

Discovering Intraday Price Patterns by Using Hierarchical Self-Organizing Maps

Chueh-Yung Tsao Chih-Hao Chou

Dept. of Business Administration, Chang Gung University

Abstract

Motivated from the financial literature about the intraday trading behavior, we use the hierarchical self-organizing maps to detect the price patterns during three trading periods, namely, the opening, the middle, and the close of the market. It is found from the empirical study that the three trading periods exhibit their unique characteristic. Furthermore, the intraday patterns in the opening and in the close are related.

Keywords: Intraday Data, Chart Analysis, Hierarchical SOMs

1. Introduction

During the trading period, any information about the market would be disclosed via trading activity. The trading activity is more prominent during the opening and the close of the market. It is found from the market microstructure literature that the trading activity is more frequent during the opening and close trading periods, results in the U-shape pattern of the intraday volatility structure (Wood, McInish, and Ord, 1985; Harris, 1986; Jain and Joh, 1988; McInish and Wood, 1990; Chan, Chan, and Karolyi, 1991).

From a practical viewpoint, professional traders like to trade at the opening in order to react the new information and modify their positions before the market closed in order to reduce the overnight risk. During the middle trading period, they examine the market and collect any

useful information. On the other hand, the market situations at the opening and at the close are usually in the headline of everyday financial newspaper.

Based on the above analysis, the unbalanced trading activity and different trading behavior might result in different price behaviors in the opening, the middle, and the close of the market. In this paper, we would like to study the intraday price patterns detected from different trading periods and see if these patterns exhibit their unique characteristic. The method we used to detect the intraday patterns is the trajectory-domain model proposed by Chen and He (2003) and Chen and Tsao (2003).

Chen and He (2003) are the first to use self-organizing maps (SOMs) to search for and identify interday price patterns. In their model, a geometric or trajectory pattern of the price series is considered to be a feature. Such a model is referred to as the trajectory-domain model. Chen and Tsao (2003) applied the same architecture and conduct a more rigorous statistical analysis of the discovered patterns.

The SOMs proposed by Kohonen (1982) are a special class of artificial neural networks. The SOMs are used for unsupervised learning to achieve auto classification, data segmentation or vector quantification. Unlike the supervised artificial neural networks, SOMs do not require the users to know in advance the exact objects that they are looking for. This convenience is particularly important when one can only effectively recognize some patterns by visual inspection rather than based on mathematical de-

scriptions. It is in this sense that the trajectory-domain model employs the SOMs as a tool to build the patterns.

This study modifies the trajectory-domain model by using a variant of SOMs, the hierarchical SOMs, due to the reason that one must subjectively decide the number of patterns (the map size in SOM) when implementing the trajectory-domain model. However, the number of patterns might differ from different markets and should be decided by the market itself. The advantages of the hierarchical SOMs are twofold. The first is that the map size and the map structure are automatically determined. The second is that the detected patterns can be presented on a hierarchical structure.

The rest of this paper is organized as follows. Section 2 gives a brief introduction to the SOMs and the hierarchical SOMs proposed in this study. The empirical analysis is presented in Section 3. We conclude the paper in Section 4.

2. Methodologies

2.1. Self-Organizing Maps

The SOMs can be understood in different disciplines: the SOM training algorithm resembles the classical vector quantization (in signal processing), SOM is one of the unsupervised learning algorithms (in pattern recognition), and the SOM can be used for data clustering (in statistical multivariate analysis). In fact, the training algorithm of SOMs is unsophisticated and intuitive. After training, the SOMs on the one hand compress the data in an ordered manner, and on the other hand preserve the topology structure of the distribution of the data.

