

[Start Here](#)[Blog](#)[Books](#)[About](#)[Contact](#)

# Time Series Forecasting as Supervised Learning

by **Jason Brownlee** on December 5, 2016 in **Time Series**



Time series forecasting can be framed as a supervised learning problem.

This re-framing of your time series data allows you access to the suite of standard linear and nonlinear machine learning algorithms on your problem.

In this post, you will discover how you can re-frame your time series problem as a supervised learning problem for machine learning. After reading this post, you will know:

- What supervised learning is and how it is the foundation for all predictive modeling machine learning algorithms.
- The sliding window method for framing a time series dataset and how to use it.
- How to use the sliding window for multivariate data and multi-step forecasting.

Let's get started.

[Get Your Start in Machine Learning](#)

Time Series Forecasting as Supervised Learning  
Photo by Jeroen Looyé, some rights reserved.

## Supervised Machine Learning

The majority of practical machine learning uses supervised learning.

Supervised learning is where you have input variables (**X**) and an output variable (**y**) and you use an algorithm to learn the mapping function from the input to the output.

1  $Y = f(X)$

The goal is to approximate the real underlying mapping so well that when you have new input data (new data), you can predict the output (new data).

Below is a contrived example of a supervised learning dataset where each row is an observation consisting of input variables (**X**) and an output variable to be predicted (**y**).

```
1 X, y
2 5, 0.9
3 4, 0.8
4 5, 1.0
5 3, 0.7
6 4, 0.9
```

It is called supervised learning because the process of an algorithm learning from the training dataset is supervised by the human learning process.

We know the correct answers; the algorithm iteratively makes predictions on the training data and is corrected by making updates. Learning stops when the algorithm achieves an acceptable level of performance.

Supervised learning problems can be further grouped into regression and classification problems.

- **Classification:** A classification problem is when the output variable is a category, such as “red” and “blue” or “disease” and “no disease.”

### Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

Get Your Start in Machine Learning

- **Regression:** A regression problem is when the output variable is a real value, such as “dollars” or *weight*. The continued example above is a regression problem.

## Stop learning Time Series Forecasting the *slow way*!

Take my free 7-day email course and discover data prep, modeling and more (with sample code).

Click to sign-up and also get a free PDF Ebook version of the course.

Start Your FREE Mini-Course Now!

## Sliding Window For Time Series Data

Time series data can be phrased as supervised learning.

Given a sequence of numbers for a time series dataset, we can restructure the data to look like a supervised learning problem by using the value at the previous time step as input variables and use the next time step as the output variable.

Let's make this concrete with an example. Imagine we have a time series as follows:

```
1 time, measure
2 1, 100
3 2, 110
4 3, 108
5 4, 115
6 5, 120
```

We can restructure this time series dataset as a supervised learning problem by using the value at the previous time step to predict the value at the next time-step. Re-organizing the time series dataset this way, the data would look as follows:

```
1 X, y
2 ?, 100
```

### Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

Get Your Start in Machine Learning

|   |      |     |
|---|------|-----|
| 3 | 100, | 110 |
| 4 | 110, | 108 |
| 5 | 108, | 115 |
| 6 | 115, | 120 |
| 7 | 120, | ?   |

Take a look at the above transformed dataset and compare it to the original time series. Here are some observations:

- We can see that the previous time step is the input (**X**) and the next time step is the output (**y**) in our supervised learning problem.
- We can see that the order between the observations is preserved, and must continue to be preserved when using this dataset to train a supervised model.
- We can see that we have no previous value that we can use to predict the first value in the sequence. We will delete this row as we cannot use it.
- We can also see that we do not have a known next value to predict for the last value in the sequence. We may want to delete this value while training our supervised model also.

The use of prior time steps to predict the next time step is called the sliding window method. For short, this is called a lag or lag method.

The number of previous time steps is called the window width or size of the lag.

This sliding window is the basis for how we can turn any time series dataset into a supervised learning problem. Here are a few things:

- We can see how this can work to turn a time series into either a regression or a classification supervised learning problem using time series values.
- We can see how once a time series dataset is prepared this way that any of the standard linear models can be applied, as long as the order of the rows is preserved.
- We can see how the width sliding window can be increased to include more previous time steps.
- We can see how the sliding window approach can be used on a time series that has more than one value, or so-called multivariate time series.

We will explore some of these uses of the sliding window, starting next with using it to handle time series with more than one observation at each time step, called multivariate time series.

## Sliding Window With Multivariate Time Series Data

### Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

The number of observations recorded for a given time in a time series dataset matters.

