

TIME SERIES STUDENT PROJECTS: HYPOTHESIS TESTING AND DISTRIBUTIONS

(The attached PDF file has better formatting.)

Updated: May 2, 2008

Jacob: What do the Durbin-Watson statistic, Bartlett's test, and the Box-Pierce Q statistic evaluate?

Rachel: These tests evaluate if the sample autocorrelations of the residuals are statistically different from zero.

Jacob: How do these tests work? I understand the formulas, but I don't get the intuition.

Rachel: To test the null hypothesis that the autocorrelations are zero, we need several assumptions about the distribution of sample autocorrelations. We explain the terms.

- The autocorrelations of the residuals are not observed. If the ARIMA process is a good model for the time series, the autocorrelations of the residuals have a mean of zero.
- The sample autocorrelation of the residuals are observed values that differ from zero. We estimate the standard error of the sample autocorrelations.

Jacob: Regression analysis uses t statistics and p -values to test hypotheses that a value is zero. Do we make assumptions about distributions?

Rachel: Hypothesis testing in regression analysis generally assumes that the residuals are normally distributed with a constant variance.

Jacob: How do the assumptions affect the hypothesis test?

Rachel: Suppose the *sample autocorrelation* of lag 1 is 8% and we want to test the hypothesis that the autocorrelation is actually zero. The null hypothesis assumes that the sample autocorrelation has a normal distribution with a mean of zero, but we don't know the variance or the standard deviation (the square root of the variance) of the distribution.

- If the standard deviation is 8%, the probability of the sample autocorrelation being greater in absolute value than 8% is about one third.
- If the standard deviation is 4%, the probability of the sample autocorrelation being greater in absolute value than 8% is about 5%.

Jacob: How do we estimate the variance of this distribution?

Rachel: Bartlett's theorem says that the sample autocorrelations of a white noise process have a normal distribution with a variance of $1/T$, where T is the number of observations. If the residuals are a white noise process, their sample autocorrelations are normally distributed with a standard deviation of $1/\sqrt{T}$.

Jacob: How does Bartlett's test work?

Rachel: If the sample autocorrelations have a normal distribution with a mean of zero and a standard deviation of σ , then the probability of an observed sample autocorrelation being greater in absolute value

- ~ than 1.96σ is 5%.
- ~ than 1.45σ is 10%.

Jacob: Suppose we find that a sample autocorrelation exceeds 1.96σ . Do we infer that the sample autocorrelations are not a white noise process?

Rachel: That depends on the lag, the type of data, and the other sample autocorrelations. Suppose $T = 400$, $1/\sqrt{T} = 5\%$, $\sigma = 5\%$, $1.96\sigma \approx 10\%$, the data are monthly interest rates, and the sample autocorrelation of lag k is 12.5%.

If $k = 1$, we avoid drawing any inference. If the interest rates are correlated with the time period, the residuals may appear to be serially correlated, even if they are not. It is common for sample autocorrelations of lag 1 to be more than zero even if the autocorrelation is zero. The Durbin-Watson statistic may be inconclusive. This is like Bayesian estimation. The *prior* distribution for the autocorrelation of lag 1 has a high probability of being greater than zero because of other reasons. An observed value of 2.5 standard deviations is inconclusive.

If we observe 10 lags ($k = 10$) and the sample autocorrelations of 9 of these lags are less than 10%, we presume the high sample autocorrelation the tenth lag is the expected random fluctuation in a normal distribution. The null hypothesis of a zero mean assumes that 5% of the sample autocorrelations have absolute values greater than 10%. One high sample autocorrelation out of ten is not unexpected.

If $k = 12$ and the sample autocorrelations of lags 1 through 11 are less than 10%, we suspect the high sample autocorrelation for lag 12 reflects annual seasonality.

Jacob: Bartlett's test sounds subjective; is this a problem?

Rachel: A skilled statistician prefers a subjective test that relies on our intuition about the time series. Seasonal correlations are more common than non-seasonal correlations. We may have subjective views on the types of autocorrelations that are most likely, such as positive vs negative autocorrelations.

Jacob: If we assume a normal distribution with a standard deviation of σ , the probability of a positive sample autocorrelation is the same as the probability of a negative sample autocorrelation. Why should the sign of the sample autocorrelation affect our inference?

Rachel: If we model monthly sales data and find a sample autocorrelation of 15% for a lag of 12 months, we presume this is annual seasonality. If the sample autocorrelation for a lag of 12 months is -15% , we may attribute this to random fluctuation.

- ~ If the actual autocorrelation is zero, the probability of positive vs negative sample autocorrelations is the same.
- ~ In practice, the incidence of positive autocorrelation is not the same as the incidence of negative autocorrelation. An AR(1) model with a negative parameter for one lag is often a spurious result.

Jacob: Do we examine the Durbin-Watson statistic, Bartlett's test, and the Box-Pierce Q statistic on the residuals or the sample autocorrelations of the residuals?

Rachel: The Durbin-Watson statistic is applied to the residuals; the Durbin-Watson statistic calculates the autocorrelation of lag 1. The Durbin-Watson statistic $\approx 2 - 2 \times$ the sample autocorrelation of lag 1.

- ~ For perfect positive autocorrelation, the Durbin-Watson statistic = 0.
- ~ For perfect negative autocorrelation, the Durbin-Watson statistic = 4.

Bartlett's test and the Box-Pierce Q statistic are applied to the sample autocorrelations of the residuals.

Jacob: How does the Durbin-Watson statistic differ from Bartlett's test and the Box-Pierce Q statistic?

Rachel: Bartlett's test and the Box-Pierce Q statistic evaluate whether the residuals have a normal distribution with a variance of $1/T$, where T is the number of observations. They examine sample autocorrelations of various lags. The Durbin-Watson statistic examines if the sample autocorrelations of lag 1 are statistically different from zero.

Jacob: Are these tests equally strict?

Rachel: The Durbin-Watson statistic considers other factors that affect the observed sample autocorrelation of lag 1. The Durbin-Watson statistic may be distorted in a lagged regression, so we don't use this statistic for hypothesis testing of ARIMA models.

Take heed: You may use the Durbin-Watson statistic in your student project, but be aware that it might a serial correlation even if none exists.

Bartlett's test and the Box-Pierce Q statistic are applied to many sample autocorrelations, not just the first one. Both tests acknowledge that sample autocorrelations for the first several lags may not follow the normal distribution or the χ -squared distribution.