National University of Ireland, Galway
Ollscoil na hÉireann, Gaillimh

Title: GARCH modelling applied to the Index of Industrial Production in Ireland

Candidate: Aoife Healy

Masters project

University Name: National University of Ireland, Galway

Supervisor: Dr. Damien Fay

Month and Year of submission: July 2007

Number of volumes: 1

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of masters is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:______________________ ID No:__________________________

Candidate

Date:______________________

TABLE OF CONTENTS **PAGE**

1.0 INTRODUCTION

The original idea for this thesis was to examine 3 times series as follows:

(i) Average New House Price in Ireland (1970 – 2006)

(ii) Average House Price in the United Kingdom (1970 – 2006)

(iii) Average House Price in the Netherlands (1970 – 2006)

A model for the conditional mean of Irish house prices was to be established and using General Autoregressive Heteroskedastistic models (herein after referred to as GARCH models) the volatility in Irish house prices over the past number years, together with forecasted volatility was to be examined. This was then to be compared with the volatility in house prices in the UK & the Netherlands in the time periods prior to property price crashes. As a result it was hoped that a pattern of volatility would emerge which will allow one to predict a crash in property prices.

However, this study was unable to continue due to the lack of available data on the property markets in question, especially the Irish Property Market. However, a large number of papers in the area of modelling property prices had been examined in preparation for this study and as such these have been outlined in section 2.1 of the literature review below.

In the absence of available data the techniques described above are applied to another time series namely the Index of Industrial Production in Ireland, hereinafter referred to as IIP. This measure is similar to a "Gross Domestic Product" (GDP) index, however unlike GDP, IIP is available on a monthly basis and thus allows for better modelling.

A study of the IIP raw data was carried out in the form of some basic statistical analysis to try and identify any underlying patterns in the data. It was found that the series was non-stationary after this analysis at which stage a number of transforms were taken to convert the series to a stationary one. The results of which are discussed in section 3.3.3. Following from these results it was found that the inverse transform of the data was most successful, although did not initially appear to convert the series to a stationary one. However on investigation it was believed that the series was possibly unit root. This was investigated and found to be true. Finally GARCH modelling was undertaken on the data and a final GARCH model decided on to model the variance in IIP.

2.0 LITERATURE REVIEW

Due to the nature of this thesis two academic areas were researched;

(i) Property Prices

(ii) GARCH modelling

2.1 PROPERTY PRICES

Upwards of 15 published papers in various Economic journals have been reviewed in order to understand the main drivers of property prices in a country and what type of modelling techniques have been used to predict property prices in the past.

From the review completed the general consensus is that the following factors are the main drivers of property prices:

(a) Real Incomes: what people earn taking inflation into account

(b) Demographics: the amount of people in the 20-29 age bracket

(c) Availability of Credit: how easy or difficult it is to get a mortgage

(d) Housing Stock: the basic economics of supply & demand

(e) Government Policy: tax & duty etc.

(f) User Cost: interest rates & inflation.

The factors listed above are referred to in property economics as fundamentals. Many of the papers reviewed addressed the problem in the Irish Property Market – "Does a speculative bubble exist?"; i.e. are houses in Ireland overpriced? The general conclusion is that although there may be a small element of inflation due to Government Policy e.g. stamp duty changes, a bubble does not exist and Irish property prices can be modelled using fundamentals.

An enormous amount of different models exist for predicting house prices. However there are a number of elements in the current situation which are unprecedented and make the modelling process difficult – these are:

i. The size and duration of the current real house price increases.

ii. The degree to which they have tended to move together across countries.

iii. The extent to which they have disconnected from the business cycle.

Another factor to consider in modelling house prices is the Business Cycle. In the past house prices and business cycle turning points roughly coincided from 1970 -2000.

The current price boom worldwide is strikingly out of step with the business cycle.

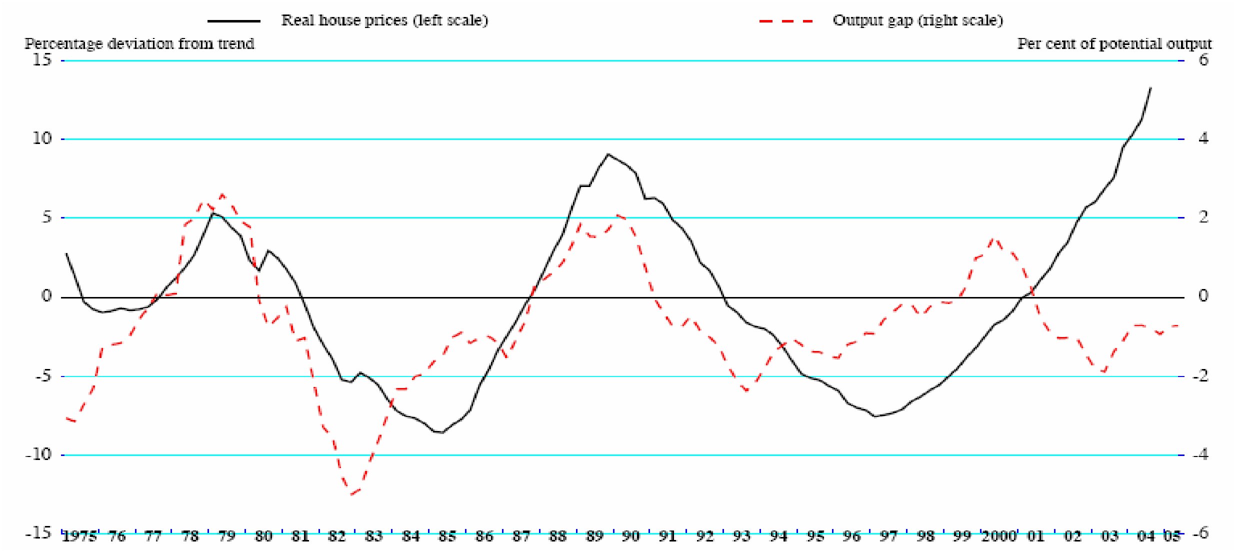


Fig 2.1: Real House Prices versus Business Cycle

Note: Real House Price Index: Calculated using purchasing power parity adjusted GDP weights. OECD (Organisation for Economic Co-operation & Development) Output gap determines the Business Cycle.

As can be seen from figure 1 above property prices in general around the world seem to be increasing (18 countries were sampled to produce figure 2).

Of these countries there appears to be a general consensus among the papers reviewed that property is overvalued in the following countries:

(i) United Kingdom (ii) Ireland (iii) Spain [1]

Contrary to this opinion, many of the papers reviewed on the Irish market state that a bubble does not exist due to the fact that the models fail because the following cannot be taken into account:

(i) Changes in regulatory conditions
(ii) Tax changes
(iii) Demographics [5]

2.2 GARCH MODELLING

There are two major papers of interest in this area. The first is by Robert Engle who developed ARCH modelling in 1982 [15]. The second by Tim Bollerslev who developed GARCH modelling and introduced it in his paper entitled "General Autoregressive Conditional Heteroskedastistic Modelling". In this paper Bollerslev introduces the extension of the ARCH process introduced by Engle (1982) to the GARCH process, which bears resemblance to the extension of the standard time series AR process to the general ARMA process. [8]

Heteroskedasticity refers to unequal variance in the regression errors, in other words time varying variance. Heteroskedasticity can arise in a variety of ways and a number of tests have been proposed to test for its presence. Typically a test is designed to test the null hypothesis of homoskedasticity (equal error variance) against some specific alternative heteroskedasticity specification.

The particular models of interest here are GARCH (General Autoregressive Conditional Heteroskedastistic) models. The best way to describe how this modelling technique will be applied is via an example. Suppose we look at the return on an asset or portfolio over time and the variance of the return represents the risk level of those returns; this is a time series application where heteroskedasticity is an issue. Looking at the returns of an asset over time will suggest that some time periods are riskier than others, that is, the expected value of the magnitude of error terms at some times is greater than at others. Moreover these risky times are not scattered randomly across quarterly or annual data, instead there is a degree of autocorrelation in the 'riskiness' of returns. The amplitude of the returns varies over time and this is described as volatility clustering. The ARCH & GARCH models are designed to deal with just this set of issues. The goal of these models is to provide a volatility measure (like a standard deviation) that can be used to make decisions relating to the data set in question. The challenge is to specify how the information can be used to forecast the mean and variance of the return conditional on the past information. Many methods exist to use the mean to forecast future returns, however until the introduction of ARCH/GARCH models virtually no methods were available involving variance.

The primary descriptive tool was the rolling standard deviation i.e. calculated using a fixed number of the most recent observations e.g. the rolling standard deviation could be calculated every day using the most recent month of observations (22 days) this can be thought of as the first ARCH model. This model assumes that the variance of tomorrows return is an equally weighted average of the squared residuals from the last 22 days. The assumption of equal weights seems unattractive as one would think that more recent events would be more relevant. Also the assumption of zero weights for observations of more than a month old is also unattractive. The ARCH model proposed by Engle (1982) let these weights be parameters to be estimated, thus the model allowed the data to determine the best weights to use in forecasting the variance. The main problem which arises with the ARCH model is that of the necessity of high order ARCH models required to catch the dynamic of the conditional variance. This high order implies that many parameters have to be estimated and the calculations of these can become prohibitive. [9]

The GARCH model was introduced by Bollerslev (1986). This model is also a weighted average of past squared residuals but it has declining weights which never go completely to zero. The most widely used GARCH specification asserts that the best predictor of the variance in the next period is a weighted average of the long run average variance, the variance predicted for this period and the new information this period which is the most recent squared residual

The GARCH model is based on an infinity ARCH specification and it allows reduction of the number of parameters to be estimated from an infinite number to just a few. In Bollerslev's GARCH model the conditional variance is a linear function of past squared innovations and earlier calculated conditional variances. [8]

3.0 STATISTICAL ANALYSIS

This section deals with background statistics in relation the raw series of data to try and establish any underlying patterns which will allow for better modelling.

3.1 ASSUMPTIONS WHEN DEALING WITH DATA

The following assumptions are used when modelling a time series:

1. Linearity in the parameters: $E(Y_t)$ is a linear function of the parameter β regardless of the relationship between Y_t and time. ($E(Y_t) = \alpha + \beta t$).
2. Homoskedasticity of ε_t: this means that the error terms have equal variance. It implies that out independent variable Y_t also possesses equal variance as it is a linear function of ε_t.
3. Normality of residuals: ε_t is an independent identically distributed random variable that is approximately normal with $E(\varepsilon_t)=0$ and $VAR(\varepsilon_t)=1$, although this assumption is often violated in time series analysis we will show that with a large enough sample size we do not need normality of the error terms.
4. Independence: Any value of the dependent variable Y_t is statistically independent from any lagged value Y_{t-k} for k=1,2,….

3.2 MODEL BUILDING WITH THE RAW DATA SET

3.2.1Data description and analysis

Data	**No. of observations (n)**	**Sample Mean**	**Sample Variance**	**Minimum**	**Maximum**
Y_t	216	63.673	1360.8	18.1	139

Table 3.1

Where Sample Mean = $\overline{Y_t} = \sum_{t=1}^{n} \frac{Y_t}{n}$ (3.1)

and Sample Variance = $Var(Y_t) = \sum_{t=1}^{n} \frac{(Y_t - \overline{Y})^2}{n-1}$ (3.2)

As can be seen from the above the IIP index has been highly volatile during the period observed.

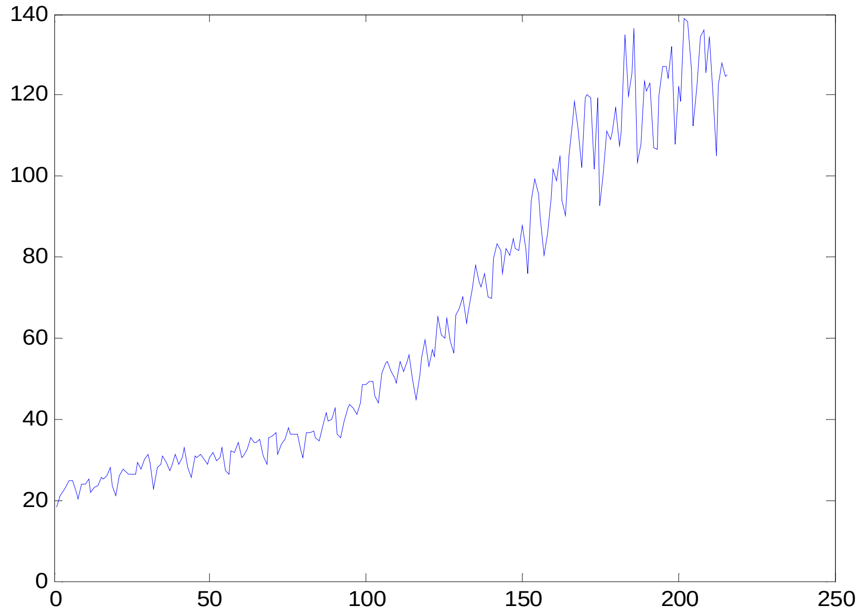


Fig 3.1: Plot of IIP

As can be seen from the above plot there is quite obviously a trend in the data, it appears to be time dependent i.e. it is increasing with time. From this we would deduce that the Y_t observed are not independent and thus non-stationary.

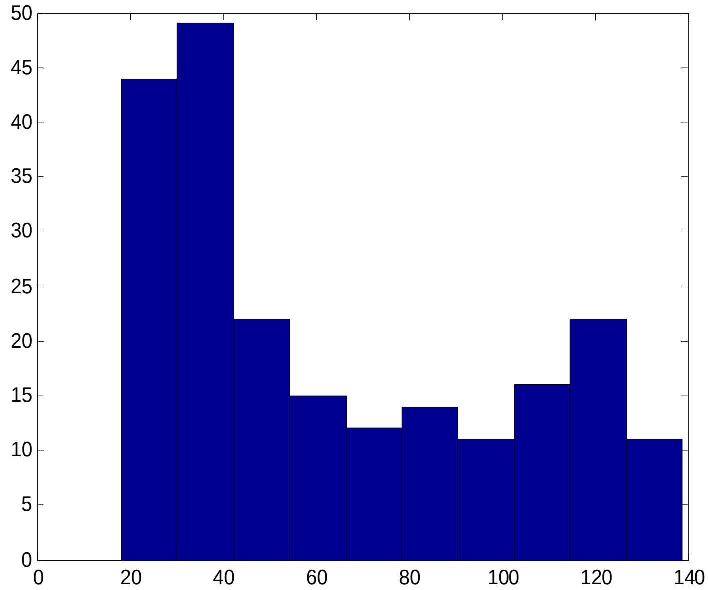


Fig 3.2: Histogram of IIP

We note here the median is 50.65 which is to the left of the mean.

3.2.2 Model of the raw data

We will test for autocorrelation by examining the sample autocorrelation function (SACF) which is given by

$$SACF = \frac{\sum_{t=1}^{n-k}(Y_t - \overline{Y_t})(Y_{t+k} - \overline{Y_{t+k}})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t+k})}} \qquad (3.3)$$

We will also test for autocorrelation by visual inspection of the SACF graphed against different lags.

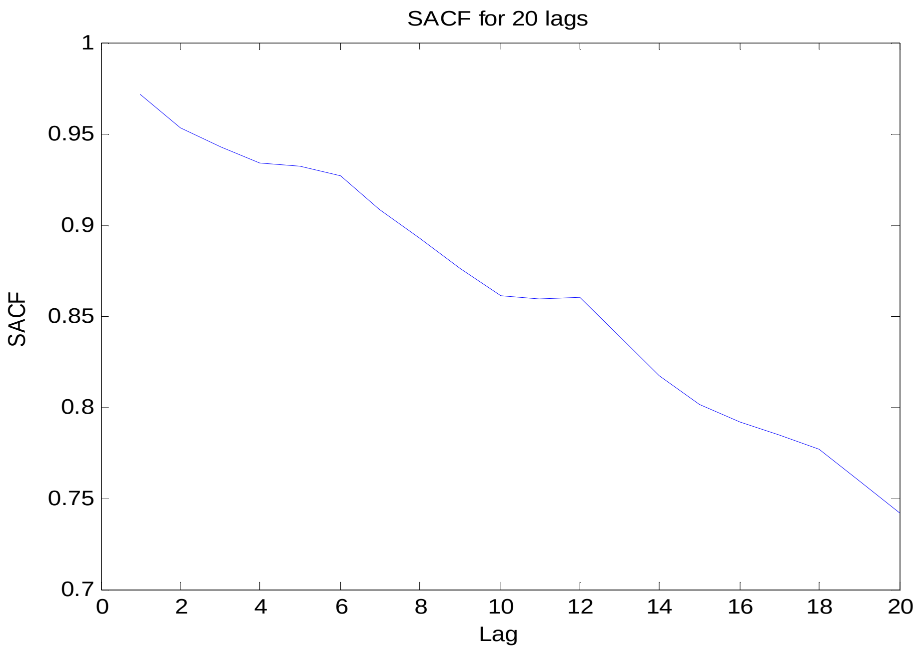


Fig 3.3: SACF for 20 lags of IIP

The above graph would suggest that the series is stationary and that an Autoregressive model with lag 1 (AR(1)) is suitable for modelling the series. There does appear to be a kick at the lag 12 which needs further examination.

We now examine the Sample Partial ACF (SPACF) at a lag of τ given by:

$$\hat{r}_x(\tau,\tau) = \begin{cases} \hat{r}_x(1), \tau = 1 \\ \dfrac{\hat{r}_x(\tau) - \sum_{j=1}^{\tau-1} \hat{r}_x(\tau-1, j)\hat{r}_x(\tau-1)}{1 - \sum_{j=1}^{\tau-1} \hat{r}_x(\tau-1, j)\hat{r}_x(j)}, \tau = 2,3,... \end{cases} \quad (3.4)$$

where $\hat{r}_x(\tau,\tau)$ is the SPACF at a lag of τ, $\hat{r}_x(\tau)$ is the SACF at a lag of τ, and $\hat{r}_x(\tau, j)$, for $j \neq \tau$ is defined (Bowerman & O'Connell, 1987) as:

$$\hat{r}_x(\tau, j) = \hat{r}_x(\tau-1, j) - \hat{r}_x(\tau,\tau)\hat{r}_x(\tau-1, \tau-j) \text{ for j=1,2,…., τ-1} \quad (3.5)$$

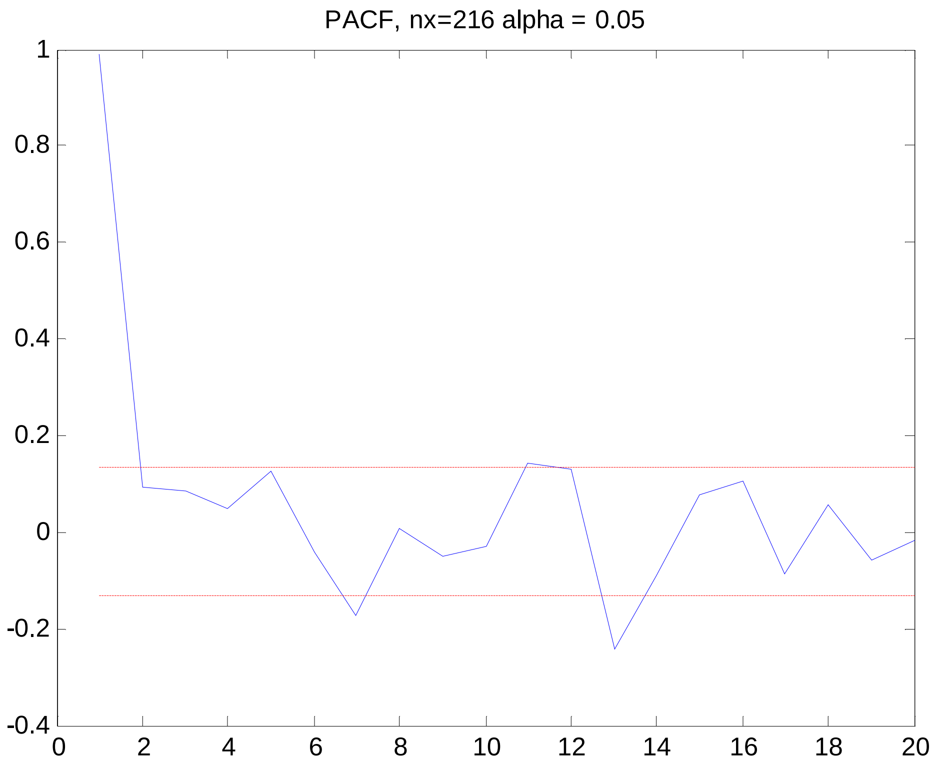


Fig 3.4: PACF of IIP

Our PACF figure also shows a significant peak at lag 12, i.e. a year lag.

As such we now use least squares to solve the following AR model for θ_1 & θ_2 :

$$x(k) = \theta_1 x(k\text{-}1) + \theta_2 x(k\text{-}12) + \varepsilon(k) \qquad (3.6)$$

Using our time series for IIP we let this represent x(k) and call it Y (a vector). We then obtain x(k-1) by delaying Y by one and call this X1 (a vector). We obtain x(k-12) by delaying Y by twelve and call this X2. We then set X = [X1, X2] a matrix. θ is obtained as follows;

$$\theta = (X^T X)^{-1} X^T Y, \text{ where } \theta = [\theta_1, \theta_2] \qquad (3.7)$$

For the above data $\theta_1 = 0.4279$ and $\theta_2 = 0.65856$

3.2.3 Analysis of errors

We now proceed to errors between the actual IIP data and the fitted model.

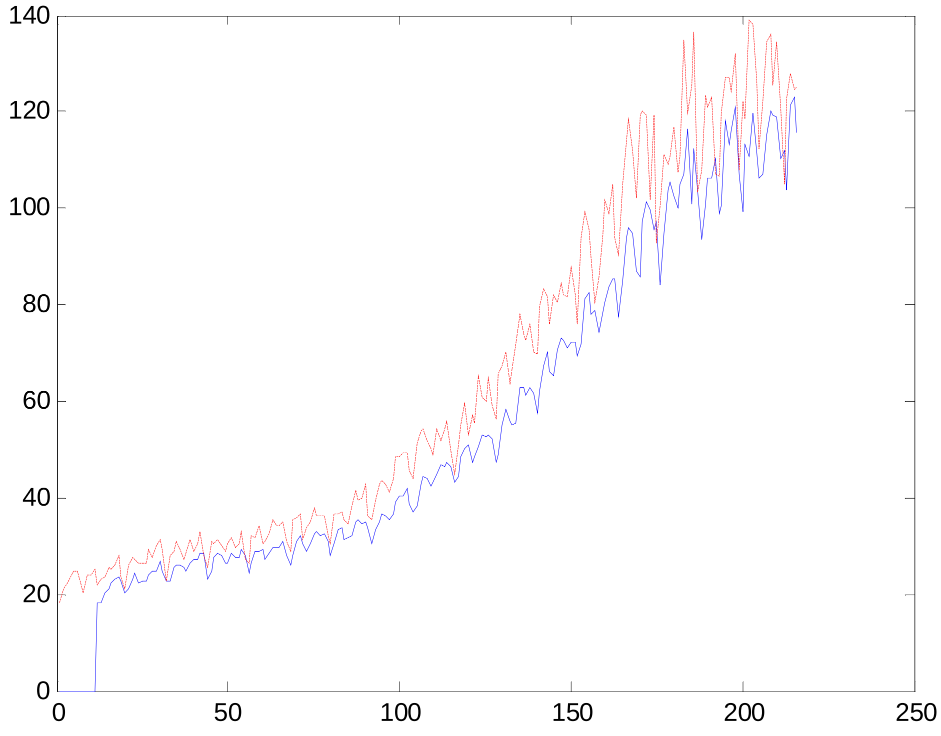


Fig 3.5: Actual IIP (dashed line) versus forecasted IIP

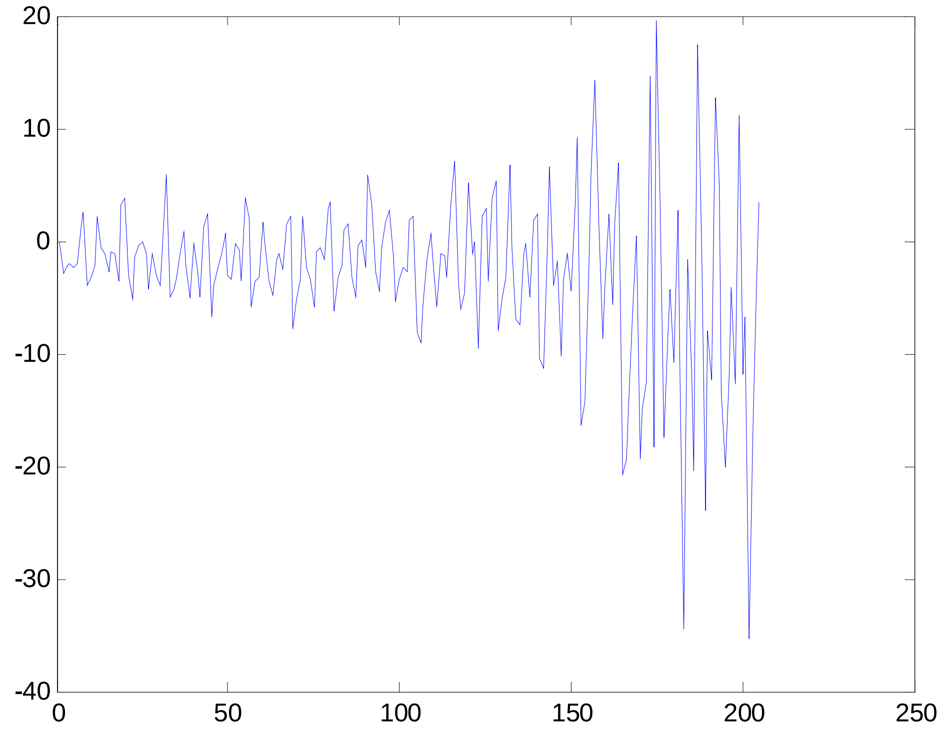


Fig 3.6: Plot of error between actual IIP and forecasted IIP

From the above plot we can see that the error varies with time i.e. it exhibits heteroskedasticity.

