

Information-Theoretic Approach for Forecasting Interval-Valued SP500 Daily Returns

T.S. Tuang Buansing¹

Fannie Mae

Amos Golan²

Department of Economics and Info-Metrics Institute, American University, and Santa Fe Institute

Aman Ullah³

Department of Economics, University of California at Riverside

Abstract

We develop an iterative and efficient information-theoretic estimator for forecasting interval-valued data. We use our estimator to forecast up to five days ahead the SP500 returns using moving windows. Our forecasts are based on 13 years of data. We show that our estimator is superior to its competitors under all the common criteria used to evaluate forecasts of interval data. Our approach is different than other methods used to forecast interval data in two major ways. First, rather than applying the more traditional methods that only use certain moments of the intervals in the estimation process, our estimator uses the complete sample information. Second, our method simultaneously selects the model (or models) and infers the models' parameters. It is an iterative approach that imposes minimal structure and statistical assumptions.

Keywords: Entropy, Information, Information Theoretic Methods, Interval-Valued Data, Iterative Process, Model Selection, SP500 Index

¹The views and conclusions expressed in this paper are those of the authors and not their employers', as applicable. Mr. T.S. Tuang Buansing's contribution to this paper was not developed while at Fannie Mae.

²Corresponding author: agolan@american.edu

³The author is associated with Info-Metrics institute as a member of advisory board.

1. Introduction

In this paper we present a new information-theoretic approach for forecasting interval-valued time series data. Rather than observing a single value for each data point, we observe an interval where the observed values are the bounds of these intervals. Such data may be a result of simple aggregation or emerging from some natural process. Examples include intraday stock prices and returns, which are routinely aggregated to obtain daily intervals within the upper and lower bounds. Similarly, daily closing prices of stocks can be aggregated to form an interval with bounds of the lowest and highest weekly prices. At times, however, the observed intervals may be simply due to the intrinsic nature of the observable information. For example, due to the difficulty in precise measurement of an individual's blood pressure, it must be measured as an interval of systolic and diastolic pressures. In other cases, the interval nature of the data emerges from necessary transformations of the original detailed data. This is commonly observed in risk analysis, high-frequency financial data, weekly oil prices, and weather patterns.

In the above cases, the observed values are the bounds of the interval. There are also cases where more information about the within-interval distribution is observed or known. Examples include data on the minute by minute temperature within one day, or data coming from the survey of professional forecasters that consist of both point forecasts and a more refined frequency, or even a histogram of frequencies. In that case, much more information is available. In this paper we concentrate on the case where the observed information is in terms of the bounds of the interval. Nothing else is observed. Our objective is to forecast these intervals (or the bounds of the intervals) with minimal distributional and structural assumptions.

Unlike the more traditional point-data, interval data have some additional features. One such feature is that the within interval distribution of the different observations need not be the same. When the observed information is only about the bounds of the interval, these distributions are unobserved. Therefore, in most estimation and forecasting models these distributions are assumed to be uniform. We avoid that assumption. Another

feature is that the relationship among these intervals (say independent and dependent, or lag interval and current time interval) may not be unique. For example, the bottom part of the X interval may be associated with the bottom part of the Y interval in one way, while the middle and upper parts of X may be associated with the middle and upper parts of Y in a different way (multiple models). Consequently, methods developed for estimation and inference of classical (point) data are often not directly applicable for estimation and inference of interval data unless strong assumptions are imposed on the estimation model.

In this paper we develop an estimation method for interval-valued data, and apply it to forecast the daily returns (up to five days into the future) of the SP500 over 13 years of data. We concentrate here on the case where the observed information is in terms of a minimum and maximum values of an interval. No other information is observed. Our method is different than other methods in two fundamental ways. First, it is based on minimal structural assumptions. For example, we do not assume that every point in the interval emerged from the same underlying process; there may be multiple models behind the process. Second, our approach uses minimal statistical assumptions. It is a semiparametric, information-theoretic approach. It is statistically and computationally efficient.

Using interval-valued data within economic, econometric and finance is a relatively new phenomenon. Most of the interval-data studies, however, focus on cross-sectional data. The development of estimators for studying Interval Time Series (ITS) data is still in its infancy. Examples include Xu et al. (2008), Han et al. (2008) and Maia et al. (2008) who extended the traditional time series analysis to interval data. These studies are extensions of Interval Least-Squares (ILS), autoregressive (AR), autoregressive integrated moving average (ARIMA) models, artificial neural network (ANN), and a hybrid methodology combining both ARIMA and ANN approaches. Each one of these approaches fits two independent models of the interval data, such as the mid-point and range of the interval. These methods were used to study the stock market indices and

the sterling-dollar exchange rate. Han et al. (2011) used their estimator for forecasting interval-valued crude oil prices with autoregressive conditional interval (ACI) models that are based on a minimum D_K -distance method (or K -th order interval distance) developed in Han et al. (2012). Depending on the kernel chosen, their D_K -distance estimator utilizes both center and range information of the interval sample. Arroyo et al. (2010), González-Rivera and Arroyo (2012), and Arroyo et al. (2011) developed forecasting methods for interval and histogram time series, which are adapted from classical algorithms such as smoothing filters and non-parametric k nearest neighbors (k-NN) methods. Another approach that builds on classical time series methods includes the work of He et al. (2010). They use a Vector Error Correction Model (VECM) with cointegrating vector constrained to (1,-1) for studying the daily highs and lows of West Texas intermediate crude oil prices. A new and different approach, proposed by Lin and González-Rivera (2016), builds on the concept of order statistic. It is based on the assumptions that the densities of the underlying stochastic process, and the number of draws, are conditional normal and negative binomial, respectively. It also assumes that the same stochastic process generates both bounds of the intervals. They applied their model to study the behavior of beef prices and the number of trades. These papers demonstrate that the different forecasting techniques used can be quite successful with financial, and other, data. We show here that our approach provides improved forecasts under all criteria

In this paper we develop an iterative, Information-Theoretic (IT) method for inference of interval-valued time series data to forecast the daily interval of the SP500 index returns. Unlike all other methods, our estimation approach (i) uses the entire sample information (rather than just the sample's moments), (ii) imposes minimal assumptions on the underlying statistical process, and (iii) allows for multiple associations (models) among the independent and dependent interval data. Our method simultaneously selects the model (or models) and infers the models' parameters. For example, in a two-variable interval regression, we allow for multiple models (relationships) between the right and left-hand-side intervals. In that case, the relationship between the lower quartile of the

independent interval may be associated with the lower quartile of the dependent interval (variable) in one way, and with the middle/upper quartile in another way.

Our inferential IT method has its roots at the intersection of information theory and statistical inference. It is based on the premise that the basic estimation problem is inherently under-determined, unless we are willing to impose a certain structure, or certain statistical assumptions, that cannot be verified. To solve this underdetermined, inverse problem, we convert the problem to a constrained optimization one where we use Shannon’s entropy as the criterion to choose the solution out of the many possible solutions that satisfy the observed interval information. See Golan (2018) for the foundations of this framework, more background and numerous applications.

The rest of the paper is structured as follows. In Section 2 we provide a brief discussion on interval estimation and describe in more details some of the competing approaches. In Section 3 we discuss our approach. We start by motivating it and summarizing the different mechanisms that may produce the observed interval data. We then describe our iterative approach and our IT method. In Section 4 we provide the statistics used to evaluate the forecasts. Section 5 presents the empirical results using the SP500 returns. We conclude in Section 6. Additional summary statistics, tables, plots, our data and a computer code are provided in the supplementary materials.

2. Interval Estimation: The Basic Problem

2.1. *Variability vs. Uncertainty*

Suppose a set of known variables \mathbf{X} is associated with (or may even have a causal effect on) the dependent variable \mathbf{y} , and both the \mathbf{y} ’s and the \mathbf{X} ’s are observed as intervals. Specifically, the only values observed are the minimum and maximum of the intervals. No other information on the distribution is known or assumed. Let (Y^L, Y^U) denote a pair of real-valued random variables such that $Y^L < Y^U$ with probability one, where L and U stand for the lower and upper bounds respectively. Then $[Y^L, Y^U]$ is an interval-valued random variable. Suppose a random sample on $[Y^L, Y^U]$ is available and is denoted

as $\{[Y^L, Y^U]_i\}_{i=1}^n$. Such data are referred to as interval data which may represent uncertainty or variability of the observed information. In the former case (uncertainty), there is a random (latent) variable of interest denoted as Y^* satisfying $Y^* \in [Y^L, Y^U]$ with probability one, but the researcher does not observe Y^* . For further discussion see Manski and Tamer (2002), Manski (2003), Beresteanu and Molinari (2008), and Beresteanu et al. (2011). In the latter case (variability), the variable of interest is the interval-valued random variable $[Y^L, Y^U]$ and the parameters of interest are functions of $[Y^L, Y^U]$.

In this paper, we focus on modeling variability and our main objective is to predict the lower and upper end points Y^L, Y^U , or more accurately, the complete interval between Y^L and Y^U . The commonly used approaches assume the existence of a single underlying model that captures the relationship among variables in the independent intervals and the dependent one (see previous references). We relax that assumption.

2.2. Discussion of Other Methods

In this section, we summarize some of the currently used methods for predicting end points of the interval-valued random variable. See Xu (2010) and Ahn et al. (2012) for a more detailed discussion.

Using the same notations as Ahn et al. (2012), let X_1, \dots, X_K be K explanatory variables and Y be the response variable. Assume that $X_{ik} = [X_{ik}^L, X_{ik}^U]$ with $X_{ik}^L \leq X_{ik}^U$ and $Y_i = [Y_i^L, Y_i^U]$ with $Y_i^L \leq Y_i^U$, for $i = 1, \dots, N$ and $k = 1, \dots, K$. Consider the following linear regression model:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{1}$$

where $\mathbf{Y} = (Y_1, \dots, Y_N)'$, $\mathbf{X} = (X_1, \dots, X_N)'$, $X_i = (1, X_{i1}, \dots, X_{iK})'$ for $i = 1, \dots, N$, $\beta = (\beta_0, \beta_1, \dots, \beta_K)'$, $\epsilon = (\epsilon_1, \dots, \epsilon_N)'$, $\epsilon_i \sim N(0, \sigma^2)$. Model (1) assumes that the effect of every covariate variable in the observed interval on every variable in $[Y^L, Y^U]$ is the same facilitating the estimation of β via classical linear regressions for real-valued random

variables. Now, we briefly review the major commonly used approaches.

