

## A NOTE ON TIME SERIES DIFFERENCING

FIRUZ KAMALOV<sup>1\*</sup>

**ABSTRACT.** Differencing is one of the key tools time series analysis. It is commonly used to obtain stationary time series. In this note, we show that the  $n$ th difference of a weakly stationary time series is weakly stationary. Similarly we prove that the  $n$ th difference of a strictly stationary time series is strictly stationary. We also consider the effect of differencing on the time series autocovariance.

### 1. INTRODUCTION

Time series analysis is traditionally carried out under the assumption of stationarity. However, in practice, the observed time series is rarely stationary. One of the main ways to deal with nonstationary series is differencing. The differencing operation involves taking the difference of two consecutive values of the time series

$$\nabla x_t = x_t - x_{t-1}. \quad (1.1)$$

It can also be expressed in terms of the backshift operator  $B$

$$\nabla x_t = (1 - B)x_t. \quad (1.2)$$

Time series differencing is mainly used to remove trend in the series. It is akin to taking the derivative. If there exists a nonlinear trend in the time series then a higher order differencing may be required. After fitting a time series model to the sample data the original scale can be recovered from Equation 1.1.

Time series differencing is a popular tool due to its efficiency in achieving stationarity. Taking the first difference often leads to substantial reduction in nonstationary behavior of the series. However, the second and higher order differencing may be required to completely eliminate nonstationary behavior. In this note, we show that differencing is a safe procedure in the sense that it does not introduce new nonstationary behavior. Concretely, we prove that the  $n$ th difference of a (weakly) stationary series is (weakly) stationary. Similarly, we show that the  $n$ th difference of a strictly stationary series is strictly stationary.

---

*Date:* Received: Nov 3, 2020; Accepted: Feb 2, 2021.

\* Corresponding author.

2010 *Mathematics Subject Classification.* Primary 62M10; Secondary 91B84.

*Key words and phrases.* differencing, stationary, weakly stationary, strictly stationary, time series, arma, autocovariance.

The paper is organized as follows. In Section 2, we provide a brief review of the existing literature. In Section 3, we present our main results. The paper is concluded with final remarks in Section 4.

## 2. LITERATURE

Time series research has experienced a tremendous growth over the last decade [7]. The direction of the research has bifurcated between the theory and applications. In particular, stationary time series analysis has gained significant track. In [15], the authors propose a new method for speeding up Markov Chain Monte Carlo for stationary time series data by efficient data subsampling in the frequency domain. The authors demonstrate a speedup of up to two orders of magnitude for certain classes of challenging problems. In [3], the authors apply model-free prediction to handle locally stationary time series. It is shown that in certain cases model-free forecast outperforms model-based forecast. A versatile bootstrap method is introduced in [16] to estimate the long-run covariance of stationary functional time series. The proposed method relies on functional principal component analysis, where principal component scores can be bootstrapped by maximum entropy. Distributional properties of a quadratic form of a stationary functional time series are studied under mild moment conditions in [4]. As a result, the authors determine consistency rates of estimators of spectral density operators. Tight boundaries for the autocovariance function of a stationary linear process with  $p$ -series coefficients are presented in [8]. Integrals inequalities are applied to produce the bounds for the ACF estimates. In addition to the classical econometric and signal processing approaches, the new generation of deep learning methods have also been applied in time series analysis [10, 11, 12]. The authors of [2] approach stochastic optimization based on streaming PCA problem for stationary time series data. In the study, the principle component of time series data with respect to the covariance matrix of the stationary distribution is estimated. The authors in [9] applied long short-term memory neural networks to develop a stock market prediction model. The model is applied to predict significant changes in stock values.

Differencing is widely used as a preprocessing step in many applications of science, engineering, and business. The effects of differencing and decomposition preprocessing is analyzed in the context of wind speed prediction by [1]. The authors discover that computational delay is significantly reduced when applying the differencing procedure. It is used in medical care analysis [5] to analyze the temporal hospital quality measures. In hydrology, it is used to analyze streamflow and water temperature series [13]. In signal processing, a low-dispersion time differencing scheme was proposed to Maxwell's equations by [14]. The proposed differencing method allows to achieve accuracy level comparable to the third-order Runge-Kutta scheme yet significantly more efficiently. More recently, fractional differencing was applied to study stock market behavior and compared to online presence [6]. The authors find that both stock market prices and online search trends require further exploration for modeling and forecasting.

