25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

GANs through the looking glass: How real is the fake financial data created by Generative Adversarial Neural Nets?

D.E. Allen[a], L. Mushunje[b] and S. Peiris[c]

[a] *School of Mathematics and Statistics, University of Sydney, School of Business and Law, Edith Cowan University, and Department of Finance, Asia University, Taiwan.*

[b] *Commercial Bank of Zimbabwe, Columbia University, New York.*
[c] *School of Mathematics and Statistics, University of Sydney.*
Email: profallen2007@gmail.com

Abstract: Generative Adversarial Neural nets (GANs), developed by Goodfellow et. al (2014) are a branch of machine learning techniques that are one of the newest and most important fields in machine learning. They have several singular characteristics. A GAN, when applied to a data set, learns to generate new data with the same statistics as the training set. This paper examines the characteristics of the time series of financial data developed when a GAN is trained on a sample financial data set comprising daily S&P500 Index values. GANs were originally proposed as a form of generative model for unsupervised learning, but their application and use have developed on several fronts into semi-supervised, supervised and reinforcement learning as well. GANs involve the application of two competing neural networks in a game theoretic context. The Generator net tries to generate pseudo data that is presented to the discriminator net which then attempts to distinguish between the real and the fake data. The process proceeds dynamically, and non-specifically, in that that the generator is not trained to minimize the distance to a specific image of the dataset, but rather to fool the discriminator. This enables the model to learn in an unsupervised manner. A known dataset serves as the initial training data for the discriminator. The generative network generates data sets while the discriminative network evaluates them. The generative network's training objective is to increase the error rate of the discriminative network. The customary procedure is to seed the generator with a randomized input that is sampled from a predefined latent space, such as a normal distribution. Then backpropagation procedures are independently applied to both networks so that the generator learns to produce more convincing samples whilst the discriminator becomes more adept at screening pseudo samples. The game reaches an equilibrium when the generator can fool the discriminator half the time. Potential convergence difficulties lead to the development of Wasserstein GANs Arjovski et al. (2017). We apply Wasser GANs in this paper and examine the characteristics of the generated fake S&P500 data set. Typically, financial data sets have long memory and fat tails. Prices series are non-stationary with variances that increase as a function of time. Fama (1965), in his development of efficient market tests, suggested that markets should have no memory, therefore the autocorrelation of return series should be approximately zero. We explore whether the generated series possess these characteristics. A related issue is how closely does the fake series mimic the real series? We explore this issue using cointegration tests on the levels of the series and regression analysis of the logarithmic first differences, or returns of the series. The results suggest that many of the characteristics of real financial series are captured in the artificial series generated by GANs.

Keywords: *GANS, Neural Nets, Machine learning, Statistical properties*

1. INTRODUCTION

Generative Adversarial Neural Networks (GANs), as developed by Goodfellow et al. (2014) can be viewed as a modified version of the Turing test, first proposed by Alan Turing, the father of modern computing, in 1950. Turing suggested a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human. Turing proposed that a human evaluator would judge natural language conversations between a human and a machine designed to generate human-like responses, via a keyboard, from a machine in the next room. If the participant could not distinguish human responses from those of a machine, the machine would have passed the test.

GANs consist of two different neural networks, a generator G and a discriminator D. The generator G is responsible for the generation of data, and the discriminator D functions to judge the quality of the generated data and provide feedback to the generator G. These neural networks are optimized under game-theoretic conditions: the generator G is optimized to generate data that deceive the discriminator D and the discriminator D is optimized to distinguish the source of the input, namely the generator G or realistic dataset. Below, we provide brief descriptions of some of the main forms of Financial time series GANS models. Unlike other time series models, GANs are maximum likelihood free. We do not necessarily implement such techniques when training GANs.

GANs are an example of generative models, the term can be used to refer to any model that takes a training set, consisting of samples drawn from a distribution pdata, and learns to represent an estimate of that distribution somehow. The result is a probability distribution pmodel. This might be estimated explicitly, or samples drawn from its distribution might be generated.

This might seem redundant but it can inform and improve our ability to represent and manipulate high dimensional probability distributions.. Generative models of time series data can be used to simulate possible future scenarios. GANs can be used to improve image resolution, create art and be used for image translation.