Traditionally, different names are used:

- **Univariate Time Series:** These are datasets where only a single variable is observed at each time, such as temperature each hour. The example in the previous section is a univariate time series dataset.
- **Multivariate Time Series:** These are datasets where two or more variables are observed at each time.

Most time series analysis methods, and even books on the topic, focus on univariate data. This is because it is the simplest to understand and work with. Multivariate data is often more difficult to work with. It is harder to model and often many of the classical methods do not perform well.

“Multivariate time series analysis considers simultaneously multiple time series. ... It is, in general, much more complicated than univariate time series analysis

— Page 1, [Multivariate Time Series Analysis: With R and Financial Applications](#).

The sweet spot for using machine learning for time series is where classical methods fall down. This is more likely with multivariate time series given the additional complexity.

Below is another worked example to make the sliding window method concrete for multivariate time series.

Assume we have the contrived multivariate time series dataset below with two observations at each time step concerned with predicting **measure2**.

```
1 time, measure1, measure2
2 1, 0.2, 88
3 2, 0.5, 89
4 3, 0.7, 87
5 4, 0.4, 88
6 5, 1.0, 90
```

We can re-frame this time series dataset as a supervised learning problem with a window width of one.

This means that we will use the previous time step values of **measure1** and **measure2**. We will also have available the next time step value for **measure1**. We will then predict the next time step value of **measure2**.

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

This will give us 3 input features and one output value to predict for each training pattern.

```

1 X1, X2, X3, y
2 ?, ?, 0.2, 88
3 0.2, 88, 0.5, 89
4 0.5, 89, 0.7, 87
5 0.7, 87, 0.4, 88
6 0.4, 88, 1.0, 90
7 1.0, 90, ?, ?

```

We can see that as in the univariate time series example above, we may need to remove the first and last rows in order to train our supervised learning model.

This example raises the question of what if we wanted to predict both **measure1** and **measure2** for the next time step?

The sliding window approach can also be used in this case.

Using the same time series dataset above, we can phrase it as a supervised learning problem where the window width of one, as follows.

```

1 X1, X2, y1, y2
2 ?, ?, 0.2, 88
3 0.2, 88, 0.5, 89
4 0.5, 89, 0.7, 87
5 0.7, 87, 0.4, 88
6 0.4, 88, 1.0, 90
7 1.0, 90, ?, ?

```

Not many supervised learning methods can handle the prediction of multiple output values without neural networks, have little trouble.

We can think of predicting more than one value as predicting a sequence. In this case, we were predicting two different output variables, but we may want to predict multiple time-steps ahead of one output variable.

This is called multi-step forecasting and is covered in the next section.

## Sliding Window With Multi-Step Forecasting

The number of time steps ahead to be forecasted is important.

### Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

Again, it is traditional to use different names for the problem depending on the number of time-steps to forecast:

- **One-Step Forecast:** This is where the next time step ( $t+1$ ) is predicted.
- **Multi-Step Forecast:** This is where two or more future time steps are to be predicted.

All of the examples we have looked at so far have been one-step forecasts.

There are a number of ways to model multi-step forecasting as a supervised learning problem. We will cover some of these alternate ways in a future post.

For now, we are focusing on framing multi-step forecast using the sliding window method.

Consider the same univariate time series dataset from the first sliding window example above:

```
1 time, measure
2 1, 100
3 2, 110
4 3, 108
5 4, 115
6 5, 120
```

We can frame this time series as a two-step forecasting dataset for supervised learning with a window

```
1 X1, y1, y2
2 ? 100, 110
3 100, 110, 108
4 110, 108, 115
5 108, 115, 120
6 115, 120, ?
7 120, ?, ?
```

We can see that the first row and the last two rows cannot be used to train a supervised model.

It is also a good example to show the burden on the input variables. Specifically, that a supervised model only has **X1** to work with in order to predict both **y1** and **y2**.

Careful thought and experimentation are needed on your problem to find a window width that results in acceptable model performance.

## Get Your Start in Machine Learning ×

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

## Further Reading

If you are looking for more resources on how to work with time series data as a machine learning problem, see the following two papers:

- [Machine Learning for Sequential Data: A Review \(2002\) \[PDF\]](#)
- [Machine Learning Strategies for Time Series Forecasting \(2013\) \(also slides PDF\)](#)

For Python code for how to do this, see the post:

- [How to Convert a Time Series to a Supervised Learning Problem in Python](#)

## Summary

In this post, you discovered how you can re-frame your time series prediction problem as a supervised learning problem using supervised learning methods.

Specifically, you learned:

- Supervised learning is the most popular way of framing problems for machine learning as a collection of supervised learning problems.
- Sliding window is the way to restructure a time series dataset as a supervised learning problem.
- Multivariate and multi-step forecasting time series can also be framed as supervised learning using supervised learning methods.

