My Store Glossary Home About Me Contact Me

Statistics By Jim

Making statistics intuitive

Graphs Basics Hypothesis Testing Regression ANOVA
Probability Time Series Fun

Autocorrelation and Partial Autocorrelation in Time Series Data

By Jim Frost — 12 Comments

Autocorrelation is the correlation between two observations at different points in a time series. For example, values that are separated by an interval might have a strong positive or negative correlation. When these correlations are present, they indicate that past values influence the current value. Analysts use the autocorrelation and partial autocorrelation functions to understand the properties of time series data, fit the appropriate models, and make forecasts.

In this post, I cover both the autocorrelation function and partial autocorrelation function. You'l learn about the differences between these functions and what they can tell you about your data In later posts, I'll show you how to incorporate this information in regression models of time series data and other time-series analyses.

Autocorrelation and Partial Autocorrelation Basics

Autocorrelation is the correlation between two values in a time series. In other words, the time series data correlate with themselves—hence, the name. We talk about these correlations using the term "lags." Analysts record time-series data by measuring a characteristic at evenly spaced intervals—such as daily, monthly, or yearly. The number of intervals between the two observations is the lag. For example, the lag between the current and previous observation is one. If you go back one more interval, the lag is two, and so on.

In mathematical terms, the observations at y_t and y_{t-k} are separated by k time units. K is the lag This lag can be days, quarters, or years depending on the nature of the data. When k=1, you're assessing adjacent observations. For each lag, there is a correlation.

The autocorrelation function (ACF) assesses the correlation between observations in a time seri for a set of lags. The ACF for time series y is given by: Corr (y_t,y_{t-k}), k=1,2,….

Analysts typically use graphs to display this function.

Related posts: Time Series Analysis Introduction and Interpreting Correlations

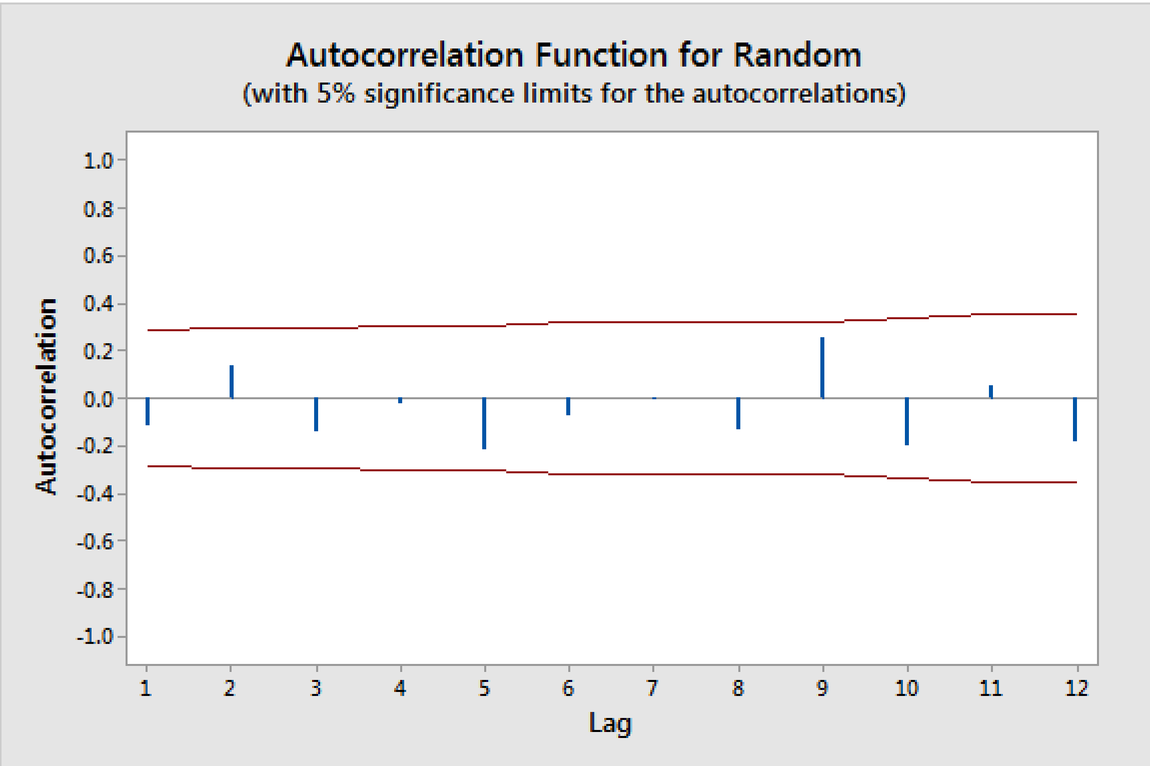
Autocorrelation Function (ACF)

Use the autocorrelation function (ACF) to identify which lags have significant correlations, understand the patterns and properties of the time series, and then use that information to model the time series data. From the ACF, you can assess the randomness and stationarity of a time series. You can also determine whether trends and seasonal patterns are present.