## 3. MAIN RESULTS

Let us set up the basic definitions and notation that will be used in the subsequent discussion. Let  $x_t$  be a times series process with mean  $\mu_t$  and autocovariance function  $\gamma(s, t)$  for  $s, t \geq 1$ . The time series  $x_t$  is called (weakly) stationary if the following conditions hold

$$\mu_t = \mu_x, \text{ for all } t \geq 1. \quad (3.1)$$

In other words, the time series has constant expected value at all time steps. And

$$\gamma(s, t) = \gamma(h), \quad h = 0, \pm 1, \pm 2, \dots \quad (3.2)$$

where  $h = s - t$ . In other words, the time series covariance depends only on the difference in time steps.

We denote the  $n^{\text{th}}$  difference of  $x_t$  by  $\nabla^n x_t$ . The mean and autocovariance function of  $\nabla^n x_t$  is denoted by  $\mu_{\nabla^n x}$  and  $\gamma_{\nabla^n x}(h)$  respectively.

**Theorem 3.1.** *Suppose  $x_t$  is a stationary process. Then the  $n^{\text{th}}$  difference of  $x_t$  is stationary for all  $n \geq 0$ .*

*Proof.* We use induction to prove the desired result. First, note that for  $n = 0$ ,  $\nabla^n x_t = x_t$ . Since  $x_t$  is stationary by the hypothesis the base case of the induction is correct. Next, assume that  $\nabla^k x_t$  is a stationary process. Our goal is to show that  $\nabla^{k+1} x_t$  is also stationary. Let  $\mu_{\nabla^{k+1} x}$  denote the mean of  $\nabla^{k+1} x_t$ . Then

$$\begin{aligned} \mu_{\nabla^{k+1} x} &= E[\nabla^k x_t - \nabla^k x_{t-1}] \\ &= \mu_{\nabla^k x} - \mu_{\nabla^k x} \\ &= 0 \end{aligned} \quad (3.3)$$

Next, we will show that the autocovariance function  $\gamma_{\nabla^{k+1} x}(s, t)$  depends only on the difference  $s - t$ ,

$$\begin{aligned} \gamma_{\nabla^{k+1} x}(s, t) &= \text{Cov}[\nabla^{k+1} x_s, \nabla^{k+1} x_t] \\ &= \text{Cov}[\nabla^k x_s - \nabla^k x_{s-1}, \nabla^k x_t - \nabla^k x_{t-1}] \\ &= \text{Cov}[\nabla^k x_s, \nabla^k x_t] - \text{Cov}[\nabla^k x_s, \nabla^k x_{t-1}] - \text{Cov}[\nabla^k x_{s-1}, \nabla^k x_t] + \text{Cov}[\nabla^k x_{s-1}, \nabla^k x_{t-1}] \\ &= \gamma_{\nabla^k x}(h) - \gamma_{\nabla^k x}(h+1) - \gamma_{\nabla^k x}(h-1) + \gamma_{\nabla^k x}(h) \\ &= 2\gamma_{\nabla^k x}(h) - \gamma_{\nabla^k x}(h+1) - \gamma_{\nabla^k x}(h-1), \end{aligned} \quad (3.4)$$

where  $h = s - t$ . It follows by induction that  $\nabla^n x_t$  is stationary for all  $n \geq 0$ .  $\square$

Next, we prove the that the  $n^{\text{th}}$  difference of a strictly stationary series is strictly stationary. A time series  $x_t$  is called strictly stationary if the following condition holds

$$F_{\mathbf{X}}(x_{t_1}, x_{t_2}, \dots, x_{t_k}) = F_{\mathbf{X}+h}(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}), \quad (3.5)$$

for any choice of  $t_1, t_2, \dots, t_k$  and  $h \geq 1$ , where  $\mathbf{X} = (X_{t_1}, X_{t_2}, \dots, X_{t_k})$  and  $\mathbf{X}+h = (X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$ . In other words, the joint distribution of the time series at any subset of time steps is time invariant. First, we will show that the result holds for  $n = 1$ .

**Lemma 3.2.** Suppose  $x_t$  is a strictly stationary process. Then the 1<sup>st</sup> difference of  $x_t$  is strictly stationary.

