Interpolation of Financial Time Series Data in a Virtual Geographic Environment

Abstract

This paper introduces a new approach to visualising and interpolating financial time series data, e.g., Bitcoin prices, in a spatial domain using the notion of spatialization: forming a spatial representation of non-spatial phenomena. The proposed algorithm first utilises the temporal components of the observations, i.e., date and time, to build a 2D virtual geographic map. It then uses the assigned coordinates to the observations and their values to estimate unknown values and construct a 3D topographic map. We assess the 3D maps using the price time series of Bitcoin with 30-minute frequency, and the results show the reliability of the 3D maps in analysing the time series data.

1. Introduction

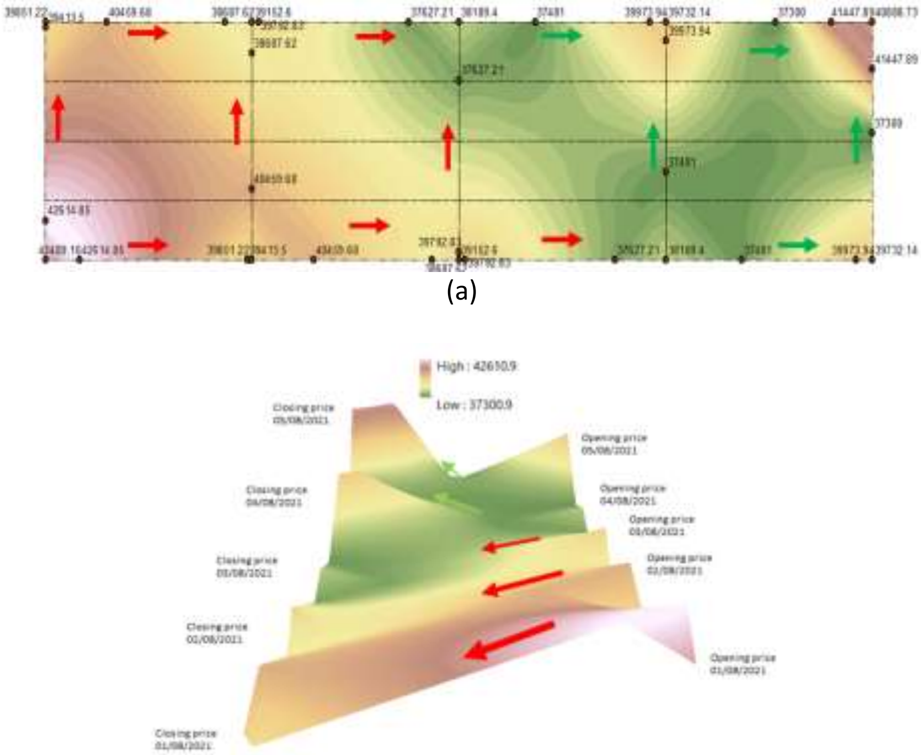
The vast majority of the algorithms in financial markets are established based on using time series data observed at equal time intervals. In this structure, an observation at time interval t, x_t, is formally defined via a n-dimensional vector, where $n \in \mathbb{N}$. A typical case of time series data is when n=1, e.g., using the closing prices of Bitcoin [1,2]. In this case, observations are usually linked and modelled using an irregular line or curve in a two-dimensional space, where the x-axis represents time, and the y-axis represents the variable of interest, e.g., the daily closing price of Bitcoin.

In the case that n>1, e.g., using the opening, high, low and closing prices of Bitcoin [3,4], the algorithms use a set of observations at each time step for modelling time series data. In this scenario, a combination of regular and irregular lines is typically applied to represent time series data, e.g., candlestick or bar charts. The irregular horizontal line is formed along the x-axis using the scalar value observed at equal time, e.g., the daily closing price. The straight vertical lines are applied to represent the opening, high, low and closing prices along the y-axis. The lengths of vertical lines are adjusted based on the difference between the values of the observation vector at each time interval.

The above time series representations show that the lines and curves are the main geometric elements in modelling and representing time series data. It also indicates that the observations can only be represented in a two-dimensional cross section view. In this paper, we propose a novel approach for analysing and visualising the time series data in a 3D virtual geographic environment. This approach is established based on spatialization: modelling a non-spatial phenomenon in a spatial domain [5,6]. The use of this model allows us to simultaneously link observations at different times and interpret patterns in a 3D virtual geographic environment in a simple and meaningful way. The use of the 3D maps can also provide a new tool for traders or financial analysers to visualize time series data using different graphical projections, e.g., isometric.

