Introduction to Quantitative Methods
Econometric Theory Review

Dr. Aris Kartsaklas
aris.kartsaklas@brunel.ac.uk

Univariate Time Series with Applications in Finance
Autumn Term 2011/2012

1 Financial Time Series and Regression Analysis (Econometric Theory Review)

- In many financial applications the relationship between two time series is of major interest

  The **market model** is an example the relates the return of an individual stock to the return of a market index

  The **term structure of interest rates** is another is another example in which the time evolution of the relationship between interest rates with different maturities is investigated.

  A simple bivariate regression model with time series errors is

  (1) $y_t = \alpha + \beta x_t + u_t \quad , t = 1, 2, ..., T$

  where y_t and x_t are two time series and u_t denotes the error term

- Simple estimation and tests of the CAPM can be conducted using an equation of the form of (1), but Arbitrage Pricing Theory (APT) does not pre-suppose that there is only a single factor affecting stock returns.

  Stock returns might be purported to depend on their sensitivity to unexpected changes in:

  (1) inflation

  (2) the differences in returns on short- and long-dated bonds

  (3) industrial production

  (4) default risks.

- It is very easy to generalise the simple model to one with k regressors (independent variables). Equation (1) becomes

  (2) $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + ... + \beta_k x_{kt} + u_t \quad , t = 1, 2, ..., T$

- So the variables $x_{2t}, x_{3t}, ..., x_{kt}$ are a set of $k-1$ explanatory variables which are thought to influence y, and the coefficient estimates $\beta_1, \beta_2, ..., \beta_k$ are the parameters which quantify the effect of each of these explanatory variables on y.

- Each coefficient measures the average change in the dependent variable per unit change in a given independent variable, holding all other independent variables constant at their average values.

Testing single hypotheses: the t-test

- Recall that the formula under a test of significance approach to hypothesis testing using a t-test for the slope parameter was

  (3) $test\ statistic = \frac{\hat{\beta}-\beta}{SE(\hat{\beta})}$

  with the obvious adjustments to test a hypothesis about the intercept

- If the test is $H_0 : \beta = 0$, $H_1 : \beta \neq 0$,

  i.e. a test that the population parameter is zero against a two-sided alternative, this is known as a t-ratio test.

  Since $\beta = 0$, the expression in (3) reduces to

  $test\ statistic = \frac{\hat{\beta}}{SE(\hat{\beta})}$

- Thus the ratio of the coefficient to its standard error is known as the t-ratio or t-statistic and it behaves as a t-distributed variable with $T - 2$ degrees of freedom

Testing multiple hypotheses: the F-test

- If you wanted to determine whether a restriction that the coefficient values for β_2 and β_3 are both unity could be imposed, so that an increase in either one of the two variables x_2 or x_3 would cause y to rise by one unit?

- Under the F-test framework, two regressions are required, known as the **unrestricted** and the **restricted** regressions.

  The unrestricted regression is the one in which the coefficients are freely determined by the data

  The restricted regression is the one in which the coefficients are restricted, i.e. the restrictions are imposed on some β's.

- Thus the F-test approach to hypothesis testing is also termed restricted least squares, for obvious reasons.

  The residual sums of squares from each regression are determined, and the two residual sums of squares are 'compared' in the test statistic

- The F-test statistic for testing multiple hypotheses about the coefficient estimates is given by
- (3) $test\ statistic = \frac{RRSS - URSS}{URSS} \times \frac{T-k}{m}$
- where

  $URSS$ = residual sum of squares from unrestricted regression

  $RRSS$ = residual sum of squares from restricted regression

  m = number of restrictions

  T = number of observations

  k = number of regressors in unrestricted regression

- Recall that OLS estimation involved choosing the model that minimised the residual sum of squares, with no constraints imposed.
- If the residual sum of squares increased not much after the restrictions were imposed

  → it would be concluded that the restrictions were supported by the data.

- If the residual sum of squares increased considerably after the restrictions were imposed

  → it would be concluded that the restrictions were not supported by the data (hypothesis should be rejected)

- The test statistic follows the F-distribution under the null hypothesis

  The value of the degrees of freedom parameters for the F-test are $F(m, T-k)$

  m is the number of restrictions imposed on the model

  $(T-k)$, the number of observations less the number of regressors for the unrestricted regression

- The appropriate critical value will be in column m, row $(T-k)$ of the F-distribution tables.

