

TIME SERIES: INDEPENDENT STUDENT PROJECTS

Updated: May 3, 2008

This posting explains the independent student projects for the time series course. It uses examples from the interest rate time series on the discussion forum. Many suggestions here are repeated in other postings.

Take heed: The student project is not a linear assignment, starting at Point A and ending at Point Z. Some candidates grasp the requirement and can start work; others need more time to understand what they must do. You don't have to read everything on the discussion forum. Select items to read: introduction, write-up, project templates, extracts, and so forth.

Background and General Information

Jacob: Is the student project required? What does the student project show?

Rachel: You must submit the student project to receive course credit. The project shows that you can apply the time series concepts to actual data.

Jacob: When is the student project due? Is it like the final exam, which is given on a set date, or like the homework assignments, which must be completed to receive credit?

Jacob: How complex is the student project? Would a loss cost trend analysis for personal auto claims be suitable for a student project?

Rachel: The student project uses time series concepts and techniques. You fit ARIMA models to a time series, using the methods taught in the on-line course.

- Fitting an exponential curve to average claim severities is a linear regression after taking logarithms. It does not use stochastic ARIMA processes or the model building methods in the time series course.
- Modeling loss cost trends with ARIMA processes is a good topic for a student project. See the project template on loss cost trends.

Jacob: Do we do the student project after finishing the course or as we do the modules?

Rachel: The time series modules build upon one another. The last five modules (Chapter 19 of the textbook) apply the concepts to actual data. You won't have a good sense of building ARIMA models until you complete the modules.

Jacob: Should we wait until the last two weeks of the course to do the student project?

Rachel: The student project deals with ARIMA models: specification, diagnostic testing, forecasting, sample autocorrelation functions, Box-Pierce Q statistic, correlograms,

autoregressive and moving average processes, structural models, stationarity, Yule-Walker equations, and integrated models.

- Learn the material over the eight weeks of the course. Focus on the textbook readings, homework assignments, and practice problems.
- During the last two weeks (last six modules) of the course, begin reading the postings on the student projects. The illustrations in Chapter 19 of the textbook are complex, using non-linear regression and partial autocorrelation functions to fit high order ARIMA processes, such as ARIMA(4,2,4) or ARIMA(6,1,8). In practice, we use simpler ARIMA processes for actuarial and financial time series.
- Until the final exam, focus on the practice problems and the statistical techniques.
- Right after the final exam, review the project templates on the NEAS web site, read several of the past student projects posted on the discussion forum, and pick a topic. Spend an hour surfing the web, looking for data on a topic that interests you. Suitable time series on thousands of topics are freely available on the web.
- The NEAS web site has hundreds of time series you can use for the student project and a wide array of project templates, illustrative workbooks, step-by-step guides, and instructions. Your student project may analyze daily temperature for your home town following the project template and illustrative workbook on the web site or it may use a time series that you pick from the internet or your company's data.

Do a student project on a topic that interests you. The time series can be sports scores, movie ticket receipts, DVD sales, voting behavior, crime rates, marriages, divorces, births, abortions, immigration, or gas prices. You can choose a project related to your work, such as claim frequency, claim severity, mortality, premium trends, or interest rates.

- Before doing the student project, most candidates fear it will be a burden.
- After completing the project, most candidates say it was the best part of the course. The time series concepts in the textbook are abstract. Using them to analyze a real time series makes them come alive.

Use the discussion forum on student projects for ideas. Read the project templates, review the past student projects, look at the data sets and time series. Explore the internet, using the search engines. You will find hundreds of sites with potential data for a student project. Discuss potential topics with other candidates.

APPLYING TIME SERIES

Jacob: What are ARIMA models used for?

Rachel: ARIMA models are used for many items:

- Business *growth*, product sales, inventory management.
- Regulatory, economic, social, and legal *interventions*.
- Macroeconomic indices: Interest *rates*, inflation, unemployment, exchange rates.
- *Seasonality*, cycles, regimes (eras); see the project template on daily temperature.

- *Residuals* of regression relations (see the project template on structural models)

The textbook emphasizes business growth, product sales, and inventory management. For time series on product sales, use the internet search engines. These data are on industry (trade association) web sites.

The *project templates* on the discussion forum use interest rates, inflation rates, daily temperatures, and insurance statistics. They discuss seasonality, time periods, and residuals.

