IEORE4742 | Deep Learning for Financial Engineering and Operations Research
Sauma Capital | Financial Time Series Generation with Frequency GANs

Pierre Counathe [1] Alexander Phillips [2]

[1]pc2977@columbia.edu [2]a.j.phillips@columbia.edu

Columbia University in the City of New York

Abstract

The generation of high quality time series data has been imperative since the advent of quantitative finance. This is especially true as models grow more abstract with more trainable parameters necessitating more training data which we do not have. Recent developments have been made in the form of temporal convolutional GANs with success. We propose a novel addition to these methods in the form of a GAN which generates data in the frequency domain rather than in the time domain. Frequency domain discriminators add value to the architectures as they help replicate the patterns in returns' autocorrelation. Frequency domain generators show some small promise as they struggle to replicate the ACF patterns of various futures time series, but the patterns in these failures show a clear path forward.

Data Exploration

Figs 2 and 1 are an example of the type of statistical feature we intend to replicate with our GAN: the auto-correlation function (ACF). The closer our GAN is to generating data that replicates the ACF of the target time series, the better.

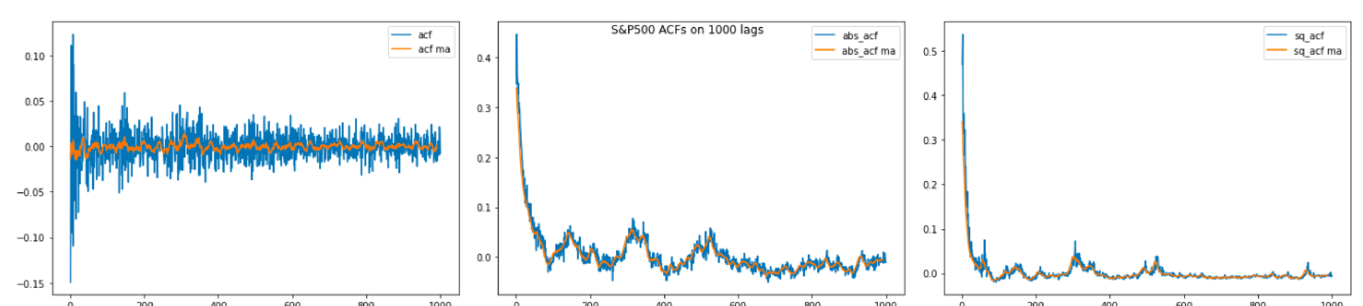


Figure 1. S&P500: ACF of log-returns, of absolute log-returns, and of squared log-returns

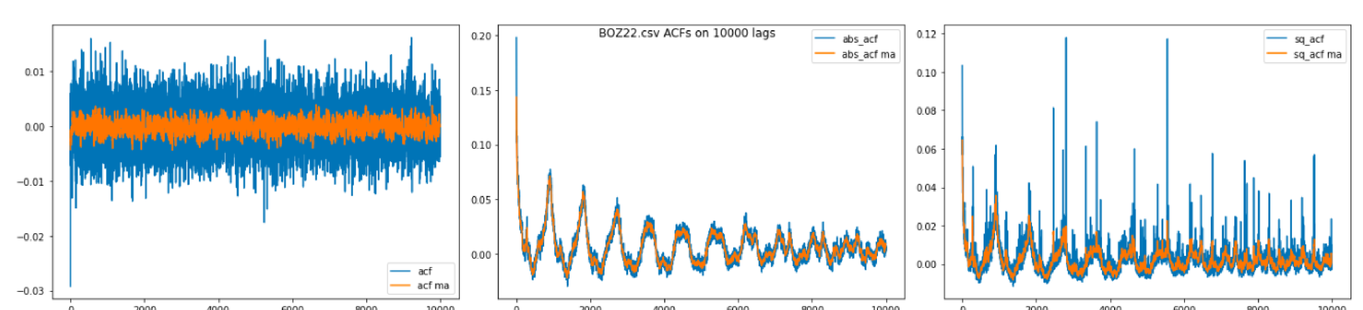


Figure 2. Features of Soybean Oil Futures 2022

Previous Work

Several GAN architectures have been explored to varying degrees of success including: Quant-GAN[5], TAGAN and TTGAN[2]. These models perform quite well on certain datasets such as S&P500 time series, but not on many others. Notably, all techniques tested as of yet seem to only approach time series data in the time domain, never in the frequency domain. For example, Fig 3 shows the performance of TAGAN on replicating returns' autocorrelations after tuning on Soy Bean Oil Futures.

