2020 March 10 - 13

Scribed by Roh

hyunwoo@uchicago.edu

1 Concepts in Time series

I will two most crucial concepts related to time series data/model

- Stationary / Non-stationary
- Memory / Memoryless

These are related to the main task of time series, which is prediction.

1.1 Definition

In many time series text book, talks about the stationarity and non-stationarity. However, are they testable?

> It is impossible to test whether a time series is non-stationary with a single path observed over a bounded time interval - no matter how long. Every statistical test of stationarity makes an additional assumption about the family of diffusions the underlying process belongs to. Thus, a null hypothesis rejection can either represent empirical evidence that the diffusion assumption is incorrect, or that the diffusion assumption is correct. **But** not the null hypothesis (e.g. the presence of a unit root) is false. The statistical test by itself is **inconclusive** about which scenario holds.

There are other points that worth spending time and thinking about them seriously.

- Memory and stationarity are nothing to do with each other.
- Iterated differentiation of a time series does not make a time series more stationary, it rather makes a time series more memoryless; a time series can be both memorlyess and non-stationary.
- Non-stationarity but memoryless time series can easily trick the statistical test of unit-root.

Stationarity refers to the property of things that do not change over time.
We would like to analyze the process that is stationary, which is usually assumed in many machine learning application. However, this is very strong assumption because things change over time. Stationarity is a wishful assumption inherent to quantitative investment management.

"Stationarity in Financial Markets is Self-Destructive"

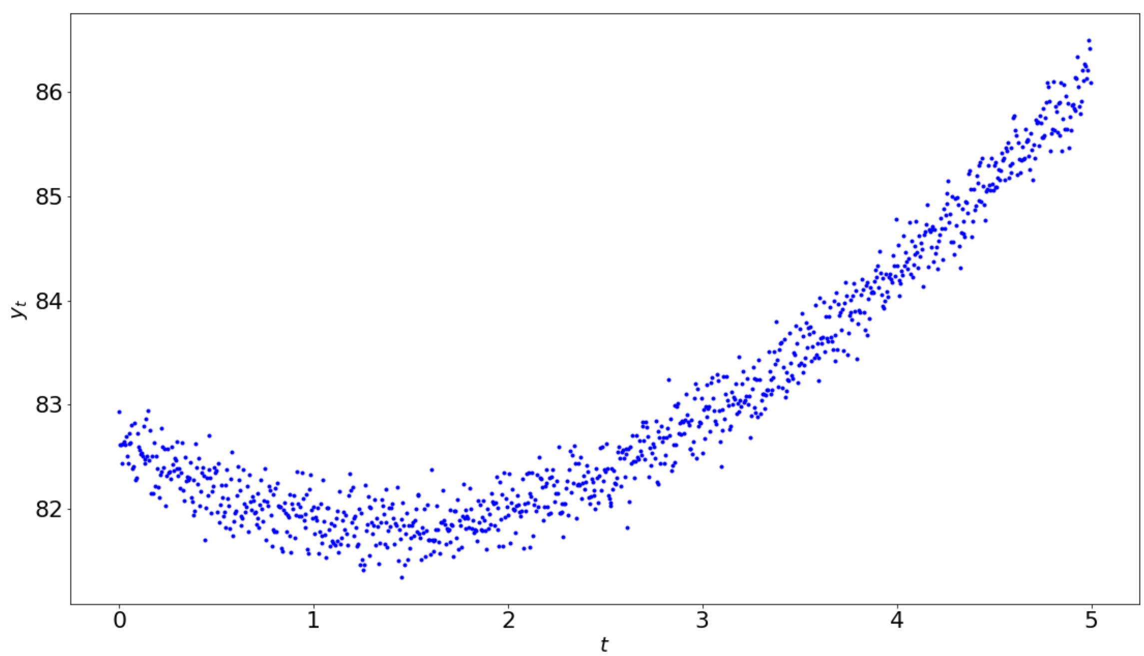
The better an alpha, the more likely it will be copied by competitors over time, and therefore the more likely it is to fade over time. Hence, every predictive pattern is bound to be a temporary or transient regime. How long the regime will last depends on the rigor used in the alpha search, and the secrecy around its exploitation.

