Time Series, Module 20: Model evaluation, historical simulation, and ex post forecasts

Homework Assignment

(The attached PDF file has better formatting.)

Updated: July 15, 2007

You construct an ARIMA process to model a time series for your student project. This homework assignment reviews the textbook method to choose an ARIMA process.

Jacob: Must we show these four steps in our student project?

Rachel: The student projects are diverse; some construct an ARIMA process, others compare two processes or two time periods, and others analyze specific elements of time series, such as seasonality. Some candidates model their student projects on this outline, but you need not follow this four step process. A separate posting on the discussion forum gives guidance for the write-up in your project.

The authors use a four step process for constructing a time series model:

- model specification
- parameter estimation
- diagnostic checking
- model evaluation

Suppose you construct an ARIMA model for one of the following time series:

- ♦ Policy applications by quarter for a single insurer
- ♦ Premium volume by quarter, either for a single insurer or for the industry
- ♦ Interest rates by month for five year Treasury notes
- ♦ Personal auto loss cost trends by quarter for the industry

Choose one time series and explain how you proceed along the four step process.

Jacob: Should we focus on the statistical items or the intuitive relation of the model to the insurance (actuarial) time series?

Rachel: The textbook shows the four step statistical procedure. The homework assignment applies these four steps to an actuarial time series. We emphasize the intuitive relation between the model and the insurance time series. The textbook cannot discuss intuitive relations since the readers may not have expertise in the subject. You are an actuary, and you will gain more if you consider how each model relates to the insurance time series.

From comments on the discussion forum:

Jacob: The homework question is open-ended; can you give more guidance? Can we use a different time series than the four listed above?

Rachel: You choose the insurance time series you want. This homework assignment applies the time series concepts to actuarial work. You may chose a different time series from your work, such as deaths or retirements or claim counts by quarter.

We note several issues that one considers when building a time series. You need not include these issues in your answer, but these are the issues you deal with in practice.

STATISTICAL ISSUES

For model specification, you determine the order of homogeneity, using first differences, second differences, and logarithms. In practice, we often use deflated figures instead of logarithms and first differences, particularly if the inflation rate is changing.

Jacob: Which method is better: deflated figures or logarithms plus first differences?

Rachel: Logarithms and first differences work well if the inflation rate is constant or if we are modeling the inflation rate itself. If we model another series, such as premium volume or stock prices in a currency that has wide swings in the inflation rate, it is better to use deflated values.

Take heed: The NEAS web site has several indices to deflate nominal dollars, such as the CPI. If you use another currency, such as Yuan, Yen, pounds, or Euros, you can usually find an inflation index on the internet.

We use the sample autocorrelation function and the partial autocorrelation function to determine the autoregressive and moving average orders. State how these two functions differ; a comparison to a derivative and a partial derivative is helpful. Say what a spike shows and what a geometric decline shows in the in the sample autocorrelation function and in the partial autocorrelation function. The textbook is concise, and it is helpful to discuss these two function on the discussion forum.

For parameter estimation, say whether we use linear estimates or non-linear estimates. The homework assignment asks you to identify the parts of the four step process.

Jacob: The discussion forum postings do not much discuss the partial autocorrelations. Do we use these in our work?

Rachel: The partial autocorrelations are harder to estimate with pencil and paper. They are not required for the student project.

INTUITIVE RELATIONS

For policies or premiums by quarter, we consider the term of the policy, the retention rate, and whether we use new business, renewal business, or both.

Renewal business for *annual* policies with a Z% retention rate has a form like $y_t = Z\% \times y_{t-4} + \epsilon_t$; for policies with six month terms, we use y_{t-2} in place of y_{t-4}.

Jacob: Do we use other ARIMA terms for the renewal business?

Rachel: The retention rate is stochastic. It may average 90%, but it may fluctuate between 88% and 92%. It may be that if the retention rate was 92% last quarter, we expect it to drift back slowly to 90%. If the volume of business is relatively even over the quarters, we might use a moving average model, with θ parameters of –0.5 for lag 1, –0.2 for lag 2, and –0.1 for lag 3. This varies; the homework assignment looks just for the autoregressive parameter of lag 4.

The new business may have either a moving average or an autoregressive form, and we have numerous possible adjustments for inflation, trends, seasonality, cycles, and so forth.

Illustration: If you model (annual term) permanent life insurance premium with a 92% retention rate by an ARMA(2,2) process of quarterly premium, something is wrong, since there is no autoregressive component of the fourth preceding quarter. For other industries, this year's sales are *not* a direct continuation of last year's sales. Movie receipts are highly seasonal, but this summer's receipts are not a direct extension of last summer's receipts. For insurance, the renewal premium this quarter is a continuation of the policy in effect last year. Insurance has new and renewal business, and each part has distinct characteristics.

Premium volume grows with inflation, loss cost trends, or volume of business. Premium volume is not stationary. Depending on the type of growth, you may use logarithms, first differences, second differences, or a combination.

Interest rates are often assumed to be mean reverting. You may use an autoregressive or a moving average model (or a combination). As the textbook shows, it may be better to model the first or second differences of interest rates.

Jacob: The models of interest rates in the textbook are unusual. At my company, we use much simpler models, such as ARIMA(1,1,0). We don't call it ARIMA(1,1,0), but it is an autoregressive model of lag 1 on the first differences of the interest rates.

Rachel: Financial economists argue about the best way to model interest rates. The model in the textbook seems complex. Many statisticians prefer parsimonious models, and think that any ARIMA model with $p + q > 2$ (except for quarterly or monthly seasonal terms) is suspect. This leaves only five models: AR(1), AR(2), MA(1), MA(2), and ARMA(1,1). We add first differences to each model to get an additional five options, and add logarithms plus first differences for another five options. These statisticians work with about 15 model before seasonal terms. This is a limited universe of models, which can be differentiated by their sample autocorrelation functions and partial autocorrelation functions.

The textbook

For *personal auto loss cost trends*, actuaries traditionally use exponential models. If claim severities are seasonal, you must include an autoregressive term of lag 4. Converting a series of average claim severities to a stationary time series requires taking logarithms and first differences.