Assumptions of the model

- Recall that five assumptions were made relating to the Classical Linear Regression Model (CLRM).

  These were required to show that the estimation technique, ordinary least squares (OLS), had a number of desirable properties, and also so that hypothesis tests regarding the coefficient estimates could validly be conducted.

  Specifically, it was assumed that

  (1) $E(u_t) = 0$

  The first assumption required is that the average value of the errors is zero

  If a constant term is included in the regression equation, this assumption will never be violated.

  (2) $\text{Var}(u_t) = \sigma^2 < \infty$

  This assumption is that the variance of the errors is constant, σ^2 – it is also known as the assumption of homoscedasticity.

  If the errors do not have a constant variance, they are said to be heteroscedastic

  White's (1980) general test is used to detect heteroscedasticity.

- **What are the consequences of using OLS in the presence of heteroscedasticity?**

  OLS estimators will still give unbiased (and also consistent) coefficient estimates, but they are no longer **BLUE** (standard error estimates could be wrong)

  In general, the OLS standard errors will be too large for the **intercept** when the errors are heteroscedastic

  The effect of heteroscedasticity on the **slope** standard errors will depend on its form.

  If the variance of the errors is positively related to the square of an explanatory variable (which is often the case in practice), the OLS standard error for the slope will be too low.

  The OLS slope standard errors will be too low when the variance of the errors is positively related to an explanatory variable

  The OLS slope standard errors will be too big when the variance of the errors is inversely related to an explanatory variable

- **How do we deal with heteroscedasticity?**

  1. If the form (i.e. the cause) of the heteroscedasticity is known, then an alternative estimation method which takes this into account is called the **generalised least squares** (GLS)

  2. Transforming the variables into logs or reducing by some other measure of 'size' has the effect of re-scaling the data to 'pull in' extreme observations.

  3. Using heteroscedasticity-consistent standard error estimates. Following White (1980) if the variance of the errors is positively related to the square of an explanatory variable the standard errors for the slope coefficients are increased relative to the usual OLS standard errors

  → this would make hypothesis testing more 'conservative', so that more evidence would be required against the null hypothesis before it would be rejected.

  (3) $\text{Cov}(u_i, u_j) = 0$ for $i \neq j$

  Assumption 3 is that the covariance between the error terms (or cross-sectionally, for that type of data) over time is zero.

  If the errors are not uncorrelated with one another, it would be stated that they are '**autocorrelated**' or '**serially correlated**'.

  In order to test for autocorrelation, it is necessary to investigate whether any relationships exist between the current value of $\hat{u}_t$, and any of its previous values, $\hat{u}_{t-1}$, $\hat{u}_{t-2}$, ...

  Durbin–Watson (DW) is a test for first order autocorrelation – i.e. it tests only for a relationship between an error and its immediately previous value

  Breusch–Godfrey test is a more general test for autocorrelation up to the r-th order.

- **What are the consequences of ignoring autocorrelation if it is present?**

  OLS coefficient estimates are still unbiased, but they are inefficient, i.e. they are not **BLUE** (standard error estimates could be wrong)

  In the case of positive serial correlation in the residuals, the OLS standard error estimates will understate their true variability.

  → this would lead to an increase in the probability of **type I error** – that is, a tendency to reject the null hypothesis sometimes when it is correct.

  Furthermore, R^2 is likely to be inflated relative to its 'correct' value if autocorrelation is present but ignored, since residual autocorrelation will lead to an underestimate of the true error variance (for positive autocorrelation)

- **How do we deal with autocorrelation?**

1. If the form of the autocorrelation is known, it would be possible to use a GLS procedure. Cochrane–Orcutt procedure works by assuming a particular form for the structure of the autocorrelation.

2. An alternative approach to dealing with residual autocorrelation would be to use appropriately modified standard error estimates. Newey and West (1987) develop a variance–covariance estimator that is consistent in the presence of both heteroscedasticity and autocorrelation

3. Sargan, Hendry and Mizon, suggests that serial correlation in the errors arises as a consequence of '**misspecified dynamics**'. Autocorrelation in the residuals is often caused by a dynamic structure in y that has not been modelled and so has not been captured in the fitted values.