Do you have any questions about the sliding window method or about this post?

Ask your questions in the comments below and I will do my best to answer.

### Get Your Start in Machine Learning ×

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

## Want to Develop Time Series Forecasts with Python?

### Develop Your Own Forecasts in Minutes

...with just a few lines of python code

Discover how in my new Ebook:  
[Introduction to Time Series Forecasting With Python](#)

[Get Your Start in Machine Learning](#)

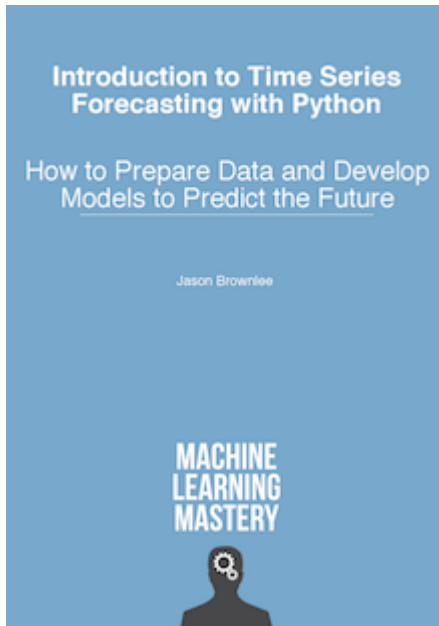


It covers **self-study tutorials** and **end-to-end projects** on topics like:  
*Loading data, visualization, modeling, algorithm tuning, and much more...*

## Finally Bring Time Series Forecasting to Your Own Projects

Skip the Academics. Just Results.

[Click to learn more.](#)



### About Jason Brownlee

Dr. Jason Brownlee is a husband, proud father, academic researcher, author, professional developer, and data scientist. He is passionate about helping developers get started and get good at applied machine learning. [Learn more.](#)

[View all posts by Jason Brownlee](#) →

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

[START MY EMAIL COURSE](#)

[◀ What Is Time Series Forecasting?](#)

[How to Go From Working in a Bank To Hired as Senior Data Scientist at Target ▶](#)

[Get Your Start in Machine Learning](#)

## 65 Responses to *Time Series Forecasting as Supervised Learning*



**Robert** December 6, 2016 at 12:32 am #

REPLY ↩

Thanks for the article. I understand the transformation. Now how do you separate the data into training and testing sets? Also, will the next article be working a simple example through to building a predictive model?



**Jason Brownlee** December 6, 2016 at 8:26 am #

Great question Robert, I will have a post on this soon.



**Leo** December 6, 2016 at 1:38 pm #

Machine learning methods are not suitable for time series analysis. They do not take into account



**Jason Brownlee** December 7, 2016 at 8:53 am #

Interesting perspective Leo.

Machine learning methods require this relationship is exposed to them explicitly in the form of a moving average, lag obs, seasonality indicators, etc. Just like linear regression does in ARIMA. No really a big leap here.

Classical methods (like MA/AR/ARMA/ARIMA/and friends) breakdown when relationships are non-linear, obs are not iid, residuals are not gaussian, etc. Sometimes the complexity of the problem requires we try alternate methods.

Finally, there are newer methods that can learn sequence, like LSTM recurrent neural networks. These methods have the potential to redefine an industry, just like has been done in speech recognition and computer vision.

### Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**Leo** December 7, 2016 at 12:12 pm #

REPLY ↩

Machine learning methods require that there is no correlation between variables. This breaks down for time series where the lagged values are correlated.

Moreover, there are many nonlinear time series methods like GARCH and its variants.



**Jason Brownlee** December 8, 2016 at 8:13 am #

REPLY ↩

Great point, thanks Leo.

The point about correlated inputs is true for many statistical methods, less true for others like trees, i (cnn and rnn).

I think you're spot on – most small univariate time series datasets will be satisfied with a classical sta lagged vars can add something, perhaps not.

When things get hairy in data with a time component (like movement prediction, gesture classification) better job of fleshing out this detail.

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**yangsp** March 17, 2017 at 8:03 pm #

I tried it half a month ago, but it didn't work well



**dirk** January 20, 2017 at 9:23 pm #

REPLY ↩

Is that not a bit Bombastic.

There are several quant hedge funds that have made and continue to make mind blowing returns th in multivariate TS data.

Get Your Start in Machine Learning

Maybe I'm missing something ?



**Jong** July 29, 2017 at 5:06 pm #

REPLY ↩

ML does NOT require that there is no correlation between variables... nor does any regression model.



