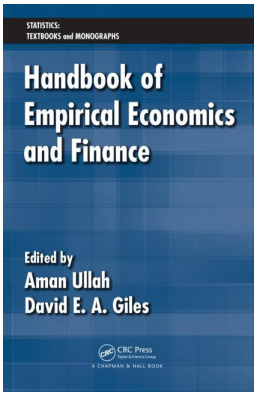

Handbook of Empirical Economics and Finance

Ullah Aman, E. A. Giles David

Forecasting with Interval and Histogram Data: Some Financial Applications

Publication details
https://test.routledgehandbooks.com/doi/10.1201/b10440-11
Arroyo Javier, González-Rivera Gloria, Maté Carlos
Published online on: 20 Dec 2010

10

Forecasting with Interval and Histogram Data: Some Financial Applications

Javier Arroyo, Gloria González-Rivera, and Carlos Maté

CONTENTS

10.1 Introduction

In economics we customarily deal with classical data sets. When we collect information on a set of variables of interest, either in a cross-sectional or/and

time series framework, our sample information is a collection of data points $\{y_i\}, i = 1 \ldots n$ or $\{y_t\}, t = 1 \ldots T$ where y_i or $y_t \in \mathbb{R}$ takes a single value in $\mathbb{R}$. In many instances, the single value is the result of an aggregation procedure, spatial or temporal, over information collected at a very disaggregated level. Some pertinent examples follow.

In financial markets the price of an asset (stocks, bonds, exchange rates, etc.) is observed at a very high frequency, i.e., tick by tick; however, there is a huge number of studies where the analysis is performed at the daily frequency using the closing price, or even at lower frequencies such as weekly or monthly. It may be claimed that tick-by-tick pricing will generate a huge amount of data from which it will be difficult to discriminate information from noise, but on the other extreme, by analyzing just closing prices we will be discarding valuable intraday information. We can think of alternative ways of collecting information, for instance, we can gather the maximum and minimum prices in a day so that the information to be analyzed will come in an interval format; or the daily interquartile prices such that the interval will run from the price at the 25% quartile to the price at the 75% quartile; or we can construct daily histograms with all the intraday prices. In these cases the data point is no longer a single value but a collection of values represented by the daily low/high interval, or the interquartile interval, or the daily histogram. The intervals or the histograms, when indexed by time, will constitute an interval time series or a histogram time series.

Another instance refers to the information collected by national statistical institutes in relation to income and population dynamics. Census surveys provide socioeconomic information on all individuals in a nation that is customarily disseminated in an aggregated format, for instance a time series of average income per capita. The objective of these national surveys is not to follow the dynamics of single individuals, which most likely will be different from one period to the next, but the dynamics of a collective. However, summarizing national information by averages, though informative, is a poor approach that throws away the internal variation provided by the disaggregated information about the single units. Once more, disseminating the data in a richer format such as intervals or histograms will provide a more complete picture of income and population dynamics. There are many other areas such as marketing, environmental sciences, quality control, medical sciences, etc. in which the information is rich enough to make the object of analysis not the single-valued variable but the interval-valued or the histogram-valued variable.

Interval- and histogram-valued data can be classified as symbolic data sets as opposed to classical data sets. Symbolic data is a proposal to deal with the massive information contained in nowadays super large data sets found across many disciplines. While the analysis of these data sets requires some summary procedure to bring them to a manageable size, the objective is to retain as much of their original knowledge as possible. An extensive review of this new field, which started in the late 1980s and early 1990s, is provided by Billard and Diday (2003, 2006), who define the complexity of symbolic data, review the current methods of analysis and state the challenges that lie ahead.

Economics and business are disciplines in which data sets are becoming consistently larger due to sophisticated information systems that collect and store huge amount of data. However, the development of new methodologies to deal with the characteristics of large data sets is moving at a slower pace. A case on time is the aforementioned high-frequency financial data and the challenges brought by it such as irregularly spaced observations with strong intraday patterns and a complex dependence structure. There are other examples in the economics literature that emphasize the richness of the data, though eventually the analysis is performed within the boundaries of classical inferential methods. For instance, the article by Zellner and Tobias (2000) provides the time series of the median and interquartile range of the industrial production growth rates of 18 countries but eventually the authors focus on the single-valued time series of the median growth rates. The article by González-Rivera, Lee, and Mishra (2008) presents a stylized time series of cross-sectional returns of the constituents of the SP500 index grouped in histograms (see Figure 10.1). However, the authors focus on the dependence structure of the single-valued time series of the time-varying cross-sectional ranks (VCR). Both of these instances could be viewed from the perspective of symbolic data: in Zellner and Tobias (2000) the data is an interval-valued time series and in González-Rivera, Lee, and Mishra (2008) is a histogram-valued time series.

There is an emergent literature in economics and statistics dealing with interval-valued data in a regression framework. Manski and Tamer (2002) examined a regression model where some regressors are interval-valued, like interval wealth and income, and some others are point-valued. Lima Neto

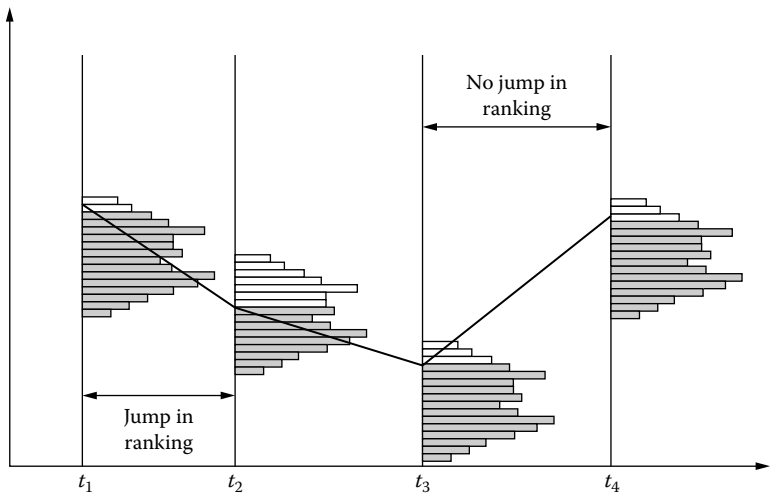

FIGURE 10.1
Stylized time series of the histograms of the cross-sectional returns of the constituents of the SP500 index. (From González-Rivera, G., T.-H. Lee, and S. Mishra. 2008. Jumps in cross-sectional rank and expected returns: a mixture model. *Journal of Applied Econometrics* 23:585–606.)

and de Carvalho (2010) proposed a constrained linear regression model for interval-valued data. Maia, de Carvalho, and Ludermir (2008) implemented ARIMA and neural networks models to forecast the center and radii of intervals. Han et al. (2008) analyzed the sterling-dollar exchange rate time series based on an interval linear model. Cheung, Cheung, and Wan (2009) analyzed the range of daily stock prices by proposing a VECM for the daily interval of high and low prices. García-Ascanio and Maté (2010) forecast monthly electricity demand with hourly interval data. A different approach to regression that treats intervals as convex compact random sets is proposed in González-Rodríguez et al. (2007) and Blanco et al. (2008). Regression models with histogram-valued data are almost nonexistent so that they offer wide opportunities for further research.

This chapter focuses on the forecasting of interval and histogram-valued data. The surveys and review articles by Diday and his coauthors focus on descriptive and multivariate methods of analysis adapted from the classical statistical methodology. To our knowledge, the development of forecasting methods for interval and histogram-valued data is in its infancy so that this chapter is a contribution to that end. We start with a preliminary section defining the structure of the data and basic descriptive statistics. There are two main sections, one for interval data and another for histogram data. In the first, we review how classical regression methods can be adapted to analyze intervals. The main insight is that the interval can be defined by its center and radius or by its minimum and maximum, so that we construct two time series to which classical methods can be applied. In this vein, we build a system, either VAR or VEC models, from which an interval forecast will be obtained. In a different approach based on the arithmetic of intervals and on notions of distances between intervals, we adapt classical filtering techniques like the exponential smoothing and nonparametric techniques like the k-Nearest Neighbors (k-NN) algorithm to produce the interval forecast. In the second main section, we deal with histogram-valued data. In this case the object of analysis is considerably more difficult to analyze and we focus exclusively on the adaptation of smoothing techniques and the k-NN. To construct a histogram forecast, we will not base our operations on the arithmetic of histograms but on the key idea of the "barycentric" histogram as the "average" measure. We should stress that no attempt has been made, either with a time series of intervals or histogram, to uncover the data generating mechanism but rather to forecast the future under the premise that it should not be very far from some average (weighted or unweighted) of the past.

10.2 Interval Data

In this section, we will define interval data and the interval random variable. As a foundation for the forthcoming analysis, we succinctly introduce the algebra of intervals. We will focus on the empirical first and second moments

of the interval random variable. The main objective of this section is to discuss (1) regression analysis with interval data, and (2) the forecasting problem. A financial application will showcase the contribution of (1) and (2) to the modeling of economic and financial data. While we will not discuss the nature of interval data, we acknowledge that there are many reasons why interval data may arise. Among others, interval data are generated when the data collection process genuinely produces intervals, or when there are not exact numerical values to quantify a variable, or when there is uncertainty of any kind in the values of the variable, or when variability of a variable is the focus of analysis, or when the measurement tools produce measurement errors. Regardless of the origin, the researcher will be facing data that come with an interval format and this is the primary object of analysis.

10.2.1 Preliminaries

We start with the basic notion of an interval following Kulpa (2006). Let $(E, \leq)$ be a partially ordered set. An interval is generally defined as follows:

Definition 10.1 *An interval* $[a]$ *over the base set* $(E, \leq)$ *is an ordered pair* $[a] = [a_L, a_U]$, *where* $a_L, a_U \in E$ *are the endpoints or bounds of the interval such that* $a_L \leq a_U$.

The interval is called degenerate when $a_L = a_U$, in which case the interval reduces to a point. An interval is the set of elements bounded by the endpoints, these included, namely, $[a] = \{e \in E \mid a_L \leq e \leq a_U\}$. When the base set E is the set of real numbers $\mathbb{R}$, the intervals are subsets of the real line $\mathbb{R}$.

An equivalent representation of an interval is given by the center (midpoint) and radius (half range) of the interval, namely, $[a] = \langle a_C, a_R \rangle$, where $a_C = (a_L + a_U)/2$ and $a_R = (a_U - a_L)/2$.