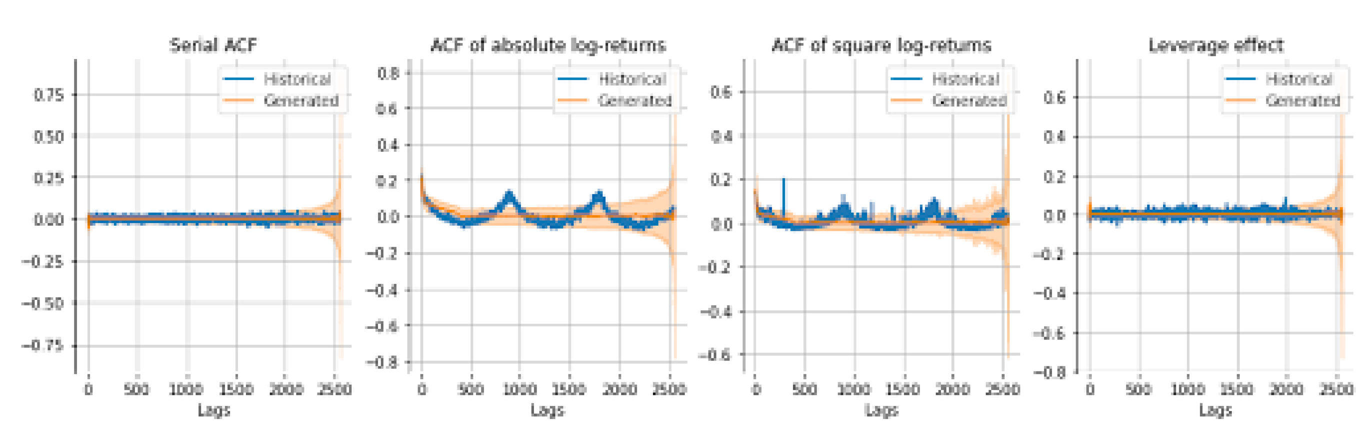


Figure 3. TAGAN Time Series ACF Evaluation on Soy Oil Futures Data

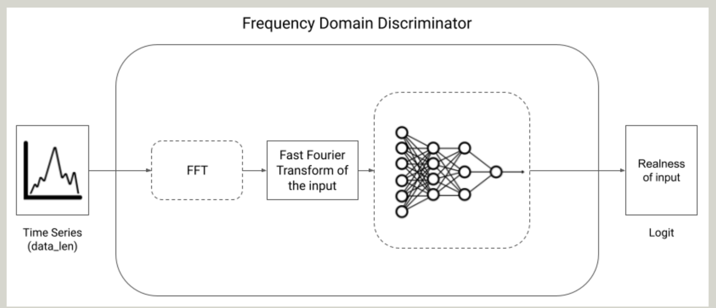
Frequency Domain GANs

Our contribution is to have some combination of the generator and discriminator do its job within the frequency domain rather than the time domain. Notably, this approach loses time-dependence in the crossover to the frequency domain and the generator is therefore only appropriate to generate snippets of a time series, never to continue a previous series. Figures 4b, 4a elaborate the architectures of the frequency domain generator and discriminator. A second discriminator in both domains is not depicted here. This discriminator is the result of the combination of the results of both the time and frequency domain discriminators in a dense layer.