*Proof.* Let  $\nabla x_t = x_t - x_{t-1}$  be the first difference at  $t$  and  $f$  denote the joint probability density function. Then

$$Pr(\nabla X_t \leq y) = Pr(X_t - X_{t-1} \leq y) = \int_D f(x_t, x_{t-1}) dA, \quad (3.6)$$

where the double integral is taken over the region  $D$  illustrated in Figure 1. Let

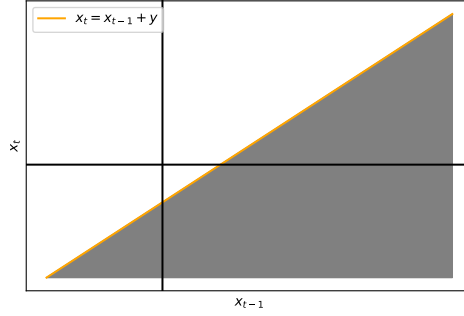


FIGURE 1. The region of integration  $D : x_t \leq x_{t-1} + y$ .

the choice of  $t_1, t_2, \dots, t_k$  and  $h \geq 1$  be given. Then

$$\begin{aligned} & Pr(\nabla X_{t_1} \leq y_1, \nabla X_{t_2} \leq y_2, \dots, \nabla X_{t_k} \leq y_k) \\ &= Pr(X_{t_1} - X_{t_1-1} \leq y_1, X_{t_2} - X_{t_2-1} \leq y_2, \dots, X_{t_k} - X_{t_k-1} \leq y_k) \\ &= \int_{D_k} \dots \int_{D_2} \int_{D_1} f(x_{t_1}, x_{t_1-1}, x_{t_2}, x_{t_2-1}, \dots, x_{t_k}, x_{t_k-1}) dA_1 dA_2 \dots dA_k \\ &= \int_{D_k} \dots \int_{D_2} \int_{D_1} f(x_{t_1+h}, x_{t_1-1+h}, x_{t_2+h}, x_{t_2-1+h}, \dots, x_{t_k+h}, x_{t_k-1+h}) dA_1 dA_2 \dots dA_k \\ &= Pr(\nabla X_{t_1+h} \leq y_1, \nabla X_{t_2+h} \leq y_2, \dots, \nabla X_{t_k+h} \leq y_k). \end{aligned} \quad (3.7)$$

It follows that the joint distribution of  $(\nabla x_{t_1}, \nabla x_{t_2}, \dots, \nabla x_{t_k})$  is shift invariant.  $\square$

We can use Lemma 3.2 to prove the general result for the strict stationarity of  $\nabla^n x_t$ .

**Theorem 3.3.** Suppose  $x_t$  is a strictly stationary process. Then the  $n^{\text{th}}$  difference of  $x_t$  is strictly stationary for all  $n \geq 0$ .

*Proof.* We use induction to prove the desired result. First, note that for  $n = 0$ ,  $\nabla^n x_t = x_t$ . Since  $x_t$  is strictly stationary by the hypothesis the base case of the induction is correct. Next, assume that  $\nabla^k x_t$  is a strictly stationary process. Our goal is to show that  $\nabla^{k+1} x_t$  is also strictly stationary. Since  $\nabla^{k+1} x_t = \nabla(\nabla^k x_t)$  then it follows from Lemma 3.2 that  $\nabla^{k+1} x_t$  is strictly stationary.  $\square$

#### 4. DIFFERENCING FOR AUTOCOVARANCE

Differencing is traditionally applied to remove the trend in a time series. A time series with a polynomial trend can be converted into a constant mean series with a few of differencing operations corresponding to the degree of the polynomial. Differencing can also be used to alleviate autocovariance issues. However, the effect of differencing on autocovariance is less clear than its effect on the mean of the time series. Consider the time series given by

$$x_t = ty, \quad (4.1)$$

where  $Y$  is a random variable with  $\mu = 0$  and  $\sigma = 1$ . Then  $\text{Cov}(x_t, x_{t+h}) = t(t+h)$ . After applying the differencing operation we obtain

$$\text{Cov}(\nabla x_t, \nabla x_{t+h}) = 1 \quad (4.2)$$

for all  $t, h$ . In certain cases, differencing can be used to improve the variance, but not the autocovariance of a series. Consider the time series

$$x_t = \alpha_t w_t, \quad \alpha_t = 2 + (-1)^t, \quad (4.3)$$

where  $w_t$  is white noise. Applying the differencing operation yields a constant variance series

$$\begin{aligned} \text{Var}(\nabla x_t) &= \text{Var}(x_t - x_{t-1}) \\ &= \text{Var}(\alpha_t w_t - \alpha_{t-1} w_{t-1}) \\ &= \alpha_t^2 + \alpha_{t-1}^2 \\ &= 10 \end{aligned} \quad (4.4)$$

for all  $t$ . However, differencing does not change the time dependence of the autocovariance. In particular, the autocovariance with time lag  $h = 1$  is not constant

$$\gamma_{\nabla x}(1) = \text{Cov}(x_t - x_{t-1}, x_{t+1} - x_t) = \alpha_t^2. \quad (4.5)$$

