

[sign up](#) [log in](#) [tour](#) [help](#)

Cross Validated is a question and answer site for people interested in statistics, machine learning, data analysis, data mining, and data visualization. Join them; it only takes a minute:

[Sign up](#)


### Here's how it works:

Anybody can ask a question


Anybody can answer

The best answers are voted up and rise to the top

## Forecasting price based on historic time series



Want to work remotely?  
Find remote jobs based on technology



Get started

I work for a material stock holder. I am trying to quantify how the probability of not getting a specified minimum of gross profit, changes over time. This problem basically boils down to assigning probabilities to the set of future market prices.


I was hoping that I could use time series analysis on the company's purchase and sales records (they span the last ten years, so I should hopefully have a decent dataset), but I am unsure about of how to go solve this. I have done some googling and from what I can see the an ARIMA modeling is what I should be looking into, but again I am unsure. If anyone could give me some pointers that would be great.

The price of the material we stock fluctuate over time, (of course, as a result of the changes in the factors that influence the market price.) Plotting the prices into a histogram, it is clear that it is not a normal distribution. (this is the case for both deflated and nominal prices) Which, from my standpoint is expected, prices are influenced by business cycles and the evolution of the production costs.

If anyone could give me a pointer in the right direction of where to start looking / reading, I would be very grateful.

[time-series](#) [arima](#)

edited Jun 17 '13 at 16:12

 **Gala**  
7,113 2 21 36

asked Jun 17 '13 at 9:04

 **Rookie**  
113 3

1 "Plotting the prices into a histogram, it is clear that it is not a normal distribution, so it is not a stationary stochastic variable I am dealing with." -- how does this follow (or how would the converse follow, in the other situation?) -- [Glen\\_b](#) ♦ Jun 17 '13 at 9:24

@Glen\_b If this statement is false, then it is my understanding of a stationary stochastic variable that is lacking, which is highly likely. I will remove this comment. -- [Rookie](#) Jun 17 '13 at 9:31

Indeed, I don't think stationarity implies normality in general, nor vice-versa -- [Glen\\_b](#) ♦ Jun 17 '13 at 9:33

@Glen\_b How about the other parts of the question, difficult to answer based on the info given? -- [Rookie](#) Jun 17 '13 at 10:04

You need minimally a more descriptive title. Best to choose for yourself before one is thrust upon you. -- [Nick Cox](#) Jun 17 '13 at 10:10

## 1 Answer

You're on the right track by acknowledging that ARIMA modeling is what you should be looking into.

I've seen ARIMA modeling applied to cases involving: inventory stock, business sales, levels of production of particular goods, and various other business related time-series. Without access to the data, I can only speculate that the data you're working with falls into this same sort of category.

Of course, ARIMA modeling is *univariate*, so any forecasts that you produce will be forecasts for the time-series under investigation. For example, if you model prices then you will derive forecasts for prices - not gross profits. Indeed, you can use price forecasts to build forecasts for profits, so choose carefully the data you want to work with if you have choice among many time-series.

It is common to see ARIMA models used as a benchmark, so even if you believe that more complex models (multiple-series & econometric models) may give you superior forecasts, ARIMA modeling is, nevertheless, still a worthwhile pursuit; if you build a number of models you have something to compare them against and this also helps decide whether or not the extra complexity is necessary.

The reason why ARIMA models are good for benchmarking is because: ARIMA forecasts are *optimal* (smaller mean-squared forecast error) univariate forecasts (if correctly built). The forecasts are optimal among forecasts from *univariate*, *linear*, *fixed-coefficient* models.

Analysis of your data may lead you to develop other models such as *multivariate* models, *non-linear* models or even *time-varying parameter* models, but starting with the simpler class of ARIMA models is a wise choice in itself because ARIMA analysis can later on complement econometric analysis. For a short discussion on this see Zellner (1978).

Obviously, the classic text to consult for ARIMA modeling (and the closely related Transfer Function models) is Box & Jenkins (1970). A good alternative is Pankratz (1983) which is basically a shorter and simpler version of Box & Jenkins' work - all of the main points are retained in Pankratz's book too.

As already mentioned, ARIMA analysis involves looking at a single time-series of past observations. At some stage, you may want to introduce other independent variables in addition to past observations of the dependent variable. This brings you into the territory of *distributed lag models* which may or may not be *autoregressive*. Extending the framework once more and these models can be *single-equation* or *multi-equation* (vector equation) models.

One of the factors to be considered when deciding to use single or vector equations will be whether or not there is possible *lagged feedback effects* among the various variables. These issues are further addressed in Pankratz (1991) which focuses on dynamic regression models.

Lastly, an excellent online time-series forecasting textbook is Rob Hyndman's [Forecasting: principles and practice](#). Furthermore, if you are an R user (or would consider becoming one) then it would be worth your time to familiarize yourself with the [R forecast package](#) (again, thanks to Rob Hyndman).

#### References:

Box, George and Jenkins, Gwilym (1970) Time series analysis: Forecasting and control, San Francisco: Holden-Day.

Hyndman, R.J. and Athanasopoulos, G. (2013) Forecasting: principles and practice. <http://otexts.com/fpp/>. Accessed on 17 June 2013.

Pankratz, Alan (1983) Forecasting with univariate Box-Jenkins models: concepts and cases, New York: John Wiley & Sons.

Pankratz, Alan (1991) Forecasting with Dynamic Regression Models, New York: John Wiley & Sons.

Zellner, Arnold, 1978. "Folklore versus Fact in Forecasting with Econometric Methods," The Journal of Business, University of Chicago Press, vol. 51(4), pages 587-93, October.

answered Jun 17 '13 at 15:59



**Graeme Walsh**

2,843 2 15 36

Thank you so much for the great set of references, and for the overview and explanation. Your answer gives me a much needed starting point. My gratitude and best wishes to you sir! – [Rookie](#) Jun 24 '13 at 11:15

@Rookie Glad to hear you found my answer helpful. If you need any help at a later stage of your forecasting project, make sure to let us know. Would be happy to help. Cheers and best of luck! – [Graeme Walsh](#) Jun 24 '13 at 11:32

@GraemeWalsh, I am curious under what assumptions your statement holds: *ARIMA forecasts are optimal*. I suspect the conditions will be too stringent to allow applying this in practice... – [Richard Hardy](#) May 16 at 7:50

@RichardHardy Good question! Yes, I will think about it and give you an update when I can. – [Graeme Walsh](#) May 16 at 7:54

---