

# Irregularly Spaced Intraday Value at Risk (ISIVaR) Models Forecasting and Predictive Abilities

Gilbert Colletaz, Christophe Hurlin and Sessi Tokpavi

*LEO, University of Orléans. Rue de Blois. BP 6739. 45067 Orléans Cedex 2.  
France. Corresponding author: sessi.tokpavi@univ-orleans.fr*

This draft, July 2007

---

## Abstract

The objective of this paper is to propose a market risk measure defined in price event time and a suitable backtesting procedure for irregularly spaced data. Firstly, we combine Autoregressive Conditional Duration models for price movements and a non parametric quantile estimation to derive a semi-parametric Irregularly Spaced Intraday Value at Risk (ISIVaR) model. This ISIVaR measure gives two information: the expected duration for the next price event and the related VaR. Secondly, we use a GMM approach to develop a backtest and investigate its finite sample properties through numerical Monte Carlo simulations. Finally, we propose an application to two NYSE stocks.

*Key words:* Value at Risk, High-frequency data, ACD models, Irregularly spaced market risk models, Backtesting.

---

<sup>1</sup> We would like to thank Pierre Giot and Renaud Beaupain for providing us with NYSE Trades And Quotes (TAQ) data, and the participants at the 14th 'Forecasting Financial Markets (FFM)' Conference in Aix-en-Provence (May, 2007) for their helpful comments. Any errors and inaccuracies are our owns.

## 1 Introduction

The availability of high-frequency (or tick by tick) data, induced by the evolution of the trading environment on the major financial places, has led to the emergence of a new category of active market participants, such as high frequency traders. The latter are characterized by very short investment horizons and then require new market risk methodology: since risk must be evaluated on shorter than daily time intervals, traditional risk measures, such as Value at Risk (VaR), must be extended to intraday data context. This new body of research receives less attention in the relevant literature compared to definition and validation of day-to-day risk measures.

To the best of our knowledge, two attempts to derive intradaily market risk models using tick by tick data are those of Giot (2005) and Dionne et al. (2006). Giot (2005) quantifies market risk at an intraday time horizon, using Normal GARCH, Student GARCH, RiskMetrics for deseasonalized tick by tick data sampled at equidistant time. He also applied the Log-ACD model on price duration to compute irregularly spaced VaR and then scale them to derive fixed-time intervals VaR. The Intraday-VaR (IVaR) of Dionne et al. (2006) is based on a rich model of price dynamics conditional on durations- known as the Ultra-High-Frequency GARCH (UHF-GARCH) model of Engle (2000)- such that unequally spaced VaR can be easily generated in a convenient way. But, the authors instead make use of a simulation-based method to infer VaR at any fixed-time horizon<sup>2</sup>. So, in both approaches, the unequally spaced nature of high-frequency market risk models is forfeited, mainly because of backtesting procedure.

This restriction obviously implies a loss of information, since durations be-

---

<sup>2</sup> It seems in the case of Dionne et al. (2006) paper, as in Giot (2005), that VaRs are rescaled in fixed-time intervals for validation purpose.

tween market events<sup>3</sup> are an essential dimension of risk when dealing with tick by tick data. A very short duration forecast thus indicates in the line of microstructure theory (Easley and O'Hara, 1992) that there are many informed traders, and this information with the level of the forecast value of VaR, will determine the market monitoring of traders. Besides, these durations allow assessing liquidity risk, with for instance the definition of Time at Risk (TaR) measures (Ghysels *et al.*, 2004).

In this context, our objective is to propose a market risk or VaR methodology defined in price events time (and not in calendar time) and a corresponding backtesting procedure. For that, we define an ISIVaR (Irregularly Spaced Intraday Value at Risk) model which consists in a couple of two measures: the forecast of the timing for the next price event (or the expected duration between two consecutive price changes) and the corresponding level of risk summarized by VaR forecast. This VaR corresponds to the maximum expected loss that will not be exceeded (at a given confidence level) at the next price event, if this event occurs. More precisely, the ISIVaR is derived from an Autoregressive Conditional Duration (ACD) model applied to deseasonalized price event durations as in Giot (2005). However, contrary to Giot, we do not impose a particular distribution on the standardized returns to derive VaR measure from price changes volatility. We use a semi parametric approach similar to that considered by Engle and Manganelli (2001) in the day-to-day VaR perspective.