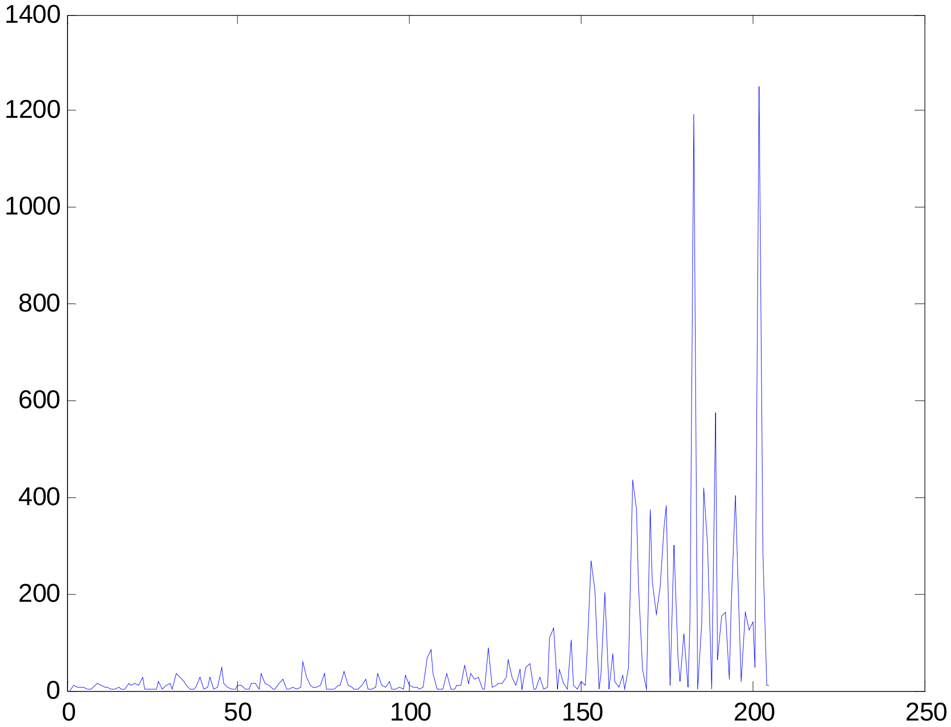


Fig 3.7: Plot of error squared

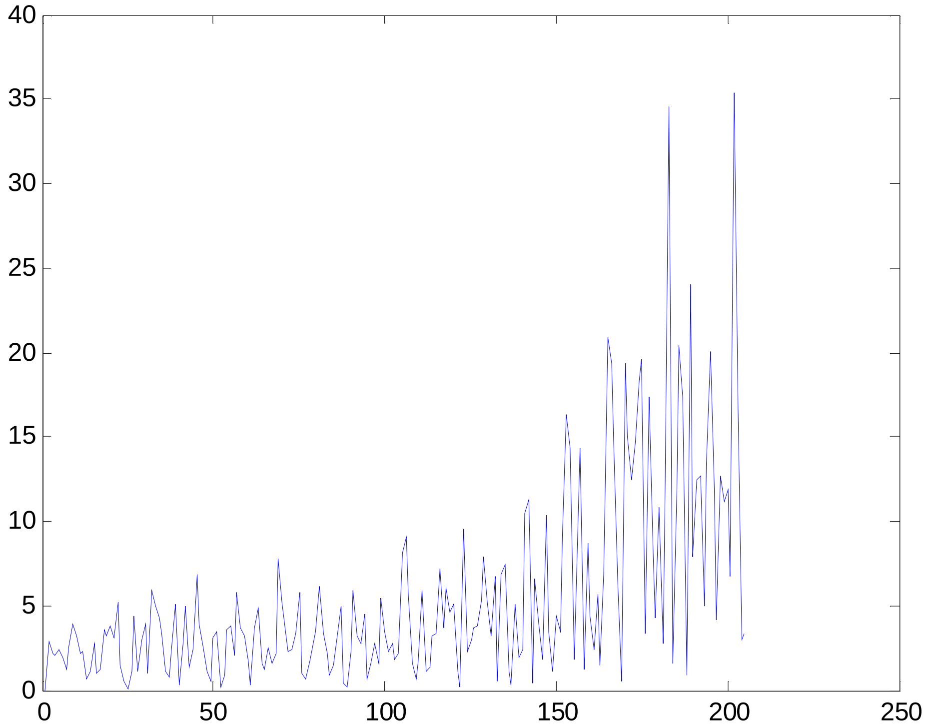


Fig 3.8: Plot of absolute value of error

We now want to model the error based on the actual value of the time series and as such we now examine the ACF of the error squared;

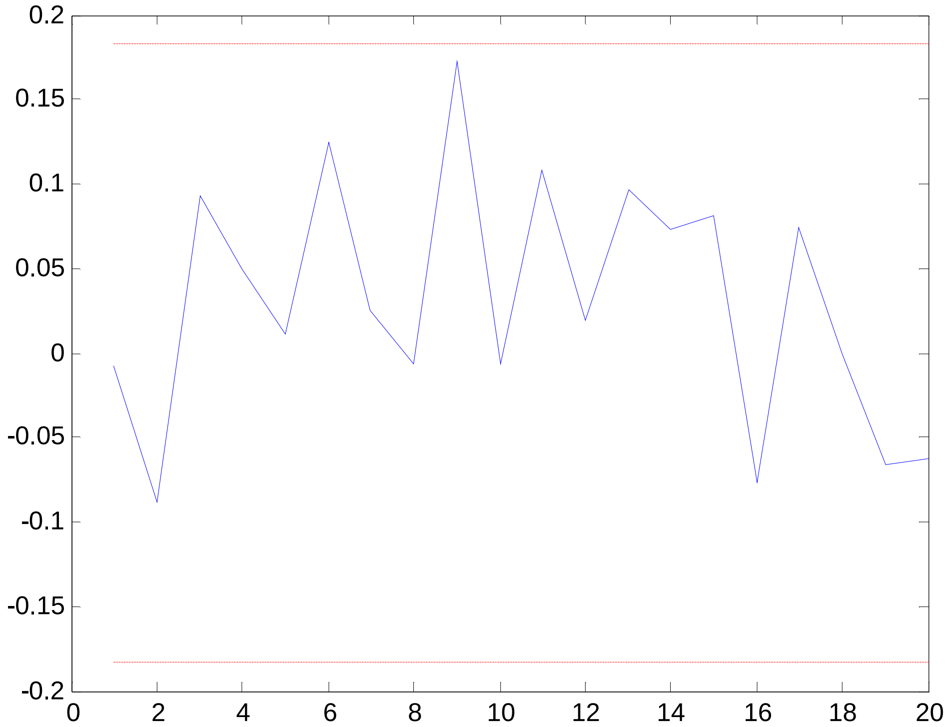


Fig 3.9: ACF of error squared

From the above plot it appears that the error looks random (which is good as it means our original forecast model of IIP performs well). However as no particular lag is evident we cannot use this as a means of building a model for the error.

The following accuracy measures are used;

(i) Mean Absolute Percentage Error (MAPE): $\dfrac{\sum_{t=1}^{n}\left|(Y_t - \hat{Y}_t)/\hat{Y}_t\right|}{n} \times 100$ (3.8)

(ii) Mean Absolute Deviation (MAD): $\dfrac{\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|}{n}$ (3.9)

(iii) Mean Squared Deviation (MSD): $\dfrac{\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|^2}{n}$ (3.10)

The MAPE measures the accuracy of the fitted time series values as a percentage value and here it equals -5.3067. The MAD expresses accuracy in the same units as the data, which helps conceptualize the amount of error, here it is 5.3286. Finally the MSD is always computed using the same denominator, n, regardless of the model, so you can compare MSD values across models. MSD is a more sensitive measure of an unusually large forecast error than MAD. Here it equals 61.077.

From Fig 5 above we see the volatility of the series tends to increase over time, particularly towards the latter stages, this pattern suggests heterogeneity of variance may be present in our time series, which violates one of our main assumptions. The series does appear to be time stationary though as the expected value doesn’t appear to vary with time in any significant way.

3.2.4 Analysis of variance

We will now look at three different residual plots as these will help us determine whether or not any more of our assumptions have been violated. The three graphs are as follows;

(i) Residuals versus the fitted value (Fig 3.10)

(ii) Histogram of the residuals (Fig 3.11)

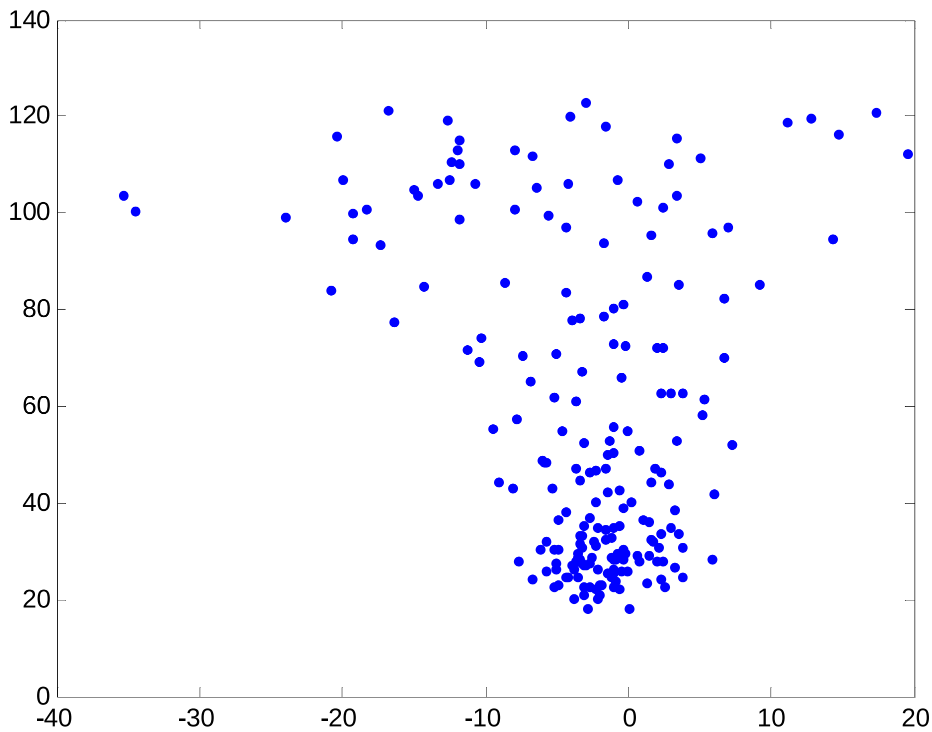


Fig 3.10: Residual versus the Fitted Values

From Fig 9 we can see a fanning out effect among the residuals which as before would lead us to conclude that our series possesses non constant variance.

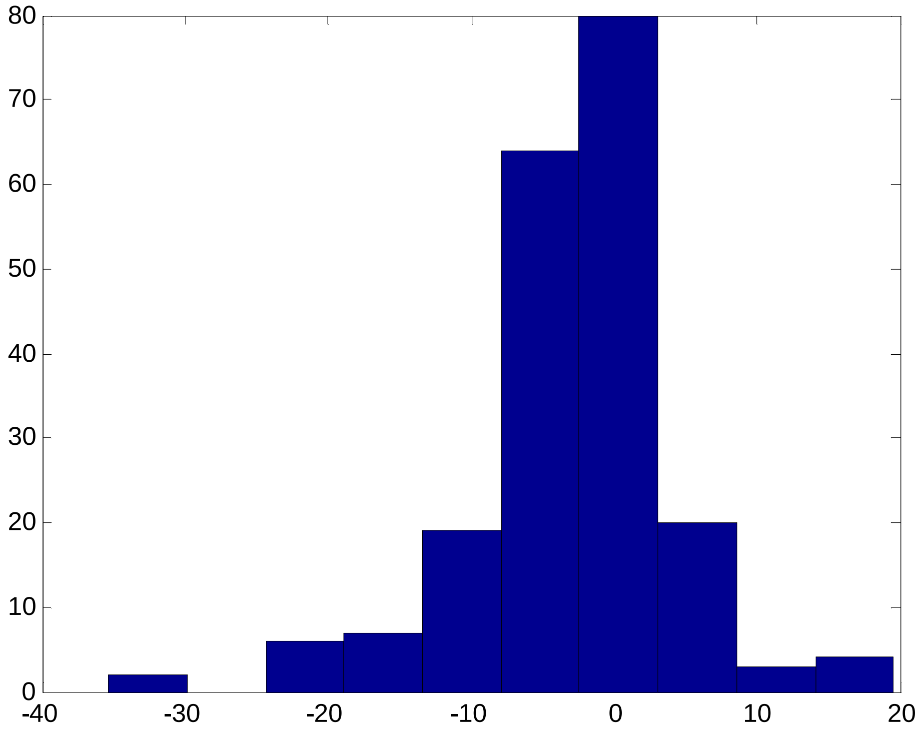


Fig 3.11: Histogram of Residuals

From Fig 10 we may be tempted to conclude that the distribution of e_t is approximately normal, however since we have already seen that our series possesses non-constant variance we can conclude that the distribution is not normal.

We will now use the method of weighted least squares to deal with non constant variance. Due to the series e_t possessing non-constant variance some of the observations in our sample provide more reliable information about the regression function than others. Observations with small variance are more reliable than observations with large variance and should therefore be given more weight. Denote the variance of the error term e_t by $E[e_t^2] = \sigma_{et}^2$. By assigning a weight $w_t = 1/ \sigma_{et}^2$ to our model we have a new regression function expressed in matrix notation of the form:

$$Y = B^T TW + E \ \& \ E(Y) = B^T TW \text{ since } E(E) = 0. \qquad (3.11)$$

Where;

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{pmatrix} \quad \underline{B} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad \underline{T} = \begin{pmatrix} 1, t_1 \\ 1, t_2 \\ ..... \\ ..... \\ 1, t_n \end{pmatrix} \quad \underline{W} = \begin{pmatrix} w_1, 0, 0, ...0 \\ 0, w_2, 0, ...0 \\ .... \\ .... \\ 0, 0, ...w_n \end{pmatrix} \quad \underline{E} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{pmatrix} \quad (3.12)$$

Such a model should rectify our violation of the assumption of constant variance and yield a variance which is constant over time. σ_{et}^2 is unknown and must be estimated from our data. By assumption σ_{et}^2=E($s_{e_t}{}^2$).

Unfortunately $s_{e_t}{}^2$ is unknown and must be estimated. We have observed that $s_{e_t}{}^2$ tends to increase over time (Fig 3.6); this means that the fits obtained from an ordinary least squares regression of e_t using time as predictor variable are unbiased estimator of $s_{e_t}{}^2$, provided that the regression function is appropriate, meaning that it accurately describes the relationship between ε_t and time. The validity of this result can be easily shown by observing that

$$s_{e_t}{}^2 = E(e_t^2) - (E(e_t))^2 \qquad (3.13)$$

and since $E(e_t)=0$ it follows that $s_{e_t}{}^2 = E(e_t^2)$.

Finally since $E(e_t^2) = E(\hat{e}_t^2)$ it follows that $E(\hat{e}_t^2)$ can be used to estimate $s_{e_t}{}^2$.

Similarly, one can estimate the standard deviation of the residual by regressing the absolute value of the error term against time. We must decide carefully whether to estimate for variance or standard deviation in our weights.

The variance is less 'forgiving' of observations with relatively large variances and is more appropriate when the discrepancies in the variance appear to be small. Since our residual plot against time (Fig 5) displayed a fanning out effect where some residuals displayed a variance of considerable greater magnitude than others the estimate of the standard deviation of the residuals is a more appropriate choice of weight in this case.

Now we need to estimate the standard deviation of the residuals by choosing a functional form which best describes the relationship between the squared value of the residuals and time.

Since it is hard to justify any theoretical relationship between the population variance and time it is best that we proceed by observing a plot of e_t^2 against time and a fitted regression line.

As expected the series does become more volatile over time. It is obvious from the graph in Fig 3.8 that the absolute value of the residuals is probably an exponential or quadratic function of time. However, we won't completely discount the possibility of the relationship being linear.

First we fit a linear model to the error squared the resulting model is $y = 2.4565 + 0.0277t$, where y is the error squared and t is time.

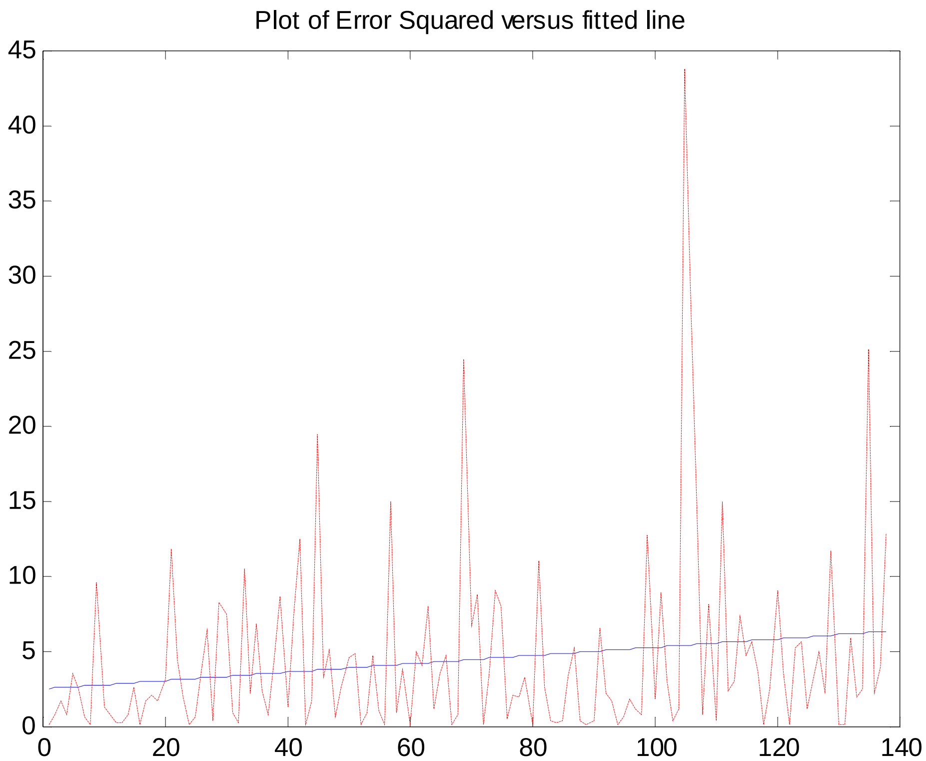


Fig 3.12: Plot of error squared versus fitted line

Next we fit a quadratic model to the error squared and the resulting model is;
$y = 1.9975 + 0.0473t-0.0001t^2$, again y is the error squared and t is time.

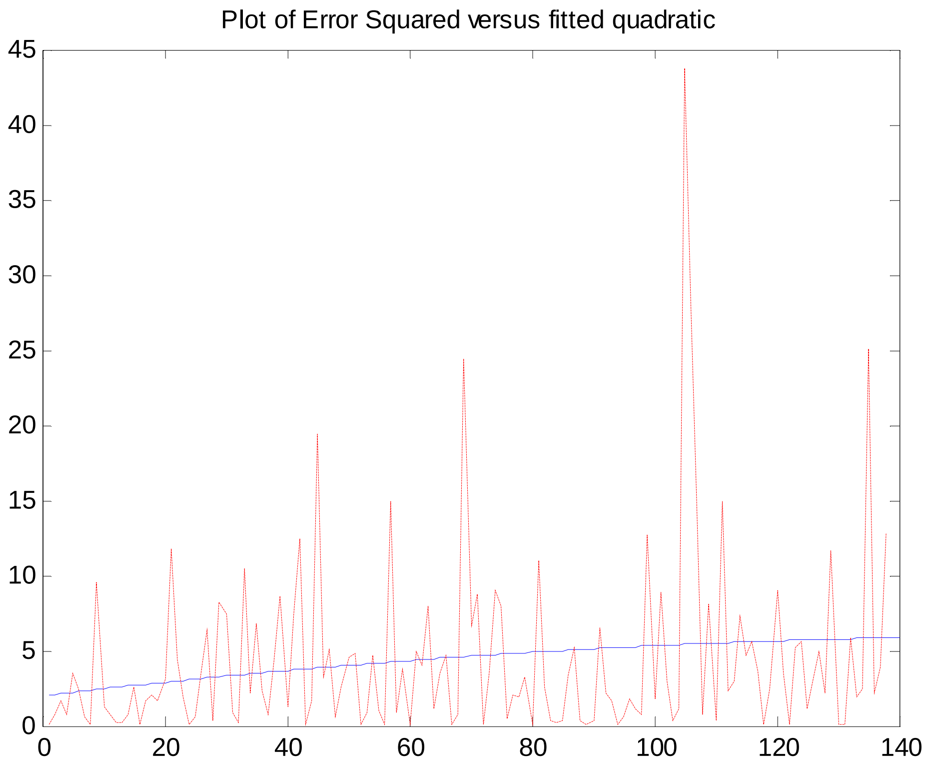


Fig 3.13: Plot of error squared versus fitted quadratic

We now fit an exponential curve to the error squared of the form: $y = ae^{bt}$. The model then becomes $\log(y) = \log(a) + bt$. The fitted model is: $\log(y) = -0.04559 + 0.00561t$. We then take the exponential of the log(y) to recover the fitted curve.