In an ACF plot, each bar represents the size and direction of the correlation. Bars that extend across the red line are statistically significant.

Randomness/White Noise

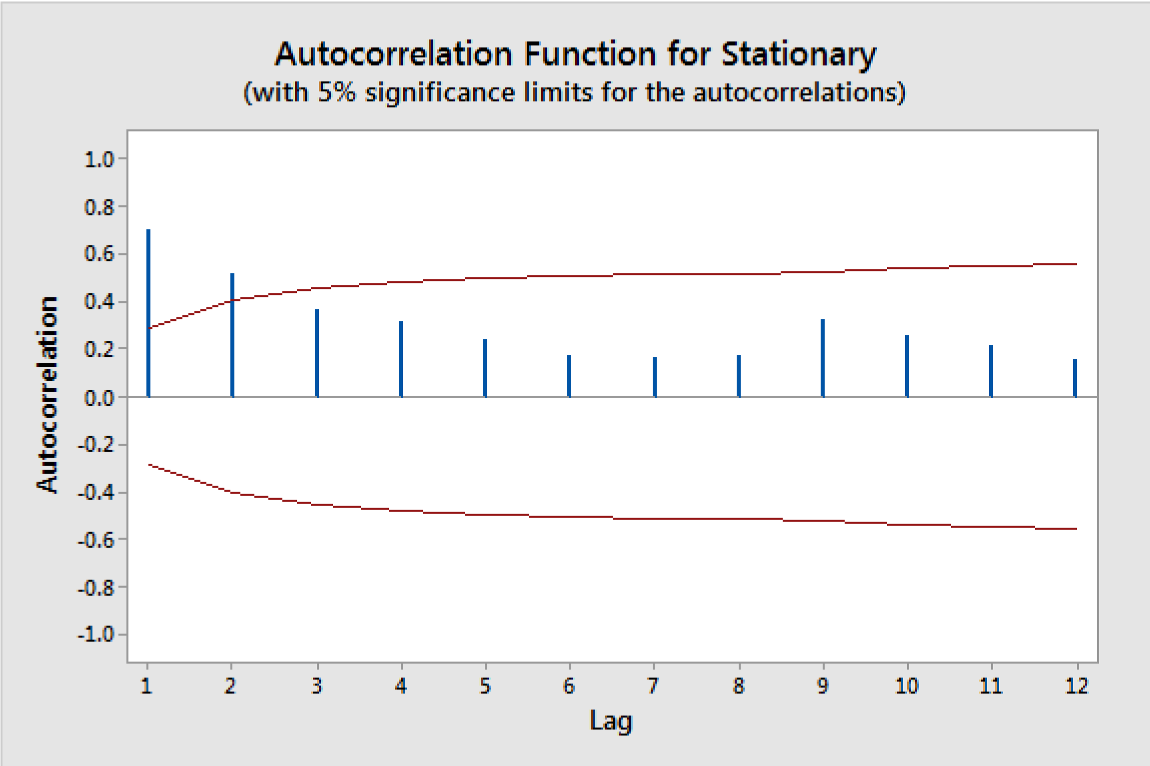
For random data, autocorrelations should be near zero for all lags. Analysts also refer to this condition as white noise. Non-random data have at least one significant lag. When the data are not random, it's a good indication that you need to use a time series analysis or incorporate lag into a regression analysis to model the data appropriately.



This ACF plot indicates that these time series data are random.