Consider a network (map) with k neurons and an input data set X with p vectors. In the training process, for an input vector $x \in X$, the weights of the winning

neuron and its close neighbors are updated according to (1),

$$v_j(n+1) = v_j(n) + \eta(n)\pi_{j,i(x)}(n)[x - v_j(n)] \quad (1)$$

where $v_j(n)$ is the weight vector of the j th neuron at the n th iteration, $\pi_{j,i(x)}(n)$ is the neighborhood function (to be defined below) of node indices j and $i(x)$,

$$i(x) = \arg \min_j \|x - v_j\|, j = 1, 2, \dots, k \quad (2)$$

and $\eta(n)$ is the learning rate at iteration n . The typical neighborhood function is the Gaussian form,

$$\pi_{j,i(x)}(n) = \exp\left(\frac{-d_{j,i(x)}^2}{2\sigma^2(n)}\right) \quad (3)$$

where $d_{j,i(x)}$ is the distance between node units j and $i(x)$ on the map grid, and $\sigma(n)$ is some suitably chosen, monotonically decreasing function of iteration times n . Here, the effective width σ decays with n linearly according to (4).

$$\sigma(n) = \sigma_0 + \frac{(\sigma_1 - \sigma_0)}{N - 1}(n - 1) \quad (4)$$

where σ_0 and σ_1 are constants ($\sigma_0 > \sigma_1$) and N is the total number of epochs (to be defined below).

Some measures are frequently used for the quality of the map. One of the most commonly used measures is the average quantization error (*aqe*) which is simply the average distance from each input vector to its best matching neuron, i.e.

$$aqe = \frac{cqe}{p} \quad (5)$$

where *cqe* is the cumulative quantization error which is defined as

$$cqe = \sum_{x \in X} \|x - v_{i(x)}\| \quad (6)$$

The *cqe* can be regarded as the variation of the input vectors due to clustering. The lower the *cqe* (*aqe*) is, the higher degree of the clustering phenomenon exists in the data.

2.2. Hierarchical Self-Organizing Maps

We have introduced the basic training algorithms for the SOMs in Section 2.1. It can be denoted as the basic SOM. In the literature, however, there are enormous variants of the SOMs developed in order to increase the efficiency, convergence, elasticity, and the capability to resolve some specific problems. For example, Dittenbach, Merkl, and Rauber (2000) propose the Growing Hierarchical Self-Organizing Map (GHSOM). The GHSOM on the one hand is a data-driven architecture and on the other hand gives an intuitive representation of hierarchical relations on the data.

In this study, we propose alternative SOMs for the intraday patterns discovery. The method is based on the hierarchical structure of the maps. We first use basic SOMs to obtain a map with the best map size (number of neurons) and the best structure of the map. We denote the map as the first-layer map. We then determine which neuron on the first layer should be further partitioned. If some neuron satisfies the pre-specified criterion, all input vectors belonging to that neuron will be clustered again by using the basic SOM. We then obtain the second-layer map. This procedure is continued until all of the neurons are not judged as the one which needs to be further clustered.

There are two issues needing to be clarified. The first is how to determine the better map size and the structure of the map. The second is how to determine which neuron should be further partitioned. For the first issue we notice that

there is an inverse relationship between the number of neuron and the quality of the map. It is easy to see that when the number of neurons (k) is larger than or equals to the number of input vectors (p), the average quantization error (*aqe*) reaches its smallest level and would be zero. This is not desirable because the over-fitting phenomena may occur. Consider the other extreme case where the number of neurons is one. Then the average quantization error (*aqe*) is largest and would approach the variance of the distance between the input vectors and their midpoint. This is also not in demand since patterns, if there is any, would be averaged out in this case. A better choice of the map size would be the one that balances between the explanation and the exploitation of the model. In this study, we use the following criterion (C_1), which is similar to the famous Akaike information criterion (Akaike, 1973), to determine the number of neurons,

$$C_1(k) = \ln(aqe^2) + \frac{2k\alpha}{p} \quad (7)$$

where α is a positive real number and is used as the penalty multiplier which equalizes the importance between the *aqe* and the map size. The map of size k with the lowest $C_1(k)$ is regarded as the best map size in this study.

Given on a map size, however, it might have several different map structures. For example, a two-dimension map with 9 neurons has three different structures, namely, the 1×9 map, the 9×1 map, and the 3×3 map, respectively. In this study, we consider all possible structures for each map size and the final choice to represent that size is the one with the smallest *aqe*.

The second issue concerns with the expansion of the map. It is intuitive to expand a neuron if it has a relative high degree of inner inconsistency, i.e. high *cqe*.

