
Using GLS to generate forecasts in regression models with auto-correlated disturbances with simulation and Palestinian market index data

Samir K. Safi¹, Ehab A. Abu Saif²

¹Dept. of Economics and Statistics, Faculty of Commerce, the Islamic University of Gaza, Gaza, Palestine

²Statistician Researcher, Gaza, Palestine

Email address:

samirsafi@gmail.com (S. K. Safi), censtat@hotmail.com (E. A. Abu Saif)

To cite this article:

Samir K. Safi, Ehab. A. Abu Saif. Using GLS to Generate Forecasts in Regression Models with Auto-correlated Disturbances with simulation and Palestinian Market Index Data. *American Journal of Theoretical and Applied Statistics*. Vol. 3, No. 1, 2014, pp. 6-17. doi: 10.11648/j.ajtas.20140301.12

Abstract: This paper involves an important statistical problem concerning forecasting in regression models in time series processes. It is well known that the most famous method of estimating and forecasting is the Ordinary Least Squares (OLS). OLS may be not the optimal in this context. So over the years many specialized estimation techniques have been developed, for example Generalized Least Squares (GLS). We are comparing the forecasting based on some estimators with the prediction using the GLS estimate. This comparison will be used by what is known as measures of forecast accuracy. We conduct an extensive computer simulation time series data, to make comparison among these methods. The similar forecasting criteria were developed and evaluated for the real data set on daily closing price in the Palestinian market index (Alquds Index). The data consists of 164 monthly observations and obtained from the website of the Palestine Stock Exchange. The main finding is that, for forecasting purposes there is not much gained in trying to identifying the exact order and form of the auto-correlated disturbances by using GLS estimation method. In addition, we noticed that the accuracy of forecasting using GLS method does not differ substantially than the other methods as Maximum Likelihood Estimation (MLE), Minimize Conditional Sum of Squares (CSS) and the combination of these two methods. Moreover, for parameter estimation, the GLS is nearly as efficient as the exact parameter estimation. On the other hand, the Ordinary Least Squares (OLS) method performs much less efficient than the other estimation methods and producing poor forecasting accuracy.

Keywords: Generalized Least Squares, Ordinary Least Squares, Maximum Likelihood, Forecasting Accuracy, Simulation

1. Introduction

Comparison of estimators and forecasting in linear regression models with autocorrelated disturbances is inspired by problems, which arise in meteorology and economics. It is well known that the most famous method of estimating and forecasting is the Ordinary Least Squares (OLS), it is maybe not the optimal in this context. So over the years many specialized estimation techniques have been developed, for example Generalized Least Squares (GLS). These methods are more complicated than OLS and are less understood. We are comparing the prediction based on some estimators with the prediction using the GLS estimate. This comparison will be used by what is known as measures of forecast accuracy.

This paper aims to study the GLS method for parameter estimation in the regression models with autocorrelated disturbances. Comparison of GLS for estimation with other well known methods based on forecasts criterion is discussed.

In this section we introduce some related studies and recall the most important findings. Shittu and Asemota (2009) were compared the performance of model order determination criteria in terms of selecting the correct order of an autoregressive model using the simulation method in small and large samples. The criteria considered are the Akaike information criterion (AIC); Bayesian information criterion (BIC), and the Hannan Quinn criterion (HQ). The results shows that BIC performs best in terms of selecting the correct order of an Autoregressive model for small

samples irrespective of the AR structure, HQ criteria can be said to perform best in large sample. Even though the AIC has the least performance among the criteria considered, it appears to be the best in terms of the closeness of the selected order to the true value.

Ojo and Olatayo (2009) compared subset Autoregressive Integrated Moving Average (ARIMA) models, with full ARIMA models. They used residual variance, AIC and BIC, to determine the performance of the models. Results revealed that the residual variance attached to the subset autoregressive integrated moving average models is smaller than the residual variance attached to the full autoregressive integrated moving average models. Subset autoregressive integrated moving average models performed better than the full autoregressive integrated moving average models.

Lee and Lund (2004) were proposed the properties of OLS and GLS estimators in a simple linear regression with stationary autocorrelated errors. Explicit expressions for the variances of the regression parameter estimators are derived for some common time series autocorrelation structures, including a first-order autoregression and general moving averages. Applications of the results include confidence intervals and an example where the variance of the trend slope estimator does not increase with increasing autocorrelation.

Koreisha and Fang (2004) were studied and described a new procedure for generating forecasts for regression models with serial correlation based on OLS. From a large simulation study they found that for finite samples the predictive efficiency of their two-step linear approach is higher than that of OLS for short and medium horizon, and very comparable to that of GLS based on $AR(\tilde{p})$ corrections with $\tilde{p} = \sqrt{T}/2$, where T is number of observation, which is also known to be very similar to the GLS estimation procedure when the error covariance matrix Ω will be estimated from data. For longer horizons OLS yields forecasts that are as efficient as those generated by GLS approaches and the two-step procedure for generating forecasts for regression models with serial correlation based exclusively on ordinary least squares (2SOLS) estimation.

Safi (2004) discussed the comparison of efficiency of the OLS estimation to alternative procedures such as GLS and estimated GLS (EGLS) estimators in the presence of first and second order autoregressive disturbances. The most important findings that the relative efficiency of the OLS estimator as compared to the GLS estimator decreases with increasing values of $|\rho|$, he found that the efficiency of the OLS estimator for estimating an intercept appears to be nearly as efficient as the GLS estimator for $|\rho| \leq 0.7$ for relatively small and moderate sample sizes. However for

large sample size, OLS appears to be nearly as efficient as the GLS estimator for the additional values of $\rho = \pm 0.9$. And The OLS estimator may often be better than assuming another incorrect truncation of the actual process. In addition, it is sometimes better to ignore the problem altogether and use OLS rather than to incorrectly assume the process is AR (1).

Findley (2003) studied properties of forecast errors and estimates of misspecified ARIMA and intermediate memory models and the Optimality of GLS for One-Step-Ahead Forecasting. Both OLS and GLS estimates of the mean function are considered. He showed that GLS has an optimal one-step-ahead forecasting property relative to OLS when the model omits a regression variable of the true mean function that is asymptotically correlated with a modeled regression variable. Some inherent ambiguity in the concept of bias for regression coefficient estimators in this situation is discussed.

The goals of this paper can be split into three main issues. Firstly, study the forecasting behavior using GLS method in regression models with auto-correlated disturbances, and compare the forecast accuracy with other estimation methods. Secondly, evaluate the forecasting for the real data set on Alquds Index for illustrative purposes, and finally conducting exhaustive simulation study setup for examining the accuracy of our findings.

This paper is organized as follows: Section 2 introduces model estimation by MLE, OLS, and GLS. In section 3, a case study on Palestinian Al-Quds index stock data is analyzed. Section 4 focuses on forecasting evaluation and present the comparisons of estimation methods for the real and simulated data. Section 5 summaries the results and offers suggestions for future research for using GLS to generate forecasts in regression models with auto-correlated disturbances.