Stationarity

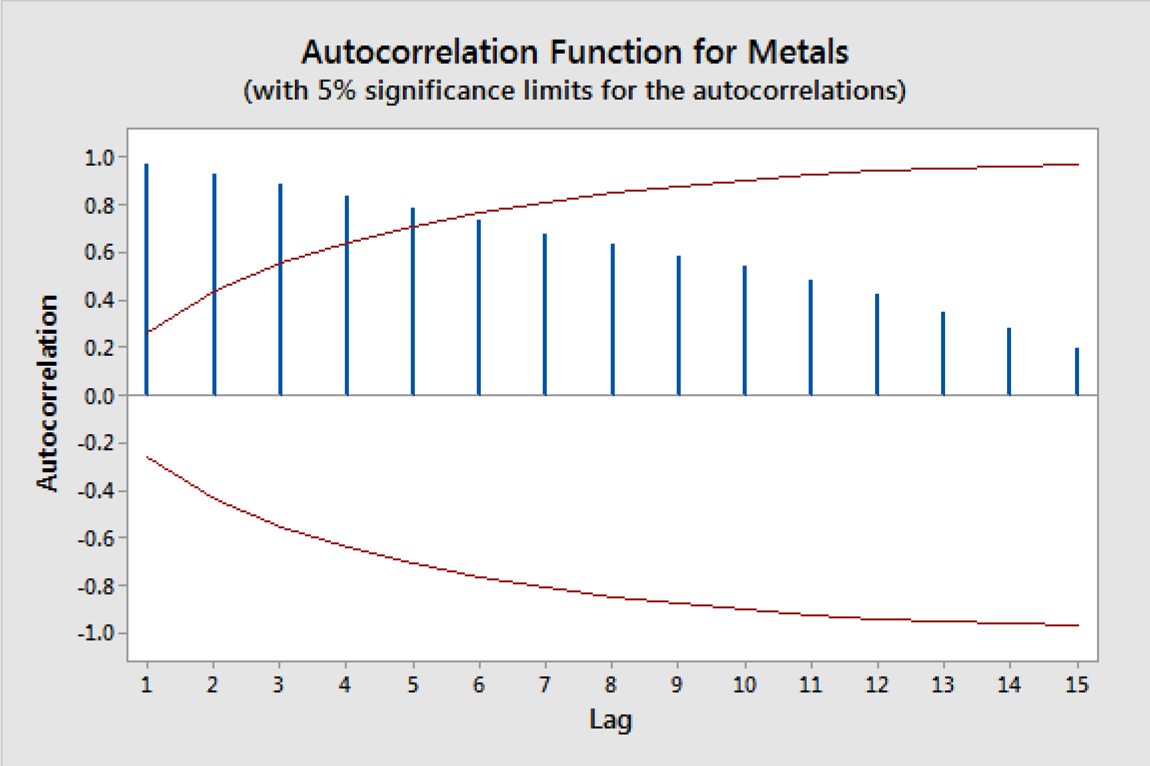
Stationarity means that the time series does not have a trend, has a constant variance, a consta autocorrelation pattern, and no seasonal pattern. The autocorrelation function declines to near zero rapidly for a stationary time series. In contrast, the ACF drops slowly for a non-stationary time series.



In this chart for a stationary time series, notice how the autocorrelations decline to non-significant levels quickly.

Trends

When trends are present in a time series, shorter lags typically have large positive correlations because observations closer in time tend to have similar values. The correlations taper off slowl as the lags increase.