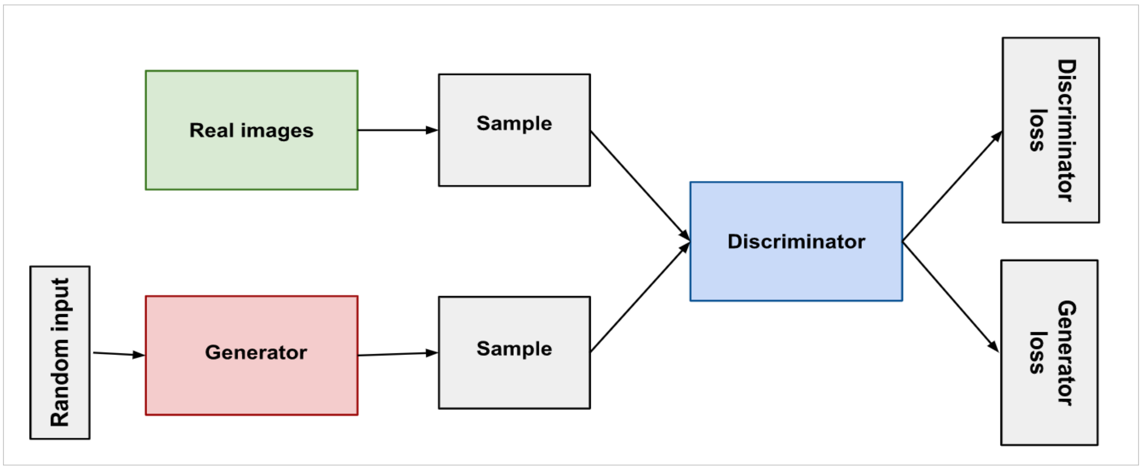


Figure 1. GAN Architecture

**GAN Architecture: The figure depicts all the states and output including the connections between all the networks that makes up a full GAN network. Firstly, the generator network takes in the random input, before passed on to the discriminator network, On the other hand, the real input is passed directly to the discriminator function. The discriminator then classifies the output as either real or fake. This comes with loss functions for each classification, as we shall further discuss.*

GANs have the advantage of using latent code, no Markov chains are required, and they are often regarded as producing the best samples, Goodfellow (2016).

Generator network $x = G(z;\Theta^{(G)})$

Where the function must be differentiable, have no invertibility requirement, be trainable for any size of z, x can be made conditionally Gaussian, given z, but there is no requirement to do this.

$$J^{(D)} = -\frac{1}{2}\mathbb{E}_{x\sim p_{data}} \log D(x) - \frac{1}{2}\mathbb{E}_z \log(1-D)G(z)))$$

$$J^{(G)} = -J(D)$$

What is the solution to *D(x)*in terms of p_{data} and $p_{generator}$? Assume both densities are non-sero everywhere. Solve for where the functional derivatives are zero.

The sample and method is discussed in section 2, the results in section 3 and the paper concludes in section 4.

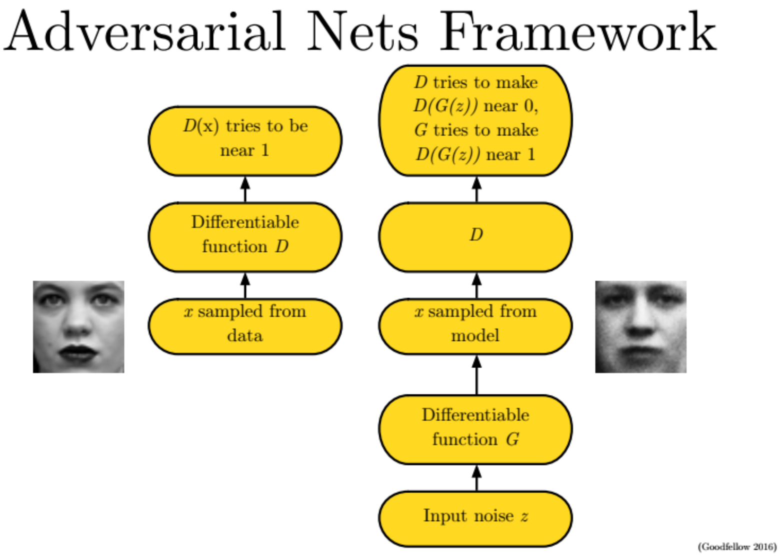


Figure 2. GAN Source: Goodfellow (2016)