**Jason Brownlee** July 30, 2017 at 7:41 am #

REPLY ↩

Regression models prefer uncorrelated input variables for model stability.

Not a requirement (we can still do it...), more of a strong preference.

See this article on Multicollinearity

<https://en.wikipedia.org/wiki/Multicollinearity>

## Get Your Start in Machine Learning



You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**Leo** December 8, 2016 at 10:38 am #

Good point Jason. I guess I need to study LSTM.



**John** January 8, 2017 at 4:47 pm #

When do you public something about the Multi-Step Forecasting? 😊



**Jason Brownlee** January 9, 2017 at 7:48 am #

REPLY ↩

They are scheduled for later this month or early next month.

Get Your Start in Machine Learning



**Dehai** January 20, 2017 at 9:19 am #

REPLY ↩

The data generated from sensors of IoT or industrial machines are also typical time series, and usually of huge amount, aka industrial big data. For this type of TS, many digital signal processing methods are used when being analysed, such as FFT, wavelet transform, euclidean distance. It seems that books discussing ML on TS usually don't cover this DSP area. What do you think?



**Jason Brownlee** January 20, 2017 at 10:25 am #

REPLY ↩

I agree Dehai.

We can view these methods as data preparation/data transforms in the project process.

Use of more advanced methods like FFT and wavelets requires knowledge of DSP which might be a bit challenging for those learning with little math background.



**Jay Urbain** January 20, 2017 at 9:22 am #

Thanks!

I had a project where I had to predict the likelihood of equipment failure from an event log. What worked was a log with temporal target features that included whether or not a piece of equipment failed in the next 30, piece of equipments past history, e.g., frequency of maintenance over different periods, variance in mean, etc. I used a machine learning technique. Test set was created from last 20% of samples.

— Jay Urbain

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**Jason Brownlee** January 20, 2017 at 10:26 am #

REPLY ↩

Very nice, thanks for sharing Jay!

Get Your Start in Machine Learning



**xjackx** February 13, 2017 at 10:25 am #

REPLY ↩

Hi Jay,

I am interested in finding out more about the predictive task you were involved with. Any chance you have a blog or can share more by email?



**Sune** October 9, 2017 at 11:45 pm #

REPLY ↩

Hi Jason and Jay

We are also trying to predict device failure based on temporal signals like temperatures, humidity, power, etc.

How does one relate 5 temporal data signals into one single fail/pass result at the end of the period?

Most examples seem to be about predicting the signal itself where as in our case we probably need to predict a binary result.

For example, if it is using a lot of power, the ambient temperature is low but the temperature is not decreasing, the compressor is failing.

Any tips would be highly appreciated.

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**Jason Brownlee** October 10, 2017 at 7:47 am #

Perhaps this example of multivariate forecasting will help as a starting point:

<https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/>



**Ziad** January 20, 2017 at 1:32 pm #

REPLY ↩

Jason, is using multi steps time lags with multivariate KNN or Random Forest equivalent to transform the feature space in similar to transform functions?

Get Your Start in Machine Learning

I will also be curious to see how SVM can be used on multivariate problems.

Thanks for the post.



**Jason Brownlee** January 21, 2017 at 10:23 am #

REPLY ↩

I don't think so Ziad, do you have a specific idea in mind?



**Kavitha Devi M K** January 20, 2017 at 4:10 pm #

REPLY ↩

In activity prediction application, the activity can be predicted only after multiple sequence of steps. How can we handle this problem for predicting the activity?



**Jason Brownlee** January 21, 2017 at 10:24 am #

Nice problem Kavitha. Sorry, I don't have any examples of activity prediction. I don't want to

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**pankaj** January 20, 2017 at 8:42 pm #

How would the time series restructuring be affected if we have 2 level or n level categorization within a time series. For example in case of sensor data we get it on each day and with-in the day say at every 5 seconds. The correlation may exist at the outer level i.e at day level but may not at internal level i.e at next sample (in seconds).

Day1 Measure

5PM 20

5PM5Sc 22

Day2 Measure

5pm 25

Get Your Start in Machine Learning

5pm5sc 27  
so on.



**Jason Brownlee** January 21, 2017 at 10:27 am #

REPLY ↩

Great question pankaj.

I would suggest resampling the data to a few different time scales and building a model on lag signals of each, then ensemble the predictions. Also build bigger models on lagged signals at each scale. You want to give your models every opportunity to exploit the temporal structure in the problem.



**Okpako A. Ejaita** January 21, 2017 at 6:10 pm #

It was a great article. My question is not really on this topic.  
how can we capture the errors in a neural network for each instance of a data and print it out in java and the errors.



