{B}$ (defined in Section 2.2, item 3), the AR(p) model may be written as

$$\alpha(\mathbf{B})X_t = \epsilon_t \tag{2.3.2}$$

where $\alpha(\mathbf{B}) = 1 - \alpha_1\mathbf{B} - \alpha_2\mathbf{B}^2 - ... - \alpha_p\mathbf{B}^p$ is a polynomial in $\mathbf{B}$ of order p. An AR(p) process is said to be stationary provided that the absolute roots of the polynomial in $\mathbf{B}$, $\alpha(\mathbf{B}) = 0$, are all greater than 1 (Wei, 2006, p40).

The ACF of a time series X_t that is generated by an AR(p) process decays exponentially with lag k. Thus, if a time series X_t is generated by an AR(p) process, then its sample ACF (see definition in Section 2.2, item 5) should decrease exponentially with lag k.

2.3.2 Moving average models of order q (MA(q))

The MA(q) model for a time series X_t is given by

$$X_t = \epsilon_t - \beta_1\epsilon_{t-1} - \beta_2\epsilon_{t-2} - ... - \beta_q\epsilon_{t-q}. \tag{2.3.3}$$

where q is a nonnegative integer, $\beta_1, \beta_2, ..., \beta_q$ are model parameters to be estimated, and ϵ_t is a series of random errors each with zero mean and constant variance σ^2. Alternatively, the MA(q) model may be written as

$$X_t = \beta(\mathbf{B})\epsilon_t \tag{2.3.4}$$

where $\beta(\mathbf{B}) = 1 - \beta_1\mathbf{B} - \beta_2\mathbf{B}^2 - ... - \beta_q\mathbf{B}^q$ is a polynomial in $\mathbf{B}$ of order q. An MA process of order q is invertible if the roots of the polynomial in $\mathbf{B}$, $\beta(\mathbf{B}) = 0$, all lie outside the unit circle (Box and Jenkins, 1970, p50).

The ACF of an MA(q) process is given by

$$\rho_k = \begin{cases} 1 & \text{if } k = 0 \\ \frac{\sum_{i=1}^{q-k}\beta_i\beta_{i+k}}{\sum_{i=0}^{q}\beta_i\beta_i} & \text{if } k \leq q \\ \rho_{-k} & \text{if } k < 0 \\ 0 & \text{if } k > q \end{cases} \tag{2.3.5}$$

Note that the ACF cuts off at lag $k = q$ which is a special feature of MA processes. Thus, if the series X_t is generated by an MA(q) process, then the sample ACF (see definition in Section 2.2, item 5) of the series should 'cut off' at lag k.

2.3.3 Autoregressive moving average models of order (p,q) (ARMA(p,q))

The model for the series X_t can be an AR(p) model or an MA(q) model or a combination of both the AR(p) and the MA(q) models. The latter model is called an autoregressive moving average of order (p,q), denoted by ARMA(p,q), and is given by

$$X_t = \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + \epsilon_t - \beta_1 \epsilon_{t-1} - ... - \beta_q \epsilon_{t-q} \tag{2.3.6}$$

where $\alpha_1, \alpha_2, ..., \alpha_p, \beta_1, \beta_2, ..., \beta_q$ are model parameters to be estimated, and ϵ_t is a series of random errors each with zero mean and constant variance σ^2 (Box, 1976, p74). Alternatively, the ARMA(p,q) model may be written as

$$\alpha(\mathbf{B}) X_t = \beta(\mathbf{B}) \epsilon_t \tag{2.3.7}$$

where $\alpha(\mathbf{B}) = 1 - \alpha_1 \mathbf{B} - \alpha_2 \mathbf{B}^2 - ... - \alpha_p \mathbf{B}^p$ is a polynomial in $\mathbf{B}$ of order p, and $\beta(\mathbf{B}) = 1 - \beta_1 \mathbf{B} - \beta_2 \mathbf{B}^2 - ... - \beta_q \mathbf{B}^q$ is a polynomial in $\mathbf{B}$ of order q. Note that the AR(p) and MA(q) models are special ARMA(p,q) models, For example, the AR(p) is the ARMA(p,0) model, and the MA(q) model is the ARMA(0,q) model.

The ACF of an ARMA(p,q) process decays exponentially after lag 1. Thus, if the series X_t is generated by an ARMA(p,q) process, then the sample ACF of the series X_t generally attenuates as the lag increases rather than 'cutting off' at some lag. For example, case ρ_1 is significantly different from zero, thus the subsequent ρ_k gets closer and closer to zero (Chatfield, 2004, p58).

2.3.4 Autoregressive integrated moving average models of order (p,d,q) (ARIMA(p,d,q))

In practice, most time series are nonstationary and, hence, ARMA(p,q) models cannot be fitted to the series despite their having nice properties. In order to exploit the nice properties of ARMA(p,q) models, it is necessary to first stationarise the nonstationary series, and then fit ARMA(p,q) models to the resultant stationary series. This is done as follows:

Suppose that the observed time series Y_t is nonstationary only in the mean, and the mean is a polynomial in t of degree d, then the time series

$$X_t = (1 - \mathbf{B})^d Y_t \tag{2.3.8}$$

is trendless and, hence, stationary. The optimal order of differencing is d, for which the standard deviation of $(1 - \mathbf{B})^d Y_t$ is the smallest. The general model for the differenced series (2.3.8) is ARMA(p,q)) since the series is stationary. Thus,

$$\alpha(\mathbf{B}) X_t = \alpha(\mathbf{B})(1 - \mathbf{B})^d Y_t = \beta(\mathbf{B}) \epsilon_t \tag{2.3.9}$$

where $\alpha(\mathbf{B}) = 1 - \alpha_1 \mathbf{B} - \alpha_2 \mathbf{B}^2 - ... - \alpha_p \mathbf{B}^p$, $\beta(\mathbf{B}) = 1 - \beta_1 \mathbf{B} - \beta_2 \mathbf{B}^2 - ... - \beta_q \mathbf{B}^q$ and ϵ_t is a series of random errors each with zero mean and constant variance σ^2. Model (2.3.9) for the series Y_t is referred to as the autoregressive integrated moving average model of order (p,d,q), and is denoted by ARIMA(p,d,q) (Diggle, 1990, p165). Note that the AR(p), MA(q) and ARMA(p,q) models are special ARIMA(p,d,q) models. For example, the ARMA(p,q) model is the ARIMA(p,0,q) model, the AR(p) model is the ARIMA(p,0,0) model and the MA(q) model is the ARIMA(0,0,q) model.

2.4 Model identification, estimation and diagnostics

The steps followed in the process of fitting ARIMA models to a given time series are

- model identification
- model estimation
- model diagnostics

These steps are discussed in this section.

2.4.1 Model identification

A tentative ARIMA model for a given time series is identified by various sample correlation functions, such as the sample ACF and the sample partial autocorrelation functions (Gooijer *et al.*, 1985). For example, the AR(p), MA(q) and ARMA(p,q) models for a given time series are identified using the sample ACFs as explained at the ends of Sections 2.3.1, 2.3.2 and 2.3.3.

2.4.2 Model estimation methods

Once a tentative ARIMA model for a given series has been identified, the next step is to estimate the model parameters. Typical methods for estimating the model parameters are either the least squares method or the maximum likelihood methods (Box, 1976, p208). These methods are briefly reviewed under the assumption that the identified tentative ARIMA model for a given time series X_t is (2.3.7), which can be rewritten as

$$\epsilon_t = \frac{\alpha(\mathbf{B})}{\beta(\mathbf{B})} X_t \qquad (2.4.1)$$

where $\alpha(\mathbf{B}) = 1 - \alpha_1\mathbf{B} - \alpha_2\mathbf{B}^2 - ... - \alpha_p\mathbf{B}^p$, $\beta(\mathbf{B}) = 1 - \beta_1\mathbf{B} - \beta_2\mathbf{B}^2 - ... - \beta_q\mathbf{B}^q$ and ϵ_t is a series of random errors each with zero mean and constant variance σ^2. Let

$$SSE(\underline{\alpha}, \underline{\beta}) = \sum_{t=1}^{n} \epsilon_t^2 \qquad (2.4.2)$$

and

$$L(\underline{\alpha}, \underline{\beta}, \sigma^2) = (2\pi\sigma^2)^{\frac{-n}{2}} exp(-\frac{1}{2\sigma^2}\sum_{t=1}^{n} \epsilon_t^2) \qquad (2.4.3)$$

where $\underline{\alpha} = (\alpha_1, \alpha_2, \ . \ . \ \alpha_p)^T$ and $\underline{\beta} = (\beta_1, \beta_2, \ . \ . \ . \ \beta_q)^T$

2.4.2.1 Least squares method

The least squares estimates of the parameters of the ARIMA model (2.3.7) are values of $(\underline{\alpha}^T, \underline{\beta}^T)^T$ which minimise the error sum of squares (2.4.2). The estimates are found using numerical methods as there are no closed formulae for the estimators of the parameters from minimising $SSE(\underline{\alpha}, \underline{\beta})$ given by (2.4.2).

2.4.2.2 Maximum likelihood method

Under the assumption that the ϵ_t in (2.4.1) are iid $N(0, \sigma^2)$, the maximum likelihood estimates of the parameters of the ARIMA model (2.3.7) are values of $(\sigma^2, \underline{\alpha}^T, \underline{\beta}^T)^T$, which maximise the likelihood function (2.4.3). As with the least squares method, the maximum likelihood estimates of the model parameters are found using numerical methods as there are no closed formulae for the estimators of the parameters from maximising $L(\underline{\alpha}, \underline{\beta}, \sigma^2)$ given by (2.4.3).

2.4.3 Model diagnostics

Although the identified tentative ARIMA model may appear to be the best among those models considered, it may be inadequate for a number of reasons which include:

- violation of the assumption of stationarity of X_t
- presence of outliers in X_t
- over- or under-parametrisation of the model

Descriptive and formal testing methods are used to check the adequacy of the fitted model. Both methods involve checking/testing

- whether or not the residuals from fitting the model are white noise,
- whether or not the model is the simplest best model

2.4.3.1 Checking/testing the white noisiness of the residuals

If the ARIMA model adequately fits the given time series, then the residuals from fitting the model should be white noise. Hence, a plot of the ACF of the residuals should show no significant spikes (correlations) at all lags k.

One formal test of the null hypothesis that the residuals from fitting the ARIMA model are white noise is the Portmanteau test (Chatfield, 2004, p68). The test for white noisiness of the residuals uses the sample ACF of the residuals to test the null hypothesis:

$$H_0 : \rho_1 = \rho_2 = ... = \rho_k = 0, \text{ for some } k > 1 \tag{2.4.4}$$

The test statistic for the hypothesis is:

$$Q = N(N-2) \sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{(N-k)} \sim \chi^2_{k-p-q} \tag{2.4.5}$$

where N is the number of observations in the differenced series (to stationarise the original series). The hypothesis that the residuals series is white noise is rejected at the $\alpha \in (0,1)$ level of significance if $Q > \chi^{2\alpha}_{k-p-q}$ - is the $(1-\alpha)100$ percentile of the χ^2_{k-p-q} distribution.

2.4.3.2 Testing the significance of the parameters in the ARIMA model

Provided that the series is long enough and that the residuals are white noise, the significance or insignificance of the parameters in the model are tested using the statistic

$$t = \frac{\text{estimate}}{se(\text{estimate})} \tag{2.4.6}$$

which has an asymptotic standard normal distribution (Chatfield, 2004, p65). The decision rule of the test is that a parameter is insignificantly different from zero at the $\alpha \in (0,1)$ level of significance if $|t| > t_{\alpha/2}$ - is the $(1-\alpha)100$ percentile of the $N(0,1)$ distribution. Parameters which are insignificant in the model are then removed and the reduced model is refitted to the series.

2.4.3.3 Choosing the best model among competing ARIMA models

There may be several competing ARIMA models that adequately describe the given time series. However, the problem that arises is choosing the best among the competing models. Two criteria for choosing the best model that are commonly used are the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) or Schwartz Bayesian criterion (SBC).

$$AIC = -2\ln(\text{Likelihood}) + 2r \tag{2.4.7}$$

where likelihood is the likelihood given by (2.4.3) evaluated at the maximum likelihood estimates of the model parameters, and r denotes the number of model parameters. The AIC increases with the number of model parameters (r), and the best model is the one with the smallest AIC.

The BIC is an extension of the AIC, and is given by

$$BIC = -2\ln(\text{ Likelihood}) + r\ln(N) \tag{2.4.8}$$

where N is the number of observations in the differenced series (to stationarise the original series). As with the AIC, the best model among competing ARIMA(p,d,q) models is one with the smallest BIC.