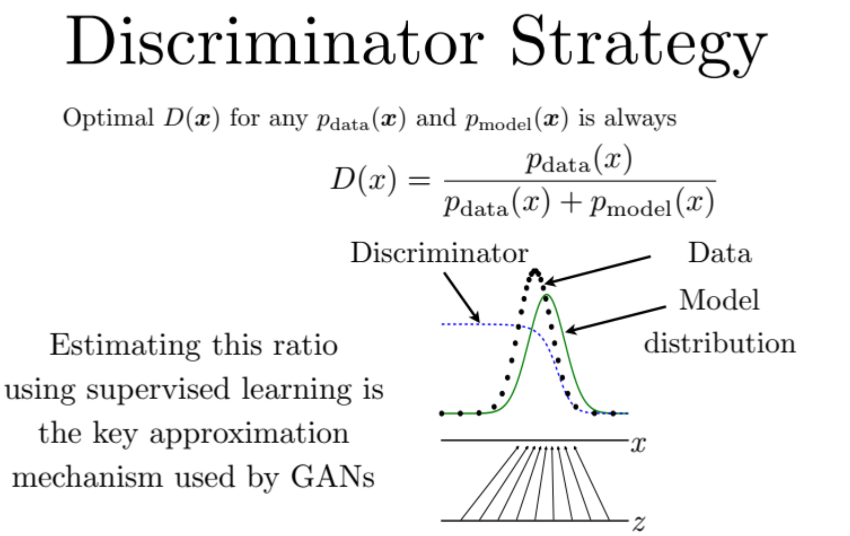


Figure 3. GAN Source: Goodfellow (2016)

2. Sample and method

We downloaded daily data for the S&P 500 index for a period from 3/1/2012 to 22/12/2022. This gave a total of 2063 daily observations. We used used the Python Pytorch library to fit the GANs analysis. The process of fitting the model generated a fake 500 index data set. We use Wasser GANs in this paper to generate the series, see Arjovsky, et al. (2017). The purpose of the current paper is to explore how closely the fake S&P 500 index series mimics the real series.

We did this by contrasting the two series using descriptive statistics. We created a return series for both by taking the logarithm of first differences. We then analysed both of the series by checking the stationarity of both series, and estimated the Hurst Index. We fitted GARCH models to the return series, estimated their periodograms and ran simple tests of market efficiency.

3. RESULTS

Table 1 presents descriptive statistics for both the fake and real series. On all metrics the real and the fake series in both levels and log differences are remarkably similar. The mean and median of the two series are very similar. Their standard deviations, skewness and excess kurtosis are also similar. KPSS unit root tests with trend on both series reject the null of non-stationarity. Their Hurst exponents are almost the same at 1.02 and1.01. This value suggests that both series have long term positive autocorrelation or long memory.

Table 1. Descriptive statistics for Real and Fake S&P500 series

	Real S&P500 Prices	**Fake S&P500 Prices**
Mean	1598.5	1568.5
Median	1471.5	1480.0
Minimum	676.53	491.25
Maximum	2690.2	2624.4
St. Deviation	459.78	491.60
Ex. Kurtosis	-0.89893	-0.99214
Skewness	0.31735	0.097534
KPSS test with trend	4.29702***	3.74915***
Hurst exponent	1.01981	1.01209
	Log difference Real S&P500 prices	**Log difference Fake S&P500 prices**
Mean	0.00023073	0.00022175
Median	0.00059556	0.00056296
Minimum	-0.094695	-0.11871
Maximum	0.10957	0.11413
St. Deviation	0.012664	0.014761
Ex. Kurtosis	10.901	12.234
Skewness	-0.34878	-0.0068651
KPSS test with trend	0.0700386	0.0531918
Hurst exponent	0.548056	0.539908

Note: *** Indicates significance at 1% level.

Plots of the levels of the two series are shown in Figure 4. They follow each other's path very closely.

