Article

Simulating Multi-Asset Classes Prices Using Wasserstein Generative Adversarial Network: A Study of Stocks, Futures and Cryptocurrency

Feng Han [1,†], **Xiaojuan Ma** [1,†] **and Jiheng Zhang** [2,*,†]

1   Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China; fhanab@cse.ust.hk (F.H.); mxj@cse.ust.hk (X.M.)
2   Department of Industrial Engineering and Decision Analytics and Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong, China
*   Correspondence: jiheng@ust.hk
†   Current address: Clear Water Bay, Sai Kung, New Territories, Hong Kong, China.

Abstract: Financial data are expensive and highly sensitive with limited access. We aim to generate abundant datasets given the original prices while preserving the original statistical features. We introduce the Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) into the field of the stock market, futures market and cryptocurrency market. We train our model on various datasets, including the Hong Kong stock market, Hang Seng Index Composite stocks, precious metal futures contracts listed on the Chicago Mercantile Exchange and Japan Exchange Group, and cryptocurrency spots and perpetual contracts on Binance at various minute-level intervals. We quantify the difference of generated results (836,280 data points) and original data by MAE, MSE, RMSE and K-S distances. Results show that WGAN-GP can simulate assets prices and show the potential of a market simulator for trading analysis. We might be the first to look into multi-asset classes in a systematic approach with minute intervals across stocks, futures and cryptocurrency markets. We also contribute to quantitative analysis methodology for generated and original price data quality.

Keywords: multi-asset classes; financial engineering; simulations; stocks; futures; cryptocurrency; precious metal futures; machine learning; financial technology

1. Introduction

Various scholars are trying to solve price prediction problems to capture short-term alphas in the markets (Ariyo et al. (2014); Foster (2002); Abraham et al. (2018)). However, a more fundamental contribution can be understanding the underlying price behavior of multiple asset classes with limited data. One of the critical processes is to generate enough data for observations and backtesting, therefore we want to learn the distribution of asset prices based on limited price information. We hope to generate richer varieties of data to simulate the original prices while preserving the original statistical features. Stock prices are considered to be random walks (and crypto as well) by Palamalai et al. (2021). To simulate the price behaviors, we adopt WGAN-GP to generate richer data with noise to discover hidden characteristics. To ensure the quality of generated real and fake data discriminated by WGAN-GP, we use MAE, MSE, RMSE and KS distance to calculate the price differences with multiple minute level intervals.

The Generative Adversarial Network(GAN) Goodfellow et al. (2014)'s success in generating realistic synthetic images has inspired machine learning for generating multi-asset class prices. The GAN models in image generations could be used in financial datasets. Inspired by the generator and discriminator ideas, we implement a Wasserstein GAN with Gradient Penalty (WGAN-GP) framework to learn the underlying original financial datasets distributions, given minute level intervals prices across stocks, futures and

cryptocurrencies markets. We explore a universal WGAN-GP model to simulate multi-asset classes datasets purchased from the Hong Kong Stock Exchanges, the Chicago Mercantile Exchange and the Japan Exchange Group. We report observations from stocks, futures and crypto markets, and evaluate generated prices against original datasets. The results showed good performance, and we can use WGAN-GP as a probabilistic prediction model on financial time series to obtain the potential distribution of original financial datasets.

Since financial datasets contain potentially sensitive information, datasets are always headaches for quantitative researchers for predictive purposes when backtesting algorithm trading strategies ranging from multiple time intervals like intraday and high frequency trading. Furthermore, based on abundant prices, reducing algorithm overfitting and enhancing the robustness of portfolios when markets are volatile are desired among professional users. Moreover, stock datasets have unique features like 'tick size' and trading mechanisms for each market including auctions and continuous trading sessions, while futures contracts have day and night sessions.

Here, we propose a practical model focusing on minute-level prices and quantifying the quality against original prices. Li et al. (2020) obtained the training tick data from OneMarketData as a financial data provider. Samuel et al. (2021) trained the WGAN-GP for daily prices downloaded from Yahoo Finance (n.d.)[1], but generated data are distinguishable. Their WGAN-GP consists of a generator and discriminator function which utilize an LSTM architecture. However, the model has space for optimization, which could be achieved by adjusting the hyperparameters. One of the solutions could be the learning rates. Samuel used an RMSprop optimizer with a fixed learning rate of 0.00005. Smith (2017) proposed training with cyclical learning rates instead of fixed values, which achieves improved classification accuracy. WGAN-GP can be trained to customize conditions through parameter settings like cyclic learning rates. To ensure the input data quality, we purchased all limit order books data including trades and order books from HKEX for stocks and CME and JPX for precious metals. We downloaded cryptos spots and perpetual futures from Binance, one of the biggest exchanges. To the best of authors' knowledge, we might be the first to look into multi-asset classes in a systematic approach with minute-level intervals observations with WGAN as in Figure 1.

We propose the following research questions:

- Can WGAN-GP be trained as a probabilistic model for financial prices simulations?
- Can WGAN-GP simulate multiple minute-level intervals prices?
- How good are the WGAN-GP when simulating multiple asset classes, e.g., stocks, futures and cryptos?
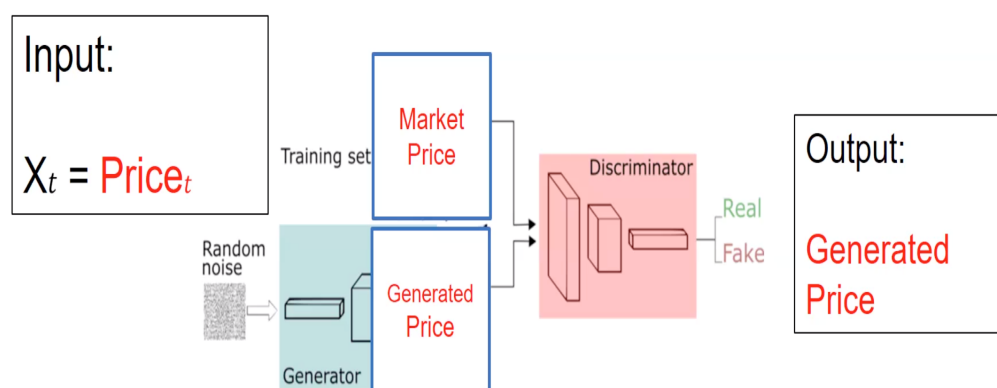
Figure 1. Given the original market prices as input, we trained the datasets using WGAN-GP and output the generated real and fake prices judged by the discriminators. The generator struggles to trick the Discriminator, while the Discriminator criticizes the generator until approaching equilibrium, keeping the training process as a zero-sum game in Algorithm 1.

The contents are organized as follows. Section 2 describes the model GAN, WGAN and WGAN-GP with their features. Section 3 describes our proposed model based on WGAN-GP and multiple asset classes datasets. Section 4 includes the results for training the Hong Kong stocks. Section 5 reports the performance of precious metal futures including gold CME Gold Futures and Option (n.d.), JPX Gold Futures and Options (n.d.) and platinum CME Platinum Futures and Options (n.d.), JPX Platinum Futures and Options (n.d.) futures products listed on CME and JPX. Section 6 focuses on training cryptocurrencies including spot and perpetual futures contracts. Section 7 discusses the results and provides implications for future research. Finally, we conclude our model performance and give insights about multi-asset classes simulations in Section 8.

Algorithm 1: WGAN-GP algorithm

> **for** i in $range(epochs)$ **do**
> > **for** j in $range(n_{critic})$ **do**
> > > sample $X_{\text{original}} \sim P_{\text{data}}$;
> > > sample $X_{noise} \sim P_{noise}$;
> > > $L_{\text{original}} = Critic(X_{\text{original}})$;
> > > $L_{\text{fake}} = Critic(Generator(X_{noise}))$;
> > > $GP = \text{gradient\_penalty}\left(L_{\text{original}}, L_{\text{fake}}\right)$;
> > > $loss_c = \frac{1}{bs}\sum_{i=1}^{bs} L_{fake_{(i)}} - \frac{1}{bs}\sum_{i=1}^{bs} L_{original_{(i)}} + GP$
> > > $\alpha_d = \text{CyclicLR}(epoch, \text{lr\_min}_c, \text{lr\_max}_c)$;
> > > $Weight_{critic} = Weight_{critic} + \alpha_d \cdot \text{RMSprop}(loss_c, Weight_{critic})$;
> > **end for**
> > sample batch noise again;
> > $L_{\text{fake}} = Critic(Generator(X_{noise}))$;
> > $loss_g = \frac{1}{bs}\left(\sum_{i=1}^{bs} L_{\text{fake}_{(i)}}\right)$
> > $\alpha_g = \text{CyclicLR}(epoch, base\_lr_g, lr\_max_g)$;
> > $Weight_{generator} = Weight_{generator}$
> > $-\alpha_{generator} \cdot \text{RMSprop}(loss_{generator}, Weight_{generator})$;
> **end for**

Algorithm 1 WGAN-GP trains each asset class with customized parameters, and we systematically examine the difference between generated data prices and original prices by MAE, MSE, RMSE and KS test in Section 4 for stocks markets, Section 5 for precious metal futures listed on CME and JPX and Section 6 for cryptos listed on Binance including spots and perpetual futures.

2. Generative Adversarial Networks

Generative adversarial networks are unsupervised learning algorithms that build two neural networks competing in a zero-sum game. The ultimate goal of GAN is to generate $P_{generated}$ close to original distribution $P_{original}$. Using GANs on images is easy and intuitive, however hard for humans to examine large-scale datasets.

2.1. GAN Structure

Generator and discriminator are the two models in a game playing against each other during the training. The generator will output an image initiated from random noise and pass it to the discriminator as a binary classifier assigning generated real or fake labels. The generator struggles to adjust according to the feedback and continues to fool the discriminator. Meanwhile, the discriminator continues to refine the decision making strategies until it cannot differentiate real and fake ones. The optimal states of the discriminator are random guesses with 0.5 accuracies.