10.2.1.1 Basic Interval Arithmetic

In order to proceed with our analysis we need an algebra to operate with intervals. Basic interval arithmetic (Moore 1966; Moore, Kearfott, and Cloud 2009) is based on the following principle: let $[a]$ and $[b]$ be two intervals and $\square$ be an arithmetic operator, then $[a]\square[b]$ is the smallest interval which contains $a \square b$, $\forall a \in [a]$ and $\forall b \in [b]$. Interval addition, subtraction, multiplication and division are particular cases of this principle and are defined by

$$[a] + [b] = [a_L + b_L, a_U + b_U] \tag{10.1}$$

$$[a] - [b] = [a_L - b_U, a_U - b_L] \tag{10.2}$$

$$\begin{aligned}[a] \cdot [b] = [&\min\{a_L \cdot b_L, a_L \cdot b_U, a_U \cdot b_L, a_U \cdot b_U\}, \\ &\max\{a_L \cdot b_L, a_L \cdot b_U, a_U \cdot b_L, a_U \cdot b_U\}]\end{aligned} \tag{10.3}$$

$$[a]/[b] = [a] \cdot (1/[b]), \text{ with } 1/[b] = [1/b_U, 1/b_L]. \tag{10.4}$$

It is worth noting that interval arithmetic subsumes the classical one, in the sense that, if the operands are degenerate intervals, the result of interval operations will be equal to the result obtained by the single number arithmetic. In interval arithmetic, addition and multiplication satisfy the associative and commutative properties. The distributive property does not always hold, but the subdistributive property is satisfied, which is defined as

$$[a]([b]+[c]) \subseteq [a][b]+[a][c]. \tag{10.5}$$

If $[a]$ is a degenerate interval, then this property becomes the distributive property. The interval arithmetic is key for the development of regression techniques and for the adaptation of forecasting methods to interval data.

10.2.1.2 Interval Random Variable

We proceed with the definition of an interval random variable. Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is the set of elementary events, $\mathcal{F}$ is the σ-field of events and $P : \mathcal{F} \to [0, 1]$ the σ-additive probability measure; and define a partition of Ω into sets $A(x)$ such $A_X(x) = \{\omega \in \Omega | X(\omega) = x\}$, where $x \in [x_L, x_U]$, then:

Definition 10.2 *A mapping $X : \mathcal{F} \to [x_L, x_U] \subset \mathbb{R}$, such that for all $x \in [x_L, x_U]$ there is a set $A_X(x) \in \mathcal{F}$, is called an interval random variable.*

10.2.1.3 Descriptive Statistics

The descriptive statistics of an interval random variable are proposed by Bertrand and Goupil (2000). For an interval random variable X, suppose that we have a sample of m individuals ($i = 1, 2, \ldots, m$) and for each i, an interval data point $[x]_i \equiv [x_{Li}, x_{Ui}]$. A key assumption for the forthcoming descriptive statistics is that the values in a given interval, i.e., $x_{Li} \leq x \leq x_{Ui}$, are uniformly distributed within the interval. Furthermore, we assume that each individual has the same probability $1/m$ of being observed. Then, the *empirical density function* $f_X(x)$ is a mixture of m uniform distributions

$$f_X(x) = \frac{1}{m} \sum_{i:x\in[x]_i} \frac{I(x \in [x]_i)}{\| [x]_i \|} = \frac{1}{m} \sum_{i:x\in[x]_i} \frac{1}{x_{Ui} - x_{Li}} \qquad x \in \mathbb{R}, \tag{10.6}$$

where $I(x \in [x]_i)$ is an indicator function that takes the value 1 when $x \in [x]_i$ and zero otherwise; and $\| [x]_i \|$ is the length of the interval $[x]_i$.

Based on the density function (Equation 10.6), the sample mean is obtained by solving the following integral

$$\begin{aligned}\bar{X} &= \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{m} \sum_{i:x\in[x]_i} \frac{1}{x_{Ui} - x_{Li}} \int_{x_{Li}}^{x_{Ui}} x dx \\ &= \frac{1}{2m} \sum_i (x_{Ui} + x_{Li}) = \frac{1}{m} \sum_i x_{C_i},\end{aligned} \tag{10.7}$$

concluding that the sample mean of an interval random variable is the average of the centers of the intervals in the sample. Analogously, the sample variance is calculated by solving the integral

$$S_X^2 = \int_{-\infty}^{\infty} (x - \bar{X})^2 f(x) dx = \left(\int_{-\infty}^{\infty} x^2 f(x) dx \right) - \bar{X}^2, \tag{10.8}$$

which can be rewritten in terms of the interval bounds as

$$S_X^2 = \frac{1}{3m} \sum_i (x_{Ui}^2 + x_{Ui} x_{Li} + x_{Li}^2) - \frac{1}{4m^2} \left[\sum_i (x_{Ui} + x_{Li}) \right]^2. \tag{10.9}$$

The sample variance combines the variability of the centers as well as the variability within each interval. When the interval is degenerate, both sample moments, the mean and the variance, collapse to the sample mean and variance of the classical data.

10.2.2 The Regression Problem

Now suppose that we have two interval random variables Y and X for which we collect a sample of intervals $([x]_i, [y]_i)$ for $i = 1, 2, \ldots, m$. The interval data point i is a rectangle centered in the centers of $[x]_i$ and $[y]_i$ and whose sides are equal to the length of the respective intervals. A graphical representation of this data is provided in Figures 10.2 to 10.4. In this section, we review the analysis of a regression model with interval data. The classical regression model can be adapted to interval data by focusing on the centers of the interval, or on the maximum and minimum of the interval, or on the center and radius of the interval. The advantage of this approach is that statistical inference is readily available.

The simplest approach to estimate a regression model with interval data is provided by Billard and Diday (2000). It consists of fitting a regression line to the centers of the intervals, $y_{Ci} = \beta' x_{Ci} + \varepsilon_{Ci}$, so that the objective function to minimize is

$$\min_{\hat{\beta}} \sum_i \hat{\varepsilon}_{Ci}^2 = \sum_i (y_{Ci} - \hat{\beta}' x_{Ci})^2, \tag{10.10}$$

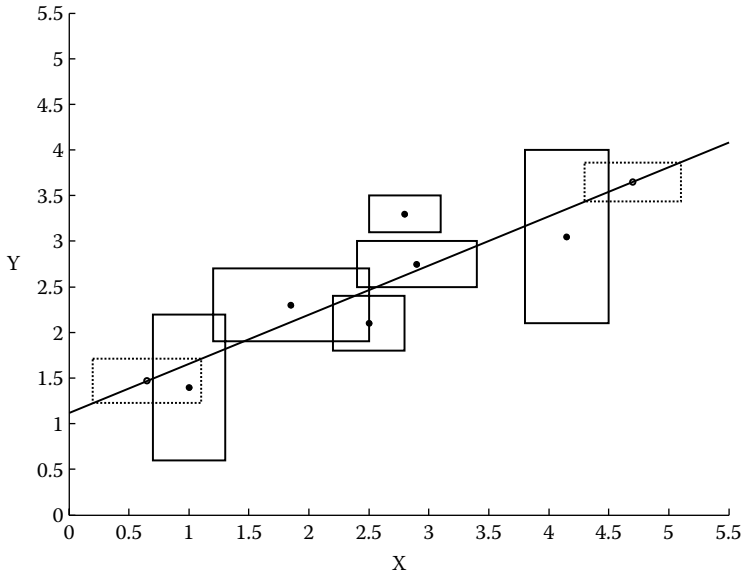

FIGURE 10.2
Fitting of a regression line to the centers of the intervals (From Billard, L., and E. Diday. 2000. Regression analysis for interval-valued data. In *Data Analysis, Classification and Related Methods: Proceedings of the 7th Conference of the IFCS, IFCS 2002*. Berlin: Springer. pp. 369–374.) The estimated rectangles according to the regression line are represented by a dashed line.

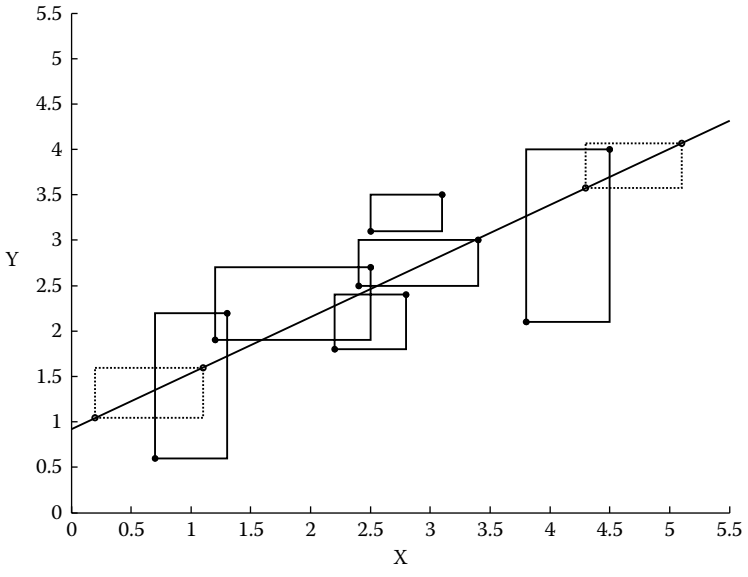

FIGURE 10.3
Regression line according to Brito (2007). The estimated rectangles from the regression line are represented by a dashed line.

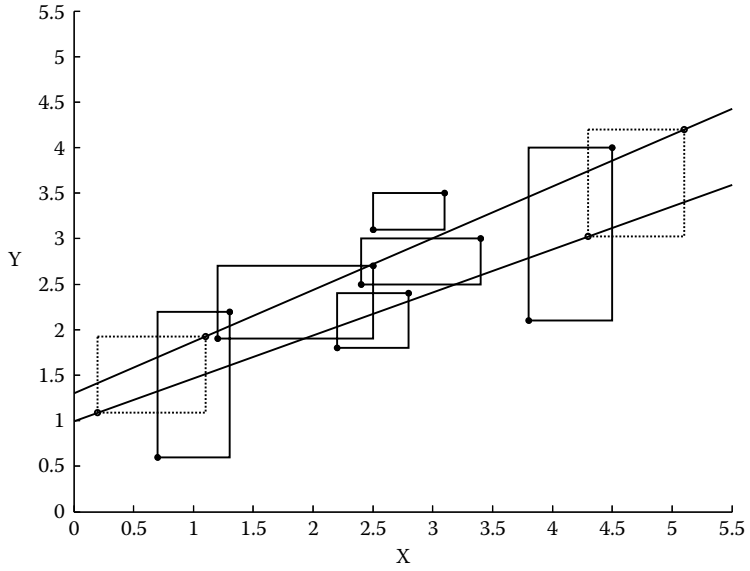

FIGURE 10.4
Regression lines fitted to the minima and maxima of the intervals (Billard, L., and E. Diday. 2002. Symbolic regression analysis. In *Classification, Clustering and Data Analysis: Proceedings of the 8th Conference of the IFCS, IFCS 2002*. Berlin: Springer. pp. 281–288.) The estimated rectangles according to the regression lines are represented by a dashed line.

the solution to this problem is the classical least squares estimator $\hat{\beta} = (X_C' X_C)^{-1} X_C' Y_C$ and standard statistical inference will apply under the standard assumptions about the error term of the regression. Though this model will provide information about the average centrality of the intervals, it disregards the range of the intervals that is an important feature of interval data.