- Many actuaries deal with interest rates, inflation rates, and similar indices. The project templates on these subjects apply the time series concepts to real actuarial work.
- The macroeconomic indices are publicly available on government web sites. Many indices are on the NEAS web site in Excel sheets that you can download.
- Daily temperatures are available for 1,221 weather stations for 100+ years. The data are in Excel (CSV) format on the NEAS web site that you can download.

Recommendation: If you want to do a student project on macroeconomic indices or daily temperatures, check the data already on the NEAS web site. Check also the suggestions for other student projects on the web site.

INTERVENTIONS

Social scientists use time series analysis to examine changes in the environment. The 1973 Middle East war led to the OPEC oil cartel. We examine if the price of oil follows a different pattern after 1973 than before 1973.

Similarly, Federal Reserve Board policy affects interest rates, minimum wage laws affect unemployment, deregulation affects prices of airline tickets and phone service, insurance regulation affects policy premiums. Examine time series before and after an intervention.

You can examine effects of home DVD sales on movie ticket sales, effects of laws and court decisions on abortions, welfare, and crime rates, effects of sports rule changes on won-loss records, and so forth.

Real research projects examine all material effects on a time series. We don't examine time series of abortion rates before and after the Roe decision because many other social and political changes also affect abortion. If you can easily include other effects in your student project, do so. But don't give up on an interesting project because you are missing data. If you have a time series of abortion rates, compare the pre-Roe vs post-Roe rates.

Take heed: For seasonality, see the project template on daily temperature and the discussion forum posting on seasonality.

Take heed: For cyclical models, see the discussion forum postings on macroeconomic indices and the project templates on real interest rates.

INDEPENDENT STUDENT PROJECT: DESIGN, DATA SETS, AND ANALYSIS

Jacob: The student project applies the time series concepts to an actual time series. Do we apply specific concepts? Do we use a specified time series? How do we know what techniques to use and which data set to apply them to?

Rachel: You can choose the time series, data set, and statistical techniques. You have just taken a course on time series analysis. Use the techniques taught in the course to analyze a time series of your choosing.

Your project can be independent or it can follow the templates on the discussion board.

You will enjoy the project more if you choose a time series that interests you. But open-ended assignments are sometimes baffling. Use the project templates and the student projects posted on the discussion forum two ways:

- The project templates give you ideas. Work through the project templates on interest rates and daily temperature. The project templates give you step-by-step instructions. Download the time series and reproduce the tables, charts, and graphics.
- You can do a student project modeled on a project template. Instead of 90 day Treasury bill yields, use a CPI index or a corporate bond yield.

Jacob: What exactly do we choose? The discussion forum has instructions and step-by-step guides. Do we follow these instructions?

Rachel: Your choice has three levels: design, data sets, and analysis. The instructions and step-by-step guides lead you through setting up and starting a student project. They don't finish the work; no project template completes the project. You decide what your project will cover, and the final form of the project depends on your choices.

You can *design* your own project. The project templates show the type of analysis, but your project can be different. You might use employer data, client data, or publicly available data showing a time series.

- *Pricing:* you might use quarterly premium volume over the past ten years.
- *Investments:* you might use a firm's weekly stock price over the past five years.
- *Social:* you might use U.S. crime rates over the past fifty years.

Take heed: Crime rates are compiled by the FBI and municipal police departments and analyzed by politicians, social scientists, and journalists.

Jacob: Could we use annual profit margins over the past ten years?

Rachel: A time series of ten values is too short for the statistical analysis. The NEAS web site shows monthly interest rates for 55 years, giving enough data to examine the effects of different eras and to test for significance of residuals.

Daily temperatures are ideal for time series analysis. The national weather service provides free data bases showing high and low daily temperatures (and much more weather data) for over a thousand locations for the past hundred years. These data have the time series relations modeled by ARIMA processes:

- Movements of cold and warm fronts cause autoregressive and moving average effects.
- Daily temperature is seasonal.
- Trends in daily temperature are unclear; you can examine possible warming or cooling in your home town over the past hundred years.

A project may also compare similar time series, such as time series models for premium volume (or loss costs) in two states or for interest rates in two periods.

Take heed: Do not worry that the results may not be significant. If you want a particular topic but the only time series you find has just 50 observations, use the data. Your write-up will note that the results are distorted by random fluctuations, but you will enjoy the work.

Jacob: What if we don't have our own data? Do you give us data and a project template?