2. Model Estimation

Building time series models involves three basic steps, model identification, model estimation and model diagnostics. In this section, we introduce the model estimation which is relevant to the purpose of this paper. We consider, Maximum Likelihood estimation, Least Squares Estimation, and GLS Estimation.

2.1. Maximum Likelihood Estimation

Basically, an *ARMA* model combines the ideas of autoregressive *AR* and moving average *MA* models into a compact form. A mixed autoregressive moving average model with p autoregressive terms and q moving average terms is abbreviated *ARMA*(p,q) and may be written as, Cryer and Chan (2008).

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

When the orders p and q of *ARMA*(p,q) model are known,

estimates of the ϕ_i s and θ_j s can be found when the data

being observations from a Gaussian *ARMA* model. Even if $\{X_t\}$ is not Gaussian, the Gaussian likelihood still is a reasonable measure of goodness of fit of the model, so maximizing it is sensible. Also, the asymptotic distribution of maximum likelihood estimators is the same whether the white noise innovations \mathcal{E}_t (and so the process itself) are Normal or not. Suppose our observed data x_1, \dots, x_n are placed in the data vector $(x_1, \dots, x_n)'$ (after appropriate differencing and mean-correction) are modeled by a stationary zero-mean Gaussian *ARMA*(p, q) process $\{X_t\}$ with parameters $\phi = (\phi_1, \dots, \phi_p)'$, $\theta = (\theta_1, \dots, \theta_q)'$ and σ^2 . Recall that σ^2 is the variance of the *WN* variables in the *ARMA* process. The covariance matrix of $X_n = (x_1, \dots, x_n)'$

$$E(XX') = \Gamma_n(\phi, \theta, \sigma^2)$$

say, is a symmetric positive definite $n \times n$ matrix, Consists from the covariances $\gamma_{i-j}, i, j = 1, \dots, n$, themselves functions of the model parameters. Because the X_t have a joint n -variate normal distribution the likelihood based on data x_n is

$$L(\phi, \theta, \sigma^2; x) = (2\pi)^{-\frac{n}{2}} |\Gamma_n|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} x_n' \Gamma_n^{-1} x_n\right)$$

and the log-likelihood up to a constant term

$$\ell(\phi, \theta, \sigma^2; x) = -\frac{1}{2} x_n' \Gamma_n^{-1} x_n - \frac{1}{2} \log |\Gamma_n|$$

In principle the maximum likelihood estimates can be obtained from this by numerical maximization. In practice however, if n is large, direct calculation of Γ_n^{-1} and $|\Gamma_n|$ could be a problem. Fortunately algorithms are available to avoid the difficulty. What is needed for a numerical maximization procedure is the ability to evaluate the log-likelihood ℓ quickly at specified values of the parameters. Given this ability a good maximization routine should be able to iterate efficiently towards the parameter values at which ℓ is largest. Thus maximization depends on the ability to evaluate ℓ easily at any given set of parameter values (See for example Abraham and Ledolter (2005).

One method of evaluation is to build up ℓ , starting as though there were very few observations, and successively calculating the changes as new observations are taken into account. To see how this works we need some further notation. For each $j = 1, 2, \dots$. Let \hat{X}_j denote the minimum variance estimate of X_j based on linear combinations of observations made before time j , $X_{j-1}, X_{j-2}, \dots, X_1$, These \hat{X}_j 's are called one-step-ahead predictors. They are linear combinations of earlier observations, with

coefficients which are functions of the covariances for the particular model we are considering. These coefficients can be expressed in terms of the model parameters, but we will not need them explicitly for the current discussion. Also let U_j denote the difference $X_j - \hat{X}_j$, called the innovation at time j . A property of the U_j 's is that they have zero expectation and are uncorrelated with each other. (This follows from the easily-proved fact that \hat{X}_j is the conditional expectation of X_j given all values observed before j .) Finally let \hat{X}_n and U_n denote the vectors of predictors $(\hat{X}_1, \dots, \hat{X}_n)$ and innovations (U_1, \dots, U_n) respectively.

Since \hat{X}_j is a function only of earlier X_i 's and is linear in them, we can write

$$U_n = X_n - \hat{X}_n = A_n X_n$$

where A_n is a matrix with 1's along its diagonal and zeroes above the diagonal, the values below the diagonal being the coefficients determining the \hat{X}_j . Being triangular with positive diagonal entries, A_n is non-singular, and so, writing $A_n^{-1} = C_n$, we have

$$X_n = C_n (X_n - \hat{X}_n)$$

Like A_n , the inverse C_n is also triangular with 1's along its diagonal and zeroes above it. Also, since the components of $X_n - \hat{X}_n$ are uncorrelated, the variance-covariance matrix of this vector, D_n say,

$$E(X_n - \hat{X}_n)(X_n - \hat{X}_n)' = D_n$$

must be diagonal, $D_n = \text{diag}(v_0, \dots, v_{n-1})$, where $v_{j-1} = \text{Var}(U_j)$ is the innovation variance at time j .

From the above

$$\begin{aligned} \Gamma_n &= E(X_n X_n') = E(C_n (X_n - \hat{X}_n)(X_n - \hat{X}_n)' C_n') = C_n D_n C_n' \\ &= (X_n - \hat{X}_n)' D_n^{-1} (X_n - \hat{X}_n) \\ &= \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{v_{j-1}} \end{aligned} \quad (2.1)$$

and

$$|\Gamma_n| = |C_n D_n C_n'| = |C_n|^2 |D_n| = v_0 \dots v_{n-1},$$

since $|C_n| = 1$ because of C_n 's unit-diagonal/triangular structure, Cochrane (2005)..

Thus the log-likelihood becomes

$$\ell(\phi, \theta, \sigma^2; x) = -\frac{1}{2} \sum_{j=1}^n \log v_{j-1} - \frac{1}{2} \sum_{j=1}^n \frac{(x_j - \hat{X}_j)^2}{v_{j-1}} \quad (2.2)$$

The great advantage of this expression is that the v_j 's and \hat{X}_j 's can be calculated very efficiently by recursion using the Innovations Algorithm. The procedure therefore gives a highly effective route to the maximum likelihood estimators.