There are several proposals aimed to incorporate the length of the interval into the analysis. Brito (2007) proposes to minimize the following objective function

$$\min_{\hat{\beta}} \sum_i \left(\hat{\varepsilon}_{Li}^2 + \hat{\varepsilon}_{Ui}^2\right) = \sum_i \left(y_{Li} - \hat{\beta}' x_{Li}\right)^2 + \sum_i \left(y_{Ui} - \hat{\beta}' x_{Ui}\right)^2, \tag{10.11}$$

which is equivalent to run two constrained (same regression coefficients) regressions on the lower bounds $y_{Li} = \beta' x_{Li} + \varepsilon_{Li}$ and the upper bounds $y_{Ui} = \beta' x_{Ui} + \varepsilon_{Ui}$ of the intervals. For the case of one regressor model, the OLS estimators have the following expression

$$\hat{\beta}_1 = \frac{\tilde{S}_{XY}}{\tilde{S}_X^2} = \frac{\frac{1}{2m}\sum_i \left[(x_{Li} - \bar{X})(y_{Li} - \bar{Y}) + (x_{Ui} - \bar{X})(y_{Ui} - \bar{Y})\right]}{\frac{1}{2m}\sum_i \left[(x_{Li} - \bar{X})^2 + (x_{Ui} - \bar{X})^2\right]} \tag{10.12}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

where $\bar{X}$ and $\bar{Y}$ are given in Equation 10.7. Brito (2007) calls the numerator $\tilde{S}_{XY}$ the co-dispersion measure and the denominator $\tilde{S}_X^2$ the dispersion measure, which is different from Equation 10.9. This regression line passes through the average center $(\bar{X}, \bar{Y})$, but the slope is guided by the range of the intervals, whose effect is summarized by the sum of the covariance between the lower bounds of $[x]_i$ and $[y]_i$ and the covariance between the upper bounds of $[x]_i$ and $[y]_i$. In other words, the researcher collects a sample of points as (x_{Li}, y_{Li}) and (x_{Ui}, y_{Ui}) and fits a unique regression line to the full sample. Equivalently, we can understand Brito's proposal as a constrained system of equations

$$\underset{2m\times 1}{\begin{bmatrix} Y_L \\ Y_U \end{bmatrix}} = \underset{2m\times k}{\begin{bmatrix} X_L \\ X_U \end{bmatrix}} \underset{k\times 1}{\beta} + \underset{2m\times 1}{\begin{bmatrix} \varepsilon_L \\ \varepsilon_U \end{bmatrix}}, \tag{10.13}$$

for which the OLS estimator is

$$\hat{\beta}_{OLS} = \left[X_L' X_L + X_U' X_U\right]^{-1} \left[X_L' Y_L + X_U' Y_U\right]. \tag{10.14}$$

However, the vector ε is likely to be heteroscedastic, i.e., $\sigma_L^2 \neq \sigma_U^2$

$$\Omega = \begin{pmatrix} \sigma_L^2 & \sigma_{LU} \\ \sigma_{LU} & \sigma_U^2 \end{pmatrix} \otimes I, \tag{10.15}$$

where $I_{m\times m}$ is the identity matrix. In this case, the GLS estimator $\hat{\beta}_{GLS} = [X'\Omega^{-1}X]^{-1}[X'\Omega^{-1}Y]$ would be more efficient than the OLS. A feasible GLS estimator will depend on the proposed model of heteroscedasticity. In the simplest heteroscedastic case, where $\sigma_L^2 \neq \sigma_U^2$, the estimated $\hat{\Omega}$ will be obtained by replacing the population moments σ_L^2, σ_U^2 and σ_{LU} with their sample counterparts.

An alternative proposal by Billard and Diday (2000, 2002) is to estimate two different regression lines, one for the minima and another for the maxima of the intervals with no restrictions across lines as in

$$\begin{aligned} y_{Li} &= \beta_L' x_{Li} + \varepsilon_{Li} \\ y_{Ui} &= \beta_U' x_{Ui} + \varepsilon_{Ui}. \end{aligned} \tag{10.16}$$

The estimation of the model proceeds by minimizing the following objective function

$$\min_{\hat{\beta}_L, \hat{\beta}_U} \sum_i \left(\hat{\varepsilon}_{Li}^2 + \hat{\varepsilon}_{Ui}^2\right), \tag{10.17}$$

which is equivalent to perform two separate minimizations, $\min_{\hat{\beta}_L} \sum_i \hat{\varepsilon}_{Li}^2$ and $\min_{\hat{\beta}_U} \sum_i \hat{\varepsilon}_{Ui}^2$ because of the absence of cross-equation restrictions. This approach can also be written as a system of seemingly unrelated regression

equations (SURE)

$$\underset{2m\times 1}{\begin{bmatrix} Y_L \\ Y_U \end{bmatrix}} = \underset{2m\times 2k}{\begin{bmatrix} X_L & 0 \\ 0 & X_U \end{bmatrix}} \underset{2k\times 1}{\begin{bmatrix} \beta_L \\ \beta_U \end{bmatrix}} + \underset{2m\times 1}{\begin{bmatrix} \varepsilon_L \\ \varepsilon_U \end{bmatrix}} \tag{10.18}$$

that is estimated by GLS, i.e., $\hat{\beta}_{GLS} = [X'\Omega^{-1}X]^{-1}[X'\Omega^{-1}Y]$. If $\Omega = I$ the GLS estimator reduces to the OLS estimator. However, given that $y_{Li} \leq y_{Ui}$ and $x_{Li} \leq x_{Ui}$, it is very likely that ε_{Li} and ε_{Ui} will be correlated and $\Omega \neq I$, thus the GLS estimator will be more efficient than the OLS. The feasible GLS will be constructed as in the previous approach. In practice, since there are not restrictions in the system, we could have some observations for which the estimated dependent variable is such that $\hat{y}_{Li} > \hat{y}_{Ui}$, which obviously contradicts the logic of interval data.

The last approach based on classical regression techniques is proposed by Lima Neto and de Carvalho (2008). It consists on running two independent regression models for the center and the radius (or range) of the intervals. Recall that $x_{Ci} = (x_{Li} + x_{Ui})/2$ and $x_{Ri} = (x_{Ui} - x_{Li})/2$. The model is

$$\begin{aligned} y_{Ci} &= \beta_C' x_{Ci} + \varepsilon_{Ci} \\ y_{Ri} &= \beta_R' x_{Ri} + \varepsilon_{Ri} \end{aligned} \tag{10.19}$$

and the objective function to minimize is

$$\min_{\hat{\beta}_C, \hat{\beta}_R} \sum_i \left(\hat{\varepsilon}_{Ci}^2 + \hat{\varepsilon}_{Ri}^2\right), \tag{10.20}$$

which, in the absence of cross-equation restrictions and with spherical disturbances, is equivalent to perform two separate minimizations, $\min_{\hat{\beta}_C} \sum_i \hat{\varepsilon}_{C_i}^2$ and $\min_{\hat{\beta}_R} \sum_i \hat{\varepsilon}_{Ri}^2$. The corresponding estimator is the classical OLS but the properties of the error term may dictate the choice of a GLS estimator, within a SURE system, as more appropriate than the OLS estimator. Other estimators as MLE or QMLE can also be implemented. However, the radius, being strictly positive, will not be normally distributed and a MLE estimator based on multivariate normality of the vector $(\varepsilon_{Ci}, \varepsilon_{Ri})'$ will be at least highly inefficient.

Figures 10.2 to 10.4 describe the graphical differences among the three regression lines proposed by Billard and Diday (2000, 2002) and Brito (2007). The proposal by Lima Neto and de Carvalho (2008) cannot be graphed in the same set of coordinates (X, Y).

10.2.3 The Prediction Problem

In this section, we define an interval-valued time series (ITS), we propose an approach to measure dissimilarities between intervals in ITS, and we implement forecasting methods for ITS based on smoothing filters and

nonparametric estimators like the k-NN. Neither of these two approaches aims to specify a model for an ITS that approximates a hidden data-generating mechanism, but rather they should be viewed as automatic procedures to extract information from a noisy signal from which eventually we can extrapolate a future value.

Definition 10.3 *An interval-valued stochastic process is a collection of interval random variables that are indexed by time, i.e.,* $\{X_t\}$ *for* $t \in T \subset \mathbb{R}$, *with each* X_t *following Definition 10.2.*

An interval-valued time series is a realization of an interval-valued stochastic process and it will be equivalently denoted as $\{[x]_t\} = \{[x_{Lt}, x_{Ut}]\} = \{\langle x_{Ct}, x_{Rt}\rangle\}$ for $t = 1, 2, \ldots, T$.

10.2.3.1 Accuracy of the Forecast

It is customary in classical time series to assess the forecast as a function of the difference between the realized value and the forecast value. In ITS, one may be tempted to calculate the difference $[x]_{t+1} - [\hat{x}]_{t+1}$ but, because the interval difference bounds all the possible results when considering single real numbers in the two operands, see property (Equation 10.2), the resulting interval will have an excessive width and thus, it will not be deemed appropriate to measure the accuracy of a forecast (Arroyo, Espínola, and Maté 2010). The following example will clarify this point.

Suppose that $[x]_{t+1} = [\hat{x}]_{t+1} = [a_L, a_U]$, $a_L < a_U$. Since the realized value is identical to the forecast, the forecast error must be zero $[x]_{t+1} - [\hat{x}]_{t+1} = [0, 0]$. If this difference is the interval difference (Equation 10.2), then it must be the case that $[A] = [a, a]$ with $a \in \mathbb{R}$, which is a contradiction with our assumption $a_L < a_U$. If $[a_L, a_U]$ is a nondegenerate interval, the result of the difference is an interval with the center in zero and with a length twice the length of the interval $[a_L, a_U]$, e.g., if $[a_L, a_U] = [1, 2]$, $[x]_t - [\hat{x}]_t = [-1, 1]$. Given these shortcomings, Arroyo and Maté (2006) propose the use of distances to quantify the dissimilarity (the forecast error) between the realized and the forecast intervals. The properties of distances, i.e., nonnegativity, symmetry, and triangle inequality, make them a suitable tool for this purpose. A distance, proposed by González et al. (2004), is defined as

$$D_K([x], [y]) = \frac{1}{\sqrt{2}}\sqrt{(x_L - y_L)^2 + (x_U - y_U)^2} = \sqrt{(x_C - y_C)^2 + (x_R - y_R)^2}, \tag{10.21}$$

which can be understood as an Euclidean-like distance considering the description of the intervals by their minimum and their maximum or, alternatively, by their center and by their radii. There is a large number of distances proposed in the literature, each with its advantages and disadvantages so that their use will depend on the needs of the researcher. In the forthcoming sections we will implement the Euclidean-type distance because of its intuitive and mathematical appeal.

Now, the assessment of a forecast will proceed by the choice of a distance measure and a loss function. Given a realized and a forecast ITS, $\{[x]_t\}$ and $\{[\hat{x}]_t\}$ with $t = 1, \ldots, T$, Arroyo, Espínola, and Maté (2010) propose the Mean Distance Error to quantify the accuracy of the forecast

$$MDE^q(\{[x]_t\}, \{[\hat{x}]_t\}) = \left(\frac{\sum_{t=1}^{T} (D^q([x]_t, [\hat{x}]_t))}{T} \right)^{\frac{1}{q}}, \tag{10.22}$$

where D is a distance such as D_K in Equation 10.21, and q is the order of the distance, such that for $q = 1$ the mean distance error is similar in spirit to the mean absolute error (MAE) loss function, and for $q = 2$ to the root mean squared error (RMSE) loss function. Other loss functions, statistical or economic/business based, can also be chosen to evaluate a forecast. The important point is that the quantification of the error should be based on a distance measure.