We also propose a backtesting procedure that allows testing the accuracy of our irregularly spaced VaR forecasts. The main advantage of this procedure is that it does not require rescaling ISIVaR forecasts to fixed-time intervals. As usual in the backtesting literature (see Campbell, 2007 for a survey) our model

---

<sup>3</sup> Market events can be either trades or defined using a particular time transformation (see for e.g. LeFol and Mercier, 1998). In this paper, we will focus on price events, *i.e.* the minimum amount of time needed for the price to have a significant change.

free backtest admits a conditional coverage null hypothesis (Christoffersen, 1998) and is based on a hit-no-hit variable<sup>4</sup>. But, in an irregularly spaced tick by tick data context, the hit-no-hit variable  $I_n$  indicates for the market event number  $n$ , if there is a hit or not. This hit-no-hit variable is irregularly spaced and then, most of usual backtesting procedures (such as the dynamic quantile test of Engle and Manganelli, 2004) can no longer be used. Consequently, we build a new test by using the fact that for a correctly specified irregularly spaced VaR models, the variable that counts the number of market events recorded before having a hit (which we call here events-hit-count variable), must have an exponential distribution. This idea is related to the duration-based test for predictive abilities of fixed-time interval VaR (Christoffersen and Pelletier, 2004). Nevertheless, there is a major difference, because the variable we focus on in the testing strategy is the events-hit-count variable, whereas in Christoffersen and Pelletier (2004), the exponential assumption is about the number of calendar time units (or days) before having a hit, or the time duration between hits.

Another contribution of this paper lies on the framework used to test exponential assumption for the observed sample of events-hit-count variable. Contrary to the Likelihood Ratio approach developed in Christoffersen and Pelletier (2004) to test the hypothesis of exponential distribution, we do not specify the form of the distribution under the alternative of misspecified ISI-VaR model. We instead use the GMM approach of distributional assumptions testing of Bontemps and Meddahi (2006). Our test is then robust to any possible specification under the alternative of inaccurate irregularly spaced intraday-VaR models.

The rest of the paper is organized as follow: in the first section, we derive a semi-parametric Irregularly Spaced Intraday-VaR (ISIVaR) model from ACD

---

<sup>4</sup> The hit-no-hit variable is generally defined as an indicator variable associated with the *ex-post* observation of a VaR violation at time  $t$ .

models applied to price movements, while in the second, we develop a test for the predictive abilities of such VaR models, and deal with its finite sample properties through monte carlo study. In a last empirical section, we illustrate the usefulness of our methodology by assessing the accuracy of ISIVaR model applied to two stocks traded on the NYSE. A last section concludes and submits further extensions.

## 2 Irregularly Spaced Intraday-VaR (ISIVaR)

Let us consider that tick by tick data for a given stock is generated by the *marked* point process  $(t_i, a_{t_i}, b_{t_i}, z_{t_i})$ ,  $i = 1, \dots, n$ , where  $t_i$  is the time occurrence of the trade number  $i$ ,  $a_{t_i}$  and  $b_{t_i}$  are respectively the ask and bid prices prevailing when the  $i^{th}$  trade occurs, and  $z_{t_i}$  a  $(k, 1)$  vector of other marks (volumes, bid-ask spreads, etc.). From this process, let us select only those points for which prices have changed<sup>5</sup>. By doing so, we are performing thinning of the original sample (defined in transaction times) by selecting a new point process *i.e.* price changes (or price events) arrival times. However, as underlined by Engle and Russell (1998), prices can sometimes move temporarily and return to their previous levels, due to quoting errors, or inventory control. To take into account those minor or insignificant changes, one has to define a pre-specified threshold  $c$  and selects only the points for which prices have increased or decreased by at least  $c$ . This leads to a new *marked* point process  $(t'_i, p_{t'_i}, z_{t'_i})$ ,  $i = 1, \dots, n_c$ , with  $n_c$  the total number of filtered quotes. The corresponding price changes returns are thus defined as  $r_{t'_i} = \ln p_{t'_i} - \ln p_{t'_{i-1}}$ .

