Department of Statistics
Bachelor's Thesis (15 credits)
Fall 2008

Birth trends and the lunar connection for Lund's community 1995-2004

Cox Communications Inc.

Authors:
Rebecca Rylance
Oxana Tarassiouk

Supervisor:
Mats Hagnell

Abstract

‘Birth trends and the lunar connection for Lund’s community 1994-2005’ analyses natural births using statistical methods. With the aid of different time series models a pattern was found, namely that most natural births occur during the spring months and the smallest number occur during the month of November. An explanation for this could be that many people take vacation during the summer months, which gives them more time to reproduce during these months.

The original idea behind this study was to examine if there was some proof behind the myth that the Full Moon influences women to go into labor. The objective with the study was to find a pattern between natural births in Lund's municipality during the period 1994-2005 and the Moon's phases. No connection could be established between natural births and the Moon's phases, neither for the Moon's four or eight phases. The analysis of the Moon's phases produced no significant differences in the phases, which means that there is more or less the same number of births during each phase.

It would be more logical to consider the synodic period, which is one day on the Moon or the equivalent of 29.53 earth days. A suggestion then for future studies would be to see if a pattern could be established between births and an average synodic period.

Sammanfattning

I 'Birth trends and the lunar connection for Lunds community 1994-2005' analyseras naturliga födslar med statistiska metoder. Med hjälp av olika tidsseriemodeller hittades ett mönster, nämligen att det var flest naturliga födslar under våren och minst under novembermånad. En förklaring till det kan vara att många personer brukar ta ut semester under sommaren, vilket ger mer tid att reproducera under dessa månader.

Den ursprungliga tanken med den här studien var att undersöka om det finns något belägg för myten om att fullmåne påverkar antalet naturliga barnafödslar. Målet med studien var att med hjälp av statistik hitta ett mönster mellan de naturliga födslarna i Lunds kommun under tidsperioden 1994-2005 och månens olika faser. Inget samband kunde fastställas mellan de naturliga födslarna och månens olika faser, varken för månens fyra eller åtta faser, vilket betyder att det var ungefär samma mängd naturliga födslar under respektive fas.

Ett förslag till framtida studier är att ta hänsyn till den sinediska omloppstiden (den tiden det tar för månen att gå ett varv runt solen, vilket är ekvivalent till 29.53 dagar på jorden) i analysen av sambandet mellan månen och födslarna.

We would like to thank Karin Källén, Tornblad Institute, for her assistance in data collection.

Table of Contents

1. Introduction

1.1. Background

Communities are interested in knowing when women are going to deliver their babies. Such groups in the community are delivery wards, midwives, families expecting babies and obstetricians. The idea for this paper comes from a discussion with two midwives at Lund's Women's clinic.

1.2. Goals of this paper

Birth Patterns

The first goal is to compare the days of the week and the months of a year in order to see if a birth pattern can be found. If a birth pattern can be established, we will develop a time series model to describe this pattern.

Lunar effects

The second goal is to examine the lunar effects on births. We will be checking for a relationship between the number of babies born and the lunar phases, both four and eight phases. This study should reveal if there is a phase or phases during the lunar cycle when more babies are born and develop a time series model to represent it.

1.3. Exclusion/Inclusion criteria

The data is from the epidemiology center in Lund and encompasses birth records from the beginning of 1995 to the end of 2004 in Lund, Sweden. An interview was held with Karin Källén, an epidemiologist who gave us access to the data and also advised us on what data modifications were needed in our study (Rylance, 2008). The data contains only spontaneous singleton births between pregnancy weeks thirty-seven to forty-one. A baby is considered premature if the baby is born before pregnancy week 37 and overdue if the baby arrives after week 41, which means all births in this study are of a normal pregnancy length. Only certain types of deliveries are included in this study. No cesarean sections[1], inductions, scheduled deliveries or deliveries that used any type of instrumentation[2] are included. Women who

[1] Cesarean section means the surgical removal of the fetus (Hamberger & Nilsson, 2003)
[2] Instrumentation vacuum extraction or forceps (Hamberger & Nilsson, 2003)

received oxytocin[3] are also excluded from the data. These modifications were done because the goal of the study is to look for a natural pattern in births.

1.4. Relevant studies

At the University of Arizona, a five-year study encompassing birth trends, the lunar cycle, and certain atmospheric conditions was done on spontaneous deliveries (Morton-Pradhan, Bay & Coonrod, 2005). Only women who were pregnant between weeks 37 to 40 weeks were included in the study. All inductions and cesarean sections were excluded. A weekly trend was found with Sundays having the least amount of births with the birth rate increasing and reaching its top number on Thursday and then decreasing again. The lunar effects on births were also analyzed in this study. The dates of the New Moon were coded '1' and checked against all other dates, which were coded '0' and the same procedure was performed with the Full Moon by coding the Full Moon '1' and comparing against all other dates being coded '0'. No connection between the number of births and the lunar cycle was found.

Another five-year study of 564,039 births out of North Carolina did not find any significant difference in the frequency of births during any one phase of the Moon (Arliss, Kaplan & Galvin, 2005). In this study, the Moon cycle was divided into eight phases comparing the number of births using one-way analysis of variance. The differences in deliveries were also divided into many categories, of which to name a few: all births, births that were not induced, vaginal deliveries, 9 categories of complicated births, caesarean deliveries, etc.

Other studies, did however find a link between the lunar cycle and the incidence of birth. One large French study of nearly six million births found a peak in births between the Last Quarter and a few days before the New Moon, which was found to be significant (Guillon et al., 1986). They also found a weekly trend with the highest number on Tuesdays, and the lowest number on Sundays and the second lowest number on Saturdays. They also found a maximum number of births during the spring months, meaning conception occurred during the previous summer months and the lowest number of births during the fall.

Two other studies found a significant number of births during the lunar cycle, but in this instance it was around the time of the Full Moon. One of those studies was over a two-year period in Brazil (Mikulecky & Lisboa, 2002) and the other in Italy (Ghiandoni et al., 1998). The Italian study used an average lunar month cycle of 29.53 days. They examined all spontaneous full term births between weeks 38-42. The study separated first time mothers and mothers delivering for the second or more times. The results were significant for women who delivered for their second time or subsequent time and multiple birth pregnancies surrounding the days of the Full Moon. Unfortunately, the significance was too weak to be able to predict which days would experience the largest amount of births.

[3] Oxytocin is a contraction-stimulating drug, given by intravenous drip (Hamberger & Nilsson, 2003). Around fifty percent of women who go into labor naturally receive oxytocin (Rylance, 2008).

1.5. Outline

Section two continues with information about the Moon. Section three gives a more detailed description of our data. Section four discusses statistical theory. Section five discusses the results. Section six gives the conclusion. Section seven provides the sources.

2. The Moon

2.1. Introduction of the Moon

The fascination with the Moon has been with us for centuries. From the great philosophers, like Plutarch, Hippocrates and Aristotle, from the Bible to the Koran, the Moon has been linked to many physiological phenomena among other things like: epilepsy, menstruation, membranes breaking, quantity of food eaten, etc. (Muñoz-Delgado, J. et al., 2000). The actual word lunacy comes from the name Luna, who is the Roman Moon goddess (Dictionary of Psychology, 2001).

How much of this is folklore and how much can be scientifically proven? Our hope is to have more science and less myth in our statistical study concerning birth trends and the lunar effects on births.

2.2. Information about the Moon

The Moon will need further explanation in order to better understand our study. If we speak in general terms, from the perspective of someone standing on the ground and looking up at the night sky, the Moon goes through a cycle from being unseen, growing, where gradually more of Moon can be seen to being a Full Moon, then diminishing in size. When the Moon is in between the sun and the earth, we cannot see the Moon with the naked eye. When it is a full Moon, the Earth is between the Sun and Moon. Moving in its own nearly circular orbit around the Earth, the Moon's mean time for it to reappear in the sky relative to the Sun and observed from the Earth is 29.53 days, almost one Gregorian calendar month, which is 30.44 days (Whipple, 1963). This is called a synodic period or month. In other words, the mean time it takes for the Moon to make one revolution around the Earth in relation to the Sun is 29.53 days. The sidereal period or month, on the other hand, is the time it takes for the Moon to return to its same spot in relationship to the stars, which is 27.32 days (Cadogan, 1981). Since the synodic month and sidereal month differ, we choose to use the synodic month over the sidereal month because the phase of the Moon is easier to establish then its relation to the stars.

The four primary lunar phases are the New Moon, First Quarter, Full Moon and Last Quarter (Meeus, 1997). The four primary phases (See Diagram 2.1) of the Moon are easily enough understood if you imagine a circle with 360 degrees, which represents the Moon's orbit around the earth, then dividing this circle into four equal parts of 90 degrees and each phase is a quarter of 360 degrees, even if the Moon does not orbit the earth in an exact circle, but a non uniform elliptic orbit (Whipple, 1963). 'By definition, the times of New Moon, First Quarter, Full Moon, and Last Quarter are the times at which the excess of the apparent longitude of the Moon over the apparent longitude of the Sun is exactly 0, 90, 180, 270 degrees' (Meeus, 1997).

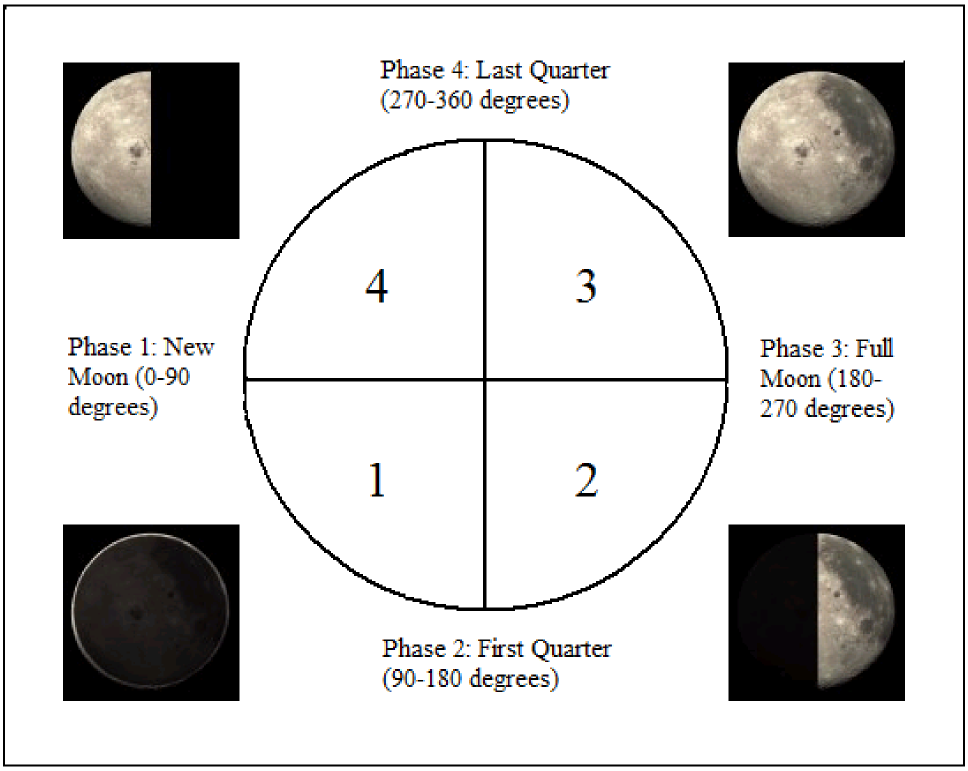


Diagram 2.1: 4 Moon phases (MoonPhases.Info, 2008; modified version)

When the Moon is divided into eight phases (See Diagram 2.2), the divisions of the phases are divided up quite differently (Quick Phase Pro, 2008). A description of each follows:

New Moon: The Moon's unilluminated side is facing the Earth (354.01 to 6 degrees).

Waxing Crescent: The Moon is less than one-half illuminated by direct sunlight (6.01 to 84 degrees).

First Quarter Moon: One-half of the Moon appears to be illuminated by direct sunlight (84.01 to 96 degrees).

Waxing Gibbous: The Moon appears to be more than one-half but not fully illuminated by direct sunlight (96.01 to 174 degrees).

Full Moon: The Moon's illuminated side is facing the Earth (174.01 to 186.0 degrees).

Waning Gibbous: The Moon is more than one-half but not fully illuminated by direct sunlight (186.01 to 264 degrees).

Third Quarter Moon or Last Quarter Moon: One-half of the Moon appears to be illuminated by direct sunlight (264.01 to 276 degrees).

Waning Crescent: The Moon is less than one-half illuminated by direct sunlight (276.01 to 354 degrees).

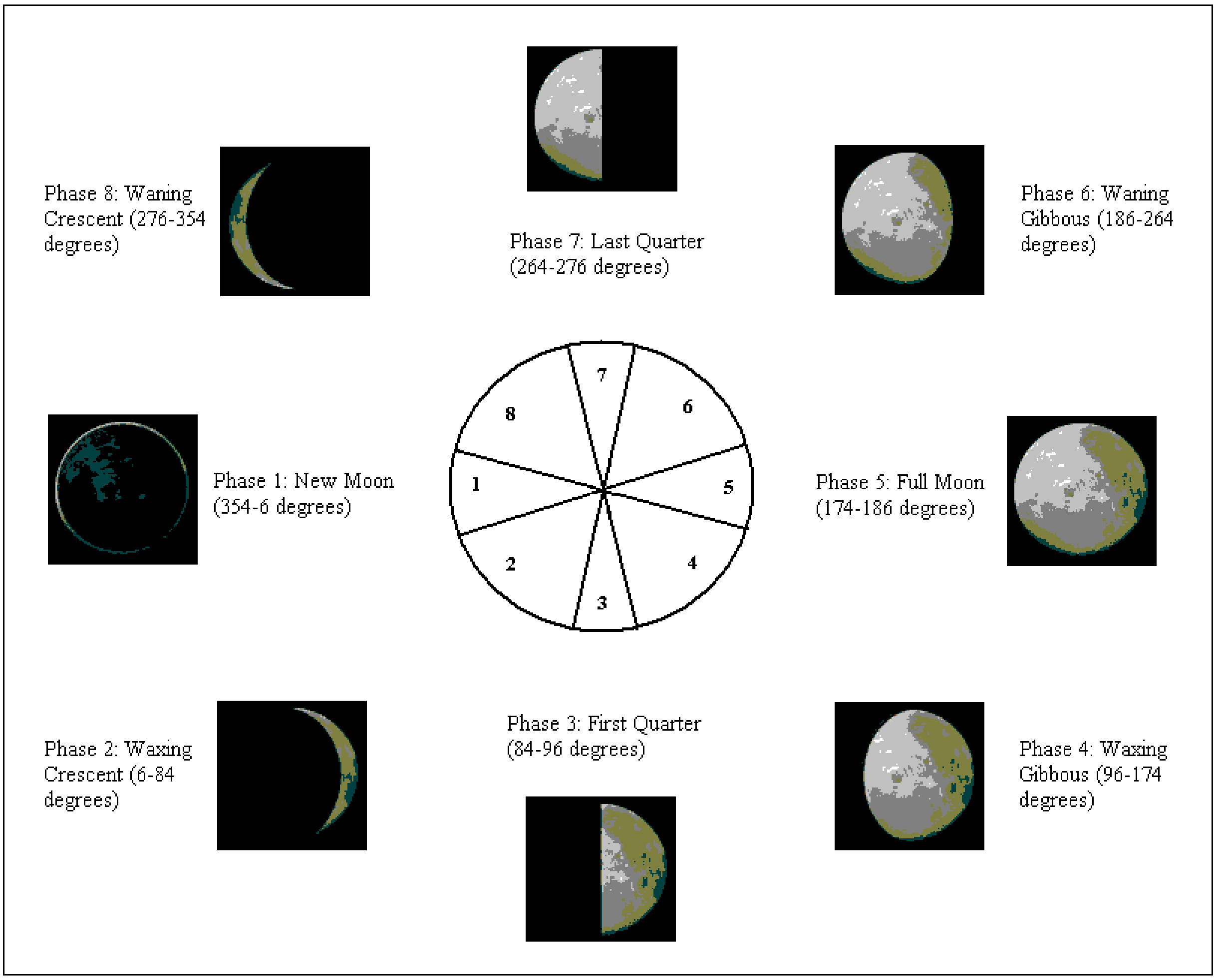


Diagram 2.2: 8 Moon phases (MoonPhases.Info, 2008; modified version)

3. Data

3.1. Data processing

In the Excel program, we coded the number of births for each day, month and year for ten years of data. More specifically, we coded each birth with a day of the week from 1-7. Sunday represents 1 and Monday 2 and so on. The same procedure was done for the months of the year. The number 1 corresponds to January and so forth. Afterwards the frequencies of births for every week, every month, and every year were calculated.