In fact the Innovations Algorithm works for any process, whether of *ARMA* form or not, stationary or not. For an *ARMA* model in particular it is found that the innovation variances v_j all have the form

$$v_j = \sigma^2 w_j$$

where the w_j depend on ϕ and θ but not on σ^2 . The coefficients defining the one-step-ahead predictors \hat{X}_j in terms of earlier observations also do not depend on σ^2 . The log-likelihood therefore becomes

$$\ell(\phi, \theta, \sigma^2; x) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{j=1}^n \log w_{j-1} - \frac{1}{2\sigma^2} \sum_{j=1}^n \frac{(x_j - \hat{X}_j)^2}{w_{j-1}} \quad (2.3)$$

and if we maximize this with respect to σ^2 (by differentiating with respect to σ^2 , setting the derivative equal to 0 and solving the resulting equation), we find that

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n \frac{(x_j - \hat{X}_j)^2}{w_{j-1}} = \frac{S(\hat{\phi}, \hat{\theta})}{n} \quad (2.4)$$

say, giving the maximum likelihood estimate of σ^2 in terms of a sum of squares S depending on the estimates of the other parameters. If this expression for $\hat{\sigma}^2$ is substituted back into the log-likelihood, thereby eliminating σ^2 , we find that $\hat{\phi}$ and $\hat{\theta}$ must be the values of ϕ and θ maximizing

$$\ell(\phi, \theta, \sigma^2; x) = -\frac{1}{2} \sum_{j=1}^n \log w_{j-1} - \frac{n}{2} \log \frac{S(\phi, \theta)}{n}$$

that is, minimizing

$$n^{-1} \sum_{j=1}^n \log w_{j-1} + \log \frac{S(\phi, \theta)}{n} \quad (2.5)$$

Minimization of $S(\phi, \theta)$ alone would be a form of least squares estimation. From (2.5) we see that if $n^{-1} \sum_{j=1}^n \log w_{j-1}$ is small or varies little with ϕ and θ the resulting estimates are likely to be close to the maximum likelihood estimates.

In practice (2.4) is little used as an estimate of σ^2 . Instead, as for the estimation of variance in regression, the

estimate $\hat{\sigma}^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n - p - q}$ is preferred. (See for example, Everitt and Hothorn, 2010 and Fox, 2002).

2.2. Least Squares Estimation

In this section, we present the least square estimation in time series and regression model.

2.2.1. The Least Squares in Time Series

The underlying idea in the fitting of a time series model by least squares, by analogy with regression, is that we should choose parameter values which minimize the sum of squared differences between the observed data and their expected values according to the model. We present least squares estimation in three models, namely: AR, MA, and ARMA.

For *AR(1)*, $X_t = \phi X_{t-1} + \varepsilon_t$: Given the observations up to time $j-1$, the expected value of X_j is ϕx_{j-1} , so it is reasonable to use

$$S(\phi) = \sum_{j=2}^n (x_j - \phi x_{j-1})^2$$

as the sum of squares to be minimized. Note that the $j = 1$ term is omitted since x_0 is not available. Minimization of S with respect to ϕ leads to the estimate

$$\tilde{\phi} = \frac{\sum_{j=2}^n x_{j-1} x_j}{\sum_{j=2}^n x_{j-1}^2}$$

The same idea works in the same way for *AR(p)* models, leaving out differences from x_j 's with $j \leq p$.

For *MA (1)*, $X_t = \varepsilon_t + \theta \varepsilon_{t-1}$: if the *MA* model is invertible then

$$X_t = -\sum_{i=1}^{\infty} (-\theta)^i X_{t-i} + \varepsilon_t$$

By analogy with the previous example and with regression this suggests basing a sum of squares on the differences

$$\varepsilon_t = X_t - \left(-\sum_{i=1}^{\infty} (-\theta)^i X_{t-i} \right)$$

However we cannot use these directly (infinitely many terms in the sums). Instead take

$$S(\theta) = \sum_{i=1}^n \varepsilon_i^2$$

and note that from the defining relation of the process we could find the terms here successively from

$$\begin{aligned}\varepsilon_1 &= X_1 - \theta\varepsilon_0 \\ \varepsilon_2 &= X_2 - \theta\varepsilon_1 \\ &\vdots \\ \varepsilon_n &= X_n - \theta\varepsilon_{n-1}\end{aligned}$$

provided we knew ε_0 . Not knowing ε_0 it's natural to replace it by its expectation, zero. S calculated in this way is referred to as a conditional sum of squares, and the estimate obtained by minimizing the conditional S is called the conditional least-squares estimate.

For a general *ARMA* model: the conditional sum of squares can be defined much as in the *MA* case,

$$\varepsilon_t = X_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} - \sum_{j=1}^p \phi_j X_{t-j}$$

to calculate ε_t from the previous ones and past data. As in the *MA* case we can take the ε 's on the right hand side to be zero for non-positive indices. As in the *AR* case we may content ourselves with calculating the ε 's only for $t > p$ so that we only use observed X 's, or we might take X 's with negative suffices to be zero too. Unconditional Least Squares means minimizing the sum of squares $S(\phi, \theta)$ in (2.5). As noted this is not quite the same as maximizing the likelihood, but will often give similar results.

2.2.2. The Least Squares in Regression Model

The regression model can be written as

$$y_t = f(x_t; \beta) + \varepsilon_t \quad (2.6)$$

where $f(x_t; \beta)$ is a mathematical function of the p independent variables $x_t = (x_{t1}, \dots, x_{tp})'$ and unknown parameters $\beta = (\beta_1, \dots, \beta_m)'$. Due to the random nature of the error terms ε_t , the dependent variable y_t itself is a random variable. The model in (2.6) can therefore also be expressed in terms of the conditional distribution of y_t given $x_t = (x_{t1}, \dots, x_{tp})'$. The regression assumptions can be written as:

1. The conditional mean, $E(y_t | x_t) = f(x_t; \beta)$, depends on the independent variables x_t and the parameters β , and the variance $Var(y_t | x_t) = \sigma^2$ is independent of x_t and time.
2. The dependent variables y_t and y_{t-k} for different time periods (or subjects) are uncorrelated

$$Cov(y_t, y_{t-k}) = E[y_t - f(x_t; \beta)][y_{t-k} - f(x_{t-k}; \beta)] = 0$$

3. Conditional on x_t, y_t follows a normal distribution with mean $f(x_t; \beta)$ and variance σ^2 , this is denoted by $N(f(x_t; \beta), \sigma^2)$.