Billard and Diday (2000) proposed the center method (CM), which is one of the first estimators developed for analyzing interval-valued data. The CM fits a linear regression line to the center points of the intervals. Let X_1^c, \dots, X_K^c be the center points of the intervals of explanatory variables X_1, \dots, X_K and Y^c be the center point of a response variable Y . The CM essentially transforms the interval linear regression model of (1) into a center-points model

$$\mathbf{Y}^c = \mathbf{X}^c \beta^c + \epsilon^c \quad (2)$$

where $\mathbf{Y}^c = (Y_1^c, \dots, Y_N^c)'$, $\mathbf{X}^c = (X_1^c, \dots, X_N^c)'$, $X_i^c = (1, X_{i1}^c, \dots, X_{iK}^c)'$ for $i = 1, \dots, N$, $\beta^c = (\beta_0^c, \beta_1^c, \dots, \beta_K^c)'$, and $\epsilon^c = (\epsilon_1^c, \dots, \epsilon_N^c)'$.

The estimator $\hat{\beta}^c$ is the usual least squares

$$\min_{\beta^c} \sum_i \hat{\epsilon}_i^c{}^2 = \sum_i (Y_i^c - \hat{\beta}^c{}' X_i^c)^2 \quad (3)$$

so $\hat{\beta}^c = (\mathbf{X}^c{}' \mathbf{X}^c)^{-1} \mathbf{X}^c{}' \mathbf{Y}^c$ and the standard statistical properties are readily adopted. Prediction of $\hat{Y} = [\hat{Y}^L, \hat{Y}^U]$ is given as:

$$\hat{Y}^q = x_0^q \hat{\beta}^c, \quad q = L, U \quad (4)$$

for a new observation (x_0^L, x_0^U) . Xu (2010) and Xu and Billard (2012) point out that a lower bound predicted response variable can be higher than an upper bound, and suggest the following modified prediction:

$$\hat{Y}^L = \min(x_0^L \hat{\beta}^c, x_0^U \hat{\beta}^c) \quad ; \quad \hat{Y}^U = \max(x_0^L \hat{\beta}^c, x_0^U \hat{\beta}^c). \quad (5)$$

Although the CM uses the ranges of predictors for predicting of the lower and upper bounds, these bounds are ignored in estimating the parameters. That is, the variations within observations are not utilized.

In an attempt to capture the internal variations as well as variations across observations, Lima Neto et al. (2004, 2008) propose the center and range method (CRM) that transforms the interval-valued data into two quantities: center and range. Then, they regress the center points and ranges separately. The CRM keeps the same model as in (2) for the center points regression. The range regression model is based on $X_{ik}^r = (X_{ik}^U - X_{ik}^L)$ and $Y_i^r = (Y_i^U - Y_i^L)$. Specifically, let X_1^r, \dots, X_K^r be the K ranges of the intervals of X_1, \dots, X_K and Y^r be the range of Y . The range regression is then

$$\mathbf{Y}^r = \mathbf{X}^r \beta^r + \epsilon^r \quad (6)$$

where $\mathbf{Y}^r = (Y_1^r, \dots, Y_N^r)'$, $\mathbf{X}^r = (X_1^r, \dots, X_N^r)'$, $X_i^r = (1, X_{i1}^r, \dots, X_{iK}^r)'$ for $i = 1, \dots, N$, $\beta^r = (\beta_0^r, \beta_1^r, \dots, \beta_K^r)'$, and $\epsilon^r = (\epsilon_1^r, \dots, \epsilon_N^r)'$. Both $\hat{\beta}^c$ and $\hat{\beta}^r$ are estimated simultaneously by minimizing the following objective function:

$$\min_{\hat{\beta}^c, \hat{\beta}^r} \sum_i (\hat{\epsilon}_i^c{}^2 + \hat{\epsilon}_i^r{}^2). \quad (7)$$

They essentially performed two separate minimizations, one for the centers and another for the ranges, assuming that mid-points and ranges are independent. However, Ahn et al. (2012) asserts that such assumption may not be true in general. Although CRM captures the interval variations through the ranges, Ahn et al. (2012) also point out that it is not clear how these variations are transferred to the estimated coefficients. Prediction of $\hat{Y} = [\hat{Y}_L, \hat{Y}_U]$ is given as

$$\hat{Y}^L = \hat{Y}^c - \frac{\hat{Y}^r}{2} \quad ; \quad \hat{Y}^U = \hat{Y}^c + \frac{\hat{Y}^r}{2} \quad (8)$$

where \hat{Y}^c and \hat{Y}^r are predicted values from (2) and (6).

Similar to CRM, Billard and Diday (2007) proposed a bivariate center and range method (BCRM) which simultaneously utilizes both the centers and the ranges as predictors in the models. So far all the approaches summarized build directly on the least

squares. Note that when all the explanatory and response intervals have the same centers and ranges, the previous methods will not work.

In contrast to the previous approaches, Xu (2010) and Xu and Billard (2012) proposed a symbolic covariance method (SCM). Their model builds on the symbolic covariance function of Billard (2007, 2008) and the symbolic sample mean of Bertrand and Goupil (2000). Ahn et al. (2012), on the other hand, proposed a Monte-Carlo method that generates a large number of samples by randomly selecting, uniformly, a single-valued data point for each observed interval, fit a classical linear regression model on each single-valued sample, and then calculates the mean estimated coefficients over the fitted models. Then, they use the means of the estimated coefficients to predict the response variable.

Among all of the above mentioned existing methods, we contrast our forecasts only with the two commonly used methods for time series data: Center and Range Method (CRM) and Constrained Center and Range Method (CCRM). Since the analysis of interval data is very much in line with the study of volatility measures, we also compare our approach with forecasts based on the classical conditional mean and variance model of AR(1) - GARCH(1,1). See Bollerslev (1986).

For a time series interval variable $R_t = [R_t^L, R_t^U]$, we define the time series of the centers and ranges as $R_t^C = (R_t^L + R_t^U)/2$ and $R_t^R = R_t^U - R_t^L$, respectively. The CRM method estimates the following system:

$$\begin{aligned} R_t^C &= \beta_0^C + \beta_1^C R_{t-1}^C + \dots + \beta_p^C R_{t-p}^C + \epsilon_t^C \\ R_t^R &= \beta_0^R + \beta_1^R R_{t-1}^R + \dots + \beta_p^R R_{t-p}^R + \epsilon_t^R. \end{aligned} \tag{9}$$

As is shown in (9), the CRM estimates two separate regressions; one for each transformed quantities of the intervals. The CCRM imposes the additional restriction $\beta_j^R \geq 0, j = 0, \dots, p$. That ensures that $\hat{R}_t^R \geq 0$, so $\hat{R}_t^L \leq \hat{R}_t^U$.

The conditional mean and variance model applied to the time series data of the centers

is basically an AR(1) - GARCH(1,1) of the centers:

$$\begin{aligned} R_t^C &= \varphi_1 R_{t-1}^C + \sigma_t \zeta_t \\ \sigma_t^2 &= \omega + \gamma \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned} \tag{10}$$

where ζ_t follows a standard normal (GARCH-Normal) or Student-t (GARCH-T) density with v degrees of freedom, and φ_1 , σ_t , ω , γ and β are parameters to be estimated. In this model, as was shown in González-Rivera and Lin (2013), $(1-\alpha)$ probability intervals can be constructed based on ζ_t . Therefore, the predicted intervals are: $[\hat{R}_t^L, \hat{R}_t^U]_\alpha = [\hat{R}_t^C - z_{\frac{\alpha}{2}} \hat{\sigma}_t, \hat{R}_t^C + z_{\frac{\alpha}{2}} \hat{\sigma}_t]$ for GARCH-Normal and $[\hat{R}_t^L, \hat{R}_t^U]_\alpha = [\hat{R}_t^C - t_{\hat{v}, \frac{\alpha}{2}} \hat{\sigma}_t, \hat{R}_t^C + t_{\hat{v}, \frac{\alpha}{2}} \hat{\sigma}_t]$ for GARCH-T. Since the original data $[R_t^L, R_t^U]$ are the observed extreme values of the process at t , one can stretch the estimated interval $[\hat{R}_t^L, \hat{R}_t^U]_\alpha$ to cover as much as 99% or 99.5% probability, so that \hat{R}_t^L and \hat{R}_t^U are far away into the tails of the distribution.

3. An Information-Theoretic Model Selection and Estimation Method

In this section, we propose an information based iterative procedure for modeling interval data. We then explore its application for predicting the interval of the dependent variable. Our proposed iterative framework has several advantages. First, it is simple and easy to implement. Second, it relaxes distributional assumptions on the within-interval behavior. Third, rather than forcing a unique relationship (model) between the (independent and dependent) intervals, we allow for multiple models, and provide a statistic to choose the optimal model (or models). Fourth, we do not impose distributional assumptions. We now discuss our iterative framework. Then, we summarize our IT estimator.

3.1. The Iterative Framework

Our interval model is formulated in its most general terms where both the dependent and independent variables are intervals. This includes the simpler case where the exogenous variables (or just some of the variables) are point-valued. In that case, the interval

of that exogenous variable reduces to a point, and therefore, the number of possible models decreases.