Let us recall that, Value at Risk is a measure of how the value of an asset or of a portfolio of assets is likely to decrease (with a given confidence level) over a certain time period, and this under usual market conditions. From a

<sup>5</sup> Generally, to avoid bid-ask bounce effect, prices are defined on the mid-point of the bid and ask prices, *i.e.*  $p_{t_i} = (a_{t_i} + b_{t_i}) / 2$ .

statistical point of view, VaR for a shortfall probability  $\alpha$  is the  $\alpha$ -quantile of asset return distribution over this period. Dealing with the thinned process, it is obvious that there is no specified period, since the events considered (price changes) occur stochastically. In this context, we can state that

**Definition 1** *Irregularly Spaced Intraday Value at Risk (ISIVaR) for a short-fall probability  $\alpha$ , is a couple  $(\psi_i, ISIVaR_i(\alpha))$  that gives simultaneously two main information, namely, the expected duration for the  $i^{\text{th}}$  price change,  $\psi_i$ , and the corresponding level of risk  $ISIVaR_i(\alpha)$  such as*

$$\Pr[r_{t'_i} < -ISIVaR_i(\alpha) \mid \mathcal{F}_{t'_{i-1}}] = \alpha \quad (1)$$

with  $\mathcal{F}_{t'_{i-1}} = \{(t'_j, p_{t'_j}, z_{t'_j}) \mid j = 1, \dots, i-1\}$ , the set of information available up the price change number  $i-1$ .

The ISIVaR is then a couple  $(\psi_i, ISIVaR_i(\alpha))$  that measures two dimensions of risk: a forecast of the market risk that will be occurred at the next price change and a forecast of the expected duration before the occurrence of this next price change, which can be interpreted as a liquidity risk. This duration forecast  $\psi_i$  can also be expressed as a Time at Risk (TaR) measure as suggested by Ghysels *et al.* (2004). Let  $x_i = t'_i - t'_{i-1}$  denotes the  $i^{\text{th}}$  duration between two price changes that occur at times  $t'_{i-1}$  and  $t'_i$ . For a given level  $\alpha$ ,  $TaR(\alpha)$  denotes the minimal duration without a price change that may occur with probability  $\alpha$  :

$$\Pr[x_{i+1} > TaR_i(\alpha)] = \alpha \quad (2)$$

We now propose a simple approach to derive the ISIVaR.

## 2.1 Modelling expected duration and price changes volatility

Without loss of generality, suppose the distribution of variable  $r_{t'_i}$  is a scale one, with zero conditional mean

$$r_{t'_i} = \sigma \left( t'_i \middle| \mathcal{F}_{t'_{i-1}} \right) \varepsilon_{t'_i} \quad (3)$$

where  $\sigma \left( t'_i \middle| \mathcal{F}_{t'_{i-1}} \right)$  is the price change volatility at time  $t'_i$  and  $\varepsilon_{t'_i}$  an *i.i.d.* innovation with zero mean and unit variance. The level of risk for the  $i^{th}$  price variation can now be expressed as

$$ISIVaR_i(\alpha) = -F^{-1}(\alpha) \sigma \left( t'_i \middle| \mathcal{F}_{t'_{i-1}} \right) \quad (4)$$

with  $F(\cdot)$ , the cumulated distribution function of variable  $\varepsilon_{t'_i}$ . As Giot (2005), we use an Autoregressive Conditional Duration (ACD) model applied to price durations variable  $x_i$  in order to generate *1-ahead out-of-sample* forecast values of both components of ISIVaR, *i.e.*,  $\psi_i$ , and  $ISIVaR_i(\alpha)$ .

The ACD model, introduced by Engle and Russell (1998), allows reproducing many empirical features such as clustering in market events, *i.e.*, durations processes are positively autocorrelated with a strong persistence (in the spirit of ARCH class models for equally spaced time series returns). Formally, ACD models treat the time between events as random, and in their formulation, scale the series of observed durations such that the new series is *i.i.d*

$$x_i = \psi_i \left( \mathcal{F}_{t'_{i-1}} \right) v_i \quad (5)$$

where  $v_i$  is an *i.i.d* positive-valued sequence with distribution  $f(\cdot)$  and  $E(v_i) = 1 \forall i$ . A recursive specification can be used to resume the dynamics of the scale function  $\psi_i$  which induces the ACD( $m, q$ ) model

$$\psi_i = w_d + \sum_{j=1}^m \alpha_{d,j} x_{i-j} + \sum_{j=1}^q \beta_{d,j} \psi_{i-j} \quad (6)$$

with  $w_d > 0$ ,  $\alpha_{d,j} \geq 0$  and  $\beta_{d,j} \geq 0$  to ensure like ARCH-type models, the