What is required is a dynamic model that allows for this extra structure in y.

Models containing lags of the explanatory variables (but no lags of the explained variable) are known as **distributed lag model**

Specifications with lags of both explanatory and explained variables are known as **autoregressive distributed lag** (ADL) models

The use of lagged variables in a regression model does, however, bring with it additional problems:

a. Inclusion of lagged values of the dependent variable violates the assumption that the explanatory variables are non-stochastic (assumption 4 of the CLRM)

b. What does an equation with a large number of lags actually mean?

(4) $\text{Cov}(u_t, x_t) = 0$

The OLS estimator is consistent and unbiased in the presence of stochastic regressors, provided that the regressors are not correlated with the error term of the estimated equation

However, if one or more of the explanatory variables is contemporaneously correlated with the disturbance term, the OLS estimator will not even be consistent.

This results from the estimator assigning explanatory power to the variables where in reality it is arising from the correlation between the error term and y_t .

(5) $u_t \sim N(0, \sigma^2)$

Recall that the **normality assumption** is required in order to conduct single or joint hypothesis tests about the model parameters

One of the most commonly applied tests for normality is the **Bera–Jarque** (hereafter BJ) test.

BJ uses the property of a normally distributed random variable that the entire distribution is characterised by the first two moments – the mean and the variance.

The standardised third and fourth moments of a distribution are known as its skewness and kurtosis.

Skewness measures the extent to which a distribution is not symmetric about its mean value

Kurtosis measures how fat the tails of the distribution are.

A normal distribution is not skewed and is defined to have a coefficient of kurtosis of 3.

It is possible to define a coefficient of excess kurtosis, equal to the coefficient of kurtosis minus 3

→ a normal distribution will thus have a coefficient of excess kurtosis of zero.

Bera and Jarque (1981) formalise these ideas by testing whether the coefficient of skewness and the coefficient of excess kurtosis are jointly zero.

- **What should be done if evidence of non-normality is found?**

  For sample sizes that are sufficiently large, appealing to a central limit theorem, the test statistics will asymptotically follow the appropriate distributions even in the absence of error normality.

  In economic or financial modelling, it is quite often the case that one or two very extreme residuals cause a rejection of the normality assumption.

  Such observations would appear in the tails of the distribution, and would therefore lead u^4, which enters into the definition of kurtosis, to be very large.

  Such observations that do not fit in with the pattern of the remainder of the data are known as **outliers**.

  If this is the case, one way to improve the chances of error normality is to use **dummy variables** or some other method to effectively remove those observations

  Dummy variables are also used in modelling seasonality and 'calendar anomalies' in financial data

  Non-normality in financial data could also arise from certain types of heteroscedasticity, known as ARCH.

  → in this case, the nonnormality is intrinsic to all of the data and therefore outlier removal would not make the residuals of such a model normal

Multocolinearity

- A problem occurs when the explanatory variables are very highly correlated with each other, and this problem is known as **multicollinearity**

  It is possible to distinguish between two classes of multicollinearity: perfect multicollinearity and near multicollinearity

  Perfect multicollinearity occurs when there is an exact relationship between two or more variables.

  Two variables are perfectly related to one another contain only enough information to estimate one parameter, not two.

  Trying to invert the $(X'X)$ matrix is difficult since it would not be of full rank (two of the columns would be linearly dependent on one another)

  $\rightarrow$ so that the inverse of $(X'X)$ would not exist and hence the OLS estimates $\hat{\beta} = (X'X)^{-1}Xy$ could not be calculated.

  Near multicollinearity would arise when there is a non-negligible, but not perfect, relationship between two or more of the explanatory variables.

  How do we deal with near multicolinearity?

  1. Ignore it, if the model is otherwise adequate, i.e. statistically and in terms of each coefficient being of a plausible magnitude and having an appropriate sign

  2. Drop one of the collinear variables unless there are strong a priori theoretical reasons for including both variables in the model.

  Also, if the removed variable was relevant in the data generating process for y, an omitted variable bias would result

  3. Transform the highly correlated variables into a ratio and include only the ratio and not the individual variables in the regression.