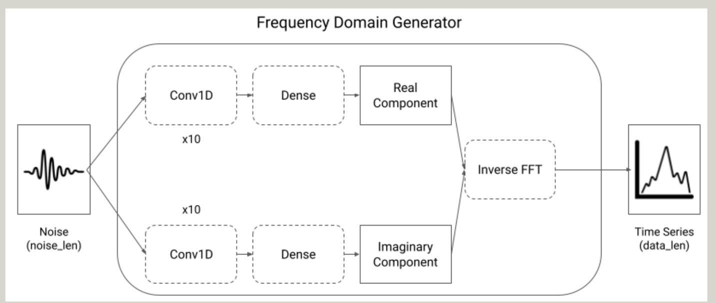
In cases where we test only one of either generator or discriminator within frequency domain, the other is from Quant-GAN. For example, when we test the frequency-time architecture, meaning the frequency domain generator4b and the Quant-GAN discriminator.

We test all of time-frequency, frequency-frequency, frequency-time, and time-both GAN architectures.

Model Architectures

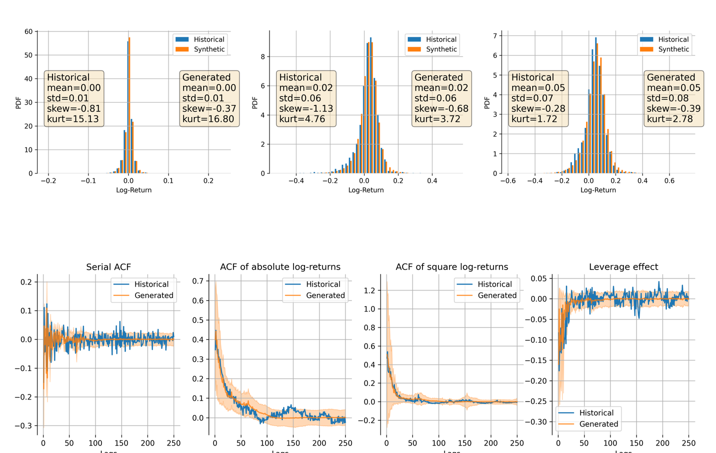


(a) Frequency Domain Discriminator

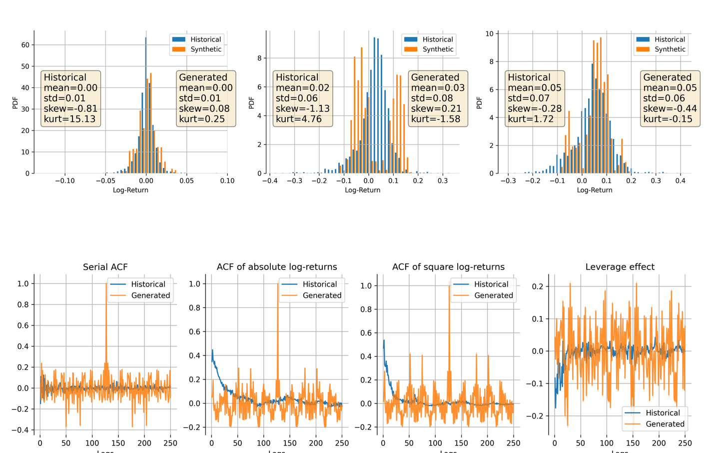


(b) Frequency Domain Generator

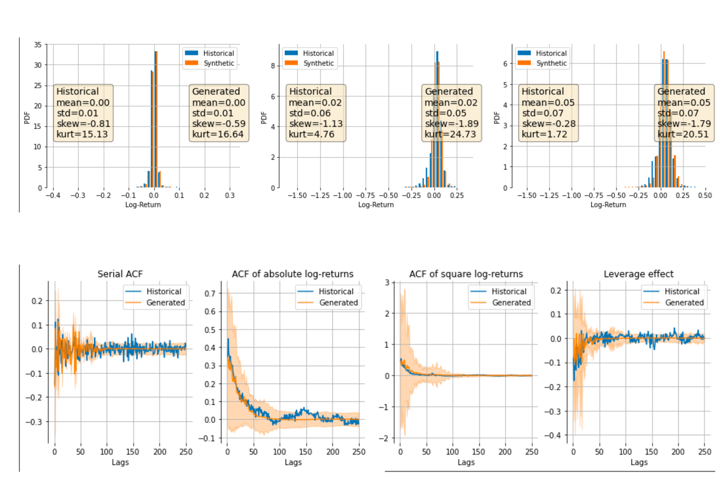
Results



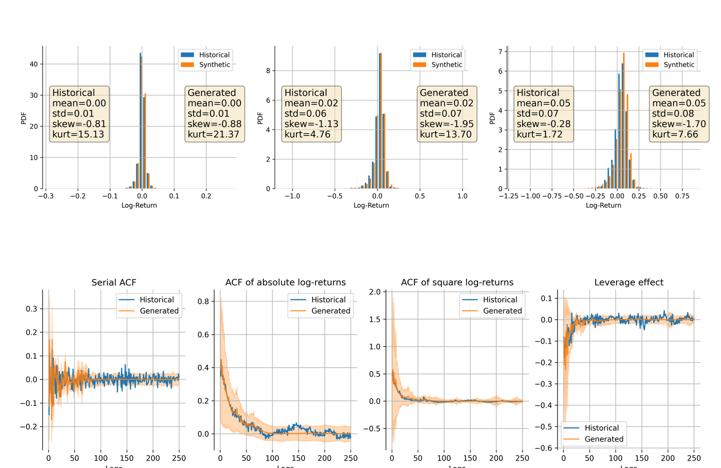
(a) Quantgan results



(b) *Frequency-Time* architecture results



(a) *Time-freq* architecture results



(b) *Time-Both* architecture results

Analysis

All results are in the same format. For every model, the seven plots in reading order, are: distributions of log-returns for 1, 50, and 100 days returns, for historical and GAN-generated data, serial ACFs, ACFs of absolute and squared log-returns, and leverage effect. A good model should replicate: historical mean, std. dev., skewness, kurtosis, ACF of absolute and square log-returns, leverage effect. Results:

- QuantGAN (Fig. 5a) reproduces well the log-returns distributions, especially their skewness and kurtosis, however, there is a slight bias in the ACF of log-returns, and it does not reproduce its cyclic pattern
- Frequency generator models (e.g. Fig. 5b)struggle to replicate log-returns distributions, with the *frequency-time* model producing bimodal distributions, different from a bell shape
- The *time-both* model (Fig. 6b) replicate the absolute and squared ACFs of log-return well, even better than what QuantGAN does. The *time-frequency* architecture struggles more with the leverage effect, but the *time-both* architecture manages to replicate it well