positivity of  $\psi_i$  and thus  $x_i$ . Another specification referred as Log-ACD model and due to Bauwens and Giot (2000), avoids the need of the above constraints on the parameters, by assuming a recursive equation for the logarithm of  $\psi_i$ , similarly to the extension of time series GARCH specification to EGARCH model of Nelson (1991). Let us precise that the conditional mean (or expected duration) and variance for the durations process  $x_i$  are respectively

$$E(x_i | \mathcal{F}_{t'_{i-1}}) = \psi_i E(v_i | \mathcal{F}_{t'_{i-1}}) = \psi_i \quad (7)$$

and

$$V(x_i | \mathcal{F}_{t'_{i-1}}) = \psi_i^2 V(v_i | \mathcal{F}_{t'_{i-1}}) = \psi_i^2 \pi^2 \quad (8)$$

with  $\pi^2$ , the variance of *i.i.d* innovations  $v_i$ . It follows that conditional dispersion (defined as the ratio of the conditional variance and square conditional mean) is equal to

$$\frac{V(x_i | \mathcal{F}_{t'_{i-1}})}{E(x_i | \mathcal{F}_{t'_{i-1}})^2} = \frac{\psi_i^2 \pi^2}{\psi_i^2} = \pi^2 \quad (9)$$

such that ACD models are in this sense enough flexible to take into account both overdispersion (resp. underdispersion) for  $\pi > 1$  (resp.  $\pi < 1$ ). The choice of  $v_i$  among parametric family of lifetime distributions yields many variants of ACD models (see Pacurar 2006 for a survey): EACD (with an Exponential disturbance) and WACD (Weibull) in Engle and Russell (1998), GACD (Generalized Gamma) in Lunde (1999) and Burr-ACD (Burr distribution) in Grammig and Maurer (2000).

Let  $N(t') = \sum_{i \geq 1} 1_{\{t'_i < t'\}}$  be a counting variable equal to the total number of events that have occurred by time  $t'$ . Then, one can generally characterizes duration models in continuous time framework, by the use of intensity function defined as

$$\lambda(t' | \mathcal{F}_{t'}, N(t')) = \lim_{\Delta \rightarrow 0} \frac{p(N(t' + \Delta) > N(t') | \mathcal{F}_{t'}, N(t'))}{\Delta} \quad \forall t' \quad (10)$$

where  $\mathcal{F}_{t'} = \{(t'_i, p_{t'_i}, z_{t'_i}) \mid t'_i < t'\}$  is the continuous counterpart of  $\mathcal{F}_{t'_i}$  (see definition 1). For ACD models, the last expression takes the following simple

form

$$\lambda(t' | \mathcal{F}_{t'}, N(t')) = \lambda_v \left( \frac{x(t')}{\psi_{N(t')+1}} \right) \frac{1}{\psi_{N(t')+1}} \quad (11)$$

where  $\lambda_v(\cdot)$  denotes the hazard function of error term  $v$ , and  $x(t') = t' - t'_{N(t')}$ . Concretely, the intensity is the probability that an event occurs in the short time interval  $t' + \Delta$ , given that it has not occurred before  $t'$ , or say differently, the arrival rate of price events as forecast at time  $t'$ . For the Weibull innovation, the functional form of the intensity is

$$\lambda(t' | \mathcal{F}_{t'}, N(t')) = \left( \frac{\Gamma(1 + \frac{1}{\gamma})}{\psi_{N(t')+1}} \right) x(t')^{\gamma-1} \gamma \quad (12)$$

where  $\Gamma(\cdot)$  is the gamma function and  $\gamma$  the Weibull parameter. The price intensity function increases (resp. decreases) for  $\gamma > 1$  (resp.  $\gamma < 1$ ) introducing enough flexibility in the modeling. The case  $\gamma = 1$  reduces to the Exponential ACD model (EACD), with a constant price intensity function  $1/\psi_{N(t')+1}$ . The interest of expressing ACD models for price durations in terms of price intensity appears clearer, when one refers to results obtained by Engle and Russell (1998). They propose a link between price intensity and instantaneous price change volatility given by:

$$\tilde{\sigma}^2(t' | \mathcal{F}_{t'}) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} E \left[ \left( \frac{p_{t'+\Delta} - p_{t'}}{p_{t'}} \right)^2 | \mathcal{F}_{t'} \right] \quad \forall t'. \quad (13)$$