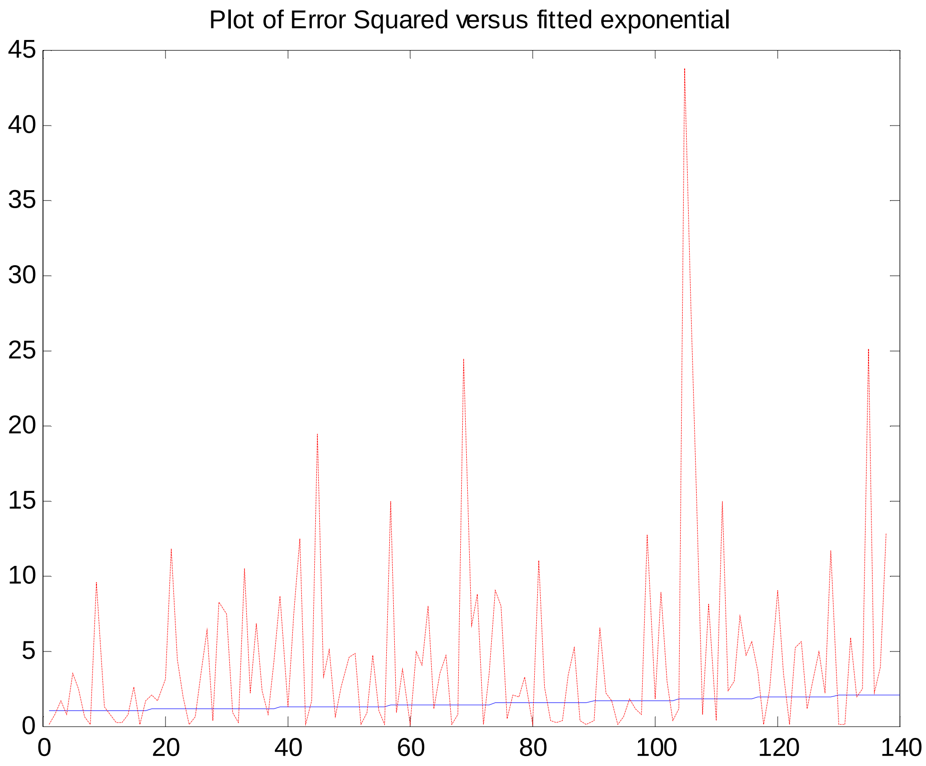
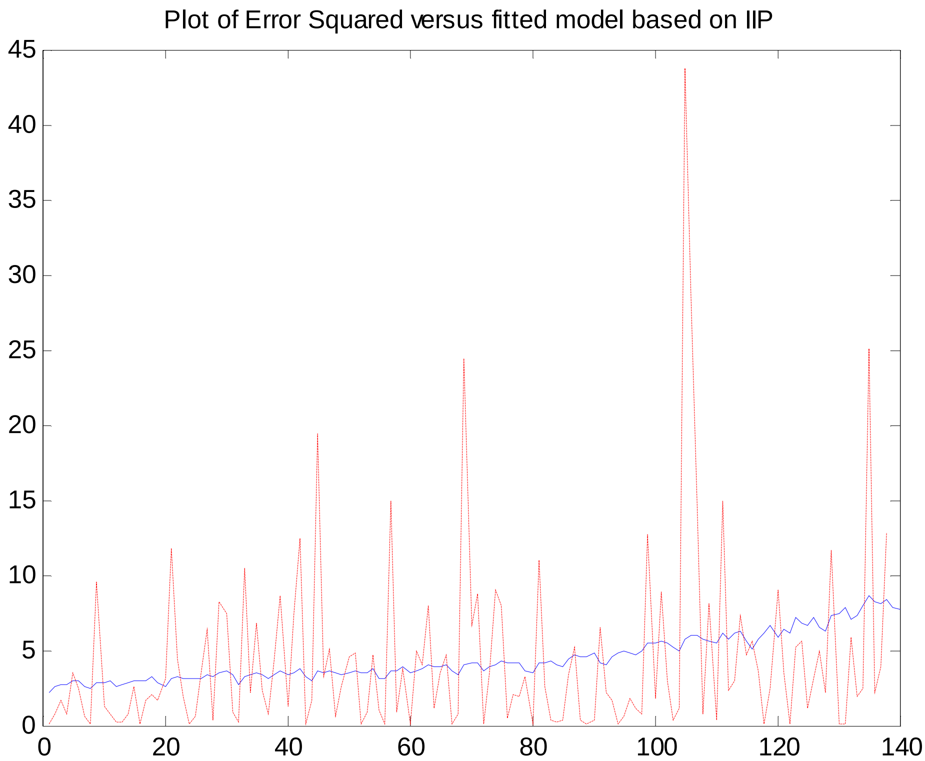


Fig 3.14: Plot of error squared versus fitted exponential

Finally we attempt to model the error squared as a function of actual IIP i.e. x(k) = a + by(k) where x is the error squared and y is actual IIP.



Plot 3.15: Plot of error squared versus fitted model based on IIP

3.2.5 Weighted Least Squares (WLS)

We now try to estimate a new AR model on the basis of WLS. We will each of the fitted models of the error above as our estimates of σ_t^2 and use this to calculate the W matrix as outlined in equation 3.12. We will then test the resulting new error for homogeneity of variance using Levenes test as outlined above.

In the first case we use the estimates of σ_t^2 as produced by the linear fitted model above. The new fitted model after carrying out WLS is:

$$y(k) = 0.3882y(k-1) + 0.42904y(k-12). \qquad (3.14)$$

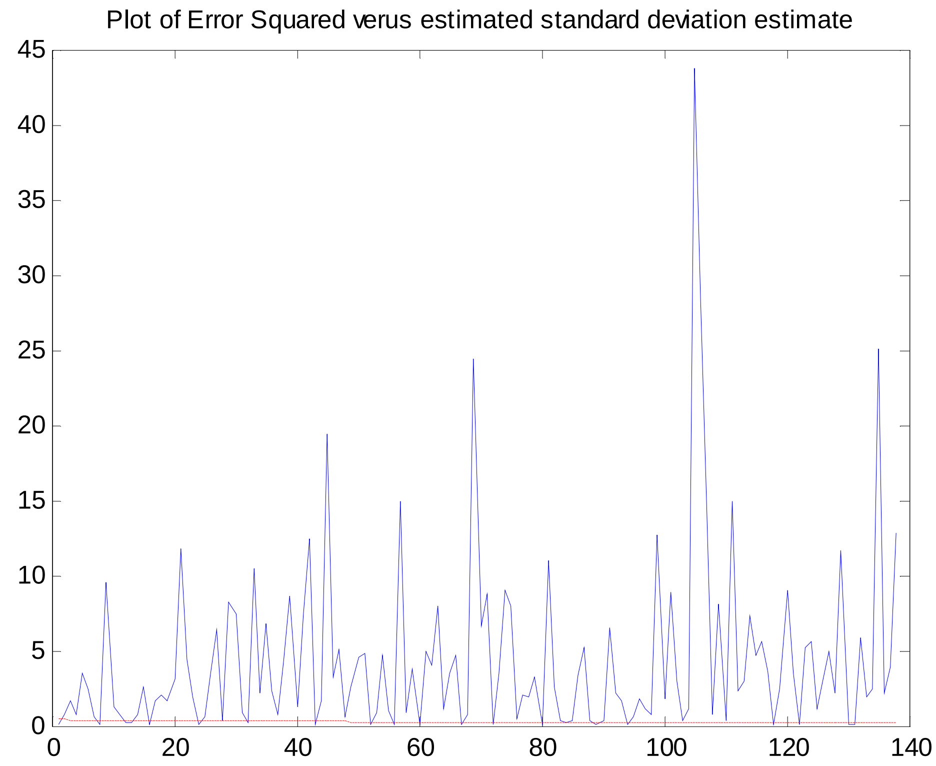


Fig 3.16: Plot of error square versus estimate standard deviation estimate

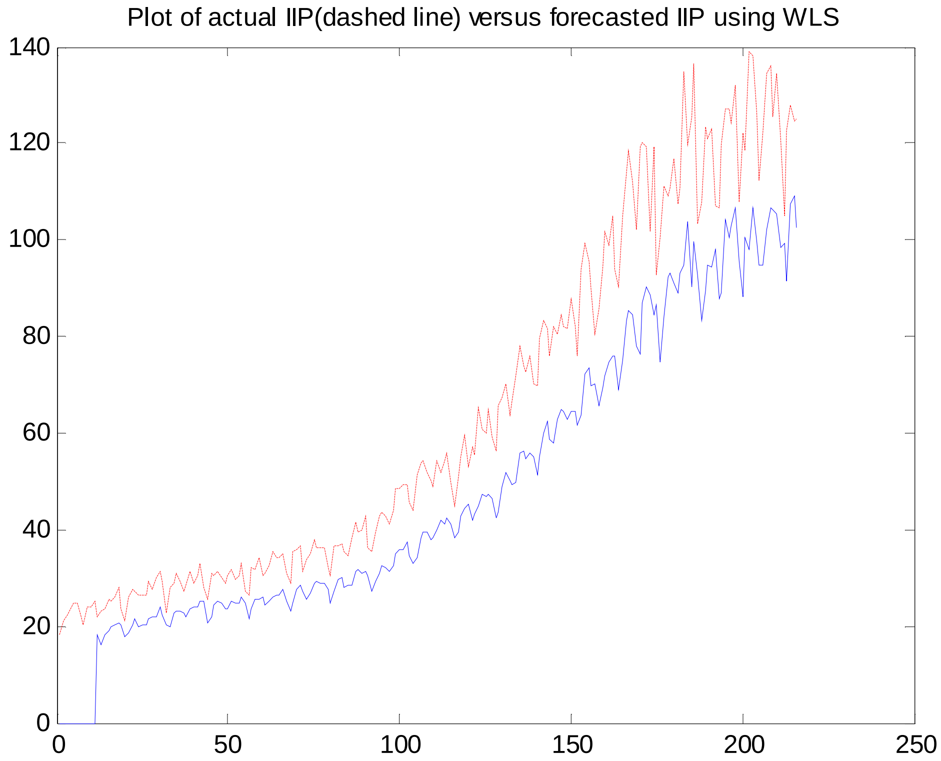


Fig 3.17: Plot of actual IIP versus fitted IIP using WLS

As can be seen from Fig 3.17 our model is much worse than the original model – see plot of the new error squared below.

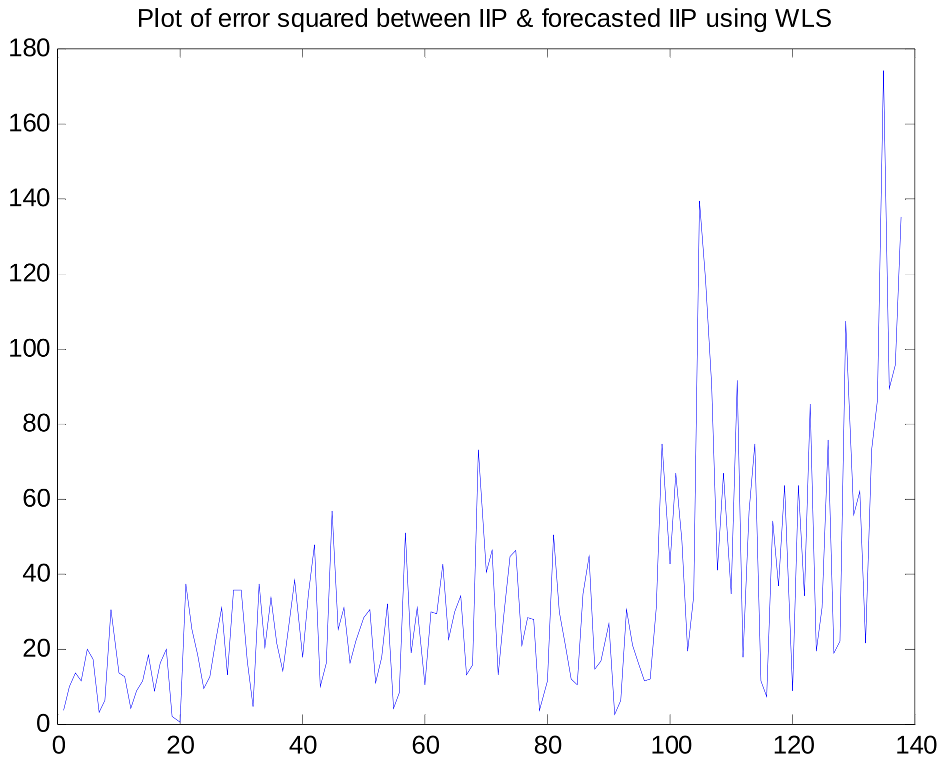


Fig 3.18: Plot of error squared.

The magnitude of the error squared is much greater than our original model. It is also noticeable from the plot that using WLS has not solved the problem of heterogeneity of variance. Please table of Levenes Statistics under results section 3.2.6 below.

In the second case we use the estimates of σ_t^2 as produced by the quadratic fitted model above. The new fitted model after carrying out WLS is

$$y(k) = 0.38798y(k-1) + 0.42861y(k-12) \qquad (3.15)$$

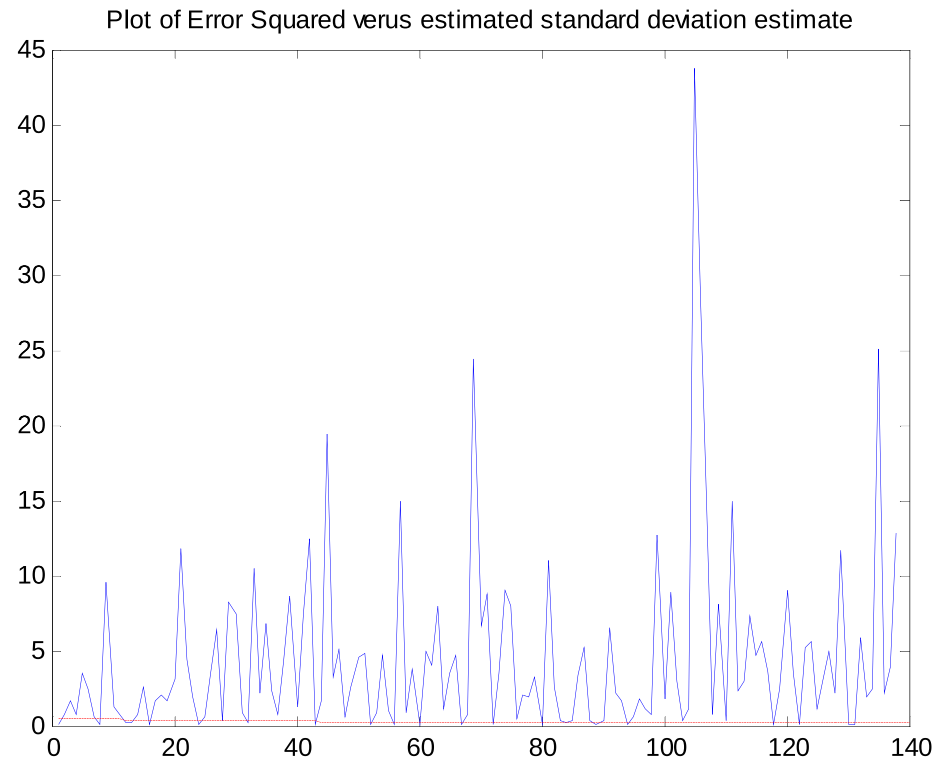


Fig 3.19: Plot of error squared versus estimated standard deviation

The results are very similar to those produced above with the forecast model much worse than before, please see plot of the error squared below:

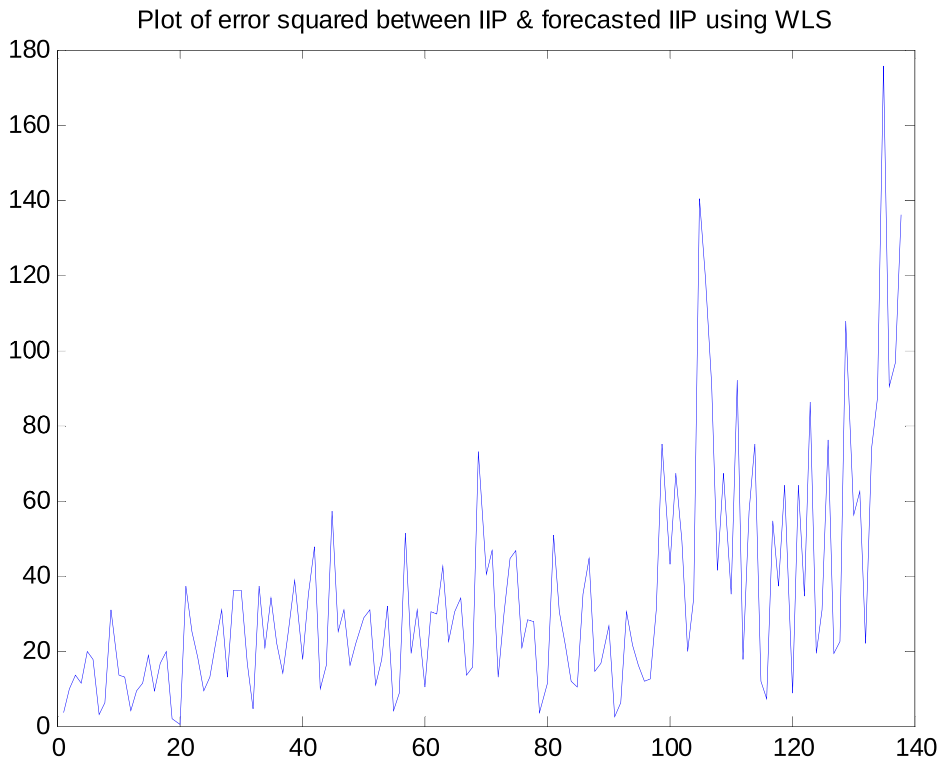


Fig 3.20 Plot of error squared

Again the magnitude of the error appears much larger and the problem of heterogeneity of variance remains.

In the third case we use the estimates of σ_t^2 as produced by the exponential fitted model above. The new fitted model after carrying out WLS is

$$y(k) = 0.38452y(k-1) + 0.4315y(k-12) \qquad (3.16)$$

We again receive very similar results to those achieved above, again with the magnitude of the error having increased and the problem of heterogeneity remaining unsolved.

In the final case we use the estimates of σ_t^2 as produced by the fitted model based on IIP. The new fitted model after carrying out WLS is:

$$y(k) = 0.38278y(k-1) + 0.43438y(k-12) \qquad (3.17)$$

Once again the same results emerge. Please see appendix

3.2.6 Levenes Test Statistics Results

There are four cases to be discussed as follows;
(i) Error squared modelled as line
(ii) Error squared modelled as quadratic
(iii) Error squared modelled as exponential
(iv) Error squared modelled with actual IIP

Please note that H_0 assumes that there is equal variance i.e. the series is homoskedastistic

Case	Sample	Sample Size	Variance	F Statistic	Significance Level	H_0
i	1	40	116.4465	27.7001	0.05	Reject
	2	60	255.5862			
	3	38	1568.64			
ii	1	40	117.6287	27.7939	0.05	Reject
	2	60	258.0327			
	3	38	1587.7177			
iii	1	40	117.7524	28.0681	0.05	Reject
	2	60	257.8663			
	3	38	1598.8305			
iv	1	40	115.0518	28.0312	0.05	Reject
	2	60	252.0099			
	3	38	1560.9044			

Table 3.2

We conclude that in all cases WLS has failed to solve the problem of heteroskedasticity. This is due to the fact that the original model was not a good enough fit in the first place. We now attempt to transform the series in an effort to achieve a series which can be modelled better than the raw data.

3.3 MODELLING TRANSFORMS

As stated at the outset of section 3.2 the raw data series is non-stationary, as such we now undertake some transforms of the data and examine the results. A well documented method of transforming the data is to use Box-Cox transforms. The Box-Cox procedure automatically identifies a transformation from the family of power transformations on Y. The family of power transformations is of the form $Y' = Y^{\lambda}$, where λ is a parameter to be determined from the data. Note that this family encompasses the following simple transformation which I have undertaken in analysing the data [13]:

λ	Y'
2	Y^2
½	$\sqrt{Y}$
0	Log_eY (by definition)
- ½	$\frac{1}{\sqrt{Y}}$
-1	$\frac{1}{Y}$

Table 3.3

The method I used to identify the best value for λ was as follows;

1. Convert the time series according to the above-mentioned transforms
2. Model the converted series using techniques outlined above i.e. using the ACF & PACF to establish which lags to use for an AR or MA model.
3. Using the series X generated by the relevant model, I took the inverse of the transform to get the modelled series X'.
4. Calculated the residuals: r=Y-X'.
5. Calculated the SSE and the MSE of the residuals in each case.

A model was constructed for each of the above transforms, in the sections below namely 3.3.1 & 3.3.2 we deal with the log transform and inverse transform respectively as these had the most promising results. The remainder of the models are discussed in Appendix A. The results and comparisons from all the models are dealt with in section 3.3.3

3.3.1 Log Transform of the Data

Given that the underlying distribution of the IIP is possibly exponential we now take the log values of the data to covert the series to a linear series. A little more time was taken with this transform due to the belief of the possible underlying distribution.

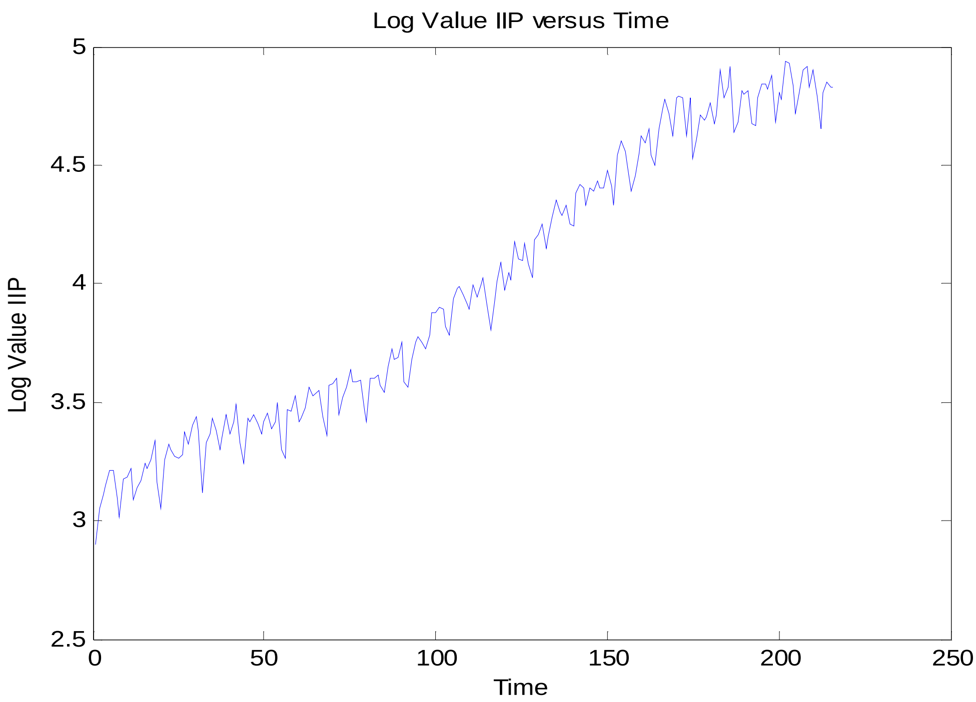


Fig 3.21: Log Value of IIP versus Time

We look at the ACF & PACF of this to try and establish a model for the series:

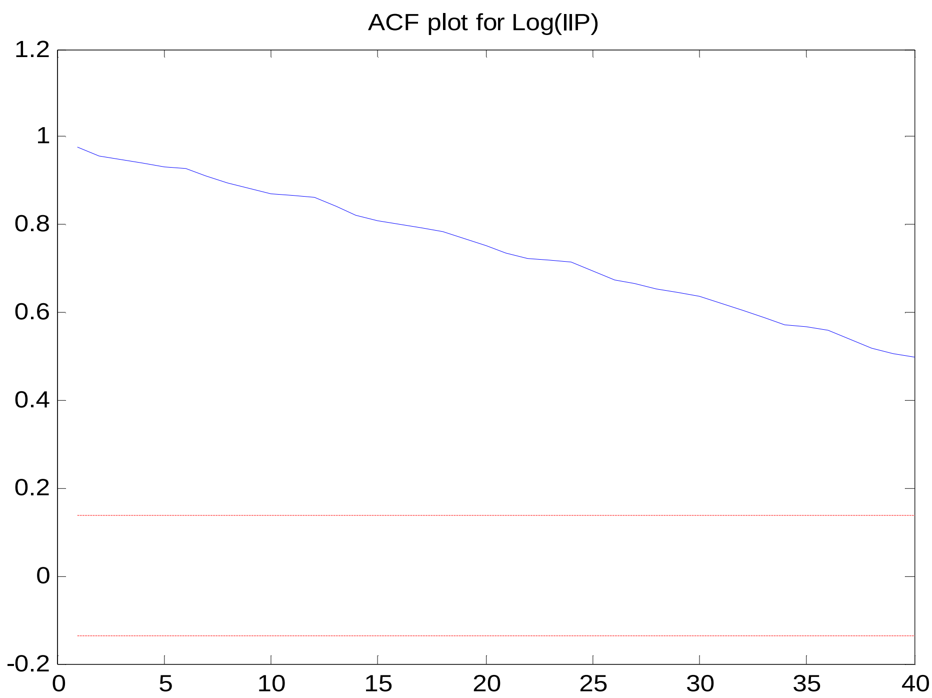


Fig 3.22: ACF of Log(IIP)