2. Data and methodology

To interpolate data and assess the proposed approach, Bitcoin data from August 1, 2021, to August 13, 2021, was downloaded from https://firstratedata.com/i/crypto/BTC and applied. We used one-minute frequency data to identify each day's opening, high, low, and closing price and the times assigned to these prices. The proposed method is implemented in two steps: creating the 2D spatialized map and constructing the topographic map. MATLAB2022a and ArcGIS 10.5 are applied to create, analyse, and render the 2D and 3D maps in a virtual geographic environment (Figure 1).



(a)

(b)

Figure 1. (a) The produced Bitcoin map using the NNI algorithm. Point features are applied to show the location of each price on the map. (b) cross-section view of Bitcoin price.

Figure 1 provides a 3D representation of Bitcoin prices for five days, between 01/08/2021 and 05/08/2021, in two views: orthographic projection (a) and cross-section (b). The use of these 3D maps makes it easier to interpret data than conventional chart-based representations, e.g., bar chart. For example, in Figure 1(a), the highest elevation is 42614 (01/08/2021), colored in white and located in the lower southwest corner of the 3D map. The lowest elevation (37301) on the map is green and placed on the east and northeast sides of the map. In Figure 1(b), an investor can easily identify when the Bitcoin market is red or green by comparing elevations in the 3D map. As the maps are modelled in a spatial domain, the coordinates of points can also be applied to generate different products, such as hill shade or slope maps. These maps can be utilized to model volatility for risk management in the crypto market.

3. Results

To quantitively assess the results of the proposed method, we use the 30-minute frequency Bitcoin prices, which are divided into two groups: Dataset 1 (1/8/2021-4/8/2021) and Dataset 2 (8/8/2021-12/8/2021). The following metrics: RMSE (root mean square error), MAPE (mean absolute percent error), MASE (mean absolute scaled error), and DA (directional accuracy) are applied to assess the accuracy and dynamic behaviour of the interpolated prices via the 3D maps [2].

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y(t) - \hat{y}(t))^2} \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y(t) - \hat{y}(t)}{y(t)}\right| \tag{3}$$

$$DA = \frac{1}{n}\sum_{t=1}^{n} a(t) \times 100\% \tag{4}$$

Where the $y(t)$ and $\hat{y}(t)$ are the actual and interpolated prices at time *t*, respectively, and *n* is the number of observations. In Equation 4, $a(t)$=0 if $(y(t+1) - y(t)) - (\hat{y}(t-1) - y(t)) < 0$, otherwise $a(t)$=1. To transform the irregularly spaced time series prices into regularly spaced time series prices, we also use the 8th degree polynomial, and the Hermite interpolation function, which is denoted as irregular data interpolation (IRDInterpolation) in Table 1. This allows us to compare the performance of the proposed method against these two conventional methods, which are usually applied to create regular time series data from irregular time series data. We also apply the Hermite interpolation function to transform the 24-hour frequency prices into 30-minute frequency prices, denoted as regular data interpolation (RDInterpolation) in Table 1. This enables us to assess the effect of using unequal time series data, namely high and low prices, in the interpolation process. Table 1 shows the results of these four methods in terms of the RMSE, MAPE and DA values. The best and worst values are highlighted in bold and underlined, respectively.

Table 1. shows the RMSE, MAPE, and DA values of the interpolation data created using four different methods: the proposed method (3D maps), IRDInterpolation, RDInterpolation, and the 8th degree polynomial function.

Data	Method	RMSE	MAPE	DA
	Proposed method	**484.75**	**0.91**	**56.25**
Dataset 1	IRDInterpolation	594.70	1.12	<u>54.17</u>
	Polynomial	<u>724.45</u>	<u>1.19</u>	55.73
	RDInterpolation	552.30	0.99	55.21
	Proposed method	**429.14**	**0.75**	**58.33**
Dataset 2	IRDInterpolation	485.76	0.89	<u>53.33</u>
	Polynomial	<u>658.99</u>	<u>1.10</u>	**58.33**
	RDInterpolation	519.98	0.91	55.00

As shown in Table 1, the proposed method provides better results compared with the other three methods for both datasets. In Dataset 1, the proposed method improves the RMSE values by more than 18%, 33%, and 12% compared with the IRDInterpolation, Polynomial, and RDInterpolation methods, respectively. For the MAEP values, an improvement of 17%, 23%, and 8% can be seen using the proposed method compared to the other methods. The highest DA value for Dataset 1 is 56.25, which belongs to the proposed method. This indicates that the 3D maps not only improve the level of accuracy of the interpolated data but also model the dynamic behaviour of time series data better in comparison with the other methods.