**NGUYEN Quang Anh** January 22, 2017 at 9:15 pm #

This is great. Though the multi-step forecast is somewhat border me. If we make a data model to show that somehow, the next step would be build upon the value of these 3 data, like  $X(t) = a_1.X(t-1) + a_2.X(t-2) + a_3.X(t-3)$  further into the future, have we extended the width of the window? In that case, as the number of features be extended right?

## Get Your Start in Machine Learning



You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

**START MY EMAIL COURSE**



**Jason Brownlee** January 23, 2017 at 8:39 am #

REPLY ↩

That is correct.

Get Your Start in Machine Learning



There are two general approaches for a multi-step forecast: direct (one model for each future time step to be predicted) and recursive (use the one-step model again and again with predictions as inputs).



**Pranab** January 23, 2017 at 3:05 pm #

REPLY ↩

Nice article. You are proposing supervised learning for complex time series, instead of classical forecasting methods. Do you have any particular supervised learning method in mind? If so, what makes you think it will work better than NN based LSTM.

You also mentioned, in response to a comment, that some ML techniques are not adversely impacted by correlated input. Can you please shed some light on your comment.



**Jason Brownlee** January 24, 2017 at 10:58 am #

Hi Pranab,

No specific method in mind, more of a methodology of framing time series forecasting as supervised nonlinear machine learning algorithms and ensemble methods. Not a new idea for sure.

Sure, often decision trees are unflappable when it comes to irrelevant features and correlated feature interactions across features, accepting pairwise multicollinearity in input features results in better pe

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**Hassine Saidane** January 27, 2017 at 2:40 am #

Hello Jason,

This is a very interesting topic. Have you considered forecasting one-step-ahead as a function of multi steps before. This will represent an output which is a function of several variables. The question of interest, by analogy to the traditional multivariate function, is how many variables (back step) to use and which ones are most significant to use through a variable selection process. Variable selection could identify which time periods influence the analysis and forecast.

This approach can greatly benefit the forecasting and analysis of time series using all of machine learning algorithms.

A colleague and I applied this approach. Four published papers on this work can be "googled using my r

Get Your Start in Machine Learning

Happy continuation and thanks for sharing the article.



**Jason Brownlee** January 27, 2017 at 12:12 pm #

REPLY ↩

Thanks for sharing Hassine.



**sam** February 8, 2017 at 11:51 am #

REPLY ↩

Hi Jason,

I am trying to predict customer attrition based on revenue trend as time series

Month1 → \$ ; month2 → \$ as training data set.

How can i use predictive algorithm to predict customer attrition based on the above training data ?

Thanks

Sam

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**Jason Brownlee** February 9, 2017 at 7:21 am #

I would encourage you to re-read this post, it sells out exactly how to frame your problem S



**Sam** February 10, 2017 at 11:03 am #

REPLY ↩

Thanks for your response Jason. I understood the above example. The above example seems to be predicting Y as regression value. But I am trying to predict Y as classification value (attrition = 1 or non attrition = 0)

Example : Below is the time series of revenue where 1,2,3.. are the months and Y tell us if the customer

Get Your Start in Machine Learning

So can i use the below format for my test data ?

revenue1 revenue2 revenue3 ...Y

100 50 -25 1

200 100 300 ..... 0

Appreciate your help.

Thanks

Sam



**Jason Brownlee** February 11, 2017 at 4:52 am #

Looks good Sam.



**Sam** February 11, 2017 at 7:16 am #

Thanks a ton Jason for your quick response.You made my day 😊



**Jason Brownlee** February 11, 2017 at 8:21 am #

I'm here to help Sam.



**Anthony from Sydney** February 22, 2017 at 3:22 pm #

Dear Dr Jason,

Two topics please

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

REPLY ↩

Get Your Start in Machine Learning

(1) On cropping data and applying the model 'to the real world'. I understand that cropping is done on the  $0:n$  and  $k:n$  data points to get a 1:1 correspondence between data values at  $t$  and  $t-1$ . I assume from previous posts that you crop say the  $(k-10)$ th to  $k$ th data points, perform the successive 1 step ahead predictions and select the model based on the min(set of mse of all selected models) of the difference between the test and predicted models.

(a) Is the idea to use the that model to predict the  $(k+1)$ th unknown.

(b) Can we assume that the model you 'trained' will be acceptable when more data is acquired. In other words, what happens if you collect another  $x$  data points, and you want to predict the  $(k + x + 1)$  data point, can we assume that the model trained at  $k$  data points will work for the model at  $k + x$  data points? Or in other words, when do you 'retrain' the model.

