TS Module 14 Model Diagnostics intuition

(The attached PDF file has better formatting.)

Diagnostic checking is especially important for the *student project*, for which you estimate one or more ARIMA processes and check which one is best. You use several methods:

- In-sample tests examine the Box-Pierce Q statistic or the Ljung-Box Q_* statistic to see if the residuals of the ARIMA model are a white noise process.
- Out-of-sample tests examine the mean squared error of the ARIMA models to see which one is the best predictor.

Diagnostic testing is both art and science. Random fluctuations and changes in the model parameters over time force us to rely on judgment in many cases.

The final exam tests objective items.

- Numerical problems test the variance or standard deviation of a white noise process, the value of Bartlett's test, or the computation of the Box-Pierce Q statistic and Ljung-Box Q_* statistic.
- Multiple choice true-false questions test the principles of diagnostic checking.

We review several topics that are often tested on the final exam.

We fit an ARMA(p,q) model to a time series and check if the model is specified correctly.

A. We compare the autocorrelation function for the simulated series (the time series generated by the model) with the *sample* autocorrelation function of the original series. If the two series differ materially, the ARMA process may not be correctly specified.

B. If the autocorrelation function of the ARMA process and the sample autocorrelation function of the original time series are similar, we compute the residuals of the model. We often assume the error terms before the first observed value are zero and the values before the first observed value are the mean.

C. If the model is correctly specified, the residuals should resemble a white noise process.

D. If the model is correctly specified, the residual autocorrelations are uncorrelated, normally distributed random variables with mean 0 and standard deviation 1/T, where T is the number of observations in the time series.

E. The Q statistic, where $Q = T \sum_{k=1}^{K} \hat{r}_k^2$, is approximately distributed as chi-square with (K – p – q) degrees of freedom. Cryer and Chan use a more exact statistic. The final exam tests both the (unadjusted) Box-Pierce Q statistic and the Ljung-Box Q_* statistic.

Statement A: Suppose the sample autocorrelations are 0.800, 0.650, 0.500, 0.400, 0.350, and 0.250 for the first six lags and we try to fit an MA(2) model.

- Use the Yule-Walker equations or nonlinear regression to estimate θ_0, θ_1, and θ_2.
- Compare the autocorrelation function for the model with the sample autocorrelation function of the original time series.

The autocorrelation function for the MA(2) model drops to zero after the second lag, but the sample autocorrelation function of the original time series does not drop to zero. We infer that the time series is not an MA(2) model.

This example is simple. Given the sample autocorrelations, we should not even have tried an MA(2) model. Other example are more complex.

This comparison does not have strict rules. No ARIMA process fits perfectly, and selecting the best model is both art and science. In a statistical project, we overlay the correlogram with the autocorrelation function of the model being tested, and we judge if the differences are random fluctuations.

Distinguish the two sides of this comparison:

- The sample autocorrelations are empirical data. They do not depend on the model.
- The autocorrelations reflect the fitted process. You select a model, fit parameters, and derive the autocorrelations.

There are different functions; be sure to differentiate them.

- The sample autocorrelations are distorted by random fluctuations. They are estimated from empirical data, with adjustments for the degrees of freedom at the later lags. This adjustment is built into the sample autocorrelation function.
- The autocorrelations are derived algebraically. If we know the exact parameters of the ARIMA process, we know the exact autocorrelations.

The time series is stochastic. The model may be correct, but random fluctuations cause unusual sample autocorrelations. Know how to form confidence intervals.

Statement B: Residuals, time series, and fitted processes.

Residuals are discussed so often it seems that time series have inherent residuals.

- The residuals of the time series are not known until we specify a model. A time series with no model has no residuals.
- The ARIMA process by itself has an error term, not residuals. The realization of the ARIMA process has residuals.

The assumptions in *Statement B* are a simple method of computing the residuals. In theory,

we can estimate slightly better residuals, but the extra effort is not worth the slight gain in accuracy. The simple assumptions cause the residuals for the first few terms to be *slightly* over-stated, but the over-statement is not material.

Statement C: White noise process

The residuals are slightly over-stated and autocorrelated for the first few terms, but this discrepancy is not material. The residuals *resemble* a white noise process; they are not exactly a white noise process. The exam problems do not harp on this distinction.

Take heed: We test the residuals to validate the fitted model. If we fit an AR(1) process, the residuals resemble a white noise process, not a random walk or an AR(1) process.

- Checking the residuals is an in-sample test.
- Out-of-sample tests are also important.

We use both in-sample and out-of-sample tests.

- In-sample tests compare the past estimates with the observed values.
- Out-of-sample tests compare the forecasts with future values.

Your student project should leave out several values for out-of-sample tests.

Illustration: For a time series of monthly interest rates or sales or gas consumption, we may use years 20X0 through 20X8 to fit the model and year 20X9 to check the model.

For final exam problems, distinguish between in-sample and out-of-sample tests. Know the tests used for each, and how we compare different models.

Statement D: The variance is $1/T$; the standard deviation is the square root of $1/T$.

Take heed: The exam problems ask about

- The distribution, which is normal, not χ-squared, lognormal, or other.
- The variance or standard deviation: we use the number of observations, not the degrees of freedom. We don't use $T - p - q$.

Keep several principles in mind:

- As T increases, the sum of squared errors of the time series increases. It is proportional to $T - p - q$. In most scenarios, p and q are small and T is large, so the sum of squared errors increases roughly in proportion to T.
- As T increases, the expected variance of the time series doesn't change. It may increase or decrease, but it is *unbiased*, so *we don't expect it* to increase or decrease.
- As T increases, the variance of the sample autocorrelations decreases in proportion to $1/T$ if the residuals are a white noise process.

Statement E: The term *approximately* is used because the residuals are not exactly a white noise process.

Take heed: Know the formula and use of the Box-Pierce Q statistic and the Ljung-Box Q_* statistic. We don't use all the residuals. If we have 200 observations in the time series, we might use sample autocorrelations from lag 5 to 35. The first few sample autocorrelations have slight serial correlation even for a white noise process and correlations of higher lags and less stable.