To introduce the idea, consider first the simplest case. Suppose that there is only one explanatory variable $\mathbf{X}_i = [X_i^L, X_i^U]$ and one response variable $\mathbf{Y}_i = [Y_i^L, Y_i^U]$ for $i = 1, \dots, N$ observations. We divide each observation of \mathbf{X}_i and \mathbf{Y}_i into some finite and discrete number of equal size sub-intervals (or points) indexed by $m = 1, 2, \dots, M$, to get $\mathbf{X}_i^{\mathbf{m}} = [X_i^1, X_i^2, \dots, X_i^M]$ and $\mathbf{Y}_i^{\mathbf{m}} = [Y_i^1, Y_i^2, \dots, Y_i^M]$, where each \mathbf{X}^m and \mathbf{Y}^m are N vectors of point-valued data generated from each sub-interval. Simply stated, M represents the number of sub-intervals within each \mathbf{X} and \mathbf{Y} , or we can view it for now as M observed points within each observation $(\mathbf{X}_i, \mathbf{Y}_i)$. For example, for observation $i = 1$ with one interval-valued regressor \mathbf{X}_i , suppose we observed $\mathbf{X}_1 = [2, 10]$ and $\mathbf{Y}_1 = [6, 22]$. We can divide them into $M = 5$ equally spread points so that we have $\mathbf{X}_1 \equiv [X_1^1, X_1^2, X_1^3, X_1^4, X_1^5] = [2, 4, 6, 8, 10]$ and $\mathbf{Y}_1 \equiv [Y_1^1, Y_1^2, Y_1^3, Y_1^4, Y_1^5] = [6, 10, 14, 18, 22]$. Alternatively, we can divide the original intervals into $M - 1$ sub-intervals so that $\mathbf{X}_1 \equiv [[2, 4], [4, 6], [6, 8], [8, 10]]$ and $\mathbf{Y}_1 \equiv [[6, 10], [10, 14], [14, 18], [18, 22]]$, and then using only endpoints of the sub-intervals would yields $\mathbf{X}_1 \equiv [2, 4, 6, 8, 10]$ and $\mathbf{Y}_1 \equiv [6, 10, 14, 18, 22]$. Both approaches would correspond to $Y_i^m = 2 + 2X_i^m$. In this example we assume that each sub-interval m of \mathbf{X} causes linearly the m th sub-interval of \mathbf{Y} and the impacts are the same across different points within the intervals.

But now, to be more realistic, assume that we know \mathbf{X} causes \mathbf{Y} linearly but we do not know the correct model or models. We do not know if each point within the interval of \mathbf{X} affects \mathbf{Y} in the same way. Maybe, for example, points on the lower part of \mathbf{X} affect \mathbf{Y} in a different way than points around the mean or the upper portion of the interval. Therefore, each sub-interval m may have a different coefficient resulting in multiple models. More precisely, we do not know which combination of intervals (or points) $\mathbf{m}(= m_1, m_2, \dots, m_K)$ of the K explanatory variables \mathbf{X} correspond to which interval (or point) m of Y^m . Given this complex setting, we want to infer the relationship between each one of the \mathbf{X} sub-intervals and the \mathbf{Y} sub-intervals. We allow for multiple a-priori

models. To do so with minimal assumptions on the underlying distribution, we apply our IT estimator for all possible combination of sub-intervals within the interval-valued random variables \mathbf{X} and \mathbf{Y} . Figure 1 presents our iterative approach for one and two independent variables.

[Insert Figure 1 here]

Consider, for example, $M = 5$ and $K = 1$ as shown in Figure 1. Then, using $M \times M^K$ combinations of pairs such as $(\mathbf{X}_1^1, \mathbf{Y}^1), (\mathbf{X}_1^1, \mathbf{Y}^2), \dots, (\mathbf{X}_1^1, \mathbf{Y}^5), (\mathbf{X}_1^2, \mathbf{Y}^1), (\mathbf{X}_1^2, \mathbf{Y}^2), \dots, (\mathbf{X}_1^5, \mathbf{Y}^5)$, where $\mathbf{X}_1^{m_1} \equiv X_{i1}^{m_1}$ and $\mathbf{Y}^m \equiv Y_i^m$ for $i = 1, \dots, N$ and $m, m_1 = 1, \dots, M$, we obtain 25 sets of estimated coefficients via a linear estimator, while $K = 2$ produces 125 sets of estimated coefficients. Since each pair of $(\mathbf{X}_1^{m_1}, \mathbf{Y}^m)$ or $(\mathbf{X}_1^{m_1}, \mathbf{X}_2^{m_2}, \mathbf{Y}^m)$ becomes our classical points-based sample, the estimated coefficients for each iteration are point identified.

Using these estimated regression coefficients, together with the empirical distribution of the estimator's objective values (discussed below) for each one of the models, we can determine the best model (or models). It is the model (models) with the highest value (values). Thus, we identify the best models (or model), and simultaneously obtain the inferred parameters of these models.

Finally, we provide two comments on the choice of M . First, theoretically, the larger the M , the more possible models one can capture. Assuming, however, that the only observed information is the lower and upper bounds of the interval, and that the within-interval is not too different than uniform, the estimates are not sensitive to that choice as long as M is 'large' enough. For all practical purposes, based on a large number sampling experiments, we find $M = 5$ to be large enough. These experiments are not presented here. Second, from computational point of view, the larger the M , the larger the number of iterations and possible models. Empirically, we find $M = 5$ to be sufficient. Our qualitative results and forecasts do not change for larger values of M . To sum, the

basic choice of M depends on the trade-off between the amount of variations obtained by increasing M and the increase in iterations required. If we expect a non-uniform distribution within the interval, we should increase M accordingly. In future work we plan to investigate M and related properties. In this paper, we use $M = 5$.

3.2. An IT Estimator: A Concise Summary

Before discussing the IT estimator we employ here, we note that any other estimator can be used within our iterative (model selection) framework. The choice of the estimator is independent of the choice of our model selection, iterative, framework. We choose to use a special IT estimator due to its superior performance and because it is based on minimal assumptions and structure. The IT approach we employ here is a member of the IT family of estimators, which is encompassed within the info-metrics framework (Golan, 2018). The estimator itself is based on the work of Golan et al. (1996) and the recent developments in Golan (2018). We use this method to estimate each one of the linear models (regressions) in our iterative framework. We now formulate that method for estimating a single regression.

Consider a linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where \mathbf{y} is a N -dimensional vector of observed random variable, \mathbf{X} is a $N \times K$ matrix of exogenous variables; $\boldsymbol{\beta}$ is K -dimensional vector of unknown coefficients; and $\boldsymbol{\epsilon}$ is a N -dimensional vector of unobserved error terms with mean zero. Rather than search for the point estimates $\boldsymbol{\beta}$, each β_k is viewed as the mean value of some well defined random variable \mathbf{z} . The unobserved error vector $\boldsymbol{\epsilon}$ is also viewed as another set of unknowns where each ϵ_i is constructed as the mean value of some random variable \mathbf{v} . Our objective here is to estimate the unknown $\boldsymbol{\beta}$ with minimal distributional assumptions. Under the info-metrics framework, however, we estimate simultaneously the full distribution of each β_k and each ϵ_i (within their support spaces) with minimal distributional assumptions. For a comprehensive discussion of info-metrics see Golan (2018). For an overview of the class of IT methods see also Judge and Mittelhammer (2011).

Without loss of generality, let each (k) element of β be bounded below by \underline{z}_k and above by \bar{z}_k :

$$B = \{\beta \in \Re^K | \beta_k \in (\underline{z}_k, \bar{z}_k), k = 1, \dots, K\}. \quad (11)$$

Let \mathbf{z}_k be an S -dimensional vector, so, $\mathbf{z}_k \equiv (\underline{z}_k, \bar{z}_k)' = (z_{k1}, \dots, z_{kS})'$ for all $k = 1, \dots, K$, and Z is a $K \times S$ matrix consisting of the individual S -dimensional vectors \mathbf{z}_k and the elements z_{km} . Let \mathbf{p}_k be an S -dimensional normalized probability distribution defined on the set \mathbf{z}_k such that

$$\beta_k = \sum_s p_{ks} z_{ks} \equiv E_{p_k}[\mathbf{z}_k] \text{ or } \beta = E_P[Z] \quad (12)$$

and E is the expectation operator. In this formulation, the observed data, \mathbf{y} , are viewed as the mean process Z with a probability distribution P that is defined on \mathbf{z}_k and is conditional on X . (Note that β can be constructed as the median or any other quantity of interest, within this framework.)

Similarly, assume that $\epsilon \in V$ where V is a convex set that is symmetric around zero. As done with the β 's above, each error term is formulated as

$$\epsilon_i = \sum_j w_{ij} v_j \equiv E_{w_i}[\mathbf{v}] \text{ or } \epsilon = E_W[V]. \quad (13)$$

That is, the observed errors are viewed as elements taken as random draws from a certain distribution with probability weights $\{w_{ij}\}$. The dimension of the matrices V and W is $N \times J$. For simplicity, the above random variables are constructed as discrete and bounded. For unbounded and continuous supports see, for example, Golan and Gzyl (2002, 2012).

Having reformulated β and ϵ , the linear model is specified as

$$y_i = \sum_{k=1}^K \sum_{s=1}^S z_{ks} p_{ks} x_{ik} + \sum_j v_j w_{ij}, \text{ or } \mathbf{y} = X E_P[Z] + E_W[V]. \quad (14)$$

The IT estimation method is

$$IT = \begin{cases} \hat{\mathbf{p}} = \underset{\{\mathbf{p}, \mathbf{w}\}}{\operatorname{argmax}} \left\{ \{H(\mathbf{p}) + H(\mathbf{w})\} \equiv -\sum_{k,s} p_{ks} \log(p_{ks}) - \sum_{i,j} w_{ij} \log(w_{ij}) \right\} \\ \text{s.t.} \\ y_i = \sum_{k,s} z_{ks} p_{ks} x_{ik} + \sum_j v_j w_{ij}, \\ \sum_s p_{ks} = 1 \quad , \quad \sum_j w_{ij} = 1 \end{cases} \quad (15)$$

where $H(\mathbf{p}) = -\sum_{k,s} p_{ks} \log(p_{ks})$ and $H(\mathbf{w}) = -\sum_{i,j} w_{ij} \log(w_{ij})$ are the entropies of \mathbf{p} and \mathbf{w} , respectively.

