

## *PROJECT TEMPLATE ON INTEREST RATES AND OTHER ECONOMIC TIME SERIES*

This project template illustrates ARIMA modeling for interest rates, inflation, unemployment rates, and other macroeconomic indices. It uses three month Treasury bills from the NEAS web site as a sample time series, and it explains the applications to other interest rates and economic indices.

This project template explains how to fit ARIMA processes to financial and economic time series. It provides guidance for many student projects, not just those on interest rates.

*Illustration:* You may use overnight LIBOR rates or corporate bond spreads for your student project. Adapt this project template to your time series.

The illustrative worksheets on sample autocorrelation functions, correlograms, Durbin-Watson statistic, and Box-Pierce Q statistic use the interest rate observations in this project template. Some of the discussion forum postings on ARIMA modeling use illustrations from this project template. Even if you choose another topic for your student project, this project template clarifies many of the statistical techniques you use.

*Take heed:* The project templates are difficult to compose. Candidates begin their student projects at various levels of expertise.

- Some understand the course material and are looking for good topics.
- Others feel lost and need guidance through all steps of the student project.

The student projects must be independent work. We can't say: "Do Step A, then Step B, and so forth," or the SOA would not grant credit for the on-line course. The on-line courses teach the statistics material so that you can apply them to real data.

The project templates on the discussion forum lead you through statistical analysis.

- If you know the statistics material, you may find some of the information to be repetitive. We repeat some items in different postings, since it takes time to absorb the concepts.
- If you are lost, you may find that the instructions seem incomplete. If you have trouble grasping the concepts, copy the data sets to an Excel workbook and recreate the analyses described in the discussion forum postings.

If you are worried about successfully completing the student project, take heart. Many candidates feel overwhelmed at first, and they think the student projects will be a drain on their time. The first year, the student projects were indeed difficult for some candidates. But our faculty has since created an excellent array of project templates and other postings that provide guidance for all levels. You need an hour or two to get oriented, and you will soon be working through statistical analysis of real data.

## LEVELS OF GUIDANCE

The project templates describe various methods. The ideal method depends on your data, your hypotheses, and your assumptions.

- The project template tells you to graph the interest rates and examine the graph and the correlogram for stationarity, seasonality, trends, and other patterns. Interest rates come in thousands of varieties, and we do not know the patterns in your time series.
- Another part of the project template suggests taking first differences to make the time series stationary. If the initial time series is a random walk, always take differences. If the initial time series is not stationary but is not a random walk, the proper method is not clear. Examine the correlogram of the first differences, but use one of the other methods if possible.
- A third part implies that taking first differences is the wrong approach. Instead, convert the nominal interest rates to real interest rates. Converting a nominal time series to real terms often makes it easier to fit an ARIMA process.
- A fourth part implies that you should divide the time series into three segments and fit different ARIMA process to each one.
- A fifth part tells you to regress the interest rates on other macroeconomic indices (GDP, inflation, money supply) and fit the ARIMA process to the residuals.

*Question:* How can we complete a student project with these instructions? *Answer:*

1. The true cause of interest rate movements is not well known. The textbook proposes several models, and financial economists propose many more. We are not hiding the answer from you; we don't know the pattern of interest rates any better than others do.
2. The student project teaches you the statistical techniques. Many methods may be used to make a time series stationary. You may use two or three methods and see which one produces the best model. We are not telling you what to do. We are explaining the statistical techniques that you may apply to the data.
3. The ARIMA process is not the true explanation of interest rates. It is a proxy that often does well for short periods. The proper analysis varies with the objective.
  - An exact financial model may regress real interest rates on GDP
  - A simple short-term proxy may use first differences of nominal interest rates.
4. If the statistician knows the expected inflation rates in coming months, projections of real interest rates are fine. If the statistician does not know expected inflation, the nominal interest rates must be projected.

*Jacob:* The textbook fits an ARIMA(8,1,4) process to interest rates. Do we fit complex ARIMA processes in the student project?

*Rachel:* For the student projects, we focus on simple processes. If you understand an ARMA(1,1) process, you understand the more complex processes as well. Focus on the basic ARIMA processes and statistical tests.

*Jacob:* For the final exam, we had to learn formulas. We don't have statistical software to compute sample autocorrelations, correlograms, the Box-Pierce Q statistic, and other items used for ARIMA fitting. Do we have to code each of these items?

*Rachel:* This is a course in statistics, not programming. We provide illustrative Excel worksheets that do the number crunching. You copy and paste cell formulas, and you use the Excel built-in functions, but do not spend time on arithmetic.

*Illustration:* We provide interest rates, cell formulas, and VBA macros for the statistical tests. You form correlograms, interpret the graphs, and fit a good model to the time series.

*Take heed:* Excel 2007 differs from previous versions in your access to add-ins. To get the *REGRESSION* add-in or the *SOLVER* add-in for previous versions of Excel, click the *TOOLS* menu and choose the add-in. In Excel 2007, click the *DATA* menu and choose the add-in from the *ANALYSIS* portion of the ribbon. The add-ins work the same way. (The add-ins are produced by other firms, not by Microsoft. They did not change with Windows Vista.) Some of the discussion forum postings have the pre-2007 Excel procedure. Nothing substantive have changed in the statistical add-ins, but the series to clicks to get the add-in differs.

## Step #1: Choose the interest rate time series

We illustrate with three month Treasury bills, for several reasons:

- The data are readily available. We give you an Excel spread-sheet with the time series in a column; you have no need to gather data.
- The data extend over many years. The ARIMA model differs by sub-period.
- The interpretation of interest rate time series is unclear. You have many ways to continue the analysis here. No two candidates will come to the same conclusion.

*Take heed:* We encourage you to search the internet for other time series. Data are easier to gather than many candidates think.

We provide several other time series of interest rates: LIBOR rates at various maturities for the U.S. dollar and some other currencies; other Treasury security rates; corporate bond yields; municipal bond yields; bank prime rates; and others.

Hundreds of other interest rate time series are available on the world wide web. Use the internet search engines (Google, MSN, Yahoo) for “interest rate” or “interest rate history.”

*Take heed:* If you find a web site with good time series data, post a message on the discussion forum. Other candidates appreciate the help.

## Step #2: Correct Missing or Erroneous Elements

The time series starts in January 1931 and continues through December 1935. Three month Treasury bills were auctioned (sold at auction) once a month in these five years. The next auction was in February 1941, leaving 61 months with no auction. We have several ways of dealing with missing periods.