Rachel: Data are available on hundreds of web sites. But if you have never used ARIMA models, you may not know what time series are best for the student project. We provide data sets and templates on the discussion board.

Recommendation: You enjoy the project more if you choose your own topic. Pick a topic and use Google or another search engine to find data on the web. Add keywords like *history* or *trend* to the search criteria.

Illustration: If you search for *daily temperature*, you get today's weather. If you search for *daily temperature history*, you get historical time series.

Jacob: What time series do the project templates use?

Rachel: We form a time series model for interest rates in the United States. We also post related time series on the web site: inflation, GNP, unemployment, exchange rates, and interest rate futures. Another project template shows ARIMA modeling of daily temperature. We show a long project template on daily temperature and shorter templates on various other topics.

Jacob: People have tried to predict interest rates for years. The interest rate models used by actuaries and economists are complex. Are we supposed to improve on these models?

Rachel: The goal of the student project is not to develop the ideal model.

- The project takes actual interest rates and applies the methods in the textbook. You are learning how to apply the concepts, not developing better interest rate models.

- The project template on daily temperature shows how to adjust for seasonality. We don't use ARIMA processes to forecast the temperature.

Jacob: Interest rates comes in dozens of forms. What time series do we use for the project template on interest rates?

Rachel: We illustrate with 90 day Treasury bills and overnight LIBOR rates, and we provide guidance for other rates. You can pick among several dimensions: long vs short rates, nominal interest rates vs real interest rates, absolute values vs residuals, and spot vs future rates. For example:

- Short rate: three month Treasury bills, overnight LIBOR
- Long rate: twenty year Treasury bonds, Moody's corporate bond rates
- Residuals of three month Treasury bills regressed on the CPI

We provide the time series on the web site along with specific project templates.

Recommendation: If you feel lost, begin with three month Treasury bills, Atlantic City daily temperature, or another project template on the NEAS web site.:

- Compare your results with the examples in the project templates.
- If you can't reproduce the time series charts, post a question on the discussion forum.

After you work through the statistical techniques (correlograms, Box-Pierce Q statistic) ARIMA modeling will make more sense. Select another time series for your student project:

- Another interest rate time series, or another macroeconomic index.
- Daily temperature for another weather station.

Once you are comfortable with the statistical techniques, you can choose any time series and do a good project.

COMPARISON STUDENT PROJECTS

Many student projects compare two or more time series.

- Several types of interest rates, several time periods, or real vs nominal rates.
- High vs low temperatures, daily temperatures in different weather stations or in different time periods.

Do not worry about choosing a correct topic. We require only that the time series is not already a white noise process or a random walk.

- No topic is inherently right or wrong for a student project.
- The keys to a good project are an interesting question on an interesting topic.

An interesting question might be

- How well can we predict Friday's daily temperature with the daily temperatures on Monday through Thursday?
- Does the time series pattern of daily temperature differ on the East Coast (New York) vs the West Coast (San Francisco)?
- Has the pattern of daily temperature changed between 1901-1950 and 1951-2000?

You may be surprised by the answers to the questions above.

STRUCTURAL MODELS VS ARIMA MODELS

Jacob: Interest rates depend on current inflation, expected future inflation, economic growth, Federal Reserve Board policy, and business cycles. Don't we have to consider these variables to forecast interest rates?

Rachel: The textbook discusses this issue several times. The best model for any time series uses all the relevant explanatory variables.

- Sales of pharmaceutical firms depend on the quality of new medications, patents for existing medications, and (perhaps) marketing. To forecast drug sales, we consider outstanding patents and the new medications under development.
- Premium volume for an insurer depends on its rate level compared to other insurers, the quality of its sales force, and special marketing efforts.
- The daily temperature depends on a host of weather variables, not just the temperature on the past few days.

The textbook gives several answers.

Answer 1: Structural models, which rely on other explanatory variables, are complex. They work well in hindsight but have little predictive power if they rely on unknown values, such as future inflation or business trends. The time series is a simple model that may perform nearly as well.

Answer 2: Most effects on a time series are gradual, hard to observe, and hard to quantify. An insurer's premium volume depends on rate level changes, competitors' actions, or the opening of a new sales office. When forecasting the insurer's sales, we may not have this information, and we don't have models linking competitors' actions to the premium volume.

These exogenous causes affect the time series values for both past and future periods. An ARIMA model may be a good forecasting tool.