Least squares estimates are minimize the sum of the squared deviations $S(\beta) = \sum_{i=1}^n [y_i - f(x_i; \beta)]^2$ and denoted by $\hat{\beta}$. The general linear regression model Linear, regression models can be written as

$$y = X\beta + \varepsilon \quad (2.7)$$

where y is a $n \times 1$ vector of observations on a dependent variable, X is a $n \times k$ matrix of independent variables of full column rank, β is a $k \times 1$ vector of parameters to be estimated, and ε is a $n \times 1$ vector of disturbances. In matrix notation the least squares criterion can be expressed as minimizing

$$S(\beta) = \sum_{i=1}^n (y_i - x_i' \beta)^2 = (y - X\beta)'(y - X\beta)$$

The minimization of $S(\beta)$ leads to the least squares estimator $\hat{\beta}$, which satisfies the $p+1$ equations

$$(X'X)\hat{\beta} = X'y \quad (2.8)$$

These are referred to as the normal equations. Since we have assumed that the design matrix X is of full column rank, the inverse of $X'X$ can be calculated. The solution of the normal equations is then given by

$$\hat{\beta} = (X'X)^{-1} X'y \quad (2.9)$$

The $(p+1) \times (p+1)$ covariance matrix of $\hat{\beta}$ is given by

$$V(\hat{\beta}) = \sigma^2 (X'X)^{-1} \quad (2.10)$$

Consider a standard linear model (2.7) with all the assumptions of the classical linear model except the assumption of constant variance, non-autocorrelated error terms. Replace this assumption with

$$Var(\varepsilon) = E[\varepsilon\varepsilon'] = \sigma^2 \Omega,$$

where Ω is an $n \times n$ symmetric and invertible matrix. Each element w_{ij} of Ω is proportional to the covariance between the errors ε_i and ε_j of observations i and j . Each diagonal element w_{ii} is proportional to the variance of ε_i . When the variance and covariance of the unobserved factors takes this form then the formula for estimating the variance covariance matrix of $\hat{\beta}$ is

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1} \quad (2.11)$$

Unfortunately we do not usually know Ω unless we put a specific structure on the process that determine the unobservable factors in the model ϵ . In general, It is impossible to estimate Ω without restrictions on its structure. Ω has $\frac{n(n+1)}{2}$ unique parameters, but we only have n observations and the number of parameters to be estimated, including those in $\hat{\beta}$, must be less than the number of observations.

For Consistent estimation of the variance covariance matrix, there are ways to estimate $X\epsilon\epsilon'X = \sigma^2X\Omega X$ which is a $k \times k$ matrix of all the cross products of the error terms and the regressors. There are various forms designed to deal with different situations. In general, what's required is an estimate of the form $\frac{1}{n} \sum_i \sum_j \hat{\epsilon}_i \hat{\epsilon}_j x_i x_j'$. But this usually restricted in some way, e.g. the The Huber-White estimate is $\frac{1}{n} \sum_i \hat{\epsilon}_i^2 x_i x_i'$ and deals with heteroskedasticity only. The Newey-West estimator extends this to serial correlation. Another popular choice is for 'clustered' standard errors, which may be useful approximations with spatial data.

2.3. Generalized Least Squares Estimation

Regressors are assumed to be non-stochastic, i.e. fixed in repeated sampling, independent and uncorrelated with the error terms in the classical linear regression model. These assumptions are not always satisfied especially in time series. OLS is not the most efficient estimator here. We can gain precision in least-squares estimates by weighting observations with 'lower' variance more heavily than those with 'higher' variance, so that the weighted error variance covariance matrix is of the standard form. The intuition is that we weight the estimator so that it places greater emphasis on observations for which the observable explanatory variables do a better job of explaining the dependent variable. This means we need to devise a $n \times n$ weighting matrix C such that:

$$Var(C\epsilon) = E [C\epsilon\epsilon'C'] = \sigma^2I$$

A matrix that does this, is a matrix C such that $C'C = \Omega^{-1}$, which implies $C\Omega C' = I$. In fact, several such matrices C exist, so that, for convenience, we can assume $C = C'$. To derive the form of the Best Linear Unbiased Estimator (BLUE) of β for the generalized regression model under the assumption that Ω is known, weighting all the variables in the model gives $Cy = CX\beta + C\epsilon$. Then, the OLS estimator applied to this gives the GLS estimator

$$\hat{\beta}^{GLS} = (X\Omega^{-1}X)^{-1} X\Omega^{-1}y$$

If the Ω are known then the GLS estimator $\hat{\beta}^{GLS}$ is BLUE, with variance covariance matrix $Var(\hat{\beta}^{GLS}) = \sigma^2 (X\Omega^{-1}X)^{-1}$. Note that this assumes that the parameters are homogenous across the sample, i.e. they do not change for different groups in the data. Weighting the data will then change the parameter estimate according to which groups are more heavily weighted.

In practice, Ω is typically unknown so that the GLS estimator is maybe not available. meaning that $\hat{\beta}^{GLS}$ is non-operational, and a estimated or feasible generalized least squares (FGLS) estimator is used. If the matrix Ω unknown, we can parameterize the matrix $\Omega = \Omega(\theta)$ in terms of a finite-dimensional parameter vector θ , and use the classical OLS residuals to obtain consistent estimators $\hat{\theta}$ and $\hat{\Omega} = \Omega(\hat{\theta})$ of θ and Ω . Then replace the unknown Ω with the estimated $\hat{\Omega}$ in the formula for GLS, yielding the feasible GLS estimator,

$$\hat{\beta}^{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1} X'\hat{\Omega}^{-1}y$$

3. Case Study: Palestinian Al-Quds Index Stock Data

In this section, we consider a real data set called Palestinian Al-Quds index stock Data. The data is obtained from Palestine Exchange (PEX) web page (www.p-s-e.ps). We consider the closing price values for the Al-Quds Index. R-statistical software is used for fitting ARIMA model for the time series.

3.1. Data Description

We consider the monthly closing price values for the Al-Quds Index in Palestine, from September 1997 to April 2011. The data is taken at the end of last trading every month, thus we have (164) observation. The closing price for the Al-Quds Index ranges between 97.01 and 1295.08 with mean 376.17 and standard deviation 233.95.

3.2. Data Processing

The Ljung-Box test statistic equals 157.53. This is referred to a chi-square distribution with one degree of freedom. This leads to a p -value $< 2.2e-16$, so we reject the null hypothesis that the error terms are uncorrelated. In other words, there is strong evidence of autocorrelation in the residuals of this data.

Figure 3.1 displays the time series plot. The series displays considerable fluctuations over time, especially since 2005, and a stationary model does not seem to be reasonable. The higher values display considerably more variation than the lower values.

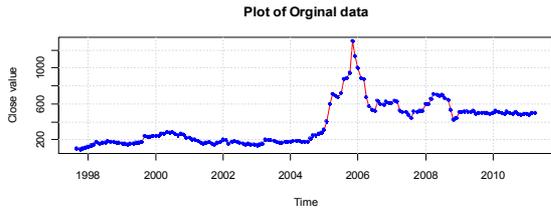


Figure 3.1. Monthly series of closing value for al-quds index: Sep. 1997 to April 2011

KPSS test for level stationarity is applied to the original series leads to a test statistic of 3.0973 and a p -value of 0.01. With stationarity as the null hypothesis, this provides strong evidence supporting the nonstationarity and the appropriateness of taking a difference of the original series.

The differences of the closing values for al-quds index are displayed in Figure 4.2. The differenced series looks much more stationary when compared with the original time series shown in Figure 4.1. On the basis of this plot, we might well consider a stationary model as appropriate.