2.5 Forecasting

One of the primary objectives of building a model for a time series analysis is forecasting future values of the series (Cryer and Chan, 2008, p191). Furthermore, the major objective of this thesis (Section 1.2) is to investigate the feasibility of accurately and precisely forecasting the number of national and provincial beneficiaries of social security grants in South Africa. In this section, we discuss the minimum mean square error forecasting method using ARIMA models (Wei, 2006, p181).

In general, given the series $X_1, X_2, ..., X_t$, the forecasting problem is to predict the value of X_{t+l} $(l = 1, 2, ...)$ of the series that will be observed at time $t + l$ in the future. According to the minimum mean square error criterion of forecasting, this value is $\hat{X}_t(l)$, which minimises the conditional mean square error

$$E[(X_{t+l} - \hat{X}_t(l))^2 | X_1, X_2, ..., X_t] \tag{2.5.1}$$

Differentiating (2.5.1) with respect to $\hat{X}_t(l)$, equating the derivative to zero, and solving the resultant equation for $\hat{X}_t(l)$ gives

$$\hat{X}_t(l) = E[X_{t+l} | X_1, X_2, ..., X_t] \tag{2.5.2}$$

as the minimiser of (2.5.1), and hence, as the minimum mean square error forecast of the value of X_{t+l}. This section is focused on the calculation of $\hat{X}_t(l)$ when the time series is generated by an ARIMA process.

Recall (Section 2.3.3) that the ARMA(p,q) model for a times series is given by

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + \epsilon_t - \beta_1 \epsilon_{t-1} - \beta_2 \epsilon_{t-2} - ... - \beta_q \epsilon_{t-q} \tag{2.5.3}$$

This means the unknown X_{t+l} is given by

$$X_{t+l} = \alpha_1 X_{t+l-1} + \alpha_2 X_{t+l-2} + ... + \alpha_p X_{t+l-p} + \epsilon_{t+l} - \beta_1 \epsilon_{t+l-1} - \beta_2 \epsilon_{t+l-2} - ... - \beta_q \epsilon_{t+l-q} \tag{2.5.4}$$

Using (2.5.2) and (2.5.4), the minimum mean square error forecast of the value of X_{t+l} is given by

$$\begin{aligned}
\hat{X}_t(l) &= E[X_{t+l} | X_1, ..., X_t] \\
&= \alpha_1 E[X_{t+l-1} | X_1, ..., X_t] + \alpha_2 E[X_{t+l-2} | X_1, ..., X_t] + ... + \alpha_p E[X_{t+l-p} | X_t, ..., X_t] \\
&- \beta_1 E[\epsilon_{t+l-1} | X_1, ..., X_t] - \beta_2 E[\epsilon_{t+l-2} | X_1, ..., X_t] - ... - \beta_q E[\epsilon_{t+l-q} | X_1, ..., X_t] \\
&= \alpha_1 \hat{X}_t(l-1) + \alpha_2 \hat{X}_t(l-2) + ... + \alpha_p \hat{X}_t(l-p) \\
&- \beta_1 E[\epsilon_{t+l-1} | X_1, ..., X_t] - \beta_2 E[\epsilon_{t+l-2} | X_1, ..., X_t] - ... - \beta_q E[\epsilon_{t+l-q} | X_1, ..., X_t]
\end{aligned} \tag{2.5.5}$$

where

$$E[\epsilon_{t+l-j} | X_1, ..., X_t] = \begin{cases} 0 & \text{if } l > j \\ \epsilon_{t+l-j} & \text{if } l \leq j \end{cases} \tag{2.5.6}$$

and

$$\hat{X}_t(l-j) = \begin{cases} \hat{X}_t(l-j) & \text{if } l > j \\ X_{t+l-j} & \text{if } l \leq j \end{cases} \tag{2.5.7}$$

For example, if $l = 2 < p = q$ then

$$\hat{X}_t(2) = \alpha_1 \hat{X}_t(1) + \alpha_2 X_t + ... + \alpha_p X_{t+2-p} - \beta_2 \epsilon_t - ... - \beta_q \epsilon_{t+l-q} \tag{2.5.8}$$

The forecast error in $\hat{X}_t(l)$ given by (2.5.5) is given by

$$\begin{aligned} e_t(l) &= X_{t+l} - \hat{X}_t(l) \\ &= \sum_{j=0}^{l-1} \psi_j \epsilon_{t+l-j} \end{aligned} \tag{2.5.9}$$

in random shock model form, where the ψ_j $(j = 1, 2, ...)$ weights are functions of the ARMA(p,q) model parameters. From (2.5.9),

$$Var(e_t(l)) = \sigma^2 \sum_{j=0}^{l-1} \psi_j^2 \tag{2.5.10}$$

which is used for constructing the confidence interval for X_{t+l} under the assumption that the white noise errors ϵ_t are independent and normally distributed with mean 0 and variance σ^2. If the assumption holds, then for $\alpha \in (0, 1)$, the $(1 - \alpha/2) \times 100\%$ confidence interval for X_{t+1} is given by

$$\hat{X}_t(l) \pm z_{\alpha/2} \sqrt{Var(\epsilon_t(l))} \tag{2.5.11}$$

where $z_{z/2}$ is the $(1 - \alpha/2) \times 100^{th}$ percentile of the standard normal distribution. In practice, the unknown parameters in the formulae for $\hat{X}_t(l)$ and $Var(\epsilon_t(l))$ are replaced with estimates.

For a given time series there maybe several competing ARIMA models for forecasting. According to (Wei, 2006, p181), the models can be compared for goodness-of-forecasting using four criteria described below. Fit the ARIMA models to the $t - l$ $(0 < l \leq t)$ observations of the time series and use the fitted models to forecast the last l observed values of the series. Calculate:

$$\epsilon_t(t - l + j) = X_{t-l+j} - \hat{X}_{t-l+j}, j = 1, 2, ..., l \tag{2.5.12}$$

where $j = 1, 2, ..., l$ $\hat{X}_{t-l+j}$ is the forecast of X_{t-l+j} using any of the competing models; and

mean percentage error/bias given by

$$MPE = \frac{1}{l} \sum_{j=1}^{l} \frac{\epsilon_t(t - l + j)}{X_{t-l+j}} \times 100\% \tag{2.5.13}$$

mean square error given by

$$MSE = \frac{1}{l} \sum_{j=1}^{l} \epsilon_t^2(t - l + j) \tag{2.5.14}$$

mean absolute error given by

$$MAE = \frac{1}{l} \sum_{j=1}^{l} |\epsilon_t(t - l + j)| \tag{2.5.15}$$

and

mean absolute percentage given by

$$MAPE = \frac{1}{l}\sum_{j=1}^{l}\left|\frac{\epsilon_t(t-l+j)}{X_{t-l+j}}\right| \times 100\% \tag{2.5.16}$$

for each of the competing models. Then the best model for forecasting is the one with the smallest MPE, MSE, MAE or MAPE depending on the criterion/criteria one chooses to use.

Chapter 3

ARIMA Forecasts of the Number of Beneficiaries of Social Security Grants in South Africa

3.1 Introduction

This chapter deals with the analysis of the data described in Section 1.3 in order to answer the questions implied by the objectives of the thesis presented in Section 1.2. The theories of the methods of analysis that are used in this chapter were briefly reviewed in Chapter 2. With the help of the Proc ARIMA procedure in SAS Version 9.2, the methods are used to fit ARIMA models to the number of national and provincial beneficiaries of selected social security grants. The fitted models are then used to forecast the number of national and provincial beneficiaries of these grants. Graphical methods and the descriptive statistics presented at the end of Section 2.5 are used to cross-validate the accuracy and precision of the forecasts, and to compare the accuracy and precision of the national forecasts obtained using the two methods. Details on how this was actually done are contained in Section 3.3.

3.2 Data selection

Recall (Section 1.3) that data consist of the monthly records of the number of national and provincial beneficiaries of social security grants, namely, old age, child support, disability, foster care, care dependency, war veteran and grant-in-aid from April 2004 to March 2010. Thus, the length of each series is 72. It should be kept in mind that for each type of grant, the national series was an aggregate of the provincial series.

For the purpose of meeting the objectives of this thesis, the old age, the child support, the foster care and the disability grants were considered. The assumption was made that any inferences made in terms of these grants would extend to the remaining grants. Besides, there are claims that there is some dependence of the grant-in-aid series on the old age, the disability, and the war veteran grant series, and some dependence of the care dependency series on the child support and the foster care grant series, which are still to be investigated.

The series are short (72). It is for this reason that ARIMA models were fitted to the first 59 observations of each series, and the remaining 13 observations were used to cross-validate the accuracy and precision of the 13 months ahead forecasts, and to compare the accuracy and precision of the national forecasts obtained using two methods. Subsequently, in Section 3.3.5 the effect of the length of the series on the accuracy and precision of the forecasts is investigated.

3.3 Methods, results and discussion

3.3.1 The monthly number of old age grant beneficiaries in South Africa

Figure 3.1 displays the time series plot of the monthly number of beneficiaries of the old age grant in South Africa from April 2004 to March 2010. Clearly, the series is not stationary as it has a nonlinear trend. This suggests differencing in order to remove the trend and hence stationarise the series.

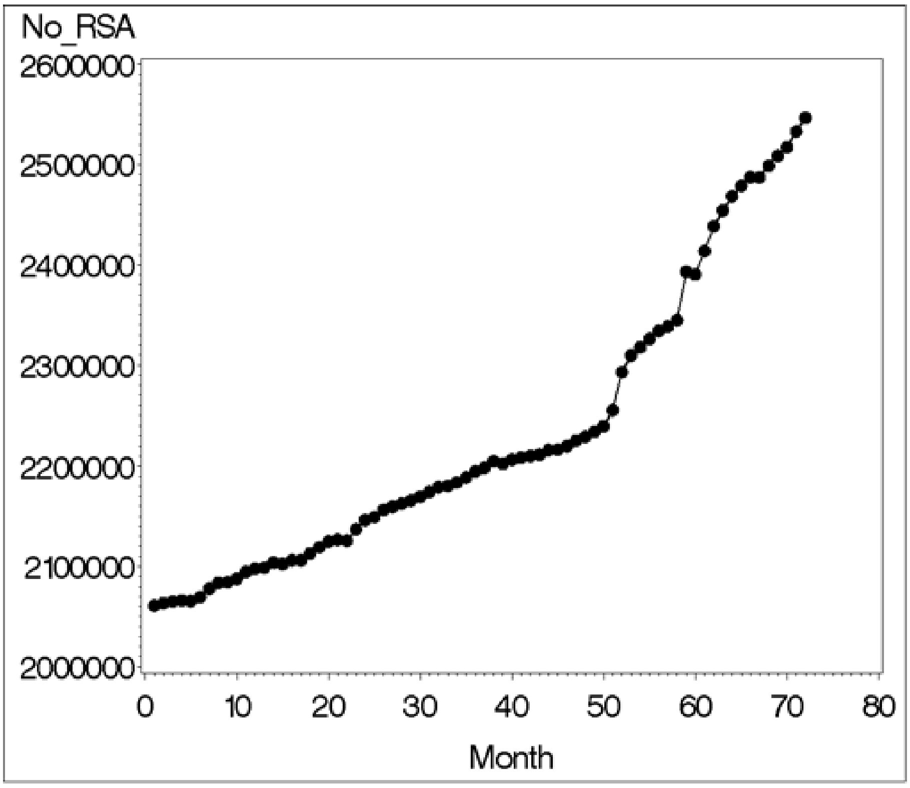

Figure 3.1: The monthly number of beneficiaries of the old age grant in South Africa versus time (Month = 1 is April 2004 and Month = 72 is March 2010). No-RSA = number of beneficiaries of the old age grant in South Africa

The last thirteen observations of the series (March 2009 to March 2010) were deleted and the remaining series (April 2004 to February 2009) ARIMA modelled. The fitted ARIMA models were used to forecast the deleted last thirteen observations of the original series, and these were used to cross-check the accuracy and precision of the forecasts.

Figure 3.1 also shows that the trend in the remaining series (Month = 1 to Month = 59) is almost linear, suggesting one as the optimal order of differencing to remove the trend in this remaining series. This order of differencing was confirmed using the method described in Section 2.3.4 (see Table 3.4).