Answer 3: The absolute level of the time series (such as interest rates) depends on other economic influences. But the remaining effect on interest rates – the residual – may not be random. We estimate a time series of the residuals. Chapter 19 of the textbook shows examples of this. Your student project may focus on residuals; we explain the intuition.

Residuals

Jacob: Suppose we forecast interest rates from inflation rates. How can ARIMA modeling of the residuals help? Don't we assume the residuals of a regression model are random? If the residuals are not random, are the results of the regression analysis still valid?

Rachel: Classical regression analysis assumes the residuals are a white noise process if the explanatory variables completely explain the dependent variable and the regression

equation is correct. But we never know all the explanatory variables, and the true relation is not exactly linear. For a time series, the residuals often have serial correlation.

The regression analysis course uses the Durbin-Watson statistic. This is similar to the autocorrelation of lag 1, but scaled from 0 to 4 instead from -1 to 1 .

The Durbin-Watson statistic tests for serial correlation. Ordinary least squares estimators are unbiased even with serial correlation, but the statistical testing is no longer accurate.

Using a time series model on the residuals is more sophisticated, in two ways:

- We examine more complex relations than just the autocorrelation of lag 1.
- We use the relation among the residuals to better forecast the values of the time series.

Illustration: Fisher Effect

Financial economists assume a Fisher effect between interest rates and inflation rates. The nominal interest rate is the inflation rate times the real interest rate, such as 1.02 , or plus the real interest rate, such as 2% .

In the long-run, interest rates and inflation rates are correlated. In the short run, the relation is weaker. If inflation rises by 50 basis points, the interest rate may rise by 25 basis points.

Jacob: Does the student project test the Fisher effect?

Rachel: We assume the Fisher effect holds, and we test whether a time series model is appropriate for the residuals. A financial economist may presume that the relation between interest rates and inflation rates has a long-term mean, such as 200 basis points, and a strong autoregressive quality in the short run.

Illustration: Suppose the long-term real interest rate is 2% for an additive model or 1.02 for a multiplicative Fisher model. If the real interest rate is 4% in January 20X6, we don't expect it to stay 4% forever or to become 2% in February 20X6. The real interest rate may drift back to 2% over the next year or two.

Your student project may compare ARIMA models for nominal and real interest rates. If inflation is volatile, an ARIMA process may forecast real interest rates well, but not nominal interest rates.

Take heed: The student project is most interesting if it examines structural relations and it allows you to include more statistical techniques.

Illustration: A time series on Chicago crime rates fulfills the VEE requirements. Adding other pieces makes the project more interesting. Look at population densities, police force policies, gang activity, and similar items.

Seasonality

Interest rates and inflation rates may vary over the year.

- In December, consumers shop for holiday gifts, and the demand for money is high.
- In January, they return to work, and the demand for money falls.

The Federal Reserve Board adjusts the money supply each month to dampen fluctuations in inflation rates and interest rates.

- The adjustments are not perfect, and some weak seasonality remains.
- The seasonality is strongest in 30 day commercial paper rates, weak but noticeable in 90 Treasury bill rates, and smoothed in smoothed in yields of one year Treasury bills and longer securities.

Your student project can examine seasonality in various types of rates. For Treasury securities, the seasonality is strongest in monthly first differences of 90 day Treasury bills.

The student project may also examine the seasonality in nominal interest rates vs real interest rates.

- The seasonality affects inflation rates and nominal interest rates, not real interest rates.
- Use the raw CPI to derive real interest rates.
- The seasonally adjusted CPI does show the seasonal fluctuation.

Your student project can examine residuals of interest rates regressed on other indices.

Illustration: Regress interest rates on GDP (also on the NEAS web site.)

- Economists do not agree on the expected regression coefficients.
- The FED often tries to moderate GDP growth by raising or lowering interest rates.
- It is unclear if the FED can influence interest rates or GDP. The macroeconomic course explains the views of different financial economists.
- Your student project may compare ARIMA models fitted to the nominal interest rates vs fitted to the residuals of this regression.

The NEAS web site shows no index for the *expected* inflation rate. Use the actual CPI as a proxy for the expected inflation rate. This is not an ideal proxy, but we are teaching statistics, not economics.