2.2. GAN Training

The GAN has iterative gradient descent for both players. The Loss function is obtained by taking a batch of real (from P_{real}) and generated samples (from $= P_{generated}$). ADAM or RMSprop optimizer is widely used for increasing training convergence. The overall value function Value (*G*, *D*), which incorporates both Loss(*D*) and Loss(*G*), is defined as:

$$\min_{G} \max_{D} Value(D, G) = \mathbb{E}_{x \sim p_{\text{original}}}\left[\log D(x)\right] + \mathbb{E}_{noise \sim p_{noise}}\left[\log(1 - D(G(noise)))\right] \quad (1)$$

Both players can be lazy from updating, leading to mode collapse, ignoring all other modes and staying in the comfort zones with few states. The generator can only create a few generated data when minimizing the loss while Discriminator remains passive feedback.

2.3. Wasserstein GAN

Wasserstein GAN solves the mode collapse problems by subtle changes on the models with earthmoving distance by Arjovsky et al. (2017). However, weight clipping is hard to enforce a Lipschitz constraint to stabilize the training while keeping diversified data varieties.

$$Wasserstein\left(P_{original}, P_{generated}\right) = \sup_{\|f\|_{Lipschitz} \leq 1} \mathbb{E}_{x \sim P_{original}}[f(x)] - \mathbb{E}_{x \sim P_{generated}}[f(x)] \quad (2)$$

2.4. Wasserstein GAN with Gradient Penalty

A derivable function is 1-Lipschtiz if and only if it has gradients of norm at most 1 everywhere. Therefore, the gradient penalty is used to constrain the gradient norm of the critic's output to be 1 everywhere in regard to its input. The gradient penalty by Gulrajani et al. (2017) is added to the objective:

$$\begin{aligned} Loss\left(P_{original}, P_{generated}\right) = &\mathbb{E}_{x \sim P_{original}}[f_w(x)] - \mathbb{E}_{noise \sim P_{generated}}[f_w(G_\theta(noise))] + \\ &\lambda \mathbb{E}_{x \sim P_{original}}\left[\left(\|\nabla_x f_w(x)\|_2 - 1\right)^2\right] \end{aligned} \quad (3)$$

Our model is based on WGAN-GP in Section 3 and customized for the following asset classes including stocks, futures, and cryptocurrencies demonstrated in Sections 4–6.

3. Our Proposed Model and Asset Classes Datasets Descriptions for Training

We initialize $bs = 64$, $n_{critic} = 5$, $lr\_min_c = 3 \times 10^4$, $lr\_max_c = 8 \times 10^4$, $lr\_min_g = 1 \times 10^4$, $lr\_max_g = 1 \times 10^3$. We used an AMD EPYC 7H12 64-Core Processor 2.6 GHz, with 1 TB RAM, and 4 Nvidia A100-SXM4-40 GB GPUs for the training tasks, since each dataset is pretty big with all limit order books information, the powerful server can better model the training performance. The training takes around 200 h to complete, and generate a total of 836,280 data points, 4140 data points for each row, respectively. For stocks we generated 372,600 points for HSI Composite, 115,920 data points including TGD, TPL, PLE and GCE for precious metal futures, 202,860 for cryptos spots and 144,900 for crypto perpetual futures. The parameters for training are outlined in Table 1 and multiple asset classes datasets in Table 2.

For the Hong Kong market, Kim and Mei (2001) showed that political developments in Hong Kong have a significant impact on its market volatility and return. The extended ARCH-jump filter with bad news has a more significant volatility effect than good news. The unique characteristics can be price jumps that misprice derivative products tracking benchmark HSI indices, leading to a dynamic hedging portfolio less effective with jumps caused by policy risks. Xu et al. (2020) argued market liberalization leads to lower quoted spread, lower effective spread, lower market depth, and higher short-term volatility for the Shanghai–Hong Kong Stock Connect program (SHHKConnect). Meyer and Guernsey (2017) discussed how high frequency trading on the HKEx compared to the SGX may derive

from an underlying conflicted approach within Hong Kong's political-economy about how HKEx relates to China's exchanges. On this occasion, we might be the first to look into the Hang Seng Index Component Stocks, a total of 60 stocks. We introduce WGAN into Hong Kong stocks simulations to better understand Hong Kong markets.

Table 1. Parameters settings for WGAN-GP, a combination of different parameter settings, can approach the optimal loss efficiently. We outlined the parameters for training considerations.

bs	Batch Sizes
lr_min	lower bound on learning rate
lr_max	higher bound on learning rate
epochs	number of training iterations
noise	probability to add instance noise
C	critic with parameters $\mathcal{W}_c$
G	generator with parameters $\mathcal{W}_g$
$P_{original}$	distribution of training price data
P_{noise}	distribution of sample noise

Table 2. We systematically examined the multiple asset classes, including shares (HSI Composite Stocks), futures (Gold, Platinum Futures on CME and JPX) and cryptos (Spot, Perpetual contracts). The data was purchased directly from the exchanges and was of the highest quality.

Asset Classes	Instrument	Symbol
Shares/ Stocks	HANG SENG INDEX Component Stocks	xxxx.HK
Futures	Gold, Platinum Futures on CME and JPX	GCE, PLE, TPL, TGD
Cryptos	Spots, Perpetual contracts	Various products

For the futures market, Kang et al. (2017) evaluated and compared optimal portfolio weights and time-varying hedge ratios from 4 January 2002 to 28 July 2016. They showed directional spillovers (DS) could transmit from one market to another. Xu and Fung (2005) indicated that pricing transmissions for these precious metals contracts are strong across the two markets. Still, information flows appear to lead from the U.S. market to the Japanese market in terms of returns, and the offshore trading information can be absorbed in 24 h. Mensi et al. (2021) suggested that gold and silver are net contributors of risk to the other markets, whereas palladium is a net receiver of risk for all the time horizons. Platinum is a net contributor of risk to the other markets in the short term and a net receiver of risk from the remaining markets in the intermediate- and long-term horizons. Precious metals provide diversification gains to currency investors for all time horizons.

For the cryptocurrency market, Lee et al. (2020) studied BTC data from January 2018 contracts to March 2019 on all CBOE and CME. They concluded that Bitcoin spot and futures are cointegrated, and Bitcoin futures are biased predictors of spot prices. Kyriazis et al. (2019) studied bearish market daily data from 1 January 2018 to 16 September 2018. They concluded that the highest capitalization digital currencies, namely Bitcoin, Ethereum and Ripple, will influence others. Baur and Hoang (2021) concluded stablecoins are relatively safe compared with BTC. We benchmarked BTC as the highest market capitalization asset from the above literature.

4. HANG SENG INDEX Components Stocks Simulations

The Hang Seng Index (HSI) is the main stock market index in Hong Kong to monitor the daily changes of the largest companies listed in the Hong Kong stock market. HSI is the

leading indicator of the overall market performance in Hong Kong. These 60 constituent companies represent about half of the capitalisation of the Hong Kong Stock Exchange.

4.1. Data Cleaning for HKEX Datasets

The complete limit order book and trade book data were purchased from HKEX. After deserializing from binary files into MongoDB to reconstruct simulations, we recorded the stock code and order book modification, including cancellation and trade orders in milliseconds. To generate an order book snapshot, we resampled by consecutive trading minutes to obtain Open, Close, High and Low prices. Finally, we obtained the minute-level original Close prices.

### 4.2. Input and Output Prices for the WGAN-GP Model

Here, we introduced the input prices of original minute close in Figure 2 for WuXi Biologics 2269.HK as an example and output as generated minute close prices in Figure 3, followed by density plot by price distributions in Figure 4 for 2269.HK.
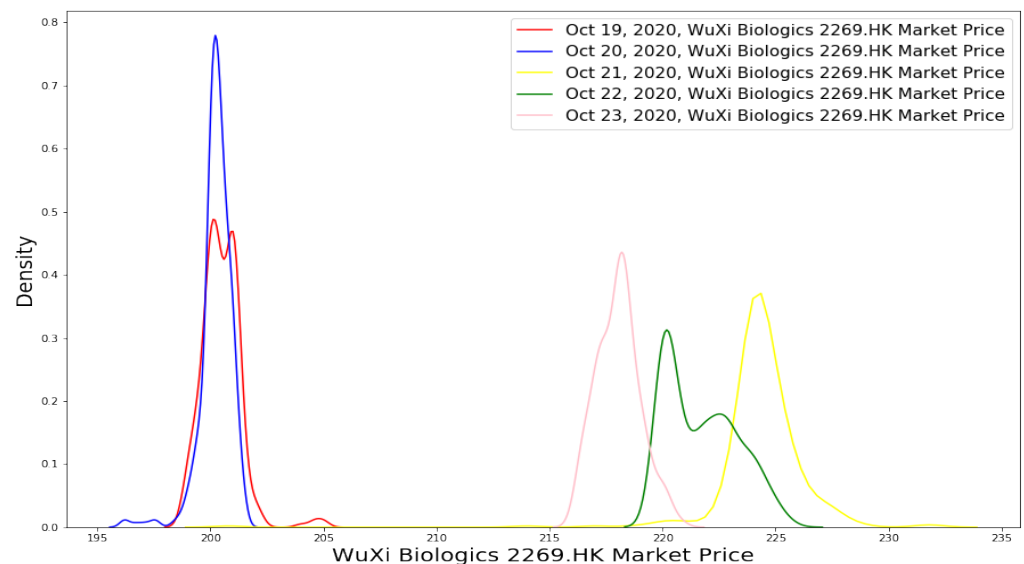
Figure 2. WuXi Biologics is a listed company on the Hong Kong Stock Exchange, one of HSI Composites. For a consecutive trading week, the prices at minute level intervals are plotted in six distributions (the total area under the curve integrates to one with each date as one distribution). The probability density (Y-axis) is the per unit on the prices (X-axis). There exist statistics properties like kurtosis and skewness, mean and standard deviations.