Constructing the Lagrangian and solving yields the estimated probabilities for the vector $\boldsymbol{\beta}$:

$$\hat{p}_{ks} = \frac{\exp(-z_{ks} \sum_i \hat{\lambda}_i x_{ik})}{\sum_s \exp(-z_{ks} \sum_i \hat{\lambda}_i x_{ik})} \equiv \frac{\exp(-z_{ks} \sum_i \hat{\lambda}_i x_{ik})}{\Omega_k(\hat{\lambda}_i)}, \quad (16)$$

the estimated probabilities for the ϵ :

$$\hat{w}_{ij} = \frac{\exp(-\hat{\lambda}_i v_j)}{\sum_j \exp(-\hat{\lambda}_i v_j)} \equiv \frac{\exp(-\hat{\lambda}_i v_j)}{\Psi_i(\hat{\lambda}_i)} \quad (17)$$

and where $\hat{\lambda}$ is the vector of estimated Lagrange multipliers associated with the data constraints y_i . The estimated values of $\boldsymbol{\beta}$ and ϵ are

$$\hat{\beta}_k = \sum_s z_{ks} \hat{p}_{ks} \quad (18)$$

$$\hat{\epsilon}_i = \sum_j v_j \hat{w}_{ij}. \quad (19)$$

$\Omega_k(\hat{\lambda}_i) = \sum_s \exp(-z_{ks} \sum_i \hat{\lambda}_i x_{ik})$ and $\Psi_i(\hat{\lambda}_i) = \sum_j \exp(-\hat{\lambda}_i v_j)$ are the partition (normalization) functions of (16) and (17) respectively.

As with other methods, it is possible to transform the constrained optimization (primal) model to a dual, concentrated model which is a function of the Lagrange multipliers λ (Golan et al., 1996; Golan, 2018):

$$\begin{aligned} \max_{\{P \in \mathcal{P}, W \in \mathcal{W}\}} H(P, W) &= \min_{\lambda \in D} \left\{ \sum_i y_i \lambda_i + \sum_k \log \Omega_k(\lambda_i) + \sum_i \log \Psi_i(\lambda_i) \right\} \\ &= \min_{\lambda \in D} \left\{ \sum_i y_i \lambda_i + \sum_k \log \left[\sum_s \exp(-z_{ks} \sum_i \lambda_i x_{ik}) \right] + \sum_i \log \left[\sum_j \exp(-\lambda_i v_j) \right] \right\}. \end{aligned} \quad (20)$$

The concentrated model is solved by minimizing with respect to λ . The optimal set of λ 's is then used to obtain the P 's via Equation (16), with which the set of β 's is inferred via Equation (18). The Hessian matrix is negative definite for $P, W \geq 0$ and thus satisfies the sufficient condition for a unique global minimum. This IT estimator minimizes the joint entropy distance between the data and the state of complete uncertainty (the uniform distribution). It is a dual-loss function that assigns equal weights to prediction and precision. Equivalently, it can be viewed as a shrinkage estimator that shrinks the data to the priors (uniform distributions) and towards the center of their supports (the midpoint between \underline{z}_k and \bar{z}_k for each β_k , and zero for each ϵ_i).

The estimated probabilities provide the full distribution (within the boundaries of the supports) of each one of the parameters of interest (β and ϵ). The β 's are direct functions of the Lagrange multipliers (λ). These multipliers reflect the marginal information of each observation. Like the empirical (or generalized) likelihood methods they capture the natural weight of each observation and convey that information in the estimated exponential distributions \hat{p}_{ks} and \hat{w}_{ij} . (For the basics of the empirical and generalized empirical likelihood methods, see for example Smith (2005), Owen (2001), Judge and Mittelhammer (2011).) For more recent development, basic properties, inference and diagnostics see Golan (2018).

4. Evaluating the Forecast Accuracy

To evaluate forecasts among the different methods, we use multi-step out-of-sample forecasts via a rolling (moving) window with RW_E as the dimension of estimation window and RW_T as the number of future periods we forecast. Thus, the entire sample size is the sum of RW_E and RW_T . We are interested in forecasting up to a week ahead ($h = 1, 2, 3, 4, 5$). After we obtain h -step out-of-sample forecasts, we move to the next window and repeat the h -step forecasts. The estimation window is rolled until we reach the last window.

Let $\hat{Y}_{t+h|t}$ denote the h -step ahead forecast of Y_{t+h} based on the information available at time t . For each one of the h -step ahead forecasts over RW_T periods, we use the following (common) forecast evaluation measures.

$$\begin{aligned}
RMSE.q &= \sqrt{\frac{1}{RW_T} \sum_{j=RW_E+1}^{RW_E+RW_T} (Y_{j+h}^q - \hat{Y}_{j+h|j}^q)^2} & MAE.q &= \frac{1}{RW_T} \sum_{j=RW_E+1}^{RW_E+RW_T} |(Y_{j+h}^q - \hat{Y}_{j+h|j}^q)| \\
numCR &= \frac{1}{RW_T} \sum_j w(Y_{j+h} \cap \hat{Y}_{j+h|j}) / w(Y_{j+h}) & numER &= \frac{1}{RW_T} \sum_j w(Y_{j+h} \cap \hat{Y}_{j+h|j}) / w(\hat{Y}_{j+h|j}) \\
numIn &= \frac{1}{RW_T} \sum_j 1(\hat{Y}_{j+h}^L < Y_{j+h}^C \ \& \ Y_{j+h}^C < \hat{Y}_{j+h}^U) & numCov &= \frac{1}{RW_T} \sum_j 1(\hat{Y}_{j+h}^L < Y_{j+h}^L \ \& \ Y_{j+h}^U < \hat{Y}_{j+h}^U) \\
numHCov &= \frac{1}{RW_T} \sum_j 1(Y_{j+h}^U < \hat{Y}_{j+h}^U) & numLCov &= \frac{1}{RW_T} \sum_j 1(\hat{Y}_{j+h}^L < Y_{j+h}^L) \\
numOverLap &= \frac{1}{RW_T} \sum_j \frac{|\min(\hat{Y}_{j+h}^U, Y_{j+h}^U) - \max(\hat{Y}_{j+h}^L, Y_{j+h}^L)|}{|\max(\hat{Y}_{j+h}^U, Y_{j+h}^U) - \min(\hat{Y}_{j+h}^L, Y_{j+h}^L)|}
\end{aligned}$$

where $q = L, U$; $Y_{j+h} \cap \hat{Y}_{j+h|j}$ is the intersection of actual and fitted intervals; and $w(\cdot)$ is the width of the interval. Less technically, these statistics include the (i) Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE) for the upper and lower bounds, (ii) coverage (numCR) and efficiency rates (numER) of the estimated intervals

(Rodrigues and Salish, 2015), and the (iii) coverage statistics proposed in Buansing et al. (2017). Rodrigues and Salish (2015)’s coverage and efficiency rates (numCR, numER) measure the average proportion of the actual interval covered by the fitted interval and the fitted interval covered by the actual interval, respectively. Although not exactly the same, they are similar in spirit with the overall coverage (numCov) and overlap measure (numOverLap). All the coverage measures are between zero and one, and a higher number suggests a better fit. In addition, they all are implemented here to reflect the recognition that potential trade-off exists between coverage and efficiency or overlap measures. In addition to the above, we also use the Diebold-Mariano test, discussed in Section 5.3.

Last, we note that our test is different than the classic test for evaluating interval forecasts developed by Christoffersen (1998), and then modified by Clements and Taylor (2003) to accommodate for the presence of periodic heteroskedasticity. In these tests the basic idea is to test whether a point forecast falls within some interval. The test builds on an indicator variable for an interval forecast: the indicator equals one if the observed data point is within the forecasted interval. In our work, we are interested in coverage and efficiency of forecasts that are naturally coming from interval data (Rodrigues and Salish, 2015), and therefore are different than the diagnostics developed for forecasting intervals resulting from point time-series data. We are not only interested in overall coverage, but rather in efficiency as well.

5. Forecasting the SP500 Returns Interval

In this section, we discuss the data and our forecasts. We use our iterative, IT method to forecast the SP500 up to five days ahead. We contrast the forecasts of our approach with that of other competing methods. We also contrast it with the well-known conditional mean and variance model of AR(1) - GARCH(1,1) which serves as a benchmark in many forecasting studies.

5.1. The Data

Following González-Rivera and Lin (2013), the interval time series of the daily low/high of SP500 index returns are defined as

$$\begin{aligned} R_t^H &= \frac{P_{high,t} - P_{close,t-1}}{P_{close,t-1}} \times 100 \\ R_t^L &= \frac{P_{low,t} - P_{close,t-1}}{P_{close,t-1}} \times 100 \end{aligned} \tag{21}$$

where $P_{high,t}$ and $P_{low,t}$ are the highest and lowest price during trading day t while $P_{close,t-1}$ is the closing price of the previous day ($t - 1$). Our sample consists of data from January 1st, 2004 to January 1st, 2016. As in González-Rivera and Lin (2013) that used similar data but their end date is April 29, 2011, the sample is split into two periods: stable and unstable periods. The stable period has 754 trading days that goes from January 1st, 2004 to January 1st, 2007 just like González-Rivera and Lin (2013), while the unstable period (2265 days) goes from January 1st, 2007 to January 1st, 2016.

Figure 2 presents the time series of the daily low and high returns, where the returns are defined according to Eq. (21). In the stable period, both low and high returns have an overall low volatilities ($\sigma_l^2 = 0.173$ and $\sigma_u^2 = 0.161$), varying roughly within a range of $[-2\%, 2\%]$. The unstable period exhibits a much higher volatility level of $\sigma_l^2 = 1.06$ and $\sigma_u^2 = 0.84$ respectively for the lower and upper levels, that ranges from -10% to 12% . The volatility in the unstable period is quite high and the correlation between low and high returns is 0.295. In the stable period, however, the correlation is much higher (0.581) while the volatility is lower. A detailed descriptive statistics for both stable and unstable periods are provided in the supplementary material.

Figure 2 also shows that both the low and high index returns exhibit volatility clustering. Furthermore, such data are also characterized by fat tails and nonlinear dependence. For example, the magnitudes of volatility are low during the first one-third of the sample, large and persistence between 2007 - 2011, and relatively low in the last part

of the sample period (2012 - 2016). In addition, both low and high returns are occasionally very large (positive/negative) and the correlation between them are high during the stable period while relatively low during the unstable period. This observation is in line with the literature (González-Rivera and Lin (2013)).