We also assigned each of the births to one of the four Moon phases. Two sources of possible lunar tables were recommended to us by Daniel Malmberg, an employee at the Astronomical Department at Lund's University (Rylance & Tarassiouk, 2008). One of the sources was the U.S. Navy and the other was NASA, but it did not matter which table we used, because they were the same. The lunar tables we used are from NASA's website (NASA, 2007). The lunar table is divided into four phases: New Moon, First Quarter Moon, Full Moon, and Last Quarter Moon. Each one of the four phases is 90 degrees, but the number of days in each phase variates, so we calculated the exact length of each individual phase for four phases for 10 years of data.

The eight phases of the Moon were calculated on-line using a Moon phase calculator because no table for 8 phases could be found (Clarke, 2008). The lengths of the phases when divided into 8 are not divided into 8 equal parts, which was described earlier, but nevertheless each of the 8 phases was also calculated exactly for each phase each year for 10 years. We accomplished this by dividing the births by the number of days that pertained to each particular phase for each year for 10 years of data.

3.2. Adjustments

Adjustments have been made to take into account the Swedish summer and winter times. In 1995, summer time was from the first Sunday in March until the last Sunday in September (Rylance & Tarassiouk, 2008). From 1996 on, summer time was from the first Sunday in March until the last Sunday in October.

4. Method

4.1. Short overview

We use an analysis of variance in our study and discuss the assumptions that need to be followed in order to perform this. We also describe time series models following the Box-Jenkins methodology.

4.2. Analysis of Variance

Analysis of variance is used to compare means of several independent groups (Kleinbaum, Kupper & Muller, 1988). The data needs to pass certain normality tests and the variances need to be deemed equal before the analysis of variance can be undertaken. The analysis of regression is a parametric test and assumes that the population has a certain distribution. The Kruskal-Wallis test can be used if the normality tests are not satisfactory (Montgomery, 1997). Kruskal-Wallis is a nonparametric test that assumes that the population does not have a specific distribution. A nonparametric test is not as powerful or effective as a parametric test.

Another assumption that needs to met when using ANOVA is that the observations need to be mutually independent. In typical time series data this assumption is not met because the time series observations are dependent on one another. The time series data that is used in this study are atypical because the number of births from one day to the next does not depend on the following day, so the ANOVA procedure is considered appropriate in this case.

4.2.1. Test of Equal Variances

In the case of testing the variances, one can use the Bartlett's test, which closely follows the chi-square distribution (Montgomery, 1997). The normality assumption must be met when using this test.

Another test, which is not so sensitive to the normality assumption, is the modified Levene's test, which is like the *F*-test, except this test uses the absolute deviation of the observations in each treatment from the treatment median to test the equality of the means (Montgomery, 1997).

The hypotheses for test of equal variances are as follows:

H_0: the variances are equal

H_1: the variances are not equal

Barlett's test

The test compares the differences in the weighted arithmetic averages and the weighted geometric averages (Minitab, 2008). If the difference is large enough, than the variances are most likely not equal. The null hypothesis that the variances are equal is rejected if the *p*-value is smaller than the chosen significance level.

The test statistic is

$$B = \frac{(\sum v_i)\ln(\sum v_i s_i^2 / \sum v_i) - \sum v_i \ln s_i^2}{1 + \left\{\sum(\frac{1}{v_i}) - 1/\sum v_i\right\}/\{3(k-1)\}} \tag{4.1}$$

where $s_i^2 = \sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}\right)^2/(n_i - 1)$, k = the number of samples, and $v_i = n_i - 1$.

Levene's test

This test in Minitab (2008), calculates the distance of the observations from their sample median instead of their sample mean. If the probability value is smaller than the significance level, the null-hypothesis is rejected.

$$L = \frac{(N-k)\sum n_i(\bar{V}_i - \bar{V})^2}{(k-1)\sum\sum(V_{ij} - \bar{V}_i)^2} \tag{4.2}$$

where $V_{ij} = \left|X_{ij} - \tilde{X}_i\right|, i = 1, \dots, k, j = 1, \dots, n_i$ and $\tilde{X}_i$ = median$\left\{X_{i1}, \dots, X_{in_i}\right\}$.

4.2.2. Normality Test

We performed the Anderson-Darling normality test in Minitab. The hypotheses for the normality test are as follows:

H_0: the data follow a normal distribution

H_1: the data do not follow a normal distribution

Anderson-Darling

The Anderson-Darling normality test compares the expected distribution if the data were normally distributed with the empirical cumulative distribution function of the sample data by measuring the area (Minitab, 2008).

Test statistic: the Anderson-Darling test statistic is defined as

(4.3) $$A^2 = -N - \left(\frac{1}{N}\right)\sum(2i-1)\left(\ln F(Y_i) + \ln\left(1 - F(Y_{N+1-i})\right)\right)$$

where F is the cumulative distribution function of the normal distribution, Y_i are the ordered observations.

4.2.3. *F*-test

The *F*-test is a parametric test and assumes that the errors have a mean of zero, are independently and normally distributed and that the variances are equal (Montgomery, 2005). One way to check if these assumptions are met is to evaluate the residuals. A normal probability plot of the residuals can be examined to see if the underlying error distribution is normal and in this case the residuals should follow a straight line. The residuals should be structureless if we are plotting the residuals versus the fitted values.

The hypotheses for the *F*-test are as follows:

H_0: the treatment means are equal

H_1: the treatment means are not equal

The *F*-test determines whether the treatment means are equal or not (Kleinbaum, Kupper & Muller, 1998). If the null hypothesis is true, then the quotient follows an *F*-distribution, which may only have positive values. The *F*-value is then calculated by taking the $MS_{Treatments}$ and dividing it by MS_E.

(4.4) $$SS_T = n\sum_{i=1}^{a}(y_{ij} - \bar{y}_{**})^2$$

(4.5) $$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i*})^2$$

(4.6) $$SS_{Treatments} = SS_T - SS_E$$

(4.7) $$MS_{Treatments} = \frac{SS_{Treatments}}{a-1}$$

where a is the number of factor levels.

(4.8) $$MS_E = \frac{SS_E}{N-a}$$

(4.9) $$F_0 = \frac{MS_{Treatments}}{MS_E}$$

A *p*-value is used to decide whether to reject the null hypothesis or not at any level of significance (Montgomery, 2005). The *p*-value is the smallest level of significance that concludes in the rejection of the null hypothesis. After the *p*-value has been established for a test, the analyst can determine how significant the data are without pre-selecting a level of significance.

4.2.4. Kruskal-Wallis

The Kruskal-Wallis test is a nonparametric test (Montgomery, 2005). With this type of test one assumes that the samples are independent random samples from different populations with the same shape (Minitab, 2008). This test measures if the population medians are equal or not. The median is a measure of central tendency among the observations. The observations are ranked in order from smallest to largest and each observation is replaced by its rank, R_i (Montgomery, 2005). If the number of observations is an even number the median is equal to the average of the middle two observations. If the number of observations is an odd number the median is simply the middle value.

The hypotheses for the Kruskal-Wallis test are as follows:

H_0: the population medians are equal

H_1: the population medians are unequal

Generally speaking, the different samples are ranked all together and then the average of each sample is calculated.

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{a} \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4}\right] \tag{4.10}$$

where

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{a} \sum_{j=1}^{n_i} R_i^2 - \frac{N(N+1)^2}{4}\right] \tag{4.11}$$

After calculating the *p*-value for the Kruskal-Wallis test, the analyst may determine how significant the data are (Montgomery, 2005).

4.3. Time Series Analysis

Time series analysis allows for a deeper understanding of data over a period of time. In theory, the observations made at different times need to be statistically dependent to one another, such is the nature of time series data (Cryer, 1985). We use the Box-Jenkins procedure to formulate models to describe our data.

4.3.1. Box-Jenkins Methodology

The Box-Jenkins methodology is used to identify a time series model where y_t represents the observations made at time *t* (Cryer, 1985). The diagnostics are then checked to find the adequacy of the identified model and then an eventual forecast of a time series (Bowerman, O'Connell & Koehler, 2004).

4.3.2. Stationarity

The Box-Jenkins method uses a stationary time series to find an appropriate model and later to forecast a future time series (Bowerman, O'Connell & Koehler, 2004). One determines if the time series is stationary by plotting $y_1,\ y_2 \dots, y_n$ against time. If there is constant variation of the n values fluctuate around a constant mean, μ, then one can assume that the time series is stationary. If the time series appears to be non-stationary, one may need to perform a transformation of the original time series, such as calculating the first differences, second differences, or a log-transformation to name a few. Here we will only display the formula for the transformation of the first differences.

First differences of the time series values $y_1, y_2, \dots, y_n$ are

(4.12) $$z_t = y_t - y_{t-1}$$

where $t = 2, \dots, n$.

One more thing that needs explanation is b, which is a value that is decided by the transformation that is performed and in this transformation is equal to 2 because the value of the first observation is not actual after the transformation.

4.3.3. General Guidelines for Box-Jenkins

The Box-Jenkins model can be identified by looking at the sample autocorrelation function (SAC) and the sample partial autocorrelation function (SPAC). The sample autocorrelation function is a listing or a graph of the sample autocorrelations at lags $k = 1, 2, \dots$
The sample autocorrelation, r_k, at lag k is

(4.13) $$r_k = \frac{\sum_{t=b}^{n-k}(z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=b}^{n}(z_t - \bar{z})^2}$$

where

(4.14) $$\bar{z} = \frac{\sum_{t=b}^{n} z_t}{(n-b+1)}$$

For nonseasonal data, a useful statistic is the

(4.15) $$t_{r_k} = \frac{r_k}{s_{r_k}}$$

The t_{r_k} statistic is a measure, which dictates whether there is a spike in the SAC (Bowerman, O'Connell & Koehler, 2004). If the t_{r_k} is larger than 2, then a spike will appear in the diagram and this means that there is autocorrelation at lag k. When analyzing the diagram a good rule of thumb is that if the SAC of the time series cuts off fairly quickly or dies down, then the time series should be considered stationary. The opposite is true if the times series dies down slowly.

4.3.4. Identifying a Box-Jenkins Model

There are two more common types of Box-Jenkins models (Bowerman, O'Connell & Koehler, 2004). The first is the nonseasonal autoregressive model of order 1, AR(1):

(4.16) $$z_t = \phi_1 z_{t-1} + a_t$$

where a_t is a random shock corresponding to time period *t*, that describes the effect of all other factors, and ϕ_1 is an unknown parameter that is estimated from the sample.

The second is the nonseasonal moving average model of order 1, MA(1):

(4.17) $$z_t = a_t - \theta_1 a_{t-1}$$

where a_t is a random shock and θ_1 is an unknown parameter that must be estimated from the sample.

4.3.5. General Seasonal Model

The backshift operator is depicted by the symbol *B* and needs to be discussed if we are going to explain the general seasonal model (Bowerman, O'Connell & Koehler, 2004). $\nabla = 1 - B$ is the non-seasonal operator and is the seasonal operator $\nabla_L = 1 - B^L$ where *L* either represents quarterly of monthly data.

The general stationarity transformation is:

(4.18) $$z_t = \nabla_L^D \nabla^d y_t^* = (1 - B^L)^D (1 - B)^d y_t^*$$

Where *d* is the degree of nonseasonal differencing and *D* is the degree of seasonal differencing used.

4.3.6. Constant in Box-Jenkins Model

A constant, δ, is needed in the model if

(4.19) $$\left|\frac{\bar{z}}{SE_{\bar{z}}}\right| > 2$$

When the absolute value in the formula above is greater than two, it means that $\bar{z}$ is statistically different than zero and a constant should be included in the model.

4.3.7. Picking a Model

A model is decided from a certain number of criteria (Cryer, 1985). The probability values in the Ljung-Box test need to be over 20%. All of the estimated parameters need to be significant, usually at or above the 5%-level. The estimated parameters need to follow certain stationarity and invertiblity conditions, usually the absolute values need to be under the value of 1 for proper forecasting (Bowerman, O'Connell & Koehler, 2004). The autoregressive has only stationarity conditions and the moving average has only invertibility conditions. The residuals, which can be described as the point estimate of the error term are also crucial when determining the best model. There should be no spikes in RSAC and RSPAC, which are the autocorrelation functions of the residuals. The *MS* value, which is the mean squared error, should be as low as possible.

4.3.8. Forecasting and measures of accuracy

One-step forecasting is used to verify model accuracy. Each forecast is calculated individually, starting with the last one and then it is removed and the next forecast is calculated in sequential order until they are all finished (Bowerman, O'Connell & Koehler, 2004). The value at the actual time of the observation is and is the forecasted value that is calculated from the previous values of right up to and including the immediate preceding value. Therefore forecasts one step ahead and thus the name one-step forecasting.

The forecast error at time *t* is defined as:

(4.20) $$e_t = y_t - \hat{y}_t$$

The mean error is used mainly to detect under or over forecasting. If you have both positive and negative errors, the errors may cancel each other out and should be as close to zero as possible.

Mean error:

(4.21) $$\bar{e} = \frac{\sum_{t=j}^{T} e_t}{n}$$

The squared errors do not allow for positive and negative forecast errors to cancel one another out and is used for the mean squared error (*MSE*) calculation (Bowerman, O'Connell & Koehler, 2004):

(4.22) $$MSE = \frac{1}{n}\sum_{t=j}^{T} e_t^2$$

MSE measures the average of the squared errors and punishes large errors much more than small errors. A perfect *MSE* value would be zero. In model comparison, the *MSE* statistic can be a helpful tool as the model with the smallest *MSE* is generally interpreted as the best explaining the variability in the observations.

5. Results

5.1. Descriptive Statistics

Variable	N	Mean	StDev	Sum	Minimum	Maximum
Days	3653	3.45	1.89	12 618	0	13

Table 5.1: Descriptive statistics for days

Table 5.1 Calculations refer to a ten year period between 1995-2004. N is equal to 3653, which are the number of days in this period. The mean result is 3.45 births per day. The standard deviation is 1.89. The sum is the total number of births, which is 12618. The minimum and maximum give the range of births per day, from 0 to 13.

Variable	N	Mean	StDev	Sum	Minimum	Maximum
Years	10	1 261.8	76.9	12 618	1 169	1 397

Table 5.2: Descriptive statistics for years

Table 5.2 The mean is 1261.8 births per year. The standard deviation was calculated to be 76.9. The minimum is 1169 and maximum is 1397 for births per year.