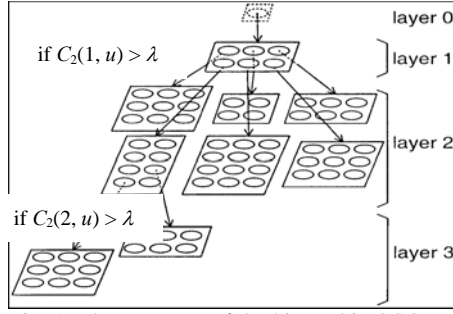


Fig. 1: The structure of the hierarchical SOM.

Let cqe_0 denote cumulative quantization error when only one neuron is used,

$$cqe_0 = \sum_{x \in X} \|x - \bar{x}\| \quad (8)$$

and $cqe_{l,j}$ denotes cumulative quantization error for neuron j on layer l ,

$$cqe_{l,j} = \sum_{\{x \in X | \hat{v}(x) = j\}} \|x - v_j\|. \quad (9)$$

It can be seen that cqe_0 is always not smaller than the summation of all the $cqe_{l,j}$ which is on the end-layer maps. Then the cqe_0 can be regarded as the total variation of the input vectors and the $cqe_{l,j}$ can be regarded as the partial variation due to clustering. The relative variation (C_2) of neuron u on layer l is defined as

$$C_2(l, u) = \frac{cqe_{l,u}}{cqe_0} \quad (10)$$

A neuron u is determined to be further partitioned if $C_2(u)$ is larger than a pre-specified level (λ). Fig. 1 is the structure of the hierarchical SOM.

The hierarchical SOM we propose in this study is similar to the GHSOM (Dittenbach, Merkl, and Rauber, 2000). However, from the experiment results not shown here we find GHSOM generates too many or too small patterns when applying to the intraday price data no matter

how we fine-tune the parameters in GHSOM.

3. Empirical Study

We apply the hierarchical SOM to two financial markets, namely, the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) and the Taiwan Futures Exchange (TAIFEX). The sample period covers from 1/2/2001 to 2/27/2007. The 1-minute data is collected in order to discover the intraday patterns. During a trading day, the first, the middle, and the last 30 minutes data are used as the sample to examine the opening, the middle and the close market patterns, respectively. We then implement six (2 markets and 3 intraday trading periods) experiments in this study.

For each experiment, the basic SOM combined with the criterion (7) is used to obtain a first-layer map. The penalty multiplier α in (7) is set to be 40 for the first layer construction. The first-layer results for the four experiments are different. We detect 1×8 , 1×8 , and 9×1 map structures for the opening, the middle, and the close of the TAIEX, respectively. In addition, we detect 5×1 , 5×1 , and 3×2 map structures for the opening, the middle, and the close of the TAIFEX, respectively. The basic SOM combined with the criterion (7) is used again for the neurons whose C_2 is lower than the threshold value λ . The value λ is set to be 3.5% in this study. It is found from the empirical study that almost all of the neurons on the first layer need to be extended further. One exception is the fourth neuron on the map of the close of the TAIEX.

It should be noticed that the penalty multiplier α for the second layer construction must be lower than that for the first layer in order to allowing for finer clustering in the second layer. The penalty multiplier α is set to be 12 for the second layer construction in this study.

Table 1: Map qualities on different trading periods.

TAIEX			
	Opening map	Middle map	Close map
Opening data	0.7036	0.8763	0.8113
Middle data	0.8053	0.7907	0.7592
Close data	0.8613	0.8670	0.7028
TAIFEX			
	Opening map	Middle map	Close map
Opening data	0.8721	0.9247	0.9731
Middle data	1.5467	0.8812	0.9638
Close data	1.5722	0.8981	0.8106

After the construction of the second layer maps, none of the second-layer neuron needs to be further partitioned. The four experiments stop at the second layer. Fig. 2 is the first-layer map of the TAIEX opening. Fig. 3 is the second-layer map of the TAIEX opening. Due to the paper size we only present the maps for the opening of TAIEX. Other results are available from the authors upon request.

Table 1 represents the map qualities in different markets and trading periods. Take the TAIEX as an example, the *aqe* for the opening, the middle, and the close of the market are 0.7036, 0.7907, and 0.7028, respectively. Given on the same setting of the experiments, it tends to have a higher quality map for the opening and the close of the market than that for the middle, which confirms the hypothesis of more information contained in the opening and close trading periods. Based on the higher quality maps in TAIEX than those in TAIFEX, the patterns in the spot market seem to be more apparent than those in the future market.

