# Evidence of Rare Events: False Positives in a Sea of Noise

*This article is authored by MOI Global instructor Sean Stannard-Stockton, President and Chief Investment Officer of Ensemble Capital, based in Burlingame, CA. For additional articles by Sean and his colleagues, visit [Intrinsic Investing](#).*

Investors are suddenly grappling with the possibility that the economic expansion is nearing its end and a recession could be just around the corner. We [discussed the context](#) for these worries in mid-October, but today let's discuss why even strong evidence of a relatively rare event such as a recession does not actually suggest a recession is likely. This does not mean a recession will not happen, only that investors need to be cautious about predicting relatively rare events *even when the evidence is strong that they might occur.*

In Nate Silver's wonderful book [*The Signal & The Noise: Why So Many Predictions Fail — But Some Don't*](#), he explains the difficulties of forecasting and introduces the reader to the [Bayes Theorem](#). Bayes Theorem offers a formula for calculating the probability of an event, based on knowledge of conditions that might be related to the event.

For instance, declining homes sales is an event that has already happened with the number of homes sold so far in 2018 coming in below the same time period in 2017. This may be evidence that a recession is coming. Existing home sales declined in the late 70's , the late 80's and the mid 2000's all in the lead up to a recession. However, in both 1995 and 2014, home sales also declined, and not only was there not a recession, but the years following 1995 saw robust economic activity. Further, home sales did not decline in 2000, yet there was a recession in 2001.

So what does this tell us about the probability that a recession is coming soon now that we know that home sales are declining? There have been four recessions in the last 40 years and three of them have been preceded by a decline in home sales. There have been five instances of home sales declining and three of them were followed by a recession. So it does seem that declines in home sales offer meaningful evidence that a recession may occur. But how much weight should we put on this evidence? That's where Bayes Theorem comes in.

You might look at the data on home sales and decide that since 75% of recessions have been preceded by a decline in home sales, this means that there is a 75% chance of a recession occurring soon. But while this line of thought is intuitively appealing, it is completely wrong. It ignores the fact that recessions are relatively rare as well as the fact that sometimes home sales decline and no recession occurs (a false positive).

There have been four recessions in 40 years (10% odds of happening). Since recessions can last for more than one year, you might also say that the US economy has been in recession for about eight of the last 40 years (20% of the time). So let's split the difference and assume that in any given year there is a 15% chance of a recession. (Note: these are estimated odds for illustrative purposes only). And don't forget that there have been five instances of home prices declining and yet only three were followed by a recession.

So this is where the math gets interesting.

- Recessions occur 15% of the time.

- When a recession occurs, 75% of the time it is precede by a decline in home sales.
- 40% of the time that home sales decline, a recession does *not* follow.

Now we can calculate the probability of recession given we know home sales have declined. It must be higher than 15% since that's the odds of a recession happening any time. But how high? Remember 75% of recessions are preceded by a decline in home sales, but 40% of the time that sales decline no recession occurs.

Using the assumed probabilities above (remember, they're being used for illustrative purposes, we're not arguing that this simple math can predict recessions) Bayes Theorem, calculates a 15% probability of recession in any given year and a 25% probability of recession in the event that you observe homes sales have declined.

You can do the math yourself using the calculator below. If you don't like the assumptions I used (15%, 75%, 40%) plug in your own. You can learn more about how to use the calculator [here](#).

Only 25%? That's higher than the 15% odds if you have no evidence either way. But why don't the odds increase more? It has to do with the relatively low probability of recession at any given time as well as the meaningful frequency with which home sales decline and yet no recession occurs.

Here's how Silver puts it:

"When we fail to think like Bayesians, false positives are a problem… As there is an exponential increase in the amount of available information, there is likewise an exponential increase in the number of hypotheses to investigate. For instance, the U.S. government now publishes data on about 45,000 economic statistics. If you want to test for relationships between all combinations of two pairs of these statistics—is there a causal relationship between the bank prime loan rate and the unemployment rate in Alabama?—that gives you literally one billion hypotheses to test. But the number of meaningful relationships in the data—those that speak to causality rather than correlation and testify to how the world really works—is orders of magnitude smaller. Nor is it likely to be increasing at nearly so fast a rate as the information itself; there isn't any more truth in the world than there was before the Internet or the printing press. Most of the data is just noise, as most of the universe is filled with empty space. Meanwhile, as we know from Bayes's theorem, *when the underlying incidence of something in a population is low (truth in the sea of data), false positives can dominate the results* if we are not careful."

Bayesian probability analysis is not intuitive. And you weren't likely to have been taught it in school where [classical](#) or "frequentist" statistics are taught (remember all those lessons about rolling dice and flipping coins?). But investing is dominated by Bayesian probability problems.

- What are the odds that Netflix subscriber growth slows given that Apple has announced a new streaming service?
- What are the odds that homes prices decline given interest rates have recently increased sharply?
- What are the odds that electric vehicles will see mainstream adoption given that Tesla has shown the ability to take very high levels of market share within the US luxury car market?

These are tough questions. But the key insight from Bayesian logic is as Silver puts it "When the underlying incidence of something in a population is low (truth in the sea of data), false positives can dominate the results." It's human nature to weigh recent experiences with a higher probability than its actual chance of occurring in reality and it's also human nature to connect events that memories recall being related. This can result in intuitive probability assessments by investors that dramatically depart

from what a more robust mathematical analysis would forecast. For this reason a tool like Bayes theorem can help us better calibrate our intuition based forecasts so that we have a better ability to judge the true odds of important events and the decisions we'd make to deal with them.