Variable	N	Mean	StDev	Sum	Minimum	Maximum
January	10	98.4	11.32	984	82	117
February	10	100.1	8.70	1 001	87	114
March	10	117.0	10.42	1 170	107	142
April	10	113.4	18.06	1 134	93	143
May	10	115.8	12.67	1 158	103	138
June	10	106.3	14.74	1 063	87	135
July	10	112.8	13.02	1 128	84	128
August	10	107.1	12.15	1 071	85	120
September	10	106.6	12.96	1 066	82	132
October	10	101.4	9.64	1 014	86	115
November	10	89.9	10.24	899	75	113
December	10	93.0	5.08	930	87	102

Table 5.3: Descriptive statistics for month

Table 5.3 divides up the births per month for the ten years of data. The mean is the average number of births per month for the ten-year period. The mean variates from 89.9 in November to 117.0 in March. The standard deviation variates from 5.08 in December to 18.06 in April.

Variable	N	Mean	StDev	Sum	Minimum	Maximum
Sunday	522	3.34	1.83	1 745	0	11
Monday	522	3.57	1.89	1 862	0	13
Tuesday	522	3.52	1.85	1 838	0	10
Wednesday	522	3.53	1.91	1 842	0	10
Thursday	522	3.45	1.93	1 799	0	13
Friday	522	3.42	1.96	1 785	0	11
Saturday	521	3.35	1.82	1 747	0	11

Table 5.4: Descriptive statistics for the days of week

Table 5.4 gives the descriptive statistics for the days of week. The mean values and standard deviations for each day are quite close to one another in value. Even so, there is tendency for more births during the week days and less births during the weekend.

Variable	N	Mean*	StDev♦
Phase 1	124	3.45	0.76
Phase 2	124	3.47	0.81
Phase 3	124	3.43	0.81
Phase 4	124	3.45	0.79

Table 5.5: Descriptive statistics for 4 Moon phases

Table 5.5 depicts the Moon's four phases: New Moon (Phase 1), First Quarter (Phase 2), Full Moon (Phase 3), Last Quarter (Phase 4). The mean values and the standard deviations are fairly similar to one another but the phase leading up to the Full Moon has the largest mean.

Variable	N	Mean*	StDev♦
Phase 1	124	3.49	1.74
Phase 2	124	3.46	0.80
Phase 3	124	3.55	1.85
Phase 4	124	3.45	0.85
Phase 5	124	3.46	1.95
Phase 6	124	3.43	0.83
Phase 7	123	3.51	1.89
Phase 8	124	3.41	0.91

Table 5.6: Descriptive statistics for 8 Moon phases

Table 5.6 depicts the Moon's eight phases: New Moon, Waxing Crescent, First Quarter, Waxing Gibbous, Full Moon, Waning Gibbous, Last Quarter and Waning Crescent. The mean is the average number of births per phase. The mean values are nearly the same, but the standard deviations variate accordingly. The phases that are around twice as long all have similar standard deviations and those phases that are shorter also have similar standard deviations.

* Mean of births per day

♦ Standard deviation of births per day

5.2. Analysis of Variance

In Table 5.7 we performed certain tests on these categories: births per day of the week, births per month, births per four phases, and births per eight phases. A plus sign means that a certain category passed either the test for equal variances or the normality test and a minus sign means the opposite.

	Tests for Equal Variances	Normality test
Day of the week	+	-
Month	+	+
4 phases	+	+
8 phases	-	-

Table 5.7: Summary table of results for normality and equal variances tests

The categories of day of the week, births per month and 4 phases passed the equal variances tests. Births per month, and births per 4 phases passed the normality test. Based on this information, we continued to perform either a one-way ANOVA or Kruskal-Wallis (See Appendix 1). The 8 phases do not have equal variances and we performed several transformations, such as square root, natural log, arcsinus, etc. and tested to see if the transformations worked without any positive results.

Accordingly, we performed an F-test on births per month and 4 phases. For births per month, the means were not equal and the null hypothesis was rejected because the probability was less than five percent. For 4 phases, the means were equal and the null hypothesis could not be rejected.

The Kruskal-Wallis test was performed on births per day of the week (See Appendix 1). The null hypothesis could not be rejected.

The only differences were found in births per month and we thus continue with this category and the Box-Jenkins methodology.

5.3. Time Series

Our investigation in finding a model began with determining a stationary time series for the category of months with the last year left out for forecasting. The time series plot depicts all the births in every month for nine years of data (See Appendix 2, Plot A2.1). The original time series plot for months appears to be fairly stationary, but a possible trend and/or seasonal component needed to be investigated further, therefore, we compared it with the plot of first differences, second differences, first seasonal differences, as well as the combined first regular and seasonal differences (See Plot A2.2 in Appendix 2). We also compared the standard deviations of several differencing transformations, but the series with the lowest standard deviation, which was the original time series could not be described by a model because of the seasonal component (See Appendix 2, Table A2.1). The first differences had the second lowest standard deviation, but could not be described by a model. Since the plot for the combined first regular and seasonal differences appeared to be stationary, we tested its model capabilities and found two models to describe it.

5.3.1. Model Comparison

The first model is the MA(1)×SMA(1) for the first regular and seasonal differences (See Appendix 3). All the parameters are significant with high p-values in the Ljung-Box test. The MS value is 142.8. The ACF and PACF for the residuals look good as well. There are no significant correlations between parameters.

The second model is the MA(1)×SAR(2) for the first regular and seasonal differences (See Appendix 4). All the parameters are significant here as well. The probability values are also high in the Ljung-Box test. The MS value is 131.9. The ACF and PACF for the residuals appear to be fine as well as the other plots. There is also no high correlation between parameters.

When we compare the models we note that both models have significant parameters and high probability values in the Ljung-Box test. Both models also show good distribution of residual plots in the normality plot, residuals vs. fits, and residuals vs. order.

The MA(1)×SMA(1) model was the simpler of the two models. The only problem with this model is that is had a higher MS value with 142.8 compared to other model, which was 131.9. Neither the MA(1)×SMA(1) or the MA(1)×SAR(2) model had significant correlation between parameters.

5.3.2. Forecasting

The MA(1)×SMA(1) proves in the end to be a better model because it has a lower MSE value than the MA(1)×SAR(2) model and the mean error value, $\bar{e}$, is also closer to zero (see Table 5.8). Both models have a negative $\bar{e}$-value, which conveys underforecasting.

Model	$\bar{e}$	MSE
MA(1)×SMA(1)	-3.249	174.54
MA(1)×SAR(2)	-5.114	292.23

Table 5.8: Comparison of forecasts

The tenth year, which was used to produce forecasts for each individual model, had a lower number of births in April, which is shown below in Diagram 5.1 so the month of April in all the forecasts is outside the confidence interval (See Appendix 3 and 4).

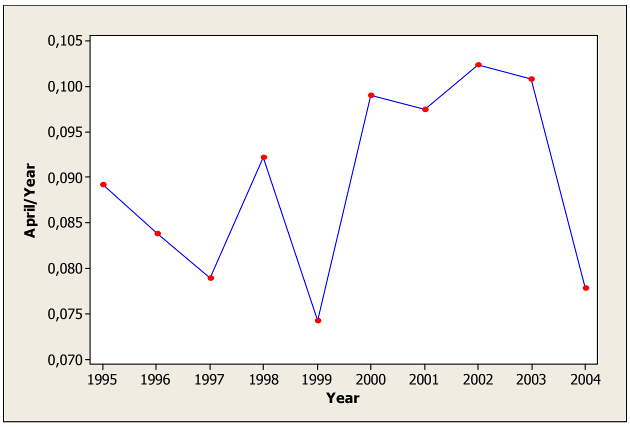


Diagram 5.1: Number of births in April, 1995-2004

6. Conclusions

Our study found a pattern in monthly birth trends for Lund with the highest number of births in March, April, May, and the lowest number in November, which was significant. This pattern may coincide with the notorious number of people that take vacation and conceive during the summer months. This result was consistent with a French study that was done with six million births.

Our study did not find any connection between the four phases of the Moon or the eight phases of the Moon, just like earlier studies at the University of Arizona and one from the Mountain Area Health Education Center in North Carolina. An explanation could be that the phases of the Moon are a man-made concept. A circle, representing the moon's path around the earth was divided into four parts and each part called a phase. There may be a different lunar pattern that coincides with when women deliver. We would suggest that a study be done on the average synodic month, which is the equivalent to one Moon day or 29.53 earth days. An Italian study performed such a study and found weakly significant results. It might also be of interest to study weather phenomena.

7. Sources

7.1. Articles

Arliss, J.M., Kaplan, E.N. & Galvin, S.L. (2005). The effect of the lunar cycle on frequency of births and birth complications. *American Journal of Obstetrics and Gynecology.* 192, pp. 1462-4.

Guillon, P., Guillon, D., Lansac, J. & Soultoul, J.H. (1986). Naissances, fertilité, rythmes et cycle lunaire. Etude statistique sur 5 927 978 naissances. *Journal de gyne'cologie, obste'trique et biologie de la reproduction.* 15, pp. 265-271.

Mikulecky, M. & Lisboa, H.R.K. (2002) Daily birth numbers in Passo Fundo, South Brazil, 1997-1999: trends and periodicities. *Brazilian Journal of Medical and Biological Research.* Vol. 35, No. 8, pp. 985-90.

Morton-Pradhan, S., Bay, R.C. & Coonrod, D.V. (2005). Birth rate and its correlation with the lunar cycle and specific atmospheric conditions. *American Journal of Obstetrics and Gynecology.* 192, p.1970-3.

Muñoz-Delgado, J., Santillán-Doherty, A.-M., Mondragón-Ceballos, R. & Erkert, H. (2000). Moon cycle effects on humans: myth or reality? *Salud Mental.* Vol. 23, No. 006, pp. 33-38.

7.2. Books

Bowerman, B-L., O'Connell, R-T. & Koehler, A-B. (2004). *Forecasting, Time series, and Regression: An Applied Approach.* Belmont: Thomson Brooks/Cole.

Cadogan, P. (1981). *The Moon - our sister planet.* Cambridge: Cambridge University Press.

Cryer, J-D. (1985). *Time Series Analysis with Minitab.* Boston, Mass.: Duxbury Press

Hamberger, L. & Nilsson, L (2003). *Ett barn blir till*. Sweden: Albert Bonniers förlag.

Kleinbaum, D-G., Kupper, L-L. & Muller K-E. (1988). *Applied regression analysis and other multivariate methods.* Boston, Mass.: PWS-Kent Publ. Co.

Meeus, J. (1997). *Mathematical Astronomy Morsels.* Virginia: Willmann-Bell, Inc.

Montgomery, D-C. (2005). *Design and Analysis of Experiments.* Noboken, N.J.: Wiley.

Whipple, F. (1941). *Earth, Moon, and Planets.* Massachusetts: Harvard University Press.

7.3. Electronic and Internet sources

A Dictionary of Psychology. Social Sciences and the Law. *Psychology of lunacy.* (2001). [Online]. Available from: http://www.encyclopedia.com/doc/1O87-lunacy.html [Accessed: 26th June 2008].

Astronomical Applications Department U.S. Naval Observatory. *Phases of the Moon.* (2008). [Online]. Available from: http://aa.usno.navy.mil/data/docs/MoonPhase.php [Accessed: 9th June 2008].

Birmingham City University Business School. *Harvard Referencing.* (2006). [Online]. Available from: http://essential.tbs.bcu.ac.uk/harvard.html [Accessed: 1st July 2008].

Clarke, K. (2008). *Moon Phase Calculator.* [Online]. Available from: http://www.inconstantmoon.com/cal_phas.htm. [Accessed: 9th June 2008].

Cox Communications Inc. *It's a Girl!* (2008). [Online]. Available from: http://members.cox.net/jecoulter/frameset/pages_portfolio/18_baby_stork.html [Accessed: 14th September 2008].

Minitab (Version 15), [Computer Program]. (2008). Minitab Inc. Quality Plaza 1829 Pine Hall Rd State College PA 16801-3008 USA.

MoonPhases.Info. *Moon Phases Calendar 2008. Images of the eight (8) Moon phases.* (2008). [Online]. Available from: http://www.moonphases.info/moon_phases.html [Accessed: 12th September 2008].

NASA (National Advisory Committee on Aeronautics). *Moon Phases: 6000 Year Catalog.* (2007). [Online]. Available from: http://eclipse.gsfc.nasa.gov/phase/phasecat.html [Accessed: 9th June 2008].

QuickPhase Pro (Version 3.3.5), [Computer Program]. (2008). ©BlueMarmot.com/ BlueMarmot, Inc. MoonConnection.com.

7.4. Interviews

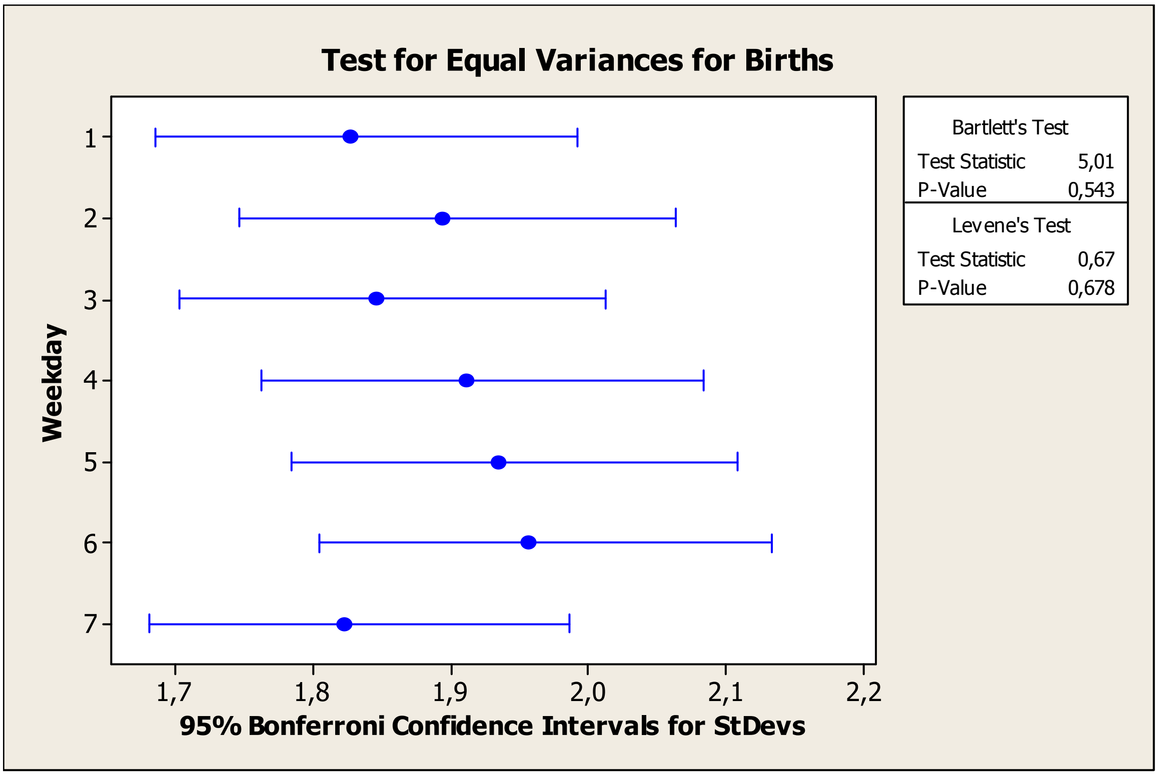
Rylance, R. (2008). Interview. [Interview with Karin Källén, 2 May 2008].