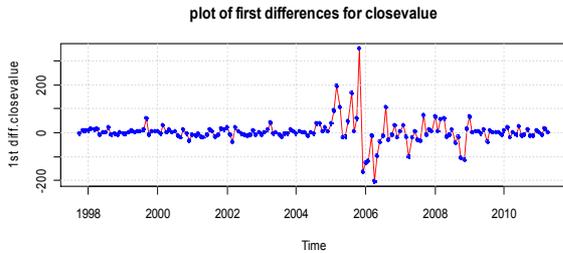


Figure 3.2. The Difference Series of the Monthly Closing Values for Al-Quds Index

KPSS test is applied to the differenced series leads to a test statistic of 0.0678 and a p -value of 0.10. That is, we do not reject the null hypothesis of Stationarity.

3.3. Model Specification

Both the sample ACF and PACF displayed in Figure 3.3, cut off after lag (1), strongly suggested an MA(1) or AR(1) appropriate model for the differenced series, respectively. Therefore, it is quite difficult to identify the MA, AR, or mixed model from this figure.

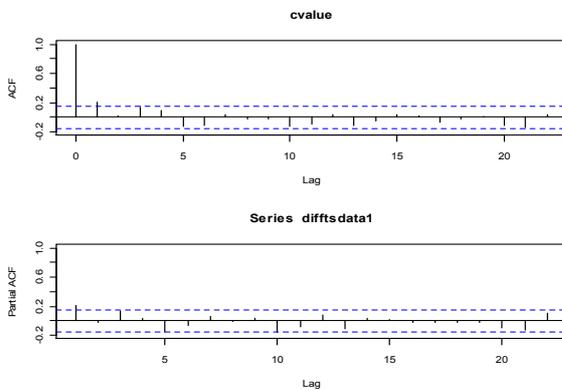


Figure 3.3. Sample ACF and PACF for Difference of the Monthly Closing Values for Al-Quds Index

The sample EACF computed on the first differences of the Monthly Closing Values for Al-Quds Index is shown in Table 3.1. In this table, an ARMA(p,q) process will have a theoretical pattern of a triangle of zeroes, with the upper left-hand vertex corresponding to the ARMA orders. Table 3.1 displays the schematic patterns for possibility of IMA(1,1), ARIMA(1,1,1), or ARIMA(2,1,2).

Table 3.1. EACF for Difference of Monthly Closing Values for Al-Quds Index Series

		MA													
		0	1	2	3	4	5	6	7	8	9	1	1	1	1
A R	0	x	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	x	0	0	0	0	0	0	0	0	0	0	0	0
	3	x	x	0	0	0	0	0	0	0	0	0	0	0	0
	4	x	x	0	0	0	0	0	0	0	0	0	0	0	0
	5	x	x	x	0	0	0	0	0	0	0	0	0	0	0
	6	x	0	0	0	x	0	0	0	0	0	0	0	0	0
	7	x	x	0	0	x	x	0	0	0	0	0	0	0	0

3.4. Model Selection

The estimate of the series mean is not significantly different from zero (P -value = 0.5467).

The three selected ARIMA models and their corresponding criteria are shown in Table 3.2. These criteria confirm the selection suggestion -IMA(1,1)- based on the smallest values of AIC, AICc, and BIC among the other ARIMA choices.

Table 3.2. Different criteria for suggested ARIMA models

Models	AIC	AICc	BIC
IMA(1,1)	1742.84	1742.99	1752.12
ARIMA(1,1,1)	1744.77	1745.03	1757.15
ARIMA(2,1,2)	1747.87	1748.4	1766.43

3.5. Model Diagnostic

Figure 3.4 displays the time series plot of the standardized residuals from the IMA(1,1) model estimated for the Al-Quds index series time series. The model was fitted using maximum likelihood estimation. There are few residuals with magnitude larger than 2. The standardized residuals don't show clusters of volatility and seem to be fairly "random" with no particular patterns.

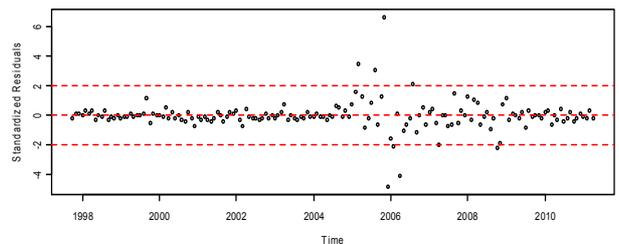


Figure 3.4. Standardized Residuals of the Fitted Model from Al-Quds index IMA (1,1) Model

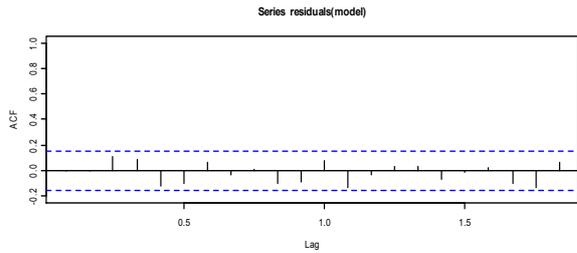


Figure 3.5. Sample ACF of Residuals of the Fitted Model IMA(1,1) Model

To check on the independence of the error terms in the model, we consider the sample autocorrelation function of the residuals. Figure 3.5 displays the sample ACF of the residuals from the IMA (1,1) model of the from Al-Quds index data. The dashed horizontal lines plotted are based on the large lag standard error of $\pm 2/\sqrt{n} = \pm 0.156$ ($n=164$). The graph does not show statistically significant evidence

of nonzero autocorrelation in the residuals. In other words, there is no evidence of autocorrelation in the residuals of this model. These residual autocorrelations look excellent.

In addition to looking at residual correlations at individual lags, it is useful to have a test that takes into account their magnitudes as a group. Figure 3.6 shows the p-values for the Ljung-Box test statistic for a whole range of values of K from 1 to 20. The horizontal dashed line at 5% helps judge the size of the p-values. The Ljung-Box test statistic with $K = 11$ is equal to 11.7298. This is referred to a chi-square distribution with 10 degrees of freedom. This leads to a p-value of 0.3035, so we have no evidence to reject the null hypothesis that the error terms are uncorrelated. The suggested model looks to fit the modeling time series very well.