[insert Figure 2 here]

5.2. Estimation

First, we study the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) of the centers of SP500 index returns, R_t^C , for stable and unstable periods. During the stable period, the autocorrelation seems to disappear after the first lag. However, autocorrelation at longer lags are observed for the unstable period. Second, we run an unrestricted AR(p) model using R_t^C and select the optimal lags by minimizing the Bayesian information criterion (BIC). The number of optimal lags is 1 for the stable period and 5 for the unstable period. These optimal number of lags are used in the iterative IT approach and two of the existing methods: CRM and CCRM. Similar trends, properties, and number of lags are chosen when we use the lower and upper bounds of the index return intervals (R_t^L and R_t^U). Given the above, we use $p = 1$ lag for the stable period and $p = 5$ lags for the unstable period in our interval autoregressive (AR) system:

$$\begin{aligned} R_t &= \beta_0 + \beta_1 R_{t-1} + \dots + \beta_p R_{t-p} + \epsilon_t \\ [R_t^L, R_t^U] &= \beta_0 + \beta_1 [R_{t-1}^L, R_{t-1}^U] + \dots + \beta_p [R_{t-p}^L, R_{t-p}^U] + [\epsilon_t^L, \epsilon_t^U]. \end{aligned} \tag{22}$$

In our iterative, IT framework we estimate a large number of models. The chosen (best) model (or models) is the one that maximizes the entropy objective function. Figure 3 demonstrates the unique relationship between the entropy value (of each model) and the estimated parameters of that model. That relationship is shown as a scatter

plot between one of the estimated slope coefficients ($\hat{\beta}_4$) and the objective value, total entropy, for the first estimation window from the unstable period. (Scatter plots for other estimated coefficients are provided in the supplementary materials.) For each estimation window, the regression model uses one lag for the stable period and five lags for the unstable period. As discussed in Section 3, under our IT estimation approach, we maximize the joint entropy of the signal and the noise subject to the observed data and normalization. Keeping in mind that the estimated errors are minimized and simultaneously, the estimated coefficients converge to their true values, when the entropy objective function is maximized, the natural statistic used for identifying the best possible (or “correct”) model is the value of the total entropy (objective function (15), or similarly (20)). Figure 3 presents this relationship. Out of all the possible models, the one with the highest entropy corresponds to the chosen model (dashed line). In this specific example and estimation window, there is no evidence of multiple “correct” models. The same is true for all other windows in the sample we studied. In addition, our comprehensive analysis shows that the fundamental market structure seems to be stable throughout the 13 years sample. These conclusions are based on all estimation windows in both stable and unstable periods which are omitted here for brevity.

Since the main purpose of this paper is forecasting extreme intervals of the daily returns, we use the modified version of intercept parameters. We shift the predicted lower and upper bounds by using the minimum and maximum of the empirical distribution of the intercepts, while keeping the slope parameters chosen by the entropy criterion. Therefore, the predicted lower and upper bounds are:

$$\hat{R}^q = R_0^q \hat{\beta}^{q*}, q = L, U \quad (23)$$

where $\hat{\beta}^{L*} = [\min(\hat{\beta}_0), \hat{\beta}_1^*, \hat{\beta}_2^*, \dots, \hat{\beta}_p^*]$ and $\hat{\beta}^{U*} = [\max(\hat{\beta}_0), \hat{\beta}_1^*, \hat{\beta}_2^*, \dots, \hat{\beta}_p^*]$. In the supplementary materials, we provide figures with the estimated coefficients chosen by our entropy criterion for all rolling windows of both periods. From these, we learn that the

estimated slope parameters are not much different from their averages and seem to be relatively stable over the rolling windows. Here, the estimation window size is 20 days with $M = 5$ and we predict up to five-day ahead for each rolling window. Although results from different window sizes such as 25, 30, or 50 give different quantitative results, the conclusions provided are qualitatively similar to those just described. These results are omitted here.

[insert Figure 3 here]

5.3. Forecasting Results

We concentrate here on multi-period ahead forecast. Further details of the one-period ahead forecasts are provided in the supplementary material. Figures 4 - 6 provide forecast evaluations for all forecast horizons (up to 5-day ahead). In the stable period, the iterative IT approach out-performs all other approaches across the forecast horizons in RMSEL and RMSEH (see Panels in Figure 4). MAEL and MAEH plots show similar results and are omitted here for brevity. As the forecast horizon increases, the RMSEs and MAEs of CRM and CCRM deteriorate while the iterative IT's lower and upper forecast errors remain stable. CRM and CCRM still perform well in numCR across the forecast horizon while iterative IT out-performs them in numER (see Panels in Figure 5) and numOverLap (see Panels in the supplement). As the forecast horizon increases, our iterative IT gets relatively better in numCRER while CRM and CCRM deteriorate a little bit. Overall, our iterative IT out-performs the other methods under the majority of the measures used and its performance remains stable or better for longer forecast horizons. In the unstable period, our iterative IT also out-performs in terms of RMSEL, RMSEH, MAEL and MAEH while competing methods become worse across the forecast horizons (see Panels in Figure 6 for the RMSEs). CRM and CCRM perform well in numCR across the forecast horizon while our iterative IT out-performs them in numER and numOverLap (see Panels in the supplement). As the forecast horizon increases, our

iterative IT is getting better relative to the other methods in numCRER. We do not show this here.

[insert Figures 4 - 6 here]

In term of forecast errors such as RMSEs, MAEs, efficiency rates (numER) and percentage of overlap between actual predicted interval (numOverLap), our iterative IT dominates the CRM and CCRM in forecasting the lower and upper bounds of the intervals of the SP500 daily return. This outperformance is getting more pronounced as the forecast horizon increases. However, most coverage statistics proposed seem to favor the CRM and CCRM. This is due to their emphasis on a complete coverage of the observed intervals at the expense of the much more informative efficiency rates statistics.

In addition to the simple autoregressive process we used to forecast the SP500, we also ran a more comprehensive scenarios where we included additional macro variables (including logarithmic transformation of these variables). Such macro variables include the Aruoba-Diebold-Scotti Business Conditions Index (Aruoba et al., 2009) and daily oil prices. Since our forecast results were not significantly improved and our iterative IT method still dominates, we do not present these results. They are available upon request.

To confirm the superiority of our forecasts over the other methods, we also used the modified Diebold-Mariano (MDM) test (Diebold and Mariano, 1995; Diebold, 2015; Harvey et al., 1997) for each pair of forecasts. The test results are quite impressive and statistically significant. It shows that our forecasts are superior to the other methods. These tests test the null hypothesis of no difference between the errors produced by the forecasts made with two different methods. Table 1 shows the MDM test results in terms of the MSE of the forecasts errors. A negative MDM statistic for Method 1/ Method 2 implies that Method 1 yields a smaller forecast MSE than Method 2, and vice versa. The statistical significance of the difference is indicated by the number of asterisks attached to the statistics. The left panel of Table 1 shows the MSE forecasts

superiority of our iterative IT method over CRM and CCRM forecasts for the stable period. It is statistically significant at the 1% level in all cases (lower and upper bounds for all one-day to five-day ahead forecast horizons). The unstable period test results are shown in the right panel. The results are similar; our approach is significantly superior to the CRM and CCRM, except for the single case of a five-day ahead forecast. Overall, the modified Diebold-Mariano tests confirm that our iterative IT has a significant edge over the CRM and CCRM for forecasting both the lower and upper bounds across the five-day ahead forecast horizons.

5.4. Economic Structure

Under our approach we can identify the best possible model, or models, for forecasting each period and horizon. It provides us with a way to identify certain structural changes, or other major changes across time. To do so, we compare the inferred estimates across times. This joint model selection and estimation capability is one of the advantages of our approach. To do this under the other approaches is much more complicated or, even impossible.

Having analyzed our results, we can say the following. The basic model of each one of the two periods (stable and unstable) is practically unchanged, though the values of the inferred intercepts are changing across periods. The slope parameters, on the other hand, do not vary much over time. From an economic point of view, it means that the changes in the basic structure of the system over time (within each one of the basic two periods we analyze) are all due to the intercept; there are no other fundamental changes. With this in mind, we must emphasize that this analysis is based on the forecasting results and not on a fundamental and causal study of the interval data.

6. Conclusion

In this paper we use an iterative, information-theoretic estimator for forecasting interval-valued data. We contrast our forecasts with those resulting from competing

approaches. These include the CRM and CCRM methods, as well as the familiar conditional mean and variance model of AR(1)-GARCH(1,1). We use our method to forecast the SP500 returns. We use a 13-year sample of SP500 interval returns data. We forecast up to five days ahead based on moving windows. Following on González-Rivera and Lin (2013), we divide our data into two distinct periods: stable and unstable. The forecasts results show that for both periods our IT approach has the lowest prediction errors, defined as RMSE and MAE, for the lower and upper bounds of the intervals. However, CRM and CCRM perform well in terms of most coverage statistics by over-emphasizing complete-coverage of the intervals at the expense of efficiency rates. These methods emphasize a complete coverage of the interval, rather than a more accurate forecasting of the upper and lower bounds; coverage on account of accuracy. Such a trade-off between coverage and efficiency rates is natural in interval forecasting models. Naturally, the relative importance of the two depends on the objective of the forecasting exercise. In this paper, our objective is to accurately forecast the lower and upper bounds of the SP500 daily returns. This is in line with the overall objectives of most forecasting models. We are not interested here in forecasting a complete-coverage of the interval. Such a forecast objective will not allow us to do more accurate predictions of the market. With this in mind, our method allows us to accurately and efficiently forecast the returns. Our forecasts dominate the other methods for all periods and up to five days into the future. In fact, the further the forecast horizon, the better our method behaves relative to the other approaches.

Acknowledgements

We thank participants at The Third International Symposium on Interval Data Modelling at Beijing, China for their fruitful comments. We also thank the Editor Dick Van Dijk, the Associate Editor and two reviewers for their very helpful suggestions and comments.