7.5. Personal communications

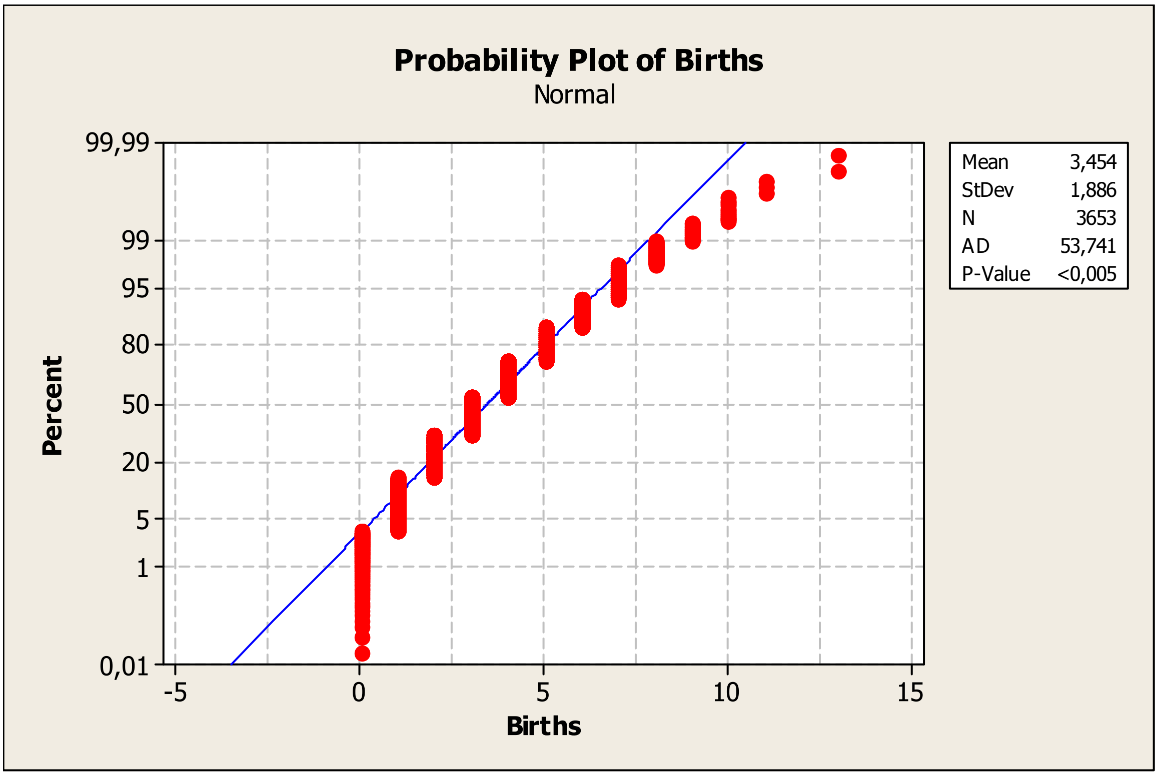
Rylance, R. & Tarassiouk, O. (2008). [Email from Daniel Malmberg. danielm@astro.lu.se, 9 June 2008].

Appendix 1: Analysis of Variance

DAYS OF THE WEEK



Plot A1.1: Test for equal variances for births per days of the week

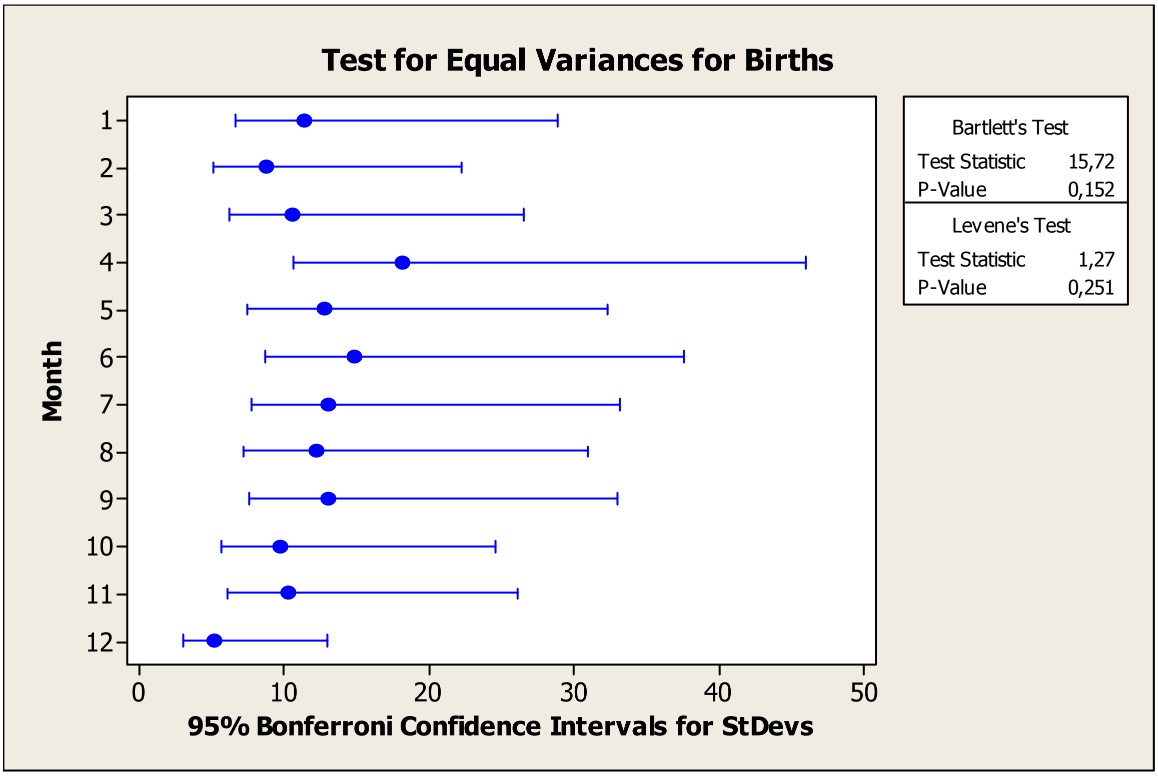


Plot A1.2: Anderson-Darling test for births per day of the week

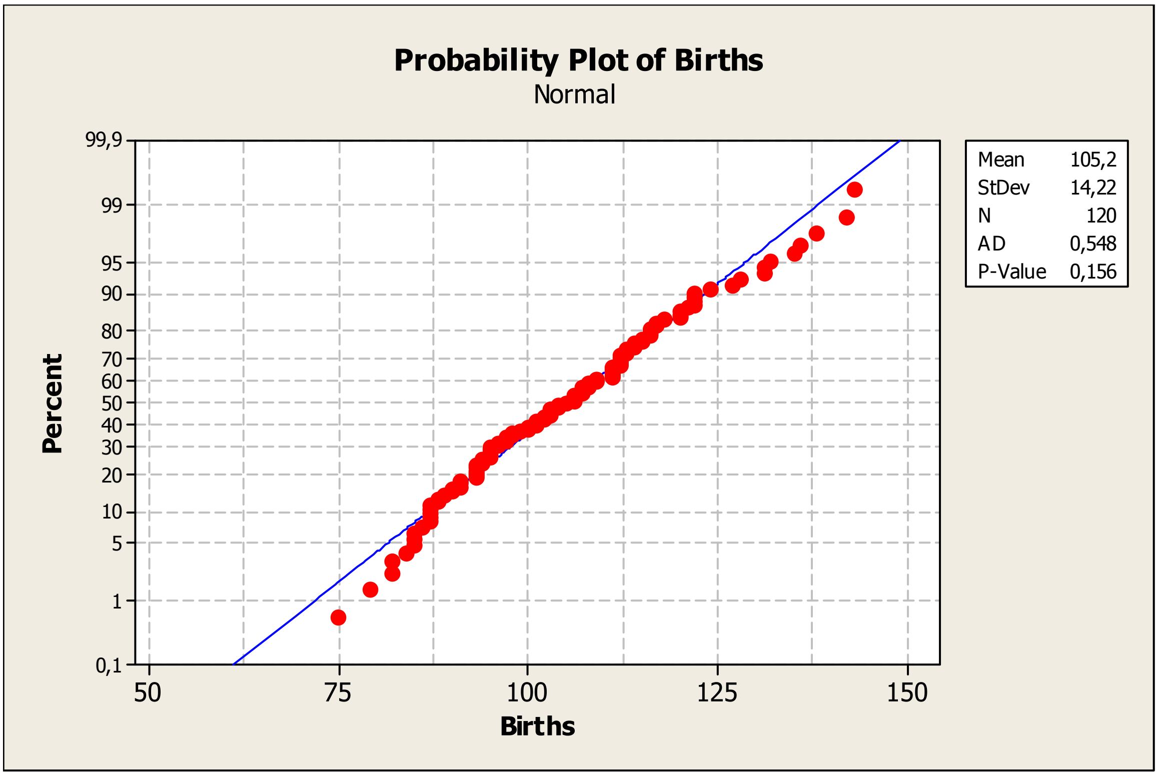
Weekday	N	Median	Ave Rank	Z
1	522	3.000	1770.8	-1.31
2	522	3.000	1895.3	1.60
3	522	3.000	1874.5	1.11
4	522	3.000	1873.0	1.08
5	522	3.000	1803.4	-0.55
6	522	3.000	1793.5	-0.78
7	521	3.000	1778.3	-1.14
Overall	3653		1827.0	
H = 7.62	DF = 6	P = 0.267		
H = 7.83	DF = 6	P = 0.251	(adjusted for ties)	

Table A1.1: Kruskal-Wallis test, births versus days of the week

MONTHS



Plot A1.3: Test for equal variances for births per month

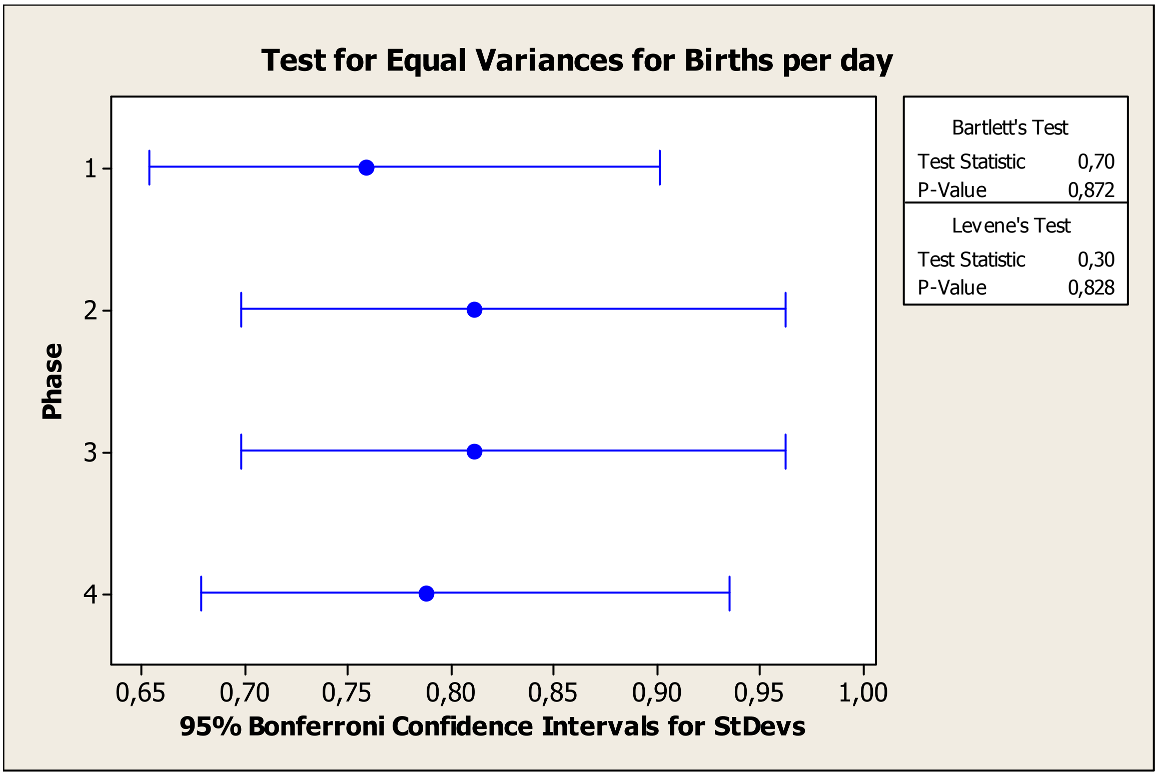


Plot A1.4: Anderson-Darling test for births per month

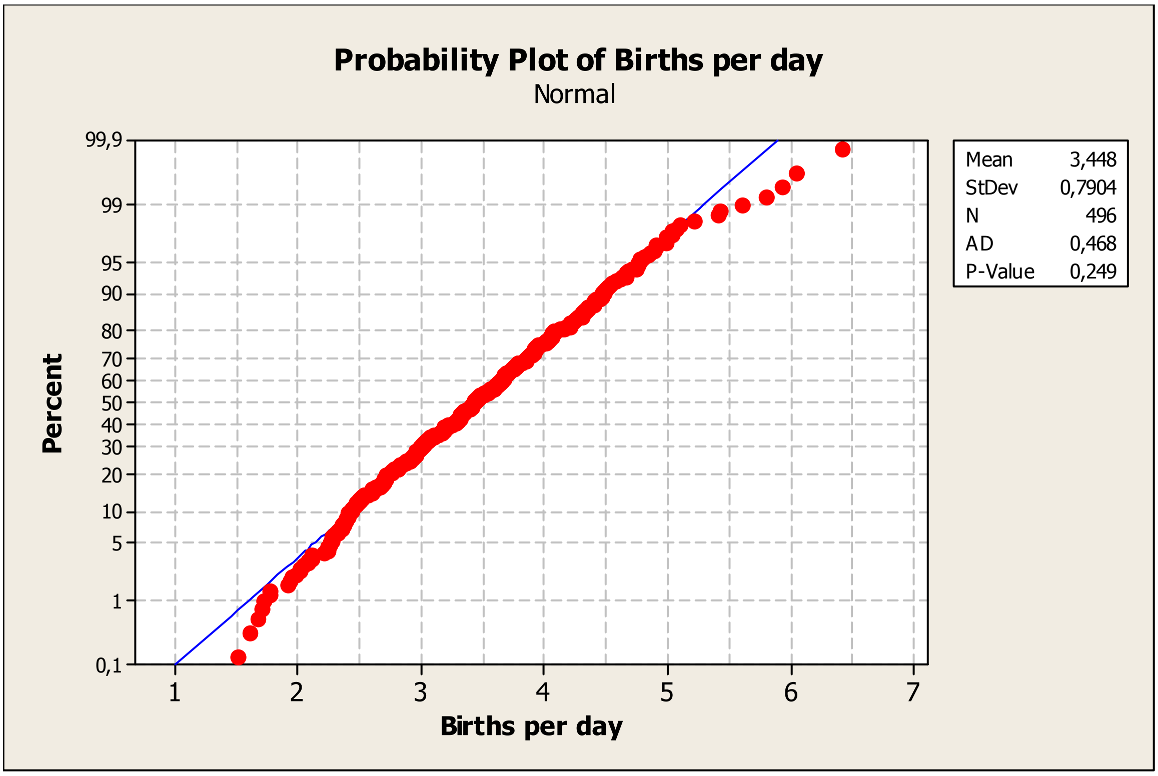
Source	DF	SS	MS	F	P
Month	11	8530	775	5.39	0.000
Error	108	15524	144		
Total	119	24053			
S = 11.99		R-Sq = 35.46%		R-Sq(adj) = 28.89%	