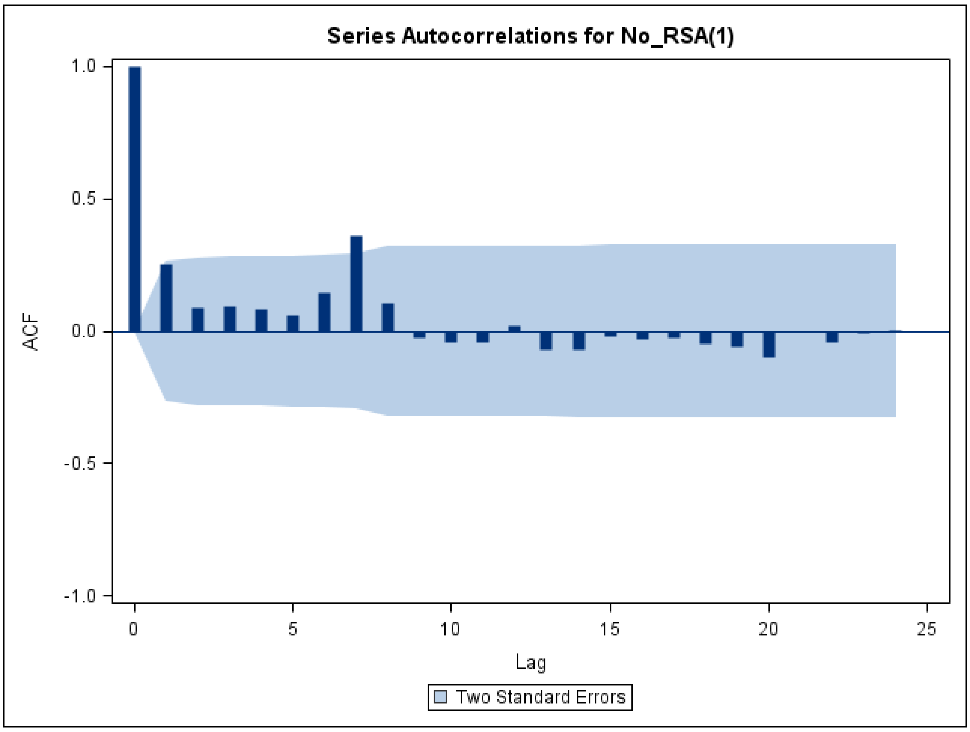

Figure 3.2: The autocorrelation function of the once differenced series of the monthly number of beneficiaries of the old age grant in South Africa (April 2004 to February 2009)

Figure 3.2 displays the autocorrelation function (ACF) of the once differenced series, and Table 3.1 displays the Portmanteau test statistics for checking/testing whether or not the differenced series is white noise. In Table 3.1, the chi-square statistics are approximately equal to the corresponding degrees of freedom, and the p-values for testing the significance of the correlations are greater than the 0.05 level of significance. This led to the conclusion that the differenced series was white noise. However, the two significant autocorrelations in the ACF displayed in Figure 3.2 suggested the moving average models (MA(q)) for the differenced series. Thus, the competing models for the differenced series $(1 - \mathbf{B})X_t$ were:

$$(1 - \mathbf{B})X_t = \mu + \epsilon_t - \beta_1\epsilon_{t-1} - \beta_7\epsilon_{t-7} \quad (3.3.1)$$

$$(1 - \mathbf{B})X_t = \mu + \epsilon_t - \beta_7\epsilon_{t-7} \quad (3.3.2)$$

$$(1 - \mathbf{B})X_t = \mu + \epsilon_t \quad (3.3.3)$$

Table 3.1: The Portmanteau test statistics for checking the white noisiness of the differenced series of the monthly number of beneficiaries of the old age grant in South Africa (April 2004 to February 2009)

Autocorrelation check for white noise

To lag	Chi-square	DF	p-value	Autocorrelations					
6	7.10	6	0.3115	0.254	0.087	0.097	0.085	0.061	0.144
12	17.09	12	0.1464	0.360	0.108	-0.026	-0.043	-0.042	0.022
18	18.20	18	0.4428	-0.069	-0.071	-0.020	-0.032	-0.023	0.046
24	19.61	24	0.7186	-0.059	-0.101	-0.005	-0.041	-0.011	0.001

Models (3.3.1), (3.3.2) and (3.3.3) were fitted to the differenced series and the model selection criteria described in Section 2.4.3.3 were used to choose the best model. Table 3.2 contains the statistics for choosing the best model, the parameter estimates, their standard errors ( ), as well as the p-values [ ] for testing the significance of the model parameters. Model (3.3.2) was the best with respect to both the AIC and the SBC criteria and was closely followed by model (3.3.1). However, (3.3.1) was unstable and its β_1 parameter was insignificantly different from zero at the 0.05 level of significance (p-value = 0.0919).

Table 3.2: The goodness-of-fit statistics from fitting models (3.3.1), (3.3.2) and (3.3.3) to the once differenced series of the monthly number of beneficiaries of the old age grant in South Africa (April 2004 to February 2009)

	Criterion		Parameter estimate $(Se)[p - value]$		
Model	AIC	SBC	$\hat{\mu}()[]$	$\hat{\beta}_1()[]$	$\hat{\beta}_7()[]$
(3.3.1)	1196	1202	5854(1674)[0.0009]	$-0.2380(0.1387)[0.0919]$	$-0.7620(0.1772)[< .0001]$
(3.3.2)	1194	1199	5864(1401)[0.0001]		$-0.6283(0.1166)[< .0001]$
(3.3.3)	1208	1210	5727(1050)[< .0001]		

The residuals from fitting model (3.3.2) to the once differenced series were checked for white noise. The ACF of the residuals had no significant spikes. Furthermore, in Table 3.3, the chi-square statistics are approximately equal to the corresponding degrees of freedom, and the associated p-values for testing the significance of the correlations are all greater than 0.1. This means that the residuals from fitting model (3.3.2) to the once differenced series were white noise, hence, the model adequately described the series. The conclusion was that the fitted model (3.3.2) given by:

$$(1-\mathbf{B})X_t = 5864 + (1+0.6283\mathbf{B}^7)\hat{\epsilon}_t \tag{3.3.4}$$

was a good candidate for use to forecast the monthly number of beneficiaries of the old age grant in South Africa.

Table 3.3: The Portmanteau test statistics for checking the white noisiness of residuals from fitting model (3.3.2) to the once differenced series of the monthly number beneficiaries of the old age grant in South Africa (April 2004 to February 2009)

Autocorrelation check for white noise

To lag	Chi-square	DF	p-value	Autocorrelations					
6	8.70	5	0.1218	0.301	0.171	0.108	0.073	-0.065	-0.003
12	10.27	11	0.5064	0.048	0.016	-0.080	-0.065	-0.058	0.071
18	10.66	17	0.8738	-0.018	0.048	0.036	0.019	-0.004	-0.023
24	15.06	23	0.8924	-0.102	-0.124	-0.110	-0.088	-0.043	0.003

The performance of model (3.3.4) in forecasting the monthly number beneficiaries of the old age grant in South Africa was checked by comparing the 13 forecasts from March 2009 to March 2010 using the model with the observed series over the same period. Furthermore, the forecasts using the model were compared with the forecasts obtained using another approach described below.

3.3.1.1 The national forecasts from aggregated provincial forecasts

The modelling procedure described above was adopted for modelling the monthly number of beneficiaries of the old age grant in each of the nine provinces of South Africa. For each province, the best ARIMA model obtained was used to forecast the monthly number of beneficiaries of the old age grant from March 2009 to March 2010. The provincial forecasts from March 2009 to March 2010 were aggregated to obtain the forecasts for South Africa as a whole. These forecasts were compared with the ARIMA forecasts for South Africa obtained using model (3.3.4).

Table 3.4 displays the optimal orders of differencing for the monthly number of beneficiaries of the old age grant in South Africa and in South Africa's provinces. The optimal orders of differencing were determined using the method described in Section 2.3.4. Table 3.4 shows that the trends in the series were all linear except for the trend in the series for Gauteng and Mpumalanga, which were quadratic.

Table 3.4: The standard deviations of the differenced monthly number beneficiaries of the old age grant in South Africa and in South Africa's provinces (April 2004 to February 2009) versus the order of differencing

Order (d)	RSA	EC	FS	GP	KZN	LP	MP	NW	NC	WC
0	82214	7789	4433	11760	19724	15628	5381	6264	6371	7355
1	**7925**	**2077**	**620**	931	**2385**	**743**	411	**1625**	**1369**	**975**
2	7925	2523	683	**883**	2785	795	**408**	2174	1974	1036

Boldfaced numbers are the smallest standard deviations associated with the optimal order of differencing (d) to remove trend, RSA = South Africa, EC = Eastern Cape, FS = Free State, GP = Gauteng, KZN = KwaZulu-Natal, LP = Limpopo, MP = Mpumalanga, NW = North West, NC = Northern Cape, WC = Western Cape

The following Table 3.5 displays the models which were found to adequately describe the monthly number of beneficiaries of the old age grant in the South Africa's provinces (April 2004 to February 2009). Clearly, the models are province specific. The models in the table were used to forecast the provincial monthly number of beneficiaries of the old age grant from March 2009 to March 2010. The provincial forecasts from March 2009 to March 2010 were aggregated to get the forecasts for South Africa as a whole. These forecasts were compared with ARIMA forecasts for South Africa which were obtained using model (3.3.4), as was mentioned above.

Table 3.5: The fitted models to the monthly number of beneficiaries of the old age grant in South Africa's provinces (April 2004 to February 2009)

Province	Fitted model
EC	$(1-\mathbf{B})X_t = 546(182)[0.0041] + (1 - 0.3501(0.1577)[0.0305]\mathbf{B}^4)\hat{\epsilon}_t$
FS	$(1-\mathbf{B})X_t = 332(114)[0.0051] + (1 + 0.7439(0.1273)[0.0001]\mathbf{B}^7)\hat{\epsilon}_t$
GP	$(1-\mathbf{B})^2X_t = 98(118)[0.4111] + \hat{\epsilon}_t$
KZN	$(1-\mathbf{B})X_t = 1366(437)[0.0028] + (1 + 0.5006(0.1775)[0.0066]\mathbf{B}^4)\hat{\epsilon}_t$
LP	$(1-\mathbf{B})X_t = 963(128)[0.0001] + (1 + 0.3833(0.1411)[0.0088]\mathbf{B})\hat{\epsilon}_t$
MP	$(1-\mathbf{B})^2X_t = 42(54)[0.4388] + \hat{\epsilon}_t$
NW	$(1-\mathbf{B})X_t = 450(215)[0.0411] + \hat{\epsilon}_t$
NC	$(1-\mathbf{B})X_t = 322(181)[0.0812] + \hat{\epsilon}_t$
WC	$(1-\mathbf{B})X_t = 604(177)[0.0012] + (1 + 0.7012(0.1194)[0.0001]\mathbf{B}^7)\hat{\epsilon}_t$

( ) = standard error of the parameter estimate, [ ] = p-value for testing the significance of the parameter, EC = Eastern Cape, FS = Free State, GP = Gauteng, KZN = KwaZulu-Natal, LP = Limpopo, MP = Mpumalanga, NW = North West, NC = Northern Cape, WC = Western Cape

Figure 3.3 displays the graph of the actual monthly number of beneficiaries of the old age grant in South Africa (March 2009 to March 2010), and the corresponding forecasts obtained using model (3.3.4) and the aggregated provincial forecasts obtained using the models in Table 3.5. It is clear that the forecasts obtained by aggregating the provincial forecasts are better since they are closer to the actual values than those obtained using model (3.3.4). This finding is confirmed by all criteria (MSE, MAE, MPE and MAPE) for choosing good forecasting methods which were presented in Section 2.5 (see Table 3.6). Therefore, it can be concluded that forecasting the national number of beneficiaries of the old age grant by aggregating the provincial ARIMA forecasts is better than forecasting via ARIMA modelling of the national number of beneficiaries of the old age grant.