We also took the logarithmic first differences of the two series to produce a real and a fake S&P500 return series. Descriptive characteristics for the two series are shown in the bottom half of Table 1. Once again in terms of their means, medians, minimums, maximums, standard deviation and excess kurtosis, they are very similar. There is less negative skewness on the fake series. KPSS unit root tests with a trend suggest that both log return series are stationary. The Hurst exponent for both series is now 0.55 and 0.54 respectively, which suggests that any autocorrelations decay rapidly.

A GARCH (1,1) model, see Bollerslev (1986), was estimated for both the real and fake series. The model failed to converge for the fake series so an ARCH (1) model, Engle (1982), was estimated instead. The results are shown in Table 2. Plots of the conditional variances for the above models described in Table 2 are shown in Figure 5.

There is evidence of slightly different behaviour in the lag structure of autocorrelations in the two return series. Figure 6 provides graphs of the periodograms of the two series. A periodogram is an estimate of the spectral density of a signal. The term was coined by Arthur Schuster in 1898. There is relatively more dependence in the fake series that then diminishes rapidly in comparison to the real series.

The relationship between the levels and of the Real S&P 500 and Fake S&P 500 series was examined using an Engle-Granger (1987) bivariate cointegration test. The results in Table 3 show the slope coefficient is close to 1 and significant at the 1 per cent level. The unit root test on the residuals of this regression rejects the null of non-stationarity at the 1 per cent level.

The relationship between the lag structure of the two return series was examined by running a simple test of market efficiency, see Fama (1965), of regressing the current return for both series on one lag of itself. The results are shown in Table 3 and are quite striking. The real series exhibit behaviour that is consistent with weak-form market efficiency. The slope coefficient is negative and significant, but the Adjusted R square suggests that the relationship only explains 1% of the variation in real returns on the S&P 500 Index.

The slope coefficient for the fake S&P 500 return series is also significant and large with a value of 0.47, and the regression has an Adjusted R Square of 22 per cent. This is not consistent with the existence of weak form market efficiency.

Figure 4. Plots of Real S&P500 and Fake S&P500

Table 2. GARCH(1,1) and ARCH(1) models fitted to Real and Fake S&P 500 return series.

Coefficient	Real S&P500 returns	Fake S&P500
Constant	0.000615600***	n.a.
Alpha(0)	2.27786e-06***	7.71906e-05***
Alpha(1)	0.119790***	0.761106***
Beta(1)	0.862700***	n.a.

Note: *** Indicates significance at 1% level.

Table 3. Engle-Granger Cointegration test

	Slope coefficient Fake Series	Adjusted R Squared
Real S&P500 series	0.930453***	0.989720

Note: *** Indicates significance at 1% level.

Table 4. Test of weak-form efficiency

	Slope coefficient	Adjusted R Squared
Real S&P500 return series	−0.104274***	0.010515
Fake S&P 500 return series	0.467936***	0.218690

Note: *** Indicates significance at 1% level.