- If the time series is 40 years of data, one year of missing data, then 40 years of data, and if the two periods have similar processes, we might ignore the 12 missing months.
- If one observation is missing, interpolate between the two surrounding values.
- Economic conditions differ by decade. Treasury bill rates decline from 3.25% at the end of 1931 to 0.09% at the end of 1935. During World War II, rates were affected by a war-time budget deficit. Financial economists often analyze post-World War II indices. Select an appropriate period, based on the data available and the time series process.

*Take heed:* The project template on daily temperature uses the first two methods above.

- Weather data are often missing on some days: interpolate for the missing values.
- The daily temperature series may have several months missing: ignore these values.

*Take heed:* Twenty minutes of checking for errors may save you hours of wasted work.

- Most time series on the discussion forum do not have errors. The data have already been checked and corrected. But the time series may have missing periods.
- If you gather data from other web sites, check the data. A missing value may be coded as -999. If you have 100 values of interest rates with an average of 5% and you include a value of -999, this one value skews the time series.
- If you use in-house data for loss cost trends, a coding error may skew your data. If a value seems erroneous, you save time by eliminating it or correcting it.

No method is right or wrong. The question is whether Treasury bill rates followed the same pattern in the early 1930's and early 1940's as in later years.

If we use post-World War II data, we have 666 months of Treasury bill rates. The first month is January 1945, and the last month is June 2000.

*Recommendation:* The choice of time series is arbitrary. Start with a long time series, such as 125 years of daily temperature or 2,500 observations of overnight LIBOR rates. As you complete the project, you may find that the time series process changes mid-way through. Sometimes the reason is clear; other times you see the change but can't explain it. This project template explains methods to make the time series stationary.

### Step #3: Graph the Time Series

Graphing the time series helps you see patterns. You form many graphs and charts for your student project. General advice on graphing:

- Use line graphs for long time series. Set your chart default to line graph (instead of bar graph). Bar graphs are used for small samples, such as a regression analysis of voter types. For time series, use line charts.
- Use deviations from the mean to see patterns. If you are not certain of a trend, graph the deviations from the mean.
- Use graphs of centered moving averages to smooth random fluctuations. If you are not certain of a trend, graph the centered moving averages.
- Label your axes, and use titles and legends, so the course instructor can read the graph. If you use an index for the observations, state the index values on the graph. If your time series uses a month index from 1 to 666, write “Jan 1945 = Month 1.”
- Callouts, arrows, autoshapes, and text boxes help readers understand your graphs.
- You may graph the time series, the first differences, or the sample autocorrelations.

The type of graph depends on the item you analyze.

- To identify trends in a seasonal process, you may graph 12 month moving averages.
- To identify seasonality, you may graph monthly averages over many years.

*Take heed:* Explain the pattern in your time series and the implications for your analysis.

*Illustration:* The graph of 3 month Treasury bill rates for January 1945 – June 2000 has three patterns:

- increasing through 1979, with either cycles or random movements
- volatile for about four years, with no clear pattern
- decreasing through the end of the time series, with cycles or random movements

*Take heed:* The three periods above are not precise. For your student project, give your impression of the time series pattern. The graph uses call-outs to identify possible changes in the time series. The worksheet uses slightly different periods for correlograms. Examine the data and decide what periods seem homogeneous.

This pattern means the time series is not stationary. Corroborate your analysis with the correlogram. A correlogram starting from February 1941 or January 1945 shows a long series of about 200 positive sample autocorrelations.

You have several ways of proceeding. We list them from easiest to preferred methods.

*Take heed:* Statisticians differ on the optimal method.

- The course textbook uses an ARIMA(8,1,4) process for interest rates.
- Other statisticians would consider this process an error: it has no intuitive rationale and the slight improvement in the in-sample fit does not offset the added complexity.

Any statistical method below is fine for the student project.

#### Step #4: DIFFERENCES

Take first differences to make the time series stationary. *Always take first differences if the time series is a random walk.* If the process is multiplicative (instead of additive), first take logarithms and then take first differences.

Many project templates discuss random walks, white noise, and mean reversion. A random walk can be hard to distinguish from a stationary AR(1) process with a high  $\phi_1$ .

The correlogram for the first differences of 3 month Treasury bills is hard to interpret. Some statisticians would say that the long string of positive sample autocorrelations in the original time series disappears from the first differences. The sample autocorrelations for the first ten lags are sometimes positive and sometimes negative. These statisticians say we can fit an ARIMA process to the first differences, not the original time series. The ARIMA(8,1,4) fit in the textbook takes this view.

But interest rates in developed countries are usually not a random walk. An economist might see three distinct patterns in the time series of 3 month Treasury bills.

If interest rates are not a random walk, first differences are not ideal. We mention methods to form a stationary time series that can be modeled by a lower order ARIMA process.

*Take heed:* The time series for three month Treasury bills has interest rates to two decimal places, such as 5.50%. If the interest rate does not change between months, the first difference is zero. The time series for LIBOR rates has more decimal places. The first differences of LIBOR rates look more like a normal distribution.

*Take heed:* You can check whether taking first differences is appropriate.

- If the sample autocorrelations of the first differences are a white noise process, taking first differences is correct.
- If the sample autocorrelations of the first differences are negative for first two or three lags, taking first differences is not correct.

The discussion forum posting on *time series simulations* explains these relations.

The sample autocorrelations of first differences of three month Treasury bills do not give a conclusive answer. We show the sample autocorrelations and the correlogram. With 665 observations, the standard deviation of a white noise process is  $1/\sqrt{665} \approx 3.88\%$ . The sample autocorrelations for the first 20 lags have many values greater than two standard deviations, but no clear pattern. Fitting an ARIMA model to the first differences is not easy.

*Take heed:* Each time series differs. Your student project may use other rates, such as LIBOR rates or bank prime rates. You need not find the correct ARIMA process or even a good fit. But you should explain your analysis: why you chose a particular approach.

## UNIT ROOTS

The textbook discussion of unit roots is concise. The homework assignments and final exam problems do not emphasize this topic, but it is important for the student project.

Interest rates may be mean reverting or random walks.

- A mean reverting time series is stationary and can be modeled by an ARMA process.
- A random walk is not stationary. It is modeled by an ARIMA process.

If your time series looks like a random walk, use three tests:

- Use a one period lagged regression and check for a unit root.
- Form a correlogram and examine the decay in the sample autocorrelation function.
- Take first differences and check the Box-Pierce Q statistic.

*Take heed:* Many economic and financial indices may be random walks. Interest rates, inflation rates, GDP growth, unemployment rates, and various similar indices look like random walks within moderate bounds.

