

TS: Note on Positive and Negative Autocorrelations

Many student projects have correlograms of the first differences showing

- ~ Slowly declining positive sample autocorrelations for the first N lags from Z to 0.
- ~ Slowly declining negative sample autocorrelations for the next N' lags from 0 to Z' .
- ~ Slowly rising negative sample autocorrelations for the next N'' lags from Z' to 0.

Often, $N' \approx \frac{1}{2} N$ and $Z' \approx -\frac{1}{2} Z$. The magnitude of N'' varies.

The correlogram has a clear pattern, but this pattern is not discussed in the textbook. Many candidates infer that the time series is not stationary, and they take second differences. They fit ARIMA processes to the second differences. Their models do not fit well and they forecast poorly.

This pattern of sample autocorrelations indicates two time series with different means. We should divide the full time series into two time periods. Taking second differences hides the problem and creates a poor model.

This error occurs most frequently in student projects on GNP (Gross National Product) and CPI (inflation). This posting explains this pattern and why it occurs.

Suppose GNP (or GDP) grows 3% a year for 15 years and then 1% a year for the next 15 years. We examine a time series by calendar quarter of annualized GDP.

This example reflects actual U.S. experience. GDP grew at 3.5% a year in the 1950's and 1960's, drifted down to about 1.5% a year by the second half of the 1970's, and then rose back to 2.5% or 3% by the late 1980's, where it remained for much of the 1990's. Actual GDP varies by quarter, begin low in recessions and high in prosperous years.

Economists assume that long-term GDP growth reflects numerous conditions, including marginal tax rates, international trade and currency restrictions, government policies, level of development, education and labor force participation of women, and a host of other items. The time series course emphasizes that we should first use a structural model, such as regression analysis, to explain GDP growth from these exogenous conditions. We then use an ARIMA process to explain the residuals.

The student project does not expect a sophisticated structural model. We encourage you to form simple regression models, but these are not required for the time series course. For GDP growth, economists do not agree on the proper explanatory variables, so a structural model is difficult to form.

For this illustration, we do not use a structural model and we use the simplified pattern of 3% and 1% per annum with low stochasticity. Stochasticity and business cycles affect the correlograms, but the pattern described here shows up in the actual results for U.S. GDP.

We examine GDP by quarter for 30 years, or quarters 1 through 120. To simplify the mathematics, we add one more value at the beginning (at quarter 0), so we have 60 values of 3% growth and 60 values of 1% growth.

We model GDP by an ARIMA process. GDP is not stationary, since it grows exponentially by improvements in worker productivity. We use real GDP, not nominal GDP, so inflation does not affect the ARIMA modeling.

Productivity growth is multiplicative (exponential), not additive. We take logarithms of real GDP to convert the exponential trend to a linear trend. We take first differences to eliminate the trend. We get a time series of 120 values: 60 of +3% and 60 of +1%.

We first convert this time series into its deviations from the mean of 2%. We get 60 values of +1% and 60 values of -1%. To simplify the mathematics, we use units of 1 percentage point, so we have 60 values of +1 and 60 values of -1.

We form the correlogram for lags of 1 through 119. The denominator of the sample autocorrelation function is $120 \times 1^2 = 120$.

The numerator of the sample autocorrelation function depends on the lag.

~ *Lag 1*: 59 values of $+1 \times +1 = +1$; 1 value of $+1 \times -1 = -1$; 59 values of $-1 \times -1 = +1$; the sum of $59 - 1 + 59 = 117$.

~ *Lag 2*: 58 values of $+1 \times +1 = +1$; 2 values of $+1 \times -1 = -1$; 58 values of $-1 \times -1 = +1$; the sum of $58 - 2 + 59 = 114$.

The sample autocorrelation decreases by 3 for each additional lag. Using the N and Z in the first lines of this posting: $N = 39$ (≈ 40) and $Z = 117/120 \approx 1$. In practice, variations in the GDP growth rates and stochasticity of GDP growth each quarter reduce Z and N. Even in clear instances of different trends for two perspective, Z is usually about 40%.

~ *Lag 40*: 20 values of $+1 \times +1 = +1$; 40 values of $+1 \times -1 = -1$; 20 values of $-1 \times -1 = +1$; the sum of $20 - 40 + 20 = 0$.

~ *Lag 41*: 19 values of $+1 \times +1 = +1$; 41 values of $+1 \times -1 = -1$; 19 values of $-1 \times -1 = +1$; the sum of $19 - 41 + 19 = -3$.

~ *Lag 59*: 1 values of $+1 \times +1 = +1$; 59 values of $+1 \times -1 = -1$; 1 value of $-1 \times -1 = +1$; the sum of $1 - 59 + 1 = -57$.

~ *Lag 60*: 0 values of $+1 \times +1 = +1$; 60 values of $+1 \times -1 = -1$; 0 values of $-1 \times -1 = +1$; the sum of $0 - 60 + 0 = -60$.

~ *Lag 119*: 0 values of $+1 \times +1 = +1$; 1 value of $+1 \times -1 = -1$; 0 values of $-1 \times -1 = +1$; the sum of $0 - 1 + 0 = -1$.

Taking second differences appears to solve the problem. We get a time series of 59 values of zero, one value of -2 , and 59 values of zero. With a stochastic time series and a slowly changing mean, the second differences may even form a stationary time series which passes Bartlett's test and the Box-Pierce Q statistic for a white noise process.

But this time series gives incorrect forecasts, since it assumes the mean first difference is $+2\%$. The true GDP growth rate in the second period is $+1\%$. We forecast 1% for the immediately following quarters, not 2% .

The proper analysis is that the time series process changed after the first 15 years, from a 3% trend to a 1% trend. The forecasts should use a 1% trend, not a 2% trend.

The textbook is not always clear about these issues, since it is hard for the statistician to judge if the time series has changed. The textbook assumes that you have examined the time series and the explanatory variables that affect the process. You have chosen a time period for which the time series is homogeneous (whether it is stationary or non-stationary). If the time series has changed because of some exogenous explanatory variable, you are using the residuals from a structural model, and these residuals form a homogeneous time series.

The time series section of this textbook does not much discuss exogenous variables that make the time series heterogeneous. Other chapters of the textbook cover this topic, such as the chapter on dummy variables that is covered in the regression analysis course. The authors assume you have applied the needed statistical techniques to make the time series homogeneous.

The student projects do not require you to use residuals from a structural model or to make the time series homogeneous by other means. We simply select a time period in which the time series is homogeneous.