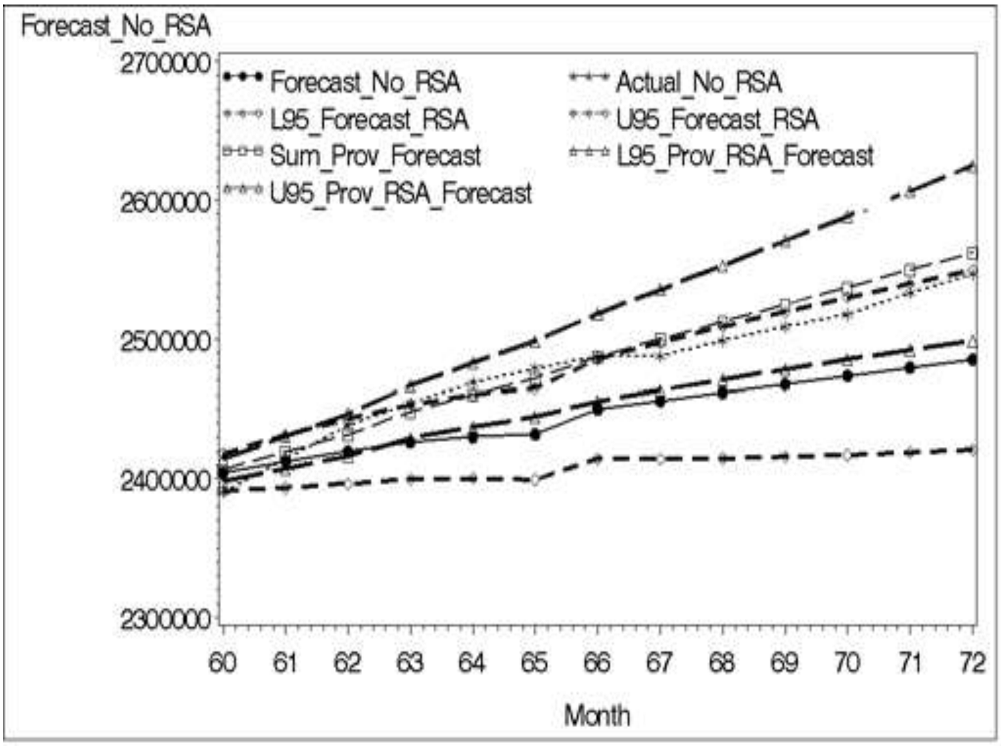

Figure 3.3: The actual monthly number of beneficiaries of the old age grant in South Africa (March 2009 to March 2010) and the corresponding forecasts obtained using model (3.3.4) and the provincial models in Table 3.5

Table 3.6: A comparison of the forecasts of the monthly number of beneficiaries of the old age grant in South Africa (March 2009 to March 2010) obtained using model (3.3.4) and those obtained by aggregating the provincial forecasts

Forecasting method	MSE	MAE	MPE	MAPE
RSA forecasts using model (3.3.4)	1465840255	34907	1.3076	1.3983
Provincial forecasts using models in Table 3.5	147236556	10914	-0.2521	0.4389

MSE = mean square error, MAE = mean absolute error, MPE = mean percentage error, MAPE = mean absolute percentage error

3.3.2 The monthly number of beneficiaries of the child support grant in South Africa

Exactly the same procedure for obtaining forecasts of the monthly number of beneficiaries of the old age grant in South Africa using model (3.3.4), and by aggregating the provincial forecasts(described in Section 3.3.1), was followed to obtain the forecasts of the number beneficiaries of the child support grant in South Africa through ARIMA modelling of the national series and the provincial series.

Figure 3.4 displays the time series plot of the monthly number of beneficiaries of the child support grant in South Africa from April 2004 to March 2010. The series is not stationary as it has a non-linear trend. A comparison of this plot with that displayed in Figure 3.1 shows that the ARIMA models for the national series of the monthly number of beneficiaries of the various types of grant beneficiary are specific to the type of grant.

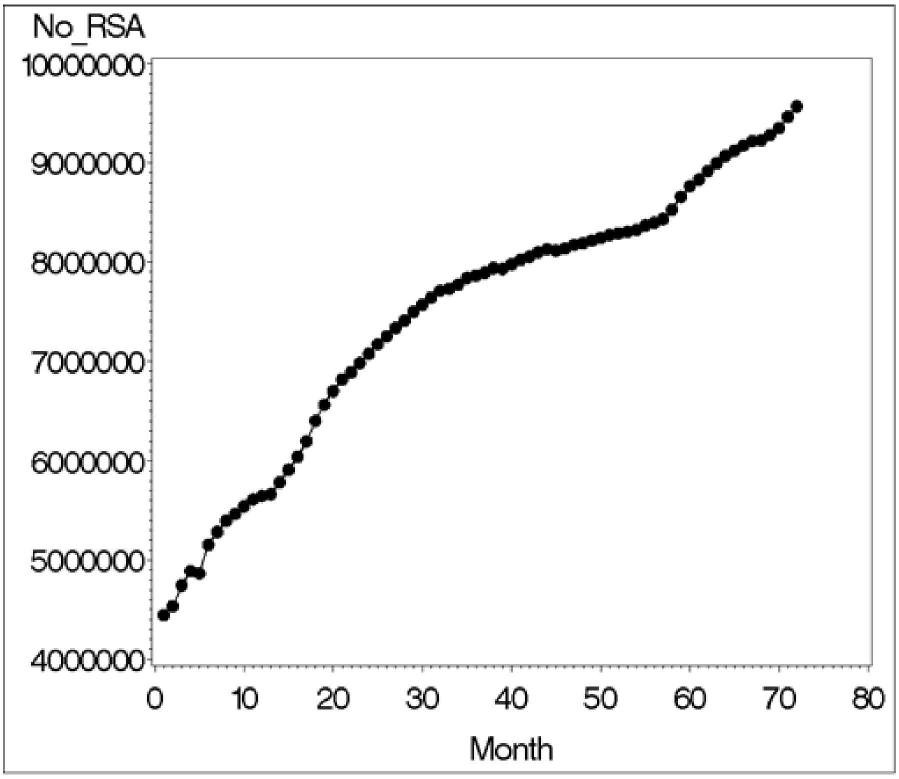

Figure 3.4: The monthly number of beneficiaries of the child support grant in South Africa over time (Month = 1 is April 2004 and Month = 72 is March 2010). No-RSA = number of beneficiaries of the child support grant in South Africa

Table 3.7 displays the optimal orders of differencing of the monthly number of beneficiaries of the child support grant in South Africa and South Africa's provinces. The removed trends were all linear except for the Eastern Cape series which was quadratic.

Table 3.7: The standard deviations of the differenced number of beneficiaries of the child support grant in South Africa and in South Africa's provinces (April 2004 to February 2009) versus the order of differencing

Order (d)	RSA	EC	FS	GP	KZN	LP	MP	NW	NC	WC
0	1225873	2489890	60585	126658	355779	159624	92053	91877	39875	65859
1	**58107**	15647	**4294**	**6122**	**14978**	**13277**	**5377**	**9639**	**5727**	**7387**
2	61285	**14968**	4397	6575	12661	15012	5489	12661	8262	12172

Boldfaced numbers are the smallest standard deviations associated with the optimal order of differencing (d) to remove trend; RSA = South Africa, EC = Eastern Cape, FS = Free State, GP = Gauteng, KZN = KwaZulu-Natal, LP = Limpopo, MP = Mpumalanga, NW = North West, NC = Northern Cape, WC = Western Cape

Table 3.8 contains the models which were found to adequately describe the monthly number of beneficiaries of the child support grant in South Africa as a whole and in South Africa's provinces (April 2004 to February 2009). Clearly, the models are province specific. Furthermore, a comparison of these models with those in Table 3.5 shows that the models are also specific to the type of grant.

Table 3.8: The fitted models to the monthly number of beneficiaries of the child support grant in South Africa's provinces (April 2004 to February 2009)

P/N.	Fitted model
RSA	$(1-\mathbf{B})X_t = \dfrac{77515(14765)[0.0001] + (1+0.5187(0.1244)[0.0001]\mathbf{B}^3)\hat{\epsilon}_t}{(1-0.3746(0.1290)[0.0053]\mathbf{B})}$
EC	$(1-\mathbf{B})^2X_t = (1-0.7143(0.0989)[0.0001]\mathbf{B})\hat{\epsilon}_t$
FS	$(1-\mathbf{B})X_t = \dfrac{3962(2685)[0.1458] + \hat{\epsilon}_t}{(1-0.2739(0.1033)[0.0104]\mathbf{B} - 0.6190(0.1090)[0.0001]\mathbf{B}^{11})}$
GP	$(1-\mathbf{B})X_t = \dfrac{8309(1861)[0.0.0001] + \hat{\epsilon}_t}{(1-0.2945(0.1280)[0.0252]\mathbf{B} - 0.3428(0.1273)[0.0094]\mathbf{B}^2)}$
KZN	$(1-\mathbf{B})X_t = \dfrac{20347(3048)[0.0001] + \hat{\epsilon}_t}{(1-0.4035(0.1248)[0.0021]\mathbf{B})}$
LP	$(1-\mathbf{B})X_t = \dfrac{13352(5063)[0.0108] + (1+0.3271(0.1422)[0.0252]\mathbf{B})\hat{\epsilon}_t}{(1-0.6488(0.1422)[0.0252]\mathbf{B}^3)}$
MP	$(1-\mathbf{B})X_t = \dfrac{5575(1305)[0.0001] + (1+0.2620(0.1227)[0.0373]\mathbf{B}^2 + 0.4214(0.1205)[0.0009]\mathbf{B}^3)\hat{\epsilon}_t}{(1-0.2808(0.1362)[0.0441]\mathbf{B})}$
NW	$(1-\mathbf{B})X_t = 5187(1277)[0.0001] + \hat{\epsilon}_t$
NC	$(1-\mathbf{B})X_t = 2145(759)[0.0064] + \hat{\epsilon}_t$
WC	$(1-\mathbf{B})X_t = 4087(664)[0.0001] + (1-0.2960(0.1278)[0.0243]\mathbf{B})\hat{\epsilon}_t$

( ) = standard error of the parameter estimate, [ ] = p-value for testing the significance of the parameter; P/N = Province/National, RSA = South Africa, EC = Eastern Cape, FS = Free State, GP = Gauteng, KZN = KwaZulu-Natal, LP = Limpopo, MP = Mpumalanga, NW = North West, NC = Northern Cape, WC = Western Cape

The fitted provincial models in Table 3.8 were used to forecast the provincial monthly number of beneficiaries of the child support grant from March 2009 to March 2010. Subsequently, these forecasts were aggregated to obtain the forecasts for South Africa as a whole to be compared with the ARIMA forecasts obtained from the model for South Africa (also in Table 3.8), as was done in the previous section.

Figure 3.5 displays the graph of the actual monthly number of beneficiaries of the child support grant in South Africa (March 2009 to March 2010) and the corresponding forecasts obtained using the national ARIMA model in Table 3.8 as well as those obtained by aggregating the provincial forecasts also obtained using the models in Table 3.8. Again, the forecasts obtained by aggregating the provincial forecasts are better since they are closer to the actual series than those obtained using the national ARIMA model.

Furthermore, the 95% confidence band for the forecasts obtained by aggregating the provincial forecasts is within that of the forecasts obtained using the national ARIMA model. This means that the forecasts obtained by aggregating the provincial forecasts are more precise than those obtained using the national ARIMA model. This finding is confirmed by all criteria (MSE, MAE, MPE and MAPE) for choosing good forecasting methods which were presented in Section 2.5 (see Table 3.9).

Therefore, it can also be concluded in this section that forecasting the national number of beneficiaries of the child support grant by aggregating the provincial ARIMA forecasts is better than forecasting by means of ARIMA modelling of the national number of beneficiaries of the child support grant.

Table 3.9: A comparison of the forecasts of the number beneficiaries of the child support grant in South Africa (March 2009 to March 2010) obtained using the national ARIMA model in Table 3.8 and those obtained by aggregating the provincial forecasts obtained using the provincial ARIMA models in Table 3.8

Forecasting method	MSE	MAE	MPE	MAPE
RSA forecasts using model in Table 3.8	28017847944	140461	-1.4527	1.5137
Provincial forecasts using models in Table 3.8	4366327091	51341	-0.3408	0.5545

MSE = mean square error, MAE = mean absolute error, MPE = mean percentage error, MAPE = mean absolute percentage error

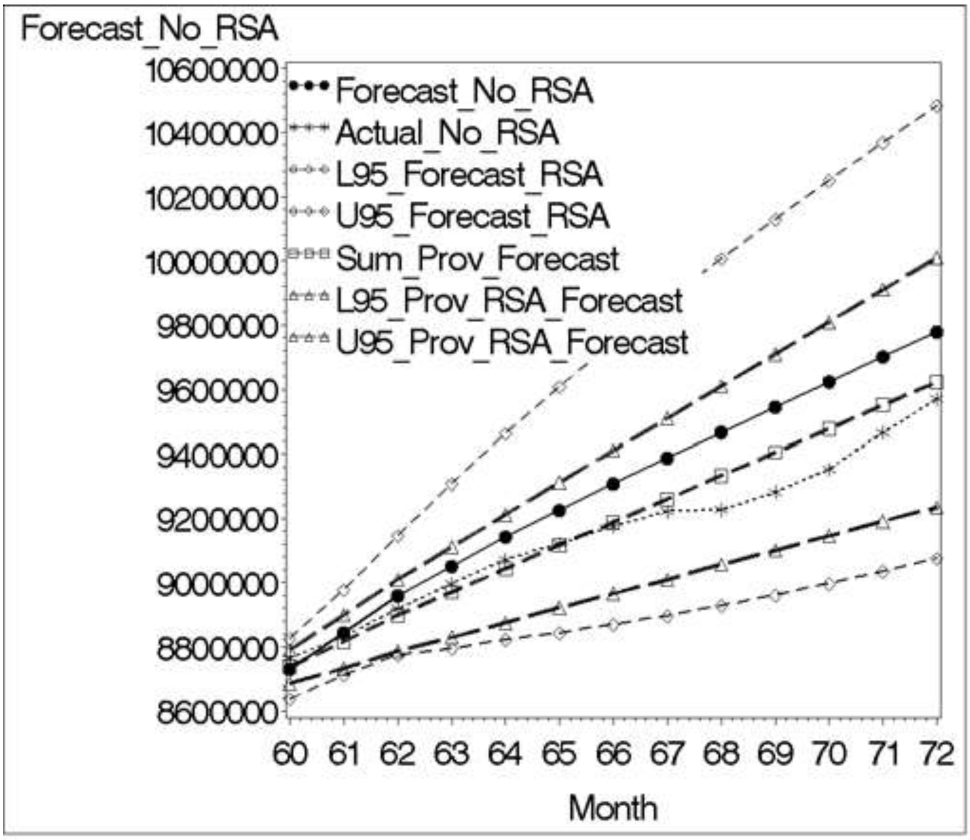

Figure 3.5: The actual monthly number of beneficiaries of the child support grant in South Africa (March 2009 to March 2010) and the corresponding forecasts obtained using the national ARIMA model and the provincial ARIMA models in Table 3.8

3.3.3 The monthly number of beneficiaries of the foster care grant in South Africa

Figure 3.6 displays the time series plot of the monthly number of beneficiaries of the foster care grant in South Africa from April 2004 to March 2010. The series is not stationary because of the presence of a linear trend and seasonality. Furthermore, this plot is very different from the plots displayed in Figures 3.1 and 3.4, which again means that the ARIMA models for the monthly number of beneficiaries of the various types of grant are specific to the type of grant.