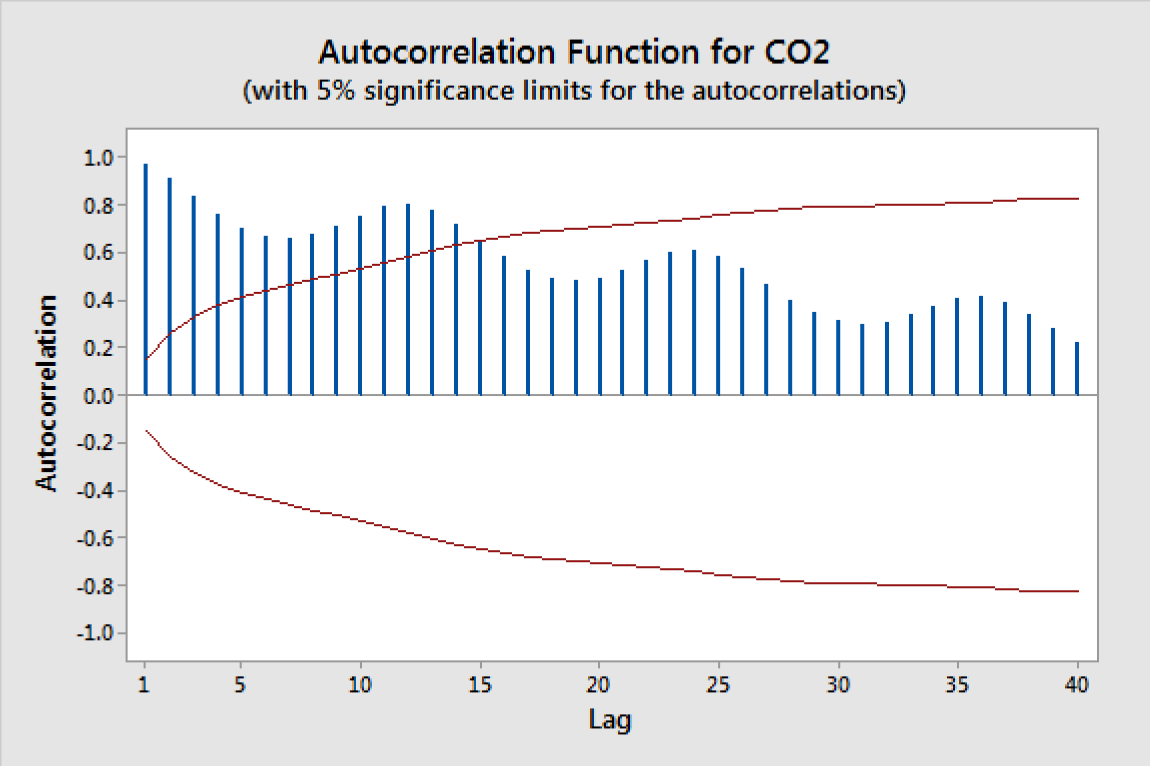


In this ACF plot for metal sales, the autocorrelations decline slowly. The first five lags are significant.

Seasonality

When seasonal patterns are present, the autocorrelations are larger for lags at multiples of the seasonal frequency than for other lags.

When a time series has both a trend and seasonality, the ACF plot displays a mixture of both effects. That’s the case in the autocorrelation function plot for the carbon dioxide (CO2) dataset from NIST. This dataset contains monthly mean CO2 measurements at the Mauna Loa Observatory. Download the CO2_Data.



Notice how you can see the wavy correlations for the seasonal pattern and the slowly diminishing lags of a trend.

Partial Autocorrelation Function (PACF)

The partial autocorrelation function is similar to the ACF except that it displays only the correlation between two observations that the shorter lags between those observations do not explain. For example, the partial autocorrelation for lag 3 is only the correlation that lags 1 and do not explain. In other words, the partial correlation for each lag is the unique correlation between those two observations after partialling out the intervening correlations.

As you saw, the autocorrelation function helps assess the properties of a time series. In contras the partial autocorrelation function (PACF) is more useful during the specification process for ar autoregressive model. Analysts use partial autocorrelation plots to specify regression models with time series data and Auto Regressive Integrated Moving Average (ARIMA) models. I'll focus on that aspect in posts about those methods.

Related post: Using Moving Averages to Smooth Time Series Data

For this post, I'll show you a quick example of a PACF plot. Typically, you will use the ACF to determine whether an autoregressive model is appropriate. If it is, you then use the PACF to he you choose the model terms.

This partial autocorrelation plot displays data from the southern oscillations dataset from NIST. The southern oscillations refer to changes in the barometric pressure near Tahiti that predicts E Niño. Download the southern_oscillations_data.

On the graph, the partial autocorrelations for lags 1 and 2 are statistically significant. The subsequent lags are nearly significant. Consequently, this PACF suggests fitting either a second third-order autoregressive model.

By assessing the autocorrelation and partial autocorrelation patterns in your data, you can understand the nature of your time series and model it!

Filed Under: Time Series

Tagged With: analysis example, conceptual, grap

Comments

Yana says
November 10, 2022 at 10:24 am

Thank you for the informative article!
Isn't stationarity a presumption of using ACF and PACF? If the data used for ACF (or PACF) has seasonality and/or trend, won't the results be invalid in that case? And if so what is the right way to check data for the seasonality/trend and measure their strength?

Reply

Nasib ullah says
November 3, 2022 at 2:55 pm

Very informative

Reply

Jim Frost says
November 3, 2022 at 3:07 pm

Thanks! I'm glad it was helpful!

Reply

Manu prakash Choudhary says
July 12, 2022 at 4:52 am

Really Great article helped me a lot thank you

Reply

Ivan says

June 3, 2022 at 2:01 am

Thank you for your reply Jim. So you mean that we need to look at the chart changing and not at the absolute chart value.