The RMSE and MAPE values confirm the better performance of the IRDInterpolation and RDInterpolation algorithms compared with the polynomial method. However, the DA values in Table 1 indicate that the polynomial algorithm performs better than the interpolation function in modelling the dynamic behaviour of the interpolated time series prices. This is because the interpolating function must pass exactly through all the observed prices. Therefore, if there is a sudden jump or drop rise in the Bitcoin prices, the algorithm might create unwanted undulations by decreasing or increasing slopes of interpolant at the interpolation points. This can change the dynamic behaviour of the time series data. The results in Table 1 for Dataset 1 also show the better performance of the RDInterpolation method compared with the IRDInterpolation, while both methods use the same interpolation function to create the data samples. The possible reason for the better performance of RDInterpolation is that the method is formed based solely on the closing price observed with the 24-hour interval. Therefore, the low and high prices cannot affect the slopes of the interpolant. While the RDInterpolation algorithm is formed based on unequally spaced time intervals. This can limit the accuracy of the interpolation function when there is a sharp rise or fall in the Bitcoin prices.

Similar to Dataset 1, the results of the proposed method show a better performance in generating the interpolated prices based on Dataset 2 compared with the other three methods. For example, the use of the 3D map generated data shows an improvement of more than 12% for the RMSE and MAPE values compared with the second-best RMSE and MAPE values in Table 1. For Dataset 2, the highest DR value is 58.33, which belongs to the polynomial and the proposed algorithm. However, the RMSE and MAPE values show that the polynomial function performs poorly compared to the proposed method. The poor results of the polynomial function can be due to the fact the function is formed by fitting a mathematical model to the prices. Therefore, it is not necessary for the function to pass precisely through the observed prices. This can reduce the model's accuracy, especially when the volatility of the Bitcoin prices is high. For both datasets, a comparison between the results of the 3D maps and RDInterpolation confirms that the use of the closing, low and high prices to interpolate the time series data can significantly improve the accuracy of the interpolated data.

Figure 2 shows the time series data and its corresponding trend components for Dataset 1 and Dataset 2, which are drawn using the actual data and the interpolated data. Figure 2(a) illustrates that the geometric behaviour of the trend created by IRDInterpolation highly coincides with the trend generated using the 3D maps. This is because, in both methods, the model is formed by passing through the observed prices. Figure 2(a) also displays that there is a gap between the trend created by the RDInterpolation and IRDInterpolation method from time 0 to time 1800, where there is a sharp drop in the Bitcoin price, while both methods use the same interpolation function. The possible reason is that the RDInterpolation method uses more information to interpolate the data. The graphs in Figure 2 (b) are formed by adding the seasonal component to the trend of the time series data. It can be seen that the proposed method shows better performance than the other methods. A comparison between graphs in Figure 2(c) and Figure 2(b) also confirms the better performance of the 3D map in estimating the stochastic component of the time series data compared with the other methods.

In contrast to Dataset 1 that there is a downward trend, Figure 2(d) shows an upward trend for the Bitcoin prices based on Dataset 2. Similar to the graphs in the first dataset, the trend of the proposed method is the closest trend to the actual data trend. This indicates the ability of the 3D maps to accurately interpolate prices under different circumstances. Figure 2(e) displays that the graphs of the 3D maps and the IRDInterpolation are similar. However, Figure 2(f) shows the created graph by the proposed method is closer to the actual data than the graph generated by the IRDInterpolation method. This means that the 3D map estimates the stochastic component of the time series data better than the IRDInterpolation method.