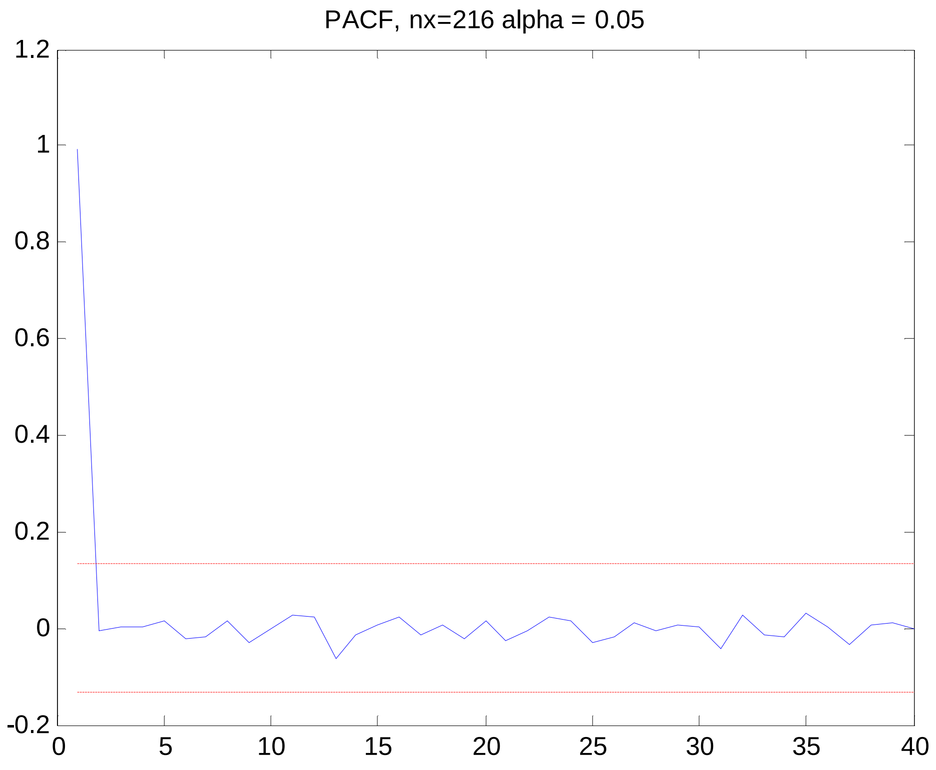


Fig 3.23: PACF of Log(IIP)

From the above figures the process appears to be AR(1):

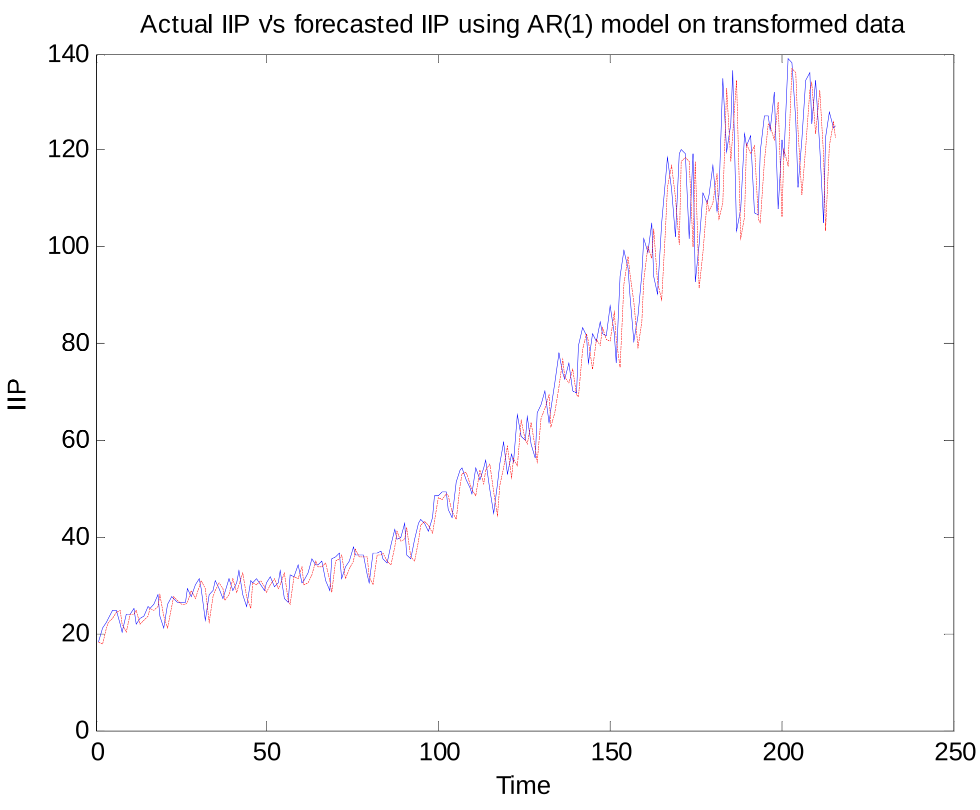


Fig 3.24: Plot of IIP versus model obtained using log transform

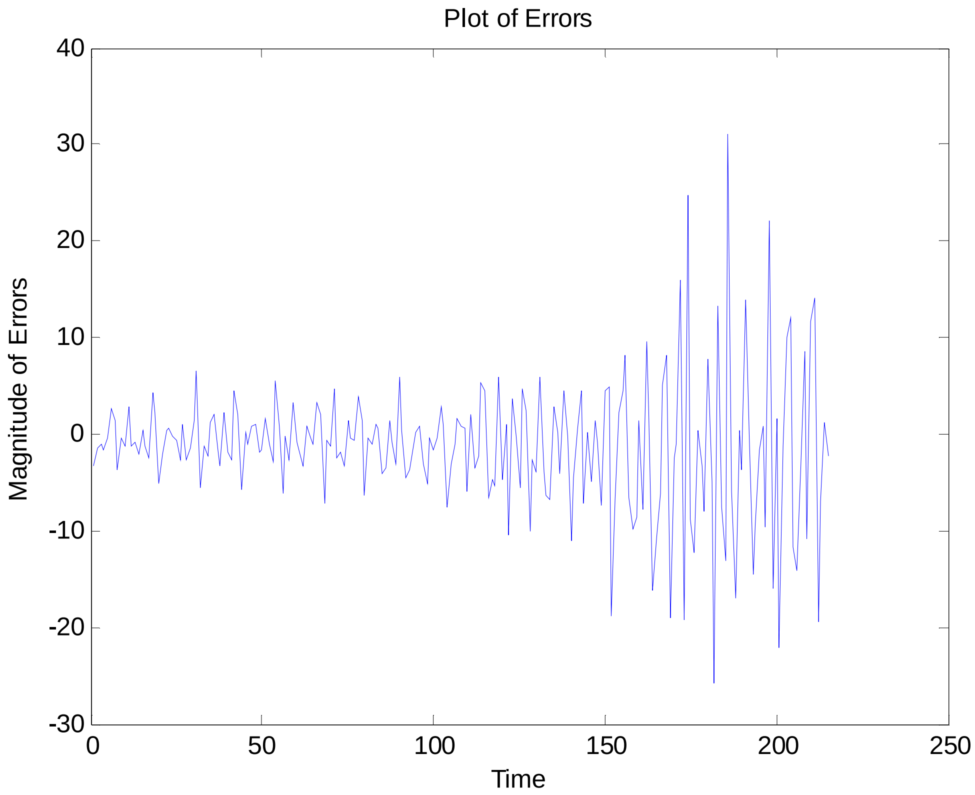


Fig 3.25: Plot of the error between actual IIP & forecasted IIP using AR(1) model on transformed data.

As can be seen there are quite a large magnitude of errors – we inspect the ACF to establish a better model:

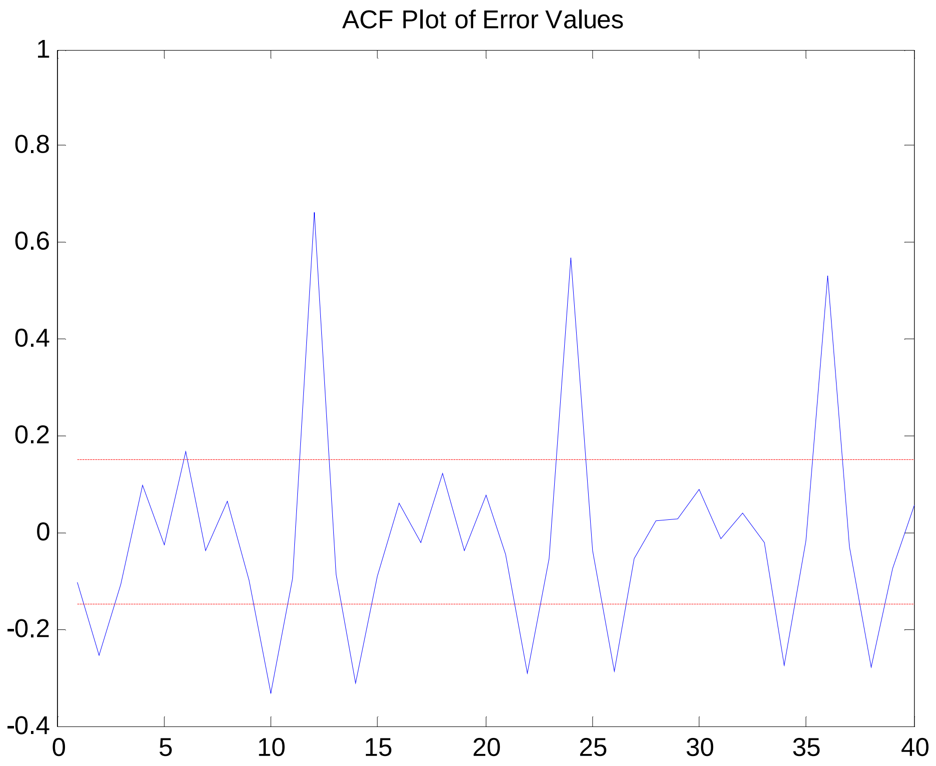


Fig 3.26: ACF of errors

After numerous attempts with different forms of models it was decided that this series could not be modelled with any degree of accuracy using these methods.

Note other models attempted:

AR model: lags 1,6,12

AR model: lags 1,2,3,6,12,36

AR model: lags 1,12.

After concluding the above series couldn't be modelled adequately a new series was looked at it which was the differences of the log(IIP);

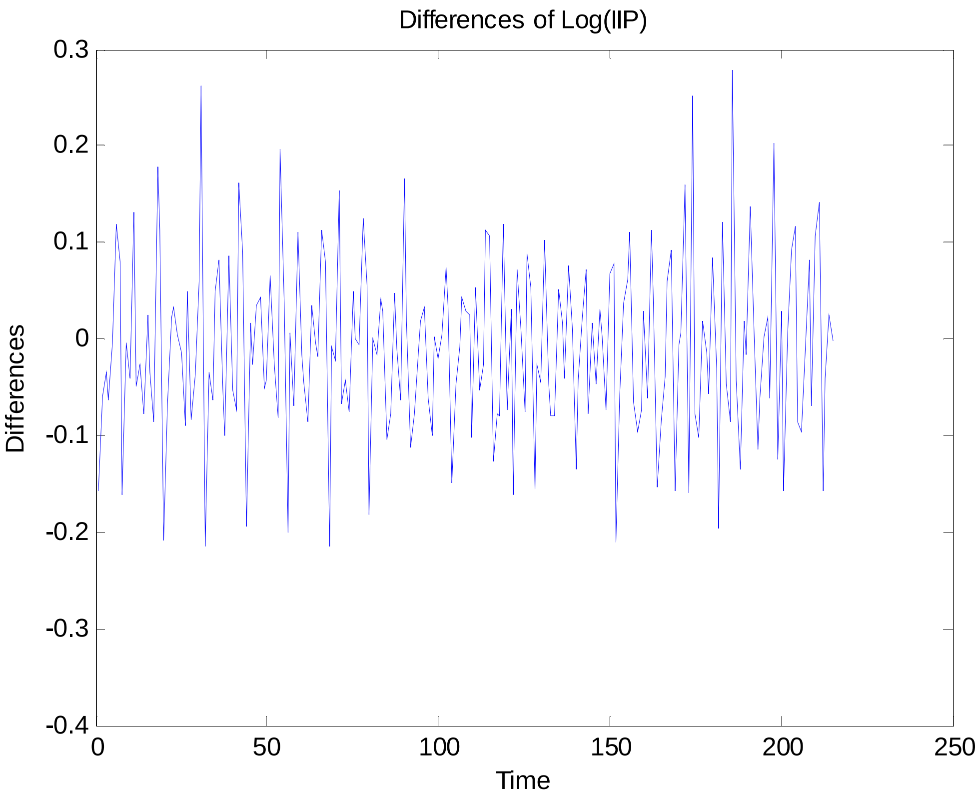


Fig 3.27: Plot of difference of the Log(IIP)

The ACF & PACF of this series are then examined to try and establish a model:

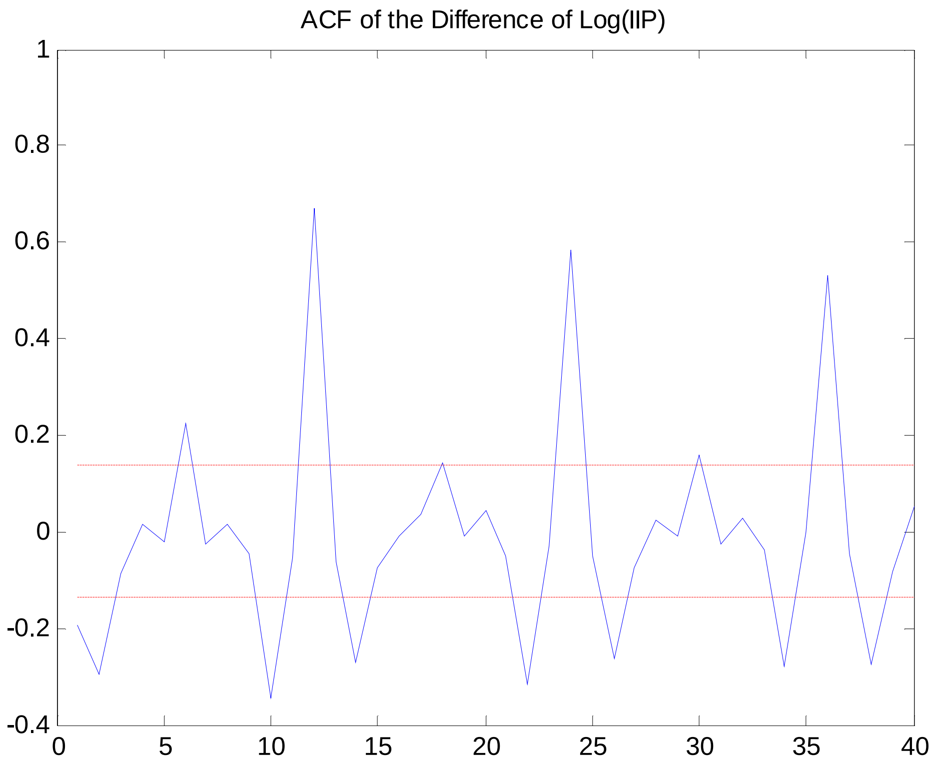


Fig 3.28: ACF of differences of Log(IIP)

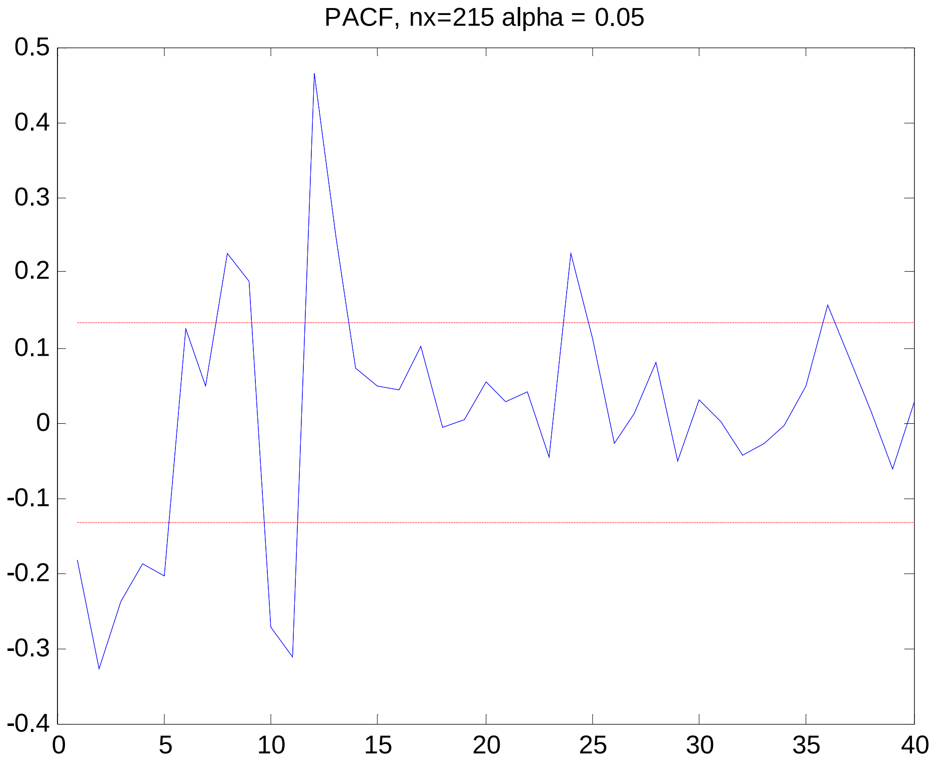


Fig 3.29: PACF of differences of Log(IIP)

There appears to be reason to use an AR model with lags at 1, 6 and 12. Firstly we will look at an AR(1) model:

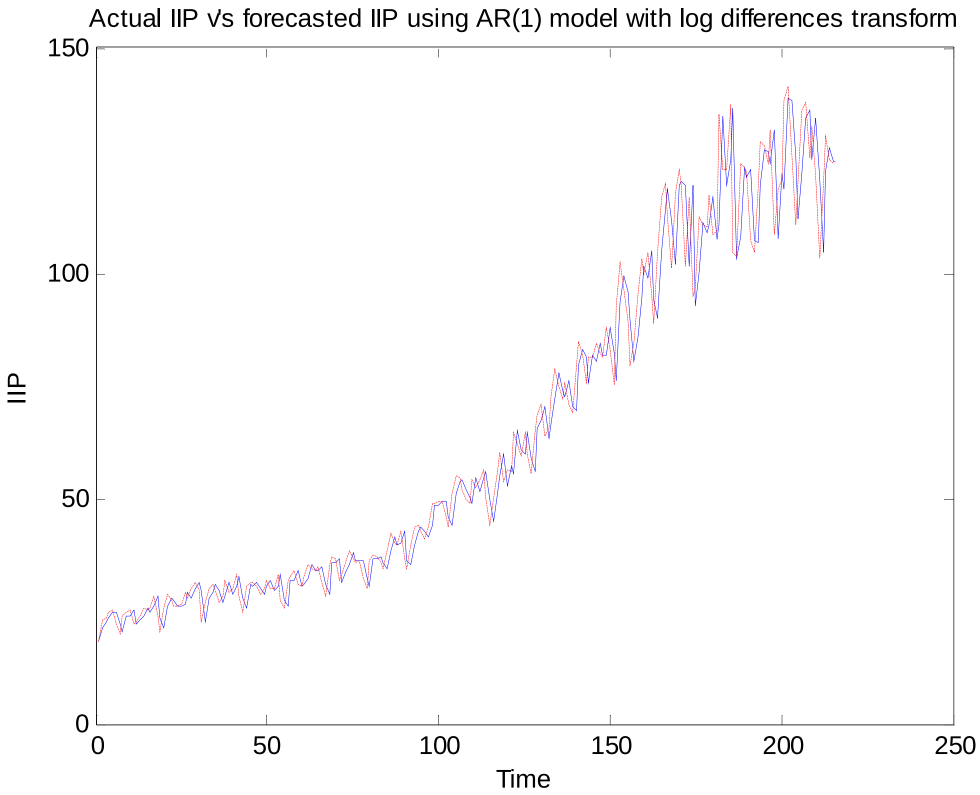


Fig 3.30: Plot of actual IIP versus forecasted IIP using an AR(1) model on transformed data using log differences.

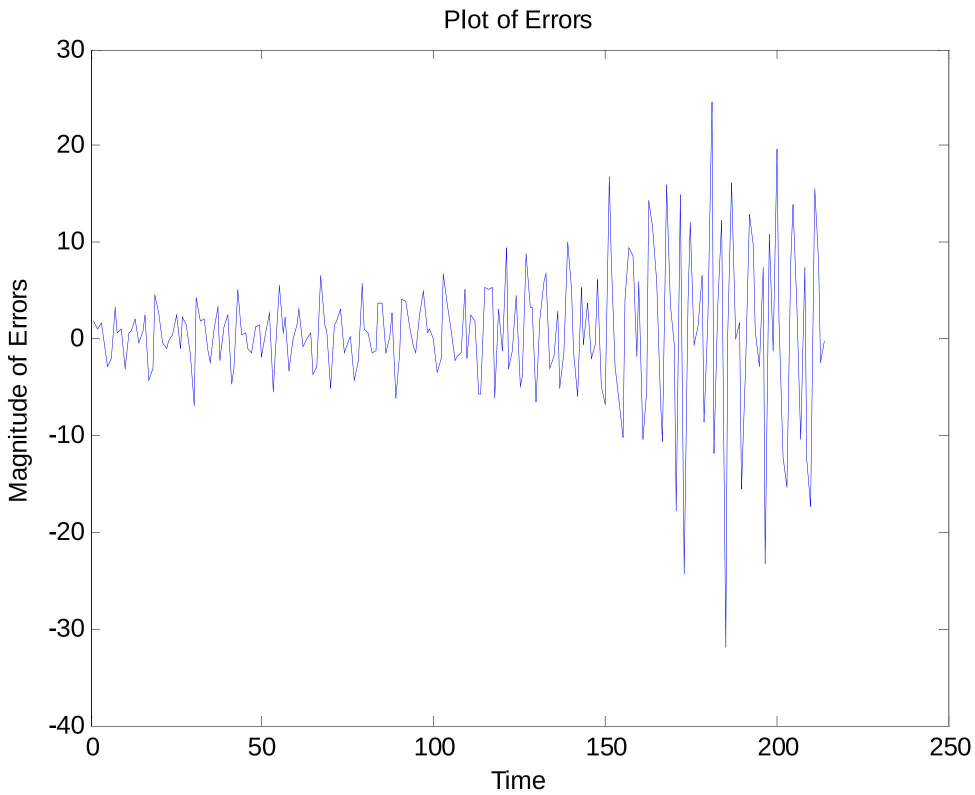


Fig 3.31: Plot of the error between actual IIP & forecasted IIP using AR(1) model on transformed data.

Now the AR model with lags at 1, 6 & 12 is examined;

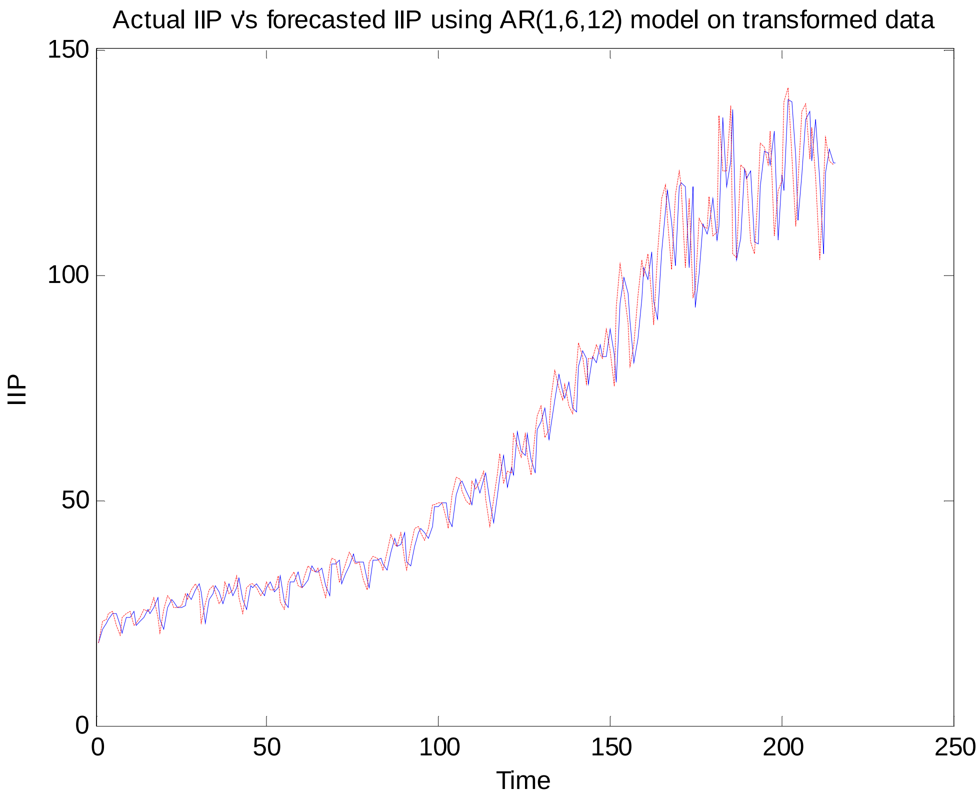


Fig 3.32: Plot of actual IIP versus forecasted IIP based on AR(1,6,12) model on the transformed data

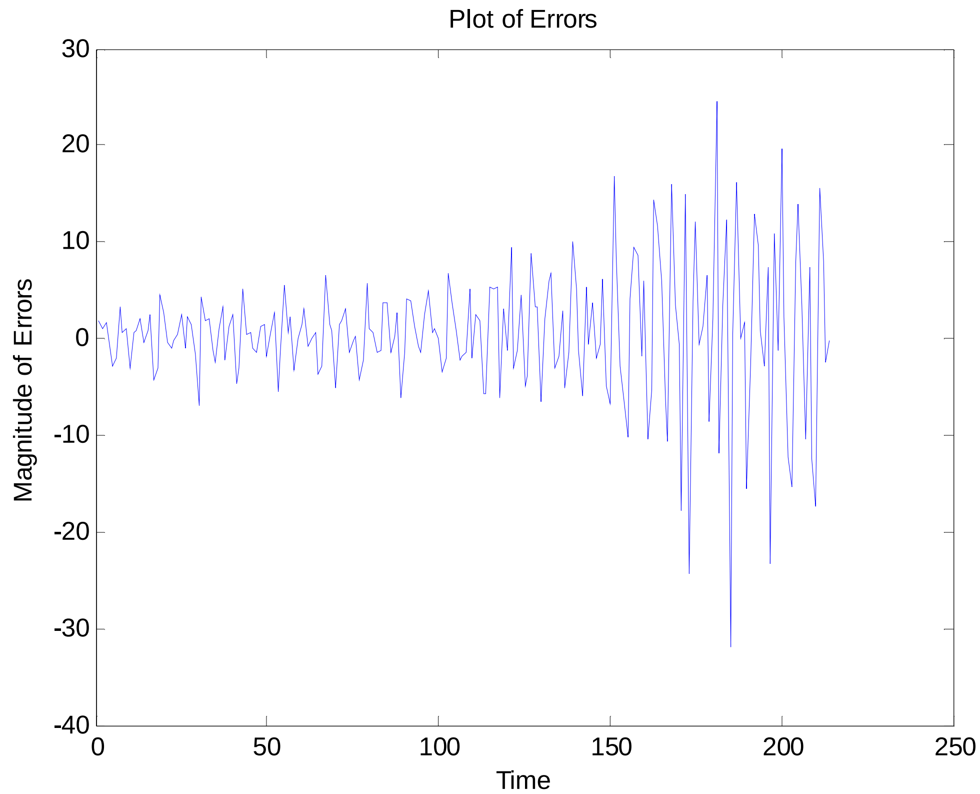


Fig 3.33: Plot of the error between actual IIP & forecasted IIP using AR(1,6,12) model on transformed data.