*Unit Root:* Regress the time series on the same values one period back. This is an AR(1) model, which is the most common ARIMA process. If  $\phi_1$  (the  $\beta$  of the regression equation) is more than 1 or less than  $-1$ , the time series is not stationary. We see this in the graph.

- If  $\phi_1 > 1$ , the time series grows continually. Random fluctuations may cause any single value to be smaller (in absolute value) than the preceding one, but the growth is clear over long periods. To correct this, we take logarithms and first differences.
- If  $\phi_1 < -1$ , the time series grows continually and oscillates. Random fluctuations may obscure the exact process, but the oscillations are evident. This type of process is rare.

If  $\phi_1$  is  $\approx 1$ , the time series is a random walk and is not stationary. Because of random fluctuations, the ordinary least squares estimator of the parameter is never exactly one.

- If we estimate  $\phi_1$  as 0.95 in a time series of 40 observations, we assume it is one and the time series is a non-stationary random walk.
- If we estimate  $\phi_1$  as 0.80 in a time series of 400 observations, we assume it is less than one and the time series is a stationary AR(1) process.

## Step #5: *PERIODS*

If an exogenous *intervention* causes the different patterns, separate the time series into two or more periods.

- Statutory, regulatory, and judicial interventions are common exogenous factors. The time series project templates have many examples.
- Federal Reserve Board policy on interest rates affected the Treasury bill time series.

If you are not aware of exogenous factors, base the periods on the time series pattern.

*Illustration:* Suppose interest rates increase for the first ten years and decrease for the second ten years.

- The time series itself is not mean reverting, so it is not stationary.
- The first differences are positive for the first ten year and negative for the last ten years. The mean first difference is not stable, so the first differences are not stationary.
- In this simple process, the second differences may be zero for all periods (except at ten years). Actual time series are less distinct. The interest rate path may be a parabola, and even the second differences are not stationary.

If the time series can be divided into two or three homogeneous processes, fit a separate ARIMA process to each period. Periods often provide interesting student projects. Distinct changes often occur, such as a new government in a country or a new CEO of a firm. You can analyze GDP growth under two governments or sales growth under two CEO's. Some past student projects posted on the discussion board analyze time series in two periods.

*Illustration:* Examine the graph of three month Treasury bill rates for January 1945 through June 2000. Note the general upward trend for 1945 through 1979, the high volatility for 1980 through 1983, and the downward trend for 1984 through 2000.

For your student project, do the following:

- Graph your time series and examine the means, trend, and volatility in different periods.
- Select periods based on either exogenous information (a change in policy, legislation, or economic environment) or the observed means, trends, and volatility.
- Separate into periods if the differences seem material, not just random fluctuation.
- Fit ARIMA processes to each period. Examine if the processes in adjoining periods are materially different.

*Take heed:* Structural models and de-trending are better than separating into periods if the time series depends on some other index. You may not be able to fit an ARIMA process to nominal interest rates, but perhaps you can model real interest rates. Unemployment rates may have a cyclical pattern, but the residuals of unemployment rates on GDP growth may be an AR(1) process.

*Take heed:* Some statisticians avoid dividing time series into periods. They say the goal of time series analysis is to forecast turning points. Periods of stability are easy to forecast. Starting a new period at each turning point does not accomplish the goal. The authors of the course textbook generally prefer to model long time series with higher order terms.

Other statisticians believe that modeling a time series composed of different processes degrades the forecasts. The interest rate process depends on monetary policy. If Federal Reserve Board policy changes, the interest rate process changes.

Periods are useful if the ARIMA process changes, not if it simply turns. Many industries show profit cycles, like the underwriting cycles in insurance.

- Corporate bond spreads may also have cycles, perhaps related to business cycles.
- If such cycles exist, the goal of ARIMA modeling is to model them, not to start a new process every time the cycle turns.

*Take heed:* Your student project may compare one ARIMA process for the entire time series vs separate ARIMA processes for each period.

For your student project, use periods cautiously. They are appropriate if the process is distinct and homogeneous for each time period. We summarize below the FED policies in the three sub-periods.

Post-World War II Federal Reserve Board policy explains the changes in the Treasury bill time series. Just as a pricing actuary knows the policy provisions, type of insurer, and extent of market competition to set optimal rates, a statistician should know the attributes of the time series to fit an ARIMA process. But the student project focuses on the statistics, just as the SOA and CAS exams focus on the actuarial procedures. You are not required to know the economic and financial effects on the time series, but this knowledge may help you fit an ARIMA process.

- From the end of World War II (1945) through the mid-1970's, the U.S. economy expanded briskly. Government officials worried about Depression-era deflation, not the mild inflation of an expanding economy. Inflation was thought to be an antidote to unemployment, which had been high during the Depression (almost one third of the labor force in 1932). The federal government and the Federal Reserve Board believed that mild inflation was beneficial, in that it restrained unemployment and did not hamper economic prosperity.

This presumed relation of inflation and unemployment was an error, but it was the prevailing macroeconomic policy in the 1960's and 1970's. To reduce unemployment, the FED used expansionary monetary policy. Inflation and interest rates had steady upward trends. The time series of interest rates is not stationary, though real interest rates and the first differences of nominal interest rates may be stationary. You select the end-point of this period: the periods flow into one another; they are not distinct.

*Take heed:* These comments apply to other interest rate time series as well, such as Treasury bonds or one year Treasury bills.

- From the late 1970's through the early 1980's, inflation and interest rates were high and volatile. See the large swings in interest rates in the graph on the illustrative worksheet. The high and volatile rate reflect (i) the mistaken macroeconomic policies of these times and (ii) the supply shocks of OPEC oil price increases (1973 and 1979).

*Take heed:* A volatile time series can be stationary, but it is hard for the statistician to distinguish trends from random fluctuations in a short, highly volatile time series.

- Paul Volcker became chairman of the Federal Reserve Board in 1981 and adopted a monetarist perspective (Milton Friedman's views). The money supply grew at a steady, slow rate. Interest rates and inflation declined. Greenspan continued Volcker's policy.

*Take heed:* We do not give Volcker credit or deny it to him. Some economic historians say Reagan was lucky to be President when Volcker chaired the FED. Others credit Reagan with stopping inflation, and Volcker was lucky to chair the FED in these years.

Most likely, a single ARIMA process is not an appropriate model for all three periods.

- You form a stationary time series several ways, as described in this project template.
- You can also combine methods, such choosing a sub-period, detrending, taking first differences, and forming a structural model.

If you select appropriate periods, state what periods you use and justify your choice.

