The Self-Organizing Maps for Seasonality Adjustment (SOM): Application to The Euro/Dollar Foreign Exchange Volatility and Quoting Activity

Walid Ben Omrane[1] Eric de Bodt[2,3]

September 6, 2004

Abstract

The existence of intraday seasonality in the foreign exchange volatility and quoting activity was confirmed by several research. To adjust raw data from seasonality many studies adopt the intradaily average observations model and the smoothing techniques to remove the day of the week effect. These methods do not remove all seasonal component, they leave a lot of bias in dependent variables. We show, however, that using the self-organizing maps model (SOM) instead is more efficient in terms of deseasonalization. SOM(p,q) is based on neural network learning and nonlinear projections and its flexibility allows to eliminate a lot of seasonality and to obtain much more adjusted data.

Keywords: foreign exchange market, self-organizing maps, seasonality, high frequency data.

JEL Classification : C13, F31

[1]Department of Business Administration (IAG), Finance Unit, Catholic University of Louvain, Place des Doyens 1, 1348 Louvain-la-Neuve, Belgium, e-mail: benomrane@fin.ucl.ac.be. Tel: +32 10 47 84 49, Fax: +32 10 47 83 24.

[2]Department of Business Administration (IAG), Finance Unit, Catholic University of Louvain, Place des Doyens 1, 1348 Louvain-la-Neuve, Belgium, e-mail: debodt@fin.ucl.ac.be. Tel: +32 10 47 84 47, Fax: +32 10 47 83 24.

[3]ESA, University of Lille 2, Place Dliot, BP 381, Lille F-59020, France.

The authors would like to thank Luc Bauwens and Christian Gouriéroux for helpful discussions and suggestions. Work supported in part by the European Community Human Potential Programme under contract HPRN-CT-2002-00232, Microstructure of Financial Markets in Europe. The usual disclaimers apply.

1 Introduction

Evidence for the intraday seasonality in the foreign exchange market was showed by several studies, recently by Degennaro and Shrieves (1997), Andersen and Bollerslev (1998), Melvin and Yin (2000), Cai, Cheung, Lee, and Melvin (2001), Bauwens, Ben Omrane, and Giot (2003) and Ben Omrane and Heinen (2004). All of them focus their investigations on FOREX volatility and quoting activity. They suggest two categories of methods in order to remove the seasonal component from endogenous variables. Some studies like Andersen and Bollerslev (1998), Cai, Cheung, Lee, and Melvin (2001) and Ben Omrane and Heinen (2004) adopt a linear projection technique to get rid of seasonality. They regress endogenous variables, including seasonal component, on exogenous ones in addition to dummies variables or flexible Fourier form, i.e. a sum of sinusoids, to detect intraday cycles. Some others adjust raw data from seasonality using a direct adjustment factors computed through the intradaily average observations adjustment (Dacorogna, Muller, Nagler, Olsen, and Pictet, 1993, Degennaro and Shrieves, 1997, Eddelbuttel and McCurdy, 1998, Melvin and Yin, 2000, and Bauwens, Ben Omrane, and Giot, 2003) or the smoothing techniques (Engle and Russell, 1998, Bauwens and Giot, 2000 and Veredas, Rodriguez-Poo, and Espasa). All of these methods eliminate a part of the cyclical component from endogenous variables and keep only the stochastic one. The more the data involve a regular seasonalily, the more the previous methods present a success in the deseasonalization process.
With respect to the previous literature on FOREX seasonality, the aim of this paper is twofold. Firstly we present a model based on self-organizing maps algorithm defined by Kohonen (1995). We adapt this model in order to deal with seasonality. The self-organizing maps (SOM), is introduced in the following study to capture cyclical components and purge endogenous variables from the seasonal component specially when it presents unregulated seasonality. The flexibility of the model is governed by two parameters, (p and q), which define its dimension.
Secondly we compare SOM model to the intradaily average observations model and to the smoothing technique, both used in the literature. All of these models adjust endogenous variables from seasonality. Indeed, we compare the results generated by implementing these models, in terms of the autoregressive correlation function (ACF) of deseasonalized variables. Our empirical framework is build on a high frequency data set of 5-minute regularly time-spaced Euro/Dollar quotes. The time period ranges from May 15, 2001 to May 15, 2002.

Our results show that the SOM(p,q) model is more relevant than both the intradaily average observations model (IAOM) and the Nadaraya-Watson kernel smoothing method. With SOM model and for a given p and q, we obtain almost the same results provided by the IAOM, in particular when we deal with five classes (i.e. (p,q)=(1,5)) which could be equivalent to the five days of the week. But SOM flexibility, through its parameters p and q, allows to go beyond the limit of IAOM and the smoothing method and improves their outputs. However, all of the three models implemented in the study do not succeed to remove totally the day of the week effect from the endogenous variables, which confirm the strongness of this seasonal component. The remaining day of the week effect is almost captured by the deterministic part of some categories of news announcement variables (Andersen and Bollerslev, 1998 and Bauwens, Ben Omrane, and Giot, 2003).

The follow-up of this paper is divided in four sections. In Section 2 we present a brief review of literature related to FOREX seasonality. In Section 3 we describe our data. We present our models and we discuss the results in Section 4. We conclude in Section 5.

2 The Foreign Exchange Seasonality

The literature on FOREX volatility stresses that the market openings/closings, news announcements and some days of the week lead to significant cyclical factors in the modelling of volatility and quoting activity (Bollerslev and Domowitz, 1993; Andersen and Bollerslev 1996, 1998; Degennaro and Shrieves, 1997; Melvin and Yin, 2000; Bauwens, Ben Omrane, and Giot, 2003 and Ben Omrane and Heinen, 2004). For instance, Andersen and Bollerslev (1998) and Bauwens, Ben Omrane, and Giot (2003) show that scheduled news announcements have a seasonal impact on volatility. These news events exhibit both a cyclical and a stochastic component, the latter being the news not properly anticipated by the market participants and thus is the 'surprise' effect.