Given the price distributions at minute intervals for 2269.HK, we started from 19 October 2020, and fed the minute interval prices into the WGAN-GP model. The generator tries to fool the discriminator by generating as many data points as possible. In contrast, the discriminator struggles to judge by labelling real and fake compared to the original data.

After obtaining the generated real data, the two curves can be plotted in the same graph (the X-axis is the price). The simulated real price curve is twisted around the original minute level close. We need to quantify the difference between two prices by MAE, MSE, RMSE and K-S distances Table 3. As shown in Tables 3 and 4, we calculated and generated 19–23 October 2020, five trading days with MAE-R, MSE-R and RMSE-R, where R denotes Generated Real, meaning the differences against original input minute close prices. To further ensure the quality of generated real and fake data, we investigate the KS test with Generated Real and Fake compared with original input prices. Usually, the smaller KS, the better results. The KS R/F ratio means the difference between KS Real and KS Fake. Similarly, the smaller, the better.
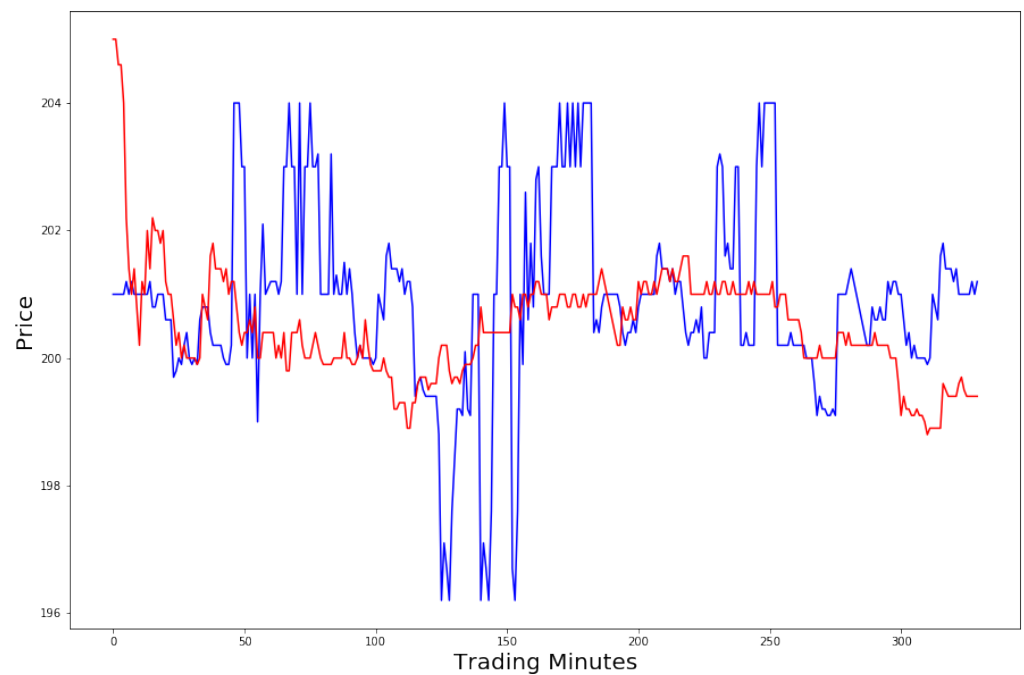
Figure 3. Generated real price at the minute interval (blue) and original market close prices (red) are plotted for 2269.HK. Starting from the first minute, the auction period will impact the following several trading minutes. The generated real prices are pretty close to the original input, as shown in tick movement.
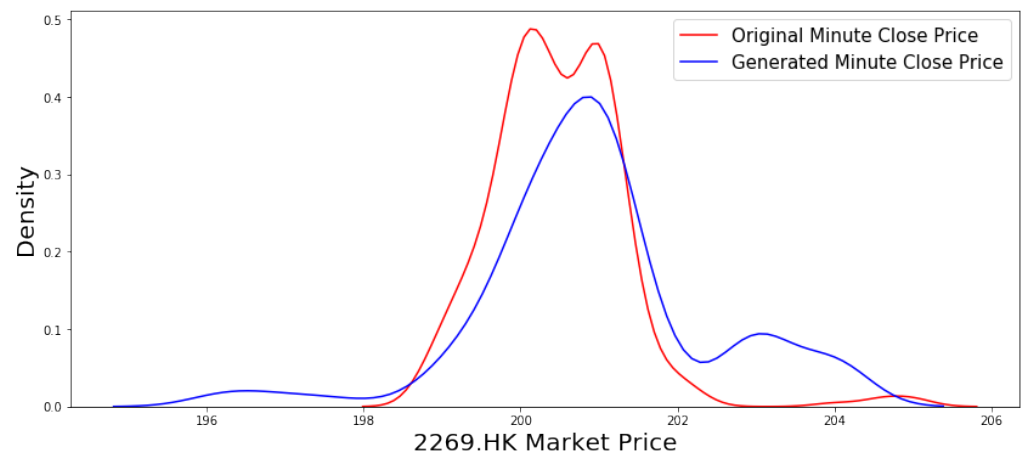
Figure 4. After mapping the price information by price, the generated prices (blue) are compared with one-minute market close prices (red).

Table 3. We listed the performance of our model by MAE-R, MSE-R, RMSE-R, and KS tests. HSI Composite stocks are grouped by tick.

HSI Composite 1 min	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
0003.HK	0.036	0.007	0.085	0.100	0.130	0.769	0.2	0.18
2382.HK	1.878	4.997	2.235	0.047	0.039	1.205	0.2	9.39
0388.HK	2.272	8.878	2.980	0.018	0.041	0.439	0.2	11.36
0700.HK	3.231	16.830	4.102	0.091	0.082	1.110	0.2	16.16
1211.HK	**6.616**	**69.563**	**8.340**	**0.036**	**0.164**	**0.220**	**0.2**	**33.08**

Table 3. *Cont.*

HSI Composite 1 min	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
0669.HK	1.099	1.947	1.395	0.023	0.043	0.535	0.1	10.99
0011.HK	1.610	5.463	2.337	0.056	0.067	0.836	0.1	16.10
2020.HK	1.706	4.520	2.126	0.038	0.050	0.760	0.1	17.06
9988.HK	1.761	4.710	2.170	0.020	0.055	0.364	0.1	17.61
2313.HK	**2.912**	**13.912**	**3.730**	**0.063**	**0.084**	**0.750**	**0.1**	**29.12**
0012.HK	0.068	0.008	0.091	0.042	0.028	1.500	0.05	1.36
1038.HK	0.108	0.018	0.136	0.029	0.072	0.403	0.05	2.16
0175.HK	0.129	0.029	0.169	0.041	0.027	**1.519**	0.05	2.58
1997.HK	0.130	0.027	0.164	0.033	0.045	0.733	0.05	2.60
0017.HK	0.146	0.036	0.189	0.026	0.089	0.292	0.05	2.92
0066.HK	0.173	0.048	0.219	0.033	0.055	0.600	0.05	3.46
2388.HK	0.190	0.068	0.260	0.035	0.037	0.946	0.05	3.80
1113.HK	0.205	0.068	0.261	**0.122**	0.117	1.043	0.05	4.10
0001.HK	0.206	0.077	0.277	0.025	0.057	0.439	0.05	4.12
0868.HK	0.224	0.091	0.301	0.022	0.055	0.400	0.05	4.48
0823.HK	0.253	0.102	0.320	0.061	0.027	2.259	0.05	5.06
1810.HK	0.279	0.139	0.373	0.053	0.086	0.616	0.05	5.58
1299.HK	0.288	0.134	0.366	0.051	0.053	0.962	0.05	5.76
0006.HK	0.334	0.173	0.416	0.041	0.043	0.953	0.05	6.68
0960.HK	0.352	0.182	0.426	0.032	0.036	0.889	0.05	7.04
1876.HK	0.372	0.227	0.476	0.054	0.047	1.149	0.05	7.44
0002.HK	0.434	0.336	0.580	0.051	0.094	0.543	0.05	8.68
2318.HK	0.459	0.407	0.638	0.020	0.077	0.260	0.05	9.18
0016.HK	0.494	0.383	0.619	0.037	0.028	1.321	0.05	9.88
6862.HK	0.498	0.432	0.657	0.042	0.048	0.875	0.05	9.96
0941.HK	0.630	0.707	0.841	0.062	0.062	1.000	0.05	12.6
1044.HK	0.673	0.807	0.898	0.032	0.056	0.571	0.05	13.46
1109.HK	0.711	0.792	0.890	0.037	0.054	0.685	0.05	14.22
2018.HK	0.806	1.033	1.016	0.047	0.037	1.270	0.05	16.12
2319.HK	0.979	1.725	1.313	0.056	0.044	1.273	0.05	19.58
3968.HK	1.026	1.674	1.294	0.037	0.043	0.860	0.05	20.52
6098.HK	1.063	1.680	1.296	0.050	0.053	0.943	0.05	21.26
0027.HK	1.209	2.432	1.560	0.032	0.057	0.561	0.05	24.18
2331.HK	**1.328**	**2.857**	**1.690**	**0.060**	**0.044**	**1.364**	**0.05**	**26.56**
2628.HK	0.190	0.060	0.244	0.049	0.054	0.907	0.02	9.5
0101.HK	0.220	0.078	0.279	0.076	0.085	0.894	0.02	11
0968.HK	0.332	0.166	0.408	0.041	0.049	0.837	0.02	16.6
1928.HK	**0.754**	**1.014**	**1.001**	**0.090**	**0.085**	**1.059**	**0.02**	**37.7**
0857.HK	0.028	0.001	0.035	0.046	0.033	1.394	0.01	2.8
3988.HK	0.040	0.002	0.049	0.047	0.059	0.797	0.01	4.0
0386.HK	0.049	0.005	0.073	0.028	0.043	0.651	0.01	4.9
0267.HK	0.058	0.006	0.075	0.031	0.028	1.107	0.01	5.8
1177.HK	0.069	0.008	0.090	0.074	0.076	0.974	0.01	6.9
0288.HK	0.085	0.012	0.110	0.035	0.052	0.673	0.01	8.5
0939.HK	0.087	0.012	0.112	0.041	0.044	0.932	0.01	8.7
0883.HK	0.122	0.026	0.161	0.059	0.066	0.894	0.01	12.2
0241.HK	0.140	0.033	0.182	0.110	**0.110**	1.000	0.01	14.0
2007.HK	0.141	0.031	0.175	0.046	0.045	1.022	0.01	14.1
1093.HK	0.156	0.040	0.200	0.039	0.038	1.026	0.01	15.6
0762.HK	**0.238**	**0.117**	**0.342**	**0.042**	**0.065**	**0.646**	**0.01**	**23.8**

Table 4. We selected several typical stocks at different tick values and reported the performance of our model at one, two, three, four, five, ten and fifteen minute intervals. Tick values include 0.2, 0.1, 0.02, 0.05 and 0.01. 2269.HK-15 min in Table 4 shows highest MAE-R/ Tick in all minute intervals compared with other selected representatives by ticks.