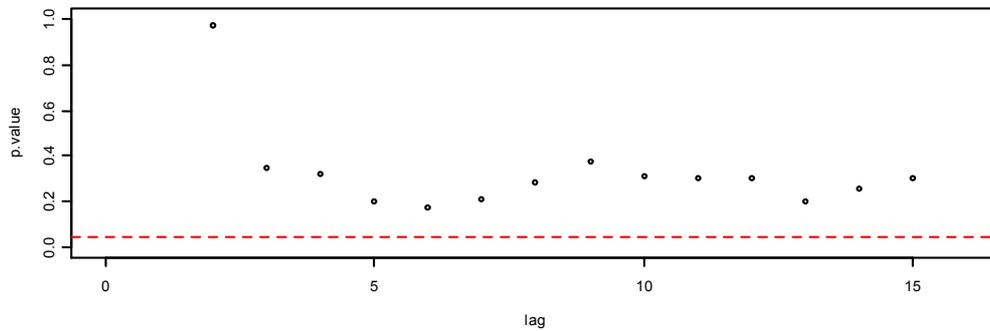


Figure 3.6. P-values for the Ljung-Box Test for the Fitted Model

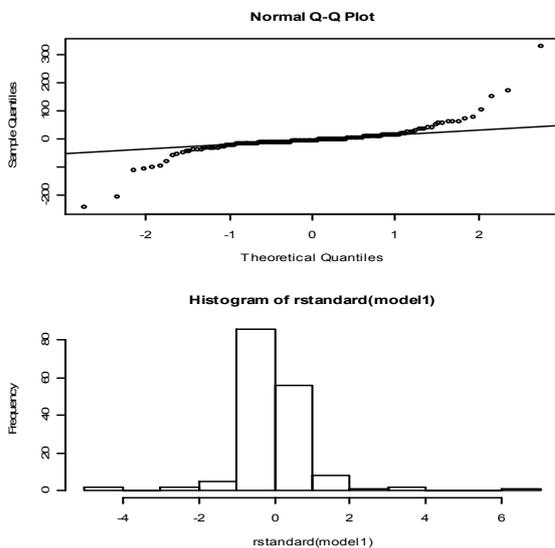


Figure 3.7. Quantile-Quantile Plot and histogram of the Residuals of the Fitted Model from Al-Quds index IMA (1,1) Model

A quantile-quantile plots are an effective tool for assessing normality. Here we apply them to the residuals of the fitted model. A quantile-quantile plot of the residuals from the IMA(1,1) model estimated for the Al-Quds index

series is shown in Figure 3.7. The points seem to follow the straight line fairly closely. This graph would not lead us to reject normality of the error terms in this model. In addition, with a few minor exceptions in the lower and upper tails, the histogram of the standardized residuals seems to be normal.

Therefore the estimated IMA(1,1) model seems to be capturing the dependence structure of the difference of Al-Quds index.

3.5. Estimation Methods

Using *forecast* package in R program, there are three estimation methods.

1. "CSS-ML" minimize conditional sum-of-squares to find starting values then maximum likelihood (the default method).
2. "ML" maximum likelihood.
3. "CSS" minimize conditional sum-of-squares.

In addition GLS will be used for comparison purposes.

Table (3.3) shows the results for the four mentioned methods for the three selected ARIMA models, namely: IMA(1,1), ARIMA(1,1,1) and ARIMA(2,1,2).

Table 3.3. result of methods with measures of estimate and measures of forecast accuracy

Method	ϕ_1	ϕ_2	θ_1	θ_2	AIC	AIC _c	BIC
IMA(1,1)							
ML			0.2184		1742.84	1742.99	1752.12
CSS-ML			0.2184		1742.84	1742.99	1752.12
CSS			0.2197		NA	NA	NA
GLS			0.22775		1737.53	NA	1749.85
ARIMA(1,1,1)							
ML	-0.1027		0.3173		1744.77	1745.03	1757.15
CSS-ML	-0.1063		0.3208		1744.77	1745.03	1757.15
CSS	-0.1062		0.3220		NA	NA	NA
GLS	-0.0458		0.2719		1739.56	NA	1754.92
ARIMA(2,1,2)							
ML	0.1789	0.250	0.0395	-0.2682	1747.87	1748.4	1766.43
CSS-ML	0.2051	0.2407	0.0141	-0.2647	1747.87	1748.4	1766.43
CSS	0.2059	0.2403	0.0145	-0.264	NA	NA	NA
GLS	0.2343	0.2807	-0.0019	-0.2955	1742.08	NA	1763.65

Based on AIC, AICc, and BIC, the results confirm IMA (1,1) is the best model among the others.

Here we see that $\hat{\theta} = 0.2184$. Noting the P-values for the estimate of the moving average coefficient, $\hat{\theta}$ (0.005) is significantly different from zero statistically, and insignificant for the intercept term intercept (0.61), consequently, it is not included in the estimated model. Therefore, The fitted model can be written as:

$$w_t = \varepsilon_t - 0.2184\varepsilon_{t-1}$$

4. Forecasting Evaluation

The crucial object in measuring forecast accuracy is the loss function. In this section we present the most widely statistical loss functions. Accuracy measures are usually defined on the forecast errors $e_{t+k,t} = X_{t+k} - \hat{X}_{t+k,t}$.

Definition 4.1. Mean Squared Error (MSE) eliminates the positive-negative problem by squaring the errors. The result tends to place more emphasis on the larger errors and therefore gives a more conservative measure than the MAE. This approach penalizes large forecasting errors. The MSE is given by

$$MSE = \frac{1}{T} \sum_{t=1}^T e_{t+k,t}^2 \tag{4.1}$$

where, T is number of periods used in the calculation.

The root mean squared error, $RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T e_{t+k,t}^2}$, is easy to interpret it is one of the most commonly used measures of forecast accuracy.

4.1. Comparisons of Estimation Methods for Real Data

We consider five estimation methods CSS-ML, ML, CSS, GLS, and classical OLS for the three selected ARIMA models, namely: IMA(1,1), ARIMA(1,1,1) and ARIMA(2,1,2). We compare the forecasting performance

for GLS estimation method with the other four methods. Table (4.1) shows the complete results for the forecast accuracy criterion RMSE. We can deduce the following:

For using the best estimation model, IMA (1, 1), GLS estimation method performs nearly as efficient as the other estimation methods. For example, the RMSE for IMA (1,1) using ML, CSS-ML, CSS, and GLS equal 49.8314, 49.8314, 49.8315, and 50.9545, respectively.

For other over estimation models, the measures of forecast accuracy criteria using the GLS estimation method do not differ substantially comparing to the other estimation methods. For example, the RMSE for ARIMA (1,1,1) using ML, CSS-ML, CSS, and GLS equal 49.8210, 49.8210, 49.8166, and 50.9545, respectively.

For the transformed data using the first difference, OLS performs as nearly as the other estimation methods. The RMSE equals 50.9544.

However, for the original data, OLS performs much less efficient than the other estimation methods, resulting poor forecasting accuracy. The RMSE equals 170.174.