Table A1.2: One-way ANOVA, births versus months

4 PHASES



Plot A1.5: Test for equal variances for the 4 phases

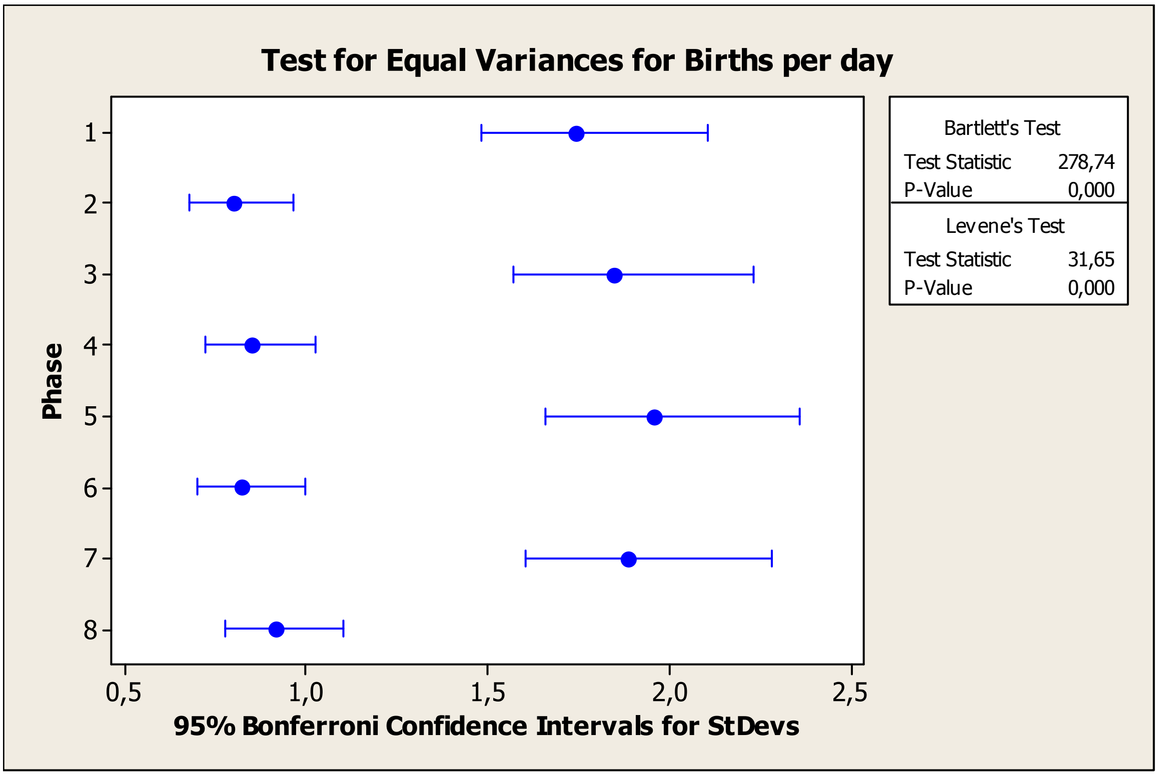


Plot A1.6: Anderson-Darling test for the 4 phases

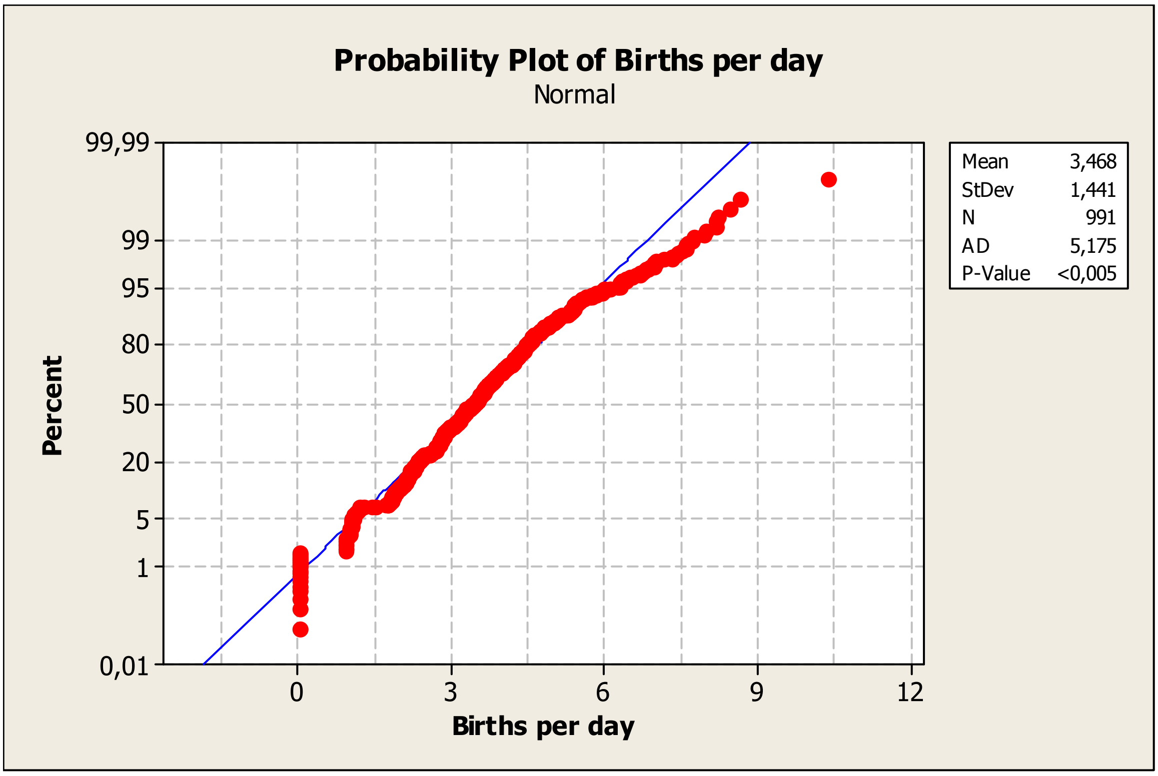
Source	DF	SS	MS	F	P
Phase	3	0.104	0.035	0.06	0.983
Error	492	309.112	0.628		
Total	495	309.216			
S = 0.7926		R-Sq = 0.03%		R-Sq(adj) = 0.00%	

Table A1.3: One-way ANOVA, births versus 4 phases

8 PHASES

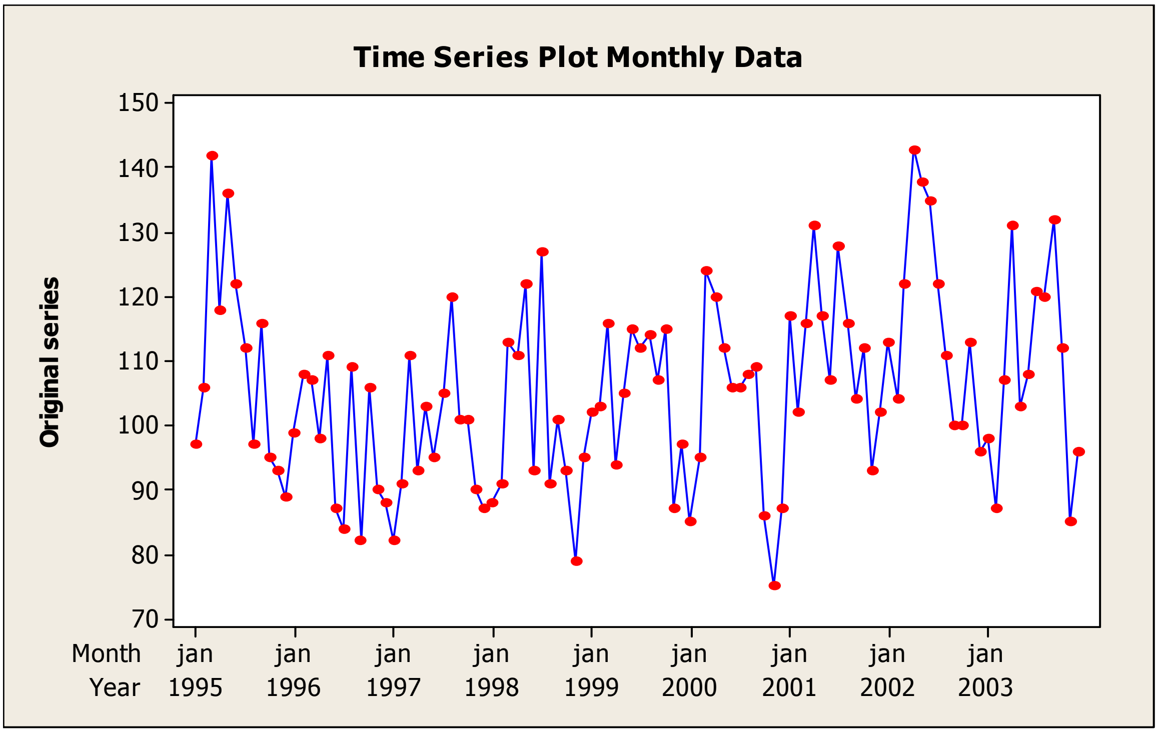


Plot A1.7: Test for equal variances for the 8 phases

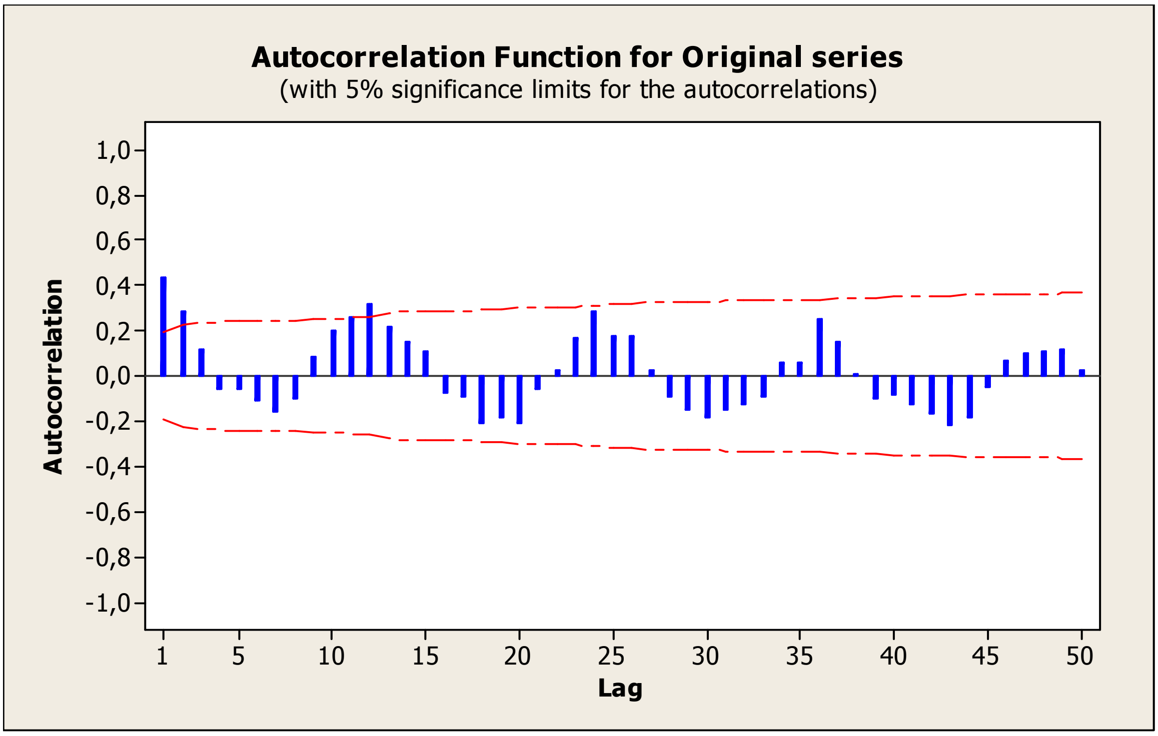


Plot A1.8: Anderson-Darling test for the 8 phases

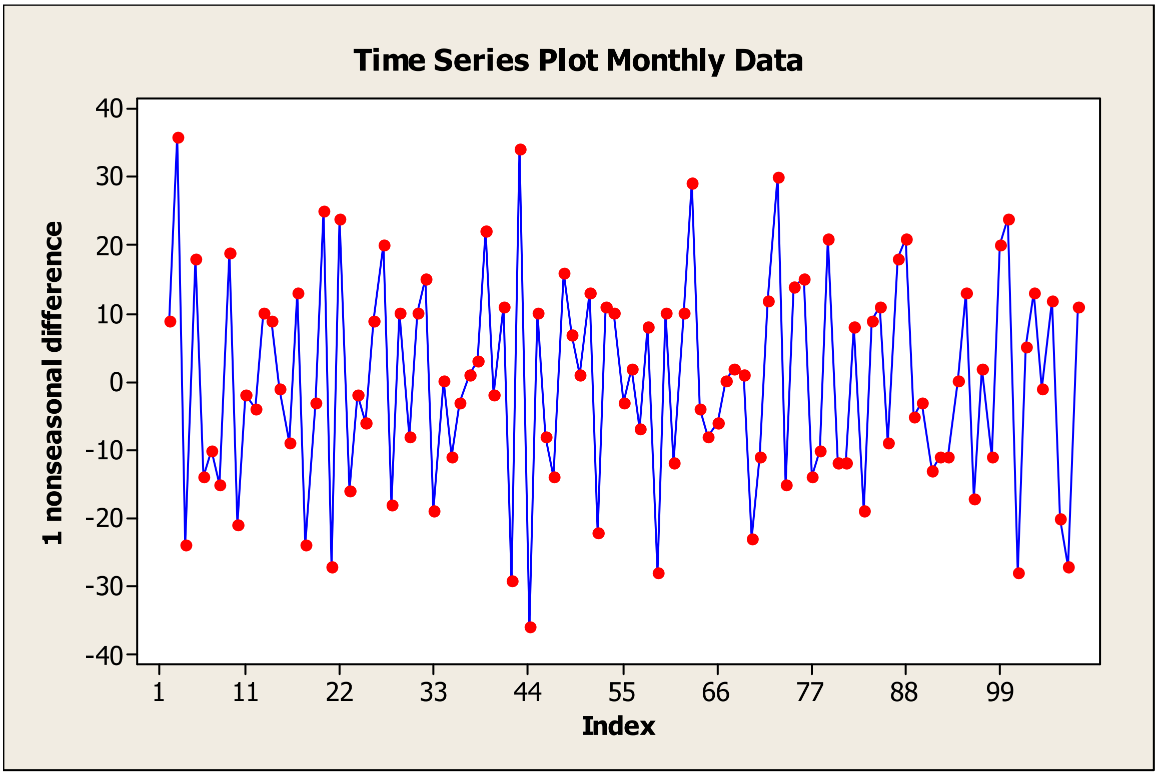
Appendix 2: Stationary and Nonstationary Time Series