References

- Ahn, J., Peng, M., Park, C., Jeon, Y., 2012. A resampling approach for interval-valued data regression. *Statistical Analysis and Data Mining* .
- Arroyo, J., González-Rivera, G., Maté, C., 2010. Forecasting with interval and histogram data: Some financial applications. *Handbook of Empirical Economics and Finance* .
- Arroyo, J., González-Rivera, G., Maté, C., San Roque, A.M., 2011. Smoothing methods for histogram-valued time series: An application to value-at-risk. *Statistical Analysis and Data Mining* 4, 216–228.
- Aruoba, S.B., Diebold, F.X., Scotti, C., 2009. Real-time measurement of business conditions. *Journal of Business & Economic Statistics* 27, 417–427.
- Beresteanu, A., Molchanov, I., Molinari, F., 2011. Sharp identification regions in models with convex moment predictions. *Econometrica* 79, 1785–1821.
- Beresteanu, A., Molinari, F., 2008. Asymptotic properties for a class of partially identified models. *Econometrica* 76, 763–814.
- Bertrand, P., Goupil-Testu, F., 2000. Analysis of symbolic data. exploratory methods for extracting statistical information from complex data, chapter descriptive statistics for symbolic data.
- Billard, L., 2007. Dependencies and variation components of symbolic interval-valued data. *Selected Contributions in Data Analysis and Classification* , 3–12.
- Billard, L., 2008. Sample covariance functions for complex quantitative data, in: *World Congress, International Association of Computational Statistics*, Yokohama, Japan.
- Billard, L., Diday, E., 2000. Regression analysis for interval-valued data. *data analysis, classification and related methods: Proceedings of the seventh conference of the international federation of classification societies, ifcs-2000, namur (belgium), kiers, hal et al. eds, vol. 1 (2000)*. Namur (Belgium), Kiers, HAL et al. Eds 1, 369–374.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 31, 307–327.
- Buansing, T.T., Fan, Y., Golan, A., Ullah, A., 2017. Information-theoretic model selection and estimation for interval data. *Working Paper* .
- Clements, M.P., Taylor, N., 2003. Evaluating interval forecasts of high-frequency financial data. *Journal of Applied Econometrics* 18, 445–456.
- Diebold, F.X., 2015. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics* 33.
- Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.
- Golan, A., 2018. *Foundations of Info-Metrics: Modeling, Inference and Imperfect Information*. Oxford University Press.

- Golan, A., Gzyl, H., 2002. A generalized maxentropic inversion procedure for noisy data. *Applied Mathematics and Computation* 127, 249–260.
- Golan, A., Gzyl, H., 2012. An entropic estimator for linear inverse problems. *Entropy* 14, 92–923.
- Golan, A., Judge, G., Miller, D., 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley and Sons.
- González-Rivera, G., Arroyo, J., 2012. Time series modeling of histogram-valued data: The daily histogram time series of s&p500 intradaily returns. *International Journal of Forecasting* 28, 20–33.
- González-Rivera, G., Lin, W., 2013. Constrained regression for interval-valued data. *Journal of Business and Economic Statistics* 31, 473–490. URL: <http://amstat.tandfonline.com/doi/abs/10.1080/07350015.2013.818004>, doi:10.1080/07350015.2013.818004, arXiv:<http://amstat.tandfonline.com/doi/pdf/10.1080/07350015.2013.818004>.
- Han, A., He, Y., Hong, Y., Wang, S., 2011. Forecasting of interval-valued crude oil prices with autoregressive conditional interval models. submitted to *International Statistical Review*.
- Han, A., Hong, Y., Lai, K., Wang, S., 2008. Interval time series analysis with an application to the sterling-dollar exchange rate. *Journal of Systems Science and Complexity* 21, 558–573. URL: <http://dx.doi.org/10.1007/s11424-008-9135-5>, doi:10.1007/s11424-008-9135-5.
- Han, A., Hong, Y., Wang, S., 2012. Autoregressive conditional models for interval-valued time series data, in: *The 3rd International Conference on Singular Spectrum Analysis and Its Applications*, p. 27.
- Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. *International Journal of forecasting* 13, 281–291.
- He, A.W., Kwok, J.T., Wan, A.T., 2010. An empirical model of daily highs and lows of west texas intermediate crude oil prices. *Energy Economics* 32, 1499–1506.
- Judge, G.G., Mittelhammer, R.C., 2011. *An Information Theoretic Approach to Econometrics*. Cambridge Univ Pr.
- Lima Neto, E.d.A., de Carvalho, F.d.A., 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., de Carvalho, F.d.A., Tenorio, C., 2005. Univariate and multivariate linear regression methods to predict interval-valued features. *AI 2004: Advances in Artificial Intelligence*, 211–229.
- Lin, W., González-Rivera, G., 2016. Interval-valued time series models: Estimation based on order statistics exploring the agriculture marketing service data. *Computational Statistics & Data Analysis* 100, 694–711.
- Maia, A.L.S., de Carvalho, F.d.A., Ludermir, T.B., 2008. Forecasting models for interval-valued time series. *Neurocomputing* 71, 3344–3352.
- Manski, C., 2003. *Partial Identification of Probability Distributions*. Springer-Verlag.
- Manski, C.F., Tamer, E., 2002. Inference on regressions with interval data on a regressor or outcome.

- Econometrica 70, 519–546.
- Owen, A.B., 2001. Empirical Likelihood. Champman &Hall/CRC.
- Rodrigues, P.M., Salish, N., 2015. Modeling and forecasting interval time series with threshold models. Advances in data analysis and classification 9, 41–57.
- Smith, R.J., 2005. Local gel methods for conditional moment restrictions. cemmap working paper, No. CWP15/05 URL: <http://www.cemmap.ac.uk/wps/cwp1505.pdf>, doi:10.1920/wp.cem.2005.1505.
- Xu, S., Chen, X., Han, A., 2008. Interval forecasting of crude oil price, in: Interval/Probabilistic Uncertainty and Non-Classical Logics. Springer, pp. 353–363.
- Xu, W., 2010. Symbolic Data Analysis: Interval-Valued Data Regression. Ph.D. thesis. University of Georgia.
- Xu, W., Billard, L., 2012. A study of symbolic intervals and its application to regression. Working Paper, University of Georgia .

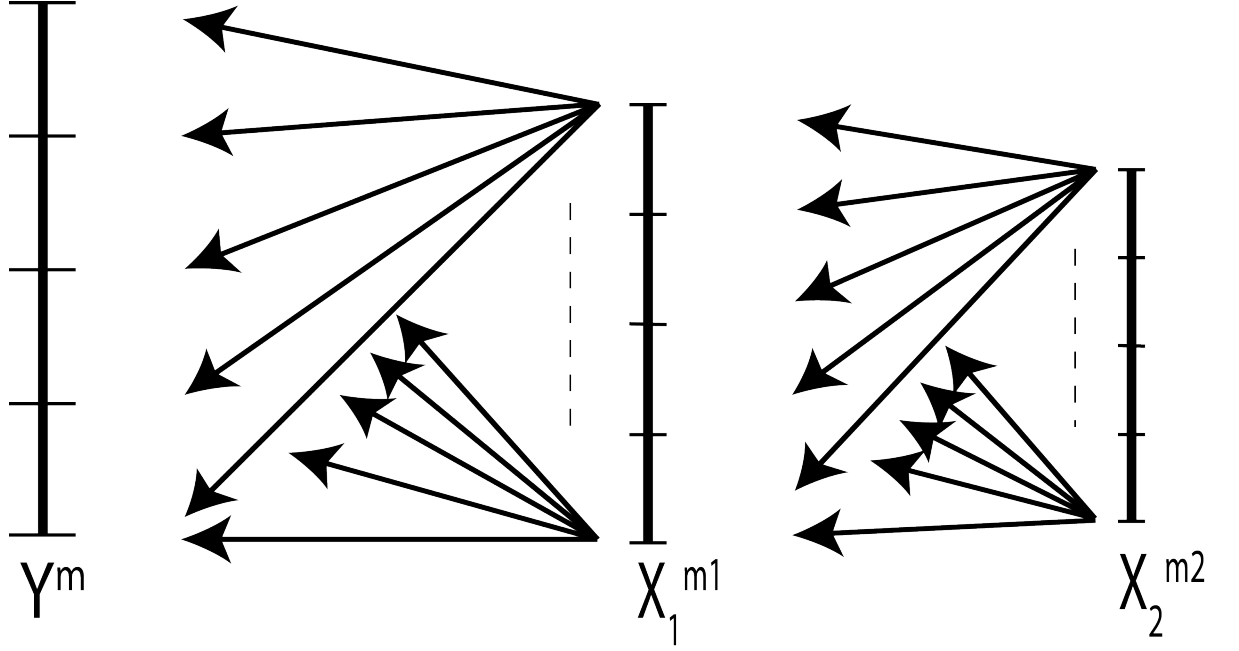


Figure 1: A graphical representation of our iterative approach. It estimates all possible combinations of the sub-intervals. For example, for $M = 5$ and $K = 1$, it iterates through 25 models: $(\mathbf{X}_1^1, \mathbf{Y}^1)$, $(\mathbf{X}_1^1, \mathbf{Y}^2), \dots, (\mathbf{X}_1^1, \mathbf{Y}^5)$, $(\mathbf{X}_1^2, \mathbf{Y}^1)$, $(\mathbf{X}_1^2, \mathbf{Y}^2), \dots, (\mathbf{X}_1^5, \mathbf{Y}^5)$. For $M = 5$ and $K = 2$, it requires 125 iterations since the number of all possible combination is $M \times M^K$.