Reply

Jim Frost says

June 3, 2022 at 5:37 pm

You want it to go towards the central line of zero.

Reply

Ivan Bukharev says

May 31, 2022 at 10:09 am

Dear Jim.

Maybe i`am dumb but for example

in "stationarity" chart red line slightly increasing both on negative and positive sides when the lag is increasing.
at the meantime you write "the autocorrelations decline to non-significant levels quickly".
So it increases or declines?

Reply

Jim Frost says

June 2, 2022 at 11:00 pm

Hi Ivan,

A large/strong correlation can be either positive or negative. Hence, when a correlation declines to non-significant levels, it means going to zero.

Reply

Ghulam mustafa says
May 17, 2021 at 11:23 pm

Dear good post
How can we choose best model using ACF and PACF????

Reply

Jim Frost says
May 17, 2021 at 11:31 pm

Hi Ghulam,

Thank you! I will dedicate the topic of a future post to showing how to use ACF and PACF to model time series data. This one serves as an introduction to the concepts.

Reply

Berns Buenaobra says
May 17, 2021 at 2:26 am

Quite clear and direct thanks! Maybe include also Cross Correlation too? I am interested for example if one ACF of one variable monitored in time series can sort modulate another variable ACF.

Reply

Jim Frost says

May 17, 2021 at 10:57 pm

Hi Berns,

That sounds like a great topic for a future blog post!

Reply

Comments and Questions

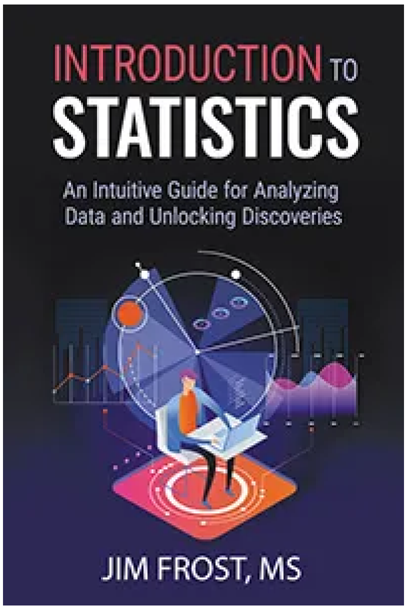
Meet Jim

I'll help you intuitively understand statistics by focusing on concepts and using plain English so you can concentrate on understanding your results.

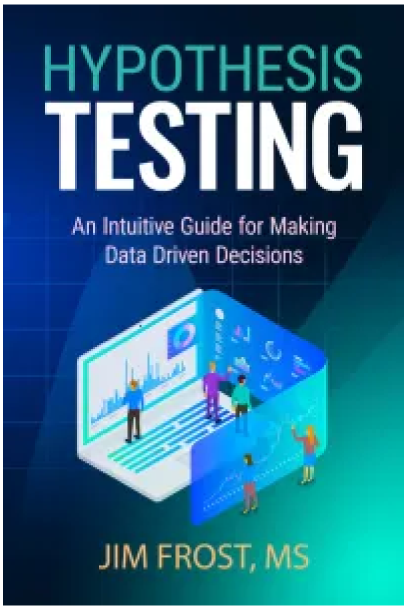
Read More...

Search this website

Buy My Introduction to Statistics Book!



Buy My Hypothesis Testing Book!



Top Posts

How to Interpret P-values and Coefficients in Regression Analysis

How To Interpret R-squared in Regression Analysis

Mean, Median, and Mode: Measures of Central Tendency

Cronbach's Alpha: Definition, Calculations & Example

Weighted Average: Formula & Calculation Examples

Z-table

Interquartile Range (IQR): How to Find and Use It

Multicollinearity in Regression Analysis: Problems, Detection, and Solutions

How to do t-Tests in Excel

F-table

Recent Posts

T Test Overview: How to Use & Examples

Wilcoxon Signed Rank Test Explained

What is P Hacking: Methods & Best Practices

Likert Scale: Survey Use & Examples

Correlation Coefficient Formula Walkthrough

Two-Way Table Explained

Recent Comments

Jim Frost on Interpreting Correlation Coefficients

Jim Frost on Covariates: Definition & Uses

Jim Frost on Choosing the Correct Type of Regression Analysis

Jim Frost on Revisiting the Monty Hall Problem with Hypothesis Testing

CH on Covariates: Definition & Uses