*Illustration:* An analysis of three month Treasury bill rates might say:

- I graphed the rates, to see if the time series has the same process in all periods. I used rates for 1982 - 2000, which seem to have a downward drift. I used real interest rates to offset the decline in inflation during these years.
- I took first differences of the rates, to see if one ARIMA process could model all the post-World War II rates. I then excluded years 1979-1982 to see if excluding years with high volatility improved the fit.

*Take heed:* Some student projects compare two time periods. You might compare abortion rates before and after Row v Wade. Your write-up might say:

I fit ARIMA models to U.S. abortion rates for the year before and after Row v Wade. I found that the drift differs in the two time periods, and separate ARIMA models are needed.

*Take heed:* You may not have the exogenous knowledge to select proper time periods. You may select time periods based on the mean, drift, or volatility of the observations.

*Illustration:* Your student project may say: “Overnight LIBOR rates for 1982 - 2006 decline for several years and then increase. (You would specify the years; see the project template on LIBOR rates.) My student project examines if the time series are really different:

- I took first differences of LIBOR rates in each period. The mean first difference differs by period, but each period can be fit by an ARIMA(2,1,0) process.
- I used real interest rates LIBOR divided by expected inflation. Each period can be fit by an ARMA(2,0) process.

*Take heed:* In short time series, drifts are hard to distinguish from random fluctuation.

*Illustration:* The years 1979-1982 is a short period of volatile interest rates. A difference of 2 or 3 months changes the observed drift. The drift is *not robust*. If you find a drift in the time series, consider also the volatility of the values and the length of the period.

*Illustration:* If a 20 year period has a drift of +2% per annum, each 10 year sub-period should have a drift of about 2% per annum. If the first ten years have a drift of +5% and the second ten years have a drift of -1%, the overall drift of +2% is not robust.

Interest rates seem to show drifts for short periods, such as several months of increasing rates followed by several months of decreasing rates. This may reflect an ARIMA(1,1,0) process and random fluctuations.

*Illustration:* Suppose interest rates are an ARIMA(1,1,0) process with  $\delta = 0$  and  $\phi_1 = 80\%$ .

- If interest rates increase 100 basis points in January 20X6 because of random fluctuations, they are expected to increase 80, 64, 51, 41, and 33 basis points in each of the next five months.
- If they then decrease 100 basis points in July 20X6 because of random fluctuations, they are expected to decrease 80, 64, 51, 41, and 33 basis points in each of the next five months.

*Illustration:* For a period of 1 month and a volatility of 0.1% per month, even a time series with no drift may show a drift of *about* 0.1% (either positive or negative)..

- The drifts of +0.02% and -0.01% in the first and third periods reflect FED policy.
- The observed drift in the middle period reflects the short time period and high volatility.

## *SUB-PERIODS AND STATIONARITY*

{Taking differences may convert a homogeneous time series to a stationary time series. Separating a time series into sub-periods may help several ways.}

*Jacob:* Are the time series stationary in each sub-period?

*Rachel:* The first and third periods, with upward or downward drifts, are not stationary, but their first differences may be stationary.

- You may compare first differences for each sub-period vs for the entire time series.
- You may compare real interest rates for each sub-period vs for the entire time series.

When analyzing sub-periods for an economic or financial time series:

- Examine the raw time series, using first and second differences.
- Detrend the time series or use real interest rates or real dollars.
- Use a structural model by regressing the time series on other indices.

{*Take heed:* These are suggestions for time series analysis. Decide how to fit an ARIMA process. We explain de-trending (real interest rates) and structural models below. They are not required for the student project, but they are good topics.}

## Step #6: *DE-TREND*

A time series may combine several patterns reflecting several explanatory variables.

*Illustration:* Corporate bond rates for auto manufacturers combine expected inflation, real interest rates, business cycles, and default probabilities for the issuing firms.

It is easier to fit an ARIMA process to each piece separately than to the combination.

*Illustration:* Expected inflation may be an ARIMA(1,1,0) process, real interest rates may be an ARMA(1,1) process, and business cycles may be an oscillatory process.

- It is often easier to model a time series in real dollars than in nominal dollars.
- The same is true for other time series that are functions of a changing measure.

*Illustration:* It is easier to fit an ARIMA process to GDP per capita than to a country's total GDP. Population growth or decline is like inflation or deflation. We fit separate ARIMA process to (i) population growth and (ii) GDP per capita.

If the time series is in dollars (or other currency), convert it to real dollars. Divide the dollars by the CPI. We provide several CPI indices on the discussion forum for deflating.

*Take heed:* Decomposing a time series and fitting ARIMA processes to its parts is a good student project. Your project may compare ARIMA processes for nominal vs real interest rates.

*Recommendation:* Choose a time series that is composed of two or more elements. To keep your student project manageable, use two pieces:

- Nominal interest rates = real interest rates + expected inflation
- Corporate interest rates = risk-free rates + default spreads
- Insurance premium = new business premium + renewal business premium
- Auto sales = new car sales + used car sales

If you want a more adventurous project, decompose the time series into 3 or 4 parts.

Compare an ARIMA process fitted to the combined time series vs ARIMA processes fitted to each part. See which model forecasts better.

*ILLUSTRATION: NOMINAL AND REAL INTEREST RATES*

To convert nominal interest rates to real interest rates, divide by the inflation rate in the previous month as a proxy for expected inflation.

*Illustration:* The nominal interest rate is 8% on June 1, 20X8. The CPI is 130 on 6/1/20X8 and 125 on 5/1/20X8. The real interest rate on June 1, 20X8, is

$$1.08 / (130 / 125) - 1 = 3.85\%$$

For your student project, do the following:

- Copy a CPI index from the discussion forum to the work-sheet with your interest rate time series. You have a choice of seasonally adjusted or not seasonally adjusted CPI.
- Compute the inflation in each month as the ratio of two CPI figures.
- The interest rate is annualized. To annualize the inflation rate, raise it to the 12<sup>th</sup> power.
- Choose a proxy for expected inflation. You might use the actual inflation the previous month or a moving average of inflation rates in the previous several months.
- Divide (1 + nominal interest rate) by (1 + expected inflation rate).

*Intuition:* The objective of ARIMA modeling is to separate trends, cycles, seasonality, and stochasticity. Time series often overlay a stationary ARIMA process on a trend, drift, or cycle. Inflation may combine a trend with seasonality, and interest rates may be a mean reverting pattern overlaid on a business cycle. ARIMA modeling separates each part.