By using the counting process  $N(t')$ , we can formulate (13) in terms of the price intensity function

$$\begin{aligned} \tilde{\sigma}^2(t' | \mathcal{F}_{t'}) &= \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \Pr \left[ |p_{t'+\Delta} - p_{t'}| \geq c | \mathcal{F}_{t'} \right] \left[ \frac{c}{p_{t'}} \right]^2 \\ &= \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \Pr \left[ (N(t' + \Delta) - N(t')) > 0 | \mathcal{F}_{t'} \right] \left[ \frac{c}{p_{t'}} \right]^2 \end{aligned} \quad (14)$$

$$= \lambda(t' | \mathcal{F}_{t'}, N(t')) \left[ \frac{c}{p_{t'}} \right]^2 \quad \forall t' \quad (15)$$

where the last equality holds, by the definition of intensity function. To conclude, the instantaneous volatility can be inferred by running an ACD model

applied to price durations variable.

It is worthy to say that what matters here is the prediction of volatility at the dates where prices have moved, *i.e.*  $\sigma^2(t'_i | \mathcal{F}_{t'_{i-1}})$  and not the continuous picture of volatility  $\tilde{\sigma}^2(t' | \mathcal{F}_{t'})$ . The task is easy and is achieved by noting as in Giot (2005) that

$$\sigma^2(t'_i | \mathcal{F}_{t'_{i-1}}) = \tilde{\sigma}^2(t'_{i-1} | \mathcal{F}_{t'_{i-1}}). \quad (16)$$

In fact,  $\tilde{\sigma}^2(t'_{i-1} | \mathcal{F}_{t'_{i-1}})$  as defined by equation (13) corresponds to volatility for each time  $t'$  between  $]t'_{i-1}, t'_i]$ , *i.e.* any time starting and ending a price duration. It turns out that

$$\sigma^2(t'_i | \mathcal{F}_{t'_{i-1}}) = \lambda(t'_{i-1} | \mathcal{F}_{t'_{i-1}}, N(t'_{i-1})) \left[ \frac{c}{p_{t'_{i-1}}} \right]^2 \quad (17)$$

is the conditional volatility for price event number  $i$ , and consequently the level of risk for the  $i^{th}$  price variation is equal to

$$ISIVaR_i(\alpha) = -F^{-1}(\alpha) \left[ \frac{c}{p_{t'_{i-1}}} \right] \lambda(t'_{i-1} | \mathcal{F}_{t'_{i-1}}, N(t'_{i-1}))^{\frac{1}{2}} \quad (18)$$

The parameters of ACD models are  $\Theta_d = (w_d, \alpha_{d,j}, \beta_{d,j}, \phi)'$  with  $\phi$  a vector of parameters related to the innovation distribution  $f(\cdot)$ . The model can be estimated with standard maximum likelihood method, if we assume a given parametric density for  $v_i$ . As noted by Engle and Russell (1998), if conditional mean duration equation (6) is not misspecified, maximizing the likelihood function with an exponential disturbance leads to consistent Quasi Maximum Likelihood (QML) estimates.

**Remark 2** *Let us precise that, in real-world applications, price durations exhibit significant diurnal patterns, i.e. a seasonal component  $\varpi(t')$  that must be first removed from the observed duration  $x_i$ , such as  $x_i = \varpi(t'_i) \tilde{x}_i$ . ACD models must then be fitted to the seasonally adjusted durations  $\tilde{x}_i$ . The volatility*

for the price change number  $i$  is thus given by

$$\sigma^2 \left( t'_i \mid \mathcal{F}_{t'_{i-1}} \right) = \lambda \left( t'_{i-1} \mid \mathcal{F}_{t'_{i-1}}, N \left( t'_{i-1} \right) \right) \left[ \frac{c}{p_{t'_{i-1}}} \right]^2 \frac{1}{\varpi \left( t'_{i-1} \right)} \quad (19)$$

and modification of equation (18) follows.