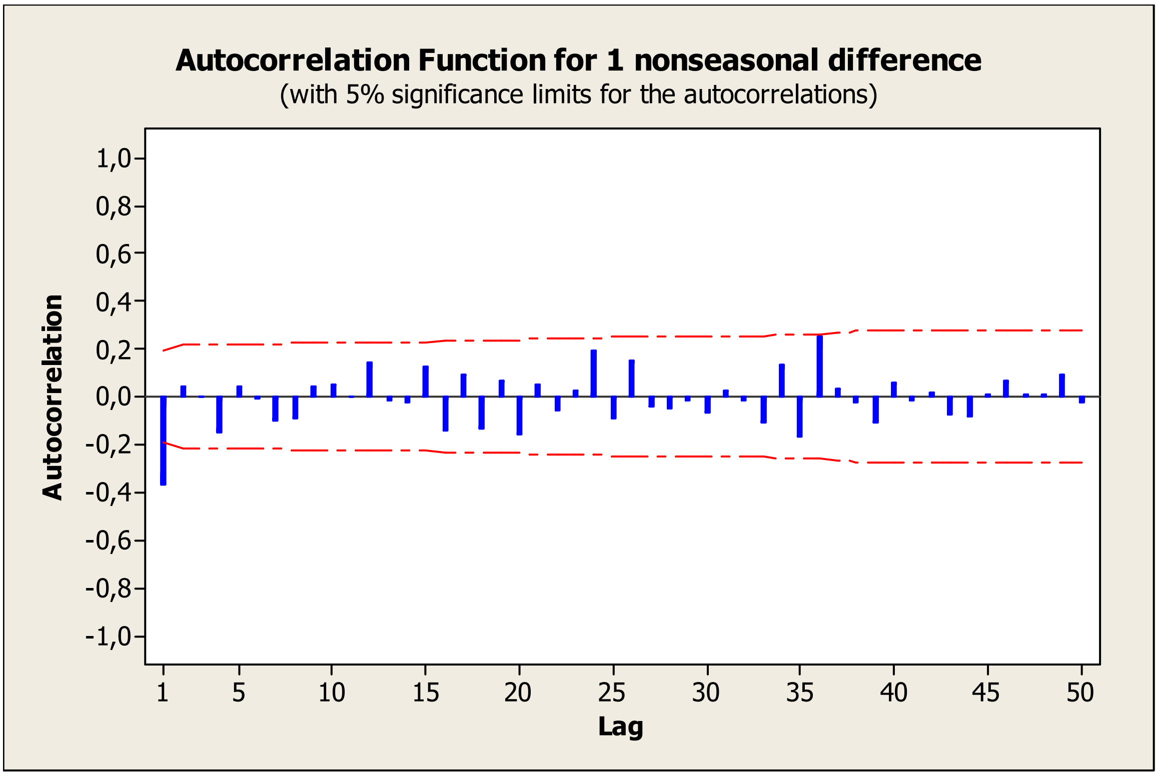
Plot A2.1: Time series plot, original data



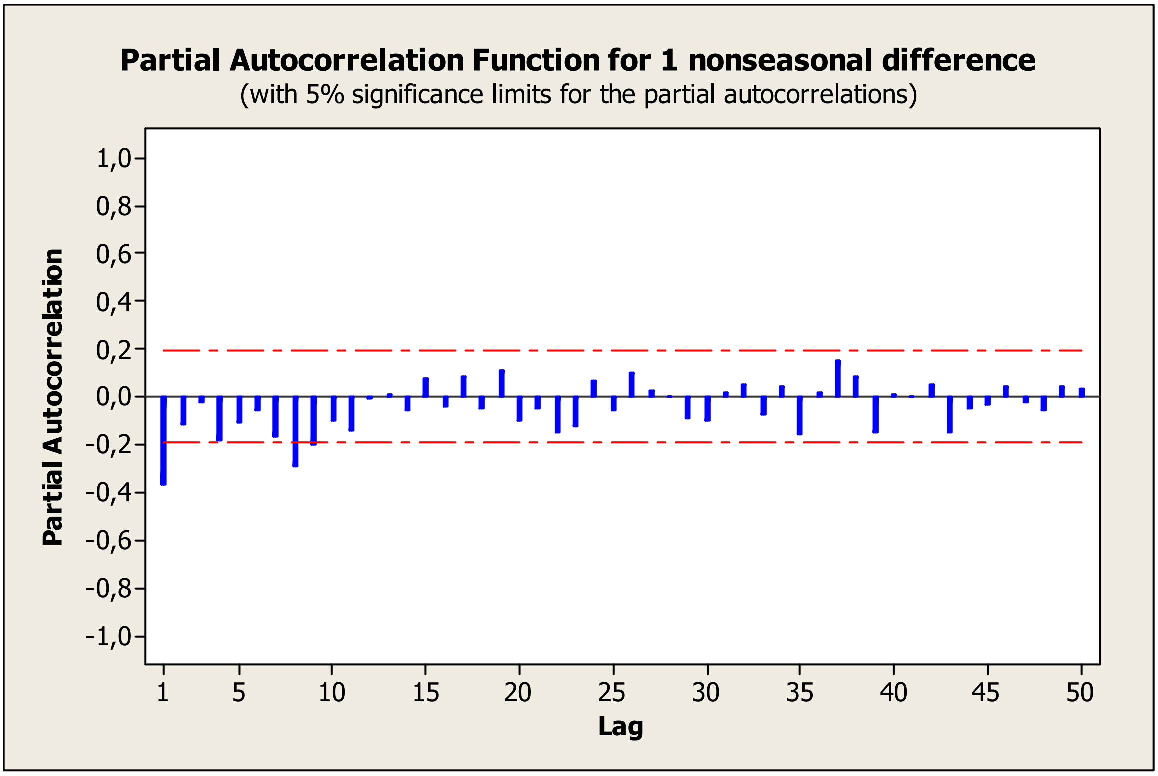
Plot A2.2: ACF, original data



Plot A2.3: Time series plot, 1st regular differences



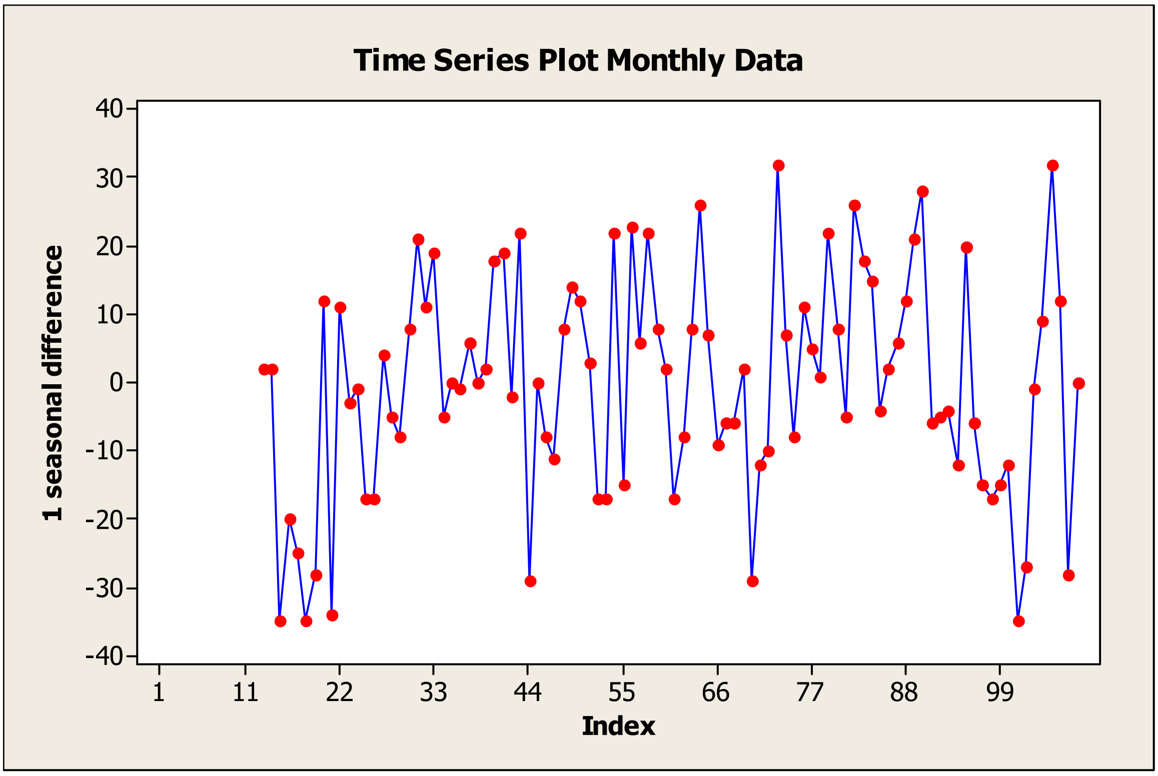
Plot A2.4: ACF, 1st regular differences



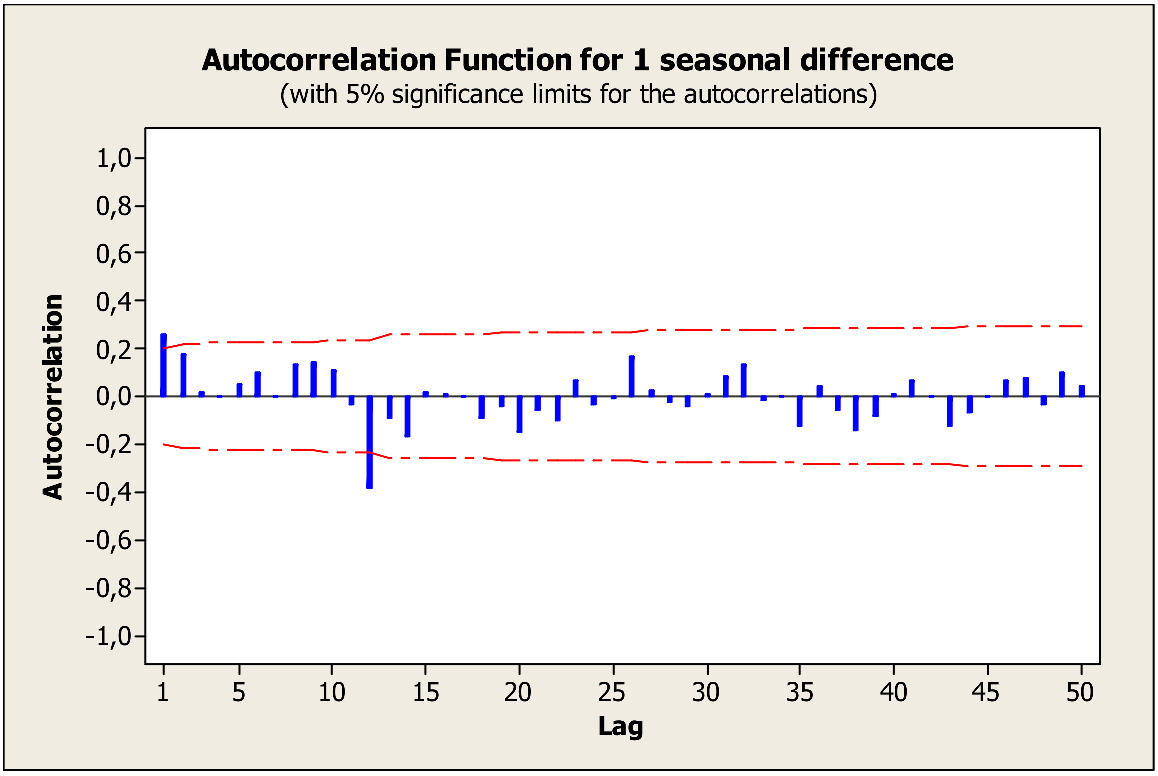
Plot A2.5: PACF, 1st regular differences

Variable	N	N*	Mean	SE Mean	StDev
Original series	108	0	105.53	1.40	14.58
1st nonseasonal differences	107	1	-0.01	1.50	15.52
2nd nonseasonal differences	106	2	0.02	2.50	25.76
1st seasonal differences	96	12	-0.24	1.65	16.18
1st nonseasonal & 1st seasonal	95	13	-0.02	2.03	19.75

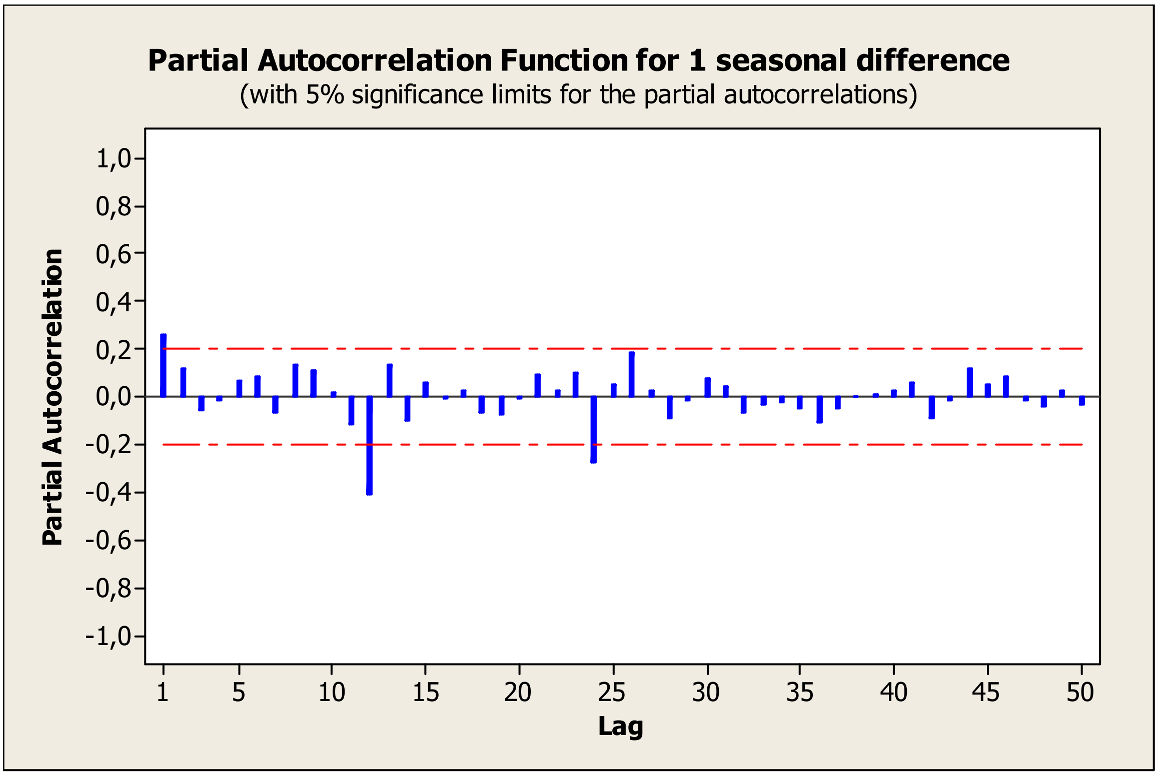
Table A2.1: Descriptive statistics for differencing transformations