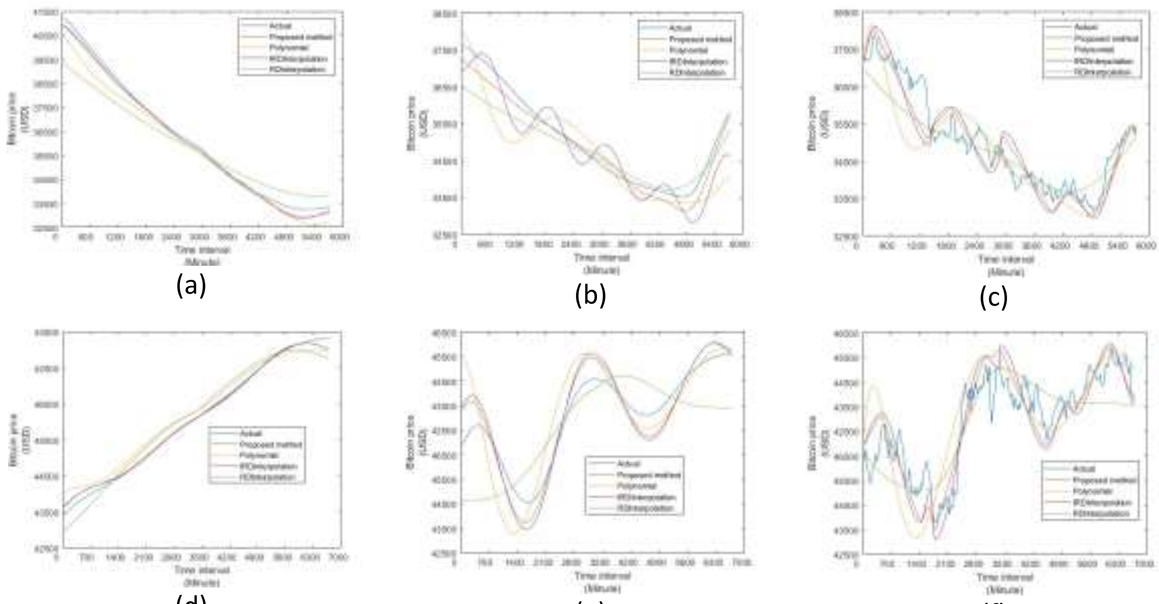


Figure 2. (a), (b), and (c) illustrate the trend component of the time series data, the combination of seasonal and trend component of the time series data, the time series data, respectively, for Dataset 1 using the actual data (blue), the proposed method (red), the polynomial function (yellow), the IRDInterpolation method (purple), and the RDInterpolation method (green). (d), (e), and (f) display the trend, combined seasonal and trend, and time-series data for Dataset 2 and using different methods, according to the colour code shown in Figure 2, respectively.

The results in Table 1 and Figure 2 confirm the high capability of the proposed model to analyse time series data and also interpolate time series data without setting any polynomial functions, which are usually applied by the conventional interpolation methods.

4. Conclusions and future works

This paper proposed a novel approach for analyzing and visualizing the time series data observed in a virtual geographic environment using spatialization. The method used the temporal elements of Bitcoin prices first to transfer data into a 2D map and then to create a 3D surface. In this study, we used the NNI algorithm to estimate the price of unknown points in 3D maps based on spatial relationships between known and unknown feature points.

The results of the generated 3D map indicate the reliability of the proposed method not only for analyzing and visualizing time series data in a meaningful and straightforward manner, but also interpolating the time series data without setting any polynomial functions. Using spatial tools, such as hillshade mapping or visibility mapping to implement financial concepts, e.g., modelling volatility or predicting return prices, would be an exciting area for future research.

References

1. Sebastião, H., Godinho, P. Forecasting and trading cryptocurrencies with machine learning under changing market conditions. Financ Innov 7, 3 (2021).
2. Mingxi Liu, Guowen Li, Jianping Li, Xiaoqian Zhu, Yinhong Yao, Forecasting the price of Bitcoin using deep learning, Finance Research Letters, Volume 40, 2021,101755.
3. Xiaolei Sun, Mingxi Liu, Zeqian Sima, A novel cryptocurrency price trend forecasting model based on LightGBM, Finance Research Letters, Volume 32, 2020.
4. Cohen, G. Forecasting Bitcoin Trends Using Algorithmic Learning Systems. Entropy 2020, 22, 838. https://doi.org/10.3390/e22080838.
5. Skupin, A., and Fabrikant, S. Spatialization Methods: A Cartographic Research Agenda for Non-Geographic Information Visualization. *Cartography and Geographic Information Science*. 2003, 30 (2): 99-119.
6. Moore A. B. & Bricker M. ‘Mountains of work’: spatialization of work projects in a virtual geographic environment, *Annals of GIS*, 21:4, 2015, 313-323.