10.2.3.2 Smoothing Methods

Smoothing is a filtering technique that consists on averaging values of a time series, and by doing that, removing noise. These methods are easy to implement and they constitute a benchmark to evaluate the forecasting ability of more sophisticated methods (Gardner 2006). With the help of the arithmetic of intervals, it is relatively easy to adapt these smoothing procedures to ITS (Arroyo, Espínola, and Maté 2010). We begin with exponential smoothing though there is an even simpler smoothing procedure provided by just a moving average of order q.

10.2.3.2.1 Exponential Smoothing Given an ITS $\{[x_t]\}$ for $t = 1, 2, \ldots, T$, the forecast for the $t + 1$ period of a simple exponential smoothing in recursive form is written as

$$[\hat{x}]_{t+1} = \alpha[x]_t + (1 - \alpha)[\hat{x}]_t, \tag{10.23}$$

where $\alpha \in [0, 1]$. This representation weights the most recent observation and its forecast. In classic time series, the simple exponential smoothing can be equivalently represented in error correction form. However, with ITS both representations are not equivalent due to the properties of the interval arithmetic. To understand this difference, let us write the error correction representation

$$[\hat{x}]_{t+1} = [\hat{x}]_t + \alpha[e]_t, \tag{10.24}$$

where $[e]_t$ would be the interval error in t, $[e]_t = [x]_t - [\hat{x}]_t$. Due to the subdistributive property (Equation 10.5) of interval arithmetic, the relation between both expressions is the following

$$\alpha[x]_t + (1 - \alpha)[\hat{x}]_t \subseteq \alpha[x]_t - \alpha[\hat{x}]_t + [\hat{x}]_t = [\hat{x}]_t + \alpha([x]_t - [\hat{x}]_t), \tag{10.25}$$

which means that the recursive form yields tighter intervals than the error correction form. Due to this fact, the error correction form should not be considered in ITS forecasting. In addition, the error correction representation is not equivalent to the ITS moving average with exponentially decreasing weights, while the recursive form is. By backward substitution in Equation 10.23, and for t large, the simple exponential smoothing becomes

$$[\hat{x}]_{t+1} \simeq \sum_{j=1}^{t} \alpha(1-\alpha)^{j-1}[x]_{t-(j-1)}, \tag{10.26}$$

which is a moving average with exponentially decreasing weights.

Since the interval arithmetic subsumes the classical arithmetic, the smoothing methods for ITS subsume those for classic time series, so that if the intervals in the ITS are degenerated then the smoothing results will be identical to those obtained with the classical smoothing methods. When using Equation 10.23, all the components of the interval — center, radius, minimum, and maximum — are equally smoothed, i.e.,

$$\hat{x}_{\Gamma,t+1} = \alpha x_{\Gamma,t} + (1-\alpha)\hat{x}_{\Gamma,t} \quad \text{where } \Gamma \in \{L, U, C, R\}, \tag{10.27}$$

which means that, in a smoothed ITS, both the position and the width of the intervals will show less variability than in the original ITS, and that the smoothing factor will be the same for all components of the interval.

Additional smoothing procedures, like exponential smoothing with trend, or damped trend, or seasonality, can be adapted to ITS following the same principles presented in this section.

10.2.3.3 k-NN Method

The k-Nearest Neighbors (k-NN) method is a classic pattern recognition procedure that can be used for time series forecasting (Yakowitz 1987). The k-NN forecasting method in classic time series consists of two steps: identification of the k sequences in the time series that are more similar to the current one, and computation of the forecast as the weighted or unweighted average of the k-closest sequences determined in the previous step.

The adaptation of the k-NN method to forecast ITS consists of the following steps:

1. The ITS, $\{[x]_t\}$ with $t = 1, \ldots, T$, is organized as a series of d-dimensional interval-valued vectors

$$[x]_t^d = ([x]_t, [x]_{t-1}, \ldots, [x]_{t-(d-1)})', \tag{10.28}$$

   where $d \in \mathbb{N}$ is the number of lags.
2. We compute the dissimilarity between the most recent interval-valued vector $[x]_T^d = ([x]_T, [x]_{T-1}, \ldots, [x]_{T-d+1})'$ and the rest of the vectors in $\{[x]_t^d\}$. We use a distance measure to assess the dissimilarity between

vectors, i.e.,

$$D_t\left([x]_T^d, [x]_t^d\right) = \left(\frac{\sum_{i=1}^{d}\left(D^q([x]_{T-i+1}, [x]_{t-i+1})\right)}{d}\right)^{\frac{1}{q}}, \tag{10.29}$$

where $D([x]_{T-i+1}, [x]_{t-i+1})$ is a distance such as the kernel-based distance shown in Equation 10.21, q is the order of the measure that has the same effect that in the error measure shown in Equation 10.22.

3. Once the dissimilarity measures are computed for each $[x]_t^d$, $t = T-1, T-2, \ldots, d$, we select the k closest vectors to $[x]_T^d$. These are denoted by $[x]_{T_1}^d, [x]_{T_2}^d, \ldots, [x]_{T_k}^d$.
4. Given the k closest vectors, their subsequent values, $[x]_{T_1+1}, [x]_{T_2+1} \ldots, [x]_{T_k+1}$, are averaged to obtain the final forecast

$$[\hat{x}]_{T+1} = \sum_{p=1}^{k} \omega_p \cdot [x]_{T_p+1}, \tag{10.30}$$

where $[x]_{T_p+1}$ is the consecutive interval of the sequence $[x]_{T_p}^d$, and ω_p is the weight assigned to the neighbor p, with $\omega_p \geq 0$ and $\sum_{p=1}^{k} \omega_p = 1$. Equation 10.30 is computed according to the rules of interval arithmetic. The weights are assumed to be equal for all the neighbors $\omega_p = 1/k \forall p$, or inversely proportional to the distance between the last sequence $[x]_T^d$ and the considered sequence $[x]_{T_p}^d$

$$\omega_p = \frac{\psi_p}{\sum_{l=1}^{k} \psi_l}, \tag{10.31}$$

with $\psi_p = (D_{T_p}([x]_T^d, [x]_{T_p}^d) + \xi)^{-1}$ for $p = 1, \ldots, k$. The constant $\xi = 10^{-8}$ prevents the weight to explode when the distance between two sequences is zero.

The optimal values $\hat{k}$ and $\hat{d}$, which minimize the mean distance error (Equation 10.22) in the estimation period, are obtained by conducting a two-dimensional grid search.

10.2.4 Interval-Valued Dispersion: Low/High SP500 Prices

In this section, we apply the aforementioned interval regression and prediction methods to the daily interval time series of low/high prices of the SP500 index. We will denote the interval as $[p_{L,t}, p_{U,t}]$. There is strand in the financial literature — Parkinson (1980), Garman and Klass (1980), Ball and Torous (1984), Rogers and Satchell (1991), Yang and Zhang (2000), and Alizadeh, Brandt, and Diebold (2002) among others — that deals with functions of the range of the interval, $p_U - p_L$, in order to provide an estimator of the volatility σ of asset returns. In this chapter we do not pursue this route. The object of analysis is the interval $[p_{L,t}, p_{U,t}]$ itself and our goal is the construction of the

one-step-ahead forecast $[\hat{p}_{L,t+1}, \hat{p}_{U,t+1}]$. Obviously such a forecast can be an input to produce a forecast $\hat{\sigma}_{t+1}$ of volatility. One of the advantage of forecasting the low/high interval versus forecasting volatility is that the prediction error of the interval is based on observables as opposed to the prediction error for the volatility forecast for which "observed" volatility may be a problem.

The sample period goes from January 3, 2000 to September 30, 2008. We consider two sets of predictions:

1. Low volatility prediction set (year 2006): estimation period that goes from January 3, 2000 to December 30, 2005 (1508 trading days) and prediction period that goes from January 3, 2006 to December 29, 2006 (251 trading days).
2. High volatility prediction set (year 2008): estimation period that goes from January 2, 2002 to December 31, 2007 (1510 trading days) and prediction period that goes from January 2, 2008 to September 30, 2008 (189 trading days).

A plot of the first ITS $[p_{L,t}, p_{U,t}]$ is presented in Figure 10.5.

Following the classical regression approach to ITS, we are interested in the properties and time series regression models of the components of the interval, i.e., p_L, p_U, p_C, and p_R. We present the most significant and unrestricted time series models for $[p_{L,t}, p_{U,t}]$ and $\langle p_{C,t}, p_{R,t} \rangle$ in the spirit of the regression proposals of Billard and Diday (2000, 2002) and Lima Neto and de Carvalho (2008) reviewed in the previous sections. To save space we omit the univariate modeling of the components of the interval but these results are available upon request. However, we need to report that for p_L and p_U, we cannot reject a unit root, which is expected because these are price levels of the SP500, and that p_C has also a unit root because it is the sum of two unit root processes. In addition, p_L and p_U are cointegrated of order one with cointegrating vector $(1, -1)$, which implies that p_R is a stationary process given

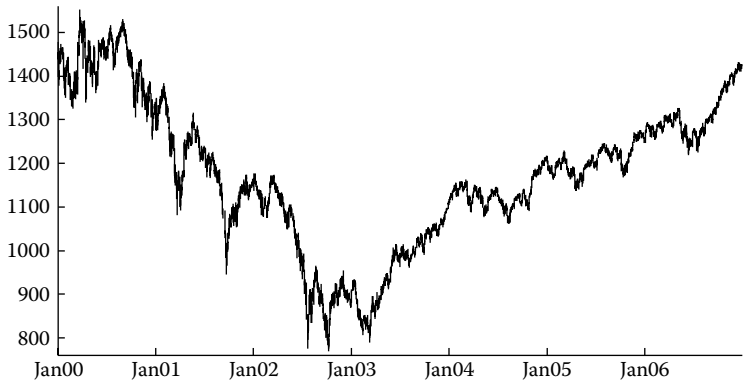

FIGURE 10.5
ITS of the weekly low/high from January 2000 to December 2006.

that $p_R = (p_U - p_L)/2$. Following standard model selection criteria and time series specification tools, the best model for $\langle \Delta p_{C,t},\ p_{R,t} \rangle$ is a VAR(3) and for $[p_{L,t},\ p_{U,t}]$ a VEC(3). The estimation results are presented in Tables A.1 and A.2 in the appendix.

In Table A.1, the estimation results for $\langle \Delta p_{C,t},\ p_{R,t} \rangle$ in both periods are very similar. The radius $p_{R,t}$ exhibits high autoregressive dependence and it is negatively correlated with the previous change in the center of the interval $\Delta p_{C,t-1}$ so that positive surprises in the center tend to narrow down the interval. On the other hand $\Delta p_{C,t}$ has little linear dependence and it is not affected by the dynamics of the radius. There is Granger causality from the center to the radius, but not vice versa. The radius equation enjoys a relative high adjusted R-squared of about 40% while the center is basically not linearly predictable. In general terms, there is a strong similarity between the modeling of $\langle \Delta p_{C,t},\ p_{R,t} \rangle$ and the most classical modeling of volatility with ARCH models for financial returns. The processes $p_{R,t}$ and the conditional variance of an asymmetric ARCH model, i.e., $\sigma^2_{t|t-1} = \alpha_0 + \alpha_1 \varepsilon^2_{t-1} + \alpha_2 \varepsilon_{t-1} + \beta \sigma^2_{t-1|t-2}$, share the autoregressive nature and the well-documented negative correlation of past innovations and volatility. The unresponsiveness of the center to the information in the dynamics of the radius is also similar to the findings in ARCH-in-mean processes where it is difficult to find significant effects of volatility on the return process.