The procedures differ for each part. We may use

- First differences for the trend
- An autoregressive process for the mean reversion
- A structural model for the business cycle
- A sine pattern for the seasonality.

*Take heed:* Pick an appropriate inflation index. For health insurance loss costs, you may use medical CPI, not total CPI. Explain if you use seasonally adjusted or non-seasonally adjusted CPI.

- If your dollar-denominated time series has the same seasonal pattern as the CPI, use non-seasonally adjusted CPI to de-trend.
- If your dollar-denominated time series is not seasonal, use seasonally adjusted CPI to de-trend.

*Take heed:* De-trending, adjusting for seasonality, and fitting the proper ARIMA process is a good student project for insurance loss cost trends. Don't presume that actuaries have already developed optimal trend models. On the contrary: actuaries are just now beginning

to use sophisticated ARIMA processes for loss cost trends. The first CAS paper on ARIMA modeling of loss cost trends was published recently.

Use spreads in the same manner as de-trending. Real interest rates are the spread over expected inflation. Corporate bond rates are the spread over risk-free rates.

*Illustration:* Instead of analyzing the Moody's AAA Bond rate, analyze the corporate bond spread to Treasury bonds. You may add a structural component (such as GDP growth) as a proxy for default probabilities.

*Intuition:* Corporate bond rates are a mix of several items:

- The term structure of interest rates, which can be modeled by Treasury securities.
- Expected inflation, which can be modeled by CPI changes and money growth.
- Real interest rates, which is best modeled by Treasuries or LIBOR rates.
- Default expectations, which can be modeled by GDP growth.

Your student project need not model all aspects of the corporate bond yield. Explain how the corporate bond spread to Treasury securities eliminates duration, risk-free rates, and expected inflation, leaving a time series governing by default expectations and business cycles. Regressing the corporate bond spreads on GDP growth may eliminate the business elements, leaving a stationary time series that can be fit to an ARIMA process.

*Note:* An early use of ARIMA processes was to model economic cycles in the U.S. and Great Britain. You have probably studied the lay version in a college economics course: prosperous years cause consumer over-confidence (or some other item) that leads to an over-heating economy and a recession. The CAS Exam 5 syllabus has a reading on underwriting cycles with a similar ARIMA perspective. The on-line NEAS macroeconomics course has a different perspective on business cycles (called real business cycle theory).

*Take heed:* The real interest rates formed on the illustrative work-sheet do not form a stationary time series. The real interest rate should be between 0.5% and 4%. A figure outside this range may mean the estimated expected inflation rate is not correct.

- Last month's inflation is sometimes greater than the nominal interest rate, implying that investors believe inflation will fall.
- Inflation may be low one month for exceptional reasons (perhaps oil prices fell because of a peace treaty in the Middle East), but investors presume future inflation will be high.

For the real interest rates on the illustrative worksheet, you still take differences, divide the time series into homogeneous periods, or de-trend the time series to fit an ARIMA process. Your student project may explain whether using real interest rates improves the ARIMA fit.

## Step #7: *STRUCTURAL MODELS*

ARIMA models are proxies for the true explanatory model. Treasury bill rates are affected by expected inflation, economic growth, consumer confidence, political stability, demand for money, other investment opportunities, and similar factors.

If we knew all the influences on Treasury bill rates and values of all explanatory variables, we could form a model  $R = f(X_1, X_2, X_3, \dots) + \epsilon$ . The model might also have lagged terms, such as economic growth last month or two months ago.

Constructing a complete model may not be possible, since we do not know the influences on Treasury bill rates. An ARIMA model is a proxy.

- The explanatory variables are also stochastic time series.
- The ARIMA process puts all the explanatory variables into one time series.

*Illustration:* Suppose nominal interest rates are a function of real interest rates, expected inflation, economic activity, and demand for money. Each of the explanatory variables can be modeled as an autoregressive or moving average process. Ideally, we might forecast each explanatory variable and then derive the indicated Treasury bill rate each month.

Structural models are a compromise between the full model and a simple ARIMA process.

*Illustration:* Suppose real interest rates depend on economic growth, consumer confidence, political stability, demand for money, other investment opportunities, and similar factors. Economic growth can be measured by GDP growth and it has a large effect on real interest rates. The other explanatory variables are hard to measure and have less effect on real interest rates.

We regress the real interest rate on real GDP growth. We fit the residuals of the regression to an ARIMA process. These residuals reflect the effects of the other explanatory variables.

A diffuse pattern of real interest rates may be a simpler ARIMA process after regressing on real GDP growth. Your student project may examine the effects of each item.

*Illustration:* Suppose your time series is the 20 year corporate bond AAA rate.

- Model the nominal corporate bond rate. Take first (and perhaps second) differences to get a stationary time series. The goodness-of-fit tests may be poor, since the ARIMA model combines several influences on corporate bond rates.
- Model the *real* corporate bond rate (offset with the change in the CPI). The in-sample fit may improve, but it may be difficult to forecast future interest rates unless you can forecast future CPI values.
- Model the corporate bond *spread* to *three month Treasury bills*. The time series may be easier to fit, and you may not need to take differences. But duration effects are

overlaid on business cycle effects, and the fit may not be good. [The duration effect is the slope of the term structure of interest rates.]

- Model the corporate bond spread to 20 year Treasury bonds. Corporate bonds and Treasury bonds have about the same duration, so the term structure of interest rates does not affect the spread. With just spread cycles and default expectations remaining, the ARIMA fit may improve.
- Regress the corporate bond spread on a measure of business activity, such as GDP growth. Economists do not agree about the influence of GDP growth on interest rates. In general, as GDP increases, bond defaults decrease, so the corporate bond spread narrows. Your time series of spread cycles may resemble a simpler ARIMA process.

Ideally, GDP growth is annualized, seasonally adjusted, converted to real terms, shown by month, and perhaps lagged. We explain each of these adjustments in the discussion forum posting on structural models. They are important for the economic analysis, not for the student project. If the explanatory variable is not in the form you want, explain the desired form in the student project write-up but do the analysis with the data you have.

Structural models are best if we have suitable explanatory variables. The structural model shows that you can decompose a time series into its pieces and use statistical tests to see if the decomposition gives a better ARIMA model. Even if you are not persuaded that a structural model is correct, you may examine it for your student project.

*Illustration:* Some economists say that higher GDP growth raises real interest rates, since firms wish to borrow (and invest) more in prosperous year and demand for cash rises. You may not be persuaded by the economic reasoning, and you would like to test it.