To remove the seasonal effects and to highlight the stochastic components, it is first necessary to detect and identify the factors that are likely to generate periodic effects. A number of methods have been recently put forward in the literature. Degennaro and Shrieves (1997), Melvin and Yin (2000) and Bauwens, Ben Omrane, and Giot (2003), amongst others, introduce the intradaily average observations model to adjust volatility and quoting activity variables from seasonality. They divide returns by the square root of the cross sectional average volatility (square returns) in order to adjust conditional volatility from the cyclical component, and they divide quoting activity by their cross sectional average to keep only the stochastic component (this method is more detailed in Section 4.2). The more the data involve a regular seasonalily, the more IAOM presents a success in the deseasonalization process. Degennaro and Shrieves (1997) use, in addition, dummy variables to allow for separate effects of each hour of the day on volatility but they do not discriminate between different days of the week. However, Andersen and Bollerslev (1998) and Bauwens, Ben Omrane, and Giot (2003) show that days of the week have different significant effects. To deseasonalize volatility, they allow for separate effects for each day of the week, but they assume that the same day of the week keep a constant effect through different weeks.
In order to model the seasonal impact, Andersen and Bollerslev (1998), Cai, Cheung, Lee, and Melvin (2001) and Ben Omrane and Heinen (2004) introduce the flexible Fourier form, i.e. a sum of sinusoids, to detect intraday cycles. Dacorogna, Muller, Nagler, Olsen, and Pictet (1993) and Eddelbuttel and McCurdy (1998) deseasonalize volatility using an adjustment factor. This factor is proportional to the (mean) absolute value of the returns over a time interval divided by the size of the time interval.
On the other hand Engle and Russell (1998) and Bauwens and Giot (2000) adjust duration variables from seasonality using the cubic splines technique. Veredas, Rodriguez-Poo, and Espasa (2002) adopt the kernel estimator to adjust duration and show that their method is more relevant than the cubic splines. However, both cubic splines and kernel estimator are based on smoothing technique.
To summarize, there are two categories of seasonality adjustment methods used in the literature. The first one is an indirect adjustment and it consists in removing seasonality through a regression in which there is the seasonality capture like dummies variables or the flexible Fourier form. The second category is a direct adjustment. It involves some methods which

adjust seasonality directly from the raw data. This category involves the smoothing techniques, for instance, cubic splines and kernel estimator, the adjustment factor proposed by Dacorogna, Muller, Nagler, Olsen, and Pictet (1993) and the intradaily average observations model detailed by Bauwens, Ben Omrane, and Giot (2003).
In the following framework we focus only on the second category of methods. We compare the performance of both the smoothing technique and the intradaily average observations model to a new type of model: the self organizing maps model. This model is often used in physics application but there are few applications in finance (de Bodt, Cottrell, Henrion, and Van Wijmeersch, 1998 and de Bodt, Lendasse, Cardon, and Verleysen, 2003). The SOM model is based on neural network learning and nonlinear projection (more details are given in section 4.1). We introduce this model in order to capture unregulated seasonality which could not be the case for the previous models.
Once seasonality is captured, we remove it from the raw data, X_i, by computing x_i which is equal to $\frac{X_i}{\phi_{t_i}}$, where ϕ_{t_i} is the deterministic intraday seasonal component captured by the three models. In order to check for the adjustment quality we compute the autocorrelation function graph of x_i. In addition we extend the test to x_i^2, and $Log(x_i)$, because adjusted variables are often used to do some studies which need the adoption of models like ARCH-type and duration-type models.

3 Data Description

The Euro/Dollar foreign exchange market is a market maker based trading system, where three types of market participants interact around the clock (i.e. in successive time zones): the dealers, the brokers and the customers from which the primary order flow originates. The most active trading centers are New York, London, Frankfurt, Sydney, Tokyo and Hong Kong. A complete description of the FOREX market is given by Lyons (2001).
To compute the returns used for the estimation of volatility, we bought from Olsen and Associates a database made up of 'tick-by-tick' Euro/Dollar quotes for the period ranging from May 15, 2001 to May 15, 2002 (i.e. one year). This database includes 6,088,382 observations. As in most empirical studies on FOREX data, these Euro/Dollar quotes are market makers' quotes and not transaction quotes (which are not widely available).[1] More specifically, the database contains the date, the time-of-day time stamped to the second in Greenwich mean time (GMT), the dealer bid and ask quotes, the identification codes for the country, city and market maker bank, and a return code indicating the filter status. According to Dacorogna, Muller, Nagler, Olsen, and Pictet (1993), when trading activity is intense, some quotes are not entered into the electronic system. If traders are too busy or the system is running at full capacity, quotations displayed in the electronic system may lag prices by a few seconds to one or more minutes. We retained only the quotes that have a filter code value greater than 0.85.[2]

[1] Danielsson and Payne (2002) show that the statistical properties of 5-minute dollar/DM quotes are similar to those of transaction quotes.

[2] Olsen and Associates recently changed the structure of their HF database. While they provided a 0/1 filter indicator some time ago (for example in the 1993 database), they now provide a continuous indicator that lies between 0 (worst quote quality) and 1 (best quote quality). While a value larger than 0.5 is already deemed acceptable by Olsen and Associates, we choose a 0.85 threshold to have high quality data. We remove however almost no data records (Olsen and Associates already supplied us with data which features a filter

From the tick data, we computed mid-quote prices, where the mid-quote is the average of the bid and ask prices. As we use five-minute returns, we have a daily grid of 288 points. At the end of each interval, we used the closest previous and next mid-quotes to compute the relevant price by interpolation. The mid-quotes are weighted by their inverse relative time distance to the interval endpoint. Next, the return at time t is computed as the difference between the logarithms of the interpolated prices at times $t-1$ and t, multiplied by 10,000 to avoid small values. Volatility is computed as the square of returns.

Because of scarce trading activity during the week-end, we excluded all returns computed between Friday 21h30 and Sunday 24h. In addition, we excluded the first return of each Monday and for each day following a holiday to avoid possible biases due to the lack of activity during the week-ends and holidays. We take into consideration the day-light saving time adjustment to account for the time changes (to winter and summer time) that occurred on October 29, 2001 and March 25, 2002. This concerns GMT hours from 6h until 21h (corresponding to market times in Europe and the USA).