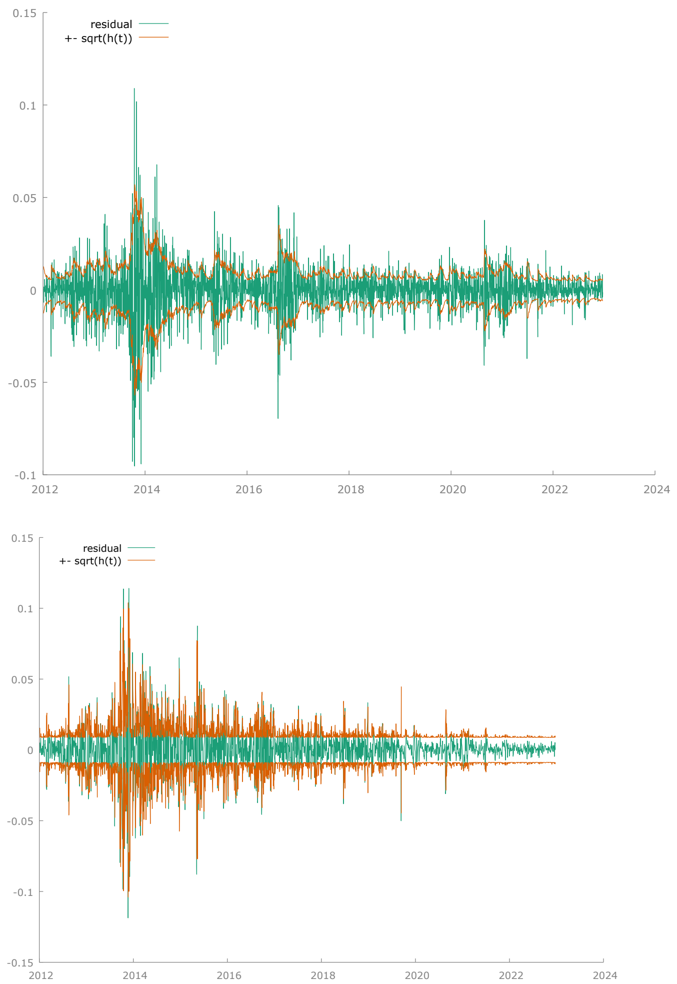
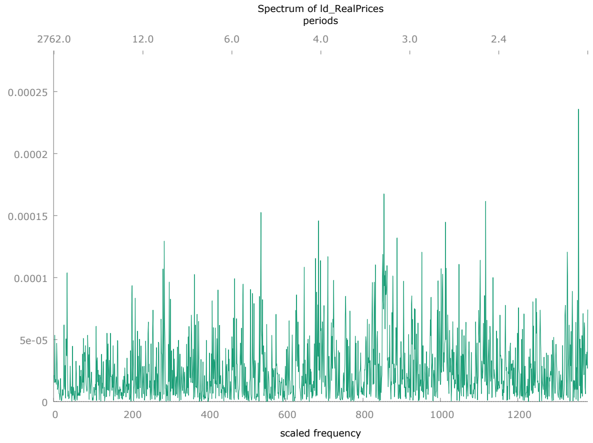


Figure 5. Plots of conditional variances

4. DISCUSSION AND CONCLUSION

The paper examines the use of GANs to generate a fake S&P500 Index series which in levels is indistinguishable from and cointegrated with the real series. It is only when the series is transformed into returns that higher-order lags of the two series behave differently. This is apparent from fitting GARCH models, periodograms, and simple tests of weak form market efficiency. These tests reveal differences in the behaviour of the two series at the higher order lags.

Goodfellow (2016) notes that there no single compelling ways to evaluate a generative model. Models with good likelihood can produce bad samples, samples themselves are hard to evaluate, and good samples can have bad likelihood. Hence the multiple metrics used in evaluation in this paper.



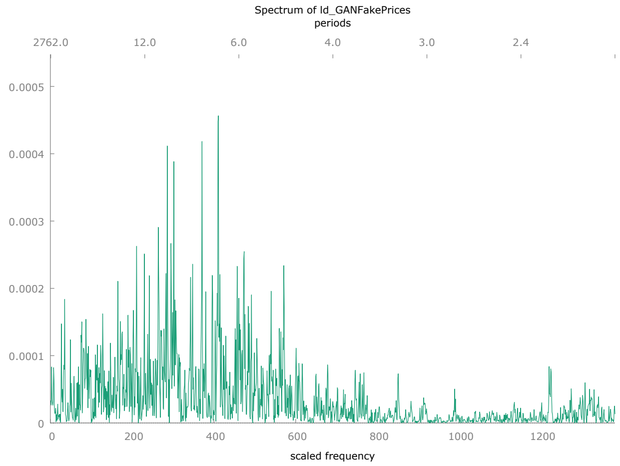


Figure 6. Periodograms

REFERENCES

Arjovsky, Martin; Chintala, Soumith; Bottou, Léon (2017-07-17). "Wasserstein Generative Adversarial Networks". International Conference on Machine Learning. PMLR: 214–223.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31 (1986) 307–327.

Engle, R. (1986). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50 (1982) 987–1007.

Engle, Robert F.; Granger, Clive W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*. 55 (2): 251–276.

Fama, E.F. (1965) The behaviour of Stock Market Prices, *Journal of Business*, 38(1): 34-105.

Goodfellow, I.J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014), Generative Adversarial Networks, arXiv:1406.2661 [stat.ML]

Goodfellow, I. (2016) NIPS Tutorial: Generative Adversarial Networks, arXiv:1701.00160

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena, Terrestrial Magnetism. 3 (1): 13 41. Bibcode: 1898TeMag...3...13S. doi:10.1029/TM003i001p00013