Again the results from each of these techniques are discussed in section 3.3.7.

3.3.2 Inverse transform of the series

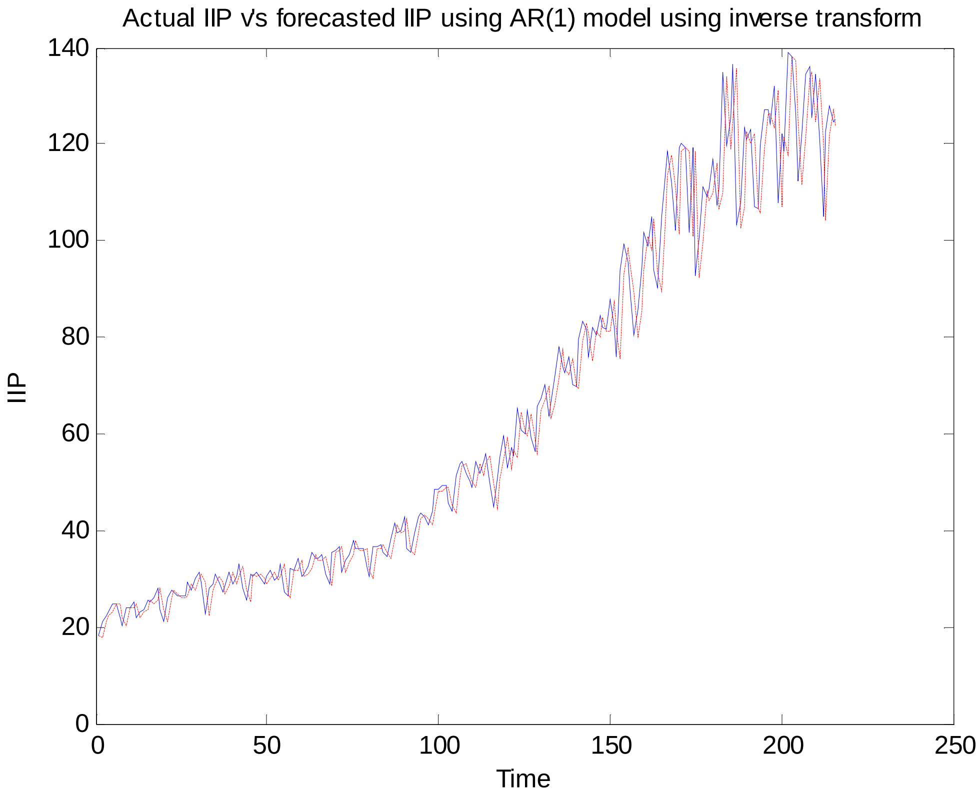


Fig 3.34: Plot of IIP versus series modelled using inverse transform

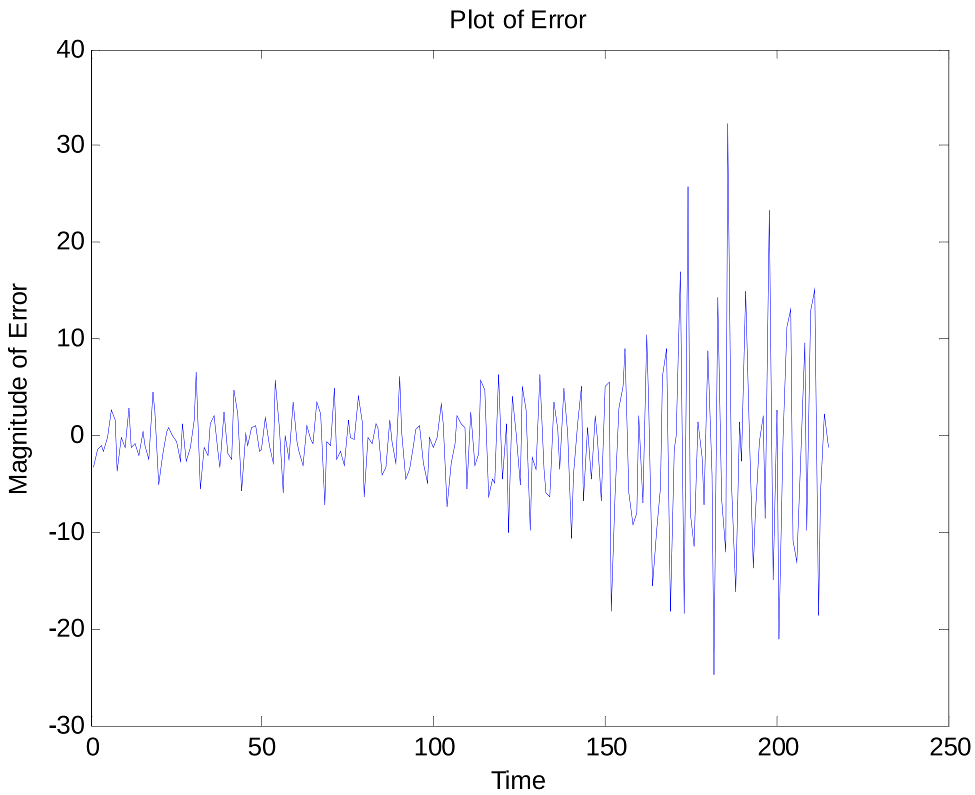


Fig 3.35: Plot of the error between actual IIP & forecasted IIP using AR(1) model on transformed data.

Again the results will be discussed in section 3.3.3 below.

3.3.3 Box-Cox Transforms Results

Model	**Transform**	**Predictive SSE**			**Predictive MSE**		
		Training	Validation	Test	Training	Validation	Test
AR(1)	Not applicable	1415.7	2915.6	6760.4	10.974	66.264	160.96
AR(1)	Y^2	1459	3005.2	6831.7	11.31	68.3	162.66
AR(1)	$\sqrt{Y}$	1398	2878.6	6736	10.837	65.422	160.38
AR(1)	$Log_e Y$	1391.5	2902.9	6747	10.787	65.976	160.64
AR(1)	1/Y	1363.3	2804.3	6702.6	10.568	63.734	159.59
AR(1)	$1/\sqrt{Y}$	1372	2823.2	6708.4	10.635	64.163	159.72
AR(1)	Log Differences	5699	3285.2	5699	10.114	74.664	3285.2
AR(1,6,12)	Log Differences	5699	3285.2	5699	10.114	74.664	3285.2

Table 3.4

From the above table of results, it appears that the model which best fits the data is the one calculated after the inverse transform is taken i.e. 1/Y and the AR(1) model.

However what is noteworthy about the results is the sharp increase in the SSE & PMSE for the test and validation sets as compared with the training sets. The reason for this we suspect is that the test and validation sets are taken from later on in the data. As discussed at the start of this section the data was divided into 3 sets, the training set being the first 130 points, the test set the next 44 points and the validation set the final 42 points. As the series progresses the time series is larger and as such the errors are larger. To investigate this further we calculate the prediction mean absolute percentage error (PMAPE) which is defined as

$$PMAPE = \frac{|error(t)|}{iip(t)} x100 = \frac{|e(t)|}{Y(t)} x100 \qquad (3.18)$$

The average of PMAPE for the training, validation and test sets are outlined in the table below:

Model	**Transform**	**PMAPE**		
		Training	Validation	Test
AR(1)	Not applicable	7.3867	7.5939	8.3058
AR(1)	Y^2	7.5015	7.6597	8.4364
AR(1)	$\sqrt{Y}$	7.3348	7.5734	8.2455
AR(1)	$Log_e Y$	7.2994	7.5803	8.4364
AR(1)	1/Y	7.2149	7.5387	8.1006
AR(1)	$1/\sqrt{Y}$	7.2477	7.5456	8.1415
AR(1)	Log Differences	7.1363	7.5115	7.2149
AR(1,6,12)	Log Differences	7.1363	7.5115	7.7721

Table 3.5

As can be seen from the above table the purported cause of the large variation in the PSSE & PMSE is true and the PMAPE does not differ significantly over the three data sets. On the basis of the above results we proceed with the model developed on the basis of the log transform due the combination of the PSSE, PMSE and PMAPE results.

3.4 ERROR ANALYSIS OF CHOSEN MODEL

We now calculate the Ljung-Box Q statistic to test for model adequacy. The above AR(1) model based on the log transform of the raw data has been fit, we now wish to test the errors between the fitted model and the actual transformed data to see if there is a departure from randomness of the errors. We are testing the null hypothesis that the model fit is adequate, i.e. there is no serial correlation at the corresponding element of lags. The results are as follows;

Error	**Lags**	**H**	**P-Value**	**Q-Statistic**	**Critical Value**
Model Error	15	1	0	134.75	22.307
Model Error	20	1	0	143.35	28.412
Model Error	25	1	0	210.05	34.382
Training Set Error	15	1	0	123.07	22.307
Training Set Error	20	1	0	128.71	28.412
Training Set Error	25	1	0	192.83	34.382
Test Set Error	15	1	0.0278	27.114	22.307
Test Set Error	20	1	0.0905	28.864	28.412
Test Set Error	25	1	0.0552	37.205	34.382
Validation Set Error	15	1	0.0175	28.718	22.307
Validation Set Error	20	1	0.0369	32.645	28.412
Validation Set Error	25	1	0.0212	41.322	34.382

Table 3.6

Notes to Table 3.6

- H = 0 indicates acceptance of the null hypothesis, H=1 rejects the null hypothesis.
- Significance level of all tests = 10%
- P-values (significance levels) at which the null hypothesis of no serial correlation at each lag in Lags is rejected.
- Critical value is of the Chi-square distribution for comparison with the corresponding element of Qstat.

As can be seen from the above table the null hypothesis has been rejected at every level. So for the moment we will accept that the errors from the model are serially correlated.

We now test the errors for ARCH effects. This test will test for the null hypothesis that the time series of residuals (errors) is i.i.d. Gaussian i.e. homoskedastistic. The results for the residuals from the fitted model are as follows;

Lags	Significance Level	H	P-Value	ARCH-stat	Critical Value
15	5%	1	$7.5968\text{x}10^{-9}$	68.703	24.996

Table 3.7

To summarise we are now in the following situation;

- We have transformed the original series Y to X=ln(Y)
- Modelled this using the following AR(1) model:

$$X(k) = 0.99685X(k-1) + e(k) \qquad (3.19)$$

- We have tested the model errors for serial correlation and it appears that e(k) is serially correlated.
- Furthermore we have tested the model errors for ARCH effects and we have found that they exhibit heteroskedastiscity.

3.5 UNIT ROOT PROCESS

At this stage we suspect that the transformed series X(k) is a unit root process I(1), that is the true model of the data is of the form:

$$X(k) = X(k-1)+e(k). \qquad (3.20)$$

Please see Appendix 7.2 for further details of this. We will use one of the Phillips-Perron tests for unit roots of which there are four cases. The case which applies here is the test done under the assumption that the true process is of the form:

$$X(k) = \alpha + X(k-1)+e(k),\ \alpha \text{ any value} \qquad (3.21)$$

and the modelled process is of the form:

$$X(k) = \hat{\alpha} + \hat{\rho}X(k-1) + \delta(k) + e(k) \qquad (3.22)$$

i.e. there is a drift in the modelled process.

We have confirmed from the Ljung-Box test and the ARCH test that the errors of the chosen model are serially correlated and heteroskedatistic and the Phillips-Perron tests allow for these situations and deal with them in testing for the unit root. As such we calculate the Philips-Perron ρ test and the Phillips-Perron t test statistics.

$$\rho - test = T(\hat{\rho} - 1) - \frac{1}{2}(T^2.\hat{\sigma}^2_{\hat{\rho}} / s^2)(\hat{\lambda}^2 - \hat{\gamma}_0) \qquad (3.23)$$

$$t - test = (\hat{\gamma}_0 / \hat{\lambda}^2)^{0.5} t - \{0.5(\hat{\lambda}^2 - \hat{\gamma}_0)(T.\hat{\sigma}_{\hat{\rho}} / s) / \hat{\lambda}\} \qquad (3.24)$$

where; T = sample size

$\hat{\rho}$ = autoregressive coefficient

$\hat{\sigma}_{\hat{\rho}}^2$=standard error of the autoregressive coefficient

s^2=standard deviation of the error (e(k))

$\hat{\lambda}^2 = \hat{\gamma}_0 + 2\{0.8(\hat{\gamma}_1) + 0.6(\hat{\gamma}_2) + 0.4(\hat{\gamma}_3) - 0.2(\hat{\gamma}_4)\}$

$\hat{\gamma}_i$= i[th] autocovariance

t = ($\hat{\rho}$ -1)/ $\hat{\sigma}_{\hat{\rho}}^2$

The following results were obtained:

Phillips-Perron	Value	Significance Level	Unit Root
ρ test	-0.16168	>-21	Yes
t test	-1.7318	>-3.44	Yes

Table 3.8

4. VARIANCE FORECASTING WITH GARCH

4.1 MODELLING VARIANCE

When modelling using least squares it is assumed that the expected value of all error terms when squared is the same at any given point, this assumption is known as homoskedasticity. Data in which the variances of the error terms is not equal, i.e. the variance varies over time is said to be heteroskedastistic. In the presence of unequal variance and when using ordinary least squares regression, the regression coefficients remain unbiased, however the standard errors and confidence intervals estimated by conventional procedures will be too narrow giving a false sense of precision.

4.1.1 ARCH- Autoregressive conditional heteroskedasticity

The ARCH-model was first introduced by Robert Engle in 1982. First consider an ordinary AR(p) model of the stochastic process y_t.

$$y_t = c + \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + u_t \qquad (4.1)$$

Where u_t is white noise. The basic AR(p)-model is now extended so that the conditional variance of u_t could change over time. One extension is that u_t^2 itself follows an AR(m)-process

$$u_t^2 = \theta_0 + \theta_1 u_{t-1}^2 + ... + \theta_m u_{t-m}^2 + w_t \qquad (4.2)$$

where w_t is a new white noise process and u_t is the error in forecasting y_t. This is the general ARCH(m)-process. (Engle 1982). For ease of calculations and for estimation, a stronger assumption about the process is added.

$$u_t = \sqrt{h_t} v_t \qquad (4.3)$$

where v_t is an i.i.d. Gaussian process with zero mean and a variance equal to one, v_t~N(0,1) and the whole model variance is now

$$\varepsilon_t \mid \psi_{t-1} \sim N(0, h_t)$$
$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \qquad (4.4)$$

where α_0>0, α_i>0, i=1,…..q and ψ_{t-1} is the information available at time t-1. Finally, when the process for modelling the variance is defined, we use an additional equation for modelling y_t.

$$y_t = c + \varepsilon_t \qquad (4.5)$$

this means that ε_t is innovations from a linear regression.

4.1.2 Generalised ARCH - GARCH

This section describes the generalisation of the ordinary ARCH model. Bollerslev introduced the GARCH(p,q) model in 1986. Since then a number of developments on the basic modelled have been invented. The existing models can be divided into two categories: *symmetric* and *asymmetric* models.

However the basic GARCH model introduced by Bollerslev is as follows and is comparable to the extension of an AR(p) model to an ARMA(p,q) model. Formally the GARCH process is written as

$$\varepsilon_t \mid \psi_{t-1} \sim N(0, h_t)$$
$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \qquad (4.6)$$

where p,q integers, α_0>0, $\alpha_i \geq 0$, i=1,….q, $\beta \geq 0$, i=1,……p. This additional feature is that the process now also includes lagged h_{t-I} values. For p=0 the process is an ARCH(q) process. For p=q=0 (an extension allowing q=0 if p=0), ε_t is white noise. (Bollerslev 1986).

Theorem 4.1: The GARCH(p,q) process as defined in equation B.23 is wide sense stationary with E(ε_t)=0, var(ε_t)=α_0(1-A(1)-B(1))$^{-1}$ and cov($\varepsilon_t \varepsilon_s$)=0 for t≠s if and only if A(1)+B(1)<0.

Proof: See Bollerslev (1986) page 323

In theorem 4.1 A(1)= $\sum_{i=1}^{q} \alpha_i$ and B(1)= $\sum_{i=1}^{p} \beta_i$. In most cases the number of parameters is rather small, e.g. GARCH(1,1) which is the focus of the next section.

4.1.3 GARCH(1,1)

In the case where p=q=1 the model becomes

$$\begin{aligned}\varepsilon_t \mid \psi_{t-1} &\sim N(0, h_t) \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}\end{aligned} \tag{4.7}$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, $\alpha_1 + \beta_1 < 1$

4.2 CRITERION OF FIT

A good performance measure of the second moment can be hard to find as variance is not directly observable. One way of dealing with this is not to rely on one measure but rather a number of measures. For the purpose of this thesis three different measures are used for evaluating the performance of variance forecasts from different GARCH models:

1. Mean Square Error (MSE)
2. Mean Absolute Error (MAE)
3. Adjusted Mean Absolute Percentage Error (AMAPE)

The MSE is calculated as follows

$$MSE = \frac{1}{h+1} \sum_{t=s}^{s+h} (\hat{\sigma}_t^2 - \sigma_t^2)^2 \tag{4.8}$$

where h = number of steps ahead we wish to predict (the case here is normally 1), $\hat{\sigma}_t^2$ = forecast volatility, σ_t^2 = "true" volatility (ε_t^2) and S = sample size.

The MAE is:

$$MAE = \frac{1}{h+1} \sum_{t=s}^{s+h} \left| \hat{\sigma}_t^2 - \sigma_t^2 \right| \tag{4.9}$$

The AMAPE is:

$$AMAPE = \frac{1}{h+1} \sum_{t=s}^{s+h} \left| \frac{\hat{\sigma}_t^2 - \sigma_t^2}{\hat{\sigma}_t^2 + \sigma_t^2} \right| \tag{4.10}$$

4.3 MODELLING THE VARIANCE OF IIP

In section 3 we modelled the underlying process of the IIP data. We found that by transforming the original series by taking the log transform we could model the series quite well on the basis of an AR(1) model with the true process shown to be unit root.

Now that we have proved that the process is unit root we can take the first difference of the transformed series and by definition of a unit root process this differenced series is stationary. In order to model the variance using the software package available (Matlab, GARCH Toolbox) we must use a mean stationary time series which we now have.

This software allows the series to be input together with a specification for the conditional mean model. The package will model the series for both conditional mean and conditional variance, returning suggested coefficients for each model. The log differenced series we now use was investigated in section 3.3 and we discovered it can be modelled quite well using an AR(1) model. As such we specify an AR(1) model as the model for our conditional mean when using the GARCH toolbox.

Please note that from here on in the transformed differenced series is referred to as y(k).

We now move on to discuss the types of GARCH modelling to be undertaken:

Case I: Conditional mean model specified as AR(1), allows the software package to produce estimates of the conditional mean and conditional variance.

Cases II involves an external variable not previously discussed. As discussed in the introduction IIP is similar to a monthly measure of GDP and numerous economic variables are combined to produce the IIP measure. One of these is electricity demand, we now use the time series of electricity demand for the same period of the IIP time series and build an ARMAX model to include this exogenous variable.

Case II: Conditional mean model specified, we specify the underlying ARX model and allow the software to estimate the coefficients of the conditional mean and conditional variance models.

4.4 RESULTS

Case I: Conditional Mean Model Specified as AR(1)

In this case we specified that an AR(1) model be fitted to the data. As with the method outlined in section 3 the above is modelled on the training set which is the first 130 points of the data. Various GARCH specification models with different orders were run on the data and it appears that GARCH(1,1) model for the purpose of the software fits the data best and returns an ARCH(1) model.

The following model was returned for the conditional mean of the input series:

$$y(k)=-0.01388-0.21385y(k-1))+\varepsilon(k) \quad (4.11)$$

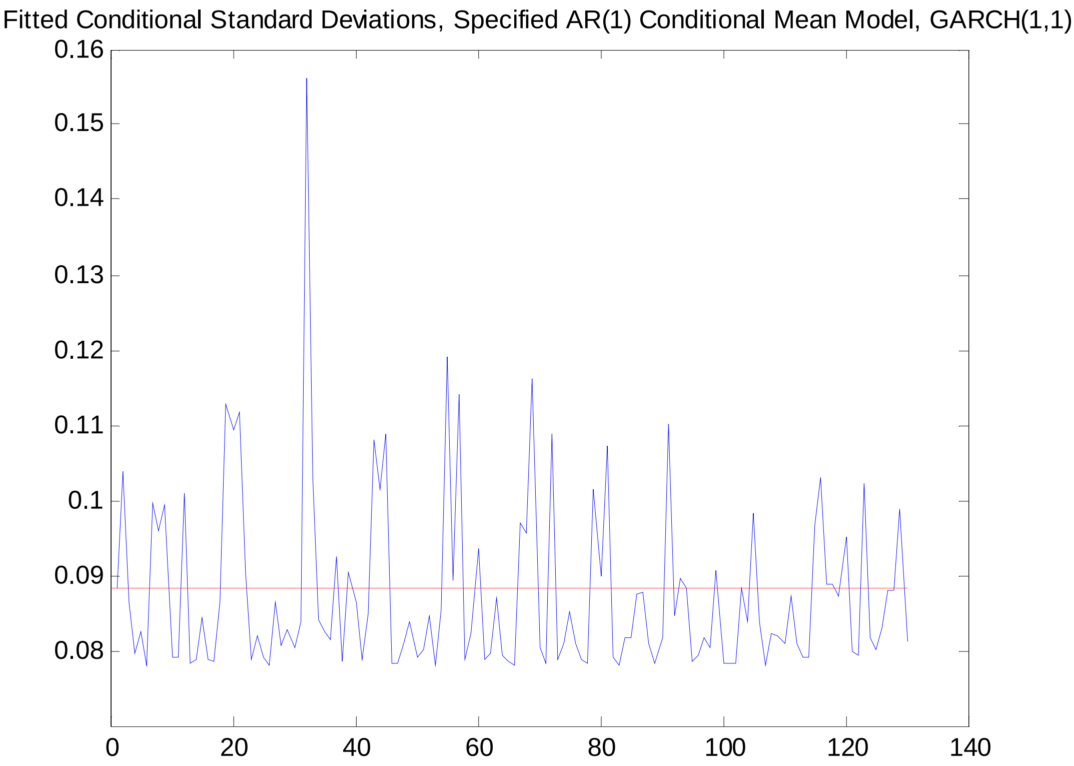


Fig 4.1: Fitted conditional variance versus unconditional variance

Case II: Conditional Mean Model Specified as ARX(1,1)

We now attempt to model the IIP data using an exogenous variable i.e. we wish to use an ARX(1,1) model to model the conditional mean. The variable used in electricity demand in the same time period. Various GARCH specification models with different orders were run on the data and it appears that GARCH(?,?) model for the purpose of the software fits the data best.