Although differencing can be used to address autocovariance issues in special cases, it is not effective in general. Consider the time series

$$x_t = x_{t-1} + \alpha_t w_t, \quad (4.6)$$

where  $w_t$  is white noise and  $\alpha_t$  is the coefficient. The series can be rewritten in the form

$$x_t = \sum_{i=0}^{\infty} \alpha_{t-i} w_{t-i}. \quad (4.7)$$

The differencing of the time series does not yield any reduction in autocovariance time dependency

$$\text{Var}(\nabla x_t) = \alpha_t^2. \quad (4.8)$$

The preceding example illustrates that differencing is not a viable tool to remove autocovariance temporal dependency. Other methods such as taking the natural logarithm of the time series can be used to obtain better results with respect to the autocovariance.

## 5. CONCLUSION

Time series differencing is an important preprocessing tool in time series analysis. It is widely used in many applications. In certain cases, multiple differencing steps must be taken in order to obtain the desired result. In this note, we provide a mathematical proof that differencing is a stationarity-neutral operation in the sense that the  $n^{\text{th}}$  order difference of a stationary time series is stationary. The result is proved both in the case of weakly and strictly stationary time series. Our results shows that it is safe to apply differencing operation as it does not lead to nonstationarity. As a future research avenue, the effects of differencing on the properties of nonstationary time series need to be investigated.

## REFERENCES

- [1] Bokde, N., Feijoo, A., & Kulat, K. (2018). Analysis of differencing and decomposition preprocessing methods for wind speed prediction. *Applied Soft Computing*, 71, 926-938.
- [2] Chen, M., Yang, L., Wang, M., & Zhao, T. (2018). Dimensionality reduction for stationary time series via stochastic nonconvex optimization. *Advances in Neural Information Processing Systems*, 31, 3496-3506.
- [3] Das, S., & Politis, D. N. (2020). Predictive inference for locally stationary time series with an application to climate data. *Journal of the American Statistical Association*, 1-16.
- [4] van Delft, A. (2020). A note on quadratic forms of stationary functional time series under mild conditions. *Stochastic Processes and their Applications*, 130(7), 4206-4251.
- [5] Devkaran, S., & O'Farrell, P. N. (2015). The impact of hospital accreditation on quality measures: an interrupted time series analysis. *BMC health services research*, 15(1), 137.
- [6] Flores-Munoz, F., Baez-Garcia, A. J., & Gutierrez-Barroso, J. (2019). Fractional differencing in stock market price and online presence of global tourist corporations. *Journal of Economics, Finance and Administrative Science*.
- [7] Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.
- [8] Kamalov, F. (2020). A note on the autocovariance of p-series linear process. *Gulf Journal of Mathematics*, 9(2), 40-45.
- [9] Kamalov, F. (2020). Forecasting significant stock price changes using neural networks. *Neural Computing and Applications*, 1-13.
- [10] Kamalov, F., Smail, L., & Gurrib, I. (2020, December). Forecasting with Deep Learning: S&P 500 index. In *2020 13th International Symposium on Computational Intelligence and Design (ISCID)* (pp. 422-425). IEEE.
- [11] Kamalov, F., Smail, L., & Gurrib, I. (2020, November). Stock price forecast with deep learning. In *2020 International Conference on Decision Aid Sciences and Application (DASA)* (pp. 1098-1102). IEEE.
- [12] Kamalov, F., Gurrib, I., & Rajab, K. (2021). Financial Forecasting with Machine Learning: Price Vs Return. *Journal of Computer Science*, 17(3), 251-264.
- [13] Moeeni, H., Bonakdari, H., & Fatemi, S. E. (2017). Stochastic model stationarization by eliminating the periodic term and its effect on time series prediction. *Journal of hydrology*, 547, 348-364.
- [14] Moustauoui, M., & Mahalov, A. (2019). A Low-Dispersion Time Differencing Scheme for Maxwell's Equations. *IEEE Transactions on Antennas and Propagation*, 67(5), 3221-3231.
- [15] Salomone, R., Quiroz, M., Kohn, R., Villani, M., & Tran, M. N. (2020, November). Spectral subsampling MCMC for stationary time series. In *International Conference on Machine Learning* (pp. 8449-8458). PMLR.
- [16] Shang, H. L. (2018). Bootstrap methods for stationary functional time series. *Statistics and Computing*, 28(1), 1-10.

<sup>1</sup> DEPARTMENT OF ELECTRICAL ENGINEERING, CANADIAN UNIVERSITY DUBAI, DUBAI, UAE.

*Email address:* [firuz@cu.ac.ae](mailto:firuz@cu.ac.ae)