INTUITION, PARSIMONY, AND STATISTICAL TESTS: FITTING ARIMA MODELS

Updated: April 19, 2006

Jacob: How do we judge which ARIMA model is best?

Rachel: We want to forecast future values. The best model is the one that gives the lowest mean squared error of the forecasts.

GOODNESS-OF-FIT TESTS

Jacob: Why do we emphasize the Durbin-Watson statistic, Bartlett's test, and Box-Pierce Q statistic? These are in-sample goodness-of-fit tests; they do not test forecast accuracy. Shouldn't we use out-of-sample tests? Should we test which ARIMA model best forecasts interest rates over the next 12 months?

Rachel: Out-of-sample tests are important, but they don't always work well.

Illustration: Suppose interest rates follow a random walk with a drift of zero and a high stochasticity. The interest rate at December 20X7 is 8%. We compare three models:

~ A random walk with a drift of +5 basis points a month.
~ A random walk with a drift of zero basis points a month.
~ A random walk with a drift of –5 basis points a month.

Consider three scenarios, all of which are reasonable.

If the stochastic error term in January 20X8 is +30 basis points, the average forecast for the rest of 20X8 is 8.30%, and Model A gives the best forecasts.

If the stochastic error term in January 20X8 is –30 basis points, the average forecast for the rest of 20X8 is 7.70%, and Model C gives the best forecasts.

If the stochastic error term in January 20X8 is 0 basis points, the average forecast for the rest of 20X8 is 8.00%, and Model B gives the best forecasts.

Jacob: What does this illustration show?

Rachel: The actual interest rates for the 12 months of 20X8 are not independent. They are one realization of an interest rate path, not 12 independent realizations of future interest rates. The actual interest rates are one scenario out of many, and they do not confirm or disprove a model.

The in-sample goodness-of-fit tests are often good predictors of forecast accuracy. A model that has a low Box-Pierce Q statistic for its residuals is likely to have a low mean squared error for its forecasts. This is not *always* true, but it is reasonable.

Jacob: For regression analysis, we use an R^2 test; why don't we use it for time series?

Rachel: For many ARIMA models, we take first differences to get a process that is close to white noise. The fitted process has a slope coefficient close to zero and a low R^2. If the slope coefficient is zero, the R^2 is zero. We do not use R^2 to compare different types of ARIMA models, and we do not use R^2 when the slope coefficient is close to zero.

INTUITION

Jacob: How does intuition affect the choice of the optimal ARIMA model? We emphasize intuition, but the concept seems so vague. Why not just use the goodness-of-fit tests?

Rachel: Suppose we examine medical malpractice average claim severity over 20 quarters (5 years). We find an upward trend, which we can model as either a linear trend or an exponential trend. We might say: Let us consider two ARIMA models:

~ We take first differences to get a linear trend
~ We take logarithms and first differences to get an exponential trend

We use the model with the better in-sample fit to forecast future values. Is this what you mean by goodness-of-fit tests?

Jacob: Yes; if the trend is exponential, the exponential trend should fit better.

Rachel: Actuaries do not do this. If a $10,000 claim this year will grow to an $11,000 claim next year, we assume that a $20,000 claim this year will grow to a $22,000 claim next year, not a $21,000 claim. We assume intuitively that the exponential trend is correct and will have the lower mean squared forecast error. If the linear trend has the better in-sample fit, we attribute that to random fluctuation. Strong intuition out-weighs the statistical tests.

Jacob: In general, which is more important: in-sample tests or intuition?

Rachel: That depends on the stochasticity of the time series values. If the stochasticity is low, we emphasize the in-sample goodness-of-fit tests. If the stochasticity is high, we emphasize our intuition about the reasonableness of each ARIMA model.

Jacob: How does this relate to interest rates? Do we know the structure of interest rates?

Rachel: We don't know the structure, but we assume that today's interest rate is a good predictor of tomorrow's interest rate.

~ Some financial economists assume we can't make better estimates.
~ Some financial economists believe that the drift in interest rates is likely to persist.
~ Some financial economists believe that interest rates are mean reverting.

These views justify random walks or AR(1) models. We begin with an AR(1) model, either of the interest rates themselves or of their first differences.

Jacob: Isn't the assumption based on the empirical data? Why not fit the empirical data to as many ARIMA models as possible and see which one fits best? That will tell us which assumption is most reasonable.

Rachel: ARIMA modeling has two parts: we specify the model and estimate parameters.

~ The intuition is most important for the specification of the model.
~ The immediate past data are most important for estimating the parameters.

Illustration: Financial economists assume that stock prices in an efficient market follow a random walk. This assumption is based on arbitrage arguments, not on the past stock prices for any given stock. We use past data to estimate the drift and volatility of the stock price, not to choose among ARIMA models.

The same is true of interest rates. A financial economist might form an ARIMA model of twenty year Treasury bond rates in 2006 from observed Treasury bond rates in 2000-2005. The economist's intuition about the structure of the model may be based on long-term risk-free rates in many time periods and many countries, and the parameters are based on the twenty year bond rates in 2000-2005.

Jacob: What exactly does intuition mean? We say that ARIMA models should be intuitive; what makes one model more intuitive (more reasonable) than another?

Rachel: Intuitive means: "If we had no data about the particular item we are forecasting, what type of model would we use?" This is similar to a Bayesian prior: what do we expect before we examine the data?

Jacob: What ARIMA models are most intuitive (most reasonable)?

Rachel: The answer depends on the time series. Financial economists dealing with efficient markets use random walks. Actuaries dealing with random fluctuations use higher order autoregressive models. Marketing personnel emphasize seasonality. Economists dealing with industry demand for durable goods use moving average models.