  Again, this may be unacceptable if financial theory suggests that changes in the dependent variable should occur following changes in the individual explanatory variables, and not a ratio of them.

4. An increase in the sample size will usually lead to an increase in the accuracy of coefficient estimation and consequently a reduction in the coefficient standard errors, thus enabling the model to better dissect the effects of the various explanatory variables on the explained variable.

A further possibility, therefore, is for the researcher to go out and collect more data – for example, by taking a longer run of data, switching to a higher frequency of sampling and pooling cross-sectional observations over time.

5. Principal components analysis is a another technique to address the issue of near multicolinearity. Specifically, if there are k explanatory variables in the regression model, PCA will transform them into k uncorrelated new variables.

Adopting the wrong functional form

- A further implicit assumption of the classical linear regression model is that the appropriate 'functional form' is linear. This means that the appropriate model is assumed to be linear in the parameters, and that in the bivariate case, the relationship between y and x can be represented by a straight line.

  Whether the model should be linear can be formally tested using **Ramsey's (1969) RESET tes**t, which is a general test for misspecification of functional form.

  Non-linear models in the parameters typically preclude the use of OLS, and require the use of a non-linear estimation technique.

Ommision of an important variable

- What would be the effects of excluding from the estimated regression a variable that is a determinant of y?

  The estimated coefficients on all the other variables will be biased and inconsistent unless the excluded variable is uncorrelated with all the included variables.

  Even if this condition is satisfied, the estimate of the coefficient on the constant term will be biased, which would imply that any forecasts made from the model would be biased.

  The standard errors will also be biased (upwards), and hence hypothesis tests could yield inappropriate inferences.

Inclusion of an irrelavant variable

- The consequence of including an irrelevant variable would be that the coefficient estimators would still be consistent and unbiased, but the estimators would be inefficient.

  This would imply that the standard errors for the coefficients are likely to be inflated relative to the values which they would have taken if the irrelevant variable had not been included.

  Variables which would otherwise have been marginally significant may no longer be so in the presence of irrelevant variables.

  In general, it can also be stated that the extent of the loss of efficiency will depend positively on the absolute value of the correlation between the included irrelevant variable and the other explanatory variables.

Parameter stability tests

- So far, regressions of a form such as

  $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$

  embody the implicit assumption that the parameters (β_1, β_2 and β_3) are constant for the entire sample, both for the data period used to estimate the model, and for any subsequent period used in the construction of forecasts.

  This implicit assumption can be tested using parameter stability tests.

  The idea is essentially to split the data into sub-periods and then to estimate up to three models, for each of the sub-parts and for all the data and then to 'compare' the RSS of each of the models.

  There are two types of test that will be considered, namely the **Chow** (analysis of variance) **test** and **predictive failure tests**.

Univariate time series models

- A class of specifications where one attempts to model and to predict financial variables using only information contained in their own past values and possibly current and past values of an error term.

  This practice can be contrasted with structural models, which are multivariate in nature, and attempt to explain changes in a variable by reference to the movements in the current or past values of other (explanatory) variables.

- Time series models are usually a-theoretical, implying that their construction and use is not based upon any underlying theoretical model of the behaviour of a variable.

  Instead, time series models are an attempt to capture empirically relevant features of the observed data that may have arisen from a variety of different (but unspecified) structural models.

- An important class of time series models is the family of **AutoRegressive Integrated Moving Average** (ARIMA) models, usually associated with Box and Jenkins (1976).

- It may be that the variables thought to drive movements of y_t are not observable or not measurable, or that these forcing variables are measured at a lower frequency of observation than y_t.

  For example, y_t might be a series of daily stock returns, where possible explanatory variables could be macroeconomic indicators that are available monthly.

- **Stock Market returns**

  There are two methods used to calculate returns from a series of prices, and these involve the formation of simple returns, and continuously compounded returns, which are achieved as follows:

  Simple Returns

  $R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100$

  Continuously compounded returns

  $r_t = 100 \times \ln\left(\frac{P_t}{P_{t-1}}\right)$

- If the asset under consideration is a stock or portfolio of stocks, the total return to holding it is the sum of the capital gain and any dividends paid during the holding period.