Differencing was used to remove the trend and seasonality from the monthly number of national and provincial beneficiaries of the foster care grant in South Africa. Table 3.10 displays the optimal order of differencing of the series. Table 3.10 shows that the optimal orders of differencing to remove the trend and seasonality in the series X_t were all $(1-\mathbf{B}^{12})(1-\mathbf{B})$, except for the series for KwaZulu-Natal, whose optimal order of differencing was $(1-\mathbf{B})$. That is, there was a linear trend and seasonality for a period of 12 months in all the series except for that for KwaZulu-Natal which had a linear trend but no seasonality.

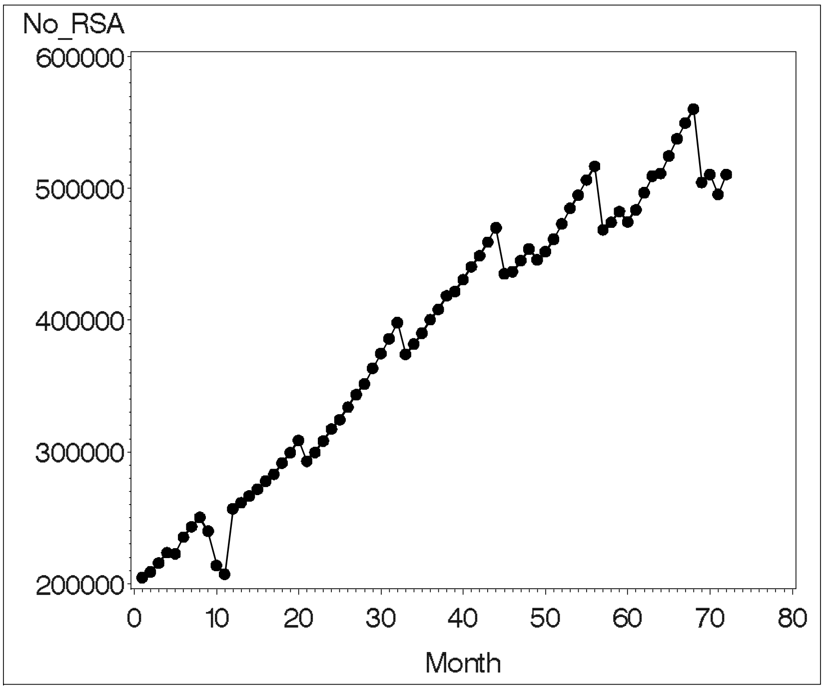

Figure 3.6: The monthly number of beneficiaries of the foster care grant in South Africa versus time (Month = 1 is April 2004 and Month = 72 is March 2010). No-RSA = number of foster care grant beneficiaries in South Africa

Table 3.10: The standard deviation of the differenced monthly number of beneficiaries of the foster care grant in South Africa and South Africa's provinces (April 2004 to February 2009) versus the order of differencing

Order (d, D)	RSA	EC	FS	GP	KZN	LP	MP	NW	NC	WC
(0,0)	97416	18733	7937	11347	28856	11457	6313	8783	2693	2507
(1,0)	19980	4006	2298	2375	**5337**	2131	943	1419	730	1619
(2,0)	27944	5633	3158	3274	7405	3003	1326	2038	991	2228
(1,12)	**19074**	**3968**	**1918**	**2345**	5529	**1977**	**902**	**1290**	**704**	**1638**

Boldfaced numbers are the smallest standard deviations associated with the optimal order of differencing (d, D) to remove trend (d) and seasonality (D), RSA = South Africa, EC = Eastern Cape, FS = Free State, GP = Gauteng, KZN = Kwazulu-Natal, LP = Limpopo, MP = Mpumalanga, NW = North West, NC = Northern Cape, WC = Western Cape

Figure 3.7 displays the autocorrelation function (ACF) of the differenced monthly number of foster care grant beneficiaries in South Africa. The one significant autocorrelation in the ACF suggested the moving average model (MA(q)) for the differenced series. Thus, the fitted model for the differenced series $(1-\mathbf{B}^{12})(1-\mathbf{B})X_t$ was

$$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = \mu + \epsilon_t + \beta_2\epsilon_{t-2} \tag{3.3.5}$$

and the result was

$$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = (1-0.7238)(0.1156)[0.0001]\mathbf{B}^2\hat{\epsilon}_t \tag{3.3.6}$$

where (0.1156) is the standard error of the estimate 0.7238 of β_2 and [0.0001] is the p-value for testing the significance of β_2. This model adequately described the series, as evidenced by the

Portmanteau test statistics (in Table 3.11) for checking/testing whether or not the residuals from fitting the model are white noise. In Table 3.11, the p-values for testing the significance of the autocorrelations of the residuals are greater than 0.9.

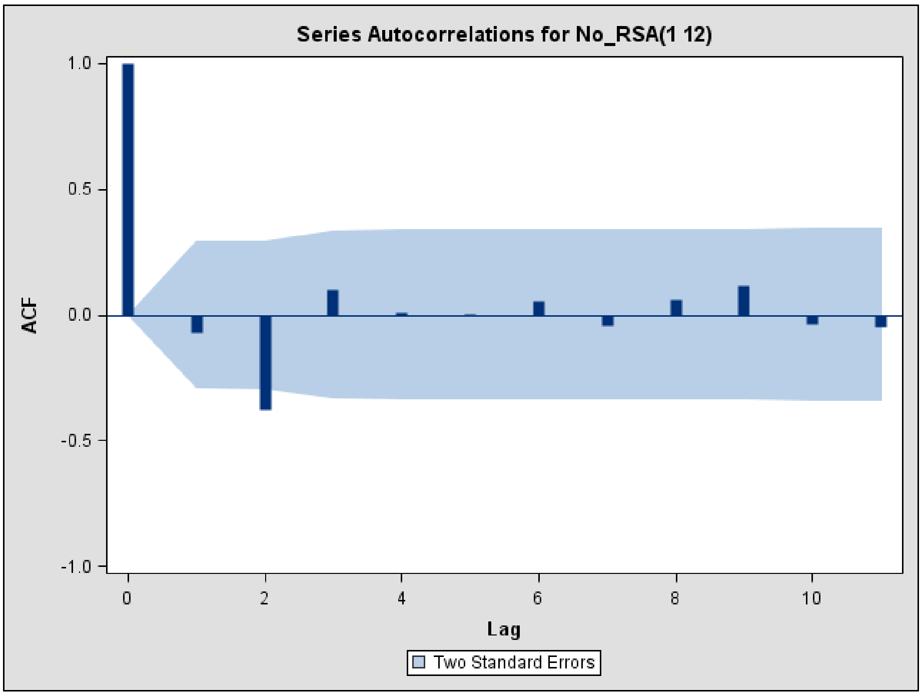

Figure 3.7: The autocorrelation function of the differenced monthly number of beneficiaries of the foster care grant in South Africa (April 2004 to February 2009)

Table 3.11: The Portmanteau test statistics for checking the white noisiness of the residuals from fitting model (3.3.5) to the differenced monthly number of beneficiaries of the foster care grant in South Africa (April 2004 to February 2009)

Autocorrelation check for white noise

To lag	Chi-square	DF	p-value	Autocorrelations					
6	1.54	5	0.9086	-0.047	0.032	0.114	0.024	0.052	-0.099
12	3.23	11	0.9873	0.022	0.104	0.125	0.025	-0.013	0.030
18	4.20	17	0.9993	-0.031	-0.027	0.041	-0.066	-0.061	-0.039
24	6.68	23	0.9996	-0.080	-0.087	0.038	-0.056	-0.089	-0.037

ARIMA models were similarly fitted to each of the differenced provincial monthly number of beneficiaries of the foster care grant in South Africa. The best fitted models are displayed in Table 3.12. The fitted models are not the same, which means that the ARIMA models for the series of the monthly number of beneficiaries of the foster care grant in South Africa are also province specific.

Table 3.12: The models fitted to the differenced provincial number of monthly beneficiaries of the foster care grant in South Africa (April 2004 to February 2009)

Province	Fitted model
EC	$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = \frac{\hat{\epsilon}_t}{(1+0.7384(0.1465)[0.0001]\mathbf{B}^2+0.5407(0.1680)[0.0024]\mathbf{B}^4+0.3081(0.1477)[0.0430]\mathbf{B}^6)}$
FS	$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = \frac{(1-0.4965(0.1785)[0.0079]\mathbf{B}^4\hat{\epsilon}_t}{(10.6761(0.1411)[0.0001]\mathbf{B}^2)}$
GP	$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = \frac{\hat{\epsilon}_t}{(1+0.6082(0.1442)[0.0001]\mathbf{B}^2+0.3076(0.1623)[0.1448]\mathbf{B}^4)}$
KZN	$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = 1383(464)[0.0042] + (1-0.3228(0.1277)[0.0143]\mathbf{B}^2)\hat{\epsilon}_t$
LP	$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = \frac{(1-0.6224(0.1820)[0.0014]\mathbf{B}^4\hat{\epsilon}_t}{(10.7714(0.1353)[0.0001]\mathbf{B}^2)}$
MP	$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = \frac{\hat{\epsilon}_t}{(1+7356(0.1389)[0.0001]\mathbf{B}^2+0.3999(0.1394)[0.0063]\mathbf{B}^4)}$
NW	$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = (1-0.4091(0.1369)[0.0045]\mathbf{B}^2)\hat{\epsilon}_t$
NC	$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = \frac{(1-0.5415(0.1314)[0.0002]\mathbf{B}^2)\hat{\epsilon}_t}{(1+0.9524(0.0551)[0.0001]\mathbf{B}^{12}}$
WC	$(1-\mathbf{B}^{12})(1-\mathbf{B})X_t = \frac{\hat{\epsilon}_t}{(1+0.7932(0.1459)[0.0001]\mathbf{B}^2+0.6396(0.1623)[0.0003]\mathbf{B}^4+0.3132(0.1450)[0.0363]\mathbf{B}^6)}$

( ) = standard error of the parameter estimate, [ ] = *p*-value for testing the significance of the parameter, RSA = South Africa, EC = Eastern Cape, FS = Free State, GP = Gauteng, KZN = Kwazulu-Natal, LP = Limpopo, MP = Mpumalanga, NW = North West, NC = Northern Cape, WC = Western Cape

Model (3.3.6) was used to forecast the national monthly number of beneficiaries of the foster care grant in South Africa from March 2009 to March 2010, and the forecasts of the provincial monthly number of beneficiaries of the foster care grant were obtained using the models in Table 3.12.

The latter provincial forecasts were aggregated to obtain the national forecasts. Figure 3.8 displays the graph of the actual monthly number of beneficiaries of the foster care grant beneficiaries in South Africa (March 2009 to March 2010), and the corresponding national forecasts obtained using model (3.3.6) and by aggregating the provincial forecasts obtained using the models in Table 3.12.

In this case, the national forecasts appear to be better, as they are closer to the actual series than the aggregated provincial forecasts. These observations were confirmed by all the criteria (MSE, MAE, MPE and MAPE) for choosing the best forecasting methods, which were presented in Section 2.5 (see Table 3.13).

Therefore, it can be concluded that forecasting using ARIMA modelling of the national number of foster care grant beneficiaries is better than forecasting by aggregating the provincial ARIMA forecasts.