Next to return volatility, a second important variable is quoting activity. FOREX quoting activity, measured by the number of quotes in five minutes time interval, is considered in many papers as a proxy for volatility and in some studies as a proxy for private information. Adjustments for week-ends and holidays are computed in the same way as for returns.The total number of observations for volatility and quoting activity is 72,675.

Table 1 presents summary statistics of the Euro/Dollar returns, and quoting activity. The returns mean is almost equal to zero, their distribution has fatter tails than the normal and feature a positive skewness coefficient. The quoting activity mean and standard deviation are relatively high. However its distribution is less leptokurtic than returns but much more asymmetric.

4 Models and Empirical Results

4.1 The Self-Organizing Maps Model (*SOM(p,q)*)

The self-organizing maps (SOM) presented by Kohonen (1995) defines a mapping from the input data space Ω, of dimension K, onto a K-dimensional array of output nodes. In order to visualize the outputs, K has not to go beyond two.

Let $x \in \Omega$ be a stochastic data vector. A vector quantization φ is an application from the continuous space Ω, endowed by a continuous probability density function $f(x)$ to a finite subset F composed by n code-vectors or quantizers or classes $m_1, \ldots, m_n$. The position of the classes are supposed to be computed as a result of a "nonlinear projection" of $f(x)$ onto F through a learning algorithm. The aim of the projection is to compress the information by replacing all elements x of a cluster C_i (subset of Ω) by a unique class m_i.

During learning, or the process in which the nonlinear projection is formed, those nodes that are topographically close in the array up to a certain geometric distance will activate each other to learn something from the same input x. This will result in a location relaxation or smoothing effect on classes in this neighborhood, which in continuous learning leads to global ordering.

For a given neighborhood structure, where $V(i)$ denotes the neighborhood of unit i, the SOM

value larger than 0.5), as most filter values are very close to 1.

algorithm is defined as follow:

- The classes $m_1, \ldots, m_n$ are randomly initialized,
- We identify each class through its map-ordinates (r_m, c_m), where $r_m = 1, \ldots, p$ and $c_m = 1, \ldots, q$ (where $p \leq q$).[3]
- At each step t:
  - A data x_{t+1} is randomly drawn according to the density $f(x)$,
  - The winning class m_t^{win} is computed by minimizing the classical Euclidean norm:

$$\|x_{t+1} - m_t^{win}\| = \min_i \|x_{t+1} - m_i\| \tag{4.1}$$

  - The class m_t^{win} and its neighbors $m_t^k, \ldots, m_t^{k+b}$, for $k+b$ in $V(i)$, are updated by

$$m_{t+1}^{win,k} = m_t^{win,k} + \varepsilon_t (x_{t+1} - m_t^{win,k}), \tag{4.2}$$

    where ε_t is an adaptation parameter which satisfies the Robbins and Monro (1951) conditions ($\sum \varepsilon_t = \infty$ and $\sum \varepsilon_t^2 < \infty$).
  - The winning classes are identified through their map-ordinates (r_m^w, c_m^w),
- Once learning algorithm is achieved, the SOM is established and it involves Q (Q$=p$ x q) classes,
- For each observation i of the vector x corresponds a winning class $m_t^{win,k}$ identified by its map ordinates (r_m^w, c_m^w).

In our paper, Ω involves a dimension K equal to 288. Each of them presents the five minutes observations for all the study period. Indeed, the winning classes which compose the map, capture the seasonal component of the observations (in our study, observations are those of the foreign exchange volatility and quoting activity). In order to remove the fitted seasonality, we divide each observation by the corresponding element of the winning class. The correspondence is done through the ordinates. The efficiency of the deseasonalization depends on the choice the map dimension parameters p and q. For instance, $(p, q) = (1,5)$ implies that we consider five (1x5) components in the seasonal part of variables which could be equivalent to the different days of the week. If we consider ten seasonal components, we have to choose between two specifications for the map-dimension parameters, either (1,10) or (2,5). However, to test the relevance for the different specification, we compute the ACF of the deseasonalized variables and we visualize its pattern to guess the adjustment of the cycles that already exists.

[3]The dimension of SOM, which depends on the array dimension K of output nodes, is governed by the two parameters p and q. If $p > 1$ and $q > 1$ then K=2, otherwise, K=1.

4.2 The Intradaily Average Observations Model (*IAOM*)

To compute the intradaily average observations at time n_k of day k (called mv_{n_k}), we divide each day into 288 five-minute intervals. We assume for simplicity that we have exactly S weeks of data. For each interval endpoint per day of the week over the S week period, we have one euro/dollar return and q observations on quoting activity. We thus compute in principle 288 values mv_{n_k} for each day of the week for both volatility ($mv^{vol}_{n_k}$) and quoting activity ($mv^{qa}_{n_k}$). Actually, as explained in Section 3, we delete the first interval of Monday and the intervals from 21h35 to 24h of Friday. Hence, we have 287 points on Monday and on the day following a holiday and 258 on Friday. Just to simplify the presentation of the model, we assume that there is no holidays, that makes a total of 1415 values over a week.

Each value $mv^{vol}_{n_k}$ and $mv^{qa}_{n_k}$ is respectively the average of the S squared returns and the average of the S observed quotes at time n_k of day k ($k = 1$ is for Monday, $k = 5$ for Friday). For example, the value of $mv^{vol}_{n_k}$ on Tuesday at 12h ($k = 4$ and $n_4 = 144$) is the average of the squared returns observed every Tuesday at 12h during the S week period. Formally,

$$mv^{vol}_{n_k} = \left(\frac{1}{S} \sum_{s=1}^{S} r^2_{f(s,k,n_k)} \right), \tag{4.3}$$

$$mv^{qa}_{n_k} = \frac{1}{S} \sum_{s=1}^{S} q_{f(s,k,n_k)}, \tag{4.4}$$

where

$$f(s,k,n_k) = 1415\,(s-1) + \sum_{j=1}^{k-1} N_j + n_k, \tag{4.5}$$

for $s = 1, \ldots, S$, $k = 1, \ldots, 5$, $N_1 = 287$, $N_2 = N_3 = N_4 = 288$, $n_1 = 2, \ldots, 288$, $n_2 = 1, \ldots, 288$, n_3 and n_4 likewise, and $n_5 = 1, \ldots, 258$ as stated above.