  Researchers often ignore any dividend payments but doing so will lead to an underestimation of the total returns that accrue to investors.

  This is likely to be negligible for very short holding periods, but will have a severe impact on cumulative returns over investment horizons of several years.

- Ignoring dividends will also have a distortionary effect on the crosssection of stock returns.

  For example, ignoring dividends will imply that '**growth**' stocks, with large capital gains will be inappropriately favoured over '**value**' stocks (e.g. utilities and mature industries) that pay high dividends.

- Alternatively, it is possible to adjust a stock price time series so that the dividends are added back to generate a total return index.

  If P_t were a total return index, returns generated using either of the two formulae presented above thus provide a measure of the total return that would accrue to a holder of the asset during time t.

- There is, however, also a disadvantage of using the log-returns. The simple return on a portfolio of assets is a weighted average of the simple returns on the individual assets
  $R_{pt} = \sum_{i=1}^{n} w_i R_{it}$
- But this does not work for the continuously compounded returns, so that they are not additive across a portfolio.

  The fundamental reason why this is the case is that the **log of a sum is not the same as the sum of a log**, since the operation of taking a log constitutes a non-linear transformation.

- Calculating portfolio returns in this context must be conducted by first estimating the value of the portfolio at each time period and then determining the returns from the aggregate portfolio values.

  Or alternatively, if we assume that the asset is purchased at time $t-K$ for price P_{t-K} and then sold K periods later at price P_t , then if we calculate simple returns for each period, R_t, R_{t+1}, . . . , R_K, the aggregate return over all K is
  $R_{kt} = \frac{P_t - P_{t-K}}{P_{t-K}} = \frac{P_t}{P_{t-K}} - 1 = \left[\frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times ... \times \frac{P_{t-K+1}}{P_{t-K}}\right] - 1 = [(1+R_t) \times (1+R_{t-1}) ... (1+R_{t-K+1})] - 1$

Stationarity

- **A strictly stationary process**

  A strictly stationary process is one where, for any t_1, t_2, . . . , $t_T \in Z$, any $k \in Z$ and $T = 1, 2, ...$

  $Fy_{t_1}, y_{t_2}, ..., y_{t_T}(y_1, ..., y_T) = Fy_{t_1+k}, y_{t_2+k}, ..., y_{t_T+k}(y_1, ..., y_T)$

  where F denotes the joint distribution function of the set of random variables

  A series is strictly stationary if the distribution of its values remains the same as time progresses, implying that the probability that y falls within a particular interval is the same now as at any time in the past or the future.

- **A weakly stationary process**

  A series is said to be weakly or covariance stationary if it has a constant mean, a constant variance and a constant autocovariance structure

  The autocovariances determine how y is related to its previous values, and for a stationary series they depend only on the difference between t_1 and t_2, so that the covariance between y_t and y_{t-1} is the same as the covariance between y_{t-10} and y_{t-11}, etc.

  The moment $E\left(y_t - E(y_t)\right)\left(y_{t-s} - E(y_{t-s})\right) = \gamma_s$, $s = 0, 1, 2, ...$is known as the autocovariance function

  It is thus more convenient to use the **autocorrelations** ($\tau_s = \gamma_s/\gamma_0$, $s = 0, 1, 2, ...$), which are the autocovariances normalised if we divide by the variance

- **A white noise process**

  A white noise process has constant mean and variance, and zero autocovariances, except at lag zero

  Another way to state this last condition would be to say that each observation is uncorrelated with all other values in the sequence

  Hence the autocorrelation function for a white noise process will be zero apart from a single peak of 1 at $s = 0$

- It is also possible to test the joint hypothesis that all m of the τ_k correlation coefficients are simultaneously equal to zero using the Q-statistic developed by **Box and Pierce (1970)**

  $Q = T\sum_{k=1}^{m} \hat{\tau}_k^2$

  where T = sample size, m = maximum lag length.

  The correlation coefficients are squared so that the positive and negative coefficients do not cancel each other out.

  The sum of squares of independent standard normal variates is itself a χ^2 variate with degrees of freedom equal to the number of squares in the sum

  Thus, it can be stated that the Q-statistic is asymptotically distributed as a χ_m^2 under the null hypothesis that all m autocorrelation coefficients are zero.