HSI-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
3690.HK-1	4.597	33.238	5.765	0.049	0.059	0.830	0.2	22.985
3690.HK-2	4.333	30.169	5.493	0.044	0.035	**1.257**	0.2	21.665
3690.HK-3	4.872	36.853	6.071	0.031	0.062	0.500	0.2	24.36
3690.HK-4	4.602	33.661	5.802	0.026	0.059	0.441	0.2	23.01
3690.HK-5	**4.943**	**37.865**	**6.153**	**0.021**	**0.079**	**0.266**	**0.2**	**24.715**
3690.HK-10	4.823	36.658	6.055	**0.094**	0.103	0.913	0.2	24.115
3690.HK-15	6.296	53.370	7.305	0.078	**0.118**	0.661	0.2	31.48
2269.HK-1	10.038	192.79	13.885	0.045	0.084	0.536	0.1	100.38
2269.HK-2	**13.358**	**279.71**	**16.724**	**0.046**	**0.059**	**0.780**	**0.1**	**133.58**
2269.HK-3	11.428	228.261	15.108	0.065	0.080	0.813	0.1	114.28
2269.HK-4	12.908	265.898	16.306	0.030	0.068	0.441	0.1	129.08
2269.HK-5	12.842	270.445	16.445	0.036	0.125	0.288	0.1	128.42
2269.HK-10	11.222	238.630	15.448	**0.075**	0.092	**0.815**	0.1	112.22
2269.HK-15	11.686	256.310	16.010	0.073	**0.129**	0.566	0.1	116.86
0005.HK-1	**0.486**	**0.521**	**0.722**	**0.047**	**0.103**	**0.456**	**0.05**	**9.72**
0005.HK-2	0.386	0.302	0.550	0.071	0.031	**2.290**	0.05	7.72
0005.HK-3	0.394	0.362	0.602	0.053	0.048	1.104	0.05	7.88
0005.HK-4	0.412	0.315	0.561	0.048	0.054	0.889	0.05	8.24
0005.HK-5	0.370	0.269	0.519	0.051	0.058	0.879	0.05	7.4
0005.HK-10	0.325	0.208	0.456	**0.104**	0.087	1.195	0.05	6.5
0005.HK-15	0.307	0.170	0.412	0.058	**0.094**	0.617	0.05	6.14
0688.HK-1	0.138	0.031	0.175	0.057	0.045	**1.267**	0.02	6.9
0688.HK-2	**0.148**	**0.035**	**0.186**	**0.069**	**0.107**	**0.645**	**0.02**	**7.4**
0688.HK-3	0.146	0.036	0.190	0.018	0.103	0.175	0.02	7.3
0688.HK-4	0.138	0.033	0.181	0.073	0.066	1.106	0.02	6.9
0688.HK-5	**0.154**	**0.038**	**0.194**	**0.075**	**0.087**	**0.862**	**0.02**	**7.7**
0688.HK-10	0.136	0.031	0.177	0.065	0.100	0.650	0.02	6.8
0688.HK-15	0.140	0.035	0.186	**0.146**	**0.146**	1.000	0.02	7.0
1398.HK-1	**0.080**	**0.011**	**0.106**	**0.052**	**0.053**	**0.981**	**0.01**	**8.0**
1398.HK-2	0.077	0.010	0.101	0.043	0.052	0.827	0.01	7.7
1398.HK-3	0.078	0.010	0.102	0.054	0.078	0.692	0.01	7.8
1398.HK-4	**0.080**	**0.011**	**0.104**	**0.084**	**0.101**	**0.832**	**0.01**	**8.0**
1398.HK-5	0.071	0.009	0.095	0.073	0.117	0.624	0.01	7.1
1398.HK-10	0.067	0.008	0.089	0.105	0.114	0.921	0.01	6.7
1398.HK-15	0.069	0.008	0.089	**0.129**	**0.133**	**0.970**	0.01	6.9

Hong Kong securities market includes 1. Auction Session Pre-opening Session 9:00 a.m.–9:30 a.m. 2. Continuous Trading Session from 9:30 a.m. to 4:00 p.m. with an hour break in the noon Hong Kong Exchanges and Clearing Limited (n.d.). HSBC continuous trading minutes include pre-opening session and continuous trading session with an closing auction every market open day shown in Figure 5.
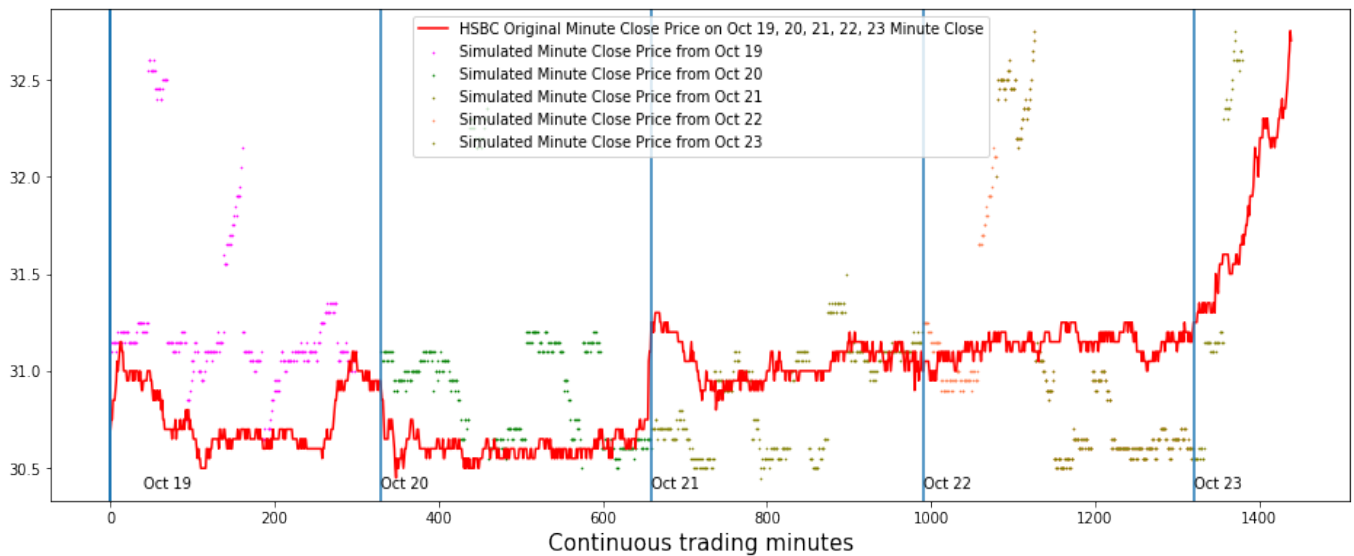
Figure 5. For the consecutive trading minutes from 19 October to 23 October 2021, we plotted the minutes trading level (X-axis) and minute interval close prices (color red), with the simulated/generated prices in other colours. The open market auction impacts the morning continuous trading session after 9:30 a.m. And starting from 3:30 p.m., there exists market impact caused by China Connect until market close.

4.3. Quantitative Results by MAE, MSE, RMSE, K-S Tests, KS R/F Ratio and MAE-R/ Tick

We measured the difference by mean absolute error (*MAE*), mean squared error (*MSE*), and root-mean-square deviation (RMSE). Due to the page limit, we only listed Generated Real data with original market close data. We calculated the Kolmogorov-Smirnov test (KS-test) between Generated Real and Generated Fake to quantify the distribution difference with original market close price data.

MAE represents the difference between the real and fake values extracted by averaging the absolute difference over the datasets. *MSE* represents the difference between the real and fake values extracted by squaring the average difference over the dataset.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |(y_i - \hat{y}_i)|$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

where y_i denotes real values and $\hat{y}_i$ denotes fake values.

The KS-test is a kind of "goodness-of-fit test". Suppose that we have an i.i.d. sample $X_1, ..., X_n$ with some unknown distribution $\mathbb{P}$ and we want to test the hypothesis that $\mathbb{P}$ is equal to a particular distribution $\mathbb{P}_0$, i.e., decide between the following hypotheses:

$$H_0 : \mathbb{P} = \mathbb{P}_0 \Leftrightarrow H_1 : \mathbb{P} \neq \mathbb{P}_0$$

The KS-test tries to determine if two datasets differ significantly. The KS-test has the advantage of not assuming the distribution of data Li et al. (2020). Li et al. proposed model Stock-GAN outperformed against recurrent conditional variational auto-encoder (VAE) and DCGAN instead of WGAN. However, they only tested the model on two stocks: a large capitalization stock, Alphabet Inc (GOOG), on one trading day in August 2017 and a small-capitalization stock, Patriot National (PN), which has relatively poor performance. The Stock-GAN model KS distance is 0.126 for GOOG, which is better than VAE (0.218) and DCGAN (0.181).