Four Step Process for ARIMA Modeling: Specification

STATIONARY INTEREST RATE SERIES

The textbook's four step process for modeling time series begins with specifying the model. Form a stationary time series. Several scenarios may occur. For interest rates:

- The time series is not stationary and cannot be converted into a stationary series. Most interest rates series are stationary or homogeneous non-stationary (so their differences are stationary). Economies with run-away inflation have non-stationary nominal interest rates and probably non-stationary real interest rates. ARIMA modeling of inflation rates in Zimbabwe does not work.
- The time series (or its logarithm) is stationary. Most mean reverting time series are stationary. Logarithms change multiplicative models into additive models. For interest rates, we don't usually take logarithms. For dollar denominated items (sales, premium, stock prices), take logarithms. See the discussion forum posting on random walks.
- The first differences of the time series (or of its logarithms) is stationary. If the time series is growing or declining (not mean reverting), take first differences. Economists disagree whether interest rates are mean reverting. Your student project examines the correlograms of the time series and its first differences. If the time series is not stationary but its first differences are, your report should explain why this occurs.
- More complex models using second differences might be needed. Second differences are not used for interest rates. Second differences might be used for a time series of the balance in an account where monthly deposits are made.

If a time series is not growing or declining, it is not seasonal, it is mean reverting, and it has no changes in its mean, it is usually stationary.

- Nominal interest rates may have a upward or downward drift (reflecting inflation) in some periods. Other times they have no drift. If your student project uses nominal interest rates (or any other index that moves with inflation), you may have to use separate periods or take differences.
- It is unclear if real interest rates are mean reverting. Examine the sample autocorrelation function (the correlogram) to determine if a time series is stationary. Use real interest rates for a better model.

A student project may examine differences between short rates, long rates, and residuals of interest rates on inflation rates.

- The short rate fluctuates more than the long rate.
- Their mean reversion is hard to predict.

Many economists assume the interest rate residuals (on inflation) are mean reverting.

- If the time series is not mean reverting, model it as a random walk.

- If interest rates have strong mean reversion, use a AR(1) or AR(2) process.

Deciding whether to model the initial time series or its first differences is critical. Graph the time series. If it has a drift, it is not stationary, so take first differences.

The interest rate graphs on the web site show drifts that vary by time period. Your student project examines if the time series changes.

Illustration: Suppose you examine the daily high temperature in City Z for 1875 - 2005.

- For 1875 – 1940, the daily temperature is recorded every six hours. The high temperature is the 12:00 noon reading.
- For 1941 – 2005, the daily temperature is recorded every hour. The high temperature is usually the 2:00 pm or 3:00 pm reading.

The temperature at 2:00 pm is 3 or 4° higher than at noon. Any single year gives a stationary time series, but the full time series of 131 years is not stationary.

Changes in the average daily temperature from smog, urbanization, ocean currents, or global warming may cause your time series to be non-stationary.

Drifts may be linear or exponential.

- If the drift is linear, take first differences.
- If the drift is exponential, take logarithms and then first differences.

Take heed: The project templates recommend detrending the time series instead of taking differences. If you eliminate the trend with an inflation index, an ARIMA process fits better.

If no drift is evident, we examine if the time series is mean reverting. A mean reverting time series moves up if it is below the mean and down if it is above the mean. A mean reverting time series can be either oscillatory or asymptotic. Many actuarial and financial time series have an exponential drift or are not mean reverting. Interest rates may or may not have a drift. Real interest rates are generally mean reverting; nominal interest rates may or may not be mean reverting.

Jacob: If the time series has no drift and is mean reverting, is it stationary?

Rachel: A time series is not stationary if the mean or variance changes. The project template on interest rates shows how FED policy affects your ARIMA modeling.

Illustration: In the 1970's, the U.S. tried to use inflation to control unemployment. In the early 1980's, the Federal Reserve Board slowed the money supply growth to restrain inflation. The mean nominal interest rate rose and then fell by over 500 basis points in a few years, and the variance of the interest rates increased and then declined significantly.

You have two possible solutions:

- Separate the time series into three eras (regimes).
- Use the residuals of interest rates on inflation.

Using second differences is not correct.

Jacob: How do we find and test for a change in regime?

Rachel: We use exogenous knowledge; we have no specific statistical test. We examine the plot of interest rates and see if a change in the mean, variance, or drift seems likely. We show an illustration on the discussion forum: the mean, variance, and drift of interest rates changes materially among three periods.

If we suspect a change in regime, we use two or more time periods, often with an interlude in between. For interest rates, we show illustrative time periods of 1945-1978, 1979-1982, and 1983-2000. For your student project, you can divide the time period differently. We re-do the time series analysis on each part.