1.2 Stationarity can't be disproved with one finite sample

In the case of time series (a.k.a stochastic processes), stationarity has a precise meaning;

1. A time series, $s(x)$, is said to be strongly stationary when all its properties are invariant by change of the origin of time, or time translation. The distribution of the random process has certain attributes that are same everywhere. **Strict stationarity** indicates that for any number k of any sites $x_1, x_2, \cdots, x_k$ the joint cumulative distribution of $(s(x_1), \cdots, s(x_k))$ remains the same under an arbitrary translation h:
$$P\left(s(x_1), \cdots, s(x_k)\right) = P\left(s(x_1 + h), \cdots, s(x_k + h)\right)$$
2. A time series is said to be second-order stationary, or weakly stationary when its mean and auto-covariance functions are invariant by change of the origin of time, or time translation
   (a) the mean of the process does not depend on x : $E[s(x)] = \mu$
   (b) the variance of the process does not depend on x : $E[s(x) - \mu]^2 = \sigma^2$
   (c) the covariance between $s(x)$ and $s(x+h)$ only depends on h :
$$\begin{aligned} Cov[s(x), s(x+h)] &= E\left[\left(s(x) - E[s(x)]\right)\left(s(x+h) - E[s(x+h)]\right)\right] \\ &= E\left[\left(s(x) - \mu\right)\left(s(x+h) - \mu\right)\right] \\ &= C(h) \end{aligned}$$
   Note that $C(0) = \sigma^2$

Because of its strong assumption on distribution over time, we wish to do modelling on stationary process. This is a reason why many of existing time series model assumes the stochastic process to be stationary. However, **the real problem is that we cannot test the stationarity of the finite sample path.** Let's talk about it why with the example below.



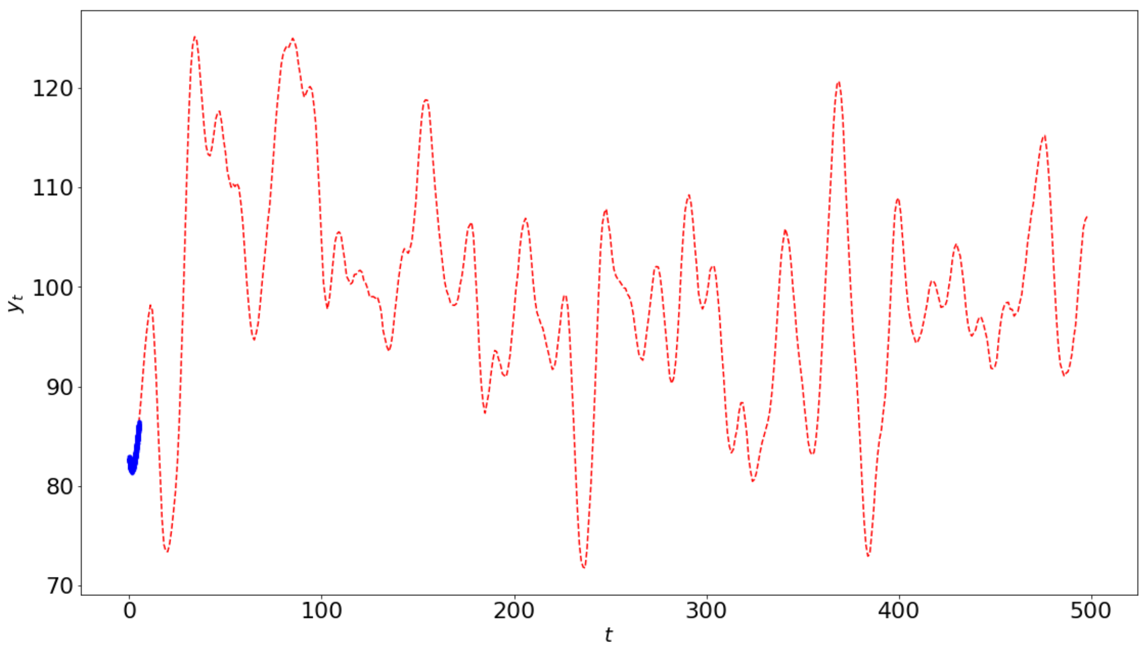
Is this the plot of stationary time series? How do we test usually? We use the stationarity test, called Augmented Dickey Fuller test (ADF test). If we run the ADF test, we get

```
ADF statistic: -2.0261
p-value: 0.2753
```

Therefore, we cannot reject the null hypothesis that the time series is an AR that has a unit root. This means that the test results says the process is not stationary. However, the actually it is a draw from a Gaussian process with mean 100 and auto-covariance function

$$\mathbb{C}\mathrm{ov}(y_u, y_v) = 100 \exp\left(-\frac{1}{2}\left(\frac{u-v}{5}\right)^2\right) + \frac{1}{25}\mathbb{I}\left(u == v\right)$$

Note that this auto-covariance function does not change over time so that we agree that it is indeed a draw from a stationary time series. To see this is right, we do draw much more sample from the sample auto-covariance function over a much longer time horizon.