4.4. Discussion for HANG SENG INDEX Components Stocks Markets

After rounding by tick size, the MAE-R/Tick can be referred as part of volatility. In Table 3, 1211.HK (Tick value = 0.2, MAE-R/Tick = 33.08), 2313.HK (Tick value = 0.1, MAE-R/Tick = 29.12), 2331.HK (Tick value = 0.05, MAE-R/Tick = 26.56), 1928.HK (Tick value = 0.02, MAE-R/Tick = 37.7), 0762.HK (Tick value = 0.01, MAE-R/Tick = 23.8) shows higher volatility in one minute interval. The highest KS REAL and KS FAKE are 0.122, 0.110 and the highest KS R/F Ratio is 1.519. Overall, the quantitative results show the WGAN-GP generated data for prices are of high quality.

5. CME and JPX Precious Metal Futures Market Simulations

The Chicago Mercantile Exchange and the Japan Exchange Group are among the most significant futures exchanges globally. Precious metals futures are the most liquid products. Gold, silver, platinum and palladium are precious metal commodities with a wide range of industrial usage. The precious metal also retains a significant role as a relatively stable investment instrument by the private firms, governments, LBMA and central banks. Precious metal futures are highly leveraged investments for hedging and speculations where margin can keep with broker or exchange.

We purchased the historical data on the website of CME Group and Japan Exchange Group covering the whole year of 2020 trade and order books. After resampling, we obtained Open, High, Low and Close at different time intervals.

5.1. Results on Japan Exchange Group Precious Metal Futures at Different Time Intervals

We selected 6 Janiary 2020, as our observation date, where we captured input prices at various minute intervals. Surprisingly, platinum futures at 15 min intervals have the least MAE-R. We generated 57,960 data points for TGD in Table 5 and TPL in Table 6 for a single day, at least three times larger than the original inputs.

Table 5. TGD-n mins (TOCOM gold) represent futures contracts listed in JPX at one, two, three, four, five, ten and fifteen minutes intervals. For KS R/F Ratio, all are smaller than one, where the highest KS REAL and FAKE are 0.077 and 0.144. Precious futures contract tick value is all one dollar.

Futures-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
TGD-1	13.213	277.513	16.659	0.040	0.042	**0.952**	1	13.213
TGD-2	14.126	302.610	17.396	0.036	0.080	0.450	1	14.126
TGD-3	13.508	283.146	16.827	0.053	0.125	0.424	1	13.508
TGD-4	**14.306**	**317.394**	**17.816**	**0.026**	**0.082**	**0.317**	**1**	**14.306**
TGD-5	12.653	250.569	15.829	0.010	0.113	0.088	1	12.653
TGD-10	12.708	245.903	15.681	**0.077**	**0.144**	0.535	1	12.708
TGD-15	13.979	306.563	16.288	0.040	0.075	0.533	1	13.979

Table 6. TPL-n mins (TOCOM platinum) represent futures contracts listed in JPX at one, two, three, four, five, ten and fifteen minutes intervals. For KS R/F Ratio, 2 out of 7 are bigger than one, where the highest KS REAL and FAKE are 0.138 and 0.143.

Futures-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
TPL-1	38.316	2505.301	50.053	0.043	0.042	**1.023**	1	38.316
TPL-2	37.885	2313.696	48.101	0.078	0.092	0.848	1	37.885
TPL-3	41.881	2715.715	52.113	0.062	0.061	1.016	1	41.881
TPL-4	39.917	2667.278	51.646	0.048	0.066	0.727	1	39.917
TPL-5	**43.670**	**2886.635**	**53.727**	**0.050**	**0.054**	**0.926**	**1**	**43.670**
TPL-10	36.285	2239.771	47.326	0.058	0.083	0.699	1	36.285
TPL-15	35.583	2080.625	45.614	**0.138**	**0.143**	**0.965**	1	35.583

5.2. Results on Chicago Mercantile Exchange Precious Metal Futures at Different Time Intervals

Platinum futures have the highest MAE-R among all contracts, whether CME or JPX. We selected 6 January 2020, as our observation date, where we captured input prices at various minute intervals. We generated 57,960 data points for PLE in Table 7 and GCE in Table 8 for a single day, at least three times larger than the original inputs. We visualized Platinum futures contract (PLE) 15 minutes results in Figure 6.

Table 7. PLE-n mins (platinum) represent futures contracts listed on CME at one, two, three, four, five, ten and fifteen minute intervals. Platinum futures have the highest MAE-R among all futures contracts. For KS R/F Ratio, two out of seven are bigger than one, where the highest KS REAL and FAKE are 0.193 and 0.188.

Futures-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
PLE-1	122.828	27,653.210	166.29	0.031	0.043	0.720	1	122.828
PLE-2	**141.040**	**33,276.476**	**182.418**	**0.042**	**0.051**	**0.820**	**1**	**141.040**
PLE-3	131.296	29,980.058	173.148	0.066	0.083	0.795	1	131.296
PLE-4	129.789	29,907.256	172.937	0.103	0.094	**1.096**	1	129.789
PLE-5	130.660	30,125.549	173.567	0.052	0.067	0.776	1	130.660
PLE-10	139.875	33,091.181	181.910	0.082	0.087	0.943	1	139.875
PLE-15	140.385	30,666.385	175.118	**0.193**	**0.188**	1.027	1	140.385

Table 8. GCE-n mins (gold) represent futures contracts (CME) at intervals of one, two, three, four, five, ten, and fifteen. For KS R/F Ratio, all are smaller than one, where the highest KS REAL and FAKE are 0.058 and 0.145.

Futures-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
GCE-1	59.168	5860.146	76.552	0.033	0.072	0.458	1	59.168
GCE-2	60.253	5704.019	75.525	0.047	**0.145**	0.324	1	60.253
GCE-3	61.642	6015.121	77.557	0.026	0.065	0.400	1	61.642
GCE-4	57.864	5590.208	74.768	0.030	0.078	0.385	1	57.864
GCE-5	60.451	5992.215	77.409	0.037	0.077	0.481	1	60.451
GCE-10	62.611	6352.792	79.704	0.058	0.069	0.841	1	62.611
GCE-15	**66.604**	**6838.208**	**82.693**	**0.054**	**0.110**	**0.491**	**1**	**66.604**
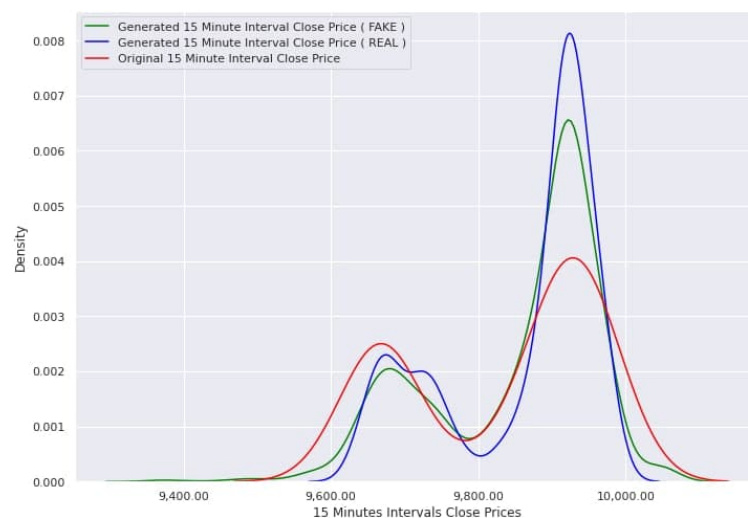
Figure 6. Platinum futures contract (PLE) 15 minutes interval has a larger MAE-R 140.385 than other contracts. Price is concentrated in a certain range and generated long-tailed prices can represent more abundant prices than the original prices.

6. Cryptocurrencies Simulations

The cryptocurrency market runs continuously all day globally. The centralized exchanges across time zones are active somewhere at any time. A cryptocurrency is a digital currency secured by cryptography, making it nearly impossible to counterfeit or double-spend. Many cryptocurrencies are decentralized networks based on blockchain technology—a distributed ledger enforced by computers networks. As an emerging asset class, we examine the following spots and perpetual futures contracts by minute level intervals.

6.1. Spot Results and Analysis

A Spot Market is a market where investors can trade assets with other traders in real-time. As the name suggests, transactions are settled immediately or "on the spot" when the buying/selling order is filled. You can purchase an asset with fiat or another cryptocurrency from a seller as a buyer.

The data is open-source on the website of Binance (n.d.)[2] where we downloaded spot and results for Bitcoin (BTCUSDT) in Table 9, Ethereum (ETHUSDT) in Table 10, Binance Coin (BNBUSDT) in Table 11, Cardano (ADAUSDT) in Table 12, Solana (SOLUSDT) in Table 13, XRP (XRPUSDT) in Table 14, Polkadot (DOTUSDT) in Table 15 on the date of 3 Septemper 2021, from 12:00 a.m. to 11:59 p.m. for a total of 1440 trading minutes. We generated 202,860 data points in total for a single trading day, 28,980 each symbol.

Table 9. We examined BTC spot prices against generated prices at intervals of one, two, three, four, five, ten and fifteen minutes. For the KS R/F Ratio, four out of seven are more significant than one, indicating hard to train, where the highest KS REAL and FAKE are 0.112 and 0.114.