The following model was returned for the conditional mean of the input series:

$$y(k)=0.015923-0.21095y(k-1) -0.00028681x(k)+\varepsilon(k) \qquad (4.12)$$

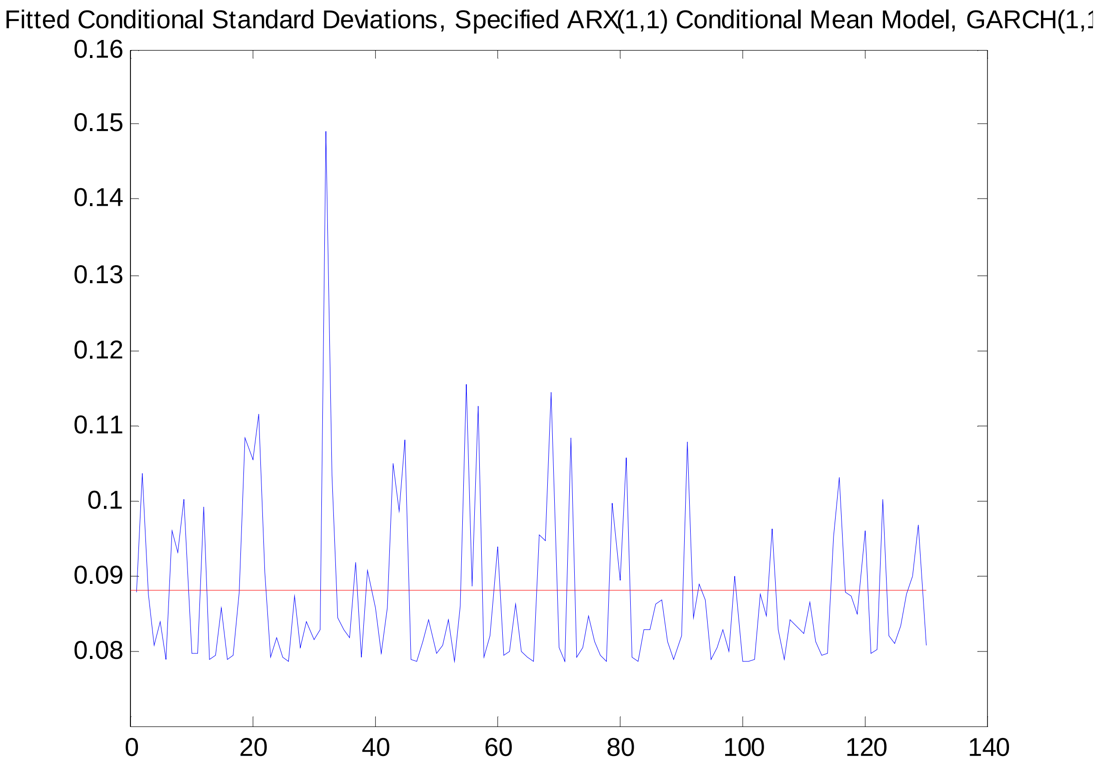


Fig 4.2: Fitted conditional variance versus unconditional variance

GARCH(P,Q)	**Case I; (1,1)**			**Case II; (1,1)**		
GARCH coefficients	α_0	α_1	β_1	α_0	α_1	β_1
	0.0061042	0.21809	0	0.006179	0.20093	0
Standard Error	0.0041519	0.2267	0.58213	0.0046124	0.22029	0.6398
T Statistic	1.4702	0.9620	0	1.3221	0.9121	0
LLF	133.22			133.45		
MSE	131.1			138.24		
MAE	619.6			643.92		
AMAPE	56.224			56.155		
Conditional Mean Model Measures						
Data Set	Training	Validation	Test	Training	Validation	Test
MSE	**9.6105**	**72.583**	**139.17**	9.9904	73.303	139.17
SSE	**1249.4**	**3193.7**	**5547.7**	1298.7	3225.3	5566.9
PMAPE	**6.8881**	**7.5263**	**7.4071**	7.2923	7.5292	7.4253
R_c^2	**0.9901**			**0.98998**		

Table 4.1

Notes to table 4.1:

1. The T-Statistic at the 5% significance level for the sample size involved is relevant for T<1.658.
2. Log Likelihood function (LLF): relates to the use of Maximum Likelihood Estimation (MLE) in the calculation of the parameters for the GARCH model, this is discussed in Appendix B.3.6.
3. The Mean Square Error (MSE), Mean Absolute Error (MAE) & Adjusted Mean Absolute Percentage Error (AMAPE) are discussed above in section 4.2. These measures relate to the variance and are used in [14]. The use of these measures is questionable as they are a measure of how well we are forecasting the variance which is taken to be the $error^2$, which is drawn from a certain distribution. We are trying to measure how well we can forecast this distribution.
4. R_c^2 is the centred sample multiple correlation coefficient: The fit of an ordinary least squares model is described by the sample multiple correlation coefficient – R^2. For R_c^2 , when the regression includes a constant term as it does in all the above cases the value must fall between 0 & 1.

$$R_c^2 = \frac{y'X(X'X)^{-1}X'y - T\bar{y}^2}{y'y - T\bar{y}^2} \qquad (4.13)$$

where y=the modelled time series, X the matrix of regression variable and T = size of the sample.

5. CONCLUSION

The time series of IIP was investigated on a number of levels, via Ordinary Least Squares (OLS), Weighted Least Squares (WLS) and GARCH modelling. It can be concluded that the method of GARCH modelling performed best.

The raw series was modelled first via OLS, and the model did not perform well. The errors from the model were investigated and heteroskedasticity was found to be an issue. As such we modelled the errors on a time basis (fitting linear, quadratic & exponential models), and on a level basis (Section 3.2.4) to carry out a WLS solution. None of the WLS solutions appeared to deal with the issue of heteroskedasticity.

As such a number of transforms of the series were investigated and modelled via OLS (Section 3.3). The log transform of the IIP series was hypothesised to be a unit root process and proved as such (Section 3.5). Thus the first difference of this series is a stationary time series. In other words we have found a method of converting the IIP time series to a stationary series which can be modelled.

The series of the log differences was investigated and found to be an AR(1) process. The software used to model the variance and conditional mean (Matlab, GARCH Toolbox) required a stationary time series to model and also allowed the specification of the conditional mean model.

The table of results in section 4.4 shows the two cases as discussed. From the table we can see that Case I, which specifies the conditional mean model as AR(1) performs best with the R_c^2 measure of 0.9901 and acceptable measures of MSE, SSE and PMAPE for the three different data sets.

It was thought that by extending the AR(1) model to an ARX(1) model by including electricity demand as an exogenous variable may further improve the fit i.e. reduce the error. As can be seen from the results in the table this was not the case, however it did not significantly reduce the fit. The R_c^2 in this case being 0.98998 and very similar results being returned for MSE, SSE and PMAPE.

In conclusion of the three methods used to model the time series in question the GARCH modelling produced the best model, the reason for this being the way in which heteroskedasticity is dealt with.

In the method of WLS the time based and level based methods were used to estimate the variance. As the series grows with time the WLS method using the time based models assumes that the variance grows with time, similarly the level based model does the same. GARCH modelling allows the variance to be modelled similar to an ARMA model and as such allows for heteroskedasticity. Empirical evidence seems to show that volatility clustering which is the only effect that GARCH takes into account that the other methods don't allows the GARCH model to perform better in forecasting our time series.

Further research could be undertaken by performing a WLS calculation on the first difference of the log transformed series, which would allow for better comparison between the GARCH modelling techniques and the WLS methods in forecasting this time series.

APPENDICES

APPENDIX A

A.1 WEIGHTED LEAST SQUARES SOLUTIONS FIGURES

In case 3 in this section we used the exponential fitted model of the error squared as our estimate of σ_t^2. The following are the figures produced from those calculations.

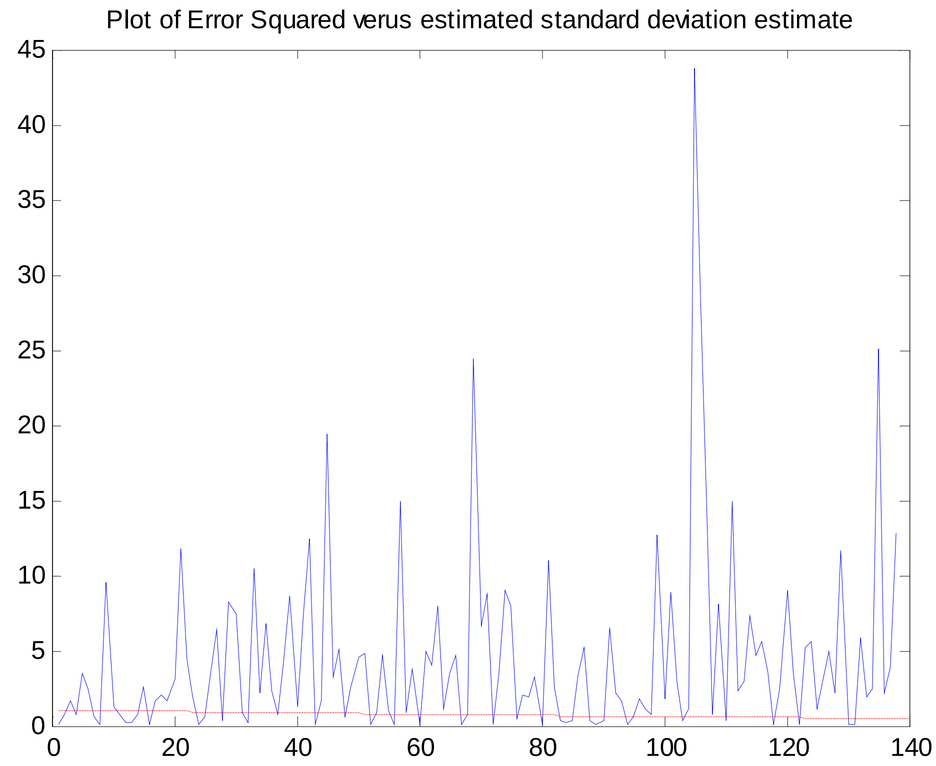


Fig A.1: Plot of error squared versus estimated standard deviation

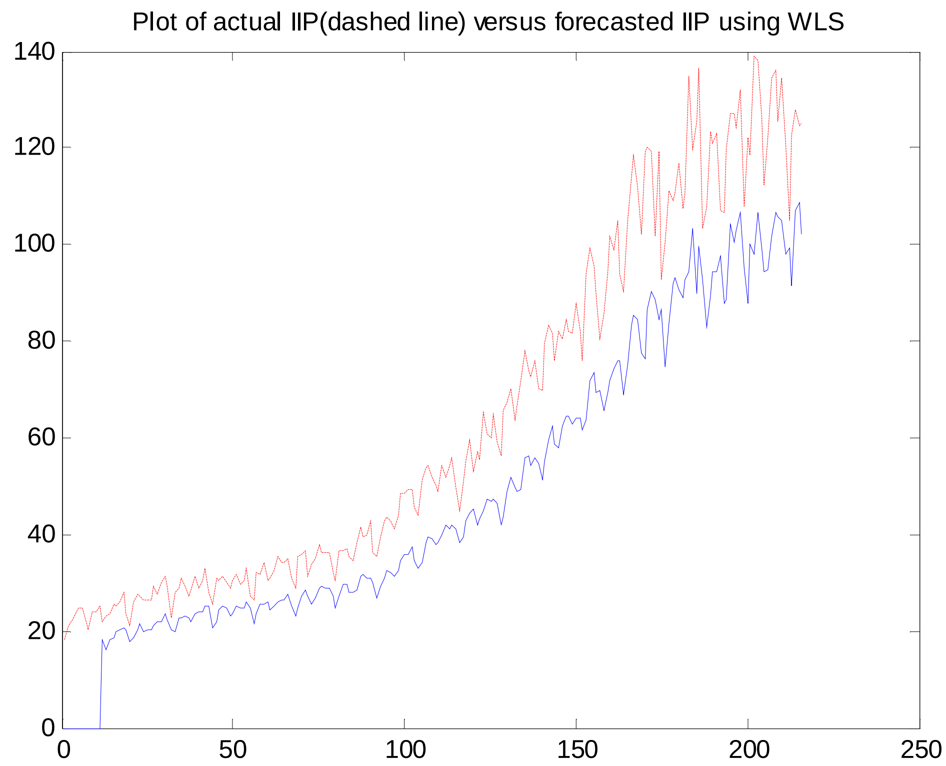


Fig A.2: Plot of actual IIP versus forecasted IIP

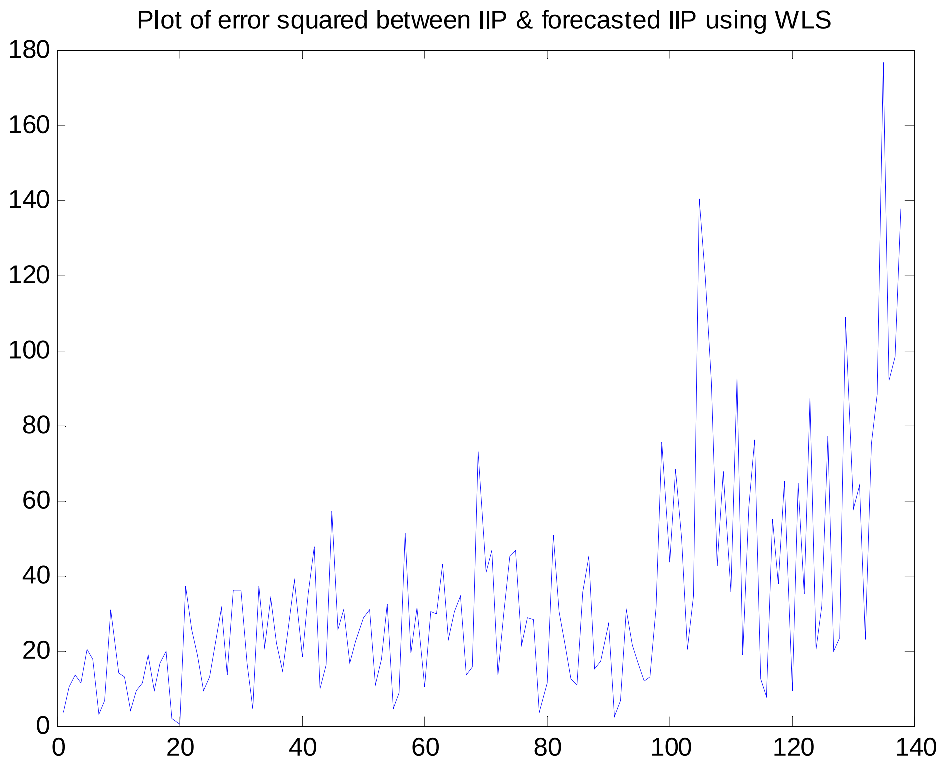


Fig A.3: Plot of error squared

In the final case of modelling using WLS we used a model fitted on actual IIP to fit the error squared, the following are the figures from the results:

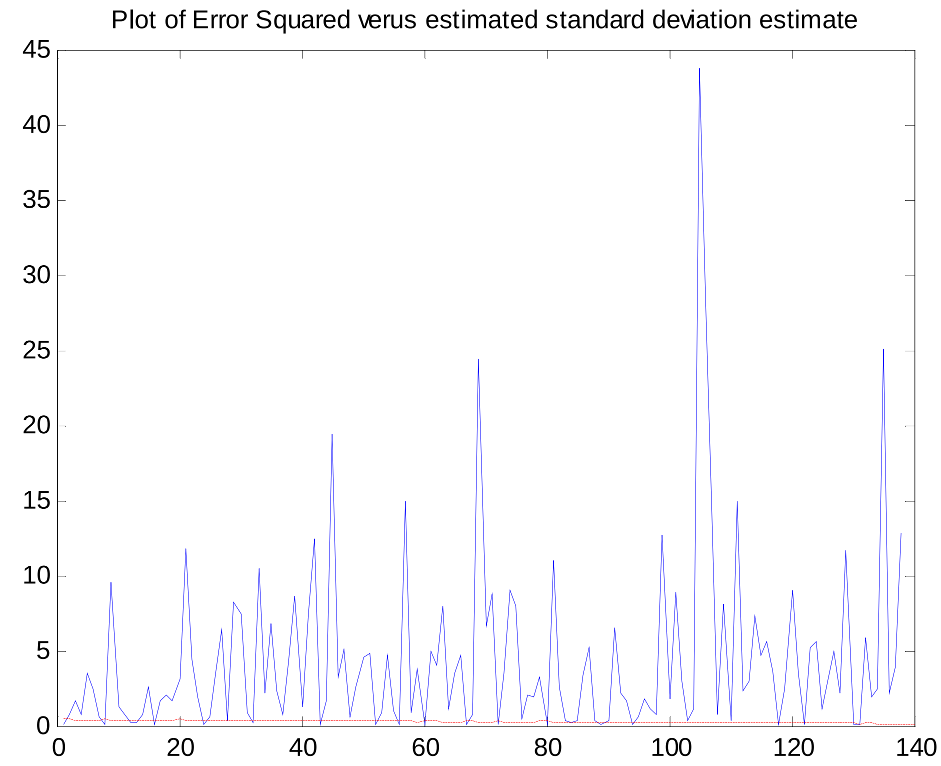


Fig A.4: Plot of error squared versus estimated standard deviation

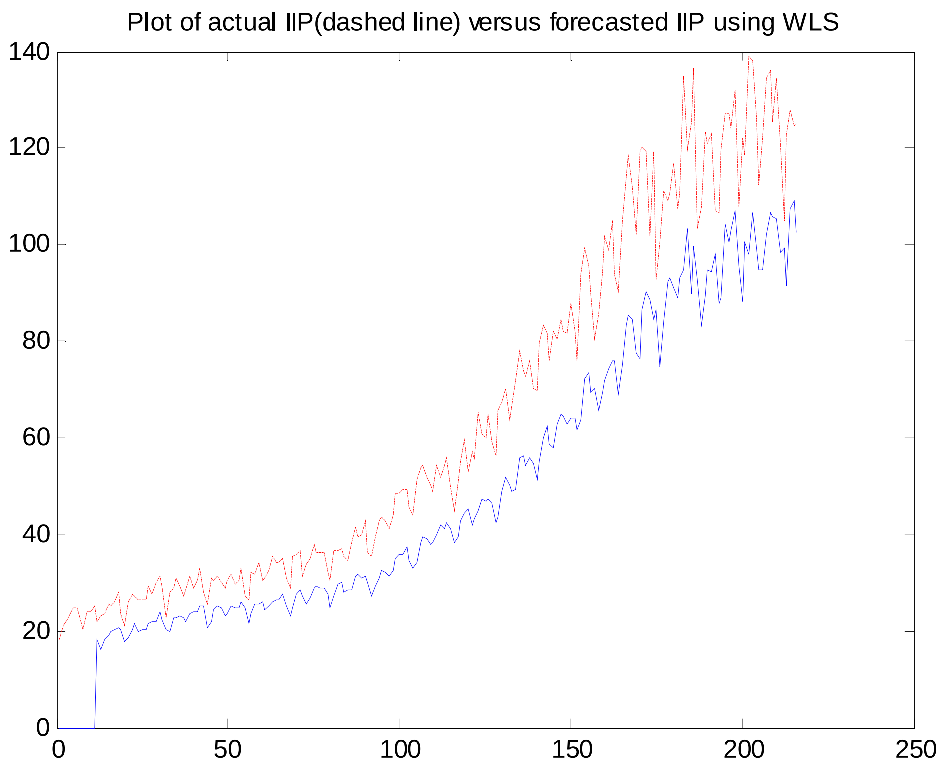


Fig A.5: Plot of actual IIP versus forecasted IIP

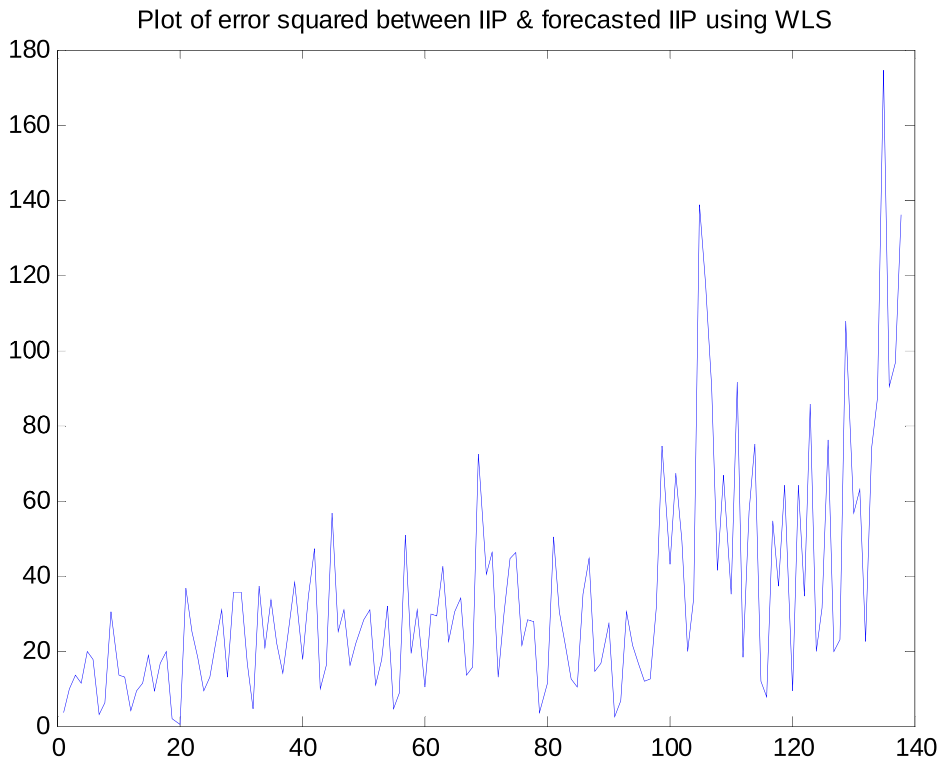


Fig A.6: Plot of error squared.

APPENDIX B

B.1 MODELLING TRANSFORMS

B.1.1 Raw Data

For comparison purposes an AR(1) model is examined using the raw data. It should be noted to construct this model and all others the data was divided into three sub-categories per standard modelling procedures:

Data Type	Number of Data Points	% of the data
Training Set	130	60
Test Set	44	20
Validation Set	42	20

The models were built on the basis of the training set, i.e. θ was calculated using the least squares method on this part of the data only. The model built on the training set was then used to forecast up to 216 steps ahead. These forecasts were then compared with the test and validation sets.