Table 4.1. RMSE for Real Data

Method	IMA(1,1)	ARIMA(1,1,1)	ARIMA(2,1,2)
ML	49.8314	49.8210	49.6809
CSS-ML	49.8314	49.8210	49.6810
CSS	49.8315	49.8166	49.6767
GLS	50.9545	50.9545	50.9545
OLS*	50.9544		
OLS**	170.174		

* OLS for the transformed data

* OLS for the original data

4.2. Comparisons of Estimation Methods for Simulated Data

In this section, we consider the robustness of the four

estimation methods. We compare the best forecasting among some of the estimation methods such as CSS-ML, ML, CSS, GLS, and classical OLS. This simulation will be conducted to examine the sensitivity of the selected estimation methods to model forecasting. In particular, what is the appropriate estimation methods for selecting the most adequate forecasting model. This section displays the results of simulation study.

Three finite sample sizes (50, 164, and 500) are generated from three different *ARIMA* models, namely: IMA(1,1), ARIMA(1,1,1) and ARIMA(2,1,2) with estimation parameters as mentioned in Table (4.2). In each case 1000 of simulations with 1000 replications were generated by R statistical software package and the value of RMSE were computed for each selected model, and sample size, and estimation method. The complete simulation results are presented in Table 4.2.

For using the most appropriate estimation model, IMA (1,1), GLS estimation method performs nearly as efficient as the other estimation methods. For example, the RMSE for IMA (1,1) using ML, CSS-ML, CSS, and GLS for N=50 equal 0.973789950, 0.973789952, 0.974568426, and 0.991164328, respectively.

For other over estimation models, the measures of forecast accuracy criteria using the GLS estimation method does not differ substantially comparing to the other estimation methods. For example, the RMSE for ARIMA (1,1,1) using ML, CSS-ML, CSS, and GLS for N=500 equal 0.997086593, 0.997084667, 0.996242168, and 1.022727195, respectively.

For the original data, OLS performs much less efficient than the other estimation methods, resulting poor forecasting accuracy. For example, when N=50 for IMA(1,1), the RMSE equals 6.623496357. However, For the transformed data using the first difference, OLS performs as nearly as the other estimation methods. The RMSE in this case equals 1.021243491.

Table 4.2. RMSE for Simulated Data

Model	Method	N=50	N=163	N=500
IMA (1,1)	ML	0.973789950	0.992241315	0.998298016
	CSS-ML	0.973789952	0.992241315	0.998298016
	CSS	0.974568426	0.992415326	0.998351545
	GLS	0.991164328	1.014788963	1.023119426
	OLS*	0.989519910	1.012996109	1.021243491
ARIMA (1,1,1)	OLS**	2.089517661	3.776981098	6.623496357
	ML	0.958924656	0.988588205	0.997086593
	CSS-ML	0.958697152	0.988560292	0.997084667
	CSS	0.938562761	0.985638071	0.996242168
	GLS	0.992092843	1.014741181	1.022727195
	OLS*	0.988605362	1.012864774	1.020507497

Model	Method	N=50	N=163	N=500
ARIMA (2,1,2)	OLS**	2.051295223	3.705580936	6.489403505
	ML	0.913104246	0.979896063	0.99505171
	CSS-ML	0.912136161	0.978561152	0.994940878
	CSS	0.836939029	0.970016181	0.992736025
	GLS	0.995749801	1.019890774	1.027462515
	OLS*	0.984947433	1.015065506	1.022514775
	OLS**	2.217033580	4.127434836	7.300410063

* OLS for the original data
 * OLS for the transformed data

Definition 4.2. The efficiency of GLS estimates relative to that of MLE in terms of the MSE of the data, $\hat{\zeta}$, is given by

$$\hat{\zeta} = \frac{\sum_{i=1}^k (X_{t+k} - \hat{X}_{t+k,t})_{GLS}^2}{\sum_{i=1}^k (X_{t+k} - \hat{X}_{t+k,t})_{MLE}^2} \quad (4.2)$$

where k is the number of simulations.

A ratio less than one indicates that the GLS estimates is more efficient than MLE, and if is close to one, then the GLS estimate is nearly as efficient as MLE estimates.

Table 4.3 shows the complete simulation results of the ratios of the GLS estimation method relative to the other methods in terms of the MSE of the data, $\hat{\zeta}$ in (4.2). The table presents the results for the three sample sizes considered, as well as all five selected estimation method for each of the selected model.

First, we see that regardless of the sample size, if the model is correctly specified, i.e. IMA(1,1), GLS estimation method performs nearly as efficiently as ML, CSS-ML, and CSS. For example when N=500, the relative efficiency of GLS to that of ML, $\hat{\zeta} = 1.02486$.

For other over estimation models, GLS estimation method does not differ significantly comparing to the other estimation methods. For example when N=50 for ARIMA (1,1,1), the relative efficiency of GLS to that of ML, $\hat{\zeta} = 1.03459$.

In addition, as the sample size increases, the efficiency of GLS mimics to that of the other estimation methods. For example for ARIMA (1,2,1), the relative efficiency of GLS to that of ML, $\hat{\zeta} = 1.09051, 1.04082, \text{ and } 1.03257$ when N=50, 163, and 500, respectively.

To further demonstrate the efficiency of GLS, consider OLS for the original data for all selected sample sizes and all selected models. GLS is much more efficient in forecasting than the OLS estimation method. In other words, OLS for the original data performs poorly as shown in Table (4.3). For example, the relative efficiency of GLS to that of OLS when T = 50, 163, and 500 for IMA(1,1)

equals $\hat{\zeta} = 0.47435, 0.26868, \text{ and } 0.15447$, respectively. The superiority of GLS over OLS is due to the fact that GLS has a smaller variance and the autocorrelated nature of disturbances is accounted for in the GLS. According to the Generalized Gauss Markov Theorem, the GLS estimator provides the *Best Linear Unbiased Estimator* (BLUE) of the regression coefficient.

However, OLS for the transformed data performs nearly as efficiently as GLS for all selected sample sizes and all selected models. This result is not surprising since the autocorrelated nature of disturbances is accounted for in the transformed data. For example, the relative efficiency of GLS to that of OLS for transformed data when $T = 50, 163, \text{ and } 500$ for IMA(1,1) equals $\hat{\zeta} = 1.00166, 1.00177, \text{ and } 1.00184$, respectively.

Table 4.3. Efficiency for RMSEs of the GLS Estimators Relative to other Methods

Model	Method	N=50	N=163	N=500
IMA (1,1)	ML	1.01784	1.02272	1.02486
	CSS-ML	1.01784	1.02272	1.02486
	CSS	1.01703	1.02254	1.02481
	OLS*	1.00166	1.00177	1.00184
	OLS**	0.47435	0.26868	0.15447
ARIMA (1,1,1)	ML	1.03459	1.02645	1.02572
	CSS-ML	1.03483	1.02648	1.02572
	CSS	1.05703	1.02953	1.02658
	OLS*	1.00353	1.00185	1.00218
	OLS**	0.48364	0.27384	0.15760
ARIMA (2,1,2)	ML	1.09051	1.04082	1.03257
	CSS-ML	1.09167	1.04224	1.03269
	CSS	1.18975	1.05142	1.03498
	OLS*	1.01097	1.00475	1.00484
	OLS**	0.44914	0.24710	0.14074

* OLS for the original data
 * OLS for the transformed data

Table 4.4 shows the actual and forecasting results with lower and upper 95% confidence interval for the daily closing price Alquds Index using IMA (1,1) model.