Spot/USDT-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
BTC-1	751.285	884,775.829	940.625	0.077	0.067	1.149	0.01	75,128.5
BTC-2	**809.114**	**1,001,496.257**	**1000.748**	**0.065**	**0.079**	**0.823**	**0.01**	**80,911.4**
BTC-3	709.722	741,981.902	861.384	0.028	0.061	0.459	0.01	70,972.2
BTC-4	697.024	770,976.836	878.053	0.073	**0.114**	0.640	0.01	69,702.4
BTC-5	705.130	754,617.954	868.687	0.088	0.056	**1.571**	0.01	70,513.0
BTC-10	707.617	768,399.041	876.584	**0.112**	0.103	1.087	0.01	70,761.7
BTC-15	678.269	712,963.929	844.372	0.084	0.061	1.377	0.01	67,826.9

Table 10. We examined ETH spot prices against generated prices at intervals of one, two, three, four, five, ten and fifteen minutes. For KS R/F Ratio, 2 out of 7 are bigger than one, where the highest KS REAL and FAKE are 0.094 and 0.186.

Spot/USDT-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
ETH-1	94.735	14,434.334	120.143	0.045	0.052	0.865	0.01	9473.5
ETH-2	93.375	14,331.716	119.715	0.054	0.052	1.038	0.01	9337.5
ETH-3	82.292	11,924.118	109.198	0.045	0.046	0.978	0.01	8229.2
ETH-4	92.459	13,608.065	116.654	0.048	0.055	0.873	0.01	9245.9
ETH-5	**97.982**	**15,446.558**	**124.284**	**0.054**	**0.046**	**1.174**	**0.01**	**9798.2**
ETH-10	85.679	12,309.505	110.948	0.048	0.064	0.750	0.01	8567.9
ETH-15	73.365	10,073.590	100.367	**0.094**	**0.186**	0.505	0.01	7336.5

Table 11. We examined BNBUSDT spot prices against generated prices at one, two, three, four, five, ten and fifteen minute intervals. For the KS R/F Ratio, two out of seven are more significant than one, where the highest KS REAL and FAKE are 0.103 and 0.113.

Spot/USDT-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
BNB-1	4.552	34.275	5.855	0.059	0.044	**1.340**	0.01	455.2
BNB-2	**4.773**	**35.867**	**5.989**	**0.018**	**0.058**	**0.310**	**0.01**	**477.3**
BNB-3	**4.784**	**39.899**	**6.317**	**0.024**	**0.097**	**0.247**	**0.01**	**478.4**
BNB-4	4.289	30.486	5.521	0.059	0.077	0.766	0.01	428.9
BNB-5	4.507	33.414	5.780	0.026	0.078	0.333	0.01	450.7
BNB-10	4.470	31.758	5.635	0.085	**0.113**	0.752	0.01	447.0
BNB-15	3.836	24.535	4.953	**0.103**	0.102	1.010	0.01	383.6

Table 12. We examined ADAUSDT spot prices against generated prices at one, two, three, four, five, ten and fifteen minute intervals. For KS R/F Ratio, two out of seven are more significant than one, where the highest KS REAL and FAKE are 0.108 and 0.101.

Spot/USDT-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
ADA-1	0.026000	0.001096	0.033106	0.059	0.044	**1.341**	0.001	26
ADA-2	0.025817	0.001030	0.032090	0.047	0.038	1.237	0.001	25.817
ADA-3	0.025094	0.000997	0.031575	0.044	0.051	0.863	0.001	25.094
ADA-4	**0.026527**	**0.001103**	**0.033213**	**0.047**	**0.053**	**0.887**	**0.001**	**26.527**
ADA-5	0.023126	0.000849	0.029142	0.058	0.070	0.829	0.001	23.126
ADA-10	0.023220	0.000844	0.029059	0.081	**0.101**	0.802	0.001	23.220
ADA-15	0.022829	0.000842	0.029011	**0.108**	0.089	1.213	0.001	22.829

Table 13. We examined SOLUSDT spot prices against generated prices at one, two, three, four, five, ten and fifteen minute intervals. For KS R/F Ratio, three out of seven are bigger than one, where the highest KS REAL and FAKE are 0.082 and 0.073.

Spot/USDT-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
SOL-1	4.277194	29.315060	5.414338	0.067	0.044	**1.523**	0.01	427.7194
SOL-2	**4.809014**	**35.177124**	**5.931031**	**0.028**	**0.050**	**0.560**	**0.01**	**480.9014**
SOL-3	4.655123	34.650182	5.886441	0.041	0.064	0.641	0.01	465.5123
SOL-4	4.539751	32.206140	5.675045	0.042	0.049	0.857	0.01	453.9751
SOL-5	4.344479	29.525525	5.433739	0.048	0.059	0.814	0.01	434.4479
SOL-10	4.241806	28.286172	5.318475	0.071	0.059	1.203	0.01	424.1806
SOL-15	3.884270	25.669128	5.066471	**0.082**	**0.073**	1.123	0.01	388.4270

Table 14. We examined XRPUSDT spot prices against generated prices at one, two, three, four, five, ten and fifteen minute intervals. For KS R/F Ratio, two out of seven are more significant than one, where the highest KS REAL and FAKE are 0.092 and 0.124.

Spot/USDT-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
XRP-1	0.021332	0.000709	0.026619	0.071	0.080	0.888	0.01	2.1332
XRP-2	**0.025705**	**0.000983**	**0.031350**	**0.054**	**0.089**	**0.607**	**0.01**	**2.5705**
XRP-3	0.024516	0.001001	0.0316399	0.038	0.087	0.437	0.01	2.4516
XRP-4	0.022586	0.000827	0.028763	0.079	0.067	1.179	0.01	2.2586
XRP-5	0.023268	0.000862	0.029365	0.063	**0.124**	0.508	0.01	2.3268
XRP-10	0.021848	0.000724	0.026909	0.087	0.096	0.906	0.01	2.1848
XRP-15	0.021128	0.000720	0.0268356	**0.092**	0.067	**1.373**	0.01	2.1128

Table 15. We compared DOTUSDT spot prices against generated prices at intervals of one, two, three, four, five, ten, and fifteen minute intervals. DOTUSDT 15 minutes KS are the largest among all time intervals.

Spot/USDT-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
DOT-1	**0.668215**	**0.795363**	**0.891831**	**0.056**	**0.072**	**0.777**	**0.01**	**66.8215**
DOT-2	0.661014	0.767141	0.875866	0.088	0.086	1.023	0.01	66.1014
DOT-3	0.638965	0.782794	0.884757	0.058	0.049	1.184	0.01	63.8965
DOT-4	0.610541	0.725811	0.851945	0.073	0.099	0.737	0.01	61.0541
DOT-5	0.603408	0.652360	0.807688	0.046	0.074	0.622	0.01	60.3408
DOT-10	0.611593	0.713514	0.844698	0.090	0.051	**1.765**	0.01	61.1593
DOT-15	0.496690	0.468474	0.684451	**0.178**	**0.166**	1.072	0.01	49.6690

6.2. Perpetual Futures Results and Analysis

Perpetual contracts are derivative contracts similar to futures with no expiration date or settlement, allowing them to be held or traded for a consistent trading time. Perpetual contracts are gaining popularity in cryptos because they allow traders to hold leveraged positions without the burden of an expiration date. Unlike traditional futures, perpetual contracts trade close to the index price of the underlying asset due to perpetual funding rates. We included BTCUSDT Perpetual in Table 16 and visualized BTC Perpetual in Figure 7, ETHUSDT Perpetual in Table 17, BNBUSDT Perpetual in Table 18, ADAUSDT Perpetual in Table 19, and XRPUSDT Perpetual in Table 20 for multiple intervals.
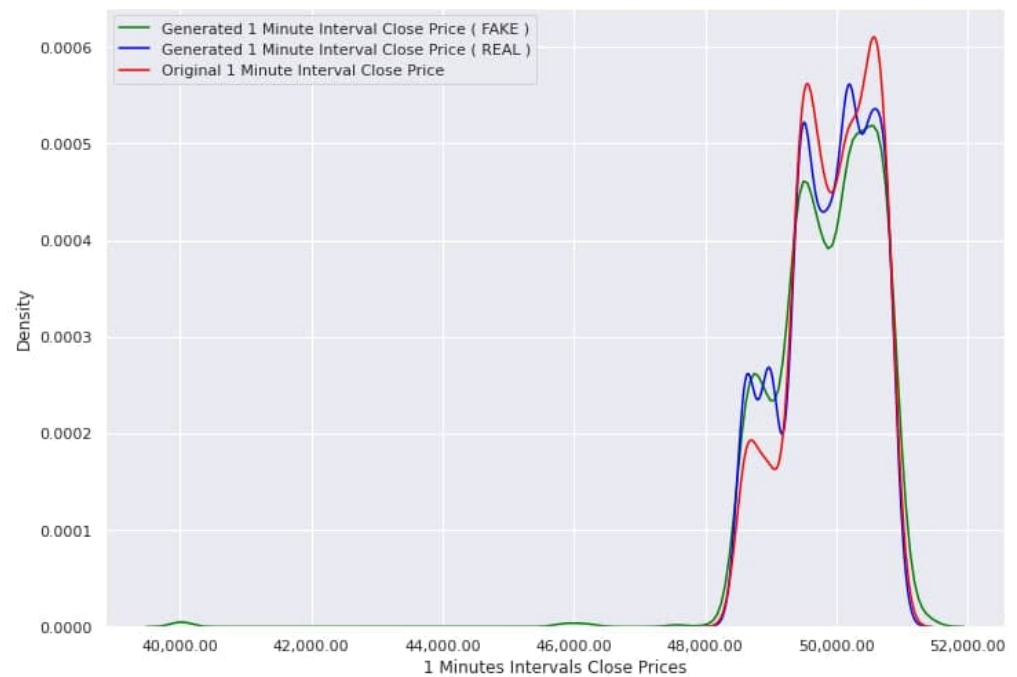
Figure 7. This graph shows the BTC one minute Interval Close prices and generated fake and real price data. BTC perpetual futures contracts are relatively volatile, with a relative large MAE of 72,840.8 compared with other perpetual contracts. The highest density from the original reaches 0.0006. The long tails from generated real and fake show the potential price volatility indicated by the original prices covering 40,000 to 48,000 USDT.