To adjust volatility and quoting activity for seasonality, we just divide them at the endpoint of each five minute interval by the corresponding value of the intradaily average volatility (using respectively the day-specific volatility and quoting activity). That means, for example, that all quoting activity at 12h on Thursday in the sample are divided by the same value (the average quoting activity at 12h on Thursday).

4.3 The Smoothing Method

It consists in smoothing the raw data using the Nadaraya-Watson kernel estimator and then adjust each observation by the correspondent value on the smooth curve. The adjustment is done through the division of the raw value by the smoothed one. We begin by assuming that for instance volatility,[4] r_t^2, satisfy the following equation:

$$r_t^2 = m(X_t) + \epsilon_t \ \ , t = 1, \ldots, T \tag{4.6}$$

where $m(X_t)$ is an arbitrary fixed but unknown nonlinear function of a state variable X_t, ($X_t = t$) and ϵ_t is a white noise.

For any arbitrary x, a smoothing estimator of $m(x)$ may be expressed as:

[4] We implement the same methodology for quoting activity variable.

$$\hat{m}(x) = \frac{1}{l}\sum_{j=1}^{l}\omega_j(x)r_t^2\,, \tag{4.7}$$

where the weight $\omega_j(x)$ is large for the prices r_t^2 with X_t near x and small for those with X_t far from x. For the kernel regression estimator, the weight function $\omega_j(x)$ is built from a probability density function $K(x)$, also called a kernel:

$$K(x) \geq 0 \;\; , \int_{-\infty}^{+\infty} K(u)du = 1\,. \tag{4.8}$$

By rescaling the kernel with respect to a parameter $h > 0$, we can change its spread:

$$K_h(u) \equiv \frac{1}{h}K(u/h) \;\; , \int_{-\infty}^{+\infty} K_h(u)du = 1 \tag{4.9}$$

and define the weight function to be used in the weighted average (4.7) as:

$$\omega_{t,h} \equiv K_h(x - X_t)/g_h(x) \tag{4.10}$$

$$g_h(x) \equiv \frac{1}{l}\sum_{j=1}^{l} K_h(x - X_t)\,. \tag{4.11}$$

Substituting (4.11) into (4.7) yields the *Nadaraya-Watson* kernel estimator $\hat{m}_h(x)$ of $m(x)$:

$$\hat{m}_h(x) = \frac{1}{l}\sum_{j=1}^{l}\omega_{t,h}(x)r_t^2 = \frac{\sum_{j=1}^{l} K_h(x - X_t)r_t^2}{\sum_{j=1}^{l} K_h(x - X_t)}\,. \tag{4.12}$$

If h is very small, the averaging will be done with respect to a rather small neighborhood around each of the X_t's. If h is very large, the averaging will be over larger neighborhoods of the X_t's. Therefore, controlling the degree of averaging amounts to adjusting the smoothing parameter h, also known as the bandwidth. Choosing the appropriate bandwidth is an important aspect of any local-averaging technique. In our case we select a Gaussian kernel with a bandwidth, h_{opt}, computed by Silverman (1986):

$$K_h(x) = \frac{1}{h\sqrt{2\pi}}e^{-\frac{x^2}{2h^2}} \tag{4.13}$$

$$h_{opt} = \left(\frac{4}{3}\right)^{1/5}\sigma_k\, l^{-1/5}, \tag{4.14}$$

where σ_k is the standard deviations for the observations. However, the optimal bandwidth for Silverman (1986) involves a fitted function which is too smooth. In other words this optimal bandwidth places too much weight on prices far away from any given time t, inducing too much averaging and discarding valuable information in local price movements. Like Lo, Mamaysky, and Wang (2000), through trial and error, we found that an acceptable solution to this problem is to use a bandwidth equal to 2% of h_{opt}.

4.4 Empirical Results

As mentioned above, the Euro/Dollar currency is almost continuously traded in FOREX markets that belong to different time zones, but the quoting activity (measured by the number of quotes per 5 minute interval) varies a lot over the 24 hours.

Figure 1 illustrates the intradaily seasonal pattern of the average volatility and quoting activity for five-minute time intervals. The top panel of the figure shows the average volatility computed by implementing respectively equations (4.3). It shows some components which exist more or less every day, although days differ from each other.

Volatility increases after midnight, i.e. at the opening of the Singapore and Hong Kong markets, one hour after the opening of the Tokyo market and two hours after Sidney. Around 4h GMT, volatility decreases because of the lunch break in the four Asian financial markets. Thereafter, volatility increases again because trading activity resumes in the Asian markets and it reaches a local maximum around 7h-8h GMT, i.e. right after the opening of the key European markets such as London and Frankfurt. This pattern of volatility increases around the opening and closing times of the regional markets are in agreement with Admati and Pfleiderer (1988), who show that these periods are characterized by a sustained level of market activity which attracts different categories of traders. In addition, Lyons (1997) shows that, because traders have to control or close their positions at the end of every day, they increase their activity right before the closing of trading and just after the market opening to get rid of unwanted risky positions. Because of the lunch break in Europe, volatility decreases around 11h30. A rebound in volatility occurs at 12h GMT as New York opens for trading. The big spike between 12h and 13h is due to US news announcements at 12h30 on Friday. Between 12h and 16h GMT volatility is generally at its highest level due to the simultaneous activity of the American and European markets. Just before the New York lunch break (which is clearly visible in the figure around 17h GMT), there is a short volatility increase due to the closing of the European markets. Volatility increases also around 21h GMT, i.e. when the New York trading session ends. Starting at 21h GMT, a short period of stability is then observed until the opening of the Sidney and Tokyo markets, which leads to an increase in volatility.