Table 4.4. Actual and Forecasting with lower and upper 95% confidence interval of IMA (1,1) model for daily closing price Alquds Index

Month	Lower	Forecast	Upper	Actual
May-11	346.51	500.46	654.41	498.8
Jun-11	308.81	503.39	697.97	492.71
Jul-11	278.23	506.32	734.40	494.77
Aug-11	251.98	509.25	766.51	491.03
Sep-11	228.72	512.18	795.63	488.2
Oct-11	207.69	515.10	822.52	NA
Nov-11	188.39	518.03	847.68	NA
Dec-11	170.49	520.96	871.43	NA

Figure 4.1 shows the data and forecasting results with lower and upper 95% confidence interval of IMA (1,1) model for the daily closing price Alquds Index.

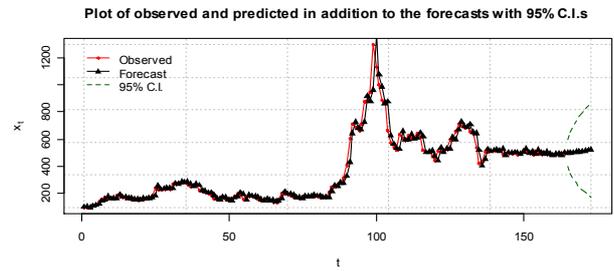


Figure 4.1. Data and Forecasting with lower and upper 95% confidence interval of IMA (1,1) model for daily closing price Alquds Index

The actual values from May 2011 to Sep. 2011 are observed and added to Figure 4.1. to see if these points fall within the confidence interval. Figure 4.2 illustrates this new data. It is clear that the new actual values located within the confidence interval indicating an excellent forecasting for IMA (1,1) model based on GLS estimation method.

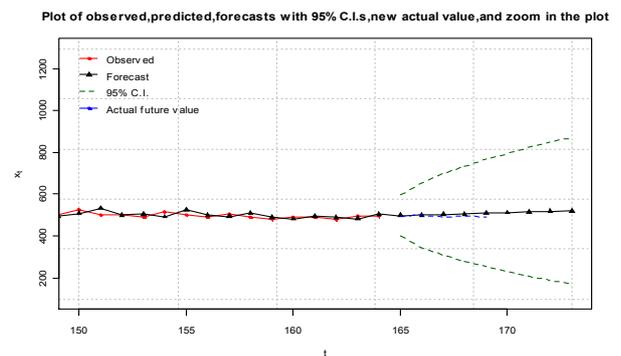


Figure 4.2. Full Data and Forecasting with lower and upper 95% confidence interval of IMA (1,1) model for daily closing price Alquds Index

5. Conclusion and Future Research

In this section, we introduce conclusion of the main findings and offer suggestions for future research for using GLS to generate forecasts in regression models with auto-correlated disturbances.

5.1. Conclusion

This paper has proposed five different estimation methods, namely: ML, CSS-ML, CSS, GLS, and OLS. We introduced the accuracy of the forecasting results based on RMSE using ARIMA model on real data for daily closing price Alquds index in Palestine and simulation technique. The main findings of this paper are as follows:

The results of both real data and simulation reveal that GLS estimation method is comparable to other complicated estimation methods such as MLE procedures which often require inversion of large matrices and hence preferable as a robust estimation and forecasting method.

If the model is correctly specified, i.e. IMA (1,1), GLS estimation method performs nearly as efficiently as the other estimation methods as ML, CSS-ML, and CSS. For other over estimation models, the GLS estimation method does not differ significantly comparing to the other estimation methods. In addition, as the sample size increases, the efficiency of GLS mimics to that of the other estimation methods.

For the original data, GLS is much more efficient in forecasting than the OLS estimation method. However, for the transformed data, OLS performs as nearly as the other estimation methods. Finally, for forecasting purposes there is not much gained in trying to identifying the exact order and form of the auto-correlated disturbances by using GLS estimation method.

5.2. Future Research

The plane for future research can be split into the following: First: Examine the effect on forecasting performance for other different models such as ARCH, and GARCH models and conduct applications in economic and financial forecasting. A second important consideration is the estimation of the standard errors of the estimators. It is unclear, however, how the variance estimators for GLS estimation behave for complicated time series models. Study the impact that the variance estimators may have on inference based on the GLS estimator.

Acknowledgment

We would deeply like to thank the referees for their valuable comments and suggestions on earlier draft of this paper.

References

- [1] Abraham, B. and Ledolter, J. (2005). *Statistical methods for forecasting*. John Wiley & Sons, Inc.
- [2] Cochrane, J. H. (2005). *Time Series for Macroeconomics and Finance*. University of Chicago.
- [3] Cryer, J. and Chan, K. (2008). *Time Series Analysis With Applications in R*, 2nd Ed. Springer Science + Business Media, LLC.
- [4] Everitt, B. S. and Hothorn, I. (2010). *A Handbook of Statistical Analyses Using R*, 2nd Ed. Taylor and Francis Group, LLC.
- [5] Findley, D. (2003). *Properties of Forecast Errors and Estimates of Misspecified RegARIMA and Intermediate Memory Models and the Optimality of GLS for One-Step-Ahead Forecasting*. Methodology and Standards Directorate U.S. Bureau of the Census Washington D.C. 20233.
- [6] Fox, J. (2002). *Time-Series Regression and Generalized Least Squares*. Appendix to An R and S-PLUS Companion to Applied Regression.
- [7] Koreisha, G. and Fang, Y. (2004). *Forecasting with serially correlated regression models*. Journal of Statistical Computation and Simulation Vol. 74(9), 625–649.
- [8] Lee, J. and Lund, R. (2004). *Revisiting simple linear regression with autocorrelated errors*, Biometrika, Vol. 91 (1): 240-245. doi: 10.1093/biomet/91.1.240
- [9] Ojo, J. and Olatayo, T. (2009). *On the Estimation and Performance of Subset Autoregressive Integrated Moving Average Models*. European Journal of Scientific Research ISSN 1450-216X Vol.28 (2), 287-293.
- [10] Safi, S. (2004). *The efficiency of OLS in the presence of auto-correlated disturbances in regression models*. PhD dissertation, American University Washington, D.C. 20016.
- [11] Shittu, O.I. and Asemota, M.J. (2009). *Comparison of Criteria for Estimating the Order of Autoregressive Process: A Monte Carlo Approach*. European Journal of Scientific Research ISSN 1450-216X Vol. 30(3), 409-416.