We further study the uniqueness of the patterns in different trading periods by measuring the map quality for one trading period, but the map is trained using another trading period. What we do is firstly apply the trajectory-domain model to one specific trading period and obtain the patterns. We then apply this trained trajec-

tory-domain model to another trading period and evaluate the performance of the model. If the two series have similar patterns, it is expected that the model will have a low *aqe* even if the model is borrowed from the other period.

Table 1 also presents the results. Take the TAIEX as an example, the qualities of the opening map in the middle and close trading period are 0.8053 and 0.8613, respectively. It is trivial to find that each map is most successfully applied to the trading period which is used to train the map. We therefore are interested in the one which has the lowest *aqe*. The general finding from Table 1 is that the least trading period for the opening map is the close. The least trading period for the close map is the opening. This finding suggests that there is a distinct price behavior between the opening and the close of the market.

From the above analysis, we have evidence to show that the intraday price pattern is quite different between the opening and the close of the market. We then study if there is any relationship between the two successive trading periods. This is motivated by the fact that the atmosphere caused by an extreme event usually envelops the market all of the day. The investors might take one specific action in the opening and one another action in the close under such environment, which results in a relationship between the opening pattern and the close pattern. If no information arrives during a day, only liquidity or noise traders are in the market. Their trading behaviors might also reflect some specific price pattern. On the other hand, the economic of Taiwan largely relies on export. Other stock markets condition might influence Taiwan. Due to the time differences between different stock markets, the market closes of other countries might affect Taiwan opening. Therefore, the investors would take actions at the opening based on both the closes of

Table 2: Pearson χ^2 test of the trading periods.

TAIEX			
	Test Stats.	df	p-value
Opening-to-Close	75.18	56	0.04**
Close-to-Opening	69.68	56	0.10*
TAIFEX			
	Test Stats.	df	p-value
Opening-to-Close	13.24	20	0.87
Close-to-Opening	22.68	20	0.30

'df' denotes the degree of freedom of the test.

** denotes the significance under 10% level.

*** denotes the significance under 5% level.

other markets and the action they took last day.

We first study if one day's opening and its following close are related. Under the trajectory-domain model considered in this study, one day's opening and its close are related if the closed pattern which follows some specific open pattern is not purely arbitrary. We then study if one day's close pattern causes next day's opening pattern. The two alternative hypotheses we face are

H_{11} : The opening pattern and the following close pattern are not independent.

H_{12} : The close pattern and the successive overnight opening pattern are not independent.

The two hypotheses can be tested by using the Pearson χ^2 test. Table 2 is the result.

It is found from Table 2 that neither the intraday opening-to-close relationship nor the overnight close-to-opening relationship is significant in the TAIFEX. The TAIEX, on the other hand, shows some dependence on both the intraday opening-to-close relationship and the overnight close-to-opening relationship. While the TAIEX is a market index which reflects the overall market situation and can not be traded directly, the TAIFEX is a tradable market and the market involves many individual and institution investors which trade the futures in order to hedge

their positions and make any arbitrage opportunity. The information in the TAIFEX market could easier be discovered and understood by the market. Therefore, the TAIFEX would more efficiently reflect the information, which results in that the price patterns are independent between the two trading periods.

4. Conclusion

This study contributes to the literature via two viewpoints. First, we modified the trajectory-domain model (Chen and He, 2003; Chen and Tsao, 2003) by proposing the hierarchical SOMs. The method can not only determine the map size and map structure automatically, but also obtain a map based on a hierarchical structure. Second, we apply the modified trajectory-domain model to TAIEX and TAIFEX in order to detect the intraday patterns during the opening, the middle, and the close of the market. It is found that the three trading periods exhibit their unique characteristic. Furthermore, the intraday patterns in the opening and in the close are related. The future research will focus on how the patterns in different trading periods reveal trading signals.

References

- [1] H. Akaike, "Statistical Predictor Identification," *Annals of the Institute of Statistical Mathematics* 22(1), pp. 203-217, 1973.
- [2] K. Chan, K.C. Chan and, G.A. Karolyi, "Intraday Volatility in The Stock Index and Stock Index Futures Markets," *Review of Financial Studies* 5, pp. 657-684, 1991.
- [3] S.-H. Chen and H. He, "Searching Financial Patterns with Self-Organizing Maps," in S.-H. Chen and P. P. Wang (Eds), *Computational Intelligence in Economics and Finance*, Springer, 2003.