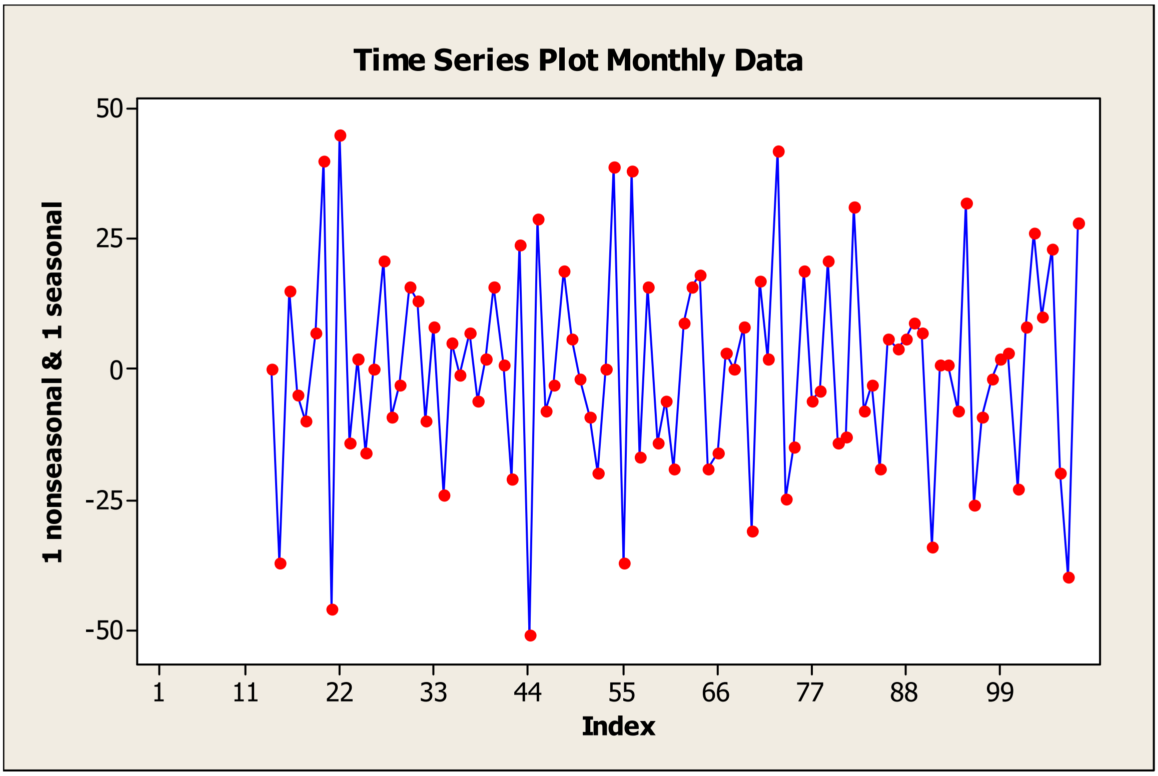
Plot A2.6: Time series plot, 1st seasonal differences



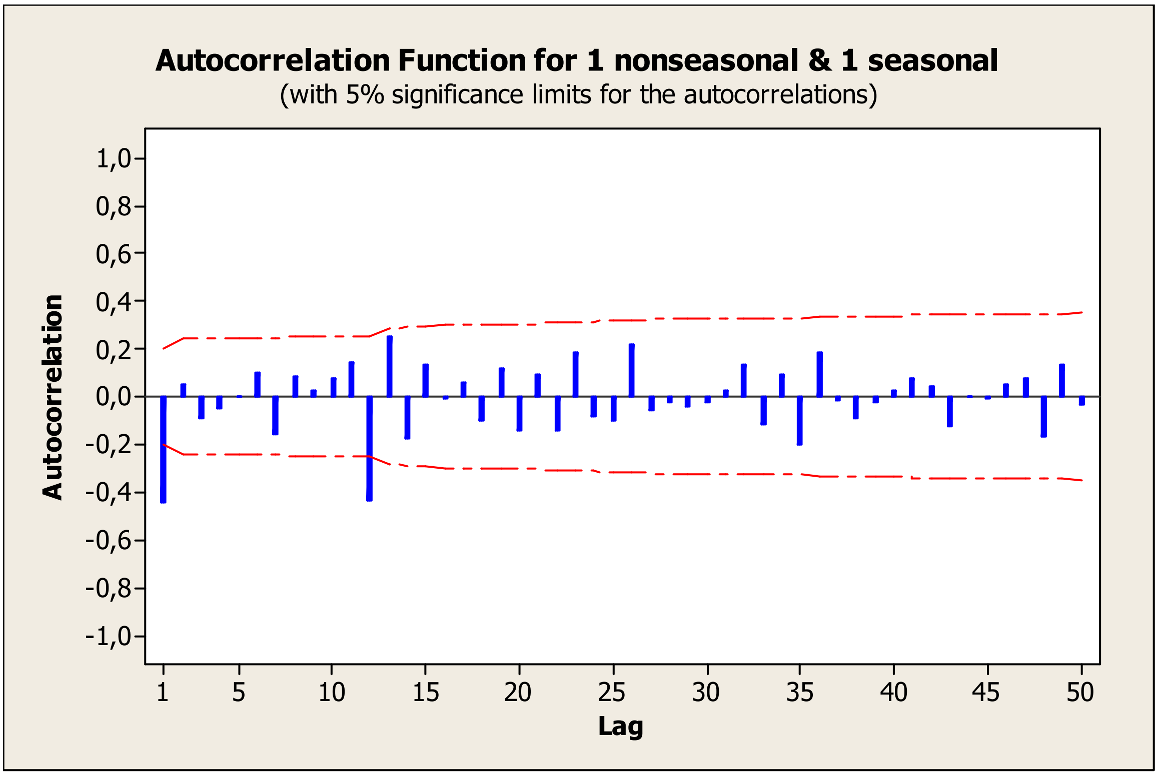
Plot A2.7: ACF, 1st seasonal differences



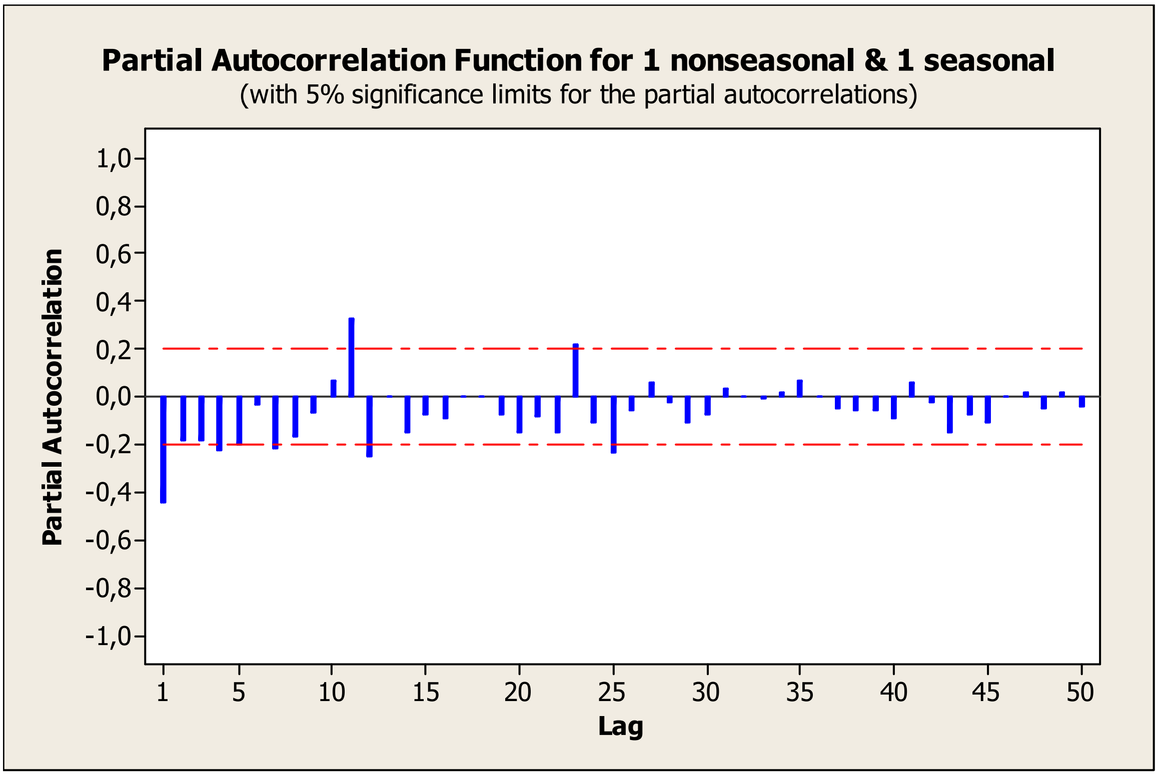
Plot A2.8: PACF, 1st seasonal differences



Plot A2.9: Time series plot, 1st regular and 1st seasonal differences



Plot A2.10: ACF, 1st regular and 1st seasonal differences



Plot A2.11: PACF, 1st regular and 1st seasonal differences

Appendix 3: MA(1)×SMA(1)

Type	Coef	SE Coef	T	P
MA 1	0.7974	0.0627	12.73	0.000
SMA 1	0.7964	0.1018	7.82	0.000

Differencing: 1 regular, 1 seasonal of order 12

Number of observations: Original series 108, after differencing 95

Residuals: SS = 13279.2 (backforecasts excluded)

MS = 142.8, DF =93

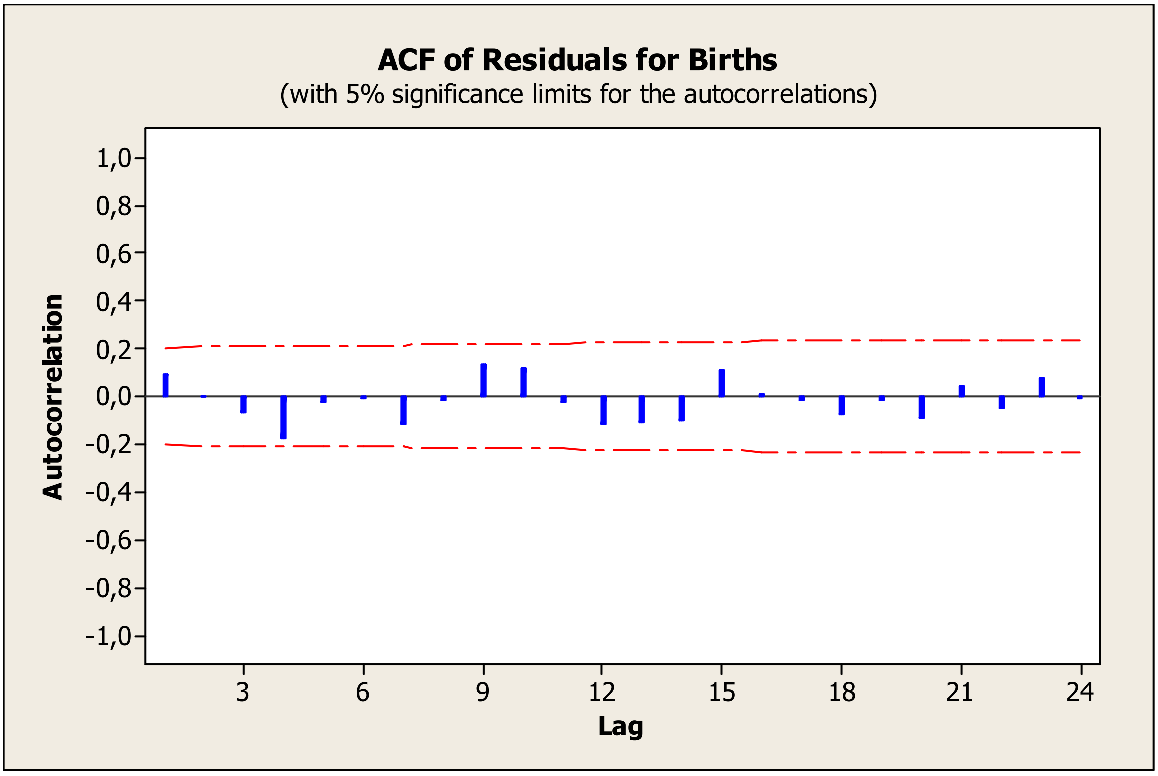
Table A3.1: Final estimates of parameters, MA(1) × SMA(1)

Lag	12	24	36	48
Chi-Square	11.0	17.8	26.2	39.2
DF	10	22	34	46
P-Value	0.357	0.717	0.828	0.750

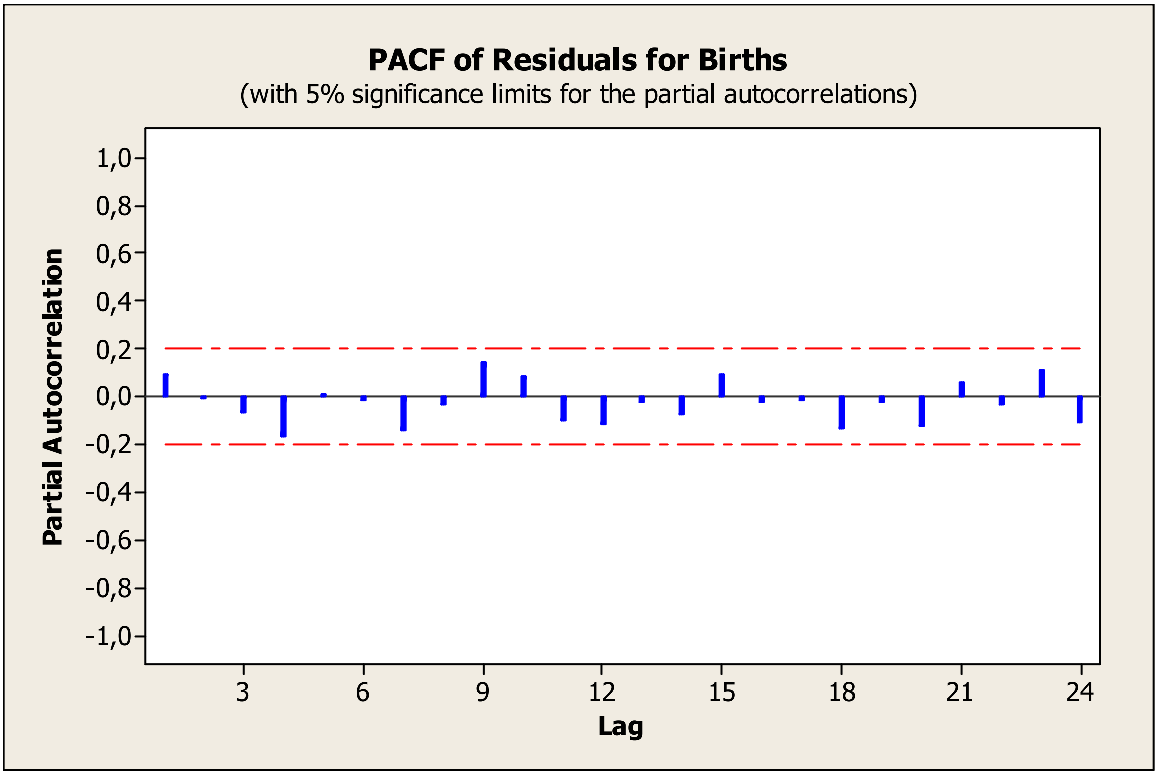
Table A3.2: Modified Box-Pierce (Ljung-Box) Chi-Square statistic, MA(1) × SMA(1)