Table 16. We compared BTCUSDT perpetual prices at multiple time intervals. BTCUSDT at one minute interval has the highest MAE-R scores. The smallest KS REAL and KS FAKE are 0.044, and 0.071. The highest KS REAL and KS FAKE are 0.113 and 0.192, respectively.

Perpetual-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
BTC-1	**728.408**	**824,327.849**	**907.925**	**0.054**	**0.082**	**0.658**	**0.01**	**72,840.8**
BTC-2	715.831	775,010.148	880.347	0.044	0.074	0.659	0.01	71,583.1
BTC-3	687.178	732,522.649	855.875	0.072	0.073	0.986	0.01	68,717.8
BTC-4	702.501	760,323.665	871.965	**0.113**	0.096	**1.177**	0.01	70,250.1
BTC-5	714.130	788,127.987	887.766	0.075	0.071	1.056	0.01	71,413.0
BTC-10	598.977	571,480.481	755.963	0.071	0.071	1.000	0.01	59,897.7
BTC-15	716.976	12,672.582	112.573	0.069	**0.192**	0.359	0.01	71,697.6

Table 17. We compared ETHUSDT perpetual prices at multiple intervals. The smallest KS REAL and KS FAKE are 0.029, and 0.054. The highest KS REAL and KS FAKE are 0.085 and 0.219 respectively. ETHUSDT perpetual 15 min intervals, KS R/F Ratio is 0.388 where KS REAL is 0.085, indicating the generated prices are indeed close to original price inputs. Only ETHUSDT 5 min intervals generated FAKE outperform the generated REAL, the R/F Ratio is 1.130.

Perpetual-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
ETH-1	93.703	14,675.114	121.141	0.038	0.060	0.633	0.01	9370.3
ETH-2	93.013	13,799.069	117.469	0.066	0.077	0.857	0.01	9301.3
ETH-3	85.099	12,511.970	111.857	0.029	0.054	0.537	0.01	8509.9
ETH-4	91.193	12,672.582	112.573	0.069	0.192	0.359	0.01	9119.3
ETH-5	**94.474**	**14,938.206**	**122.222**	**0.061**	**0.054**	**1.130**	**0.01**	**9447.4**
ETH-10	83.122	11,923.743	109.196	0.058	0.087	0.667	0.01	8312.2
ETH-15	84.010	12,094.374	109.974	**0.085**	**0.219**	0.388	0.01	8401.0

Table 18. We compared BNBUSDT perpetual futures at multiple intervals. Interestingly, BNB, ETH, BTC, ADA and XRP perpetual futures all show the highest KS FAKE score in 15 min intervals due to high volatility.

Perpetual-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
BNB-1	4.202	31.258	5.591	0.086	0.099	0.868	0.01	420.2
BNB-2	4.334	31.809	5.640	0.049	0.098	0.500	0.01	433.4
BNB-3	4.571	33.465	5.785	0.034	0.067	0.507	0.01	457.1
BNB-4	**4.846**	**38.627**	**6.215**	**0.051**	**0.089**	**0.573**	**0.01**	**484.6**
BNB-5	4.528	34.421	5.867	0.020	0.047	0.426	0.01	452.8
BNB-10	4.375	33.382	5.778	0.078	0.076	1.026	0.01	437.5
BNB-15	4.660	35.617	5.968	**0.110**	**0.108**	1.019	0.01	466.0

As shown in Table 16, BTC Perpetuals, for KS R/F Ratio, three out of seven are more significant than one. From Table 17 ETHUSDT Perpetuals, only one ratio is over one. From Table 18 BNBUSDT Perpetuals, two out of seven are over one. Also, for Table 19 ADAUSDT Perpetuals, two out of seven are more significant than one. For Table 20 XRPUSDT Perpetuals, one out of seven is more significant than one. Overall, the REAL is much closer to the original price inputs, ensuring the data quality we generate.

Table 19. We compare ADAUSDT perpetual futures at multiple time intervals.

Perpetual-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
ADA-1	**0.025316**	**0.001064**	**0.032616**	**0.044**	**0.046**	**0.956**	**0.0001**	**253.16**
ADA-2	0.023791	0.000922	0.030358	0.024	0.060	0.400	0.0001	237.91
ADA-3	0.024922	0.001012	0.031813	0.054	0.052	1.038	0.0001	249.22
ADA-4	0.023527	0.000891	0.029857	0.069	0.100	0.690	0.0001	235.27
ADA-5	0.024623	0.00101	0.031774	0.067	0.073	0.918	0.0001	246.23
ADA-10	0.024315	0.000940	0.030662	0.091	0.103	0.883	0.0001	243.15
ADA-15	0.023182	0.000950	0.030822	**0.113**	**0.107**	**1.056**	0.0001	231.82

Table 20. We compared XRPUSDT perpetual futures at multiple time intervals. XRP-3 has the largest MAE-R among all intervals.

| Perpetual-n mins	MAE-R	MSE-R	RMSE-R	KS REAL	KS FAKE	KS R/F Ratio	Tick	MAE-R/ Tick
XRP-1	0.021806	0.000781	0.028072	0.066	0.074	0.891	0.0001	218.06
XRP-2	0.021128	0.000780	0.027932	0.050	0.052	0.962	0.0001	211.28
XRP-3	**0.024460**	**0.001067**	**0.032661**	**0.039**	**0.024**	**1.625**	**0.0001**	**244.60**
XRP-4	0.022315	0.000794	0.028179	0.073	0.104	0.702	0.0001	223.15
XRP-5	0.021052	0.000714	0.026724	0.034	0.073	0.466	0.0001	210.52
XRP-10	0.020315	0.000699	0.026433	0.067	0.068	0.985	0.0001	203.15
XRP-15	0.022898	0.000850	0.029157	**0.083**	**0.107**	0.776	0.0001	228.98

7. Discussion and Future Directions

Each asset class shows its unique features due to granular trading conditions, market microstructure, liquidity and volatility. A few directions to explore:

1. There exists the possibility to extend more asset classes into Forex and other emerging markets like China A markets. The exchange will only release price quotations once every three seconds, including price and volume, within three seconds are hidden from the markets. Therefore, the simulation is useful for generating abundant hidden information and understanding the blackbox between the intervals for understanding the China A markets.

2. From the asset classes perspective, the Hong Kong stock market has less liquidity due to policy risks and expensive transaction fees of around 0.1477% of total transaction value per side, including buying and selling. Lezmi et al. (2020) conducted experiments including risk parity strategy and augmenting the investment universe with market regime indicators. Lezmi addressed RBMs and GANs being used for estimating the probability distribution of performance and risk statistics, which can improve the risk management of quantitative investment strategies. Rosolia and Osterrieder (2021) discussed different asset classes such as commodities, forex, futures, index and shares. For trading hours, the Hong Kong stock market had 5.5 h of trading sessions with an hour lunch break, almost one-quarter of the futures market (Sunday–Friday 6:00 p.m.–5:00 p.m. (5:00 p.m.–4:00 p.m./CT) with a 60-min break each day beginning at 5:00 p.m. (4:00 p.m. CT)) and crypto market (24 h). Interestingly, the 2269.HK-1 min test result is around 100.38 for MAE-Real/Tick size, which is five times larger than the 3690.HK-1 min test result, ten times larger compared to 0005.HK-1 min test result, and 14 times larger than 0688.HK-1 min test result. Some underlying features (e.g., volume, volatility, etc.) might exist during our observations. Our simulator could discover the potential abnormal stock behaviour and underlying market impacts by comparing generated prices against original input price behaviour. Further investigation can focus on evaluating and clustering MAE values to discover market microstructure and trading conditions.

3. Volume simulation can be another further work. Institutional investment banks and trading firms predict volume curves for normal days for optimizing their portfolio

positions. Special dates like index rebalancing days, Korean exam days, Japan Nikkei 225 component weights change days, festival half days, Triple witching hour days on the third Friday of every March, June, September, and December are volume sensitive. Many volume parameters of passive and active trading executions need to be discovered.

4. From the trading frequency perspective, simulating tick data, minute-level data, and daily-level data will enhance the robustness and scalability for multiple trading frequencies. Some asset managers prefer to buy and hold strategies, and others prefer transactions at a weekly frequency. Kakushadze (2016) presented alpha signals with an average holding period ranging from approximately 0.6 to 6.4 days. Thus, a scalable observation of pricing the inventory assets would be desired. Simulating order books, including best bid and ask orders with inter-arrivals, especially on extreme market conditions like COVID 19 or rockets up like Trading Curb situations, will be helpful topics to explore. The figure shows the simulated volume-weighted average price (VWAP) and time-weighted average price (TWAP) against original VWAP and TWAP. The spread is getting narrow when trading continues until 1200 minutes in Figure 8.
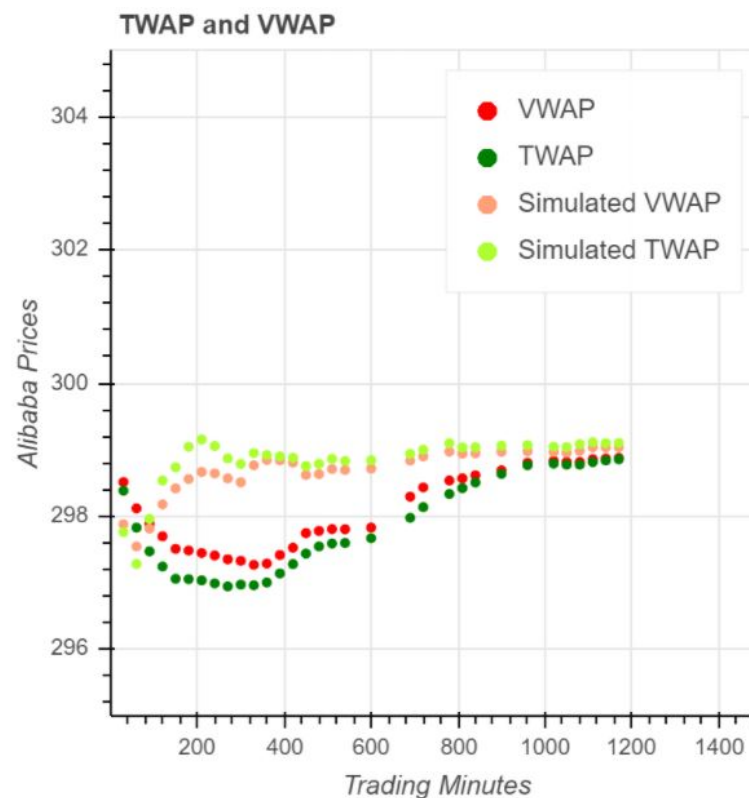
Figure 8. One of the implications can be backtesting execution algorithm parameters to optimize filling rates and portfolio positions. The left figure shows the simulated volume-weighted average price and time-weighted average price against original VWAP and TWAP. The spread gets narrow when trading continues until 1200 minutes. The prices are concentrated in a particular range, and WGAN-GP learns from the original prices and can even show the potential market uptrends.