(2) On windowing the data: based on this blog, is the purpose of windowing the data to find the differences and train the differenced data to find the model. How can we make the assumption that the  $(k+1)$ th differenced observation can be predicted from the  $k$ th differenced observation.

Thank you,

Anthony from Sydney Australia



**Jason Brownlee** February 23, 2017 at 8:51 am #

Hi Anthony,

Sorry, I don't understand what you mean by cropping. Perhaps you could give an example?

Generally, we use all available historical data to make a one-step prediction ( $t+1$ ) or a multi-step prediction. We then evaluate the model and when new data becomes available.

Windowing is about framing a univariate time series into a supervised learning problem with lag observations. We use supervised learning algorithms to model the problem and make predictions.

I hope that helps.

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**Anthony from Sydney** February 23, 2017 at 10:02 am #

REPLY ↩

Dear Dr Jason,

I will rephrase both (1) and (2) into one.

Perhaps I wasn't very clear at all.

Get Your Start in Machine Learning

Cropping. by cropping I mean remove the earliest, the 0th and the latest kth data points because there are no corresponding lagged values by virtue of lagging.

eg

data point value lagged data point array reference

1 ? – this is cropped/pruned 0

2 1 1

3 2 2

44 3 3

5 4 4

..

560 1234 k-1.

? X – this is cropped/pruned. k

dataset available for processing

datapoint lagged data point (array ref based on original data)

2 1 1

3 2 2

44 3 3

5 44 4

..

560 1234 k-1

This is the above dataset with the 0th and kth elements cropped/pruned from the original.

I should have been clearer. I apologise.

My questions

(a) Based on the 'new' lagged dataset, how can you make a prediction for the  $(k + 1)$ th dataset point? Or in other words how can we assume that differencing or windowing as in this tutorial/blog will be the basis of our training model?

(b) Perhaps I'm missing something, having read the other posts on ARIMA. How can we make the assumption that predicting the next data point is based on the previous data point when there may well be MA or AR or other kinds processes on the data? Or in other words how can we assume that differencing or windowing as in this tutorial/blog will be the basis of our training model?

(c) Suppose you trained your model based on the original dataset. Suppose that as your system acquires more datapoints, won't the original model that you trained become invalid. Say you got an extra 10 or 1000 datapoints, do you have to retrain your data because the coefficients of the original model may not be an adequate predictor for a larger dataset.

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

Get Your Start in Machine Learning

Thank you again and I hope I have been clearer,  
Anthony of Sydney Australia



**Jason Brownlee** February 24, 2017 at 10:08 am #

REPLY ↩

Hi Anthony,

What is  $k$ ? Is that a time step  $t$ ? I think it is given context.

If you want to forecast a new data point that is out of sample ( $t+1$ ) beyond the training dataset, your model will use  $t-1, \dots, t-n$  as inputs to make the forecast.

This applied regardless of the type of model used. E.g. if you are using an AR, the inputs will be the autoregression of the lagged error series.

If differencing is performed in the preparation of the model, it will have to be performed on all data. The model will adjust is based on the data itself and your analysis of temporal structure like trends and seasonality.

Yes, as new data comes in the model will need to be refit. This is not a requirement for all problems. In general, for most expectation, we evaluate models in the same way using walk-forward validation that does not require refitting the model and predicts the next out of sample observation.

I hope this helps. I do cover all of this in my book, lesson by lesson.

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**Anthony from Sydney** February 23, 2017 at 10:06 am #

Dear Dr Jason, apologies again, my original spaced data set example did not appear neat.

In both the original and the cropped/pruned/windowed datasets, there are meant to be three columns consisting of the data, data lagged by 1, and the array index based on the original dataset.

I don't know how to get nicely spaced tabbed data when posting replies on this blog

Regards

Anthony of Sydney

Get Your Start in Machine Learning



**Jason Brownlee** February 24, 2017 at 10:10 am #

REPLY ↩

You can use the pre HTML tag, e.g.:

```
1 this, is, pre, formatted, data
```



**Anthony from Sydney** February 24, 2017 at 11:36 am #

REPLY ↩

[src]<https://en.wikipedia.org/wiki/BBCode>[/src]

On how to insert BBCode in forum replies

[list]

\* 1 ?

\* 2 1

\* 3 2

\* 4 3

\* ? 4

[/list]

This is an experiment in inserting HTML code on a forum reply.

I hope this works,

[b] Anthony [/b] [i] from Sydney [/i]

## Get Your Start in Machine Learning



You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

REPLY ↩



**Anthony of Sydney** February 26, 2017 at 2:06 pm #

Testing using the 'pre' enclosed in ", inserting "this is a test message", then "

```
1 this is a test message
```

Hope it works

Get Your Start in Machine Learning



**Nirikshith** March 14, 2017 at 2:09 pm #

REPLY ↩

Dear Jason,

have you planned any blog on forecasting Multivariate Time Series? I went through your ARIMA post and it was good start point for me.