- Your regression analysis student project may examine several macroeconomic indices that may affect real interest rates, such as GDP growth, budget deficits, foreign interest rates, and employment rates.
- Your time series student project may fit ARIMA processes to the residuals of each regression line.

{Structural models are discussed in a separate discussion forum posting, and an example is in a separate illustrative worksheet.}

## Step #8: CORRELOGRAMS

*Terms:* Compute sample autocorrelations of the time series. The student project fits an ARIMA process to the observed time series. We test the fit several ways:

- The sample autocorrelations of the time series should have the same pattern as the implied autocorrelations of the ARIMA process.
- The residuals of the time series from the fitted ARIMA process should be random, so the sample autocorrelations of the residuals should be those of a white noise process.

*Terms:* The sample autocorrelation depends on the lag, such as 30% for a lag of 1, and 20% for a lag of 2. The sample autocorrelation as a function of the lag is the sample autocorrelation function.

The graph of the sample autocorrelation function is the correlogram. When you first start ARIMA fitting, the graph (correlogram) is easier to read than a table of autocorrelations. But you use the table of sample autocorrelations for the statistical tests.

The raw time series may not be stationary. Explain the pattern of sample autocorrelations. The correlogram should confirm the pattern you see in the graph.

Your student project proceeds in two directions.

- The sample autocorrelations reflect trends, seasonality, cycles, and other patterns in the time series. Adjust the data to remove trends, seasonality, and cycles, and compute again the sample autocorrelations.
- After fitting an ARIMA process to the time series, determine the sample autocorrelation function of the residuals. The residuals should be a white noise process, which we test by the pattern of the sample autocorrelation function.

*Illustration:* Inflation is seasonal, so short interest rates, such as overnight LIBOR, may also show seasonality. If you believe the rates are seasonal, you may de-seasonalize the data and re-compute the sample autocorrelations.

*Take heed:* We provide several time series of LIBOR rates. These are excellent time series for student projects, since the short rates (overnight, one week, and two weeks) have daily observations. We also show LIBOR rates in other currencies, such as the Euro, and we show inflation and foreign currency exchange rates. These indices are related and they are all stochastic, so you can form various structural models.

*Take heed:* Examining the seasonality of overnight LIBOR rates is not easy, since trends, seasonality, cycles, and random fluctuations are overlaid on each other.

- Convert each rate to its difference from a centered moving average of one year. The LIBOR time series shows rates for business days only, so a year is somewhat less than 250 days. Use ratios for the differences or first take logarithms.
- Compute long-term averages for a given day. Use all the years on the NEAS web site.
- Graph these averages to see if they show a seasonal pattern.
- Confirm your results with a correlogram. If LIBOR rates are seasonal, the correlogram should start high (reflecting the autoregressive process), decline to about 125 days (half a business year), and then rise to a local maximum at slightly less than 250 days.

You learn how to test for seasonality in a complex time series.

*Illustration:* The daily temperature, after adjusting for seasonality, may have a trend. Even if the trend is obscured by random fluctuations, it may be seen in the correlogram. After detrending the temperature, the correlogram may indicate a stationary time series. You then fit an ARIMA process so that the residuals show a white noise process.

We discussed earlier segmenting a time series into periods. The periods are indicated if the full time series is not stationary but each period is stationary.

*Illustration:* Treasury bill rates may show patterns, such as increasing or decreasing rates. You may divide the time series into periods, each of which has its own pattern.

- The first differences in each period may be stationary processes with different means.
- De-trended interest rates, real interest rates, or the interest rates regressed on another index may be distinct ARIMA processes in each period.

## Step #9: Fit a Model

The model depends on the sample autocorrelation function. For the student project, look at four models: AR(1), AR(2), MA(1), and ARMA(1,1), along with their first differences: ARIMA(1,1,0), ARIMA(2,1,0), ARIMA(0,1,1), and ARIMA(1,1,1). You may fit more complex models if you believe they are needed, but we do not require more complex models for the student project.

*Take heed:* If the sample autocorrelation function (the correlogram) indicates that an AR(1) or AR(2) model is appropriate, and if the fitted model passes the Box-Pierce Q statistic, you need not fit MA(1) or ARMA(1,1) processes.

*Take heed:* Fit the ARMA models or the ARIMA models, not both.

- If the time series is stationary, use the ARMA models, not the ARIMA models.
- If the time series is not stationary, use the ARIMA models, not the ARMA models.

The model fitting is described in separate discussion forum postings.

- For the AR(1) and AR(2) processes, fit a lagged regression.
- For the MA(1) and ARMA(1,1) processes, use the Yule-Walker equations.

After fitting the model:

- Compare the sample autocorrelation function with the autocorrelation function.
- Use Bartlett's test and the Box-Pierce Q statistic to test the goodness-of-fit.
- For forecasts and evaluate the out-of-sample goodness-of-fit.

Each of these is described in separate postings.

*Take heed:* Document your work as you do it. When you finish the student project, edit your documentation using the guidelines on the discussion forum. The documentation is the write-up for the student project.

## Step #10: AUTOCORRELATIONS AND SAMPLE AUTOCORRELATIONS

After you fit an ARIMA process, compare the implied autocorrelations from that process with the sample autocorrelations of the time series.

- The correlogram shows the pattern of the observed time series.
- The actual time series is stochastic, so the correlogram is not smooth.
- You fit an ARIMA process, which has a smooth autocorrelation function.
- Form a graph overlaying this smooth function on the correlogram.
- If the smooth function is close to the sample autocorrelations, the fit is good.

*Question:* Why do we compare the correlogram to the autocorrelation function? Why not compare the observed time series to the pattern of the ARIMA process?

*Answer:* An ARMA process becomes a straight line at the mean after a few lags, and an ARIMA process becomes a diagonal line with a slope equal to the drift after a few lags. The forecasts from any ARMA process look the same.

*Question:* Why not compare the graph of the one month forecasts from the ARIMA process to the actual time series?

*Answer:* This comparison is the sum of squared deviations, which we use to choose the best model. The algebra gives a figure that varies by ARIMA process. On a graph, it is hard to see which process fits best. An AR(1) process looks about the same whether  $\phi_1 = 40\%$  or 60%.

The graph of the autocorrelation function looks different for each ARIMA process. It is a good marker of the ARIMA process, and it is easy to compare with the correlogram.

*Take heed:* The discussion forum posting on *time series simulations* has correlograms for several ARIMA processes. The correlogram reveals the time series process. Know what the correlogram of each ARIMA process looks like, so you can identify a reasonable model for your time series.