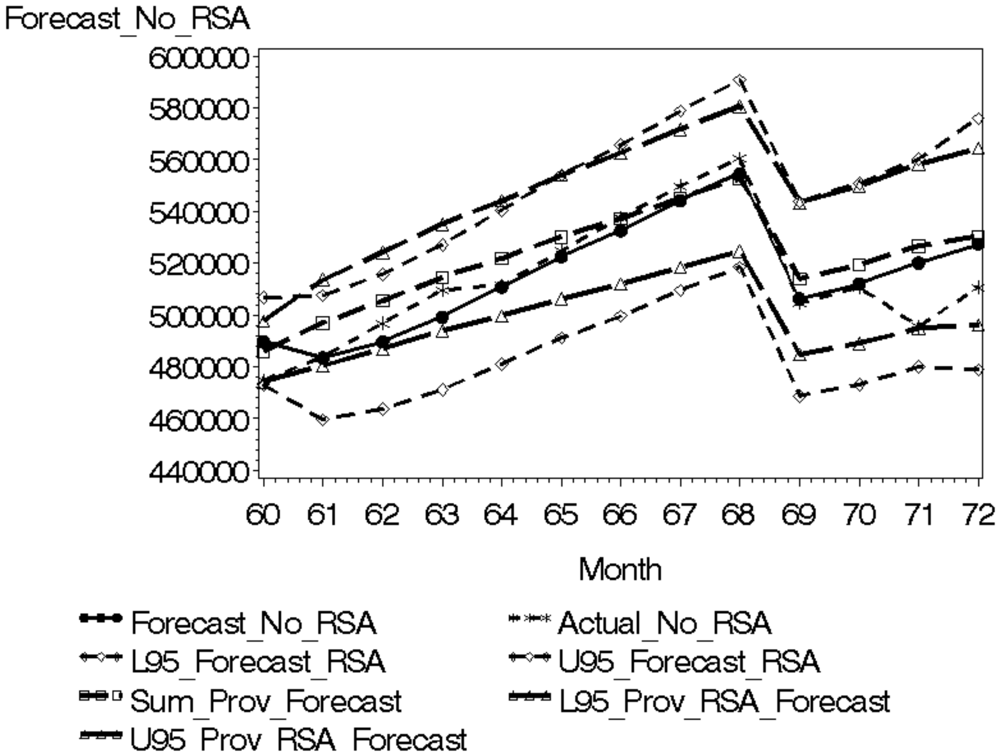

Figure 3.8: The actual monthly number of beneficiaries of the foster care grant in South Africa (March 2009 to March 2010) and the corresponding national forecasts obtained using model (3.3.6) and the provincial models in Table 3.12

Table 3.13: A comparison of forecasts of the monthly number of beneficiaries of the foster care grant in South Africa (March 2009 to March 2010) obtained using the national ARIMA model (3.3.6) and those obtained by aggregating the provincial forecasts obtained using the provincial ARIMA models in Table 3.12

Forecasting method	MSE	MAE	MPE	MAPE
RSA forecasts using model in Table 3.8	13838545	7824	-1.3733	1.5671
Provincial forecasts using models in Table 3.8	300491343	14446	-2.8248	2.8698

MSE= mean square error, MAE = mean absolute error, MPE = mean percentage error, MAPE = mean absolute percentage error

3.3.4 The monthly number of beneficiaries of the disability grant in South Africa

The same procedure for obtaining the forecast number of beneficiaries of the old age grant in South Africa (April 2004 to February 2009) obtained using model (3.3.4) and by aggregating the provincial forecasts (described in Section 3.3.1) was followed to obtain forecasts of the number of disability grant beneficiaries in South Africa (April 2004 to February 2009) through ARIMA modelling of national and provincial series.

Figure 3.9 displays the time series plot of the monthly number of disability grant beneficiaries in South Africa from April 2004 to March 2010. The series is not stationary as it has a nonlinear trend. Thus, differencing is done in order to remove the trend and, hence, stationarise the series.

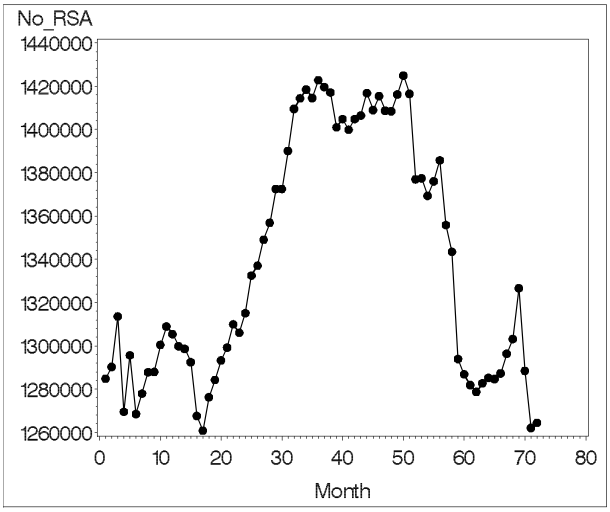

Figure 3.9: The monthly number of beneficiaries of the disability grant in South Africa versus time (Month = 1 is April 2004 and Month = 72 is March 2010). No-RSA = number of disability grant beneficiaries in South Africa

Table 3.14 displays the optimal orders of differencing for the series of the number of beneficiaries of the disability grant in South Africa and its provinces. The trends in the series were all linear even though Figure 3.9 suggested quadratic trends.

Table 3.14: The standard deviation of the differenced number of beneficiaries of the disability grant in South Africa and South Africa's provinces (April 2004 to February 2009) versus the order of differencing

Order (d)	RSA	EC	FS	GP	KZN	LP	MP	NW	NC	WC
0	54760	29181	5194	11024	42632	8781	6585	9365	4590	6909
1	**15192**	**18903**	**2062**	**7199**	**14313**	**1884**	**3366**	**3836**	**1539**	**2153**
2	21181	27856	2571	9996	19791	2557	5676	4570	2162	3209

Boldfaced numbers are the smallest standard deviations associated with the optimal order of differencing (d) to remove trend/seasonality, RSA = South Africa, EC = Eastern Cape, FS = Free State, GP = Gauteng, KZN = KwaZulu-Natal, LP = Limpopo, MP = Mpumalanga, NW = North West, NC = Northern Cape, WC = Western Cape

Figure 3.10 displays the autocorrelation function (ACF) of the once differenced series of the monthly number of beneficiaries of the disability grant in South Africa (April 2004 to February 2009). The

ACF suggests a moving average model for the series. The fitted national model and the provincial models to the once differenced series of the monthly number of beneficiaries of the disability grant in South Africa (April 2004 to February 2009) are given in Table 3.5. These are all different.

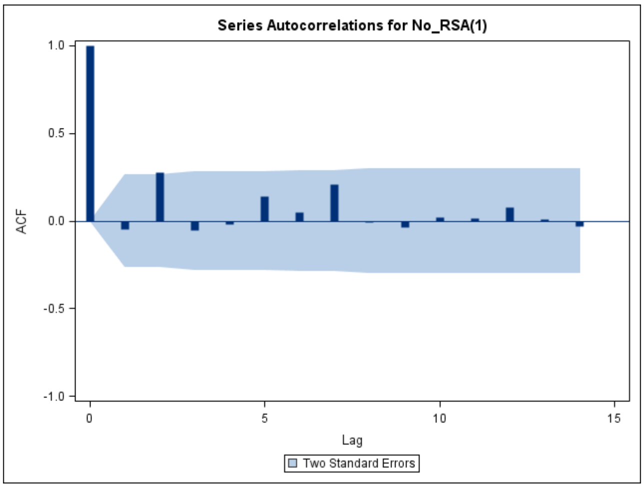

Figure 3.10: The autocorrelation function of the once differenced series of the monthly number of beneficiaries of the disability grant in South Africa (April 2004 to February 2009)

Table 3.15: The fitted models to the monthly number of beneficiaries of the disability grant in South Africa (April 2004 to February 2009)

P/N.	Fitted model
RSA	$(1-\mathbf{B})X_t = (1+0.4192(0.1255)[0.0015]\mathbf{B}^2)\hat{\epsilon}_t$
EC	$(1-\mathbf{B})X_t = (1-0.3523(0.1247)[0.0065]\mathbf{B}^2)\hat{\epsilon}_t$
FS	$(1-\mathbf{B})X_t = \dfrac{\hat{\epsilon}_t}{(1-0.3096(0.1315[0.0220]\mathbf{B}^2}$
GP	$(1-\mathbf{B})X_t = (1-0.6015(0.1161)[0.0001]\mathbf{B}^2)\hat{\epsilon}_t$
KZN	$(1-\mathbf{B})X_t = (1-0.2782(0.1390)[0.0396]\mathbf{B}^2)\hat{\epsilon}_t$
LP	$(1-\mathbf{B})X_t = \hat{\epsilon}_t$
MP	$(1-\mathbf{B})X_t = 294(134)[0.0349] + (1-0.6624(0.1077)[0.0001]\mathbf{B})\hat{\epsilon}_t$
NW	$(1-\mathbf{B})X_t = \dfrac{\hat{\epsilon}_t}{(1-0.2564(0.1247)[0.0445]\mathbf{B}+0.2700(0.1246)[0.0346]\mathbf{B}^3)}$
NC	$(1-\mathbf{B})X_t = \hat{\epsilon}_t$
WC	$(1-\mathbf{B})X_t = \frac{\hat{\epsilon}_t}{(1+0.3200(0.1335)[0.0198]\mathbf{B}^4)}$

( ) = standard error of the parameter estimate, [ ] = p-value for testing the significance of the parameter, P/N = Province/National, RSA = South Africa, EC = Eastern Cape, FS = Free State, GP = Gauteng, KZN = KwaZulu-Natal, LP = Limpopo, MP = Mpumalanga, NW = North West, NC = Northern Cape, WC = Western Cape

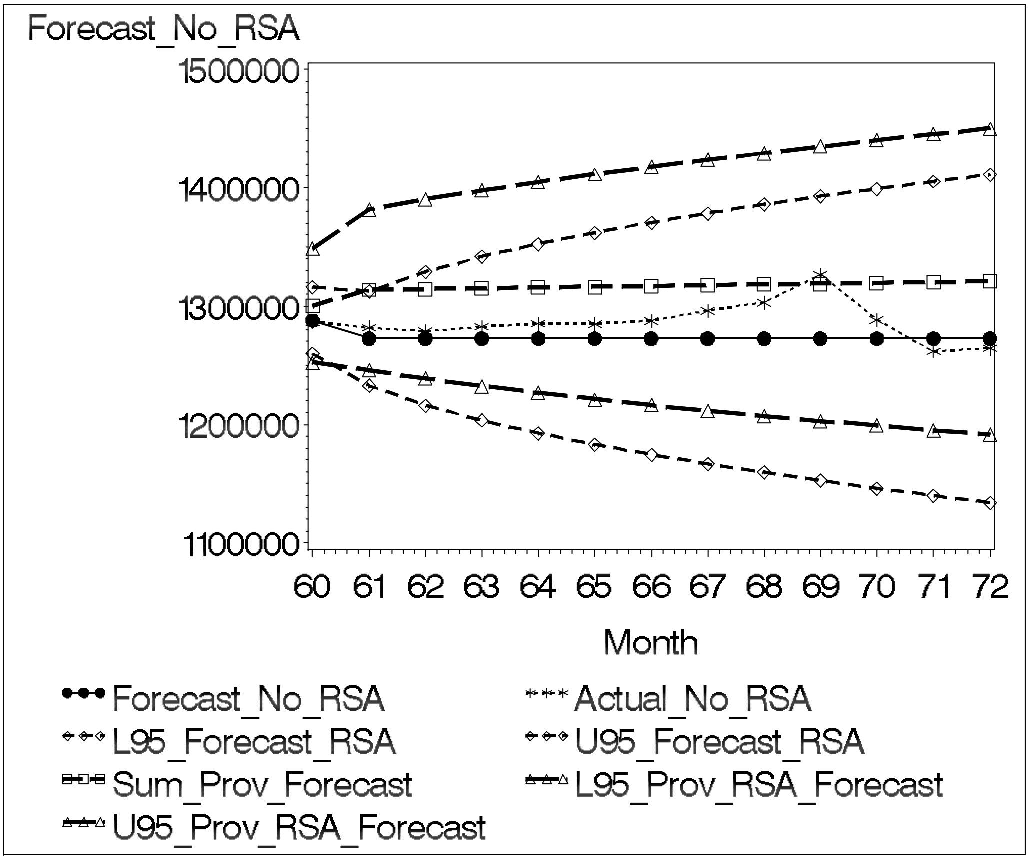

Figure 3.11: The actual monthly number of beneficiaries of the disability grant in South Africa (March 2009 to March 2010) and the corresponding forecasts obtained using the national and provincial models in Table 3.15

Figure 3.11 shows the graph of the actual monthly number of beneficiaries of the disability grant in South Africa (March 2009 to March 2010) and the corresponding forecasts obtained using the national ARIMA model in Table 3.15 and the aggregated provincial forecasts obtained using the provincial models, also in Table 3.15. It is clear that the national forecasts are better since they are closer to the actual series than the aggregated provincial forecasts. This observation is confirmed by the MSE, MAE, MPE and MAPE statistics in Table 3.16. Thus, national forecasting using the ARIMA model of the national number of beneficiaries of the disability grant is better than forecasting by aggregating the provincial ARIMA forecasts.