## 2.2 An algorithm for ISIVaR forecasting

At this stage, all information necessary to compute irregularly spaced intraday VaR have been computed. In this section, we combine ACD models applied to price durations, and a non parametric quantile method to generate *1-ahead out-of-sample* ISIVaR forecast for the price change number  $n_c + 1$ , using the information contained in  $\mathcal{F}_{t'_{n_c}}$ . The principle is rather simple:

- (1) First of all, we use the available sample of seasonally adjusted durations  $\{\tilde{x}_i\}_{i=1}^{n_c}$  to estimate ACD model (equations 5-6) with for example an exponential disturbance  $v_i$ <sup>6</sup>. We thus obtain estimates for the parameters vector  $\hat{\Theta}_d$ , and compute the *1-ahead out-of-sample* expected duration  $\hat{\psi}_{n_c+1}$  for the price change number  $n_c + 1$

$$\hat{\psi}_{n_c+1} = \hat{w}_d + \sum_{j=1}^m \hat{\alpha}_{d,j} x_{n_c+1-j} + \sum_{j=1}^q \hat{\beta}_{d,j} \hat{\psi}_{n_c+1-j} \quad (20)$$

We also predict the *1-ahead* price intensity equal (for the exponential disturbance) to

$$\hat{\lambda} \left( t'_{n_c} \mid \mathcal{F}_{t'_{n_c}}, N \left( t'_{n_c} \right) \right) = \frac{1}{\hat{\psi}_{n_c+1}} \quad (21)$$

and the forecast value of volatility is given by

$$\hat{\sigma}^2 \left( t'_{n_c+1} \mid \mathcal{F}_{t'_{n_c}} \right) = \hat{\lambda} \left( t'_{n_c} \mid \mathcal{F}_{t'_{n_c}}, N \left( t'_{n_c} \right) \right) \left[ \frac{c}{p_{t'_{n_c}}} \right]^2 \frac{1}{\hat{\varpi} \left( t'_{n_c} \right)} \quad (22)$$

<sup>6</sup> We choose the exponential disturbance for the ease of presentation, but one can instead rely on other distributions for the duration model.

- (2) With the results of the above ACD model, compute the series of *in-sample* volatilities as follow

$$\hat{\sigma}^2(t'_i | \mathcal{F}_{t'_{i-1}}) = \hat{\lambda}(t'_{i-1} | \mathcal{F}_{t'_{i-1}}, N(t'_{i-1})) \left[ \frac{c}{p_{t'_{i-1}}} \right]^2 \frac{1}{\widehat{\varpi}(t'_{i-1})} \quad i = 1, \dots, n_c \quad (23)$$

and approximate non-parametrically  $F^{-1}(\alpha)$  by  $q$ , the empirical  $\alpha$ -quantile of *in-sample* standardized series of returns  $\widehat{\varepsilon}_{t'_i}$

$$q = \text{percentile} \left( \{\widehat{\varepsilon}_{t'_i}\}_{i=1}^{n_c}, 100\alpha \right) \quad (24)$$

where  $\widehat{\varepsilon}_{t'_i} = r_{t'_i} / \hat{\sigma}(t'_i | \mathcal{F}_{t'_{i-1}})$   $i = 1, \dots, n_c$ .

- (3) Finally, the value of ISIVaR for the next price change (with a short-fall probability  $\alpha$ ) is given by  $(\widehat{\psi}_{n_c+1}, ISIVaR_{n_c+1}(\alpha))$  where the second component is equal to

$$ISIVaR_{n_c+1}(\alpha) | \mathcal{F}_{t'_{n_c}} = -q\hat{\sigma}(t'_{n_c+1} | \mathcal{F}_{t'_{n_c}}) \quad (25)$$

It is worth noting that our method to estimate ISIVaR is semi-parametric in the sense that we do not need to specify the distribution of price event returns  $(\varepsilon_{t'_i})$ , but only the one related to the durations series *i.e.*  $v_i$ . A misspecification of the latter will however leads to inconsistent estimates of the intensity function and then volatilities. In order to be free of any source of misspecification, one can consider the estimation of duration model under the exponential disturbance as Quasi Maximum Likelihood. But in that case, the intensity function should be estimated non parametrically, by using for *e.g.* a k-nearest neighbour method as in Engle (2000). Note also that the ISIVaR model as described above is very simple, and can be widened, by considering a richer specification for the duration mean equation (6) as in Engle and Russell (1998). Indeed, the recursive equation can be extended by additional exogenous variables to capture some market microstructure effects. One can for *e.g.* introduces as explanatory variables, the number of transactions per

second, volume per transaction or bid and ask spread. This can help to capture Easley and O'Hara hypotheses, which advance that information-based trading predict lower durations, and thus higher volatilities.