In Table A.2, we report the estimation results for $[p_{L,t},\ p_{U,t}]$ for both periods 2000–2005 and 2002–2007. In general, there is much less linear dependence in the short-run dynamics of $[p_{L,t},\ p_{U,t}]$, which is expected as we are modeling financial prices. There is Granger-causality running both ways, from Δp_L to Δp_U and vice versa. Overall, the 2002-2007 period seems to be noisier (R-squared of 14%) than the 2000–2005 (R-squared of 20%–16%).

Based on the estimation results of the VAR(3) and VEC(3) models, we proceed to construct the one-step-ahead forecast of the interval $[\hat{p}_{L,t+1|t},\ \hat{p}_{U,t+1|t}]$. We also implement the exponential smoothing methods and the k-NN method for ITS proposed in the above sections and compare their respective forecasts. For the smoothing procedure, the estimated value of α is $\hat{\alpha} = 0.04$ in the estimation period 2000–2005 and $\hat{\alpha} = 0.03$ in 2002–2007. We have implemented the k-NN with equal weights and with inversely proportional as in Equation 10.31. In the period 2000–2005, the numbers of neighbors is $\hat{k} = 23$ (equal weights) and $\hat{k} = 24$ (proportional weights); in 2002–2007 $\hat{k} = 18$ for the k-NN with equal weights and $\hat{k} = 24$ for proportional weights. In both estimation periods, the length of the vector is $\hat{d} = 2$ for the k-NN with equal weights and $\hat{d} = 3$ for the proportional weights. The estimation of α, k, and d has been performed by minimizing the mean distance MDE (Equation 10.22) with $q = 2$. In both methods, smoothing and k-NN, the centers of the intervals have been first-differenced to proceed with the estimation and forecasting. However, in the following comparisons, the estimated differenced centers are transformed back to present the estimates and forecasts in levels. In Table 10.1 we show the performance of the five models measured by the MDE ($q = 2$) in the estimation and prediction

TABLE 10.1

Performance of the Forecasting Methods: MDE ($q = 2$)

	Period 2000–2006		Period 2002–2008	
Models	**Estimation 2000–2005**	**Prediction 2006**	**Estimation 2002–2007**	**Prediction 2008**
VAR(3)	9.359	6.611	7.614	15.744
VEC(3)	9.313	6.631	7.594	15.766
k-NN (eq.weights)	9.419	6.429	7.625	15.865
k-NN (prop.weights)	9.437	6.303	7.617	16.095
Smoothing	9.833	6.698	7.926	16.274
Naive	10.171	7.056	8.231	16.549

periods. We have also added a "naive" model that does not entail any estimation and whose forecast is the observation in the previous period, i.e., $[\hat{p}_{L,t+1|t},\ \hat{p}_{U,t+1|t}] = [p_{L,t},\ p_{U,t}]$.

For both low- and high-volatility periods the performance ranking of the six models is very similar. The worst performer is the naive model followed by the smoothing model. In 2006, the k-NN procedures are superior to the VAR(3) and VEC(3) models, but in 2008 the VAR and VEC systems perform slightly better than the k-NNs. The high-volatility year 2008 is clearly more difficult to forecast, the MDE in 2008 is twice as much as the MDE in the estimation period 2002–2007. On the contrary, in the low volatility year 2006, the MDE in the prediction period is about 30% lower than the MDE in the estimation period 2000–2005. A statistical comparison of the MDEs of the five models in relation to the naive model is provided by the Diebold and Mariano test of unconditional predictability (Diebold and Mariano 1995). The null hypothesis to test is the equality of the MDEs, i.e., $H_0 : E(D^2_{(\text{naive})} - D^2_{(\text{other})}) = 0$ versus $H_1 : E(D^2_{(\text{naive})} - D^2_{(\text{other})}) > 0$. If the null hypothesis is rejected the other model is superior to the naive model. The results of this test are presented in Table 10.2.

In 2006 all the five models are statistically superior to the benchmark naive model. In 2008 the smoothing procedure and the k-NN with proportional weights are statistically equivalent to the naive model while the remaining three models outperform the naive.

TABLE 10.2

Results of the Diebold and Mariano Test

	T-Test for $H_0 : E(D^2_{(\text{naive})} - D^2_{(\text{other})}) = 0$	
Models	**2006**	**2008**
VAR(3)	2.86	2.67
VEC(3)	2.26	2.46
k-NN(eq.weights)	3.55	2.43
k-NN(prop.weights)	4.17	1.79
Smoothing	5.05	1.15

We also perform a complementary assessment of the forecasting ability of the five models by running some regressions of the Mincer–Zarnowitz type. In the prediction periods, for the minimum p_L and the maximum p_U, we run separate regressions of the realized observations on the predicted observations as in $p_{L,t} = c + \beta \hat{p}_{L,t} + \varepsilon_t$ and $p_{U,t} = c + \beta \hat{p}_{U,t} + \upsilon_t$. Under a quadratic loss function, we should expect an unbiased forecast, i.e., $\beta = 1$ and $c = 0$. However, the processes $p_{L,t}$ and $\hat{p}_{L,t}$ are $I(1)$ and, as expected, cointegrated, so that these regressions should be performed with care. The point of interest is then to test for a cointegration vector of $(1, -1)$. To test this hypothesis using an OLS estimator with the standard asymptotic distribution, we need to consider that in the $I(1)$ process $\hat{p}_{L,t}$, i.e., $\hat{p}_{L,t} = \hat{p}_{L,t-1} + \nu_t$, the innovations ε_t and ν_t are not independent; in fact because $\hat{p}_{L,t}$ is a forecast of $p_{L,t}$ the correlation $\rho(\nu_{t+i}, \varepsilon_t) \neq 0$ for $i > 0$. To remove this correlation, the cointegrating regression will be augmented with some terms to finally estimate a regression as $p_{L,t} = c + \beta \hat{p}_{L,t} + \sum_i \gamma_i \Delta \hat{p}_{L,t+i} + e_t$ (the same argument applies to $p_{U,t}$). The hypothesis of interest is $H_0 : \beta = 1$ versus $H_1 : \beta \neq 1$. A t-statistic for this hypothesis will be asymptotically standard normal distributed. We may also need to correct the t-test if there is some serial correlation in e_t. In Table 10.3 we present the testing results.

We reject the null for the smoothing method for both prediction periods and for both $p_{L,t}$ and $p_{U,t}$ processes. Overall the prediction is similar for 2006 and 2008. The VEC(3) and the k-NN methods deliver better forecasts across the four instances considered. For those models in which we fail to reject $H_0 : \beta = 1$, we also calculate the unconditional average difference between the realized and the predicted values, i.e, $\bar{p} = \sum_t (p_t - \hat{p}_t)/T$. The magnitude of this average is in the single digits, so that for all purposes, it is insignificant given that the level of the index is in the thousands. In Figure 10.6 we show the k-NN (equal weights)-based forecast of the interval low/high of the SP500 index for November and December 2006.

TABLE 10.3

Results of the t-Test for Cointegrating Vector $(1, -1)$

	Asymptotic (Corrected) t-Test $H_0 : \beta = 1$ versus $H_1 : \beta \neq 1$ $p_t = c + \beta \hat{p}_t + \sum_i \gamma_i \Delta \hat{p}_{t+i} + e_t$			
	2006		**2008**	
	min: $p_{L,t}$	**max: $p_{U,t}$**	**min: $p_{L,t}$**	**max: $p_{U,t}$**
VAR(3)	3.744*	−1.472	3.024*	−2.712*
VEC(3)	1.300	0.742	2.906*	−2.106
k-NN (eq.weights)	0.639	−4.191*	1.005	−2.270
k-NN (prop.weights)	3.151*	−2.726*	1.772	−1.731
Smoothing	−3.542*	−2.544*	2.739*	−3.449*

*Rejection of the null hypothesis at the 1% significance level.

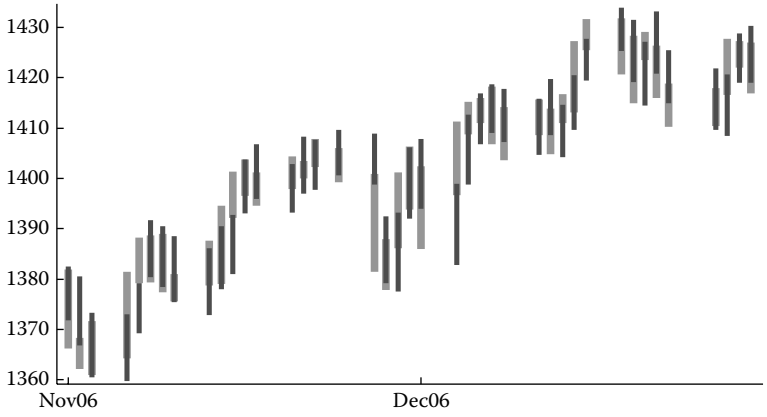

FIGURE 10.6
k-NN based forecast (black) of the low/high prices of the SP500; realized ITS (grey).

10.3 Histogram Data

In this section, our premise is that the data is presented to the researcher as a frequency distribution, which may be the result of an aggregation procedure, or the description of a population or any other grouped collective. We start by describing histogram data and some univariate descriptive statistics. Our main objective is to present the prediction problem by defining a histogram time series (HTS) and implementing smoothing techniques and nonparametric methods like the k-NN algorithm. As we have seen in the section on interval data, these two methods require the calculation of suitable averages. To this end, instead of relying on the arithmetic of histograms, we introduce the barycentric histogram that is an average of a set of histograms. The choice of appropriate distance measures is key to the calculation of the barycenter, and eventually of the forecast of a HTS.

10.3.1 Preliminaries

Given a variable of interest X, we collect information on a group of individuals or units that belong to a set S. For every element $i \in S$, we observe a datum such as

$$h_{X_i} = \{([x]_{i1}, \pi_{i1}), \ldots, ([x]_{in_i}, \pi_{in_i})\}, \quad \text{for } i \in S, \tag{10.32}$$

where π_{ij} , $j = 1, \ldots, n_i$ 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 $x_{Ui\,j-1} \leq x_{Lij}$ $\forall i, j$, for $j \geq 2$. The datum h_{X_i} is a histogram and the data set will be a collection of histograms $\{h_{X_i}, i = 1, \ldots, m\}$.

As in the case of interval data, we could summarize the histogram data set by its empirical density function from which the sample mean and the sample variance can be calculated (Billard and Diday 2006). The sample mean is

$$\bar{X} = \frac{1}{2m} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_{Uij} + x_{Lij}) \pi_{ij}, \tag{10.33}$$

which is the average of the weighted centers for each interval; and the sample variance is

$$S_X^2 = \frac{1}{3m} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left(x_{Uij}^2 + x_{Uij} x_{Lij} + x_{Lij}^2\right) \pi_{ij} - \frac{1}{4m^2} \left[\sum_{i=1}^{m} \sum_{j=1}^{n_i} (x_{Uij} + x_{Lij}) \pi_{ij} \right]^2,$$

which combines the variability of the centers as well as the intra-interval variability. Note that the main difference between these sample statistics and those in Equations 10.7 and 10.9 for interval data is the weight provided by the frequency $\pi_{i,j}$ associated with each interval $[x]_{i,j}$.