Table 3.16: A comparison of the forecasts of the monthly number of beneficiaries of the disability grant in South Africa (March 2009 to March 2010) obtained using the national ARIMA model in Table 3.15 and those obtained by aggregating the provincial forecasts obtained using the provincial ARIMA models in Table 3.15

Forecasting method	MSE	MAE	MPE	MAPE
RSA forecasts using model in Table 3.15	430626637	16059	0.9989	1.2377
Provincial forecasts using models in Table 3.15	971676740	28376	-2.0826	2.2160

MSE = mean square error, MAE = mean absolute error, MPE = mean percentage error, MAPE = mean absolute percentage error

3.3.5 Effect of length of series on the accuracy and precision of the forecasts

This is the subsection of the section that follows: (1) the first 27 observations (April 2004 to June 2006) of the original series of the monthly number of beneficiaries of a grant in South Africa were deleted from the series; and (2) the next 32 observations (July 2006 to February 2009) were used to estimate an ARIMA model for forecasting the last 13 observations of the series (March 2009 to March 2010). The accuracy and the precision of the forecasts (March 2009 to March 2010) obtained using the ARIMA model estimated from the 32 observations (July 2006 to February 2009) were compared with those obtained from ARIMA modelling the first 59 observations (April 2004 to February 2009), as was done in the previous section.

3.3.5.1 The monthly number of old age grant beneficiaries in South Africa

The ARIMA model which fitted the 32 observations (July 2006 to February 2009) of the series of the monthly number of old age grant beneficiaries in South Africa is model (3.3.7). This model was used to obtain the forecasts from March 2009 to March 2010. These forecasts were compared with the actual observed values of the series during the same period, and also with the forecast of the series from April 2004 to February 2009 (all 59 observations).

$$(1-\mathbf{B})X_t = \frac{\hat{\epsilon}}{(1-0.3610(0.0009)[0.0001]\mathbf{B}^1 - 0.6390(0.0009)[0.0001]\mathbf{B}^7)} \tag{3.3.7}$$

Table 3.17: A comparison of the forecasts of the monthly number of beneficiaries of the old age grant in South Africa (March 2009 to March 2010) obtained from ARIMA modelling of the series from April 2004 to February 2009 (all 59 observations) and the series from July 2006 to February 2009 (last 32 observations)

Forecasting method	MSE	MAE	MPE	MAPE
RSA forecasts using model (3.3.4)	1465840255	34907	1.3076	1.3983
RSA forecasts using model (3.3.7)	1429877970	31503	-1.1640	1.2626

MSE = mean square error, MAE = mean absolute error, MPE = mean percentage error, MAPE = mean absolute percentage error

Figure 3.12 displays the graph of the actual monthly number of beneficiaries of the old age grant in South Africa (March 2009 to March 2010) and the corresponding forecasts obtained using model (3.3.7). The confidence limits of the forecasts in this figure appear to be wider than the corresponding confidence limits displayed in Figure 3.3. However, the statistics in Table 3.17 indicate that the forecasts from ARIMA modelling the series from July 2006 to February 2009 (last 32 observations) are closer to the actual series than those obtained from ARIMA modelling the series from April 2004 to February 2009 (all 59 observations).

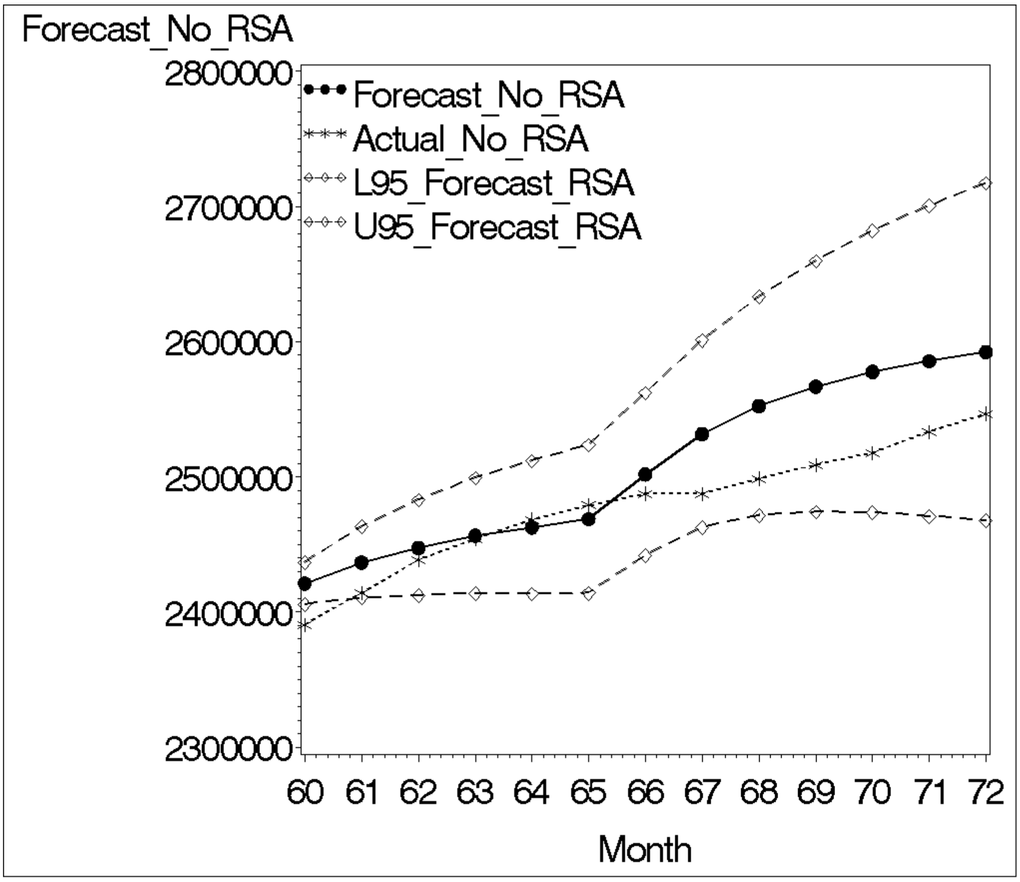

Figure 3.12: The actual monthly number of beneficiaries of the old age grant in South Africa (March 2009 to March 2010) and the corresponding forecasts obtained using model (3.3.7)

3.3.5.2 The monthly number of beneficiaries of the child support grant in South Africa

When the ARIMA models were fitted to the 32 observations (July 2006 to February 2009) of the series of the monthly number of beneficiaries of the child support grant in South Africa, model (3.3.8) was obtained. This model was used to obtain the forecasts from March 2009 to March 2010. These forecasts were compared with the actual observed values of the series during the same period, and also with the forecasts of the series from April 2004 to February 2009 (all 59 observations).

$$(1 - \mathbf{B})X_t = 42112(7703)[0.0001] + (1 + 0.6866(0.1446)[0.0001]\mathbf{B}^1)\hat{\epsilon} \tag{3.3.8}$$

The statistics in Table 3.18 indicate that the forecasts from ARIMA modelling the series from April 2004 to February 2009 (all 59 observations) are closer to the actual series than those obtained from ARIMA modelling the series from July 2006 to February 2009 (last 32 observations). A comparison of the graphs of the actual monthly number of beneficiaries of the child support grant in South Africa (March 2009 to March 2010) and the corresponding forecasts displayed in Figures 3.5 and 3.13 confirms this observation.

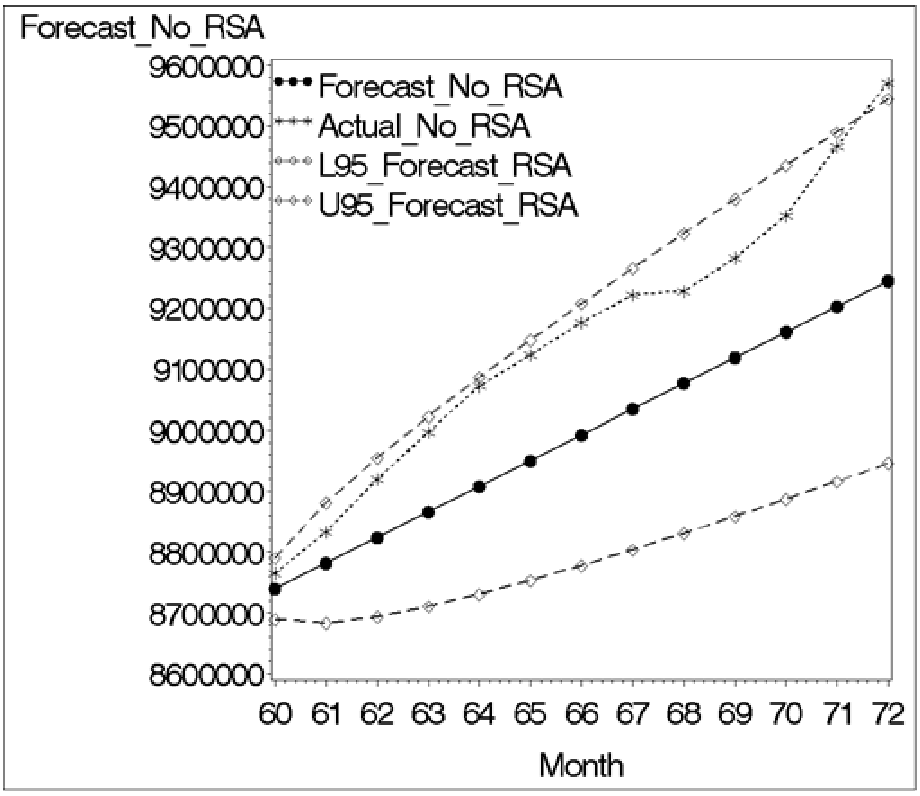

Figure 3.13: The actual monthly number of beneficiaries of the child support grant in South Africa (March 2009 to March 2010) and the corresponding forecasts obtained using model (3.3.8)

Table 3.18: A comparison of the forecasts of the number beneficiaries of the child support grant in South Africa (March 2009 to March 2010) obtained from ARIMA modelling of the series from April 2004 to February 2009 (all 59 observations) and the series from July 2006 to February 2009 (last 32 observations)

Forecasting method	MSE	MAE	MPE	MAPE
RSA forecasts using model in Table 3.8	28017847944	140461	-1.4527	1.5137
RSA forecasts using model (3.3.8)	32080281982	162131	1.7524	1.7524

MSE = mean square error, MAE = mean absolute error, MPE = mean percentage error, MAPE = mean absolute percentage error

3.3.5.3 The monthly number of beneficiaries of the foster care grant in South Africa

ARIMA modelling of 32 observations (July 2006 to February 2009) of the series of the monthly number of beneficiaries of the foster care grant in South Africa obtained model (3.3.9), which was then used to obtain the forecasts from March 2009 to March 2010. These forecasts were compared with the actual observed values of the series during the same period, and also with the forecasts of the series from April 2004 to February 2009 (all 59 observations).

$$(1 - \mathbf{B}^{12})(1 - \mathbf{B})X_t = \hat{\epsilon} \tag{3.3.9}$$

The statistics in Table 3.19 indicate that the forecasts from ARIMA modelling the series from July 2006 to February 2009 (last 32 observations) are closer to the actual series than those obtained from ARIMA modelling the series from April 2004 to February 2009 (all 59 observations). A comparison of the graphs of the actual monthly number of beneficiaries of the foster care grant in South Africa (March 2009 to March 2010) and the corresponding forecasts displayed in Figures 3.8 and 3.14 is inconclusive other than that the confidence limits of the forecasts appear to be the same.