8. Conclusions

The contributions can be summarized in the following:

1. A complete WGAN-GP framework was proposed through extensive and systematic experiments for multiple asset classes, including stocks, futures, and cryptocurrency. Also, an in-depth analysis was given to fine-tune important hyperparameters. We trained our model on various datasets, including the Hong Kong Stock Market Hang Seng Index composite stocks, precious metal futures contracts listed on the Chicago

Mercantile Exchange, the Japan Exchange Group, and cryptocurrency spots and perpetual contracts on Binance at various minute-level intervals. After generating 836,280 data points from 90,900 epochs, we showed that WGAN-GP can be trained as a probabilistic model with stability, scalability and robustness for price simulations.

2. From an asset classes perspective, we find that 2269.HK has the largest MAE-R/Tick among Hang Seng Index stocks. Platinum assets like TPL (TOCOM Platinum on the Tokyo Commodity Exchange) and PLE (Platinum (Globex) NYMEX on the Chicago Mercantile Exchange) have relatively larger MAE-R/Tick than a gold asset like GCE and TGD. BTC, ETH, and BNB spots and perpetual futures have the largest MAE-R/Tick values. Many interesting topics remain to work on. First, we can further extend to other markets like The Standard and Poor's 500, Nikkei 225, KOSPI 200 Index, CSI Small 500 Index, and the DAX Performance Index. Second, derivatives, especially non-linear products like option pricing, can be simulated against historical data.

3. From a trading frequency perspective, WGAN-GP can be used to simulate possible price datasets at various minute-level intervals. We proposed to evaluate the outcomes by comparing the generated synthetic distribution against the original data distribution at multiple time intervals, including one, two, three, four, five, ten and fifteen minutes. When the model goes to the production environment, for execution considerations, some statistics for evaluation such as the quantity of executed orders, prediction of inter-arrival orders, cancellation rates and the best bid and best ask can be implemented.

To sum up, our results show that WGAN-GP can simulate asset prices and show a market simulator's potential for trading analysis. We might be the first to look into multi-asset classes in a systematic approach with minute level intervals across stocks, futures and cryptocurrencies markets. We also contribute to quantitative analysis methodology for generated and original price data quality.

Author Contributions: Conceptualization, F.H. and J.Z.; methodology and visualization, F.H.; resources and curation, J.Z.; writing—original draft preparation, F.H.; writing—review and editing, X.M.; supervision, J.Z. and funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data that supports the findings of this study are available from the corresponding author upon reasonable request. Restrictions apply to the availability of data.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

TGD	TOCOM Gold on Tokyo Commodity Exchange
TPL	TOCOM Platinum on Tokyo Commodity Exchange
GCE	Gold (Globex) COMEX on Chicago Mercantile Exchange
PLE	Platinum (Globex) NYMEX on Chicago Mercantile Exchange
CME	COMEX on Chicago Mercantile Exchange
JPX	Japan Exchange Group
WGAN-GP	Wasserstein Generative Adversarial Networks-Gradient Penalty
HSI	Hang Seng Indexes

Notes

1   https://help.yahoo.com/kb/SLN2311.html/, accessed on 8 May 2021.
2   https://www.cryptodatadownload.com/data/binance/, accessed on 8 May 2021.

References

Abraham, Jethin, Daniel Higdon, John Nelson, and Juan Ibarra. 2018. Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review* 1: 1.

Ariyo, Adebiyi A., Adewumi O. Adewumi, and Charles K. Ayo. 2014. Stock price prediction using the ARIMA model. Paper presented at the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, March 26–28; pp. 106–12.

Arjovsky, Martin, Soumith Chintala, and Léon Bottou. 2017. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*. Sydney: PMLR; pp. 214–23.

Baur, Dirk G., and Lai T. Hoang. 2021. A crypto safe haven against Bitcoin. *Finance Research Letters* 38: 101431. [CrossRef]

Binance. n.d. Available online: https://www.binance.com/en (accessed on 7 January 2022).

CME Gold Futures and Options. n.d. Available online: https://www.cmegroup.com/markets/metals/precious/gold.html (accessed on 8 May 2021).

CME Platinum Futures and Options. n.d. Available online: https://www.cmegroup.com/markets/metals/precious/platinum.html (accessed on 8 May 2021).

Foster, Ernest Allen. 2002. Commodity Futures Price Prediction: An Artificial Intelligence Approach. Doctoral dissertation, University of Georgia, Athens, GA, USA.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in Neural Information Processing Systems* 27. Available online: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf (accessed on 8 May 2021).

Gulrajani, Ishaan, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. 2017. Improved training of wasserstein gans. *arXiv*, arXiv:1704.00028.

Hong Kong Exchanges and Clearing Limited. n.d. Available online: https://www.hkex.com.hk/ (accessed on 8 May 2021).

JPX Gold Futures and Options. n.d. Available online: https://www.jpx.co.jp/english/derivatives/products/precious-metals/gold-standard-futures/index.html (accessed on 8 May 2021).

JPX Platinum Futures and Options. n.d. Available online: https://www.jpx.co.jp/english/derivatives/products/precious-metals/platinum-standard-futures/01.html (accessed on 8 May 2021).

Kakushadze, Zura. 2016. 101 formulaic alphas. *Wilmott* 84: 72–81. [CrossRef]

Kang, Sang Hoon, Ron McIver, and Seong-Min Yoon. 2017. Dynamic spillover effects among crude oil, precious metal, and agricultural commodity futures markets. *Energy Economics* 62: 19–32. [CrossRef]

Kim, Harold Y., and Jianping P. Mei. 2001. What makes the stock market jump? An analysis of political risk on Hong Kong stock returns. *Journal of International Money and Finance* 20: 1003–16. [CrossRef]

Kyriazis, Nikolaos A., Kalliopi Daskalou, Marios Arampatzis, Paraskevi Prassa, and Evangelia Papaioannou. 2019. Estimating the volatility of cryptocurrencies during bearish markets by employing GARCH models. *Heliyon* 5: e02239. [CrossRef] [PubMed]

Lee, Seungho, Nabil El Meslmani, and Lorne N. Switzer. 2020. Pricing efficiency and arbitrage in the bitcoin spot and futures markets. *Research in International Business and Finance* 53: 101200. [CrossRef]

Lezmi, Edmond, Jules Roche, Thierry Roncalli, and Jiali Xu. 2020. Improving the Robustness of Trading Strategy Backtesting with Boltzmann Machines and Generative Adversarial Networks. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3645473 (accessed on 8 May 2021).

Li, Junyi, Xintong Wang, Yaoyang Lin, Arunesh Sinha, and Michael Wellman. 2020. Generating realistic stock market order streams. Paper presented at the AAAI Conference on Artificial Intelligence, Taipei, Taiwan, April 11–13; vol. 34, pp. 727–34.

Mensi, Walid, Jose Arroeola Hernandez, Seong-Min Yoon, Xuan Vinh Vo, and Sang Hoon Kang. 2021. Spillovers and connectedness between major precious metals and major currency markets: The role of frequency factor. *International Review of Financial Analysis* 74: 101672. [CrossRef]

Meyer, David R., and George Guernsey. 2017. Hong Kong and Singapore exchanges confront high frequency trading. *Asia Pacific Business Review* 23: 63–89. [CrossRef]

Palamalai, Srinivasan, K. Krishna Kumar, and Bipasha Maity. 2021. Testing the random walk hypothesis for leading cryptocurrencies. *Borsa Istanbul Review* 21: 256–68. [CrossRef]

Rosolia, Antonio, and Joerg Osterrieder. 2021. Analyzing Deep Generated Financial Time Series for Various Asset Classes. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3898792 (accessed on 8 May 2021).

Samuel, Rikli, Bigler Daniel Nico, Pfenninger Moritz, and Osterrieder Joerg. 2021. Wasserstein GAN: Deep Generation applied on Bitcoins financial time series. *arXiv*, arXiv:2107.06008.

Smith, Leslie N. 2017. Cyclical learning rates for training neural networks. Paper presented at the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, March 24–31; pp. 464–72.

Xu, Ke, Xinwei Zheng, Deng Pan, Li Xing, and Xuekui Zhang. 2020. Stock Market Openness and Market Quality: Evidence from the Shanghai–Hong Kong Stock Connect Program. *Journal of Financial Research* 43: 373–406. [CrossRef]

Xu, Xiaoqing Eleanor, and Hung-Gay Fung. 2005. Cross-market linkages between US and Japanese precious metals futures trading. *Journal of International Financial Markets, Institutions and Money* 15: 107–24. [CrossRef]

Yahoo Finance. n.d. Available online: https://help.yahoo.com/kb/SLN2311.html (accessed on 7 January 2022).