Next, we proceed with the definition of a histogram random variable. Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is the set of elementary events, $\mathcal{F}$ is the σ-field of events and $P : \mathcal{F} \to [0, 1]$ the σ-additive probability measure; and define a partition of Ω into sets $A_X(x)$ such that $A_X(x) = \{\omega \in \Omega | X(\omega) = x\}$, where $x \in \{h_{X_i}, i = 1, \ldots, m\}$.

Definition 10.4 *A mapping $h_X : \mathcal{F} \to \{h_{X_i}\}$, such that, for all $x \in \{h_{X_i}, i = 1 \ldots .m\}$ there is a set $A_X(x) \in \mathcal{F}$, is called a histogram random variable.*

Then, the definition of stochastic process follows as:

Definition 10.5 *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 10.4.*

A histogram-valued time series is a realization of a histogram-valued stochastic process and it will be equivalently denoted as $\{h_{X_t}\} \equiv \{h_{X_t}, \ t = 1, 2, \ldots, T\}$.

10.3.2 The Prediction Problem

In this section, we propose a dissimilarity measure for HTS based on a distance. We present two distance measures that will play a key role in the estimation and prediction stages. They will also be instrumental to the definition of a barycentric histogram, which will be used as the average of a set of histograms. Finally, we will present the implementation of the prediction methods.

10.3.2.1 Accuracy of the Forecast

Suppose that we construct a forecast for $\{h_{X_t}\}$, which we denote as $\{\hat{h}_{X_t}\}$. It is sensible to define the forecast error as the difference $h_{X_t} - \hat{h}_{X_t}$. However, the difference operator based on histogram arithmetic (Colombo and Jaarsma 1980) does not provide information on how dissimilar the histograms h_{X_t} and $\hat{h}_{X_t}$ are. In order to avoid this problem, Arroyo and Maté (2009) propose the mean distance error (MDE), which in its most general form is defined as

$$MDE^q(\{h_{X_t}\}, \{\hat{h}_{X_t}\}) = \left(\frac{\sum_{t=1}^{T} D^q(h_{X_t}, \hat{h}_{X_t})}{T} \right)^{\frac{1}{q}}, \tag{10.34}$$

where $D(h_{X_t}, \hat{h}_{X_t})$ is a distance measure such as the Wasserstein or the Mallows distance to be defined shortly and q is the order of the measure, such that for $q = 1$ the resulting accuracy measure is similar to the MAE and for $q = 2$ to the RMSE.

Consider two density functions, $f(x)$ and $g(x)$, with their corresponding cumulative distribution functions (CDF), $F(x)$ and $G(x)$, the Wasserstein distance between $f(x)$ and $g(x)$ is defined as

$$D_W(f, g) = \int_0^1 |F^{-1}(t) - G^{-1}(t)| dt, \tag{10.35}$$

and the Mallows as

$$D_M(f, g) = \sqrt{\int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt}, \tag{10.36}$$

where $F^{-1}(t)$ and $G^{-1}(t)$ with $t \in [0, 1]$ are the inverse CDFs of $f(x)$ and $g(x)$, respectively. The dissimilarity between two functions is essentially measured by how far apart their t-quantiles are, i.e., $F^{-1}(t) - G^{-1}(t)$. In the case of Wasserstein, the distance is defined in the L_1 norm and in the Mallows in the L_2 norm. When considering Equation 10.34, $D(h_{X_t}, \hat{h}_{X_t})$ will be calculated by implementing the Wasserstein or Mallows distance. By using the definition of the CDF of a histogram in Billard and Diday (2006), the Wasserstein and Mallows distances between two histograms h_X and h_Y can be written analytically as functions of the centers and radii of the histogram bins, i.e.,

$$D_W(h_X, h_Y) = \sum_{j=1}^{n} \pi_j |x_{Cj} - y_{Cj}| \tag{10.37}$$

$$D_M^2(h_X, h_Y) = \sum_{j=1}^{n} \pi_j \left[(x_{Cj} - y_{Cj})^2 + \frac{1}{3}(x_{Rj} - y_{Rj})^2 \right]. \tag{10.38}$$

10.3.2.2 The Barycentric Histogram

Given a set of K histograms h_{X_k} with $k = 1, \ldots, K$, the barycentric histogram h_{X_B} is the histogram that minimizes the distances between itself and all the K histograms in the set. The optimization problem is

$$\min_{h_{X_B}} \sum_{k=1}^{K} \left[D^r(h_{X_k}, h_{X_B}) \right]^{1/r}, \tag{10.39}$$

where $D(h_{X_k}, h_{X_B})$ is a distance measure. The concept is introduced by Irpino and Verde (2006) to define the prototype of a cluster of histogram data. As Verde and Irpino (2007) show, the choice of the distance determine the properties of the barycenter.

When the chosen distance is Mallows, for $r = 2$, the optimal barycentric histogram $h^*_{X_B}$ has the following center/radius characteristics. Once the k histograms are rewritten in terms of n^* bins, for each bin $j = 1, \ldots, n^*$, the barycentric center x^*_{Cj} is the mean of the centers of the corresponding bin in each histogram and the barycentric radius x^*_{Rj} is the mean of the radii of the corresponding bin in each of the K histograms,

$$x^*_{Cj} = \frac{\sum_{k=1}^{K} x_{Ckj}}{K} \tag{10.40}$$

$$x^*_{Rj} = \frac{\sum_{k=1}^{K} x_{Rkj}}{K}. \tag{10.41}$$

When the distance is Wasserstein, for $r = 1$ and for each bin $j = 1, \ldots, n^*$, the barycentric center x^*_{Cj} is the median of the centers of the corresponding bin in each of the K histograms,

$$x^*_{Cj} = \text{median}(x_{Ckj}) \quad \text{for } k = 1, \ldots, K \tag{10.42}$$

and the radius x^*_{Rj} is the corresponding radius of the bin where the median x^*_{Cj} falls among the K histograms. For more details on the optimization problem, please see Arroyo and Maté (2009).

10.3.2.3 Exponential Smoothing

The exponential smoothing method can be adapted to histogram time series by replacing averages with the barycentric histogram, as it was shown in Arroyo and Maté (2008).

Let $\{h_{X_t}\}\, t = 1, \ldots, T$ be a histogram time series, the exponentially smoothed forecast is given by the following equation

$$\hat{h}_{X_{t+1}} = \alpha h_{X_t} + (1 - \alpha)\hat{h}_{X_t}, \tag{10.43}$$

where $\alpha \in [0, 1]$. Since the right-hand side is a weighted average of histograms, we can use the barycenter approach so that the forecast is the solution

to the following optimization exercise

$$\hat{h}_{X_{t+1}} \equiv \arg\min_{\hat{h}_{X_{t+1}}} \left(\alpha D^2(\hat{h}_{X_{t+1}}, h_{X_t}) + (1-\alpha) D^2(\hat{h}_{X_{t+1}}, \hat{h}_{X_t})\right)^{1/2}, \quad (10.44)$$

where $D(\cdot, \cdot)$ is the Mallows distance. The use of the Wasserstein distance is not suitable in this case because of the properties of the median, which will ignore the weighting scheme (with the exception of $\alpha = 0.5$) so intrinsically essential to the smoothing technique. For further developments of this issue see Arroyo, González-Rivera, Maté and Muñoz-San Roque (2010).

For t large, the recursive form (Equation 10.43) can be easily rewritten as a moving average

$$\hat{h}_{X_{t+1}} \simeq \sum_{j=1}^{t} \alpha(1-\alpha)^{j-1} h_{X_{t-(j-1)}}, \quad (10.45)$$

which in turn can also be expressed as the following optimizations problem

$$\hat{h}_{X_{t+1}} \equiv \arg\min_{\hat{h}_{X_{t+1}}} \left[\sum_{j=1}^{t} \alpha(1-\alpha)^{j-1} D^2(\hat{h}_{X_{t+1}}, h_{X_{t-(j-1)}})\right]^{1/2}, \quad (10.46)$$

with $D(\cdot, \cdot)$ as the Mallows distance. The Equations 10.44 and 10.46 are equivalent.

Figure 10.7 shows an example of the exponential smoothing using Equation 10.44 for the histograms $h_{X_t} = \{([19, 20), 0.1), ([20, 21), 0.2), ([21, 22], 0.7)\}$ and $\hat{h}_{X_t} = \{([0, 3), 0.35), ([3, 6), 0.3), ([6, 9], 0.35)\}$ with $\alpha = 0.9$ and $\alpha = 0.1$. In both cases, the resulting histogram averages the location, the support, and the shape of both histograms h_{X_t} and $\hat{h}_{X_t}$ in a suitable way.

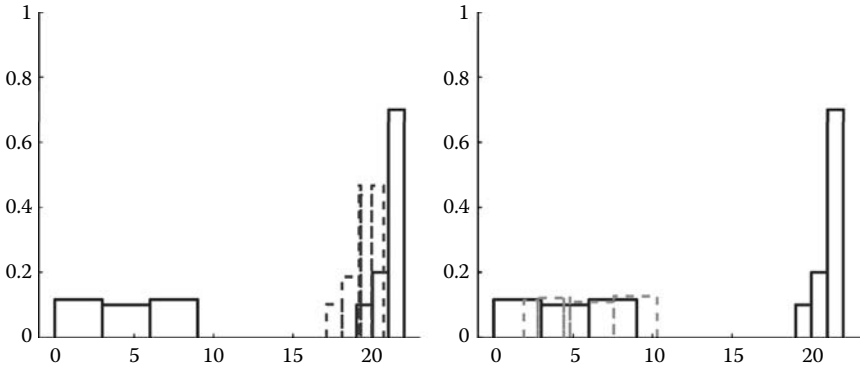

FIGURE 10.7
Exponential smoothing of histograms using the recursive formulation with $\alpha = 0.9$ (left) and $\alpha = 0.1$ (right). In each part of the figure, the barycenter is the dash-lined histogram.