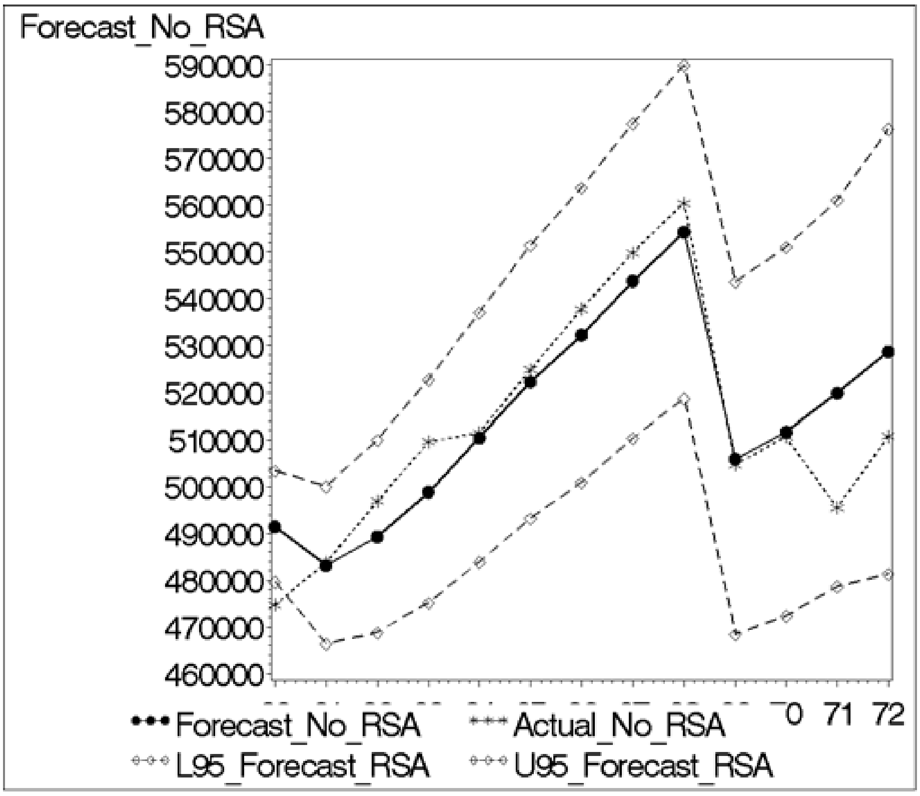

Figure 3.14: The actual monthly number of beneficiaries of the foster care grant in South Africa (March 2009 to March 2010) and the corresponding national forecasts obtained using model (3.3.9)

Table 3.19: A comparison of forecasts of the monthly number beneficiaries of the foster care grant in South Africa (March 2009 to March 2010) obtained from ARIMA modelling of the series from April 2004 to February 2009 (all 59 observations) and the series from July 2006 to February 2009 (last 32 observations)

Forecasting method	MSE	MAE	MPE	MAPE
RSA forecasts using model in Table 3.8	13838545	7824	-1.3733	1.5671
RSA forecasts using model (3.3.9)	113961920	7803	-0.3617	1.5424

MSE = mean square error, MAE = mean absolute error, MPE = mean percentage error, MAPE = mean absolute percentage error

3.3.5.4 The monthly number of beneficiaries of the disability grant in South Africa

Finally, ARIMA modelling of 32 observations (July 2006 to February 2009) of the series of the monthly number of beneficiaries of the disability grant in South Africa obtained model (3.3.10), which was then used to obtain the forecasts from March 2009 to March 2010. These forecasts were compared with the actual observed values of the series during the same period, and also with the forecasts of the series from April 2004 to February 2009 (all 59 observations).

$$(1-\mathbf{B})X_t = \hat{\epsilon} \tag{3.3.10}$$

The statistics in Table 3.20 indicate that the forecasts from ARIMA modelling the series from July 2006 to February 2009 (last 32 observations) are closer to the actual series than those obtained from ARIMA modelling of the series from April 2004 to February 2009 (all 59 observations). A comparison of the graphs of the actual monthly number of beneficiaries of the disability grant in South Africa (March 2009 to March 2010) and the corresponding forecasts displayed in Figures 3.11 and 3.15 indicates that the confidence limits of the forecasts in Figure 3.15 are more precise than the corresponding confidence limits in Figure 3.11.

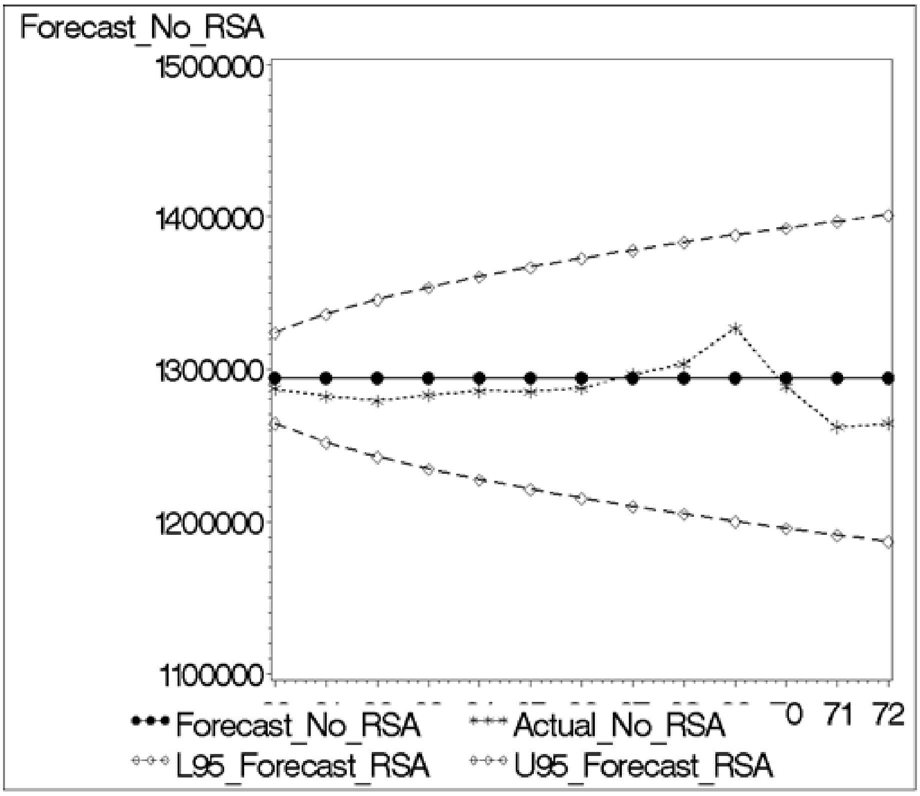

Figure 3.15: The actual monthly number of beneficiaries of the disability grant in South Africa (March 2009 to March 2010) and the corresponding forecasts obtained using (3.3.10)

Table 3.20: A comparison of the forecasts of the monthly number of beneficiaries of the disability grant in South Africa (March 2009 to March 2010) obtained from ARIMA modelling of the series from April 2004 to February 2009 (all 59 observations) and the series from July 2006 to February 2009 (last 32 observations)

Forecasting method	MSE	MAE	MPE	MAPE
RSA forecasts using model in Table 3.15	430626637	16059	0.9989	1.2377
RSA forecasts using model (3.3.10)	295170049	13959	-0.56901	1.08612

MSE = mean square error, MAE = mean absolute error, MPE = mean percentage error, MAPE = mean absolute percentage error

Chapter 4

Conclusion and Recommendations

To address disparity in the awarding of social security grants and to improve the welfare system, the National Department of Social Development (NDSD) appointed the South African Social Security Agency (SASSA) as its agency for the purpose of implementing the norms and standards set by the NDSD for improving the delivery of social assistance grants to deserving South Africans. Success to date includes a huge increase in the number of social security grant beneficiaries from 6.4 million in 2004 to 14 million in 2010.

Given the crude analysis of social security grants in South Africa, the main objective of this thesis was to investigate the feasibility of accurately and precisely forecasting the number of both national and provincial beneficiaries of social security grants in South Africa using autoregressive integrated moving average (ARIMA) models. In particular, the objectives of this thesis were to: (1) find simple ARIMA models for accurately and precisely forecasting the number of both national and provincial beneficiaries of various types of social security grant to be used by the South African government in planning and making policy decisions; (2) confirm the hypothesis that the simple ARIMA models for accurately and precisely forecasting the number of provincial beneficiaries of various types of social security grant are specific to province and the type of social security grant; (3) compare the accuracy and precision of the forecasts of the number of national beneficiaries of social security grants obtained from ARIMA modelling with those obtained by aggregating forecasts from ARIMA modelling of the number of provincial beneficiaries of social grants; and (4) to investigate the effect of the length of the series of the number of beneficiaries of social security grant series on the accuracy and precision of the forecasts of the future values of the series.

In this thesis, the social security grant data series of the monthly number of beneficiaries of the old age, child support, foster care and disability grants from April 2004 to March 2010 were used to achieve the objectives of the thesis. Graphs of actual series versus forecasts, the mean square error, the mean absolute error, the mean percentage error, and the mean absolute percentage error of the forecasts were used to evaluate the accuracy and precision of the forecasts and the forecasting methods. The results of the data analysed confirmed the hypothesis that: (1) the ARIMA models for forecasting the number of provincial beneficiaries of various types of social security grant are specific to the province and type of social security grant and; (2) for some social security grants (e.g. old age and child support) the forecasts of the national number of beneficiaries of the grant obtained by aggregating the provincial ARIMA forecasts are better than those obtained from ARIMA modelling the national series. However, forecasts of social grants (foster care and disability) obtained by national ARIMA forecasts are better than those obtained by aggregated provincial ARIMA forecasts.

The effect of length of old age grant, child support grant, foster care grant and disability grant series on the accuracy and precision of the forecast number of national and provincial beneficiaries was investigated. It is also concluded that for some social security grants (e.g. old age, foster care and disability), the forecasts of the national number of beneficiaries obtained from ARIMA modelling of the latest half of the national series are better than those obtained from ARIMA modelling of the full series. However, forecasts of the national number of child support grant beneficiaries obtained from ARIMA modelling of the full series of the national beneficiaries are better than those obtained from ARIMA modelling of the latest half series.

Given the success in terms of the accuracy and precision of the forecast number of national ARIMA forecasts on the social security grants, namely, (foster care, disability) and the aggregated provincial ARIMA forecasts on the old age grant and child support grant, it is recommended that the South African government should consider implementing methods for forecasting the number of social security grants for both aggregated provincial ARIMA forecasts and national ARIMA forecasts to achieve better results in the distribution of social security grants.

It is also recommended that the government should consider the length of the social security grant series forecasts of the number of beneficiaries obtained from ARIMA modelling of the latest half series and full series of the national beneficiaries in order to choose the best ARIMA forecast result. The findings of this report illustrate that three social security grants, namely, the old age, foster care and disability, provide success in terms of the accuracy and precision of the forecast number of beneficiaries of ARIMA modelling the latest half series; whereas the full series of child support grant beneficiaries is better. With all these findings, it is recommended that the South African government should consider using ARIMA forecast results that are accurate and precise for planning and making policy decisions in future.

There are number of limitations related to the study of this thesis. For example, the first differencing was used to accomplish stationarity in the series modelling. However, differencing only removes polynomial trends and seasonality in the series, it does not remove nonlinear trends. As a result, areas for further research include investing more highly developed methods to remove nonlinear trends in the series. This further research will also include transformation to linearise nonlinear trends to stabilise variances. Further direction of this study is also to use seasonal modelling and the length of the series to analyse the data which need to be long enough.

Bibliography

Box G.E.P. 1976. *Time series analysis: forecasting and control.* San Fransciso, Calif: Holden Day.

Box G.E.P., and Jenkins G.M. 1970. *Time series analysis: forecasting and control.* San Fransciso, Calif: Holden Day.

Constitution of the Republic of South Africa Act 108. 1996. *Republic of South Africa.* Pretoria: Government Printers.

Chatfield C. 2004. *The analysis of time series: an introduction.* Sixth edition. Florida: CRC Press.

Criminal procedure Act 51. 1997. *Republic of South Africa.* Pretoria: Government printer.

Cryer J.D., and Chan K.S. 2008. *Time series analysis with applications in R.* Second edition. Springer Science + Business media: USA.

Department of Social Welfare. 1997. *White paper for Social Welfare.* Pretoria: Government Printers.

Diggle P.J. 1990. *Time series: a biostatistical introduction.* Oxford: Clarendon Press.

Gooijer J.G., Abraham B., Gould A. and Robinson L. 1985. Methods for determining the order of an autoregressive−moving average process: A survey. *Int. Statist. Rev,* 53:301−29.

Montgomery D.C., Jenning C.L., and Kulanci M. 2008. *Introduction to time series analysis and forecasting.* Canada: Wiley and Hoboken.

Reikart G. 2009. Prediction of radiation at high resolutions: A comparison of time series forecast. *Solar Energy,* 83:342−349.

SAS Version 9.2. 2004. *SAS Institute Inc.* Cary, North Carolina

Social Assistance Act 13. 2004. *Republic of South Africa.* Pretoria: Government Gazette.

Social Assistance Act 6. 2008. *Republic of South Africa.* Pretoria: Government Gazette.

South African Social Security Agency. 2010. *Social Pension System Database.* www.sassa.gov.za.

Wei W.W.S. 2006. *Time series analysis: Univariate and multivariate methods.* Second edition. USA: Addison-Wesley.