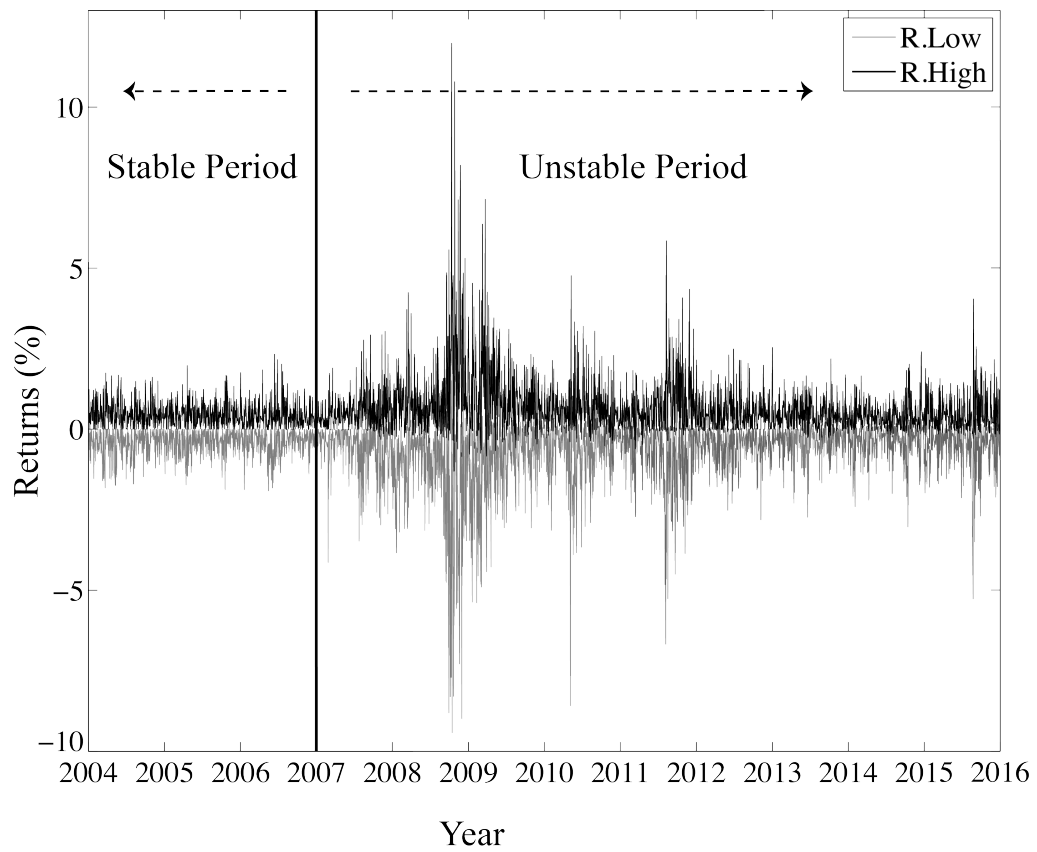


Figure 2: Low (R.Low) and high (R.High) returns of daily SP500 index from 01/01/2004 - 01/01/2016. The stable period is defined as the trading days period between 01/01/2004 and 01/01/2007. The unstable period is from 01/01/2007 through 01/01/2016.

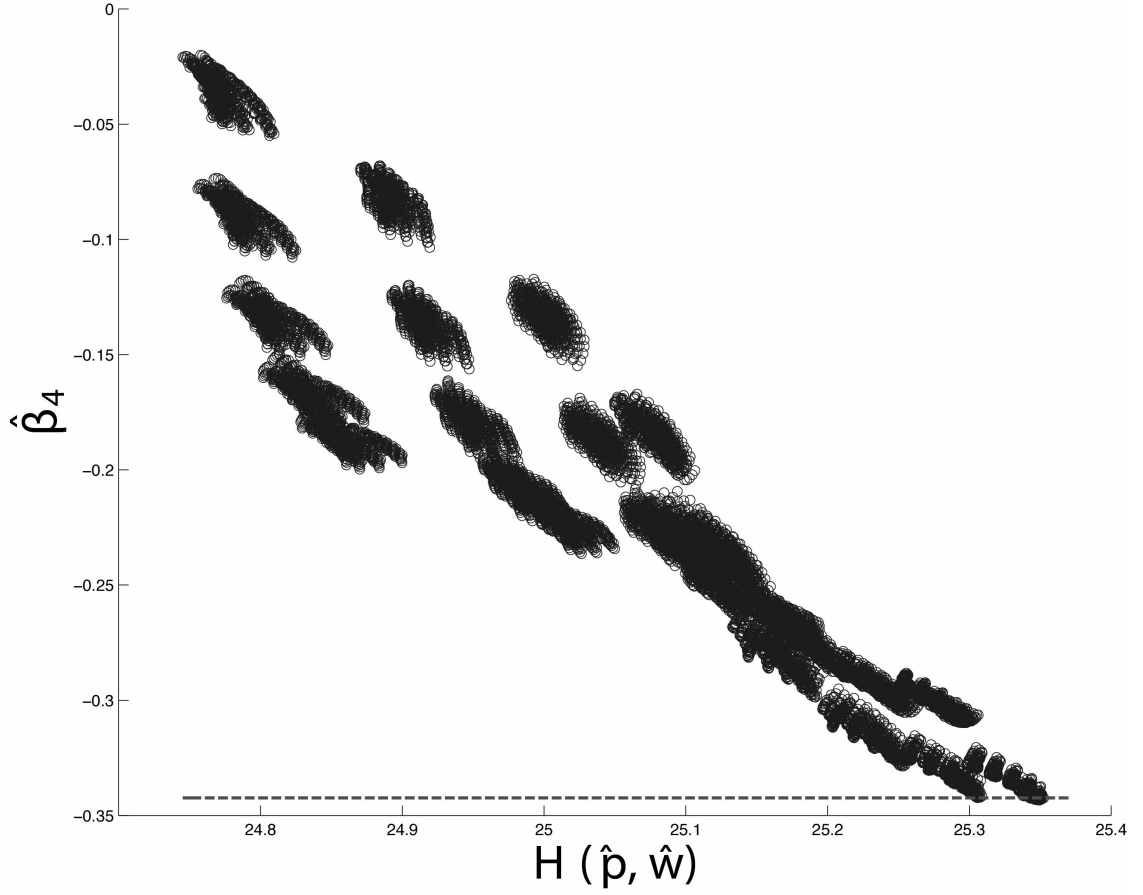
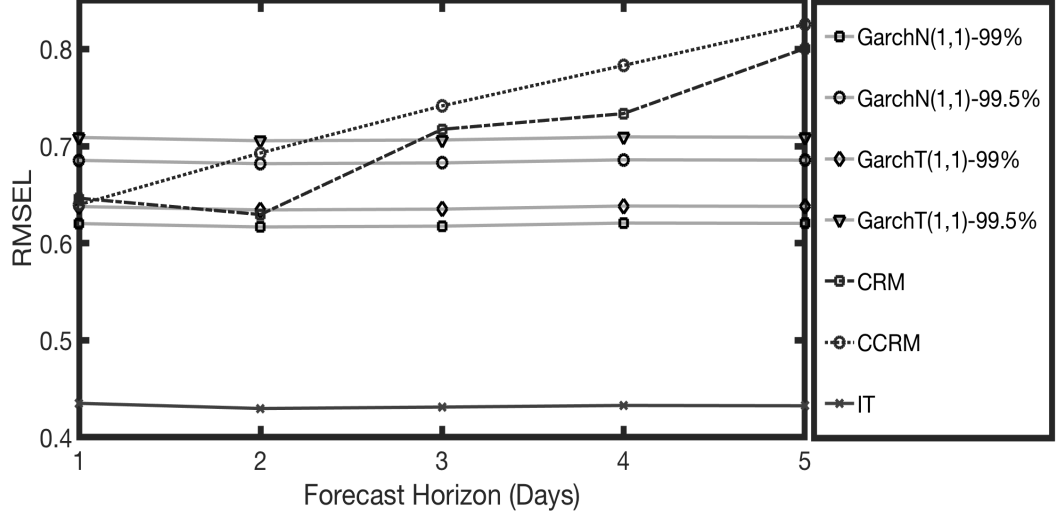
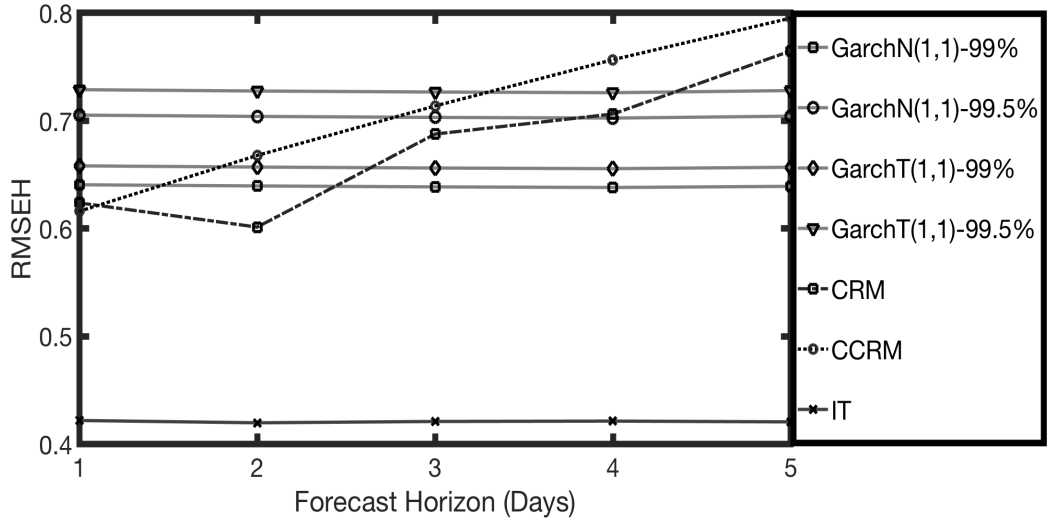


Figure 3: Scatter plots of the estimated $\hat{\beta}_4$ coefficient vs. the estimator's objective values for the first rolling window of the iterative, IT method for the unstable period. The estimation window size is 20 days with $M = 5$ sub-intervals, $\mathbf{z} = (-500, 250, 0, 250, 500)$ and $\mathbf{v} = (-3\hat{\sigma}_Y, 0, 3\hat{\sigma}_Y)$. See Section 3 for more details. Dashed lines show the coefficients chosen by the entropy criterion. It suggests that the iterative IT can consistently pick the most probable model. That same behavior is observed for all the periods analyzed.