Jacob: If the ARIMA models fit better by part, do we use separate models?

Rachel: We get more stable results in each part even if there are no real changes. Unless the change is large, we don't use separate parts.

Illustration: If we get a mean interest rate of 8% in one period and 7% in the second period, we assume this is random fluctuation and use a single time series. If the means are 11% and 5%, we assume a regime change and use two periods.

We form correlograms for each period. *The correlogram should validate the intuition from looking at the plots.* For the student project, you form the plot, make a guess about the type of time series, and then form a correlogram to validate your inference.

Speed of Decline

If the sample autocorrelations do not decline to zero, or if the decline is *too slow*, the time series is not stationary. We take first differences (and perhaps logarithms) to form a stationary time series.

Jacob: What is *too slow*? Are sample autocorrelations of $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, ..., $\frac{1}{n}$, ..., too slow?

Rachel: *Too slow* means the time series is not autoregressive or moving average.

- If the time series is moving average, the autocorrelation drops to zero after period Q, where Q is the order of the moving average model.
- If the time series is autoregressive, the autocorrelations have a geometrically declining envelop after period P, where P is the order of the model.

The sample autocorrelations in your example show too slow a decline.

Jacob: Should we examine also the second and third differences?

Rachel: We stop taking differences as soon as we get a stationary series. Few time series are homogeneous of order more than one. It is sometimes worthwhile to look at the second differences, but that is not required for the student project.

Jacob: Are the interest rate time series all similar?

Rachel: The time series differ by type of interest rate and by era. The short rate may have a varying mean, the long rate may be a random walk, and the real interest rate (the residual) may be white noise. Your student project can compare the time series.

Length of Periods

Jacob: How long a period do we use? Some interest rates on the web site start in 1945. Do we use all 60 years?

Rachel: The length of the time series depends on two items:

(1) For a short period, the model parameters depend on factors we don't want to include.

- A single year of monthly interest rates is not enough. The short run fluctuations depend on economic factors specific to that year. The ARIMA model will not forecast well.
- Random fluctuations distort the sample autocorrelations for short periods. For a single year of monthly rates, the standard deviation of the sample autocorrelations is $1/\sqrt{12} = 28.87\%$ even if the time series is white noise. Even an autocorrelation of 50% may be random fluctuation. The observed time series doesn't tell us much.
- Seasonal patterns require a series of several years. The student project should test for seasonality, so we use several years.

Jacob: Don't the statistical tests take random fluctuations into account?

Rachel: If the period is too short, the time series is affected by other factors and is not correctly specified. *In-period* goodness-of-fit tests are good: R^2 is high and the Box-Pierce Q statistic is low. But the time series forecasts poorly; we get poor out-of-period results.

(2) For a long period, other factors affect the model. The depression years, World War II, and the modern period have different interest rate models. Putting them together gives a non-stationary time series. Monthly interest rates show one pattern when the Federal Reserve Board focuses on rate stability and a different pattern when the Federal Reserve Board focuses on unemployment.

Choosing Periods

Part of the student project is choosing the proper length of the time series. We illustrate with post World War II years: 1945 and onward. You can choose other periods if you want.

Jacob: How do we use to choose the proper length?

Rachel: We look for stability of the model coefficients. Suppose we use historical data through December 2005 to predict interest rates in 2006. In-period goodness-of-fit tests don't tell us if the parameters are stable, so we use out-of-period tests. We form three time series models:

- Data through December 2002 to predict interest rates in 2003.
- Data through December 2003 to predict interest rates in 2004.
- Data through December 2004 to predict interest rates in 2005.

We examine two things:

- The ARIMA model gives a forecast variance. If the actual interest rates are sufficiently close to the forecasted interest rates, we don't reject the model. *Close to* means within one or two standard deviations.
- The ARIMA model gives standard errors for the time series coefficients: the ϕ_1 , ϕ_2 , and so forth. We examine if the coefficients estimated in different periods are the same. For data from 1985 through 2005, we might use three periods: 1985-1991, 1992-1998, and 1999-2005. Each period has seven years of monthly rates, or 84 data points. The samples are large enough that a change in the time series coefficients indicates the model is not good. If the standard errors of the regression parameters are 5 basis points, a ϕ_1 of 20% in one period and 30% in another period, we use separate ARIMA models. (Standard errors of the regression parameters are covered in the regression analysis course.)