Thanks,

#student #aspring data analyst



**Jason Brownlee** March 15, 2017 at 8:07 am #

REPLY ↩

Thnaks Nirikshith.

Yes, I hope to cover multivariate time series forecasting in depth soon.



**Bruce Anthony** March 31, 2017 at 2:23 pm #

Jason,

I am new to machine learning. I have a problem type and I was wondering if you could point me to the right appropriate model/technique. I have a set of time series data(rows), composed of a number of different hundreds of sensors, measured each second. I have a hunch that there is a relationship between the column 1 at time t1 in column 1 and 10 seconds later there is a change in column 2. My desire is to find the columns when a change in one column is reflected in the related column(s). My goal would be to then train a model earlier in time variable prior to the later in time variable changing. Your article is helpful to understand how to find relationships between single column, but how do I train or dig out the relationships between columns?

If you could point me to what parts of machine learning I should focus my learning efforts I would appreciate it.

Thanks

Bruce

## Get Your Start in Machine Learning



You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

Get Your Start in Machine Learning





**Jason Brownlee** April 1, 2017 at 5:51 am #

REPLY ↩

Hi Bruce, time series analysis is a big field. I'd recommend picking up a good practical book.

Generally, consider looking for correlations between specific lags and your output variable. (e.g. correlation plots).

I hope that helps as a start.



**Bruce Anthony** April 1, 2017 at 2:11 pm #

REPLY ↩

Jason,

Thank you, do you have a suggestion for a good book to start with?

Bruce



**Jason Brownlee** April 2, 2017 at 6:24 am #

Yes, my book:

<https://machinelearningmastery.com/introduction-to-time-series-forecasting-with-python/>

If you are interested in R, here are more books:

<http://machinelearningmastery.com/books-on-time-series-forecasting-with-r/>

## Get Your Start in Machine Learning



You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**HP** June 24, 2017 at 5:33 am #

REPLY ↩

Hi Jason,

Very Nice Article, Just had a question whether there is a forecasting technique for Region/Branch based forecasting.

Get Your Start in Machine Learning



**Jason Brownlee** June 24, 2017 at 8:05 am #

REPLY ↩

What do you mean exactly?



**Rishi** July 31, 2017 at 5:28 pm #

REPLY ↩

Hi Jason,

Thanks for this article. I have 2 questions:

1. Is there a way to avoid removing the rows altogether? If we are creating lag (t-2), (t-3) etc then we will have to remove more rows. I have seen Kaggle masters use XGB with missing = NA option so that it handles missing data but not sure what can be done.
2. Can you please shed some light on the fact that data may not be i.i.d.  $P(Y|X)$  (may be identical but  $y|x$  models should fail in this scenario. Am i thinking in the right direction? Also is there a way to check the ii

Regards,  
Rishi

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**Jason Brownlee** August 1, 2017 at 7:54 am #

Thanks Rishi.

Yes, you can mark the values as NaN values, some algorithms can support this, or set them to 0.0 and

Great point. Classical methods would not fail, but may fair worse than methods that are adjusted for the dependence. I'd still recommend spot checking a suit of methods on a problem as a baseline. ARIMA is corrected for the dependence (as far as I remember).



**Rishi** August 1, 2017 at 2:06 pm #

REPLY ↩

Get Your Start in Machine Learning

Thanks for the reply Jason. I was reading up on auto correlation correction in regression ( detected using Durbin Watson) but that was applicable for continuous data – Cochrane orcutt. Is there in general any way to correct for it? I think most of the problems that we work on in real world are time series such as customer churn etc. And I feel time series regression is what we (unknowingly) do as well, as in use X such as performance in last month etc. Please suggest some material.

Regards,  
Rishabh



**Jason Brownlee** August 2, 2017 at 7:44 am #

REPLY ↩

Sorry, I'm not sure what you're asking, can you restate your question Rishi?



**Rishi** August 2, 2017 at 4:06 pm #

Let's say we pick a real life case study, predict customer's retail spend this month. In this case on whether he had a big spend large month or not. Obviously we can have lagged y as X in the model to Residual analysis should give some insight into it for sure (Durbin Watson should also help detect that).

Also problems like customer churn, I always use this approach: fix a timeline lets say 1 Jan, Target is churn information from past (spend in last 2 months Dec and Nov for all customers). Variables used are like spend right for time series kind of classification?