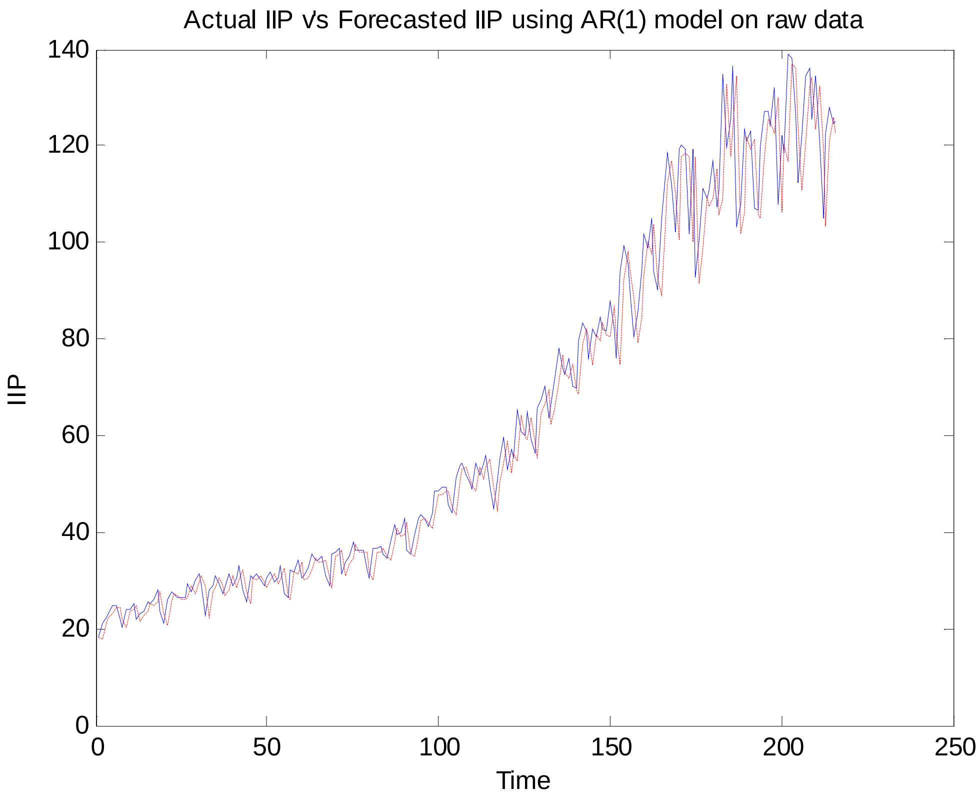


Fig B.1: Actual IIP versus forecasted IIP using an AR(1) model with the raw data

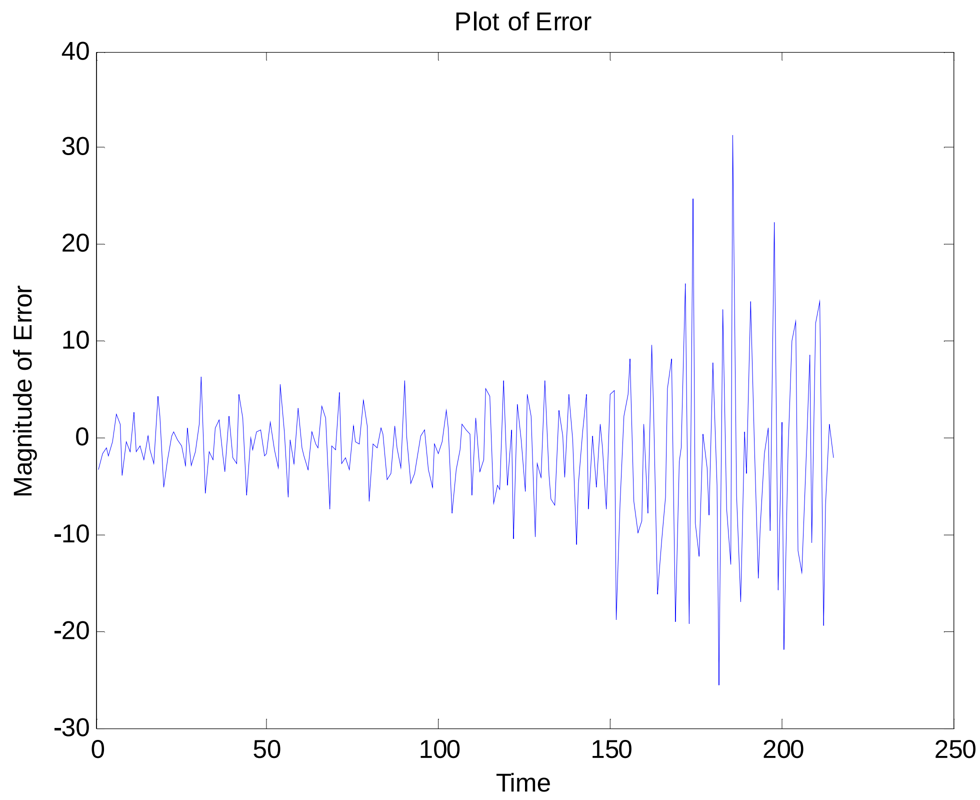


Fig B.2: Plot of the error between actual IIP & forecasted IIP using AR(1) model on raw data.

As can be seen from the plot the errors grow larger the further ahead we forecast which is to be expected as the model is built on the first 130 data points. The results such as SSE and MSE are discussed in more detail in section 3.3.3.

B.1.2 Modelling with the Squared Transform of the Data

Using the method outlined above in steps 1-5 the following was observed:

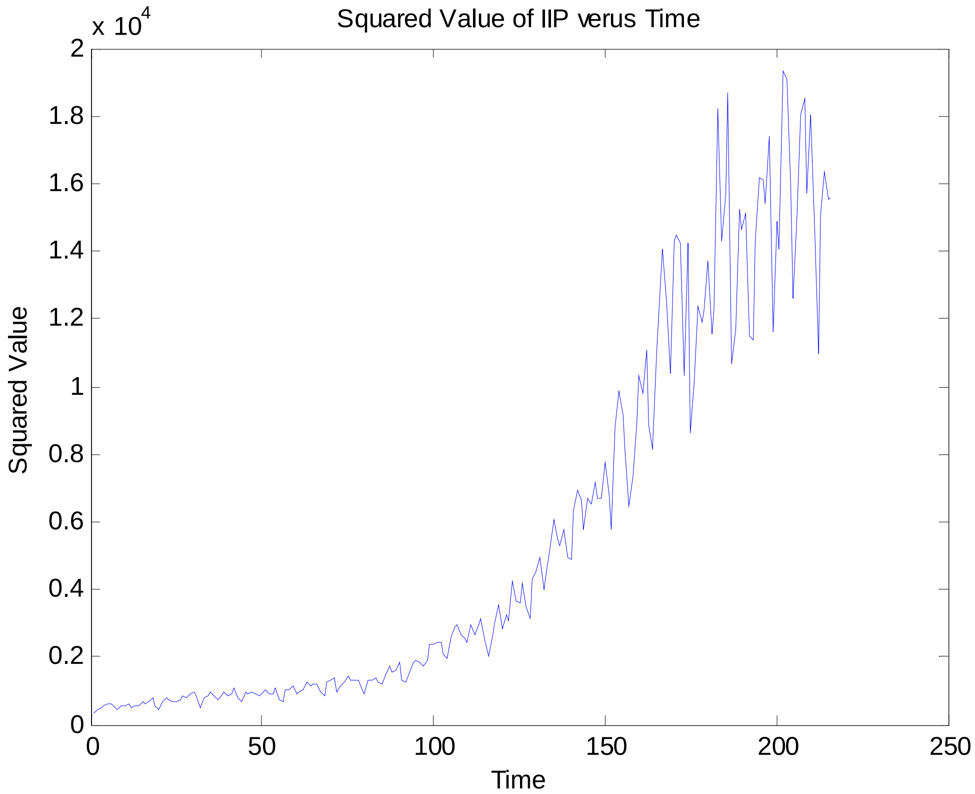


Fig B.3: Plot of Squared value of IIP time series

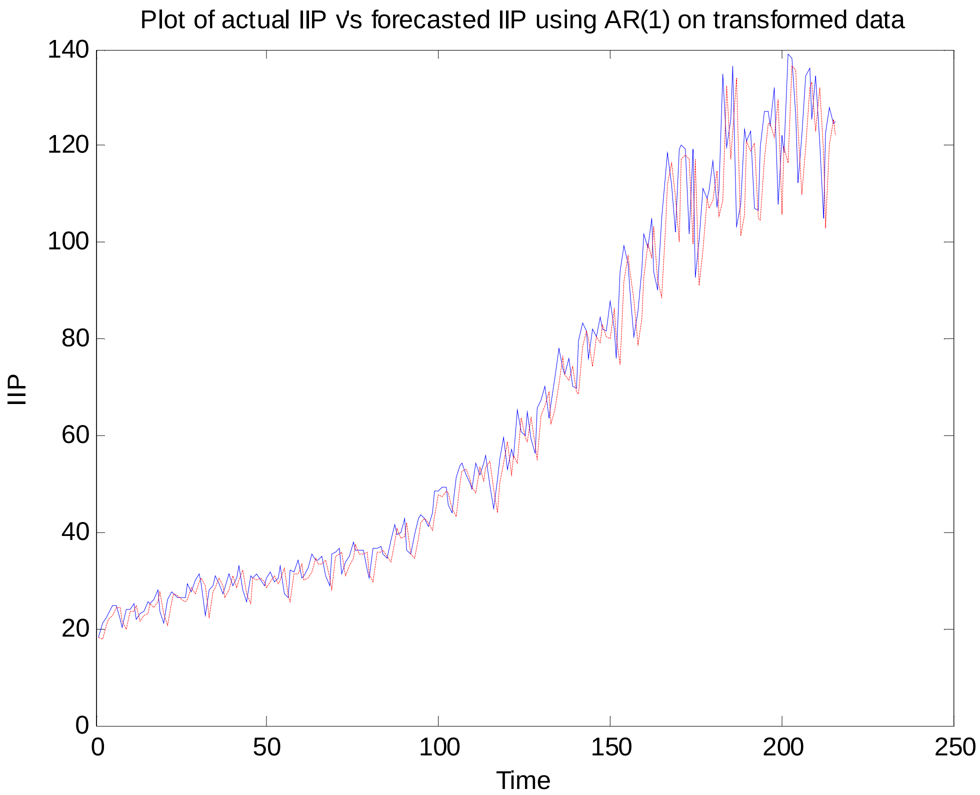


Fig B.4: Plot of IIP time series versus modelled series using squared transform

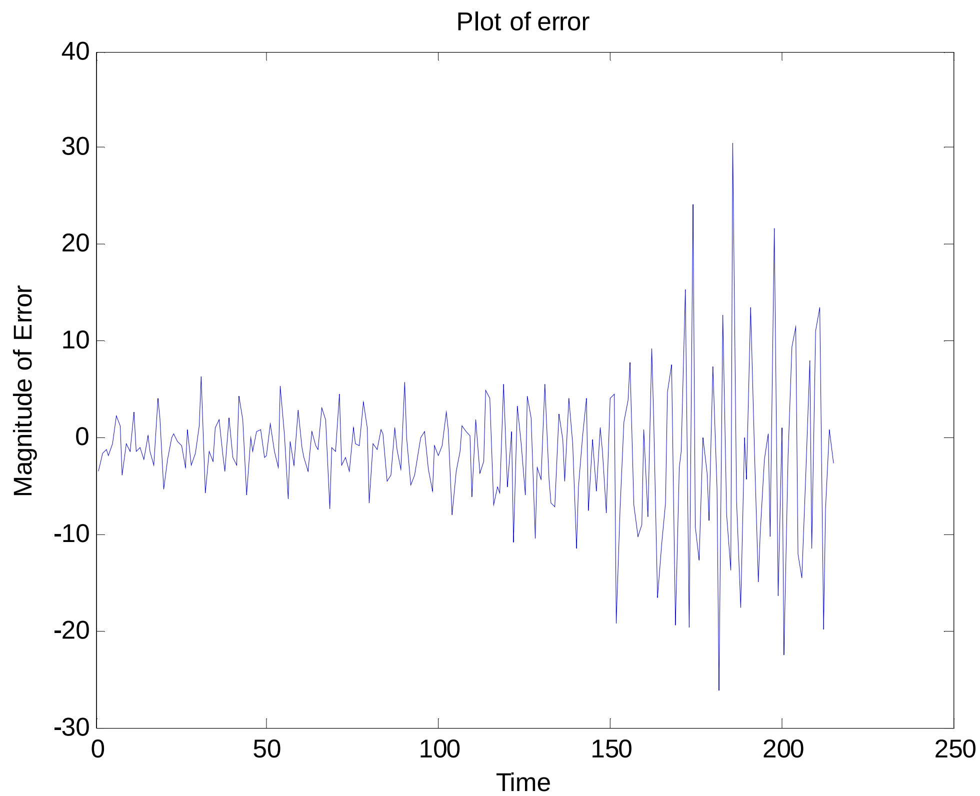


Fig B.5: Plot of the error between actual IIP & forecasted IIP using AR(1) model on transformed data.

The results from each of these techniques are discussed in section 3.3.3.

B.1.3 Modelling with the Square Root Transform of the Data

Using the method outlined above in steps 1-5 the following was observed:

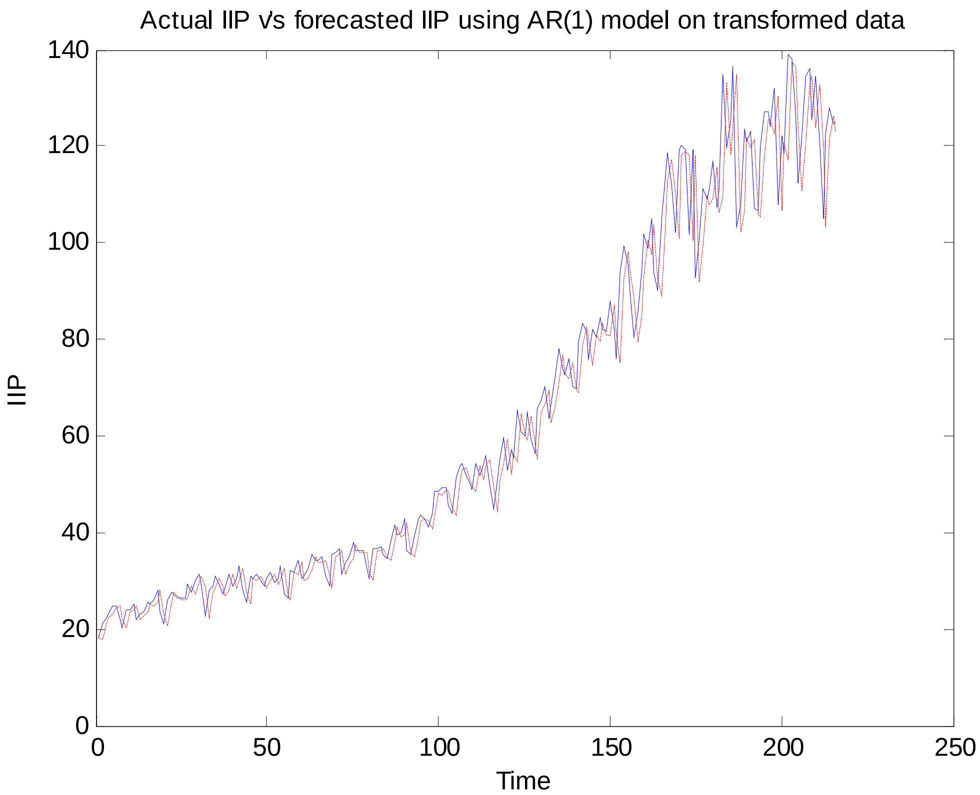


Fig B.6: Plot of IIP versus model obtained using square root transform

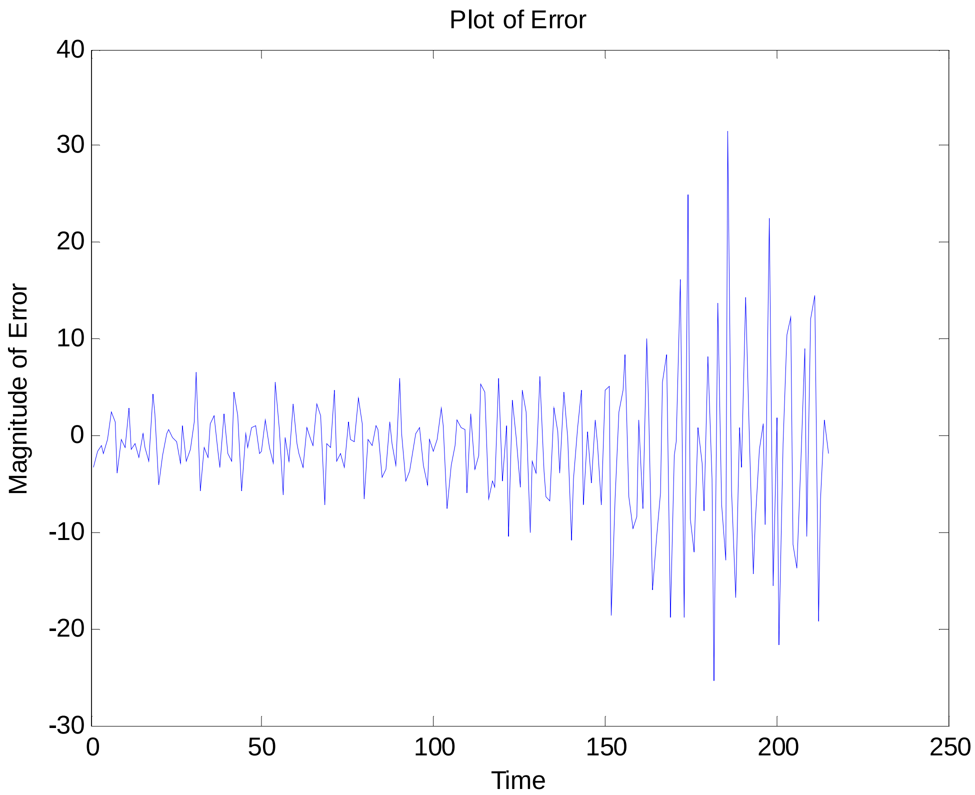


Fig B.7: Plot of the error between actual IIP & forecasted IIP using AR(1) model on transformed data.

The results from each of these techniques are discussed in section 3.3.3.

B.1.4 Inverse square root transform

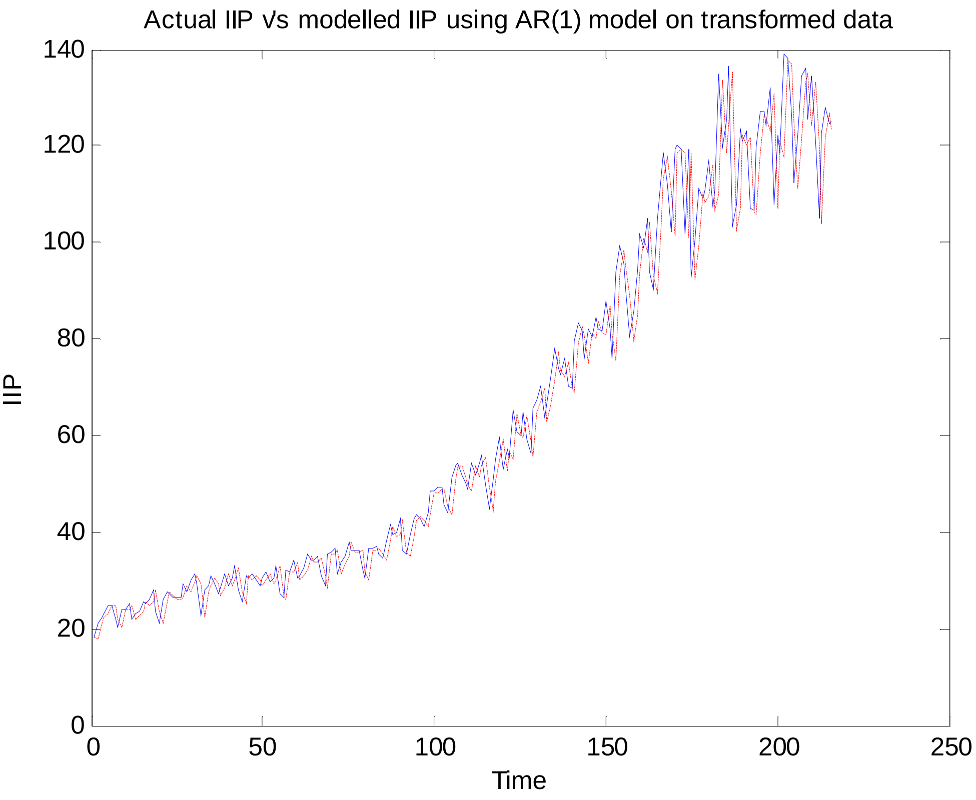


Fig B.8: Plot of IIP versus modelled series using inverse square root transform

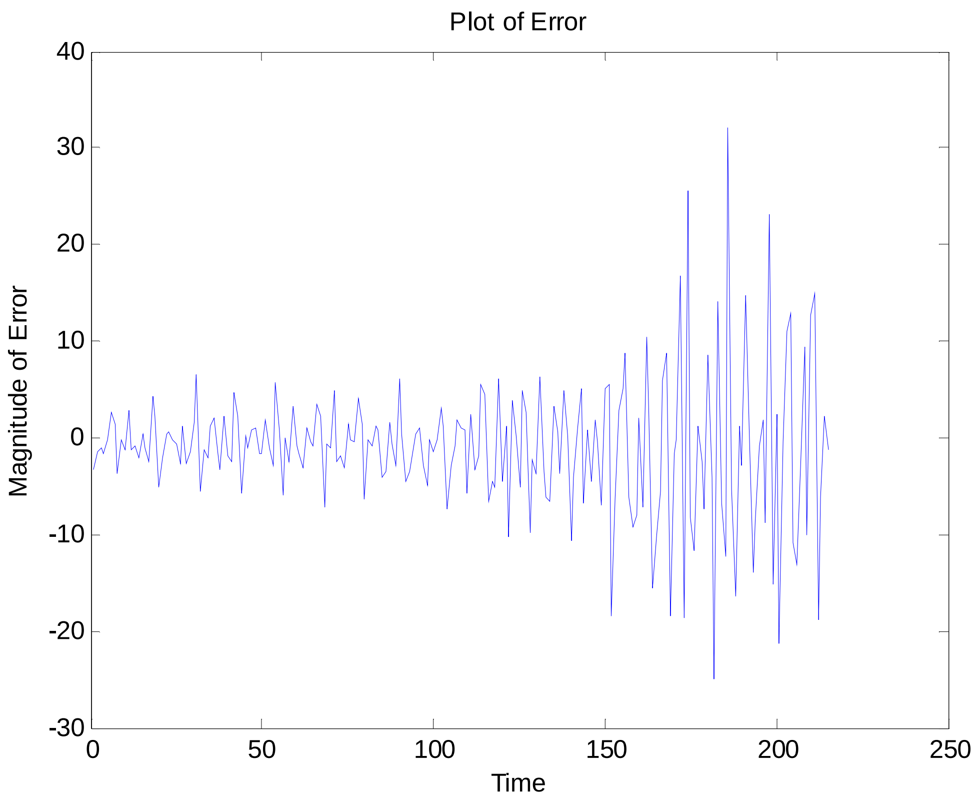


Fig B.9: Plot of the error between actual IIP & forecasted IIP using AR(1) model on transformed data.

Again, please see results section 3.3.3 above for discussion

APPENDIX C

C.1 STATISTICAL THEORY

C.1.1 Statistics

A brief description of the statistics used in the course of this thesis are presented below. In all cases below, X is a discrete valued stochastic variable, k is the summation index and $p_x(k)$ is the probability that X takes the value k. [12] Ch 3.

The first moment is the population mean and is defined as

$$E(X) = \sum_k k p_x(k) = \mu \qquad \text{(C.1)}$$

The non-central second moment is then defined as

$$E(X^2) = \sum_k k^2 p_x(k) = Var(X) + E^2(X) \qquad \text{(C.2)}$$

Non-central moments are then in the general case defined as

$$E(X^r) = \sum_k k^r p_x(k) \qquad r = 1,2,3\ldots \qquad \text{(C.3)}$$

Skewness is defined as

$$\frac{E((X-\mu)^3)}{(Var(X))^{3/2}} \qquad \text{(C.4)}$$

A variable with positive skewness is more likely to have values far above the mean value than far below. For a normal distribution the skewness is zero.

Kurtosis is defined as

$$\frac{E((X-\mu)^4)}{(Var(X))^2} \qquad \text{(C.5)}$$

For a normal distribution the kurtosis is 3. A distribution with a kurtosis greater than 3 has more probability mass in the tails, so called “fat tails” or leptokurtic.

C.1.2 Correlation

The population correlation between two different random variables X and Y is defined by

$$Corr(X,Y) \equiv \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \qquad \text{(C.6)}$$

[12] Page 743

C.1.3 Autocorrelation

The jth autocorrelation is defined as the jth autocovariance divided by the variance:

$$Corr(X_t, X_{t-j}) \equiv \frac{Cov(X_t, X_{t-j})}{\sqrt{Var(X_t)Var(X_{t-j})}} \qquad \text{(C.7)}$$

[12] Page 49

C.1.4 Sample autocorrelation function

The autocorrelation function (ACF) is a plot of the auto correlations of a time series versus the lag at which the correlations are calculated. The ACF is the correlation function using an average operator:

$$\hat{r}_x(\tau) = \frac{\sum_{k=1}^{n-\tau}(x(k)-\bar{x})(x(k-\tau)-\bar{x})}{\sum_{k=1}^{n}(x(k)-\bar{x})^2} \qquad \text{(C.8)}$$

where $\hat{r}_x(\tau)$ is the ACF value for a lag of τ, n is the number of observations used and $\bar{x}$ is the average value of x(k).

The ACF is useful to determine the order of the lag in a moving average process.

C.1.5 Partial autocorrelation function

The partial autocorrelation function (PACF) is defined as the last coefficient in a linear projection of Y on the m most recent values. Letting this be denoted $\alpha_m^{(m)}$, if the constant for the process is zero then the equation becomes:

$$Y_{t+1} = \alpha_1^{(m)}Y_t + \alpha_2^{(m)}Y_{t-1} + \ldots.. + \alpha_m^{(m)}Y_{t-m+1} \qquad \text{(C.9)}$$

If the process were a true AR(p) process the coefficients with lags greater than m would be zero. [12]

C.2 HYPOTHESIS TESTING

A hypothesis test is a procedure for analysing data to address the question of whether a certain criterion is fulfilled or not. This can be tested in a number of different ways and this section presents the hypothesis tests that will be used in this thesis.