Autocorrelations

Jacob: After choosing between the interest rates and the first differences and selecting a length for the series, what is the next step?

Rachel: We examine the sample autocorrelations and the partial autocorrelations.

Jacob: How many lags do we use? Do we focus on the sample autocorrelation for one lag at a time or for all lags?

Rachel: To specify a time series model, we examine one lag at a time, using the principle of parsimony. For diagnostic testing, we examine the Box-Pierce Q statistic for a large number of lags.

Jacob: Is there an order for the models we try? Or do we try a bunch of ARIMA models to see which fits best?

Rachel: The textbook implies that the sample and partial autocorrelation functions indicate the proper type of model. The textbook authors examine various models, comparing the R^2 and the Box-Pierce Q statistic for each.

In practice, we use only four or five models. Most ARIMA models are AR(1). We use an AR(1) model and examine the sample autocorrelations of the residuals from this model. If these autocorrelations are not statistically significant and the Box-Pierce Q statistic is not significant, we assume the time series can be modeled as AR(1).

Out-of-Sample Tests

Jacob: The AR(1) model may not be optimal, even if the fit is reasonable. Suppose the sample autocorrelations are 50% for lag 1 and 30% for lag 2. If the time series were AR(1) with a ϕ_1 of 50%, the sample autocorrelation of lag 2 would be 25%. Since we get a 30% autocorrelation for lag 2, we assume a small but positive value for ϕ_2 .

Rachel: If the standard error of the sample autocorrelations is 15%, the difference between 25% and 30% is not significant.

Jacob: Even if it is not significant, the AR(2) model fits better. Is there any harm in using the AR(2) model?

Rachel: Using a higher order increases the in-sample goodness-of-fit, but the model may not be better. In the example you gave, most statisticians would bet that even though the AR(2) model has the better in-sample fit, the AR(1) model forecasts better.

Jacob: How do we test the out-of-sample goodness-of-fit?

Rachel: Suppose we use monthly interest rates for 1985 through 2005 to forecast future interest rates. Our final model uses all the historical data. To build the model, we use interest rates for 1985 through 2004. We may specify two or three models. The in-sample goodness-of-fit tests are not decisive. The optimal model is the one that minimizes the mean squared error of the forecast.

Jacob: Do we choose the model with the lowest mean squared error for the forecast?

Rachel: It is easy to pick the model that minimizes the *observed* squared error; it is hard to pick the model that minimizes the *mean* squared error.

Jacob: What is the difference?

Rachel: Interest rates are stochastic. Suppose two models fit equally well in-sample. We may have several scenarios for next year's interest rates. For some scenarios, one model fits better; for other scenarios, the other model fits better.

Jacob: How do we solve this problem?

Rachel: One solution is to examine various periods, or interest rates in various countries, or various interest rate series (Treasury bills, LIBOR rates, ...).

Alternatively, we rely on intuition and experience with time series models. The intuition for an AR(1) model is strong. Most time series show stability; if this period's value is high, next period's value may also be high. Other relations are rare: negative serial correlation and ARIMA models with $p+q > 2$.

Seasonality

Jacob: How do we include seasonality? The textbook uses various seasonal adjustments.

Rachel: Seasonal correlations should be reasonable. High-renewal annual policies have a 12 month autocorrelation in premium volume. Interest rates depend on the demand for money, which varies seasonally. For any time series, we begin with AR(1) and a seasonal adjustment, such as AR(4) for quarterly data or AR(12) for monthly data.

Illustration: Suppose we model interest rates as a ARIMA(2,1,0) time series, using 20 years of quarterly data. We also model the rates with a four quarter seasonal autocorrelation. The textbook calls this ARIMA(4,1,0). Other statisticians use notation to specify that the non-seasonal part is ARIMA(2,1,0) and the seasonal part is ARIMA(4,1,0), with a *single* autocorrelation for lag 4.

Jacob: How do we test for seasonality?

Rachel: We examine the quarterly sample autocorrelations of lags 4, 8, 12, and 16. Seasonality appears as spikes in the correlogram. To test whether the 4 quarter seasonal lag helps, we examine the Box-Pierce Q statistic for the seasonal vs non-seasonal models.

Jacob: Do we use the time series model with the lowest Box-Pierce Q statistic?