*Take heed:* The discussion forum posting on *time series techniques* and the attached Excel work-sheet gives cell formulas and a VBA macro for sample autocorrelations. That posting uses the interest rates in this project template as the illustration.

Review the illustrative worksheet on time series techniques. Make sure you understand the cell formulas and the correlograms.

The VBA macro is optional. We have not made the macro a screen based facility, so you don't treat the macro as a black box. Open the macro in the VBE (the editor) and choose the length of the time series and the number of lags. The defaults are the entire time series

and all lags. Choose a smaller number of lags if the time series has so many observations that the number crunching takes too much time (e.g., more than 10,000 observations).

Choose a different length for the time series if you want the correlogram for a sub-period.

*Illustration:* Suppose the time series has 240 observations of monthly interest rates. You want to form correlograms separately for the first 10 years and the second 10 years.

- Place the cursor on the first observation. Choose a length of 120 to form a correlogram from the first 120 observations.
- Place the cursor on the 121<sup>st</sup> observation. The default is to the end of the time series. This forms a correlogram from the second 120 observations.

If you don't want to use macros, you can use the cell formulas. The macro is more efficient. If you have a slow machine and a time series with tens of thousands of observations, you must use the macro. For interest rates, you can use cell formulas.

## Step #11: LAGGED REGRESSIONS

For the AR(1) and AR(2) processes, fit a lagged regression. Copy the stationary time series to a second column and use Excel's *REGRESSION* add-in.

*Take heed:* If the original time series is not stationary and you took first differences, do the lagged regression on the stationary first differences.

*Take heed:* Excel comes with the analysis tool-pack, but you must load it on your copy.

*Illustration:* The time series values are in Cells A11:A200 and we fit an AR(2) process.

- Copy Cells A11:A200 to Cells B12:B201 and also to Cells C13:C202. *Take heed:* Column B has the values lagged one period; Column C has values lagged 2 periods.
- Invoke Excel's *REGRESSION* add-in. For Excel versions before 2007, click the *TOOLS* menu  $\Rightarrow$  *DATA ANALYSIS*  $\Rightarrow$  *REGRESSION*. For Excel 2007, click *DATA*  $\Rightarrow$  *ANALYSIS*  $\Rightarrow$  *DATA ANALYSIS*  $\Rightarrow$  *REGRESSION*. The sequence of clicks may vary for your version of Excel.
- On the regression screen, select Cells A13:A200 as the Y values and Cells B13:C200 as the X values. Do not include headers, since the values in Cells A13:C13 are time series observations, not labels for the variables. Label the output of the regression.
- Select *SHOW RESIDUALS* (not standardized residuals). If the time series is stationary, the mean does not vary, so the standardized residuals should be almost the same as the ordinary residuals.
- Place the output on a new worksheet. Name the new worksheet as the ARIMA process, such as AR(2). This work-sheet will also show the correlogram of the residuals, their Box-Pierce Q statistic, and a comparison with the theoretical autocorrelations.

### GOODNESS-OF-FIT

The goodness-of-fit measures used in regression analysis, such as the  $R^2$  of the regression and the significance of the ordinary least squares estimators, are not the most important measures for ARIMA fitting.

- If you have taken the regression analysis course, comment on these items.
- If you have not taken the regression analysis course, you may ignore these items and comment just on Bartlett's test and the Box-Pierce Q statistic.

You can use the  $\bar{R}^2$  (the  $R^2$  adjusted for degrees of freedom) to compare AR(1) and AR(2) processes. If the adjusted  $R^2$  does not increase from AR(1) to AR(2), use AR(1). The  $R^2$  (or adjusted  $R^2$ ) won't help you decide between AR(1) and MA(1).

Check that the ordinary least squares estimators are significant. For a lagged regression with two periods – an AR(2) model – the correlation of the explanatory variables affects the  $t$  values.

Form the sample autocorrelations of the residuals, and use the Durbin-Watson statistic, Bartlett's test, and the Box-Pierce Q statistic. The illustrative worksheet on time series techniques provides cell formulas and a VBA macro. The cell formulas and macro do all the arithmetic. Your write-up should explain how you use and interpret these tests.

## Step #12: MOVING AVERAGE MODELS

For MA(1) and ARMA(1,1) processes, the textbook uses nonlinear regression to estimate the ARIMA parameters and Yule-Walker equations for initial estimates of the parameters.

- Nonlinear regression is not covered in the on-line courses.
- The Yule-Walker estimates are close enough for the student project.
- Excel's *SOLVER* add-in makes it simple to use the Yule-Walker equations.

*Take heed:* You can solve the Yule-Walker equations by hand; you don't need solver. For MA(1) and ARMA(1,1) processes, solve a quadratic equation for  $\theta_1$ .

*Illustration:* Suppose the sample autocorrelations in a time series of 10,000 observations are 45% for lag 1 and 10% for lag 2. The sharp drop from lag 1 to lag 2 suggests a moving average parameter, so we test an ARMA(1,1) model. Your write-up explains why you use an ARMA(1,1) process:

- With 10,000 observations, the standard error of the sample autocorrelations of a white noise process is 1%.
- For an AR(1) process,  $45\%^2 = 20.25\% > 10\%$ .
- Even if the sample autocorrelations are off by one standard deviation,  $46\%^2 = 21.16\% > 11\%$ .

Examine both an AR(2) process and an ARMA(1,1) process.

- For an AR(2) process, we fit the autoregressive parameters with linear regression. An AR(2) process has  $\phi_1 > 0$  and  $\phi_2 < 0$ , as might occur in a cyclical time series.
- For MA(1) and ARMA(1,1) process, we use the Yule-Walker equations.

For an MA(1) process,

$$\theta_1 = \frac{-1 \pm \sqrt{1 - 4\rho^2}}{2\rho} = (-1 + \sqrt{1 - 4 \times 0.45^2}) / (2 \times 0.45) = -0.627$$

An MA(1) process would have a zero autocorrelation of lag 2. The MA(1) process fits no better than the AR(1) process. Once you have fit all four models, use Bartlett's test and the Box-Pierce Q statistic to see which fits best. The fit depends on all observations; we can't decide which process fits best from the first two sample autocorrelations alone.

Equations 17.58 and 17.59 on page 536 gives the autocorrelations for lags 1 and 2 of an ARMA(1,1) process.

$$\rho_1 = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - 2\phi_1\theta_1 + \theta_1^2}$$

$$\rho_k = \phi_1\rho_{k-1} \quad \text{for } k \geq 2$$

We know the sample autocorrelations. We solve the equations for the ARIMA parameters.