10.3.2.4 k-NN Method

The adaptation of the k-NN method to forecast HTS was proposed by Arroyo and Maté (2009). The method consists of similar steps to those described in the interval section:

1. The HTS, $\{h_{X_t}\}$ with $t = 1, \ldots, T$, is organized as a series of d-dimensional histogram-valued vectors $\{h^d_{X_t}\}$ where

$$h^d_{X_t} = (h_{X_t}, h_{X_{t-1}}, \ldots, h_{X_{t-(d-1)}})', \tag{10.47}$$

   where $d \in \mathbb{N}$ is the number of lags and $t = d, \ldots, T$.
2. We compute the dissimilarity between the most recent histogram-valued vector $h^d_{X_T} = (h_{X_T}, h_{X_{T-1}}, \ldots, h_{X_{T-(d-1)}})'$ and the rest of the vectors in $\{h^d_{X_t}\}$ by implementing the following distance measure

$$D_t(h^d_{X_T}, h^d_{X_t}) = \left(\frac{\sum_{i=1}^{d} \left(D^q(h_{X_{T-i+1}}, h_{X_{t-i+1}}) \right)}{d} \right)^{\frac{1}{q}}, \tag{10.48}$$

   where $D^q(h_{X_{T-i+1}}, h_{X_{t-i+1}})$ is the Mallows or the Wasserstein distance of order q.
3. Once the dissimilarity measures are computed for each $h^d_{X_t}$, $t = T - 1, T - 2, \ldots, d$, we select the k closest vectors to $h^d_{X_T}$. These are denoted by $h^d_{X_{T_1}}, h^d_{X_{T_2}}, \ldots, h^d_{X_{T_k}}$.
4. Given the k closest vectors, their subsequent values, $h_{X_{T_1+1}}, h_{X_{T_2+1}}, \ldots, h_{X_{T_k+1}}$, are averaged by means of the barycenter approach to obtain the final forecast $\hat{h}_{X_{T+1}}$ as in

$$\hat{h}_{X_{T+1}} \equiv \arg\min_{\hat{h}_{X_{T+1}}} \left[\sum_{p=1}^{k} \omega_p D^r(\hat{h}_{X_{T+1}}, h_{X_{T_p+1}}) \right]^{1/r}, \tag{10.49}$$

   where $D(\hat{h}_{X_{T+1}}, h_{X_{T_p+1}})$ is the Mallows distance with $r = 2$ or the Wasserstein distance with $r = 1$, $h_{X_{T_p+1}}$ is the consecutive histogram in the sequence $h^d_{X_{T_p}}$, and ω_p is the weight assigned to the neighbor p, with $\omega_p \geq 0$ and $\sum_{p=1}^{k} \omega_p = 1$. As in the case of the interval-valued data, the weights may be assumed to be equal for all the neighbors $\omega_p = 1/k \ \forall p$, or inversely proportional to the distance between the last sequence $h^d_{X_T}$ and the considered sequence $h^d_{X_{T_p}}$.

The optimal values, $\hat{k}$ and $\hat{d}$, which minimize the mean distance error (Equation 10.34) in the estimation period, are obtained by conducting a two-dimensional grid search.

10.3.3 Histogram Forecast for SP500 Returns

In this section, we implement the exponential smoothing and the k-NN methods to forecast the one-step-ahead histogram of the returns to the constituents of the SP500 index. We collect the weekly returns of the 500 firms in the index from 2002 to 2005. We divide the sample into an estimation period of 156 weeks running from January 2002 to December 2004, and a prediction period of 52 weeks that goes from January 2005 to December 2005. The histogram data set consists of 208 weekly equiprobable histograms. Each histogram has four bins, each one containing 25% of the firms' returns.

For the smoothing procedure, the estimated value of α is $\hat{\alpha} = 0.13$. We have implemented the k-NN with equal weights and with inversely proportional as in Equation 10.31 using the Mallows and Wasserstein distances. With the Mallows distance, the estimated numbers of neighbors is $\hat{k} = 11$ and the length of the vector is $\hat{d} = 9$ for both weighting schemes. With the Wasserstein distance, $\hat{k} = 12$, $\hat{d} = 9$ (equal weights), and $\hat{k} = 17$, $\hat{d} = 8$ (proportional weights). The estimation of α, k, and d has been performed by minimizing the Mallows MDE with $q = 1$, except for the Wasserstein-based k-NN which used the Wasserstein MDE with $q = 1$. In Table 10.4, we show the performance of the five models measured by the Mallows-based MDE ($q = 1$) in the estimation and prediction periods. We have also added a "naive" model that does not entail any estimation and for which the one-step-ahead forecast is the observation in the previous period, i.e., $\hat{h}_{X_{t+1|t}} = h_{X_t}$.

In the estimation and prediction period, the naive model is clearly outperformed by the rest of the five models. In the estimation period, the five models exhibit similar performance with a MDE of 4.9 approximately. In the prediction period, the exponential smoothing and the Wasserstein-based k-NN seem to be superior to the Mallows-based k-NN. We should note that the MDEs in the prediction period are about 11% lower than the MDEs in the estimation period.

For the prediction year 2005, we provide a statistical comparison of the MDEs of the five models in relation to the naive model by implementing the Diebold and Mariano test of unconditional predictability (Diebold and Mariano 1995). The null hypothesis to test is the equality of the MDEs, i.e., $H_0 : E(D_{(\text{naive})} - D_{(\text{other})}) = 0$ versus $H_1 : E(D_{(\text{naive})} - D_{(\text{other})}) > 0$. If the null

TABLE 10.4

Performance of the Forecasting Methods: MDE ($q = 1$)

Models	Estimation 2002–2004	Prediction 2005
Mall. k-NN (eq.weights)	4.988	4.481
Mall. k-NN (prop.weights)	4.981	4.475
Wass. k-NN (eq.weights)	4.888	4.33
Wass. k-NN (prop.weights)	4.882	4.269
Exp. Smoothing	4.976	4.344
Naive	6.567	5.609

TABLE 10.5

Results of the Diebold and Mariano Test

	t-Test for $H_0 : E(D_{(\text{naive})} - D_{(\text{other})}) = 0$
Models	**2005 Prediction Year**
Mall. k-NN(eq.weights)	2.32
Mall. k-NN(prop.weights)	2.69
Wass. k-NN(eq.weights)	2.29
Wass. k-NN(prop.weights)	2.29
Exp. smoothing	3.08

hypothesis is rejected, the "other" model is superior to the naive model. The results of this test are presented in Table 10.5.

In 2005, all the five models are statistically superior to the benchmark naive model, though the rejection of the null is stronger for the exponential smoothing and the Mallows-based k-NN models with proportional weights.

In Figure 10.8, we present the 2005 one-step-ahead histogram forecast obtained with the exponential smoothing procedure and we compare it to the realized value. For each time period, we draw two histograms: the realized histogram (the right one) and the forecast histogram (the left one). Overall the forecast follows very closely the realized value except for those observations that have extreme returns. The fit can be further appreciated when we zoom in the central 50% mass of the histograms (Figure 10.9).

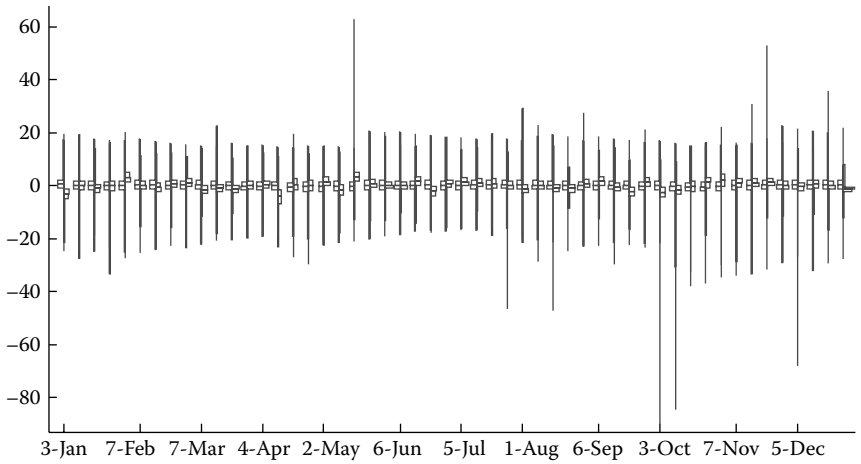

FIGURE 10.8
2005 realized histograms (the right ones) and exponential smoothed one-step-ahead histogram forecasts (the left ones) for the HTS of SP500 returns. Weekly data.

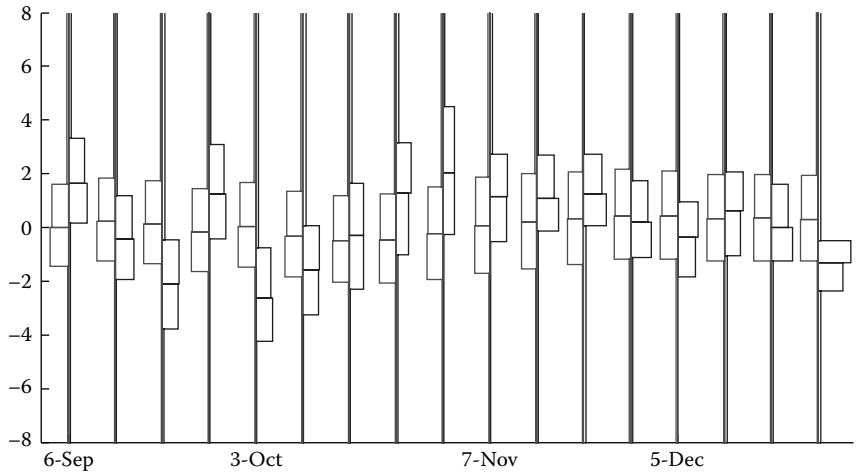

FIGURE 10.9
Zoom of Figure 10.8 from September to December 2005.

10.4 Summary and Conclusions

Large databases prompt the need for new methods of processing information. In this article we have introduced the analysis of interval-valued and histogram-valued data sets as an alternative to classical single-valued data sets and we have shown the promise of this approach to deal with economic and financial data.

With interval data, most of the current efforts have been directed to the adaptation of classical regression models as the interval is decomposed into two single-valued variables, either the center/radius or the min/max. The advantage of this decomposition is that classical inferential methods are available. Methodologies that analyze the interval per se fall into the realm of random sets theory and though there is some important research on regression analysis with random sets, inferential procedures are almost nonexistent. Being our current focus is the prediction problem, we have explored two different venues to produce a forecast with interval time series (ITS). First, we have implemented the classical regression approach to the analysis of ITS, and secondly we have proposed the adaptation to ITS of filtering techniques, such as smoothing, and nonparametric methods, such as the k-NN, to ITS. The latter venue requires the use of interval arithmetic to construct the appropriate averages and the introduction of distance measures to assess the dissimilarity between intervals and to quantify the prediction error. We have implemented these ideas with the SP500 index. We modeled the

center/radius time series and the low/high time series of what we called interval-valued dispersion of the SP500 index and compared their one-step-ahead forecasts to those of a smoothing procedure and k-NN methods. A VEC model for the low/high series and the k-NN methods have the best forecasting performance.

With histogram data, the analysis becomes more complex. Regression analysis with histograms is in its infancy and the venues for further developments are large. We have focused exclusively in the prediction problem with smoothing methods and nonparametric methods. A key concept for the implementation of these two procedures is the introduction of the barycentric histogram that is a device that works as an average (weighted or unweighted) of a set of histograms. As with ITS, the introduction of the appropriate distances to judge dissimilarities among histograms and to assess forecast errors are fundamental ingredients in the analysis. The collection over time of cross-sectional returns of the firms in the SP500 index provides a nice histogram time series (HTS), on which we have implemented the aforementioned methods to eventually produce the one-step-ahead histogram forecast. Simple smoothing techniques seem to work remarkably well.

There are still many unexplored areas in ITS and HTS. A very important question is the search for a model. This will require the understanding of the notion of dependence in ITS and HTS. A first step in this direction is provided by González-Rivera and Arroyo (2010) who construct autocorrelation functions for HTS and ITS. From an econometric point of view, model building requires further research on identification, estimation, testing, and model selection procedures. Economic and financial questions will benefit greatly from this new approach to the analysis of large data sets.