Now it looks more like what we'd expect from a stationary time series (e.g. it is visually mean-reverting). Let's confirm that with our ADF test:

```
ADF statistic: -4.2942
p-value: 0.0005
```

Indeed, we can reject the null hypothesis that the time series is non-stationary at a 0.05% p-Value, which gives us strong confidence. However, the process did not changed between the two experiments. In fact even the random path used is the same, and both experiments have enough points (at least a thousand each). So what's wrong? Intuitively, although the first experiment had a large enough sample size, it didn't spend long enough a time interval to be characteristic of the underlying process, and there is no way we could have known that beforehand! For this reason, it is simply impossible to test whether a time series is stationary from a single path observed over a finite time interval, without making **any additional assumption.**

1.3 What assumption?

1.3.1 Stationary is about the whole process not of an finite observed path

Stationarity is a property of a **stochastic process, not of a path**. Attempting to test stationarity from a single path ought to implicitly rely on the assumption that the path at hand is sufficiently informative about the nature of the underlying process. As we saw above, this might not be the case and, more importantly, one has no way of ruling out this hypothesis. Because a path does not look mean-reverting does not mean that the underlying process is not stationary. You might not have observed enough data to characterize the whole process.

Along this line, any financial time series, whether it passes the ADF test or not, can always be extended into a time series that passes the ADF test (hint: there exist stationary stochastic processes whose space of paths are universal). Because we do not know what the future holds, strictly speaking, saying that financial time series are non-stationary is slightly abusive, at least as much so as saying that financial time series are stationary.

In the absence of evidence of stationarity, a time series should not be assumed to be non-stationary — we simply can't favor one property over the other statistically. This works similarly to any logical reasoning about a binary proposition A: **no evidence that A holds is never evidence that A does not hold.**

1.3.2 Assumption on the class of diffusions

Every statistical test of stationarity relies on an assumption on the class of diffusions in which the underlying process's diffusion must lie. Without this, we simply cannot construct the statistic to use for the test. Commonly used (unit root) tests typically assume that the true diffusion is an AR process, and test the absence of a unit root as a proxy for stationary.

> The implication is that such tests do not have as null hypothesis that the underlying process is non stationary, but instead that the underlying process is non-stationry AR process!

Here the empirical evidence leading to reject the null hypothesis could point to

- Underlying process is not an AR
- Underlying process is not a stationary
- Underlying process is both not AR and stationary

Unit root tests by themselves are not enough to rule out the possibility that the underlying process might not be an AR process. The same holds for other testes of stationarity that place different assumptions on the underlying diffusion. Without a model there is no statistical hypothesis test, and no statistical hypothesis test can validate the model assumption on which is based.

1.3.3 Seek Stationary Alphas, Not stationary Time series

Given that we cannot test whether a time series is stationary without making an assumption on its diffusion, we are faced with two options:

- Make an assumption on the diffusion and test stationary
- Learn a predictive model, with or without assuming stationarity

The former approach is the most commonly used in the econometrics literature because of the influence of the Box-Jenkins method, whereas the latter is more consistent with the machine learning spirit consisting of flexibly learning the data generating distribution from observations.. Modeling financial markets is hard, very hard, as markets are complex, almost chaotic systems with very low signal-to-noise ratios. Any attempt to properly characterize market dynamics — for instance by attempting to construct stationary transformations — as a requirement for constructing alphas, is brave, counter-intuitive, and inefficient.

1.4 Thought on Memory? What is memory in time series? Its relation to stationarity?

Intuitively, a time series should be thought to have memory when its past values are related to its future values.

> Standard stationairty transformations, like integer differentiation, further recrudesce the signal by removing memory. **Prices series have memory** , because every value is dependent upon a long history of previous levels. In contrast, integer differentiated series, like returns, have a memory cut-off, in the sense that history is disregarded entirely after a finite sample window. Once stationarity transformations have wiped out all memory from the data, statistician resort to complex mathematical techniques to extract whatever residual signal remains. Not surprisingly, applying theses complex techniques on memory-erased series likely leads to false discoveries. In this chapter, we introduce data tranformation method that ensures the stationarity of the data while preserving as much memory as possible.

It is common in finance to find non-stationary time series. What makes these series non-stationary is the presense of memory, i.e., a long history of previous levels that shift the series' mean over time. In order to perform inferential analyses, researchers need to work with invariant processes, such as returns on prices (or changes in logprices), changes in yield, or changes in volatility. These data transformations make the series stationary, at the expense of removing all memory from the original series (Alexander [2001], chapter 11).