  As for any joint hypothesis test, only one autocorrelation coefficient needs to be statistically significant for the test to result in a rejection.

  However, the Box–Pierce test has poor small sample properties, implying that it leads to the wrong decision too frequently for small samples.

- A variant of the Box–Pierce test with better small sample properties is the **Ljung–Box (1978)** statistic

  $Q = T(T+2)\sum_{k=1}^{m} \hat{\tau}_k^2$

  It should be clear from the form of the statistic that asymptotically (that is, as the sample size increases towards infinity), the $(T+2)$ and $(T-k)$ terms in the Ljung–Box formulation will cancel out, so that the statistic is equivalent to the Box–Pierce test.

  This statistic is very useful as a portmanteau (general) test of linear dependence in time series.

Moving Average processes (MA)

- The simplest class of time series model that one could entertain is that of the moving average process.

  Let u_t $(t = 1, 2, 3, ...)$ be a white noise process with $E(u_t) = 0$ and $\text{Var}(u_t) = \sigma^2$. Then

  $y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + ...+ \theta_q u_{t-q}$

  is a q-th order moving average mode, denoted $MA(q)$. This can be expressed using sigma notation as

  $y_t = \mu + \sum_{i=1}^{q} \theta_i u_{t-i} + u_t$

  A moving average model is simply a linear combination of white noise processes, so that y_t depends on the current and previous values of a white noise disturbance term.

AutoRegressive processes (AR)

- An autoregressive model is one where the current value of a variable, y, depends upon only the values that the variable took in previous periods plus an error term

  An autoregressive model of order p, denoted as $AR(p)$, can be expressed as

  $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ...+ \phi_p y_{t-p} + u_t$

  where u_t is a white noise disturbance term. A manipulation of previous expression will be required to demonstrate the properties of an autoregressive model.

  This expression can be written more compactly using sigma notation

  $y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t$

ARMA processes

- By combining the $AR(p)$ and $MA(q)$ models, an $ARMA(p,q)$ model is obtained. Such a model states that the current value of some series y depends linearly on its own previous values plus a combination of current and previous values of a white noise error term.

  The model could be

  $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ...+ \ \phi_p y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + ... + \ \theta_q u_{t-q} + u_t$

  or

  $y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i u_{t-i} + u_t$

  with $E(u_t) = 0$, $E(u_t^2) = \sigma^2$, $E(u_t u_s) = 0$, $t = s$

  The characteristics of an $ARMA$ process will be a combination of those from the autoregressive (AR) and moving average (MA) parts.

  For more information about the invertibility and stationarity conditions of AR, MA and ARMA processes please see Chapter 5 of Chris Brook's book 'Introductory Econometric for Finance'

- **Forecasting**

  Some examples in finance of where forecasts from econometric models might be useful include

  Forecasting tomorrow's return on a particular share

  Forecasting the price of a house given its characteristics

  Forecasting the riskiness of a portfolio over the next year

  Forecasting the volatility of bond returns

  Forecasting the correlation between US and UK stock market movements tomorrow

  Forecasting the likely number of defaults on a portfolio of home loans.

- It is useful to distinguish between two approaches to forecasting:

  Econometric (structural) forecasting – relates a dependent variable to one or more independent variables.

  Such models often work well in the long run, since a long-run relationship between variables often arises from no-arbitrage or market efficiency conditions.

  Examples of such forecasts would include return predictions derived from arbitrage pricing models, or long-term exchange rate prediction based on purchasing power parity or uncovered interest parity theory.

  Time series forecasting – involves trying to forecast the future values of a series given its previous values and/or previous values of an error term.

  A one-step-ahead forecast is a forecast generated for the next observation only, whereas multi-step-ahead forecasts are those generated for 1, 2, 3, . . . , s steps ahead, so that the forecasting horizon is for the next s periods.

  For more information about how to forecast with AR, MA and ARMA models please see Chapter 5 of Chris Brook's book 'Introductory Econometric for Finance'

References

- Chris Brooks, Introductory Econometrics for Finance, Cambridge University Press, Second Edition.

  Chapters 1-5