Rachel: We consider also the principle of parsimony. More complex models have lower sample autocorrelations. If the Box-Pierce Q statistic adjusted for degrees of freedom is not lower, we use the simpler model. If it is lower, the choice depends on how much lower.

First Differences and Seasonality

Jacob: How do we combine first differences and seasonality?

Rachel: We have several methods.

(1) We take first differences of the time series and examine the four quarter lag of these first differences. If we examine toy sales, the first differences are positive for third to fourth quarter, negative for fourth to first quarter, and zero elsewhere. Similarly, if the Federal Reserve Board did not adjust the money supply for the higher demand in the fourth quarter, we would expect a large seasonal effect on interest rates and inflation rates.

Jacob: Would interest rates be higher or lower in the fourth quarter of the year?

Rachel: Economists disagree about the effects of monetary policy on interest rates. Keynesian economists assume that a demand for money exceeding the supply of money causes real interest rates to rise, as consumers sell bonds and seek loans. Neo-classical economists assume money is neutral and has no effect on real interest rates, but a demand for money exceeding the supply of money reduces the price level, pushing down nominal interest rates. Your student project may examine the seasonality of interest rates.

(2) We take first differences of the interest rate from the rate lagged twelve monthly periods. The time series is the change in the interest rate from its level 12 months ago.

Illustration: Suppose interest rates are about 150 basis points higher in the fourth quarter of each year. The time series we examine is the 1/20X6 interest rate minus the 1/20X5 interest rate, the 2/20X6 interest rate minus the 2/20X5 interest rate, and so forth.

(3) We de-seasonalize the time series and take first differences from month to month. We first compute factors to de-seasonalize the interest rates, such as subtracting 150 basis points from the fourth quarter rates. We then take month to month first differences.

(4) We take both month to month differences and 12 month differences. This is a second difference, of the form $y_t - y_{t-1} - (y_{t-12} - y_{t-13})$.

(5) We take using first differences and an autoregressive model with non-zero parameters for lags 1 and 12.

Jacob: Which method should we use?

Rachel: Your student project can compare several methods of adjusting for seasonality. Often the simplest method is best. The textbook uses the last method above.

Mean Reversion

Jacob: When do we estimate an autoregressive model from the interest rates themselves and when do we use first differences?

Rachel: If mean reversion is strong, we use the interest rates. If mean reversion is weak (or zero), we use first differences.

Jacob: From 1990 through 2005, interest rates have been low; from 1975 to 1985, interest rates were higher. If there were strong mean reversion, interest rates should have regressed toward the mean in both periods.

Rachel: We infer that the mean changed, not that mean reversion is weak. A student project may compare various periods to see whether different models are appropriate.

Interest Rates on the Web Site

Jacob: What interest rates are on the web site?

Rachel: You can use any interest rate you want, but it is difficult to discuss the methods on the discussion forum unless others are using the same rate. We have selected a few interest rates that are useful for the student projects:

- Three month Treasury bills by month from January 1931 through June 2000. Some years are missing, when auctions were not held.
- Twenty year Treasury bonds by month from April 1953 through December 2005. Some years are missing, when auctions were not held.
- Moody's seasoned AAA bond rates, by day, from January 3, 1983, through January 16, 2005.
- The CPI from January 1913 through December 2005, both (SA) seasonally adjusted and (NSA) not seasonally adjusted. There are many CPI indices; we show U.S. city average, all items.

Whatever series you use, choose a period during which the time series is stationary but a long enough period for statistical testing. We will put up additional series on the web site as time goes on.

Jacob: Can we take another interest rate time series from the internet?

Rachel: You can use any interest rate time series you want. Many financial economists believe that the time series models work better for interest rate *futures* than for the interest rates themselves. An excellent student project may compare a model for 20 year Treasury bonds with a model for Treasury bond futures. You can do the same for LIBOR rates and Eurodollar futures.

Jacob: Is the grading of the student project easier or harder if we choose a different time series or a different test than those in the project templates?

Rachel: The grading is easier. If you follow the project template, you should demonstrate that you understand the concepts by performing the various statistical tests. If you design another project, we are less concerned with performing all the statistical tests. If you select a reasonable item to test and you show how this is done, that is sufficient.

DIFFERENCES

Take heed: Do not take differences unless they are needed.

Jacob: Do the first differences of a mean reverting time series give a better model?

Rachel: Taking first differences of a stationary time series makes the forecasts less robust. Once we have a stationary time series, we use it.