### 3 Testing the Accuracy of ISIVaR models

In this section, we present a general setup for the predictive abilities of Irregularly Spaced Intraday VaR (ISIVaR) models, like the one exposed in previous sections. We begin by defining the testable hypothesis and after we build a test statistic and deal with its asymptotic distribution.

#### 3.1 The null hypothesis

Let  $\{r_{t'_i}\}$  be an univariate stochastic process  $r \equiv \{r_{t'_i} : \Omega \rightarrow R\}$  defined on the probability space  $(\Omega, \mathcal{F}, P)$ , with  $\mathcal{F} \equiv \{\mathcal{F}_{t'_i}, t'_i \in \mathbb{N}^*\}$  and  $\{\mathcal{F}_{t'_i}\}$  an increasing family of sub- $\sigma$ -algebras, such that  $\mathcal{F}_{t'_i} = \sigma\{t'_s, r_{t'_s}, z_{t'_s} \mid s \leq i\}$ . In this setup,  $r_{t'_i}$  is the high-frequency return of price change number  $i$  for a given asset and  $z_{t'_i}$  a vector of marked data. Let us suppose that we observe a sample path of variables  $r_{t'_i}$  and  $z_{t'_i}$ , and produce via a given model, say  $\mathcal{M}$ , a sequence of  $N$  1-ahead out-of-sample ISIVaRs ( $ISIVaR_n(\alpha)$ ,  $n = 2, \dots, N + 1$ ), each of them conditional on the information available up the price change number  $n - 1$ , using for *e.g.* the above algorithm. We then have, as already mentioned

$$\Pr[r_{t'_n} < -ISIVaR_n(\alpha) \mid \mathcal{F}_{t'_{n-1}}] = \alpha \quad \forall n \quad (26)$$

with  $\alpha \in (0, 1)$  the nominal shortfall probability level. Following Christoffersen (1998), we define the hit-no-hit variable as

$$I_n = \begin{cases} 1 & \text{if } r_{t'_n} < -ISIVaR_n(\alpha) \\ 0 & \text{if } r_{t'_n} \geq -ISIVaR_n(\alpha) \end{cases} \quad (27)$$

which informs when a price change occurs, if the observed return is lower or higher than the *ex-ante* level of ISIVaR. Any testable hypothesis concerning the accuracy of model  $\mathcal{M}$  can be formulated using the so-called *conditional coverage* hypothesis, due to Christoffersen (1998)

$$E \left[ I_n - \alpha \mid \mathcal{F}'_{n-1} \right] = 0 \quad \forall n \quad (28)$$

or say differently

$$I_n \sim i.i.d \text{ Bernoulli}(\alpha) \quad \forall n. \quad (29)$$

To explain more, let us recall that  $ISIVaR_n(\alpha)$  is statistically the  $\alpha$ -quantile of the conditional distribution of the  $n^{th}$  price change. Thus, the probability of having a hit (or an *ex-post* loss higher than the *ex-ante* reported level of risk) must be equal to  $\alpha$ . Each price event can then be viewed as a trial, with probability of success ( $I_n = 1$ ) equal to  $\alpha$ . It follows that, independently to the time occurrence of any price change, having a hit or not is nothing but a Bernoulli trial, and the conditional coverage hypothesis applied for the indicator variable  $I_n$ . However, a practical question remains: can we use available tests (for conditional coverage hypothesis with equally spaced VaR)?

To answer, let us indicate that those existing tests are mainly the  $LR_{cc}$  test of Christoffersen (1998), the DQ test of Engle and Manganelli (2004), the duration-based test of Christoffersen and Pelletier (2004) and the tests of Berkowitz et al (2005) based on martingale difference property. For a brief review, the principle of the test of Christoffersen (1998) consists in postulating that the hit-no-hit process follows a two states markov chain. From then on, he deduces very easily a conditional coverage test by testing the estimated parameters of the transition matrix in a likelihood ratio framework. Engle and Manganelli (2004) propose a test based on the projection of the centered process of hit-no-hit on its  $K$  last values, a constant and exogenous variables. The derived test of conditional coverage then brings back to a joined nullity test of the parameters of this linear model. As for Christoffersen and Pelletier

(2004), they use the insight that if a VaR model is correctly specified, then the time between two consecutive hits or hit-duration should have no memory and a mean duration of  $1/\alpha$  days.