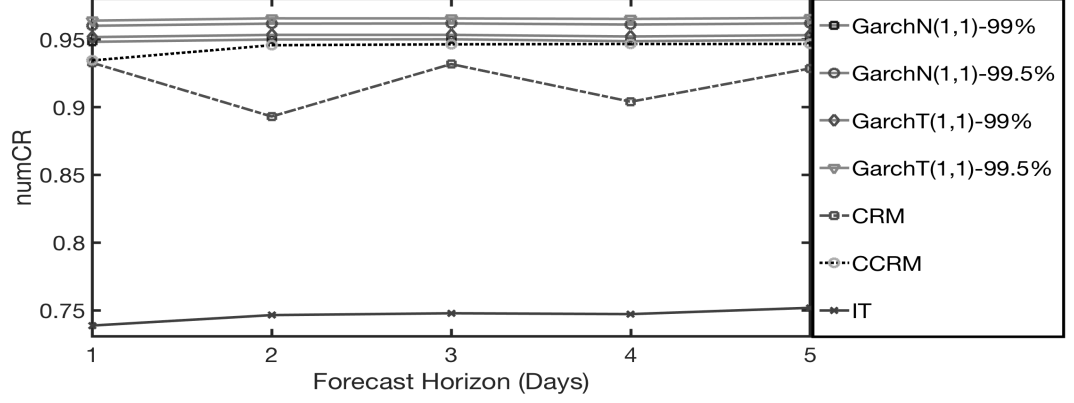


(a)

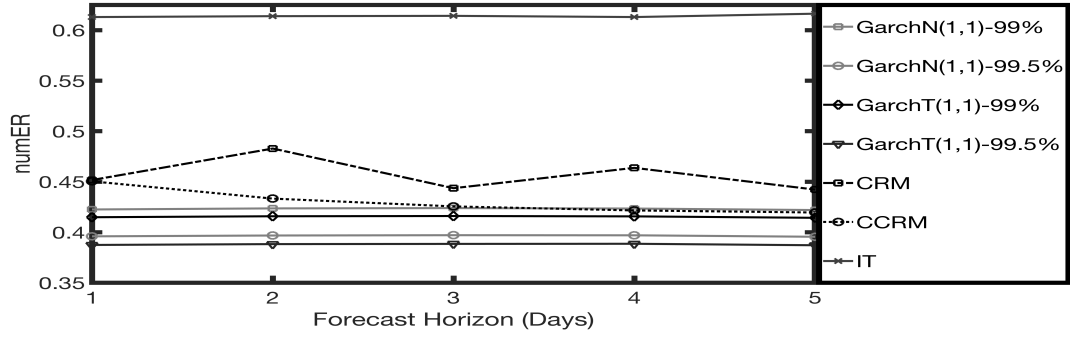


(b)

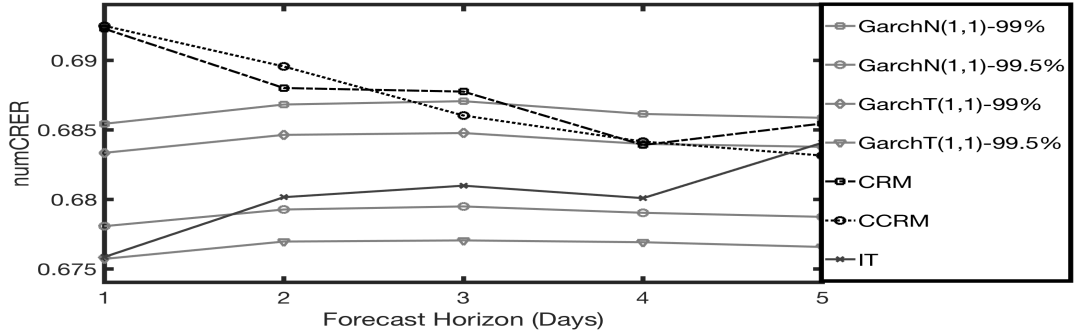
Figure 4: Five-day ahead forecast evaluations (RMSEL, RMSEH) over all rolling windows ($RW_E = 20$ days) for the iterative IT, CRM and CCRM during the stable period. Panel (4a) presents the RMSEL (Root Mean Squared Errors of the lower bound) while Panel (4b) presents the RMSEH (Root Mean Squared Errors of the upper bound). $RW_E = 500$ days for GARCH(1,1)-Normal and GARCH(1,1)-T. For iterative IT estimator, we choose the model with the highest entropy (objective) values for the slope parameters; $\hat{\beta}_0^L$ and $\hat{\beta}_0^U$ are obtained from minimum and maximum of the empirical distribution of $\hat{\beta}_0$, respectively (Eq. (23)). The iterative IT uses $M = 5$ sub-intervals, $\mathbf{z} = (-500, 250, 0, 250, 500)$ and $\mathbf{v} = (-3\sigma_Y, 0, 3\sigma_Y)$. Evaluation statistics are defined in Section (4). Our IT method outperforms the other methods under these measures.



(a)



(b)



(c)

Figure 5: Five-day ahead forecast evaluations (numCR -Panel a, numER -Panel b, numCRER -Panel c) over all rolling windows ($RW_E = 20$ days) for iterative IT, CRM and CCRM during the stable period. $RW_E = 500$ days for GARCH(1,1)-Normal and GARCH(1,1)-T. Evaluation statistics are defined in Section (4). Existing methods over-emphasize coverage and hence perform better under the numCR and numCRER statistics. However, that advantage deteriorates as forecast horizon increases. In terms of the efficiency rate (numER) our IT method dominates.

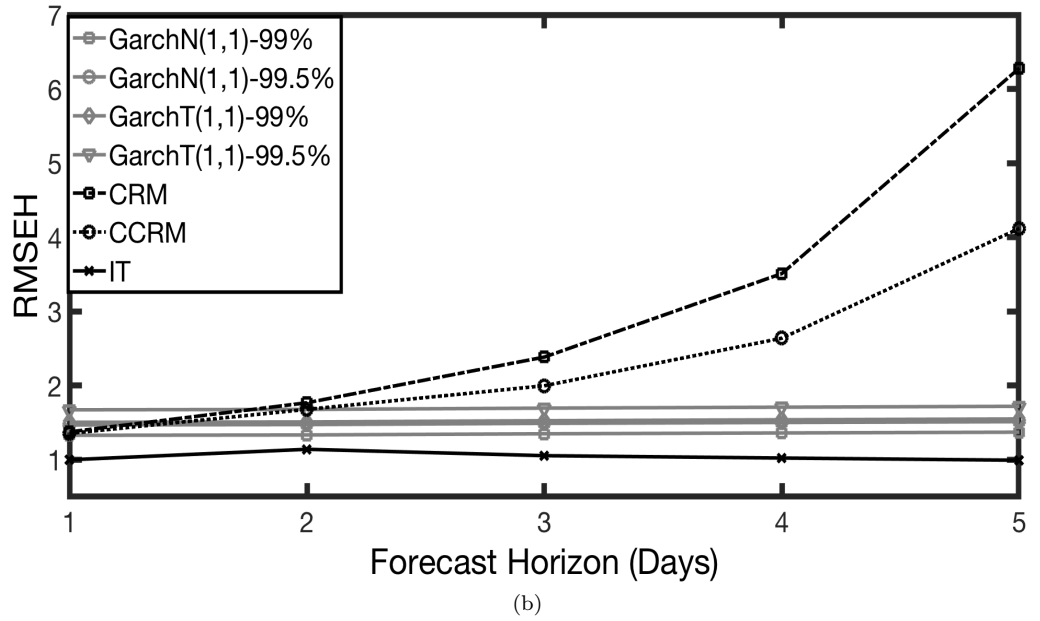
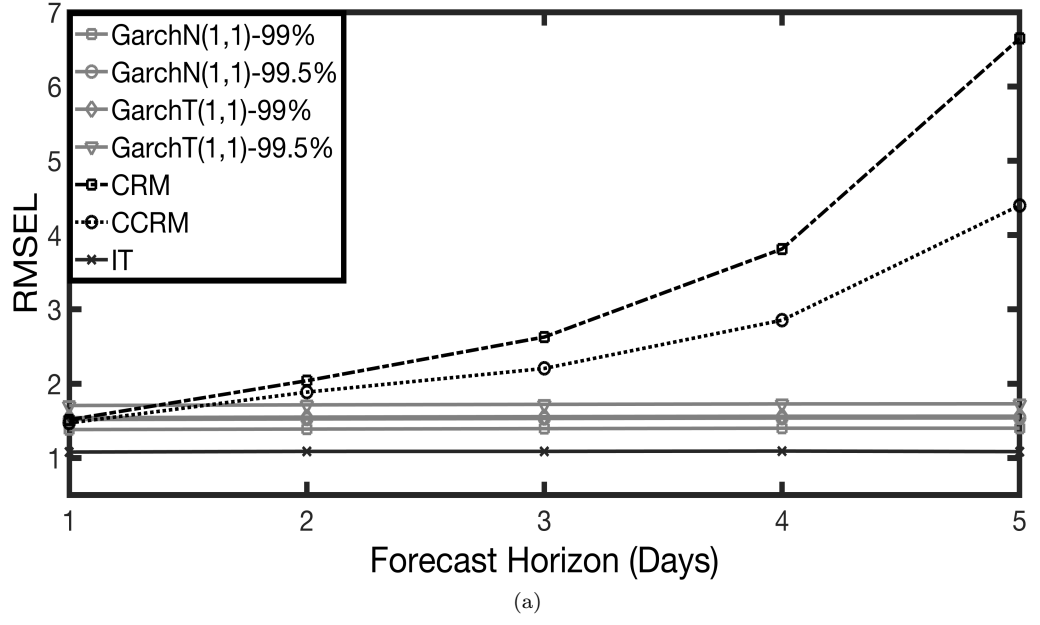


Figure 6: Five-day ahead forecast evaluations (RMSEL -Panel a and RMSEH -Panel b) over all the rolling windows ($RW_E = 20$ days) for iterative IT, CRM and CCRM during the unstable period. $RW_E = 500$ days for GARCH(1,1)-Normal and GARCH(1,1)-T. Evaluation statistics are defined in Section (4).

Stable Period				
	IT/CRM		IT/CCRM	
	Low	High	Low	High
h=1	-14.54**	-14.98**	-14.62**	-15.00**
h=2	-9.44**	-10.35**	-12.99**	-15.61**
h=3	-9.13**	-9.90**	-10.31**	-11.44**
h=4	-6.12**	-5.86**	-7.72**	-7.61**
h=5	-5.18**	-4.95**	-5.80**	-5.57**

Unstable Period				
	IT/CRM		IT/CCRM	
	Low	High	Low	High
h=1	-8.24**	-8.27**	-7.71**	-8.44**
h=2	-5.30**	-5.25**	-5.81**	-5.97**
h=3	-4.75**	-5.28**	-4.25**	-4.44**
h=4	-3.34**	-3.58**	-2.48**	-2.48**
h=5	-2.27*	-2.32*	-1.47	-1.46

Table 1: Modified Diebold-Mariano Test Results for Stable and Unstable Periods. Negative values correspond to superiority of the IT approach. The asterisks correspond to the level of significance: * and ** indicate, respectively, significance at the 5%, and 1% levels.