Although stationarity is a necessary property for inferential purpose, it is rarely the case in signal processing that we wish all memory to be erased, as that memory is the basis for the model's predictive power. For example, equilibrium (stationary) models need some memory to assess how far the price process has drifted away from the long-term expected value in order to generate a forecast. The dilemma is that returns are stationary, however memory-less and prices have memory, however they are non-stationary.

What is the minimum amount of differentiation that makes a price series stationary while preserving as much memory as possible?

There is a wide region between these two extremes (fully differentiated series on one hand, and zero differentiated series on the other) that can be explored through fractional differentiation for the purpose of developing a highly predictive ML model. Supervised learning algorithms typically require stationary features. The reason is that we need to map a previously unseen (unlabeled) observation to a collection of labeled examples, and infer from them the label of that new observation. If the features are not stationary, we cannot map the new observation to a large number of known examples. But stationarity does not ensure predictive power. Stationarity is a necessary, non-sufficient condition for the high performance of an ML algorithm. **The problem is there is a trade-off between stationarity and memory**

It this true?

Part of it is true now that we can always make a series more stationary through differentiation, but it will be at the cost of erasing some memory, which will defeat the forecasting purpose of the ML algorithm. Virtually all the financial time series literature is based on the premise of making non-stationary series stationary through integer transformation. This raises two questions:

1. Why would integer 1 differentiation (like the one used for computing returns on log-prices) be optimal?

2. Is over differentiation one reaason why the literature has been so biased in favor of the efficient markets hypothesis?

The notion of fractional differentiation applied to the predictive time series analysis dates back at least to Hosking [1981]. In that paper, a family of ARIMA processes was generalized by permitting the degree of differencing to take fractional values. This was useful because fractionally differenced processes exhibit long-term persistence and antipersistence, hence enhancing the forecasting power compared to the standard ARIMA approach. In the same paper, Hosking states: "Apart from a passing reference by Granger (1978), fractional differencing does not appear to have been previously mentioned in connection with time series analysis.

1.5 The Method

Considering the backshift operator, B, applied to a matrix of real-valued features $\{X_t\}$, where $B^k X_t = X_{t-k}$ for any integer $k \geq 0$. For example, $(1-B)^2 = 1 - 2B + B^2$, where $B^2 X_t = X_{t-2}$, so that $(1-B)^2 X_t = X_t - 2X_{t-1} + X_{t-2}$. Note that $(x+y)^n = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} x^k y^{n-k}$, for n a positive integer. For a real number d,

$$(1+x)^d = \sum_{k=0}^{n} \begin{pmatrix} d \\ k \end{pmatrix} x^k$$

the binomial series.

In a fractional model, the exponent d is allowed to be a real number, with the following formal binomial series expansion:

$$\begin{aligned}
(1-B)^d &= \sum_{k=0}^{n} \begin{pmatrix} d \\ k \end{pmatrix} (-B)^k \\
&= \sum_{k=0}^{n} \frac{\prod_{i=0}^{k-1}(d-i)}{k!} (-B)^k \\
&= \sum_{k=0}^{n} \frac{(-B)^k}{k!} \prod_{i=0}^{k-1} (d-i) \\
&= 1 - dB + \frac{d(d-1)}{2!} B^2 - \frac{d(d-1)(d-2)}{3!} B^3 + \cdots
\end{aligned}$$

1.5.1 Long Memory

Let us see how a real (non-integer) positive d preserves memory. This arithmetic series consists of dot product

$$\widetilde{X}_t = \sum_{k=0}^{\infty} w_k X_{t-k}$$

whose coefficients are determined based on the notion of fractional differentiation with a fixed-window, as an alternative to log-returns (order 1 differentiation on log-prices). The weights w is

$$w = \left\{ 1, -d, \frac{d(d-1)}{2!}, -\frac{d(d-1)(d-2)}{3!}, \cdots, (-1)^k \prod_{i=0}^{k-1} \frac{d-i}{k!}, \cdots \right\}$$

and values X

$$X = \{X_t, X_{t-1}, X_{t-2}, \cdots, X_{t-k}, \cdots\}$$

When d is a positive integer number and memory beyond that point is cancelled. For example, $d = 1$ is used to compute returns and