The five day-specific panels of Figure 1 indicate that the seasonality of volatility is also dependent on the day of the week: shocks are mainly observed on Tuesday and Wednesday around 8h, 10h and 16h30, Thursday and Friday around 12h, 12h30 and 16h30. Bauwens, Ben Omrane, and Giot (2003) attribute these shocks to the pre-announcement of some categories of scheduled news. Volatility increases every Monday around 0h, 8h, 13h, and 22h, which are respectively the opening hours of the Asian, European, American, and Australian markets. These increases during the first minutes of trading of every week are linked to the control of positions: FOREX traders who accumulate customers' orders at the end of the Friday session and who could not settle their positions have to keep them during the week-end. To minimize the risk of these positions, dealers are keen on executing their remaining orders in the first minutes of the Monday session. They do so by quoting attractive prices to attract counterparts and quickly close their positions.

The intradaily average quoting activity, illustrated in the bottom panel of Figure 1, shows the cross sectional average activity during the 24-hour trading day. The day specific effect is weak for the American and European markets and it is almost absent for the Asian and Australian markets. Moreover, Monday is the least active day of the week. The figure clearly shows the significant difference between market activity in Europe and the USA, with respect to the trading activity in the Asian and Australian markets.

Using five classes instead of days, computed by implementing SOM(1,5), we obtain almost the same seasonal pattern. Figure 2 illustrates the "map" or the seasonal class pattern which presents few differences to the day of the week pattern. For instance, volatility class seasonality presents the different peaks generated by some categories of news announcements. Which means that SOM model does not under-evaluate the shocks that happen on some days of the week, despite that these latter are not taken into consideration in the SOM algorithm computation. However, quoting activity class seasonality is different from the daily one. The effect of Classes is different from a class to another which is not the case for the days of the week.

Indeed, the adjustment of raw data from seasonality using both IAOM and SOM(1,5) , generates two different outputs, mainly when we deal with quoting activity. Figures 3 and 4 show the ACF of the deseasonalized volatility and quoting activity by implementing respectively IAOM and SOM(1,5). The autocorrelation function for both deseasonalized volatility is quite the same, which is an expected result because the two methods capture almost the same seasonality for volatility. Both techniques have succeeded to remove only few seasonality component but they keep a lot of deterministic bias. However, SOM(1,5) remove much more seasonality than IAOM when we adjust quoting activity. Although, the Nadaraya-Watson smoothing method generates a result which is better than both previous methods in terms of volatility adjustment but it produces almost the same output as IAOM when we adjust quoting activity variable (see Figure 5). These results are confirmed by the autocorrelation coefficient computed at the frequencies of one, two and three days and presented in Table 2.

It is worth to pointing out that according to ACF patterns of the deseasonalized variables, using different methods, the residual seasonality presents a daily peaks. Which means that seasonality is involved mainly by the day of the week effect. This is an evidence for the strongness effect of different days of the week in the foreign exchange market. Which is not a surprise as the currency market is sensitive to some scheduled news announcements that happen during some specific days of the week (Andersen and Bollerslev, 1998 and Bauwens, Ben Omrane, and Giot, 2003).

Nevertheless, when we change the two parameters of the SOM we improve the results given by both the smoothing technique and SOM(1,5). We select SOM(2,5) for volatility adjustment and SOM(6.6) to get rid of seasonality component in quoting activity variable. Figure 6 shows that seasonality, for both volatility and quoting activity, is almost removed. The adjustment does not remove the whole seasonality but an important part is kicked out. The residual seasonality can be easily captured by dummies variables that represent categories of news announcements (Bauwens, Ben Omrane, and Giot, 2003). To be sure that there is no important cyclical pattern in adjusted variables, Figure 7 shows the time of the day pattern for deseasonalized volatility and quoting activity respectively by SOM(2,5) and SOM(6,6). We observe clearly that there is no specific shape for seasonality, the two patterns are quite stationary.

Finally, it is important to note that we have done the same study, in terms of seasonality adjustment using the three methods, for the square and the logarithm of both volatility and quoting activity. We find almost the same results as those of the previous study. That is why we do not create an independent subsection which shows and displays these results[5].

[5]These results are available and can be provided by the authors.

5 Conclusion

In this paper, we apply the self-organizing maps (SOM) model to adjust two FOREX microstructure variables from seasonality. This model is based on neural network learning and non linear projections. We compare the obtained results with those obtained when we deal with two other methods already used in the literature: the intradaily average observations model and the Nadaraya-Watson kernel smoothing method. We show that the flexibility of SOM model, through its dimension parameters, allows to go beyond the limits of the previews methods and improves their outputs. However, all of the three models do not succeed to remove the whole seasonality but the SOM adjustment is the best in terms of deseasonalization.

References

Admati, A., and P. Pfleiderer (1988): "A theory of intraday patterns: volume and price variability," *The Review of Financial Studies*, 1, 3–40.

Andersen, T., and T. Bollerslev (1996): "Heterogeneous information arrivals and returns volatility dynamics: uncovering the long-run in high frequency rendements," NBER Working paper 5752.

——— (1998): "Deutsche mark-dollar volatility: intraday volatility patterns, macroeconomic announcements and longer run dependencies," *The Journal of Finance*, 1, 219–265.

Bauwens, L., W. Ben Omrane, and P. Giot (2003): "News Announcements, Market Activity and Volatility in the Euro/Dollar Foreign Exchange Market," *Journal of International Money and Finance, forthcoming.*

Bauwens, L., and P. Giot (2000): "The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks," *Annales d'Economie et Statistique*, 60.

Ben Omrane, W., and A. Heinen (2004): "The Foreign Exchange Quoting Activity as an Informative Signal," *CORE Discussion Paper*.

Bollerslev, T., and I. Domowitz (1993): "Trading patterns and prices in the inter-bank foreign exchange market," *The Journal of Finance*, 4, 1421–1443.

Cai, J., Y. Cheung, R. Lee, and M. Melvin (2001): "Once in a generation yen volatility in 1998: fundamentals, intervention and order flow," *Journal of International Money and Finance*, 20, 327–347.

Dacorogna, M., U. Muller, R. Nagler, R. Olsen, and O. Pictet (1993): "A geographical model for the daily and weekly seasonal volatility in the foreign exchange market," *Journal of International Money and Finance*, 12, 413–438.

Danielsson, J., and R. Payne (2002): "Real trading patterns and prices in the spot foreign exchange markets," *Journal of International Money and Finance*, 21, 203–222.

de Bodt, E., M. Cottrell, E. Henrion, and C. Van Wijmeersch (1998): "Self-Organizing Maps for Data Analysis : an Application to the Belgium Leasing Market," *Journal of Computing Intelligence in Finance*, 6, 5–24.

de Bodt, E., A. Lendasse, P. Cardon, and M. Verleysen (2003): "Self-organizing feature maps for the classification of investment funds," *Journal of Economic and Social Systems, forthcoming.*

Degennaro, R., and R. Shrieves (1997): "Public information releases, private information arrival and volatility in the foreign exchange market," *Journal of Empirical Finance*, 4, 295–315.

Eddelbuttel, D., and T. McCurdy (1998): "The impact of news on foreign exchange rates : evidence from high frequency data," Working paper, University of Toronto.

Engle, R., and J. Russell (1998): "Autoregressive conditional duration: a new approach for irregularly spaced transaction data," *Econometrica*, 66, 1127–1162.

Kohonen, T. (1995): *Self-Organizing Maps*. Springer.

Lo, A., H. Mamaysky, and J. Wang (2000): "Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation," *Journal of Finance*, 56, 1705–1765.

Lyons, R. (1997): "A simultaneous trade model of the foreign exchange hot potato," *Journal of International Economics*, 45, 181–211.

——— (2001): *The Microstructure Approach to Exchange Rates*. MIT Press.

Melvin, M., and X. Yin (2000): "Public information arrival, exchange rate volatility and quote frequency," *The Economic Journal*, 110, 644–661.

Robbins, H., and S. Monro (1951): "A stochastic approximation model," *Annals of Mathematical Statistics*, 22, 400–407.

Silverman, B. (1986): *Density estimation for statistics and data analysis*. Chapman and hall.

Veredas, D., J. Rodriguez-Poo, and A. Espasa (2002): "On the (Intradaily) Seasonality and Dynamics of a Financial Point Process: A Semiparametric Approach," CORE DP, 2002/23.

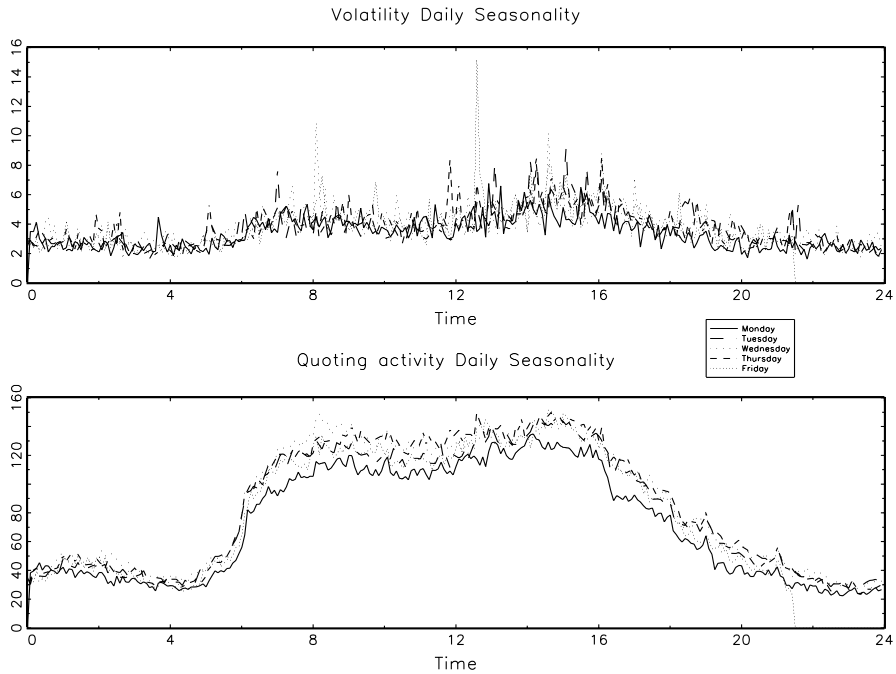

Figure 1: Intradaily, day-specific average volatilities and quoting activity (see Equations (4.3) and (4.4)).

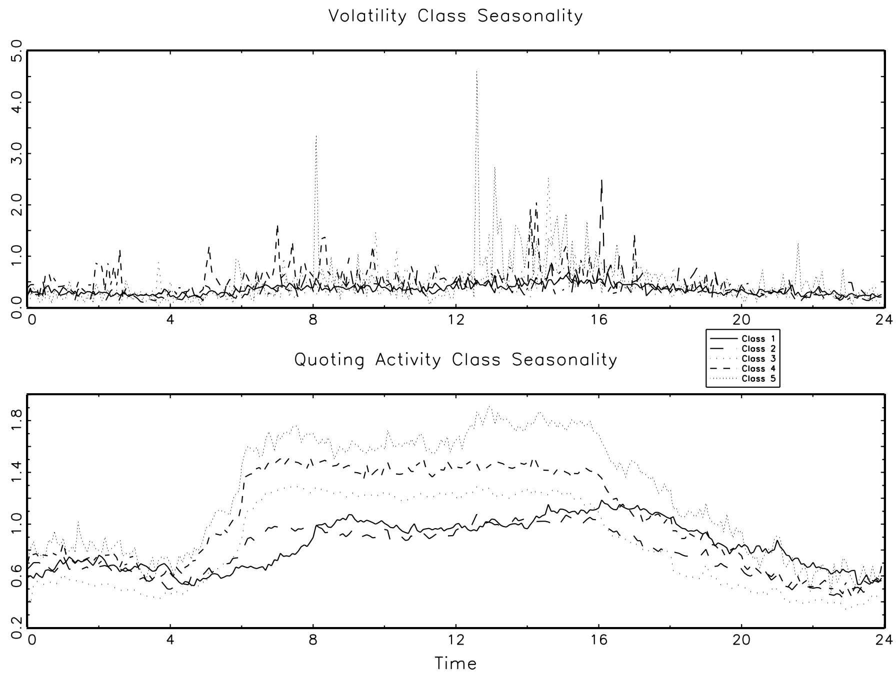

Figure 2: Five seasonal classes for volatility and quoting activity obtained through SOM(1,5) (see Equations (4.1) and (4.2)).

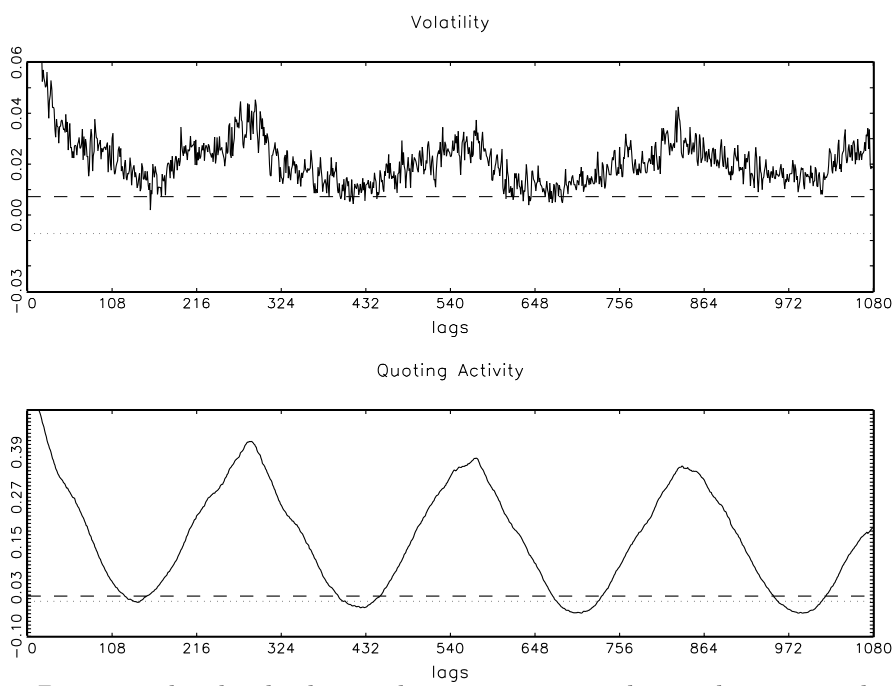

Figure 3: Deseasonalized volatility and quoting activity by implementing the IAOM.

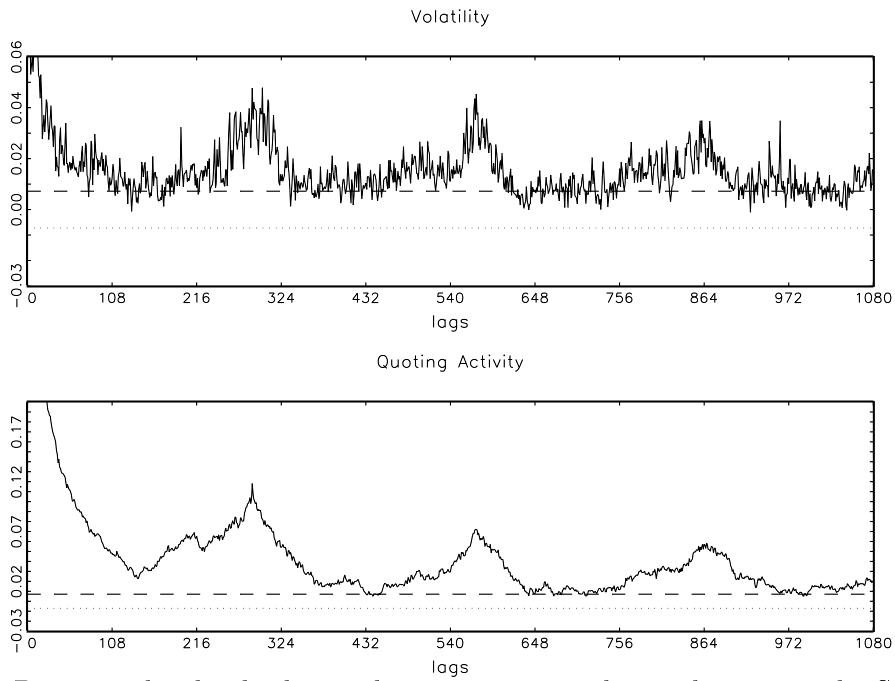

Figure 4: Deseasonalized volatility and quoting activity by implementing the SOM(1,5).

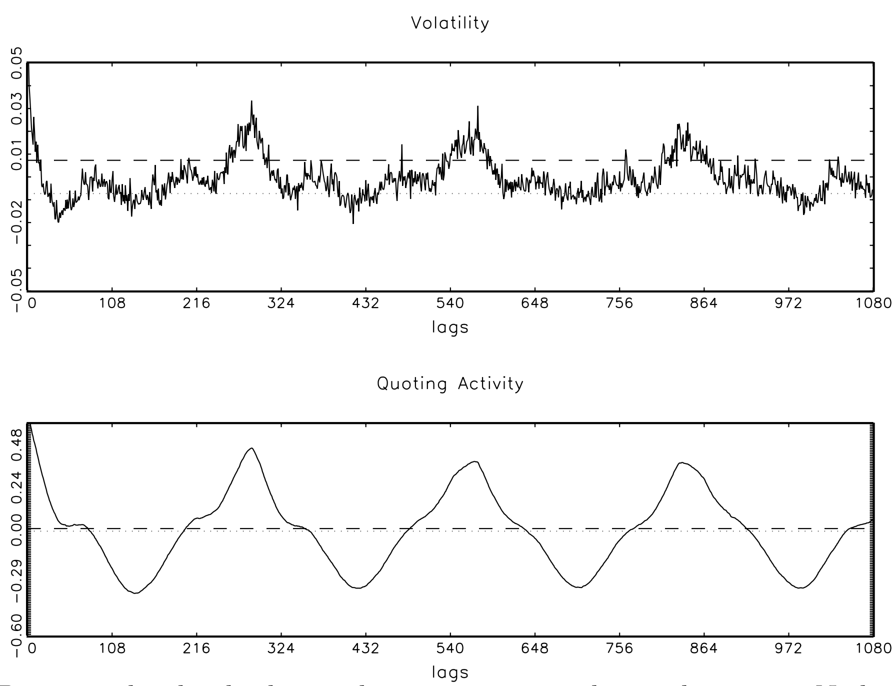

Figure 5: Deseasonalized volatility and quoting activity by implementing Nadaraya-Watson kernel smoothing method.

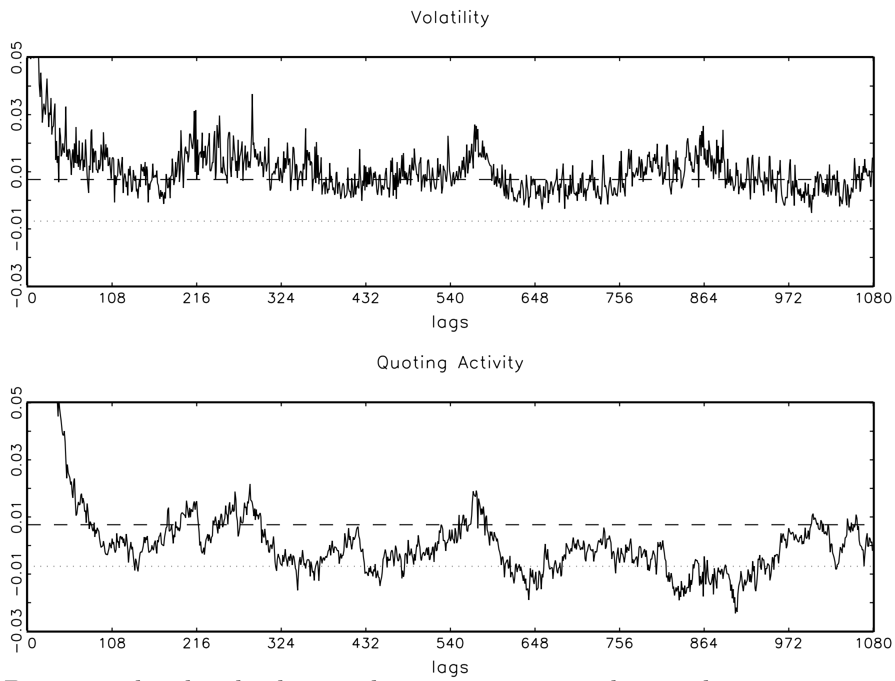

Figure 6: Deseasonalized volatility and quoting activity by implementing respectively the SOM(2,5) and the SOM(6,6).

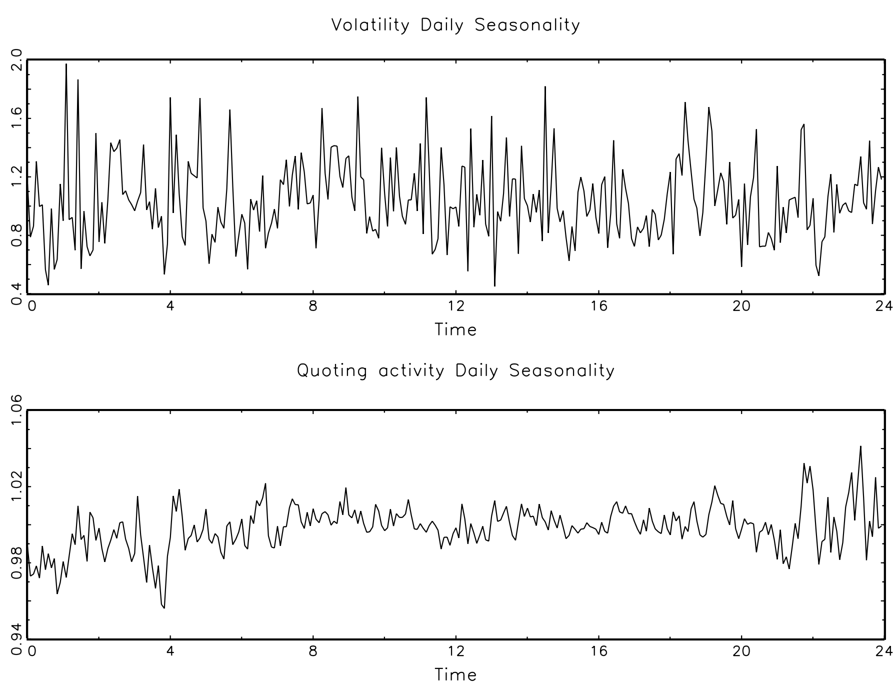

Figure 7: Time of the day pattern for deseasonalized volatility and quoting activity respectively by SOM(2.5) and SOM(6.6).

Table 1: Moments of the Euro/Dollar returns and quoting activity

	Returns	Quoting activity
Mean	0.007	82.31
Standard deviation	3.91	60.11
Skewness coefficient	0.21	1.23
Kurtosis coefficient	15.0	5.43

The 5-minute returns have been pre-multiplied by 10,000 (to avoid small values). The number of observations is 72,675, corresponding to the period from May 15, 2001 to May 15, 2002.

Table 2: Moments and autocorrelations of the Euro/Dollar deseasonalized volatility and quoting activity

	IAOM		*SOM(1,5)*		*SOM(2,5)*	*SOM(6,6)*	*Kernel*	
	V	QA	V	QA	V	QA	V	QA
μ	0.999	1.000	1.021	1.000	1.038	0.998	0.918	0.923
σ	2.180	0.554	2.584	0.449	2.629	0.387	2.045	0.440
ρ_{288}	0.037	0.417	0.048	0.119	0.037	0.014	0.027	0.461
ρ_{576}	0.032	0.372	0.038	0.067	0.025	0.006	0.031	0.382
ρ_{864}	0.030	0.320	0.030	0.054	0.026	-0.017	0.011	0.295

The number of observations is 72,675, corresponding to the period from May 15, 2001 to May 15, 2002. The seasonality adjustment was done by implementing the SOM model presented in Section 4.1, the intradaily average observations one (iaom) presented in Section 4.2 and the kernel smoothing method detailed in section 4.3.