10.5 Acknowledgment

We thank the referees and the editors for useful and constructive comments. Arroyo acknowledges support from the Spanish Council for Science and Innovation (grant TIN2008-06464-C03-01) and from the *Programa de Creación y Consolidación de Grupos de Investigación UCM-Banco Santander*. González-Rivera acknowledges the financial support provided by the University Scholar Award. Maté acknowledges the financial support provided by the project "Forecasting models from symbolic data (PRESIM)" from Universidad Pontificia Comillas.

Appendix

Estimation Results for ITS SP500 Index

TABLE A.1

Estimation of the VAR(3) Model for the Differenced Center and Radius Time Series

Estimation Sample 2000–2005			Estimation Sample 2002–2007		
VAR	**D(Cen)**	**Rad**	**VAR**	**D(Cen)**	**Rad**
D(Cen(−1))	0.33218	−0.09764	**D(Cen(−1))**	0.279225	−0.074092
	0.0262	0.00997		0.02619	0.00978
	[12.6803]	[−9.79410]		[10.6611]	[−7.57934]
D(Cen(−2))	−0.181348	−0.001809	**D(Cen(−2))**	−0.092471	−0.010534
	0.02742	0.01043		0.02713	0.01012
	[−6.61378]	[−0.17332]		[−3.40879]	[−1.04037]
D(Cen(−3))	0.050564	0.00429	**D(Cen(−3))**	0.006178	−0.013364
	0.02616	0.00996		0.02629	0.00981
	[1.93281]	[0.43091]		[0.23500]	[−1.36214]
Rad(−1)	0.066659	0.150616	**Rad(−1)**	−0.00284	0.152907
	0.06593	0.02509		0.06731	0.02512
	[1.01103]	[6.00287]		[−0.04219]	[6.08652]
Rad(-2)	−0.049629	0.313259	**Rad(-2)**	0.046537	0.27345
	0.06319	0.02405		0.0649	0.02422
	[−0.78541]	[13.0270]		[0.71705]	[11.2886]
Rad(-3)	0.129442	0.285272	**Rad(-3)**	−0.01386	0.276629
	0.0648	0.02466		0.06635	0.02477
	[1.99747]	[11.5678]		[−0.20888]	[11.1699]
C	−1.319847	2.088036	**C**	−0.045805	2.074405
	0.60607	0.23064		0.5355	0.19987
	[−2.17772]	[9.05315]		[−0.08554]	[10.3788]

TABLE A.2

Estimation of the VEC(3) Model for Low/High Time Series

Estimation Sample 2000–2005			Estimation Sample 2002–2007		
Error Correction:	**D(Low)**	**D(High)**	**Error Correction:**	**D(Low)**	**D(High)**
CointEq1	−0.438646	0.007023	**CointEq1**	−0.124897	0.121926
	0.05364	0.04758		0.04103	0.03692
	[−8.17770]	[0.14761]		[−3.04419]	[3.30283]
D(Low(-1))	0.112549	0.515586	**D(Low(-1))**	−0.165406	0.425054
	0.05429	0.04816		0.0489	0.044
	[2.07293]	[10.7050]		[−3.38238]	[9.66024]
D(Low(-2))	−0.093605	0.193326	**D(Low(-2))**	−0.314249	0.130253
	0.0505	0.0448		0.04863	0.04375
	[−1.85344]	[4.31532]		[−6.46233]	[2.97698]
D(Low(-3))	0.026446	0.112943	**D(Low(-3))**	−0.15041	0.061275
	0.0396	0.03512		0.0399	0.0359
	[0.66790]	[3.21547]		[−3.76992]	[1.70691]

TABLE A.2 (Continued)

Estimation of the VEC(3) Model for Low/High Time Series

Estimation Sample 2000–2005			Estimation Sample 2002–2007		
Error Correction:	**D(Low)**	**D(High)**	**Error Correction:**	**D(Low)**	**D(High)**
D(High(-1))	0.313542	−0.287591	**D(High(-1))**	0.524179	−0.221533
	0.05905	0.05238		0.05188	0.04668
	[5.30959]	[−5.49018]		[10.1046]	[−4.74625]
D(High(-2))	−0.073453	−0.382411	**D(High(-2))**	0.248088	−0.239401
	0.05604	0.04971		0.05323	0.04789
	[−1.31078]	[−7.69307]		[4.66085]	[−4.99871]
D(High(-3))	0.04646	−0.065429	**D(High(-3))**	0.182654	−0.073329
	0.04356	0.03864		0.04262	0.03835
	[1.06663]	[−1.69337]		[4.28593]	[−1.91234]
C	−0.064365	−0.118124	**Cointegrating Eq:**	**CointEq1**	
	0.28906	0.25642	**Low(-1)**	1	
	[−0.22267]	[−0.46068]	**High(-1)**	−1.002284	
Cointegrating Eq:	**Co-intEq1**			0.00318	
Low(-1)	1			[−315.618]	
High(-1)	−1.001255		**C**	16.82467	
	0.00268			3.81466	
	[−373.870]			[4.41053]	
@TREND(1)	−0.012818				
	0.00105				
	[−12.1737]				
C	27.97538				

References

Alizadeh, S., M. Brandt, and F. Diebold. 2002. Range-based estimation of stochastic volatility models. *Journal of Finance* 57(3):1047–1091.

Arroyo, J., R. Espínola, and C. Maté. 2010. Different approaches to forecast interval time series: a comparison in finance. *Computational Economics* Forthcoming.

Arroyo, J., G. González–Rivera, C. Maté, and A. Muñoz–San Roque. 2010. Smoothing methods for histogram-valued time series. An application to value-at-risk. Working paper, Department of Computer Science and Artificial Intelligence, Universidad Complutense de Madrid, Madrid. Spain.

Arroyo, J., and C. Maté. 2006. Introducing interval time series: accuracy measures. In *COMPSTAT 2006, Proceedings in Computational Statistics*. Heidelberg: Physica-Verlag, pp. 1139–1146.

Arroyo, J., and C. Maté. 2008. Forecasting time series of observed distributions with smoothing methods based on the barycentric histogram. In *Computational Intelligence in Decision and Control. Proceedings of the 8th International FLINS Conference*, pp. 61–66. Singapore: World Scientific.

Arroyo, J., and C. Maté. 2009. Forecasting histogram time series with k-nearest neighbours methods. *International Journal of Forecasting 25*(1):192–207.

Ball, C., and W. Torous. 1984. The maximim likelihood estimation of security price volatility: theory, evidence, and an application to option pricing. *Journal of Business* 57(1):97–112.

Bertrand, P., and F. Goupil. 2000. Descriptive statistics for symbolic data, *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*. Berlin: Springer. pp. 103–124.

Billard, L., and E. Diday. 2000. Regression analysis for interval-valued data. In *Data Analysis, Classification and Related Methods: Proceedings of the 7th Conference of the IFCS, IFCS 2002*. Berlin: Springer. pp. 369–374.

Billard, L., and E. Diday. 2002. Symbolic regression analysis. In *Classification, Clustering and Data Analysis: Proceedings of the 8th Conference of the IFCS, IFCS 2002*. Berlin: Springer. pp. 281–288.

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*. 1st ed. Chichester: Wiley & Sons.

Blanco, Á., A. Colubi, N. Corral, and G. González-Rodríguez. 2008. On a linear independence test for interval-valued random sets. In *Soft Methods for Handling Variability and Imprecision. Advances in Soft Computing*. Berlin: Springer. pp. 111–117.

Brito, P. 2007. Modelling and analysing interval data. In *Proceedings of the 30th Annual Conference of GfKl*. Berlin: Springer. pp. 197–208.

Cheung, Y.-L., Y.-W. Cheung, and A. T. K. Wan. 2009. A high-low model of daily stock price ranges. *Journal of Forecasting* 28(2):103–119.

Colombo, A., and R. Jaarsma. 1980. A powerful numerical method to combine random variables. *IEEE Transactions on Reliability* 29(2):126–129.

Diebold, F. X., and R. S. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13(3):253–263.

García-Ascanio, C., and C. Maté. 2010. Electric power demand forecasting using interval time series: a comparison between VAR and iMLP. *Energy Policy* 38(2):715–725.

Gardner, E. S. 2006. Exponential smoothing: the state of the art. Part 2. *International Journal of Forecasting* 22(4): 637–666.

Garman, M. B., and M. J. Klass. 1980. On the estimation of security price volatilities from historical data. *Journal of Business* 53(1):67–78.

González, L., F. Velasco, C. Angulo, J. A. Ortega, and F. Ruiz. 2004. Sobrenúcleos, distancias y similitudes entre intervalos. *Inteligencia Artificial, Revista Iberoamericana de IA* 8(23):111–117.

González-Rivera, G., and J. Arroyo. 2010. 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.-H. Lee, and S. Mishra. 2008. Jumps in cross-sectional rank and expected returns: a mixture model. *Journal of Applied Econometrics* 23:585–606.

González-Rodríguez, G., A. Blanco, N. Corral, and A. Colubi. 2007. Least squares estimation of linear regression models for convex compact random sets. *Advances in Data Analysis and Classification* 1:67–81.

Han, A., Y. Hong, K. Lai, and S. Wang. 2008. Interval time series analysis with an application to the sterling-dollar exchange rate. *Journal of Systems Science and Complexity* 21(4):550–565.

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: Springer. pp. 185–192.

Kulpa, Z. 2006. A diagrammatic approach to investigate interval relations. *Journal of Visual Languages and Computing* 17(5):466–502.

Lima Neto, E. A., and F. d. A. T. de Carvalho. 2008. Centre and range method for fitting a linear regression model to symbolic interval data. *Computational Statistics and Data Analysis* 52:1500–1515.

Lima Neto, E. d. A., and F. d. A. T. de Carvalho. 2010. Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics & Data Analysis* 54(2):333–347.

Maia, A. L. S., F. d. A. de Carvalho, and T. B. Ludermir. 2008. Forecasting models for interval-valued time series. *Neurocomputing* 71(16–18):3344–3352.

Manski, C., and E. Tamer. 2002. Inference on regressions with interval data on a regressor or outcome. *Econometrica* 70(2):519–546.

Moore, R. E. 1966. *Interval Analysis*. Englewood Cliffs, NJ: Prentice Hall.

Moore, R. E., R. B. Kearfott, and M. J. Cloud (eds.) 2009. *Introduction to Interval Analysis*. Philadelphia, PA: SIAM Press.

Parkinson, M. 1980. The extreme value method for estimating the variance of the rate of return. *The Journal of Business* 53(1):61.

Rogers, L., and S. Satchell. 1991. Estimation variance from high, low, and closing prices. *Annals of Applied Probability* 1(4):504–512.

Verde, R., and A. Irpino. 2007. Dynamic clustering of histogram data: using the right metric. In *Selected Contributions in Data Analysis and Classification*. Berlin: Springer. pp. 123–134.

Yakowitz, S. 1987. Nearest-neighbour methods for time series analysis. *Journal of Time Series Analysis* 8(2):235–247.

Yang, D., and Q. Zhang. 2000. Drift independent volatility estimation based on high, low, open, and close prices. *Journal of Business* 73(3):477–492.

Zellner, A., and J. Tobias. 2000. A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting* 19:457–469.