- Estimate  $\phi_1$  as  $10\% / 45\% = 2/9 = 22.22\%$ .
- Estimate  $Z = \theta_1$  as  $45\% = [(1 - Z \times 2/9) \times (2/9 - Z)] / (1 + Z^2 - 2 \times Z \times 2/9)$

We solve the quadratic equation for  $Z = \theta_1 = -0.290855735 \approx -0.291$ .

We estimate  $\delta$  from the ARIMA parameters and the mean of the time series.

- For the MA(1) process,  $\delta$  is the mean of the time series.
- For the ARMA(1,1) process,  $\delta$  is the mean of the time series times the complement of the autoregressive parameter.

*Take heed:* Instead of solving the quadratic equation by hand, we can use Excel's *SOLVER* add-in. Excel's *GOAL SEEK* can also do the arithmetic.

Choose cells for  $\theta_1$  and  $\phi_1$ , such as Cells B2 and B3. Name these cells one of three ways:

- Use the *DEFINE NAMES* dialogue box. (Select Cell B3, type *phis1* in the dialogue box, and press *ADD*.)
- Use the name box on the left end of the formula toolbar. (Select Cell B3, type *phis1* in the name box, and press enter.)
- Place the strings "theta1" and "phis1" in Cells A2 and A3. Use the *CREATE NAMES* dialogue box to assign these names to the values in Cells B2 and B3.

*Take heed:* You don't have to name the cells. You can refer to the cells with absolute references, such as  $\$B\$3$ . But your cell formulas become hard to understand.

- Choose cells for the sample autocorrelations of lags 1 and 2, such as Cells C2 and C3.
- Name these cells "rhos1" and "rhos2".

Type cell formulas into Cells C2 and C3.

- For Cell B3, type "= $\text{rhos1} / \text{rhos2}$ ".
- For Cell C2, type "= $((1 - \text{phis1} * \text{theta1}) * (\text{phis1} - \text{theta1})) / (1 - 2 * \text{phis1} * \text{theta1} + \text{theta1}^2)$ ".

Use solver to find  $\theta_1$ . We show the figures in the illustrative worksheet *YULEWALKER*.

*Take heed:* In Excel versions before 2007, we use names *phi1*, *rho1*, and *rho2*, not *phis1*, *rhos1*, *rhos2*. Excel 2007 changes the names *phi1*, *rho1*, and *rho2* to *phi1\_*, *rho1\_*, and

rho2\_, since phi1, rho1, and rho2 can refer to cells in Excel 2007. Typing underscores can be a hassle, so use names with 4 letters and a digit. "phis1" stands for phi-sub-1.

*Take heed:* If *SOLVER* has a poor starting figure for  $\theta_1$ , it may not find a solution.

- If Cell B2 has 1.000 when you invoke *SOLVER*, no solution is found.
- Start with a reasonable value for  $\theta_1$ . A good starting value is  $\phi_1 - \rho_1 = 0.2222 - 0.45 = -2278 = -22.78\%$ .

{This step-by-step guide fits MA(1) and ARMA(1,1) processes with Yule-Walker equations. With better statistical software, you can solve a nonlinear regression, as discussed in the textbook (not on the time series syllabus). Alternatively, you can fit an MA(1) or ARMA(1,1) process with Excel's *SOLVER* add-in. We show the *SOLVER* technique in a separate discussion forum posting.}

### Step #13: GOODNESS-OF-FIT TESTS

Test the goodness-of-fit of the moving average models using Bartlett's test and the Box-Pierce Q statistic, just as for autoregressive processes.

- Excel's *REGRESSION* add-in gives residuals. The VBA macro in the illustrative worksheet gives the sample autocorrelations and the Box-Pierce Q statistic.
- Excel does not have an add-in for the residuals of a moving average process.
- Compute the residuals manually. The cell formulas are simple, and the process shows you understand the goodness-of-fit test.

*Illustration:* Suppose the time series values are in Cells A11:A200.

- Column B is the forecasts. In Cell B11, type “=A11.”
- Column C is the residuals. In Cell C11, type “=A11-B11.”

*Take heed:* Column B is the forecast made in the previous period.

- Cell A12 has the time series value for Period 2.
- Cell B12 has the forecast for Period 2 made in Period 1.
- Cell C12 has the residual for Period 2.

Name three variables as delta, theta1, and phis1, using the methods described earlier.

*Take heed:* You don't have to place these values in the work-sheet. You can name constants in the “define names” dialogue box.

- Place these value in work-sheet cells and name the cells if these values change as you analyze the data. This is the more common scenario.
- Use named constants if these values do not change.

*Take heed:* For the MA(1) process, the phis1 variable is zero.

- In Cell B12, type “=delta + A11 × phis1 – C11 × theta1.”
- Copy Cell B12 to Cells B13:B200.
- Copy Cell C12 to Cells C13:C200.

Excel shows residuals for 190 observations. The residual for the first observation is zero (by definition), so we do not use it. We set B11 equal to A11 because we have no values for A10 and C10.

For the other 189 residuals, form the sample autocorrelations. Use Bartlett's test and the Box-Pierce Q statistic to test the quality of the ARIMA fit.

The VBA macro does most of the number crunching. Place the cursor in Cell C12 and run the macro. The macro also computes the values for the Box-Pierce Q statistic. Bartlett's test is subjective; you must count the sample autocorrelations above a critical value.

*Take heed:* This procedure slightly overstates the residuals by assuming the first residual is zero. The distortion is not material.

Step #14: *FORECASTS*.

The ARIMA model helps forecast future interest rates. A student project that fits an ARIMA process may have a section testing the fit with an out-of-sample test.

*Take heed:* A forecasting section is not required for all student projects. It is useful, since it shows that you understand the purpose of ARIMA modeling. If your student project does a good analysis of other topics, you need not include forecasting.

To test the quality of the forecasts, leave out the last N observations and forecast them.

*Illustration:* For a time series of three month Treasury bills ending June 2000, you might

- use observations through December 1999 and forecast the next six months.
- use observations through December 1998 and forecast the next 18 months.

A separate discussion forum posting covers forecasts and out-of-sample goodness-of-fit tests. For your student project, leave out the last year of observations from the data used to fit the ARIMA model. Do the ARIMA fitting using the techniques in this project template.

Once you have fit two or more models, compare their forecasts of the final observations. Keep in mind that an out-of-sample fit is easily distorted by random fluctuations. The ARIMA modeling and forecasting completes your student project.

*Take heed:* The project template discusses many techniques that you can apply to the time series. You need not use all the techniques. The illustrative worksheet shows some of the techniques, not all of them.