Sorry for a long post, just wanted to clarify my thoughts.

## Get Your Start in Machine Learning



You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE



**Jason Brownlee** August 3, 2017 at 6:46 am #

REPLY ↩

Yes, I would encourage you to test it empirically rather than getting too bogged down in analysis.

You cannot pick the best algorithm for a specific prediction problem analytically.

Get Your Start in Machine Learning



**Varun** August 9, 2017 at 5:48 pm #

REPLY ↩

Hi Jason,

Superb post!

I have a query. I am working on a real life problem of forecasting 3 days sales for a Retail store.

I am thinking of applying a hybrid model (ARIMAX + Neural network) i.e. Dynamic regression with regressors using `auto.arima`, then fitting Neural network model on the residuals. The final forecast will be  $y = L + N$  where  $L$  = forecast from ARIMAX and  $N$  = forecast of residuals from NNETAR. What do you think of this approach?

Also, I need your input on applying the cross validation techniques. I have daily sales data from Jan14-June17. Would it be worth to tune the parameters using cross validation techniques (Adding months/quarters) or should I go ahead training the model only once (Let's say from Jan14-Dec16) and measure the accuracy on the rest? (Test & Validation)? What could be the best approach as I need only 3 days forecast?



**Jason Brownlee** August 10, 2017 at 6:52 am #

It does not matter what I think, use data to make decisions – e.g. the model or combination harness.

Use walk forward validation on time series, more here:

<http://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/>



**Muthu Kalyan** September 22, 2017 at 3:38 am #

Jason

Excellent article about time series forecast. I have a fair understanding of statistical traditional ML techniques and its application. I have couple of questions on applying NN/LSTM to time series forecast

1. To what an extent we need to worry about over fitting?
2. Are there ensemble techniques that apply different models for different time horizons?

## Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

Get Your Start in Machine Learning



**Jason Brownlee** September 22, 2017 at 5:39 am #

REPLY ↩

Overfitting is always a problem in applied machine learning.

Not sure I follow. If you have different time horizons, then you will need different models to make those predictions. Perhaps you can use outputs from one model as inputs to another, but I have not seen a structured way to do this – I'd encourage you to experiment.

## Leave a Reply

 Name (required) Email (will not be published) (required) Website

SUBMIT COMMENT

## Get Your Start in Machine Learning ×

You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

 Email Address

START MY EMAIL COURSE

Get Your Start in Machine Learning

## Welcome to Machine Learning Mastery



Hi, I'm Dr. Jason Brownlee.

My goal is to make practitioners like YOU awesome at applied machine learning.

[Read More](#)

### Get Good at Time Series Forecasting

Need visualizations and forecast models?

Looking for step-by-step tutorials?

Want end-to-end projects?

[Get Started with Time Series Forecasting in Python!](#)

#### Introduction to Time Series Forecasting with Python

How to Prepare Data and Develop  
Models to Predict the Future

Jason Brownlee

MACHINE  
LEARNING  
MASTERY



## Get Your Start in Machine Learning ×

You can master applied Machine Learning  
**without the math or fancy degree.**

Find out how in this *free* and *practical* email  
course.

[START MY EMAIL COURSE](#)

POPULAR

[Get Your Start in Machine Learning](#)

**Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras**

JULY 21, 2016

**Your First Machine Learning Project in Python Step-By-Step**

JUNE 10, 2016

**Develop Your First Neural Network in Python With Keras Step-By-Step**

MAY 24, 2016

**Sequence Classification with LSTM Recurrent Neural Networks in Python with Keras**

JULY 26, 2016

**How to Setup a Python Environment for Machine Learning and Deep Learning with Anaconda**

MARCH 13, 2017

**Time Series Forecasting with the Long Short-Term Memory Network in Python**

APRIL 7, 2017

**Multi-Class Classification Tutorial with the Keras Deep Learning Library**

JUNE 2, 2016

**Regression Tutorial with the Keras Deep Learning Library in Python**

JUNE 9, 2016

**Multivariate Time Series Forecasting with LSTMs in Keras**

AUGUST 14, 2017

**How to Implement the Backpropagation Algorithm From Scratch In Python**

NOVEMBER 7, 2016

## Get Your Start in Machine Learning



You can master applied Machine Learning **without the math or fancy degree.** Find out how in this *free* and *practical* email course.

[START MY EMAIL COURSE](#)[Get Your Start in Machine Learning](#)

© 2017 Machine Learning Mastery. All Rights Reserved.

[Privacy](#) | [Contact](#) | [About](#)

## Get Your Start in Machine Learning ×

You can master applied Machine Learning **without the math or fancy degree.**

Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE