

# Smoothing Methods for Histogram-valued Time Series An Application to Value-at-Risk

by

**Javier Arroyo**

Departamento de Ingeniería del Software  
e Inteligencia Artificial  
Universidad Complutense de Madrid  
28040 Madrid, Spain

**Gloria González-Rivera<sup>1</sup>**

Department of Economics  
University of California, Riverside  
Riverside, CA 92521

**Carlos Maté**

Instituto de Investigación Tecnológica  
Universidad Pontificia Comillas  
28015 Madrid, Spain

**Antonio Muñoz San Roque**

Instituto de Investigación Tecnológica  
Universidad Pontificia Comillas  
28015 Madrid, Spain

## Abstract

We adapt smoothing methods to histogram-valued time series (HTS) by introducing a barycentric histogram that emulates the "average" operation, which is the key to any smoothing filter. We show that, due to its linear properties, only the Mallows-barycenter is acceptable if we wish to preserve the essence of any smoothing mechanism. We implement a barycentric exponential smoothing to forecast the HTS of daily histograms of intradaily returns to both the SP500 and the IBEX 35 indexes. We construct a one-step-ahead histogram forecast, from which we retrieve a desired  $\gamma$ -Value-at-Risk forecast. In the case of the SP500 index, a barycentric exponential smoothing delivers a better forecast, in the MSE sense, than those derived from vector autoregression models, especially for the 5% Value-at-Risk. In the case of IBEX35, the forecasts from both methods are equally good.

*Key Words:* Symbolic data, Exponential Smoothing, Barycenter, High frequency data, Value-at-Risk

*Acknowledgement:* Arroyo acknowledges the financial support of Programa de Creación y Consolidación de Grupos de Investigación UCM-BSCH, GR58/08 and of the project with grant TIN2008-06464-C03-01, funded by the Spanish Council for Science and Innovation. González-Rivera acknowledges the financial support of UCR University Scholar Award and the UCR Academic Senate grants. Maté acknowledges the financial support of PRESIM project funded by Universidad Pontificia Comillas.

*JEL Classification:* C22, C53

---

<sup>1</sup>

Corresponding author: [gloria.gonzalez@ucr.edu](mailto:gloria.gonzalez@ucr.edu), tel. (951) 827 1590, fax (951) 827-5685.

# 1 Introduction

With the advent of very large data sets, there is a pressing need for the development of new statistical and econometric methodologies. Symbolic data analysis (Billard and Diday, 2006) provides a path in this direction. The key insight is that the data can be organized into objects that provide information beyond that of a simple scalar variable. Intervals and histograms are examples of symbolic data; both provide clear graphical displays to describe dispersion as in the case of intervals, or more generally as in the case of a histogram, centrality, dispersion and shape of the frequency distribution of the realizations of a random variable.

In this paper, we work with a histogram as a symbolic object. This means that some random experiment produces realizations of a random variable that are observed as frequency distributions, or in other words, the “datum” is a histogram. From a cross-sectional perspective, several authors have adapted multivariate techniques to the analysis of a set of histograms, for instance, principal components analysis (Rodríguez et al., 2000; Nagabhushan and Pradeep Kumar, 2007), and cluster analysis (Irpino and Verde, 2006; Verde and Irpino, 2007). From a time series perspective, it is possible to define a histogram-valued time series (HTS) as a collection of histograms ordered over time. A classical modelling strategy will call for evaluating the dependence of the data and searching for the most appropriate model. Unfortunately, the state of the art for HTS has not yet reached that stage. However, the absence of a model –as the basis for building a forecast– should not block the exploration of more primitive approaches for the construction of a forecast. Going back to basics, our view is that a forecast based on well-defined averages of past information should provide a sensible forecast of the future under the assumption that the world is somewhat stable. Following this argument, Arroyo and Maté (2009) adapted the k-Nearest Neighbors (k-NN) algorithm and Arroyo et al. (2010) applied this work to forecast financial aggregates. González-Rivera and Arroyo (2011) went further along these lines and implemented the empirical autocorrelation function for a HTS of financial intradaily returns. In the present work, we will focus on the analysis of smoothing techniques to forecast HTS. Though in the aforementioned papers, we sketched how to adapt smoothing filters, now we will provide a detailed analysis of the issues that arise in a forecasting environment.

On dealing with histograms, the fundamental question behind any smoothing mechanism is how to “sum” histograms. There is a well-defined arithmetic of histograms proposed by Colombo and Jaarsma (1980). Their goal was to operate with random variables for which their

“randomness” is represented by a histogram. We will argue that this arithmetic is not suitable for forecasting purposes. With any smoothing method, there are two key operations in the construction of a forecast: the addition of past information and the definition of forecast errors (subtraction operation). We provide some examples to show that, based on the arithmetic rules, smoothing histogram data produces histograms with non-desirable features.

Given the drawbacks of histogram arithmetic, we propose an alternative approach based on the notion of a barycentric histogram, which is an average of histograms. The barycenter minimizes the sum of the distances between itself and any other histogram in the set. We entertain two distance measures: Mallows and Wasserstein. Both measures evaluate the distances between the quantiles of any two histograms but Mallows uses the  $L^2$ -norm and Wasserstein uses the  $L^1$ -norm. As a result, the Mallows-barycentric histogram is formed by averaging quantiles of the histograms in the set, and the Wasserstein-barycentric histogram by finding their median. Based on the barycenter approach, we adapt smoothing methods either as moving averages or as exponential smoothing. We show that, due to its linear properties, only the Mallows-barycenter is acceptable if we wish to preserve the natural smoothness, which is the essence of any smoothing mechanism.

We offer a financial application of the smoothing techniques for histograms. For a portfolio of assets, Value-at-Risk (VaR) is a probabilistic measure defined as the loss that a portfolio will face in a given period of time with a specific confidence level. The  $\gamma$ -VaR is defined as the conditional quantile of the distribution of returns such that the probability of returns below  $\gamma$ -VaR is  $\gamma\%$ . Thus, it seems very natural to analyze histograms of returns within a symbolic data approach. The main advantage of our proposal resides on the absence of parametric assumptions on the conditional density as well as on the conditional mean and variances of the return process. We will implement a barycentric exponential smoothing to eventually construct a one-step-ahead histogram forecast, from which we will retrieve the desired  $\gamma$ -VaR. We will compare the HTS forecasts with those obtained from classical autoregressive models.

The organization of the paper is as follows. In Section 2, we set preliminary concepts and notation. In Section 3, we discuss histogram arithmetic, barycentric histograms, and the adaptation of smoothing methods. In Section 4, we predict Value-at-Risk with a barycentric exponential smoothing filter, and in Section 5, we conclude.

## 2 Preliminaries

In this section, we set the notation and review the basic notions that support the forthcoming smoothing methods.

**Definition 1.** Consider a histogram random variable  $X$ , and a sample of  $m$  units ( $i = 1, \dots, m$ ), where each  $i$  is described by a histogram such as

$$h_{X_i} = \{([x]_{i1}, \pi_{i1}), \dots, ([x]_{in_i}, \pi_{in_i})\}, \quad (1)$$

where  $\pi_{ij}$ ,  $j = 1, \dots, n_i$ , is a frequency that satisfies  $\pi_{ij} \geq 0$  and  $\sum_{j=1}^{n_i} \pi_{ij} = 1$ ; and  $[x]_{ij} \subseteq \mathbb{R}$ ,  $\forall i, j$ , is an interval (also known as bin) defined as  $[x]_{ij} \equiv [x_{Lij}, x_{Uij})$  with  $-\infty < x_{Lij} \leq x_{Uij} < \infty$  and  $x_{Uij} \leq x_{Lij}$   $\forall i, j$ , for  $j \geq 2$ .

Assuming that values in each given bin, i.e.,  $x \in [x]_{ij}$ , are uniformly distributed within the interval (Billard and Diday, 2003), the cumulative distribution function,  $H_{X_i}(x)$ , of a histogram  $h_{X_i}$  is defined as follows

$$H_{X_i}(x) = \int_{-\infty}^x h_{X_i}(x) dx = \begin{cases} 0, & \text{if } x < x_{Li1}; \\ \sum_{l=1}^{r-1} \pi_l + \frac{x-x_{Lir}}{x_{Uir}-x_{Lir}} \pi_{ir}, & \text{if } x \in [x]_{ir}; \\ 1, & \text{if } x \geq x_{Uin_i}. \end{cases} \quad (2)$$

The histogram definition is flexible enough to cover any kind of binned density estimator. Some interesting types of histogram representations are

- Equispaced histograms, i.e., histograms with bins of the same width.
- Equifrequency histograms, i.e., histograms where the frequency associated with each bin is the same. Boxplots, proposed by Tukey (1977), are a popular example of this kind of histogram offering a nice representation of the subjacent distribution (Benjamini, 1988).
- Histograms with a specific partition of the histogram domain. This representation focuses on the parts of the domain that are of interest to the researcher.
- Histograms defined according to an ordered sequence of quantiles. This representation can provide a more detailed view of some parts of the distribution such as the tails. An example of this could be offered by the sequence of quantiles .05, .1, .9 and .95.

Since our interest lies on forecasting, we need a notion of stochastic process and time series.

**Definition 2.** A histogram-valued stochastic process is a collection of histogram random variables that are indexed by time, i.e.,  $\{h_{X_t}\}$  for  $t \in T \subset \mathbb{R}$ , with each  $h_{X_t}$  following Definition 1.

**Definition 3.** A histogram-valued time series is a realization of a histogram-valued stochastic process and it will be equivalently denoted as  $\{h_{X_t}\} \equiv \{h_{X_t}, t = 1, \dots, T\}$ .

### 3 Smoothing methods for HTS

Given a histogram time series (HTS)  $\{h_{X_t}\} \equiv \{h_{X_t}, t = 1, \dots, T\}$ , our objective is to form a one-step-ahead forecast  $\hat{h}_{X_{T+1}|T}$  conditioning on the available information up to time  $T$ . A classical modelling strategy will call for evaluating the dependence of the data and searching for the most appropriate model. Unfortunately, the state of the art for HTS has not yet reached that stage. However, the absence of a model – as the basis for building a forecast – should not block the exploration of more primitive approaches for the construction of a forecast. Going back to basics, our view is that a forecast based on well-defined averages of past information should provide a sensible forecast of the future under the assumption that the world is somewhat stable. In this line, a smoothing mechanism seems to be appropriate. In this section, we propose an adaptation of smoothing methods – moving averages and exponential smoothing – to HTS.

The one-step-ahead forecast of HTS based on a moving average of order  $q$  is a weighted (or unweighted) mean of the  $q$  past histogram-valued observations

$$\hat{h}_{X_{t+1}} = \omega_1 h_{X_t} + \dots + \omega_q h_{X_{t-(q-1)}}, \quad (3)$$

where  $\omega_i$  is the weight assigned to observation  $i$ .

The one-step-ahead forecast of HTS based on an exponential smoothing filter is a weighted average of the most recent observations and its forecast

$$\hat{h}_{X_{t+1}} = \alpha h_{X_t} + (1 - \alpha) \hat{h}_{X_t}, \quad (4)$$

where  $\alpha \in [0, 1]$ .

By backwards substitution in (4), we could find an “equivalent” moving average with exponentially decreasing weights as

$$\hat{h}_{X_{t+1}} = \sum_{j=1}^t \alpha(1 - \alpha)^{j-1} h_{X_{t-(j-1)}}, \quad (5)$$

or alternatively, rearranging terms in (4), we could write the error correction equation as

$$\hat{h}_{X_{t+1}} = \hat{h}_{X_t} + \alpha(h_{X_t} - \hat{h}_{X_t}). \quad (6)$$

In all these expressions (3), (4), (5), and (6), the key operation is the “sum” of histograms that is essential to find the “average histogram”, which is the basis of any smoothing procedure. We will analyze the problem in two stages:

1. A natural proposal is to use histogram arithmetic to find the average histogram. We will argue that, within a forecasting environment, the available histogram arithmetic has some disadvantages that will be apparent in the forthcoming definitions of forecast and forecast error.

2. We propose an alternative approach based on the notion of a barycentric histogram, which is an average of histograms. This approach requires the use of distance measures. We will show that the choice of distance is crucial to the equivalence between expressions (4) and (5).

### 3.1 Histogram arithmetic

Colombo and Jaarsma (1980) proposed a histogram arithmetic that is an extension of the interval arithmetic proposed by Moore (1966), as histograms can be understood as a set of intervals with an associated weight. Though their method was geared to operate with equifrequency histograms, it is general enough to operate with any other binned density estimator. As interval arithmetic is key to understanding histogram arithmetic, first we show the basics of interval arithmetic.

Given two intervals  $[a] = [a_L, a_U]$  and  $[b] = [b_L, b_U]$ , interval arithmetic operations are given by

$$[a] + [b] = [a_L + b_L, a_U + b_U], \quad (7)$$

$$[a] - [b] = [a_L - b_U, a_U - b_L], \quad (8)$$

$$[a] * [b] = \left[ \min\{a_L * b_L, a_L * b_U, a_U * b_L, a_U * b_U\}, \max\{a_L * b_L, a_L * b_U, a_U * b_L, a_U * b_U\} \right], \quad (9)$$

$$[a]/[b] = [a] * (1/[b]), \text{ with } 1/[b] = [1/b_U, 1/b_L]. \quad (10)$$

The reciprocal of an interval,  $1/[b]$ , is defined if and only if  $0 \notin [b]$ , otherwise it is not defined because the resultant interval contains  $\infty$ . In summary, for  $+$ ,  $-$ , and  $*$ , the upper and lower bounds of the resultant interval are the maximum and minimum of all combinations of the endpoints of the interval operands. Interval addition and multiplication enjoy the associative and commutative properties but they do not satisfy the distributive property. However, the subdistributive property always

holds, that is,

$$[a]([b] + [c]) \subseteq [a][b] + [a][c], \quad (11)$$

which implies that in practice  $[a]([b] + [c])$  will be preferred because it is a tighter interval than  $[a][b] + [a][c]$ .

Colombo and Jaarsma's histogram arithmetic considers all possible pairs of intervals in any two histogram operands and combines them according to the rules of interval arithmetic. More precisely, given two histograms,  $h_A = \{([a]_i, \pi_{a,i})\}$  with  $i = 1, \dots, n$  and  $h_B = \{([b]_j, \pi_{b,j})\}$  with  $j = 1, \dots, m$ , both representing a pair of *independent* random variables  $A$  and  $B$ , and  $\square$  being some arithmetic operator in  $\{+, -, *, /\}$ ,  $C = A \square B$  can be represented by the unsorted histogram  $h_C = \{([c]_k, \pi_{c,k})\}$  with  $k = 1, \dots, n \cdot m$ , where

$$c_{L,(i-1)m+j} = \min\{a_{L,i} \square b_{L,j}, a_{U,i} \square b_{L,j}, a_{L,i} \square b_{U,j}, a_{U,i} \square b_{U,j}\}, \quad (12)$$

$$c_{U,(i-1)m+j} = \max\{a_{L,i} \square b_{L,j}, a_{U,i} \square b_{L,j}, a_{L,i} \square b_{U,j}, a_{U,i} \square b_{U,j}\}, \text{ and} \quad (13)$$

$$\pi_{C,(i-1)m+j} = \pi_{A,i} \cdot \pi_{B,j}. \quad (14)$$

As a result, the number of bins in  $h_C$  is  $n \cdot m$ , which may be potentially high, and the  $[c]_k$  intervals may overlap, which is not desirable.

We provide an instance for which a direct application of the Colombo and Jaarsma histogram arithmetic will not deliver suitable results in a smoothing environment, where the average of histograms play a key role.

The expressions (3), (4), (5) are all weighted averages of histograms. Without loss of generality, let us consider a simple average of histograms such as

$$h_{\bar{X}} = \frac{h_{X_1} + \dots + h_{X_n}}{n}, \quad (15)$$

where the value  $n$  is also considered as a histogram of the following form  $h = \{([n, n], 1)\}$ . Suppose that we consider five histograms, all of them identical to each other. We would like to obtain the average  $h_{\bar{X}}$ . Let us say that  $h_{X_i} = \{([1, 2), .1), ([2, 3), .15), ([3, 4), .25), ([4, 5), .5]\}$ ,  $\forall i$  and  $i = 1, \dots, 5$ . By applying the rules of the Colombo and Jaarsma arithmetic, the average histogram  $h_{\bar{X}}$  is given by  $h_{\bar{X}} = \{([1, 2), .002), ([2, 3), .115), ([3, 4), .615), ([4, 5), .268)\}$ . In Figure 1 we plot the five histograms and their average  $h_{\bar{X}}$ .

[FIGURE 1]

The average histogram has the same four bins as those of each histogram in the sum but the frequency associated with each bin has changed. The average histogram has more mass in the central bins and less in the extreme bins, and as a result, it is less skewed than any of the

components of the average. In a forecasting environment, if these five identical histograms were the realizations of the last five periods, the most natural one-step-ahead forecast would be an *identical* histogram to those observed in the most recent past. However, if we were to use (3) with  $\omega_i = 1/5$  as the forecasting filter, we would obtain a histogram as that described in Figure 1.

Moreover, histogram arithmetic is not suitable to represent the forecast error,  $h_{X_t} - \hat{h}_{X_t}$ . For example, for the case where the prediction is exact such that  $\hat{h}_{X_t} = h_{X_t}$ , histogram arithmetic does not deliver satisfactory results. Due to these reasons, we need to search for alternative approaches to operationalize the sum of histograms.

We should also mention that a very important assumption in the Colombo and Jaarsma arithmetic is that the random variables behind the histograms are independent. This assumption is rather strong since we should expect some dependence between the elements  $h_{X_t}$  of a histogram-valued time series. There are more sophisticated methods to operate with histograms of random variables characterized by some types of dependence, for instance, Li and Hyman (2004) and Berleant and Zhang (2004). However, the proposed operations do not deliver a random variable with a single distribution but rather some upper and lower bounds that enclose the unknown solution.

### 3.2 The barycentric histogram

The barycenter of a set of histograms was proposed by Irpino and Verde (2006) in a clustering context where the histogram barycenter represents the centroid of a cluster of histogram-valued data. An interesting precedent can be found in the work by Delon (2004a,b) that proposes a mean histogram using the Monge-Kantorovich distance for image equalization.

The barycentric histogram  $h_{X_B}$  is defined as the histogram that minimizes the distance between itself and a set of histograms  $h_{X_i}$  for  $i = 1, \dots, n$ ,

$$\min_{h_{X_B}} \left( \sum_{i=1}^n \omega_i D^p(h_{X_i}, h_{X_B}) \right)^{\frac{1}{p}}, \quad (16)$$

where  $\omega_i$  is the weight associated with histogram  $h_{X_i}$  such that  $\omega_i \geq 0$  and  $\sum_i \omega_i = 1$ ;  $p$  is a positive integer and  $D(h_{X_i}, h_{X_B})$  is a distance measure. The barycentric histogram  $h_{X_B}$  is also understood as a convex combination of the  $n$  histograms  $h_{X_i}$  considered.

Verde and Irpino (2007) analyzed different distance measures. Some, like the Hellinger Coefficient and the Total Variation distance, are rejected because they do not guarantee that the sum of weights of the barycenter is equal to 1. Others, like the Kolmogorov or the Prokhorov



distances, produce inconclusive solutions. Only the Mallows distance ( $p = 2$ ) is considered suitable for constructing barycenters. Following upon the ideas of Irpino and Verde, Arroyo and Maté (2009) proposed a barycenter using the Wasserstein distance ( $p = 1$ ) to construct a histogram forecast based on a k-NN algorithm.

Interestingly, both the Mallows and the Wasserstein distances have a clear interpretation as they are related to the Earth Movers Distance used in computer vision (Levina and Bickel, 2001). This distance, proposed by Rubner et al. (2000), is a solution to the Monge-Kantorovich transportation problem that determines the most efficient way of turning one pile of dirt into another (Rachev, 1984). In our case, the probability distributions are considered as piles of dirt and the considered distances are the minimum cost of turning one of them into the other.

The Mallows distance between two histograms  $h_1$  and  $h_2$  is defined as

$$D_M(h_1, h_2) = \sqrt{\int_0^1 (H_1^{-1}(l) - H_2^{-1}(l))^2 dl}, \quad (17)$$

and the Wasserstein distance as

$$D_W(h_1, h_2) = \int_{\mathfrak{R}} |H_1(x) - H_2(x)| dx = \int_0^1 |H_1^{-1}(l) - H_2^{-1}(l)| dl, \quad (18)$$

where  $l \in [0, 1]$  and  $H_i^{-1}(l)$  is the quantile function of the histogram  $h_i$ .

**Proposition 1.**

Let  $D^p(h_{X_i}, h_{X_B})$  be the Mallows distance and  $p = 2$ . The barycentric histogram  $h_{X_B}$  that solves the optimization problem

$$\min_{h_{X_B}} \left( \sum_{i=1}^n \omega_i \int_0^1 (H_{X_B}^{-1}(l) - H_{X_i}^{-1}(l))^2 dl \right)^{\frac{1}{2}} \quad (19)$$

has a quantile function  $H_{X_B}^{-1}$  that satisfies

$$H_{X_B}^{-1}(l) = \sum_{i=1}^n \omega_i H_{X_i}^{-1}(l) \text{ for } l \in [0, 1]. \quad (20)$$

*Proof.* For fixed  $l$ , the solution (20) follows directly from the first order condition  $\sum_{i=1}^n \omega_i (H_{X_B}^{-1}(l) - H_{X_i}^{-1}(l)) = 0$ .

**Proposition 2.**

Let  $D^p(h_{X_i}, h_{X_B})$  be the Wasserstein distance and  $p = 1$ . Define  $n_i = \omega_i N$ , where  $n_i$  is the number of times that the quantile value

$H_{X_i}^{-1}(l)$  is repeated and  $N = \sum_{i=1}^n n_i$ . The barycentric histogram  $h_{X_B}$  that solves the optimization problem

$$\min_{h_{X_B}} \sum_{i=1}^n \frac{n_i}{N} \int_0^1 |H_{X_B}^{-1}(l) - H_{X_i}^{-1}(l)| dl \quad (21)$$

has a quantile function  $H_{X_B}^{-1}$  that equals the median over all  $N$  values of  $\{H_{X_i}^{-1}(l)\}$  for all  $n_i$ ,  $i = 1, \dots, n$ .

*Proof.* Let  $H_{X_B}^{-1}(l)$  be an increasingly monotonic function on  $l$ . For fixed  $l$ , and for an odd<sup>A</sup> number of observations  $H_{X_i}^{-1}(l)$ , a global minimum exist. Let  $H_{X_i}^{-1}(l)$ ,  $i = 1, \dots, n$ , and let  $\mathbb{I}(\cdot)$  be an indicator function, the solution of the derivative of (21),

$$\frac{d \frac{n_i}{N} |H_{X_B}^{-1}(l) - H_{X_i}^{-1}(l)|}{d H_{X_B}^{-1}(l)} = \sum_{i=1}^n \frac{n_i}{N} \mathbb{I}(H_{X_B}^{-1}(l) > H_{X_i}^{-1}(l)) - \sum_{i=1}^n \frac{n_i}{N} \mathbb{I}(H_{X_B}^{-1}(l) < H_{X_i}^{-1}(l)), \quad (22)$$

is the number of observations  $H_{X_i}^{-1}(l)$  below the value  $H_{X_B}^{-1}(l)$  minus the number of observations above  $H_{X_B}^{-1}(l)$ . Thus, the sample median is the only value of  $H_{X_B}^{-1}(l)$  for which the number of observations above it equals the number of observations below. In practical terms, we order the values of  $H_{X_i}^{-1}(l)$  from low to high and choose  $H_{X_B}^{-1}(l)$  as the median quantile over all  $N$  ordered quantile values  $H_{X_i}^{-1}(l)$ <sup>B</sup>.

In Figure 2, we illustrate the construction of the barycentric histogram with the Mallows and Wasserstein distances.

[FIGURE 2]

Let  $h_1 = \{([1, 2), 0.7), ([2, 3), 0.2), ([3, 4), 0.1)\}$  and let  $h_2 = \{([11, 12), 0.1), ([12, 13), 0.2), ([13, 14), 0.7)\}$  be two histograms and  $\omega_1 = \omega_2 = 0.5$ . In the top left panel, we show the histograms  $h_1$ ,  $h_2$  and the barycenter  $h_B$  when the distance is Mallows; and in the top right panel, we show their corresponding distribution functions  $H_1, H_2$ , and  $H_B$ . Note

<sup>A</sup>With an odd number of observations the definition of the median is unambiguous. With an even number of observations, the solution is not unique as the median will fall between the two middle observations. The technical complications of considering an even number of observations have no practical consequences.

<sup>B</sup>In the singularity point, i.e.,  $H_{X_B}^{-1}(l) = H_{X_i}^{-1}(l)$  for some  $i$ , the derivative obviously does not exist. We can have a case in which  $H_{X_B}^{-1}(l) = H_{X_i}^{-1}(l)$  for several  $i$ , and it is possible that the median falls into one of these values.

that for any fixed  $l$ , the quantile function  $H_B^{-1}(l)$  is the average of the quantiles functions  $H_1^{-1}(l)$  and  $H_2^{-1}(l)$ , and consequently, the barycentric histogram  $h_B$  is an average of the position, range and shape of the histograms  $h_1$  and  $h_2$ . In the bottom panel, we perform a similar exercise using the Wasserstein distance. Since the number of histograms is even, the Wasserstein barycenter is not unique. Thus, any quantile  $H_B^{-1}(l)$  enclosed within the quantile functions  $H_1^{-1}(l)$  and  $H_2^{-1}(l)$  can be a valid Wasserstein barycenter for  $h_1$  and  $h_2$ . This behavior illustrates that only the Mallows distance will be suitable for the adaptation of smoothing techniques using the barycenter approach. Further explanation is provided in the forthcoming sections.

### 3.3 The barycenter and the moving average

Given the relation between the concepts of average and barycenter, the histogram forecast  $\hat{h}_{X_{t+1}}$  based on a moving average as in (3) can be reformulated as the barycentric histogram that solves the following optimization problem

$$\min_{\hat{h}_{X_{t+1}}} \sum_{i=1}^q \left( \omega_i D^p(\hat{h}_{X_{t+1}}, h_{X_{t-(i-1)}}) \right)^{1/p}, \quad (23)$$

where  $q$  is the order of the moving average,  $D(\cdot, \cdot)$  is an appropriate distance, and  $\omega_i$  is the weight assigned to  $h_{X_{t-(i-1)}}$ .

If  $D(\cdot, \cdot)$  is a Mallows distance and  $p = 2$ , then  $\hat{h}_{X_{t+1}}$  will be such that

$$\hat{H}_{X_{t+1}}^{-1}(l) = \sum_{i=1}^q \omega_i H_{X_{t-(i-1)}}^{-1}(l) \text{ for } l \in [0, 1]. \quad (24)$$

If  $D(\cdot, \cdot)$  is a Wasserstein distance and  $p = 1$ , depending on whether the number of elements in the sum (23) is even or odd, and depending on the values of the weights  $\omega_i$ , the median barycenter may not be unique or it may be based on fewer than  $q$  histograms. For instance, suppose that in the expression  $\hat{h}_{X_{t+1}} = \omega_1 h_{X_t} + \dots + \omega_q h_{X_{t-(q-1)}}$ , we have that  $\omega_1 > 0.5$ . Then, the Wasserstein barycenter is  $\hat{h}_{X_{t+1}} = h_{X_t}$  because it is calculated as the median of the quantile functions associated with the histograms  $h_{X_{t-(i-1)}}$  for  $i = 1, \dots, q$ . This median behavior induced by the Wasserstein distance implies that the natural smoothing associated with a moving average may be lost, and because of that, the Wasserstein distance is less suitable than the Mallows distance.

### 3.4 The barycenter and exponential smoothing

By following the same arguments as in the previous section, the histogram forecast  $\hat{h}_{X_{t+1}}$  based on an exponential smoothing filter as in (4)

can be reformulated as the barycentric histogram that solves the following optimization problem

$$\min_{\hat{h}_{X_{t+1}}} \left( \alpha D^p(\hat{h}_{X_{t+1}}, h_{X_t}) + (1 - \alpha) D^p(\hat{h}_{X_{t+1}}, \hat{h}_{X_t}) \right)^{1/p}, \quad (25)$$

where  $\alpha \in [0, 1]$ . The initial value required for  $t = 1$  can be as simple as  $h_{\hat{X}_1} = h_{X_1}$  or it can be obtained by a backcasting procedure using the reversed HTS and with an initial value  $h_{\hat{X}_T} = h_{X_T}$ , where  $T$  is the length of the time series.

If the distance measure is a Wasserstein distance and  $p = 1$ , the histogram forecast will be either  $h_{X_t}$  or  $\hat{h}_{X_t}$  depending on the value of  $\alpha$ . There are three possible outcomes: (i) if  $\alpha > 0.5$ , then  $\hat{h}_{X_{t+1}} = h_{X_t}$ ; (ii) if  $\alpha < 0.5$ , then  $\hat{h}_{X_{t+1}} = \hat{h}_{X_t}$ , which implies that, for any  $t$ ,  $\hat{h}_{X_{t+1}} = h_{X_1}$ ; and (iii) if  $\alpha = 0.5$ , then the result is any histogram whose cumulative distribution function is enclosed within the cumulative distribution functions of  $h_{X_t}$  and  $\hat{h}_{X_t}$ . A reasonable choice in this case would be to take a “mean” histogram such as the Mallows barycenter (see the Wasserstein distance example in Figure 2). In cases (i) and (ii), the purpose of smoothing is again lost because of the median effect.

The Mallows distance with  $p = 2$  is preferable regardless of the value of  $\alpha$ . When  $\alpha > 0.5$ , the Mallows barycenter will assign more weight to  $h_{X_t}$  than to  $\hat{h}_{X_t}$ , and when  $\alpha < 0.5$  the opposite will happen. In Figure 3, we report two examples with  $\alpha = 0.9$  and  $\alpha = 0.1$ .

[FIGURE 3]

It is important to note that the Mallows barycentric histogram is identical in the two optimization problems stated above, either as an exponential smoothing formulation (25) or as a moving average (23) with exponentially decreasing weights, that is,

$$\begin{aligned} \arg \min_{\hat{h}_{X_{t+1}}} & \left( \alpha D_M^2(\hat{h}_{X_{t+1}}, h_{X_t}) + (1 - \alpha) D_M^2(\hat{h}_{X_{t+1}}, \hat{h}_{X_t}) \right) \quad (26) \\ & = \arg \min_{\hat{h}_{X_{t+1}}} \sum_{j=1}^t \alpha (1 - \alpha)^{j-1} D_M^2(h_{X_{t-(j-1)}}, \hat{h}_{X_{t+1}}). \end{aligned}$$

This equivalence is due to Proposition 1. The solution to the left hand side of equation (26) is the Mallows barycenter with  $p = 2$ , which is obtained from the exponential smoothing equation (25). In this way, for each  $l \in [0, 1]$ , the  $\hat{H}_{X_{t+1}}^{-1}(l)$  satisfies

$$\hat{H}_{X_{t+1}}^{-1}(l) = \alpha H_{X_t}^{-1}(l) + (1 - \alpha) \hat{H}_{X_t}^{-1}(l). \quad (27)$$

In turn, the solution to the right hand of equation (26) is the Mallows barycenter with  $p = 2$  obtained from the moving average representation (23). In this way, for each  $l \in [0, 1]$ , the  $\hat{H}_{X_{t+1}}^{-1}(l)$  satisfies

$$\hat{H}_{X_{t+1}}^{-1}(l) = \sum_{j=1}^t \alpha(1-\alpha)^{j-1} H_{X_{t-(j-1)}}^{-1}(l). \quad (28)$$

By backward substitution in (27), we obtain (28). Thus, the solutions to equations (27) and (28) are equivalent.

On the contrary, this equivalence will not hold with the Wasserstein barycenter because of Proposition 2, and the solution to (25) may be different from the solution to (23).

Finally, the exponential smoothing in its error correction formulation (6) cannot be adapted within the barycenter approach because it is not possible to adapt the subtraction operation between two histograms.

#### 4 Value-at-Risk (VaR) with histogram data

The Bank for International Settlements (BIS) require that the banking industry, investment firms, and other financial institutions maintain certain levels of capital to guard against market risk that arises from their financial operations. The standard measurement of market risk, supported by the Basel Committee on Banking Supervision, is Value-at-Risk (VaR), which is not only a regulatory device but also a tool to guide investment decisions and capital allocation. The VaR is a probabilistic measure defined as the loss that a portfolio will face in a given period of time with a specific confidence level or probability. Let  $\{y_t\}$  be a time series of returns to a portfolio. The  $VaR_t^\gamma$  is defined as the (conditional) quantile of the distribution of  $y_t$  such that

$$\Pr(y_t \leq VaR_t^\gamma | \mathcal{F}_{t-1}) = \gamma, \quad (29)$$

where  $\mathcal{F}_{t-1}$  is the information set up to time  $t-1$ . The problem of interest is forecasting  $VaR_t^\gamma$ . The simplicity of the measure is deceiving because, in order to obtain the time-varying quantile  $VaR_t^\gamma$ , we need to search for an adequate model that exploits efficiently the information set  $\mathcal{F}_{t-1}$ . There are numerous proposals in the literature, most of them relying on a set of parametric assumptions about the probability distribution of portfolio returns. For instance, if the density belongs to the location-scale family,  $VaR_t^\gamma$  can be estimated from

$$VaR_t^\gamma = \mu_t(\theta) + F_t^{-1}(\gamma)\sigma_t(\theta), \quad (30)$$

where  $\mu_t(\theta)$  is the conditional mean of returns, depending upon a parameter vector  $\theta$ ;  $\sigma_t^2(\theta)$  is the conditional variance, also parameterized

by  $\theta$  or a subset of  $\theta$ ; and  $F_t^{-1}(\gamma)$  is the  $\gamma$ -quantile of a standardized cumulative distribution function  $F_t$  of returns. Common choices of  $F$  are the standard normal and the standardized Student-t distributions, though other densities, parametric and nonparametric, may be also suitable (González-Rivera et al., 2004). Other approaches are based on historical simulations of portfolio returns (Boudoukh et al., 1998); on the projection of the portfolio returns on a set of factors for which conditional means and conditional correlations must be forecasted (J.P.Morgan, 1996); on the modelling of the tails of the density appealing to extreme value theory (McNeil and Frey, 2000); and, on directly estimating the conditional quantile appealing to quantile estimation theory (Engle and Manganelli, 2004).

We propose an alternative or complementary measurement of  $VaR_t^\gamma$  that shares some aspects of the just mentioned approaches but it departs from them in an important dimension: the organization of the data set. We move the analysis from classical data to symbolic data. The symbolic object is a time series of histograms of returns. For a given frequency, say daily, we collect the daily returns and grouped them in a histogram at a lower frequency, say monthly. Then, the time series of monthly histograms of daily returns is the object of analysis. For  $VaR_t^\gamma$  calculations, the interest resides on the tails of the distribution of returns. It is customary to analyze the 1%, 5%, 95%, and 99% percentiles. These will guide the choice of bins and frequencies in the construction of the histogram. It is also possible to analyze just the quantile of interest, either with classical methods (i.e., the  $\gamma$ -quantile values form a classical time series) or with symbolic data methods (i.e., the  $\gamma$ -quantile intervals form an interval-valued time series). However, we will proceed by analyzing the full histogram because, as González-Rivera and Arroyo (2011) have shown, there are prediction gains from carrying information on many quantiles as opposed to just the information contained in a specific quantile.

In addition to the symbolic data approach, the main advantage of our proposal resides on the absence of parametric assumptions on the conditional density as well as on the conditional mean and variances of the return process. On the other hand, a criticism of our approach is that it will be substantially more difficult to find the data generating process behind a time series of histograms as we do not have yet the tools for model specification and testing for this type of objects. Nevertheless, our proposal – exponential smoothing approach based on barycentric histograms – provides a sensible exploratory tool. It is a low cost and parsimonious tool as it relies on the estimation of just one parameter, which helps to understand the weight of past information, and it provides

a one-step-ahead forecast that can be evaluated with classical tests.

We analyze the time series of daily histograms of the 5-minute returns<sup>C</sup> of two different stock indexes from May 1, 2008, to April 30, 2009. The stock market indexes considered are the SP500 and the IBEX35 with 252 and 253 observations, respectively. This exercise is of interest for those institutions engaged in high frequency trading. For each day and for each index, we construct a daily histogram with the 5-minute returns. There are 77 returns in each day for the case of the SP500 and 109 returns for the case of the IBEX35. Each histogram consists of 5 bins, partitioned according to the 5%, 30%, 70% and 95% quantiles. We do not include the 1% and 99% quantiles because there is not sufficient data in their corresponding intervals. We are interested mostly in the one-day-ahead forecast of the 5% and 95%  $Var_{t+1|t}^\gamma$  but we also include mid-quantiles to exploit all available information. We split the sample into an estimation period with 200 observations (from the beginning of May 2008 to mid-February 2009) and a forecasting period (from mid-February 2009 to the end of April 2009) reserved for the assessment of the one-step-ahead forecasts. The number of observations in the forecasting period are 52 in the SP500 and 53 in the IBEX 35.

In Figure 4, we plot the daily histograms in mid-October 2008 of the SP500 time series. We should note that this sample period is characterized by the greatest financial turmoil since the Great Depression, in particular, the months from September 2008 to March 2009 are times of high volatility.

[FIGURE 4]

We proceed with the estimation of the parameter  $\alpha$  in the exponentially smoothed average; see Eq. (25). We implement a one-dimensional grid search over  $\alpha \in [0, 1]$ . For the estimation sample and for every  $\alpha$ , we obtain the barycentric histogram  $\hat{h}_{X_t}(\alpha)$ . For each estimated histogram  $\hat{h}_{X_t}(\alpha)$ , we compute the value of the objective function, which is the Mean Distance Error function, defined as

$$MDE_M(\alpha) = \frac{1}{R} \sum_{t=1}^R D_M(h_{X_t}, \hat{h}_{X_t}(\alpha)), \quad (31)$$

---

<sup>C</sup>The 5-minute frequency is standard in the construction of realized volatility measures. See, for example the work in Bollerslev et al. (2000). It gives a nice trade-off between noise, present in ultra-high frequency observations, and information, most likely present in lower frequencies. The construction of the 5-min VaR is geared towards a computerized trading desk. Investment banks and other trading houses are allowed to deploy computer-based trading desks in the major exchanges, where trades happen in seconds and milliseconds.

where  $D_M(\cdot, \cdot)$  is the Mallows distance and  $R$  is the number of observations in the estimation sample. The optimal value  $\hat{\alpha}$  is the value that minimizes the loss function  $MDE_M(\alpha)$ . To obtain the one-day-ahead forecast, we fix  $\hat{\alpha}$  and construct the forecast one day at a time by expanding the information set one observation at a time until we exhaust the forecasting sample. Since our interest is on specific quantiles, we have retrieved the one-day-ahead  $\gamma$ -quantile forecast  $\hat{q}_{t|t-1}^\gamma$  from the one-day-ahead histogram forecast, and so we can compute the classical Mean Squared Forecast Error  $MSFE = (1/P) \sum_{t=1}^P (q_t^\gamma - \hat{q}_{t|t-1}^\gamma)^2$  for  $\gamma = 5, 30, 70$ , and 95% quantiles.

We also implement classical forecasting methods for the time series of  $\{q_t^\gamma\}$  for  $\gamma = 5, 30, 70, 95$ . The four quantile time series are modeled by an unrestricted Vector Autoregression (VAR) model (Lütkepohl, 2006). The observed quantile time series are plotted in Figures 5 and 6 for the SP500 and IBEX35 indexes, respectively. The visual inspection of both charts clearly reveals that the SP500 was more volatile than was the IBEX35 during the period considered.

[FIGURES 5 and 6]

In the SP500 case, model selection criteria indicate that the optimal number of lags is between 1 and 8. The Akaike Information Criterion (Akaike, 1974) (AIC) suggests 8, Schwarz Information Criterion (Schwarz, 1978) (SIC) 1, and Hannan-Quinn (Hannan and Quinn, 1979) (HQ) 2 lags. Given that the estimation sample is 200 observations, a VAR(8) model will be excessively parameterized, thus we have chosen two intermediate specifications, VAR(2) and VAR(4) models. Two interesting findings deserve some commentary. First, the lower quantile  $q_t^5$  Granger-causes (Granger, 1969) the rest of the quantiles, but not *vice versa*. Secondly, the lower quantiles are negatively correlated with the upper quantiles; thus, a large (in magnitude) return in the lower quantile (i.e., in the lower tail, returns are negative) is associated with a large positive return in the upper quantile (i.e., in the upper tail, returns are positive) making the histogram more spread out. On the contrary, the histogram tends to narrow when either the returns in the lower or in the upper tails are smaller in magnitude. This means that bad news increases the histogram spread, but good news reduces it. In the IBEX35 case, the VAR(1) model is the best choice according to the SIC, while VAR(5) is the model suggested by the AIC.

After estimation, we proceed to compute the one-day-ahead forecast of the four quantile time series in the validation period and their corresponding  $MSFE$  (Mean Square Forecast Error). In Tables 1 (SP500) and 2 (IBEX35), we report the  $MSFE$  for the four quantiles estimated



	VAR(2)	VAR(4)	Exp.Sm.	Naive
5% quantile	6.33E-03	5.91E-03	4.5E-03	7.14E-03
30% quantile	1.42E-03	1.26E-03	1.31E-03	2.1E-03
70% quantile	0.92E-03	0.93E-03	0.65E-03	1.11E-03
95% quantile	7.44E-03	7.71E-03	5.73E-03	10.67E-03

Table 1: Mean Square Forecast Error of the SP500 (February 17 to April 30, 2009)

	VAR(1)	VAR(5)	Exp.Sm.	Naive
5% quantile	5.21E-03	4.41E-03	4.5E-03	6.89E-03
30% quantile	0.67E-03	0.64E-03	0.58E-03	0.91E-03
70% quantile	0.49E-03	0.59E-03	0.65E-03	1.11E-03
95% quantile	4.67E-03	6.17E-03	5.71E-03	5.59E-03

Table 2: Mean Square Forecast Error of the IBEX35 (February 13 to April 30, 2009)

with four methods: two VAR models, the exponential smoothing method with barycenters, and a naive forecast, which does not entail estimation and claims that  $\hat{h}_{X_t|t-1} = h_{X_{t-1}}$ .

For the SP500, the exponential smoothing method delivers the smallest *MSFE* for the 5, 70, and 95% quantiles. For the 30% quantile, the *MSFE* of the VAR(4) model is the smallest (0.00126) but very close to that of the exponential smoothing method (0.0013). For the IBEX35, the exponential smoothing outperforms the naive method except in the 95% quantile and delivers mixed results compared to those from the VAR models.

We statistically assess the differences in *MSFE* by implementing a test of unconditional predictive ability (Diebold and Mariano, 1995). The null hypothesis of this test claims that there is not any difference in the value of the *MSFE* function between any two methods. On rejecting the null hypothesis, we will claim that one method is preferred to the other because, on average, it delivers a smaller loss over the prediction period. This test is very suitable in this environment because the methods considered, under the null hypothesis, are not nested within each other. The hypotheses of the test are  $H_0 : E(e_{other}^2 - e_{exp}^2) = 0$  and  $H_1 : E(e_{other}^2 - e_{exp}^2) > 0$ . The results for the SP500 and the IBEX35 are reported in Tables 3 and 4, respectively.

In the SP500 case, the test rejects very strongly the “naive” approach as the null hypothesis is rejected at significance levels lower than the customary 5 and 1% levels. When the exponential smoothing method

	VAR(2)-Exp.Sm.	VAR(4)-Exp.Sm.	Naive-Exp.Sm.
5% quantile	3.395 (0.0003)	1.469 (0.070)	3.108 (0.001)
30% quantile	1.137 (0.127)	-0.237 (0.406)	3.303 (0.0009)
70% quantile	1.624 (0.052)	1.511 (0.065)	2.747 (0.004)
95% quantile	1.146 (0.126)	1.465 (0.071)	3.417 (0.006)

Table 3: Value of the t-statistic and p-value (in parentheses) of the test of unconditional predictive ability in the SP500.

	VAR(1)-Exp.Sm.	VAR(5)-Exp.Sm.	Naive-Exp.Sm.
5% quantile	1.106 (0.136)	-0.037 (0.485)	2.168 (0.017)
30% quantile	1.135 (0.13)	0.7 (0.243)	2.41 (0.009)
70% quantile	-1.365 (0.089)	-0.484 (0.315)	3.144 (0.001)
95% quantile	-1.226 (0.112)	0.471 (0.319)	-0.108 (0.457)

Table 4: Value of the t-statistic and p-value (in parentheses) of the test of unconditional predictive ability in the IBEX35.

is compared to the VAR specifications, there are mixed results. The exponential smoothing method generates a smaller loss than does the VAR(4) model for the 5, 70, and 95% quantiles with a significance level of about 6%. The 30% quantile seems to be predicted equally well by the VAR models and by the exponential smoothing method. There is a strong advantage when modeling the lower tail – the 5% quantile that is regulated by the Basle Committee – by exponential smoothing. The test strongly rejects the VAR (2) model with a p-value of 0.0003 and it rejects the VAR (4) model with a p-value of 7%. The latter p-value is slightly higher than the customary 5% level but we should take into consideration that the VAR(4) model carries 68 parameters to be estimated and that the equation for the 5% quantile carries many regressors that are statistically insignificant given that the 30, 70, and 95% quantiles do not Granger-cause the 5% quantile. This affects the precision and the efficiency properties of the estimators that contribute to the power of the tests.

The results of the tests in the IBEX35 are similar for the exponential smoothing method and the VAR models. According to Table 4, the exponential smoothing method clearly dominates the naive method for the 5, 30, and 70% quantiles, but delivers an average loss similar to those from the VAR models. This is an interesting result taking into account that the forecasts of the exponential smoothing method are driven by just one parameter, while VAR models could be highly parametrized and require model selection and specification tests.

In summary, the exponential smoothing method has shown a better performance for the SP500 which is more volatile than the IBEX35. In Figure 7, we present a graphical comparison between the 5% and 95% quantile predictions of the VAR (4) model and the exponential smoothing method together with the actual time series. Visual inspection of this figure shows that the VAR predictions are more ragged than those offered by exponential smoothing, which follows very closely the sharp peaks and troughs of the realized data.

[FIGURE 7]

## 5 Conclusion

We have shown how to adapt smoothing techniques to forecast HTS. Our proposal is based on the introduction of a barycentric histogram that emulates the “average” operation, which is the key in any smoothing filter. We have also shown that, due to its linear properties, only the Mallows-barycenter is acceptable if we wish to preserve the essence of any smoothing mechanism. We have also reviewed the available histogram arithmetic and we have shown that, in a forecasting environment, it is not suitable to calculate averages and forecast errors.

In the absence of a data generating process for HTS, exponential smoothing methods are low-cost and parsimonious exploratory tools as they require the estimation of just one parameter. In our empirical application to forecast Value-at-Risk, a barycentric exponential smoothing method shows a strong forecasting performance – especially for the 5% quantile – when it is compared with forecasts from vector autoregression models, which require a large number of parameters to estimate.

## Appendix

### Conversion of a histogram into a quantile histogram

Let  $h_D = ([d_i], \pi_{d,i})$  with  $i = 1, \dots, m$  be a histogram of  $m$  disjoint bins and let us assume that inside each bin data points are uniformly distributed. Histogram  $h_D$  can be rearranged as a histogram  $h_S = ([s_j], \pi_{s,j})$  with  $j = 1, \dots, n$  defined by an ordered sequence of  $n + 1$  quantiles  $\{q_j\}$  where  $q_1 = 0$  and  $q_{n+1} = 1$ , and where frequencies of  $h_S$  are determined by  $\pi_{s,j} = q_{j+1} - q_j$ . The rearrangement of  $h_D$  as  $h_S$  can be done by means of the following pseudocode.

```
s_L[1]:=d_L[1];
i:=1;
for j:=1 to n do{ //Loop for each [s_j] bin
  //Frequency of [s_j] bin
  freq:=q(j+1)-q(j);
  //Current frequency of [s_j] bin
  pi_s[j]:=0;
  full:=false;
  //Loop for each [d_i] bin
  while (i<= m) and (not full) do{
    //New frequency falls behind the limit
    if (pi_s[j]+pi_d[i]<freq) then{
      pi_s[j]:=pi_s[j]+pi_d[i];
      i:=i+1;
    }
    //New frequency is precisely the limit
    elseif (pi_s[j]+pi_d[i]= freq) then{
      full:=true;
      pi_s[j]:=freq;
      i:=i+1;
      //Upper bound of the [s_j] bin
      s_U[j]:=d_U[i];
      //If [s_j] is not the last bin then...
      if (j+1<= n) then{
        //Lower bound of the [s_{j+1}] bin
        s_L[j+1]:=d_U[i];
      };
    }
  }
  //New frequency exceeds the limit
  else{
    full:=true;
    pi_s[j]:=freq;
```

```

//Required frequency
req_freq:=freq- $\pi$ _s[j];
//Upper bound of the [s_j] bin
s_U[j]:=d_L[i]+(req_freq/ $\pi$ _d[i])*(d_U[i]-d_L[i]);
//Updating the lower bound of the [d_i] bin
d_L[i]:=s_U[j];
//Updating the frequencie of the [d_i] bin
 $\pi$ _d[i]:= $\pi$ _d[i]-req_freq;
//Upper bound of the [s_j] bin
s_U[j]:=d_U[i];
//If [s_j] is not the last bin then...
if (j+1<= n) then{
    //Lower bound of the [s_{j+1}] bin
    s_L[j+1]:=d_U[i];
};
};
};

```

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.
- Arroyo, J., G. González-Rivera, and C. Maté (2010). *Handbook of Empirical Economics & Finance*, Chapter Forecasting with interval and histogram data. Some financial applications, pp. 247–280. Chapman & Hall/CRC.
- Arroyo, J. and C. Maté (2009). Forecasting histogram time series with k-nearest neighbours methods. *International Journal of Forecasting* 25, 192–207.
- Benjamini, Y. (1988). Opening the box of a boxplot. *The American Statistician* 42, 257–262.
- Berleant, D. and J. Zhang (2004). Representation and problem solving with the distribution envelope determination (DEnv) method. *Reliability Engineering and System Safety* 85(1–3), 153–168.
- Billard, L. and E. Diday (2003). From the statistics of data to the statistics of knowledge: Symbolic data analysis. *Journal of the American Statistical Association* 98(462), 470–487.
- Billard, L. and E. Diday (2006). *Symbolic Data Analysis: Conceptual Statistics and Data Mining* (1 ed.). Chichester: Wiley & Sons.
- Bollerslev, T., J. Cai, and F. M. Song (2000). Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the us treasury bond market. *Journal of Empirical Finance* 7(1), 37–55.
- Boudoukh, J., M. Richardson, and R. Whitelaw (1998). The best of both worlds: A hybrid approach to calculating Value at Risk. *Risk* 11, 64–67.
- Colombo, A. and R. Jaarsma (1980). A powerful numerical method to combine random variables. *IEEE Transactions on Reliability* 29(2), 126–129.
- Delon, J. (2004a). *Fine comparison of images and other problems*. Ph. D. thesis, École Normale Supérieure de Cachan.
- Delon, J. (2004b). Midway image equalization. *Journal of Mathematical Imaging and Vision* 21, 119–134.

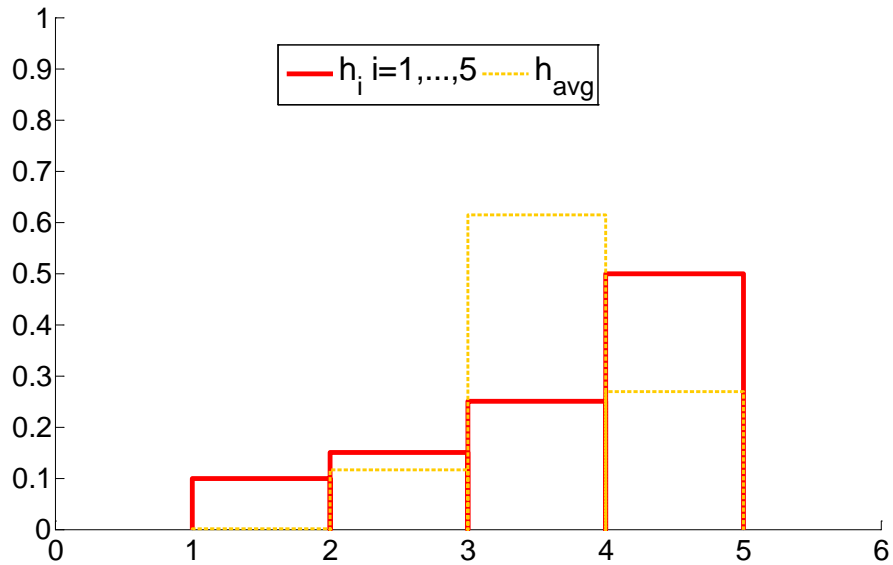
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.
- Engle, R. and S. Manganelli (2004). Caviar: Conditional autoregressive Value-at-Risk by regression quantiles. *Journal of Business and Economic Statistics* 22, 367–381.
- González-Rivera, G. and J. Arroyo (2011). Time series modeling of histogram-valued data. The daily histogram time series of SP500 intradaily returns. *International Journal of Forecasting* (forthcoming).
- González-Rivera, G., T. Lee, and S. Mishra (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting* 20, 629–645.
- Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.
- Hannan, E. J. and B. G. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society B* 41, 190–195.
- Irpino, A. and R. Verde (2006). A new Wasserstein based distance for the hierarchical clustering of histogram symbolic data. In *Data Science and Classification, Proceedings of the IFCS 2006*, Berlin, pp. 185–192. Springer.
- J.P.Morgan (1996). Riskmetrics. Technical document, J.P. Morgan.
- Levina, E. and P. J. Bickel (2001). The Earth Mover’s Distance is the Mallows distance: Some insights from statistics. In *Proceedings of the 8th International Conference on Computer Vision*, pp. 251–256.
- Li, W. and J. M. Hyman (2004, July). Computer arithmetic for probability distribution variables. *Reliability Engineering & System Safety* 85(1-3), 191–209.
- Lütkepohl, H. (2006). *Palgrave Handbook of Econometrics. Volume 1: Econometric Theory*, Chapter Vector Autoregressive Models, pp. 477–510. Houndmills: Palgrave Macmillan.
- McNeil, A. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7, 271–300.

- Moore, R. E. (1966). *Interval Analysis*. Englewood Cliffs, N.J.: Prentice Hall.
- Nagabhushan, P. and R. Pradeep Kumar (2007). Histogram PCA. In *ISNN '07: Proceedings of the 4th International Symposium on Neural Networks*, Berlin, Heidelberg, pp. 1012–1021. Springer-Verlag.
- Rachev, S. T. (1984). The Monge-Kantorovich mass transference problem and its stochastic applications. *Theory of Probability and its Applications* 29, 647–676.
- Rodríguez, O., E. Diday, and S. Winsberg (2000). Generalization of the principal components analysis to histogram data. In *Workshop on Symbolic Data Analysis at the European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD 2004*, Pisa.
- Rubner, Y., C. Tomasi, and L. J. Guibas (2000). The Earth Mover’s Distance as a metric for image retrieval. *International Journal of Computer Vision* 40(2), 99–121.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.
- Tukey, J. W. (Ed.) (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Verde, R. and A. Iripino (2007). Dynamic clustering of histogram data: using the right metric. In *Selected Contributions in Data Analysis and Classification*, pp. 123–134. Springer.