Overall, these results are very promising, and it seems like a work on the *time-both* architecture to balance the importance of time and frequency discriminators could be useful and provide great results.

Conclusion

Generating time series in the frequency domain is complex. Our model struggles to replicate positive ACFs of absolute and squared log-returns, and the nature of generating time series in the frequency domain implies spikes in the ACFs as well as an impossibility of generating time series with even a small trend.

The frequency domain discriminator is easy to combine with QuantGAN's generator. It improves the GAN's ability to replicate ACFs, but hinders its capability to replicate log-returns distributions. The discriminator in both the time and frequency domains seems promising as its frequency part helps replicate ACFs and its time part helps replicate log-returns distributions. The combination of both also helps reduce the ACF's variance over multiple generated time series.

References

[1] Yuanqi Chen, Ge Li, Cece Jin, Shan Liu, and Thomas Li. Ssd-gan: Measuring the realness in the spatial and spectral domains, 2020. URL `https://arxiv.org/abs/2012.05535`.

[2] Weilong Fu, Ali Hirsa, and Jörg Osterrieder. Simulating financial time series using attention. 2022. doi:10.48550/ARXIV.2207.00493. URL `https://arxiv.org/abs/2207.00493`.

[3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014. URL `https://arxiv.org/abs/1406.2661`.

[4] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio, 2016. URL `https://arxiv.org/abs/1609.03499`.

[5] Magnus Wiese, Robert Knobloch, Ralf Korn, and Peter Kretschmer. Quant GANs: deep generation of financial time series. *Quantitative Finance*, 20(9):1419–1440, apr 2020. doi:10.1080/14697688.2020.1730426. URL `https://doi.org/10.1080%2F14697688.2020.1730426`.