Jacob: Can you explain the difference between financial economists and actuaries? *(Financial economists dealing with efficient markets use random walks. Actuaries dealing with random fluctuations use higher order autoregressive models.)*

Rachel: The stock market is efficient. The consensus among financial economists is that an ARIMA(0,1,0) model is optimal for stock prices. Economists examine past data to judge the drift and volatility of stock price movements, not to specify the type of model. We don't use stock prices for the student projects because the arbitrage arguments for an ARIMA(0,1,0) process are so strong that any other model is hard to justify.

Jacob: Why do actuaries often use higher order ARIMA models?

Rachel: Financial economists modeling interest rates do not worry about sampling error. The interest rate may be rounded to the nearest basis point, and fluctuations in supply and demand may cause another 2 or 3 basis point errors, but this is minor. Actuarial statistics have great random fluctuations. Last year's personal auto claim frequency may be the best forecast for this year's claim frequency, but last year's frequency may be distorted by random fluctuations. Instead of a random walk, the actuary may use a three year weighted average, such as a 20%-30%-50% weighting. For low frequently / high severity items, the actuary may use a high order autoregressive process.

Jacob: What about interest rates? Are they like stock prices?

Rachel: The intuition for interest rates is less clear. Financial economists disagree about the optimal model, so interest rates are ideal for student projects.

Jacob: Why are interest rates different from stock prices? For stock prices, we assume the first differences are a white noise process (after taking logarithms). If the long-term average change in stock prices is 1% a month, our forecast for the February stock price change is 1%. It stock prices rose 2% in January, we don't change our forecast for February.

Interest rates have a long-term average movement of zero. The average risk-free interest rate over the past decade is about the same as it was 50 years ago and 100 years ago. Just because interest rates rise (or fall) in one month, why should we assume they will rise (or fall) in the next month?

Rachel: Compare interest rates and the weather. The long-term average temperature on May 1 in Chicago doesn't change much from decade to decade. The simplest forecast for the May 1 temperature is the average May 1 temperature over the past 100 years. If this average temperature is 45°, our time series is 45° + ϵ, which is a white noise process.

An ARIMA model does better than the white noise forecast. If the temperature on April 31 was 55°, we might forecast 50° for May 1. If the temperature on April 31 was 35°, we might forecast 40° for May 1. This is an AR(1) model with ϕ_1 = 50%.

Jacob: This justifies an AR(1) model, which is a reasonable model for both interest rates and local temperature. Is there any justification for a higher order model?

Rachel: If the temperature on April 31 was 45°, the AR(1) model forecasts 45° for May 1. Perhaps we can do better.

~ If the temperature on April 30 was 40°, we might presume a warm front is moving in, and our forecast for May 1 may be higher than 45°.
~ If the temperature on April 30 was 50°, we might presume a cold front is moving in, and our forecast for May 1 may be lower than 45°.

We might form an AR(2) model to represent this process. Interest rates may be similar.

~ The average interest rate may be 8% over the past century. The simplest model is a white noise process, with a mean of 8%.
~ Interest rates change slowly, so we may give 80% weight to last year's interest rate and 20% weight to the long-term average of 8%. This is an AR(1) model with ϕ_1 = 80%.
~ Interest rates are not traded in an efficient market. If interest rates were 8.4% last month, 8.3% the previous month, and 8.2% the month before, we might presume that interest rates are rising. Numerous economic reasons may cause this: the economy may be recovering from a recession with high demand for investment funds or a budget deficit may create an over-supply of Treasury securities. An AR(2) process may forecast more accurately than an AR(1) process.

Jacob: If the cause of the rising interest rate affects our forecasts, shouldn't we use an econometric model, not an ARIMA model?

Rachel: ARIMA models are proxies for the true economic, financial, or other explanation. An economist who knows why interest rates are rising may make better forecasts than a statistician who uses ARIMA models. But economists understand interest rate movements imperfectly, and statisticians using ARIMA models may make reasonable forecasts.

PARSIMONY

Jacob: How does parsimony fit into this?

Rachel: First differences, second differences, AR(1), AR(2), and MA(1) models have intuitive rationales. They reflect economic influences on interest rates. If these influences had certain effects in past years, we assume they have similar effects in future years.

We have no intuition for the effects of residuals or interest rates from 3 or 4 periods back. We assume the past relations are random fluctuations. If we include them in the model, they are more likely to raise the mean squared error than to lower it.

Jacob: It sounds like parsimony is a part of intuition; is that true?

Rachel: Yes. Parsimony does not mean that we prefer a linear trend to an exponential trend. We base the choice on the intuition for linear vs exponential trends. Parsimony means: if we have no explanation for a moving average or autoregressive factor, leave it out.

INTUITION VS DATA

Jacob: If the intuition is so important, why do we focus on graphs and correlograms?

Rachel: We rarely understand the causes of interest rate movements. Many economic factors affect interest rates, and the possible ARIMA process are varied. This makes interest rates ideal for the student projects.

An example is seasonality. Some candidates say there is no justification for seasonality in interest rates, so they treat any patterns as random fluctuations. A financial economist would observe the high demand for money in December and wonder why the seasonality is not stronger. An experienced economist would explain that interest rates were highly seasonal 80 years ago, but the Federal Reserve Board now adjusts the supply of money to mitigate the fluctuations in interest rates.

The same is true of many economic indices. Inflation, unemployment, GDP growth, and similar items are seasonal. A student project on any of these items might explore how the correlogram allows us to measure the seasonality. We are expanding the project templates to include other subjects. We have a conservative perspective: we introduction a new project template when candidates are comfortable with the current templates. Too rapid expansion leads to unnecessary confusion.