	1
2	-0.100

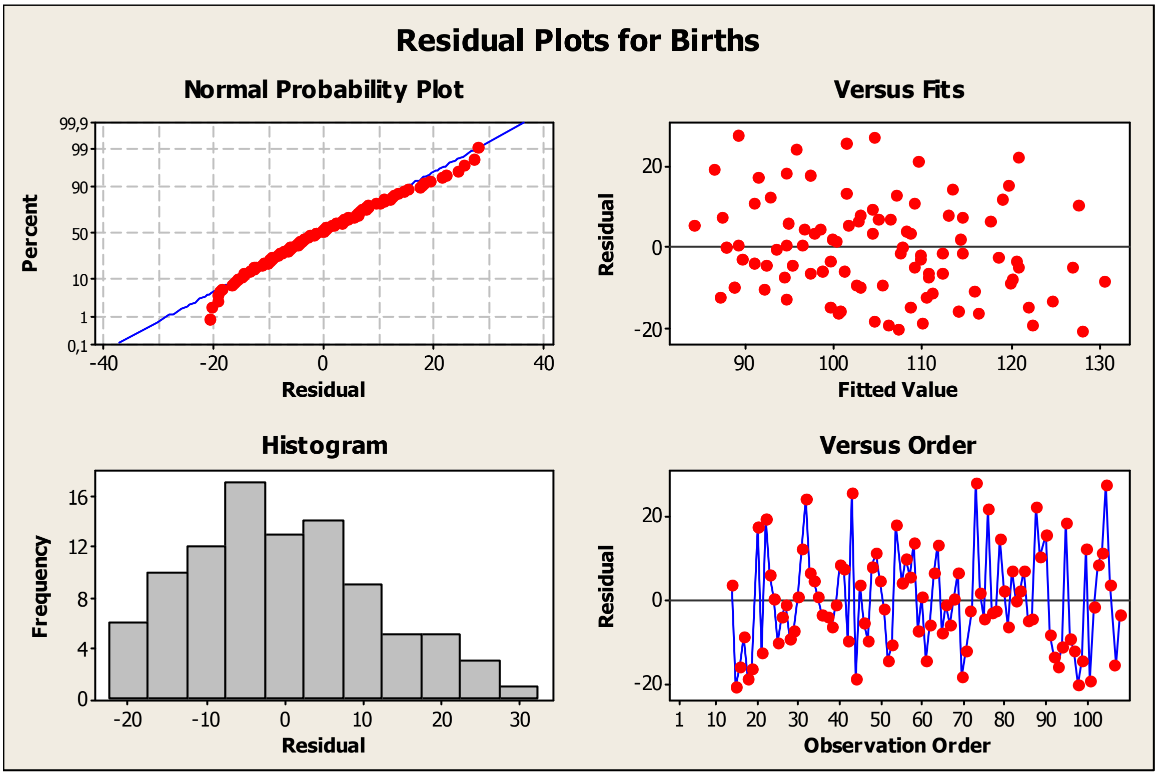
Table A3.3: Correlation matrix of the estimated parameters, MA(1) × SMA(1)



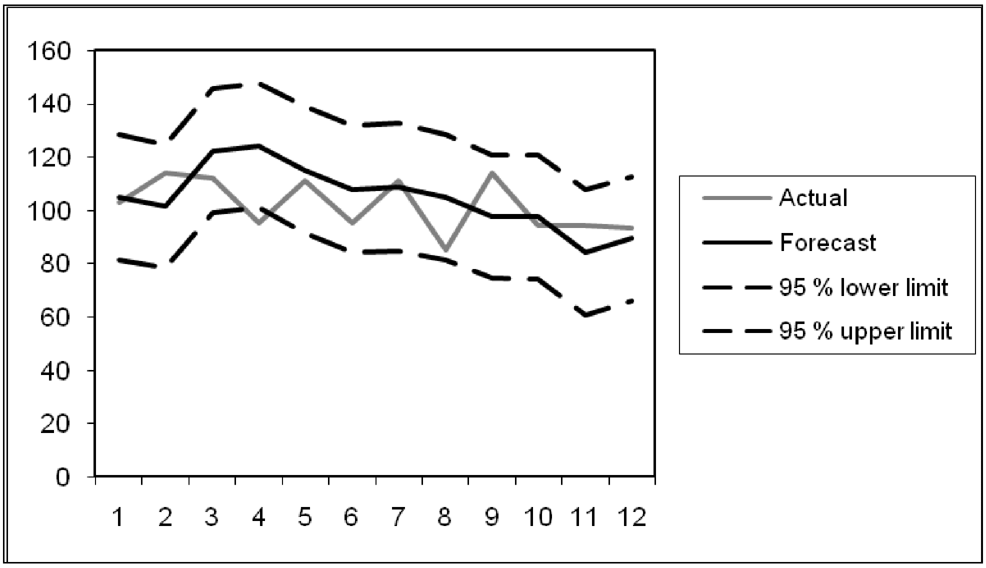
Plot A3.1: ACF of residuals, MA(1) × SMA(1)



Plot A3.2: PACF of residuals, MA(1) × SMA(1)



Plot A3.3: Residual plot, MA(1) × SMA(1)



Plot A3.4: One-step forecasts for year 2004, MA(1) × SMA(1)

Appendix 4: MA(1)×SAR(2)

Type	Coef	SE Coef	T	P
SAR 12	-0.8375	0.0945	-8.86	0.000
SAR 24	-0.5015	0.0994	-5.05	0.000
MA 1	0.7993	0.0606	13.20	0.000
Differencing: 1 regular, 1 seasonal of order 12				
Number of observations: Original series 108, after differencing 95				
Residuals: SS = 12132.5 (backforecasts excluded)				
MS = 131.9, DF = 92				

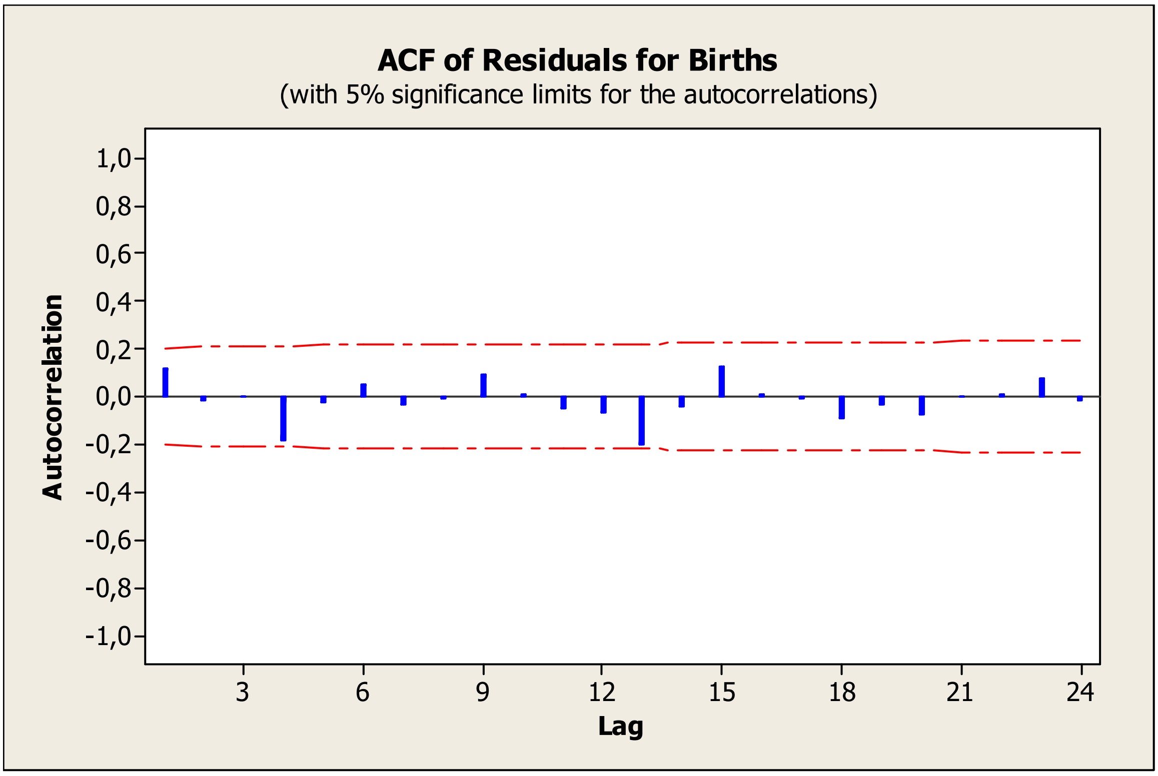
Table A4.1: Final estimates of parameters, MA(1) × SAR(2)

Lag	12	24	36	48
Chi-Square	7.1	16.0	28.6	42.7
DF	9	21	33	45
P-Value	0.631	0.772	0.688	0.571

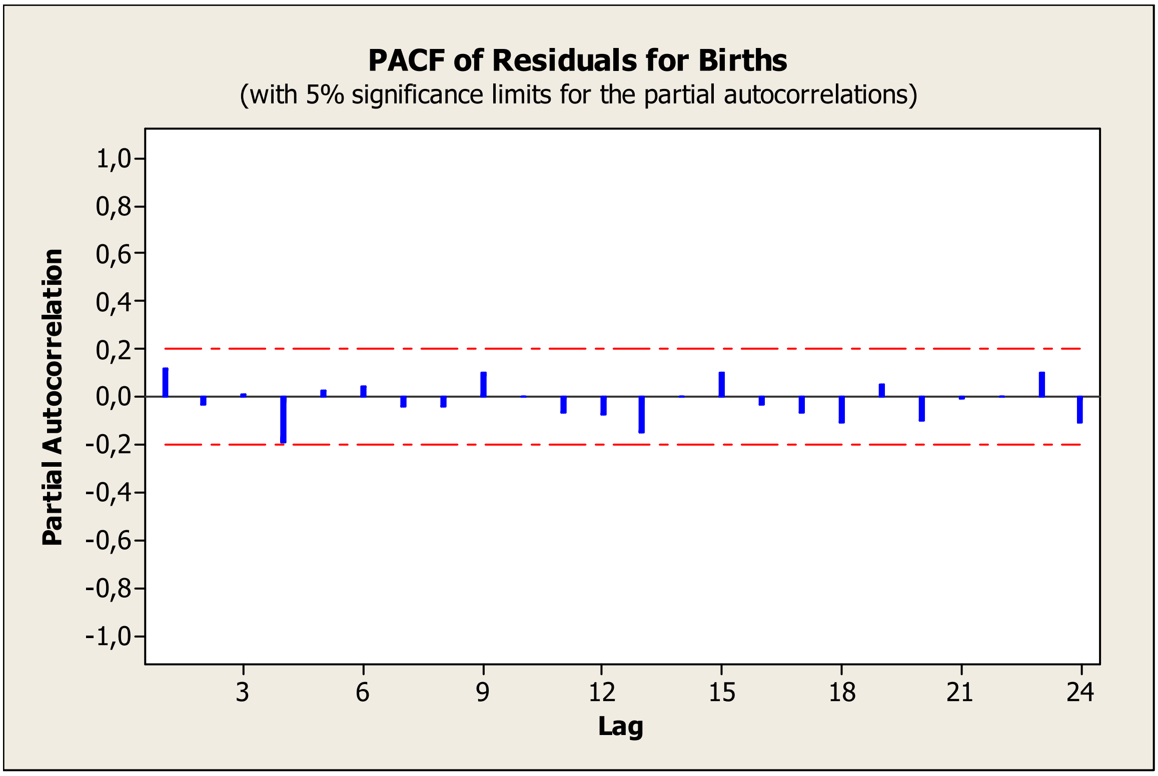
Table A4.2: Modified Box-Pierce (Ljung-Box) Chi-Square statistic, MA(1) × SAR(2)

	1	2
2	0.561	
3	0.129	0.082

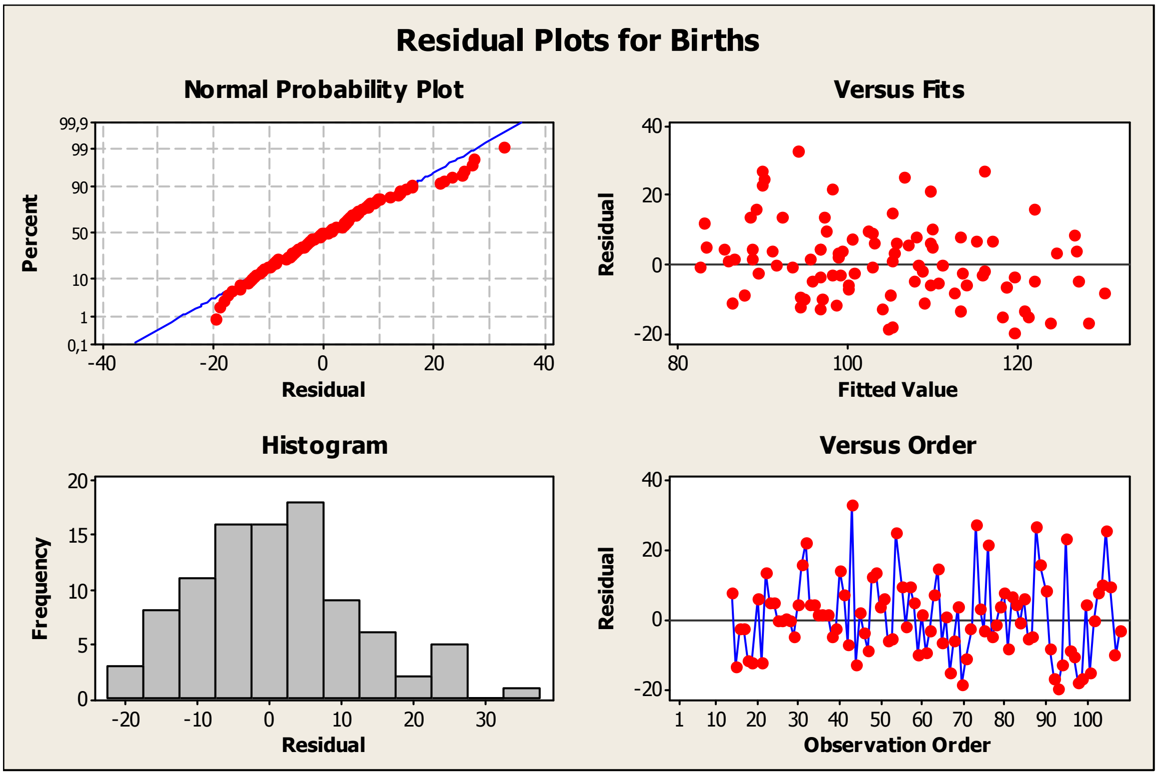
Table A4.3: Correlation matrix of the estimated parameters, MA(1) × SAR(2)



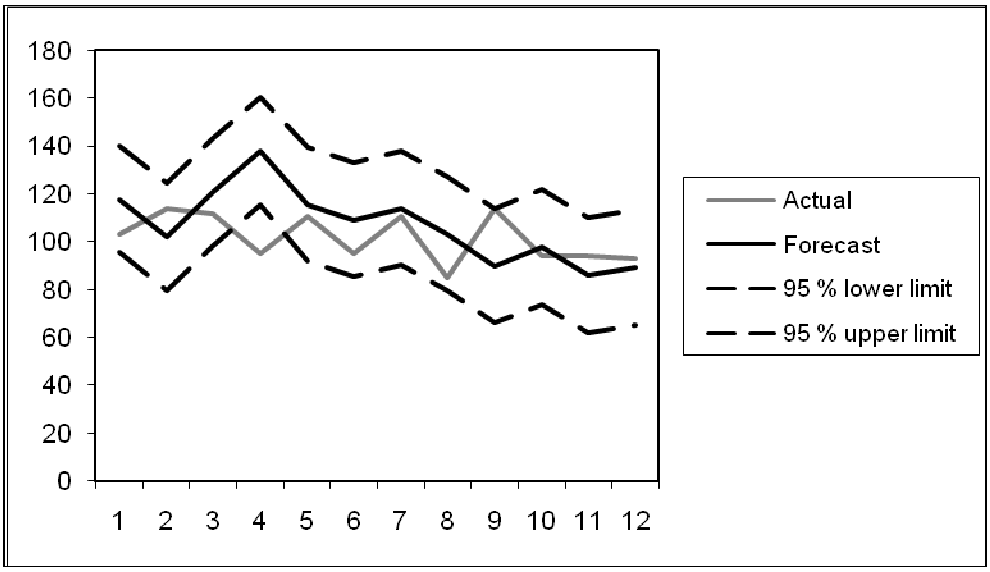
Plot A4.1: ACF of residuals, MA(1) × SAR(2)



Plot A4.2: PACF of residuals, MA(1) × SAR(2)



Plot A4.3: Residual plot, MA(1) × SAR(2)



Plot A4.4: One-step forecasts for year 2004, MA(1) × SAR(2)