All tests have a corresponding p-value. This p-value under the assumption of the null hypothesis, is the probability of observing the given sample result.

At a significance level of 95%, which in our notation corresponds with a critical value of 0.05, then we reject the null hypothesis if the p-value is lower than this critical value. A p-value greater than 0.05 corresponds to insufficient evidence for rejecting the null hypothesis. (Mathworks 2002).

C.2.1 Ljung-Box Test

The Ljung-Box test is performed to test whether a series has significant autocorrelation or not. The Ljung-Box Q-Statistic is a lack of fit hypothesis test for model misspecification. The Lbq-value is calculated by:

$$Q_k = T(T+2)\sum_{i=1}^{k}\frac{r_i^2}{T-i} \qquad \text{(C.10)}$$

where T is the number of samples, k is the number of lags and r_i the i^{th} autocorrelation. If Q_k is large then the probability that the process has uncorrelated data decreases. The null hypothesis for the test is that there exists no correlation and under that hypothesis, Q_k is χ^2 with k degrees of freedom.

C.2.2 Levene's Test

Levene's Test is used to see what functional form if any equalises the variance throughout a time series. The test is performed on the residuals of a chosen model. The residuals are plotted and observed and divided in groups depending on how many areas of differing variance appear. The t-test statistic is then obtained as:

$$t^* = \frac{\sum_{i=1}^{\tau}(d_i - \bar{d})^2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}+\ldots+\frac{1}{n_\tau}}} \qquad \text{(C.11)}$$

where $\bar{d}_k = \sum \frac{|e_{ik} - \tilde{e}_k|}{n_k}$. Here $\tilde{e}_k$ is the sample median of the residuals for group k and

$$s = \sqrt{\frac{\sum(d_{i1}-\bar{d}_1)^2 + \sum(d_{i2}-\bar{d}_2)^2 + \ldots + \sum(d_{i\tau}-\bar{d}_\tau)^2}{n-\tau}}.$$

Although the distribution of the absolute value of the residuals may not be normal it has been shown that the t-statistic approximately follows the t-distribution with a large enough sample size. In order to make the test more robust against departures from normality we use the absolute deviations of the residuals from the sample median. Large absolute values of t* indicate that the error terms do not have constant variance.

C.2.3 ARCH Test

It is relatively simple to test whether the residuals from a regression have conditional heteroskedasticity or not. The test is based on ordinary least squares (OLS) regression, where the OLS residuals $\hat{u}_t$ from the regression are saved. $\hat{u}_t^2$ is thereafter regressed on a constant and it own m-lagged values. This is done for all samples t = 1,2,…T. This regression has a corresponding R^2-value. TR^2 is then asymptotically χ^2-distributed with m degrees of freedom under the null hypothesis that $\hat{u}_t$ is i.i.d. $N(0,\sigma^2)$. (Engle 1982)

The ARCH-test can also be performed as a test for GARCH-effects. The ARCH-test for a lag (p+q) is locally equivalent to a test for GARCH effects with lags (p,q). (Mathworks 2002).

The null hypothesis, H_0, is that no ARCH effects exist. This is tested for lags up to T.

C.3 STOCHASTIC PROCESS

A stochastic process is a system which evolves in time according to probabilistic equations, that is, the behaviour of the system is determined by one or more time-dependent random variables.

C.3.1 White Noise

One of the basic building blocks when modelling stochastic processes is the white noise process which is a sequence $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ whose elements have zero mean and variance σ^2, that is:

$$E(\varepsilon_t) = 0$$
$$E(\varepsilon_t \varepsilon_s) = \begin{cases} \sigma^2, t = s \\ 0, t \neq s \end{cases} \quad \text{(C.12)}$$

C.3.2 Stationarity

A process Y_t is said to be covariance-stationary or weakly stationary if the neither the mean u_t nor the autocovariance γ_{jt} depend on the time t and if the given moment exists.

$$\begin{aligned} &E(Y_t) = \mu, \forall t \\ &E((Y_t - \mu)(Y_{t-j} - \mu)) = \gamma_j, \forall t, any"\ j" \end{aligned} \quad \text{(C.13)}$$

C.3.3 Moving Average Process

A moving average process of order one MA(1) is described as

$$Y_t = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \varepsilon_t \quad \text{(C.14)}$$

where α_0 and α_1 could be any real constants, and ε_t is a white noise process described in equation 3.12. In the general MA(q) instance the equation becomes

$$Y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j} + \varepsilon_t \qquad (C.15)$$

C.3.4 Autoregressive Process

The "autoregressive" property in principle means that old events leave waves behind a certain time after the actual time of action, i.e. the process depends on its past. Firstly consider the AR(1) process

$$Y_t = \alpha_0 + \beta_1 Y_{t-1} + \varepsilon_t \qquad (C.16)$$

Where α_0 and β_1 can be any real constants, and ε_t is a white noise process described in equation 3.12. An AR(1) process can also be generalised to an AR(p) process

$$Y_t = \alpha_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} + \varepsilon_t \qquad (C.17)$$

When deciding which lags i.e. p & q in an MA and AR process the autocorrelation and the partial autocorrelation functions, as described in sections 3.1.3 and 3.1.4 above. The table below indicates when to use which type of model, this table is available at www.itl.nist.gov 2002.

Shape of ACF	**Indicated Model**
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero.	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, remainder essentially zero	Moving average model. Order identified by where plot becomes zero.
Decay, starting after a few lags	Mixed autoregressive and moving average model.
All zero or close to zero	Data is essentially random
High values at fixed intervals	Include seasonal autoregressive term
No decay to zero	Series is not stationary

C.3.5 Least Square Estimation

Least squares estimation is a well-known and much used technique for calculating model parameters. The method involves fitting a model to the data in question and then minimising the squared error, the error representing the deviation of each point in the model from the regression line.

Given a model of the form $Y = \theta X + \varepsilon$ we wish to solve this for θ, X and Y are known. Θ is given by $\theta = (X^T X)^{-1} X^T Y$.

This technique is used throughout this thesis to estimate the parameters of AR models – see section 4.

C.3.6 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is another well-known technique for estimating model parameters. MLE can be used for any type of model for which least squares estimation cannot be used.

It is based on the probable distributions of the parameters given the data set at hand. This data comes from a data generating process with a probability distribution function f(x(1),x(2),…..; θ), i.e. the data is generated by a probability distribution function that is conditioned on the parameters of the data generator model. MLE seeks and estimator which given the data x(1), x(2),… will give an estimate of θ, which should be unbiased, efficient and consistent.

The approach is to invert f(x(1),x(2),…..; θ), and look for some g(θ ; x(1),x(2),…..). For example given a set of data points we seek to maximise the probability that this set will occur. The values of θ that maximises this value are the most likely values of θ in the data generator: $\theta=\max L(\theta;X)$, where L is the likelihood function. MLE returns the value of θ under which the data set is most likely to occur. To calculate the MLE we require the probability distribution function the data generator.

Finally the MLE may not exist and if it does it may not be unique, and there is no closed form solution for MLE a search algorithm must be used.

MLE is used in the MATLAB algorithm which estimates the GARCH parameters.

C.4 UNIT ROOTS

A unit root process is a case where a forecasting model is of the form y(t) = y(t-1) + e(t) i.e. the co-efficient of the regressor equals 1. This type of model is often denoted I(1) – i.e. integrated of order 1.

Many forecasting methods used on time series involve the following;

- the unconditional expectation of the variable is a constant independent of time: $E(y_t) = \mu$.
- As one tries to forecast farther into the future the forecast converges to this mean: $\lim_{s \to \infty} y_{t+s/t} = \mu$

These are not appealing assumptions when forecasting economic variables as generally an underlying trend can be identified which would like to take into account when forecasting into the future e.g. there is an underlying constant increase in GDP over time and when forecasting into the future we would like to take account of this and not have forecasted GDP equal to the mean of past GDP i.e. we would like it to be higher.

One way to avoid the occurrence of the above is to use a unit root process:

$(1-L)y(t) = \delta + \psi(L)\varepsilon(t), where\, \psi(1) \neq 0$

The mean of (1-L)y(t) is denoted by δ & (1-L) = Δ

$$\Rightarrow \Delta y(t) = y(t) - y(t-1) \qquad \text{(C.18)}$$

If we set Ψ(L)=1 get

$$\Delta y(t) = \delta + \varepsilon(t) \Rightarrow y(t) = y(t-1) + \delta + \varepsilon(t) \qquad \text{(C.19)}$$

- this process is known as a random walk with drift delta.

Note that $$\psi(z) = 1 + \psi_1 z^1 + \psi_2 z^2 + ... \qquad \text{(C.20)}$$

It is convenient to work with the following representation of the unit root process;

$$y(t) = \alpha + \delta(t) + u(t) \qquad \text{(C.21)}$$

where u(t) follows a zero mean Autoregressive Moving Average (ARMA) process:

$$(1 - \phi_1 L - \phi_2 L^2 - .... - \phi_p L^p)u(t) = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)\varepsilon(t) \qquad \text{(C.22)}$$

where the moving average operator $(1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)$ is invertible.

We can factorise the autoregressive operator $(1-\phi_1 L-\phi_2 L^2-........-\phi_p L^p)$ as follows;

$$(1-\phi_1 L-\phi_2 L^2-........-\phi_p L^p)=(1-\lambda_1 L)(1-\lambda_2 L).......(1-\lambda_p L) \qquad \text{(C.23)}$$

If all the eigenvalues $\lambda_1, \lambda_2, ...\ \lambda_p$ are inside the unit circle then (5) can be expressed as

$$u(t)=\frac{1+\theta_1 L+\theta_2 L^2+......+\theta_q L^q}{(1-\lambda_1 L)(1-\lambda_2 L)......(1-\lambda_p L)}\varepsilon(t)\equiv\psi(L)\varepsilon(t) \qquad \text{(C.24)}$$

with $\sum_{j=0}^{\infty}\left|\psi_j\right|<\infty$ & roots of $\Psi(z)=0$ outside the unit circle. Thus when $\left|\lambda_i\right|<1\forall i$ the process $y(t)=\alpha+\delta(t)+u(t)$ (4) would just be a special case of the trend stationary process of $y(t)=\alpha+\delta(t)+\psi(L)\varepsilon(t)$.

Suppose instead that $\lambda_1=1$ & $\left|\lambda_i\right|<1\forall i$ then (6) above would yield

$$(1-L)u(t)=\frac{1+\theta_1 L+\theta_2 L^2+......+\theta_q L^q}{(1-\lambda_1 L)(1-\lambda_2 L)......(1-\lambda_p L)}\varepsilon(t)\equiv\psi^*(L)\varepsilon(t) \qquad \text{(C.25)}$$

With $\sum_{j=0}^{\infty}\left|\psi_j\right|<\infty$ & roots of $\Psi^*(z)=0$ outside the unit circle.

Thus if $y(t)=\alpha+\delta(t)+u(t)$ is first differenced the result is;

$$(1-L)y(t)=(1-L)\alpha+[\delta(t)-\delta(t-1)]+(1-L)u(t)=0+\delta+\psi^*(L)\varepsilon(t) \qquad \text{(C.26)}$$

which is the form of our original equation.

This representation explains the use of the tern “unit root process” i.e. one of the roots or eigenvalues (λ_1) is 1 (of the autoregressive polynomial) and all other eignevalues are inside the unit circle. As stated above this process is known as I(1).

If the process has 2 eigenvalues λ_1 & λ_2 that are both 1 and all other eigenvalues are inside the unit circle the second difference of the data will have to be taken before getting to a stationary time series:

$$(1-L)^2 y(t) = k + \psi(L)\varepsilon(t) \qquad \text{(C.27)}$$

with y(t) ~ I(2).

A general process written in the form of (4) & (5) is called an Autoregressive Integrated Moving Average Process denoted ARIMA(p,d,q) where;
p = number of autoregressive lags (not counting unit roots)
d = order of integrations (number of differences)
q = number of moving average lags.
From this if we take the d^{th} difference of an ARIMA(p,d,q) process we arrive at a stationary ARMA(p,q) process.

C.5 FIGURES RE MODELLED VARIANCE

Further to section 4.1 and Cases I & II.

Case I : AR(1) model specified for conditional mean and GARCH(1,1) model

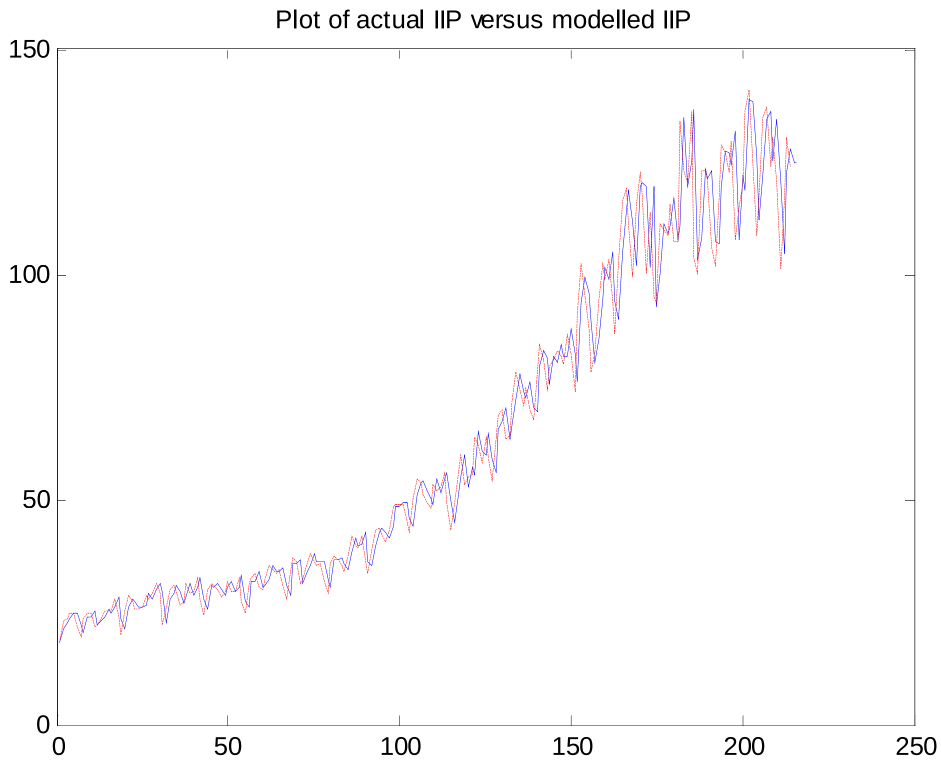


Fig C.1: Actual IIP versus modelled IIP

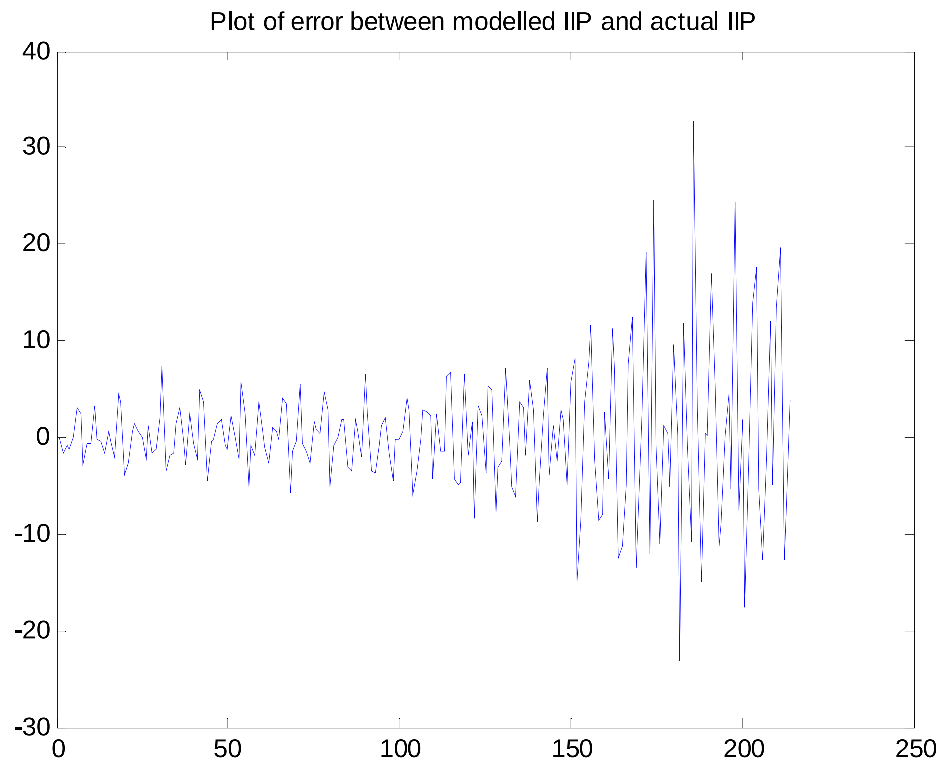


Fig C.2: Plot of error between actual IIP and modelled IIP for AR(1) model

Case II : ARX(1,1) model specified for conditional mean and GARCH(1,1) model

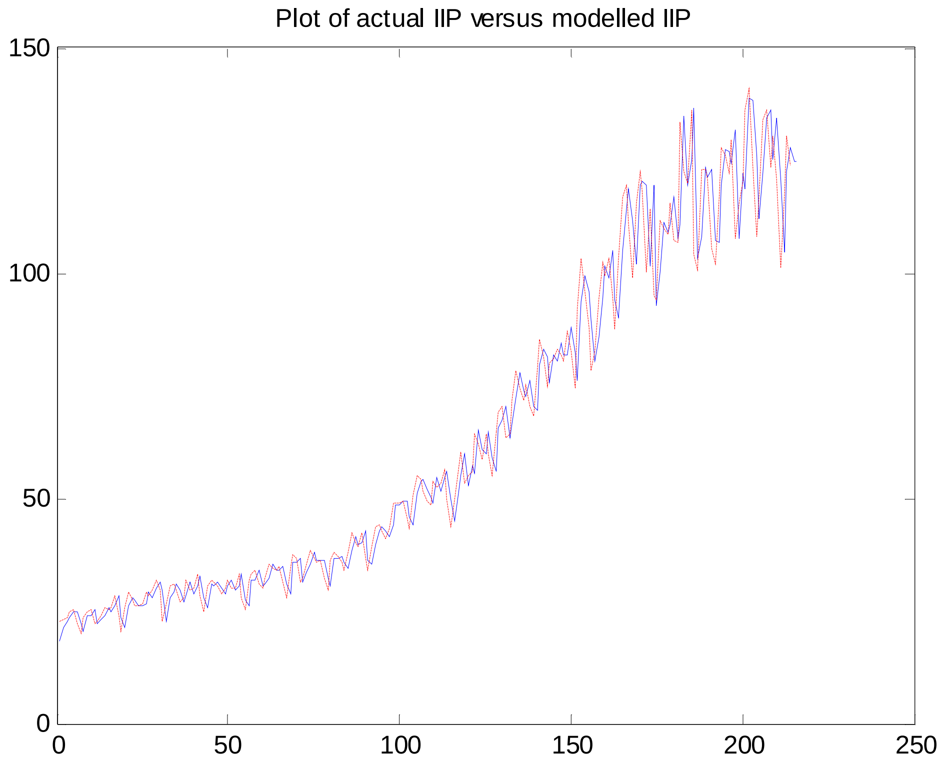


Fig C.3: Actual IIP versus modelled IIP using ARX(1,1) model

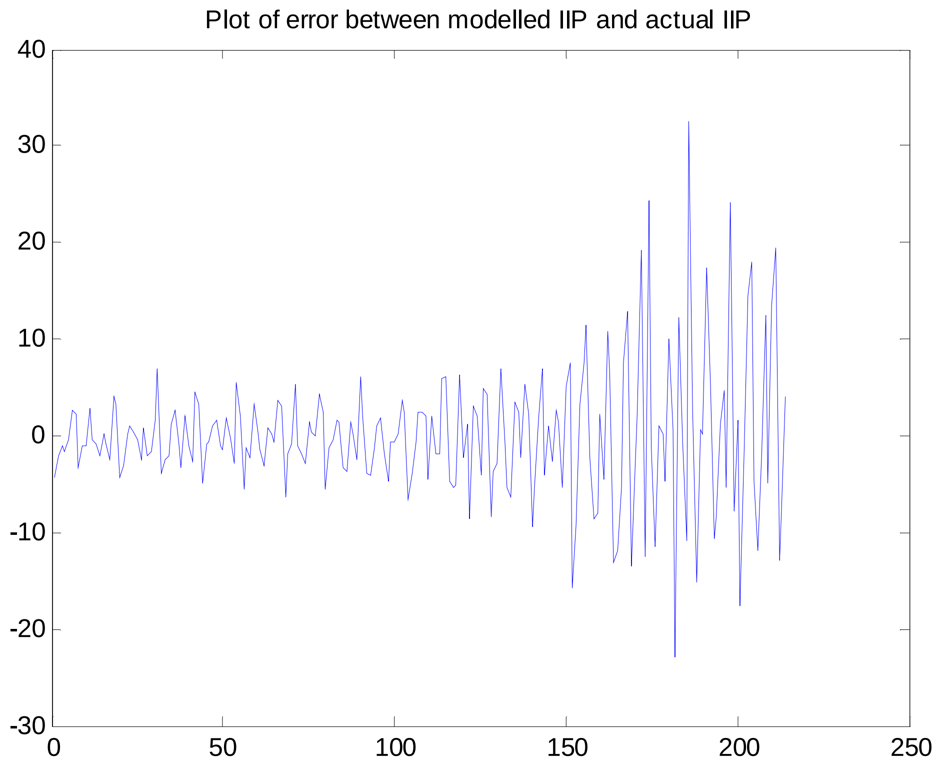


Fig C.4: Plot of error between actual IIP and modelled IIP for ARX(1,1) model

BIBLIOGRAPHY

[1] Recent House Price Developments: The Role of Fundamentals, OECD Economic Outlook No. 78 (Preliminary Edition)

[2] ID Wilson, SD Paris, JA Ware, DH Jenkins, Residential property price time series forecasting with neural networks, School of Technology, University of Glamorgan, Pontypridd, Mid Glamorgan, Wales. Elsevier, Knowledge Based Systems 15 (2992) Pg 335-341.

[3] HL Chan, SK Lee, KY Woo, Detecting rational bubbles in the residential housing markets of Hong Kong, Department of Economics, Hong Kong Baptist University, Kowloon, Hong Kong, Elsevier, Economic Modelling 18 (2001) Pg 61-73

[4] G Kenny, Modelling the demand and supply sides of the housing market: evidence from Ireland, Research Department, Central Bank of Ireland, Dame Street, Dublin 2. Elsevier, Economic Modelling 16 (1999) Pg 389-409.

[5] MJ Roche, The rise in house prices in Dublin: bubble, fad or just fundamentals, Department of Economics, National University of Ireland, Maynooth, Co. Kildare. Elsevier, Economic Modelling 18 (2001) Pg 281-295.

[6] S Holly, N Jones, House prices since the 1940's: cointegration, demography and asymmetrics, Department of Applied Economics, University of Cambridge, Sedgwick Avenue, Cambridge. Elsevier, Economic Modelling 14 (1997) Pg 549-565.

[7] S Stevenson, New empirical evidence on heteroskedaticity in hedonic housing models, Centre for Real Estate Research, Michael Smurfit Graduate Business School of Business, University College Dublin, Carysfort Avenue, Blackrock, Co. Dublin. Elsevier, Journal of Housing Economics 13 (2004) Pg 136-153.

[8] T Bollerslev, General Autoregressive Conditional Heteroskedasticity, University of California at San Diego, La Jolla CA 92093, USA, Journal of Econometrics 31 (1986) Pg 307-327.

[9] R Engle, GARCH 101: An introduction to the use of ARCH/GARCH models in Applied Econometrics, Journal of Economic Perspectives, Vol 15, No. 4 Fall 2001, Pg 157-168.

[10] CW Cheong, AHS Mohd Nor, Z Isa, Asymmetry and long memory volatility: Some empirical evidence using GARCH, Faculty of Information Technology, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia. Elsevier, Physica 373 (2007) Pg 651-664.

[11] C. Gouriéroux, ARCH Models and Financial Applications, Springer-Verlag New York 1997.

[12] JD Hamilton, Time Series Analysis, Princeton University Press 1994.

[13] MH Kutner, CJ Nachtsheim, J Neter, W Li, Applied Linear Statistical Models, 5th edition, McGraw-Hill Irwin 2005.

[14] J Nasstrom, Volatility Modelling of Asset Prices using GARCH models, Department of Electrical Engineering, Linkoping University, Sweden 2003.

[15] RF Engle, Autoregressive Conditional Heteroskedasticity with estimates of the variance of United Kingdom inflation, Econometrica, Vol 50, No. 4, July 1982 Pg 987-1007