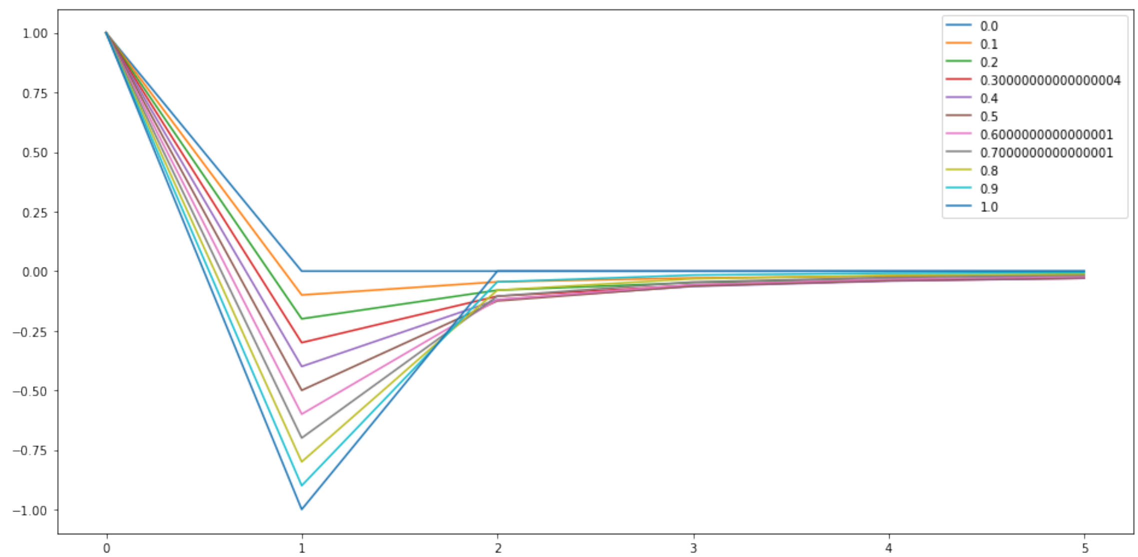
$$w = \{1, -1, 0, 0, \cdots\}$$

1.5.2 Iterative Estimation

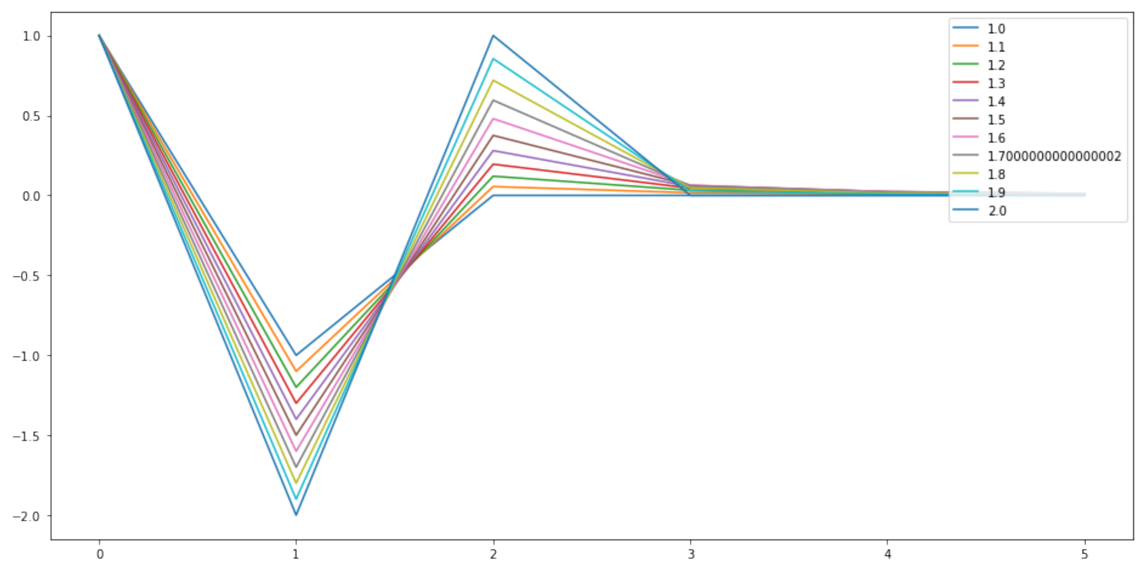
Looking at the sequence of weights, w, we can appreciate that for $k = 0, ..., \infty$, with $w_0 = 1$, the weights can be generated iteratively as:

$$w_k = -w_{k-1} \cdot \frac{d-k+1}{k}$$

Here is the plot.



The plot of weights where $d \in [1, 2]$ is



What do we want to do?

We want to preserve the memory and at the same time we want to make time series “stationary”. The author recommends choosing the smallest degree of fractional differentiation $0 < d < 1$ for which the moving average time series passes the ADF stationary test (at a given p-value).

Reference:

1. https://medium.com/kxytechnologies/non-stationarity-and-memory-in-financial-markets-4b8d1200667c

2. Book: Advances in Financial Machine
3. https://github.com/jjakimoto/finance_ml/blob/master/examples/miscellaneous/Ch%205.ipynb