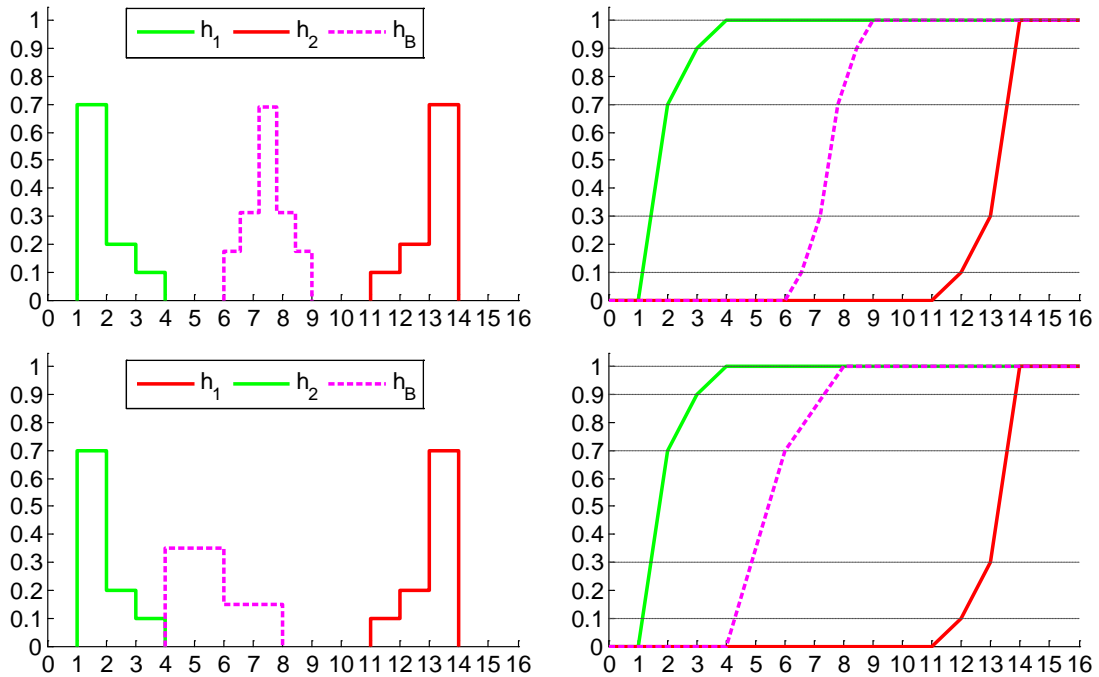
**FIGURE 1**

Average histogram of a set of five identical histograms



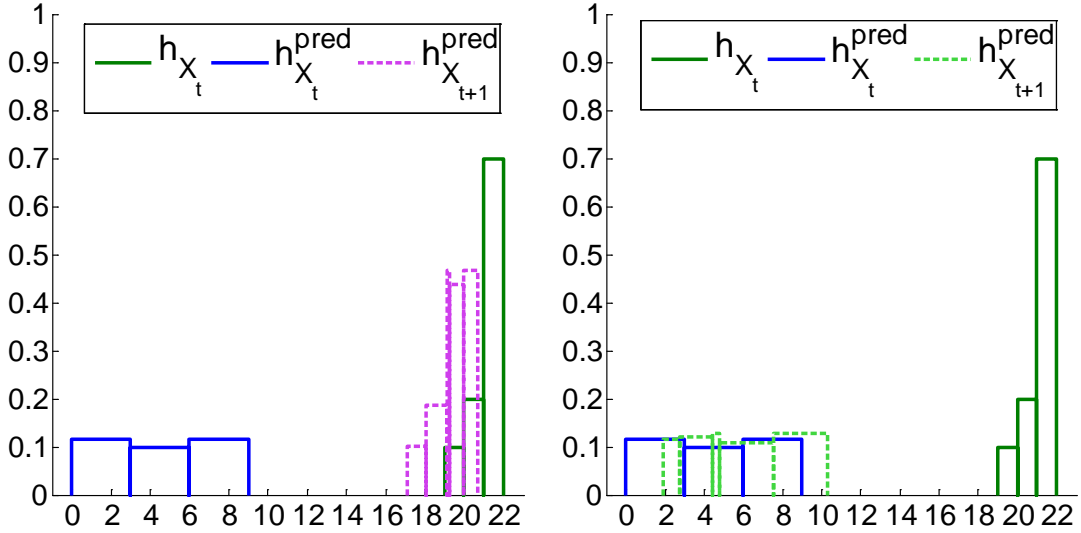
**FIGURE 2**

The Mallows barycenter (top) and a Wasserstein barycenter (bottom) for histograms  $h_1$  and  $h_2$  (histograms on the left and cumulative distribution functions on the right)



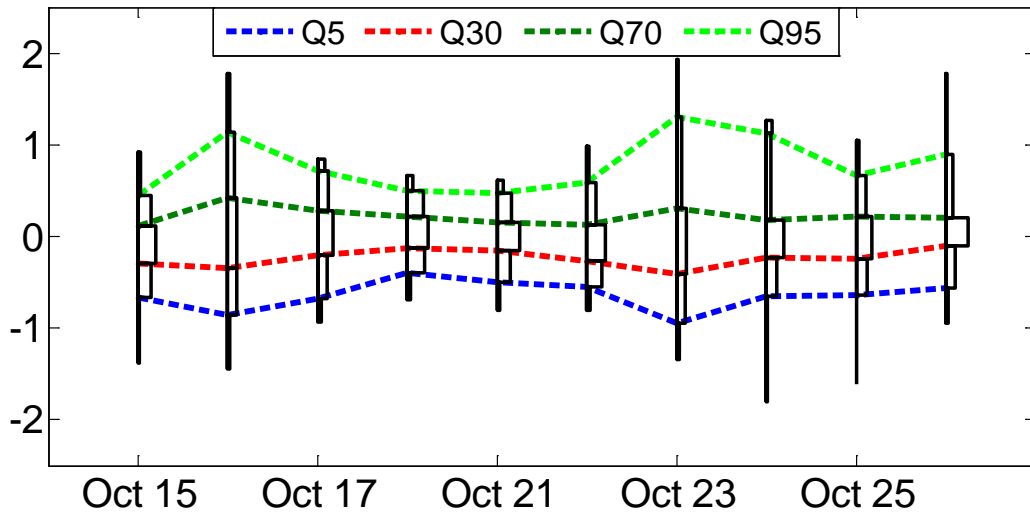
**FIGURE 3**

Mallows barycenter as the histogram forecast based on an exponential smoothing filter  $\alpha=0.9$  (left) and  $\alpha=0.1$  (right)

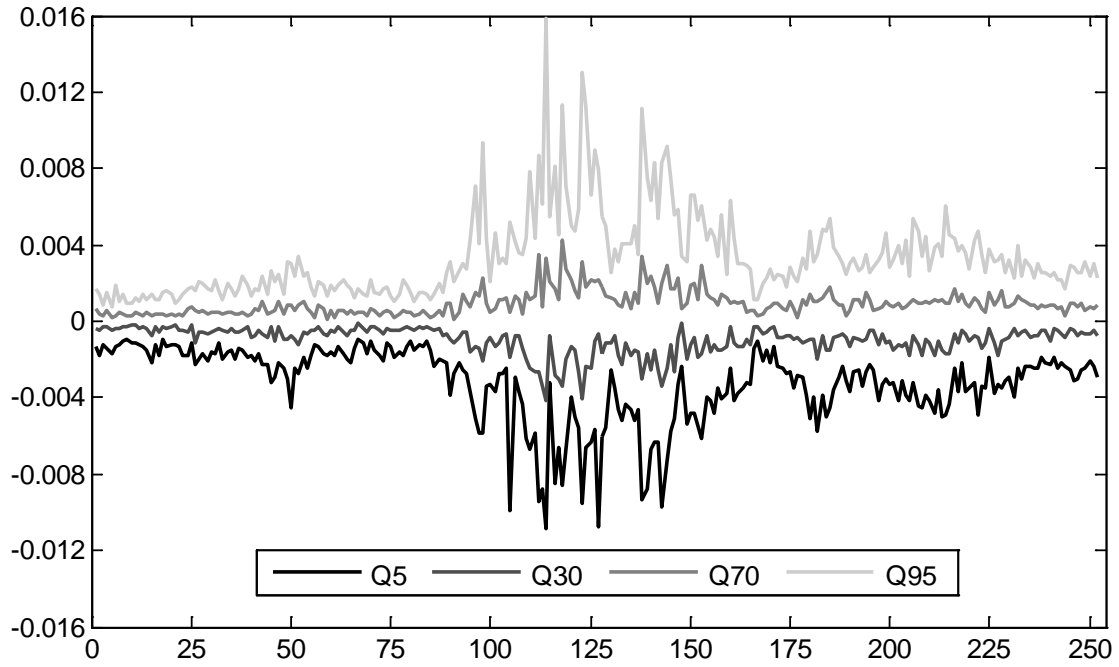


**FIGURE 4**

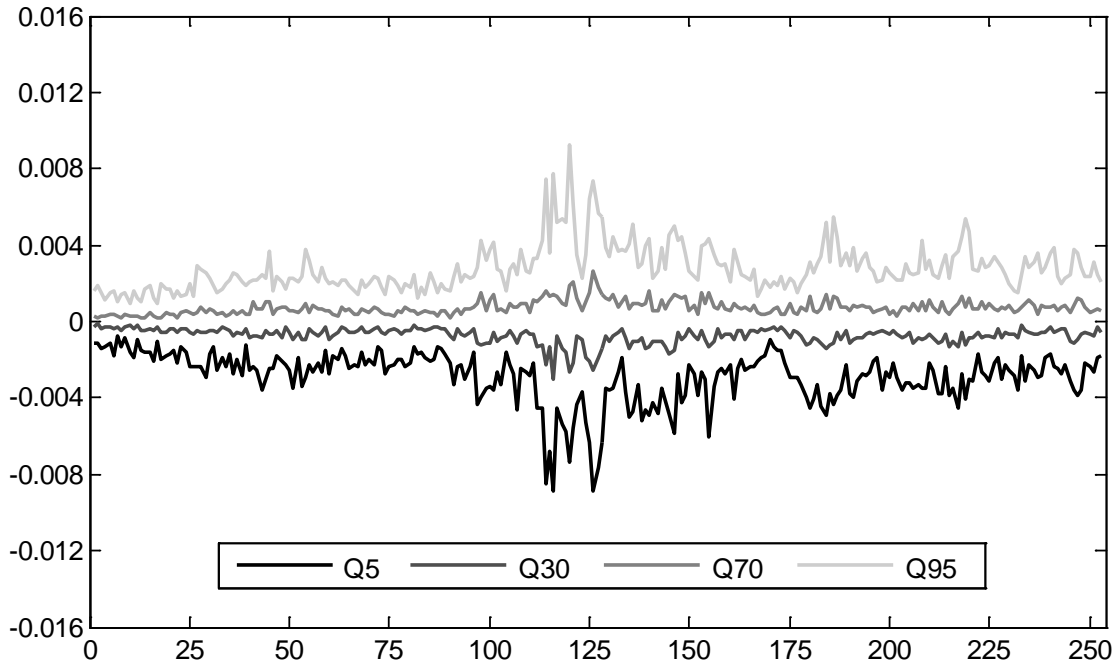
Excerpt of the intra-daily returns HTS of the SP500 in October 2008



**FIGURE 5**  
Observed quantiles of the SP500



**FIGURE 6**  
Observed quantiles of the IBEX35



**FIGURE 7**  
**Quantile prediction**