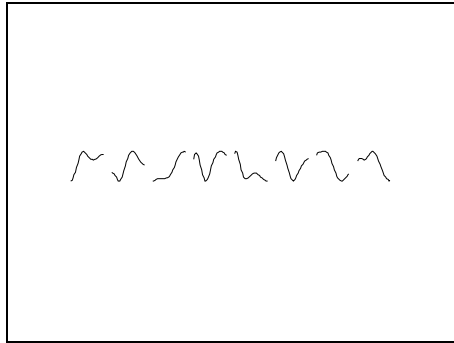


Fig. 2: The first-layer map of the TAIEX opening.

- [4] S.-H. Chen and C.-Y Tsao, "Financial Modelling Based on the Trajectory Domain," *Working paper*, 2003.
- [5] M. Dittenbach, D. Merkl and A. Rauber, "The Growing Hierarchical Self-Organizing Map," In S.-I. Amari, C. L. Giles, M. Gori, and V. Piuri(Eds.), *Proceedings of the International Joint Conference on Neural Networks(IJCNN 2000)*, Vol. 6, pp. 15-19, 2000.
- [6] L. Harris, "A Transaction's Data Study of Weekly and Intradaily Patterns in Stock Returns," *Journal of Financial Economics* 16, pp. 99-117, 1986.
- [7] P.C. Jain and G.H. Jon, "The Dependence Between Hourly Prices And Trading Volume," *Journal of Financial and Quantitative Analysis* 23, pp. 269-283, 1988.
- [8] T. Kohonen, *Self-Organizing Maps*, 3rd Edition, Springer, 2001.
- [9] T. H. McInish and R. A. Wood," An Analysis of Transaction Data for the Toronto Stock Exchange: Return Patterns and End-of the Day Effect," *Journal of Banking and Finance* 14, pp. 441-458, 1990.
- [10] R.A. Wood, T.H. McInish, and J.K. Ord, "An Investigation of Transactions Data for NYSE Stocks," *Journal of finance* 40, pp. 723-741, 1985.

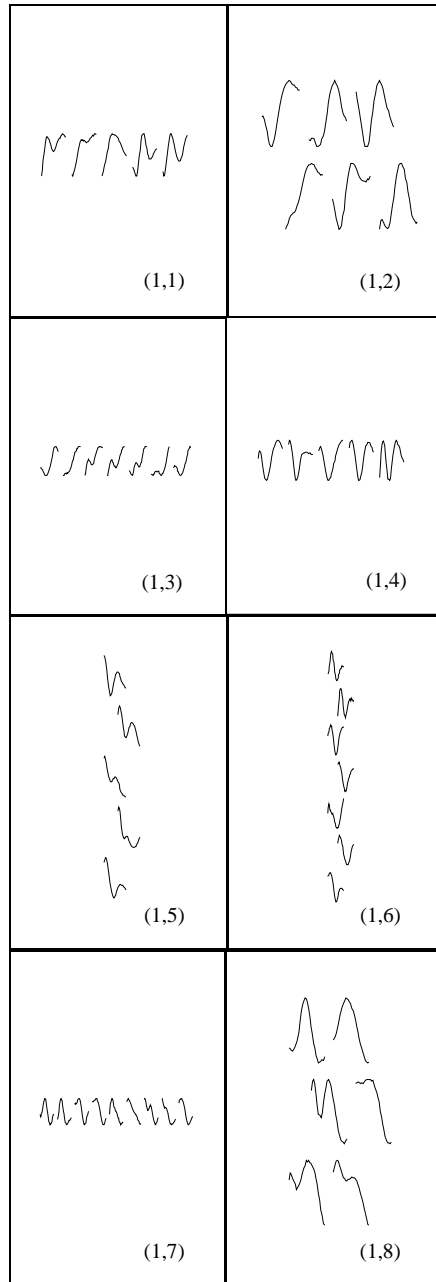


Fig. 3: The second-layer map of the TAIEX opening.