The great disappointment when dealing with the new defined variable  $I_n$  lies in the fact that the most powerful test among those mentioned, namely the DQ test is no longer adapted. In fact, this test is a regression-based one (using standard projections methods) and its relevance should be questioned with unequally spaced data. Nevertheless, the duration-based approach of Christoffersen and Pelletier (2004) continues to apply, but with a new sense given to the testing variable. Formally, we let  $C$  a variable we call events-hit-count, defined as the number of price changes recorded before having a hit, or say differently, the number of price events between two consecutive hits

$$C_i = n_i - n_{i-1} \quad (30)$$

where,  $n_i$  denotes the number of the price change at which the  $i^{th}$  hit occurs. Under the null hypothesis that the sequence of variable  $I_n$  is i.i.d Bernoulli( $\alpha$ ), the discrete probability function of variable  $C_i$  is given by

$$f(c, \alpha) = \Pr [C_i = c] = \Pr [I_{n_{i-1}+1} = 0, I_{n_{i-1}+2} = 0, \dots, I_{n_{i-1}+c} = 1]. \quad (31)$$

Using the fact that variables  $I_n$  are independently distributed and  $\Pr [I_j = 1] = \alpha \forall j$  we have

$$f(c, \alpha) = (1 - \alpha)^{c-1} \alpha \quad (32)$$

which is the lifetime distribution of a geometric variable. It follows that under the null of a well calibrated ISIVaR model, we should have

$$H_0 : C_i \sim \text{geometric}(\alpha) \quad \forall i \quad (33)$$

Following Christoffersen and Pelletier (2004), we use the continuous analogue of the geometric distribution, i.e. the exponential variable, to test the null hypothesis. Since variable  $C$  is defined in event times, this approximation will

introduce a discreteness bias, and its impact will be evaluated when dealing with monte carlo simulations. The exponential distribution which reaches the conditional coverage hypothesis ( $E [I_n - \alpha | \mathcal{F}'_{n-1}] = 0 \quad \forall n$ ) has two major implications, namely the unconditional coverage and independence hypothesis:

- The unconditional coverage prediction implies that the probability of an ex-post loss exceeding ISIVaR forecast (for any recorded price event) must be equal to the shortfall probability  $\alpha$

$$\Pr [I_n = 1] = E [I_n] = \alpha \quad (34)$$

*i.e.* the occurrence of losses exceeding ISIVaR forecasts must then correspond to the total number of price changes for which ISIVaRs are forecast. For a 5% ISIVaR, used as a reference measure over 1000 price events, the expected number of hits should be equal to 50. If this number is significantly higher or lower than 50, then the ISIVaR model fails the test. In term of variable  $C$ , this is equivalent to say that its mean should be equal to  $1/\alpha = 20$  *i.e.* the mean of an exponential distribution with parameter  $\alpha$ . However, the unconditional coverage property does not give any information about the temporal dependence of hits or equivalently a memoryless variable  $C$ .

- The independence prediction of hits is nevertheless an essential property, because it is related to the ability of a ISIVaR model to accurately model the higher order dynamics of high-frequency returns. In fact, a model which does not satisfy the independence property can lead to clusterings of hits (for a given group of price events), even if it has the correct average number of hits (see, Berkowitz and O' Brien (2002) for an illustration for daily VaR). So, there must not be any dependence in the hit-no-hit sequence, prediction that is summarized under the null by a distribution  $C$  with a lack of memory property.

Another contribution of this paper rises from the framework used to test exponential assumption about variable  $C$ . Contrary to the LR methodology of Christoffersen and Pelletier (2004), we do not specify the distribution of variable  $C$  under the alternative hypothesis. One robust approach which leads to a new test statistic for the accuracy of ISIVaR models is the distributional assumptions testing of Bontemps and Meddahi (2006). They derive a set of moment conditions that must hold for a given distribution and then proposed testing the null that a sample of observations is driven by the postulated distribution. For example, one can test normality by using a set of moment conditions (known as Stein Equation, see Bontemps and Meddahi (2005)) and this without defining the form of the distribution under the alternative hypothesis, like for example a Gaussian autoregressive model as in Berkowitz (2001). Here, we use this framework by testing directly the hypothesis of exponential distribution for variable  $C$ , robust to any specification under the alternative. In the next section, we present the methodology and derive our test statistics and its asymptotic distribution.

### 3.2 Test statistics and asymptotic distribution

Let  $Y$  be a stationary random variable with density function  $q(\cdot)$  and finite squared moments. Then, it exists a sequence of orthonormal polynomials  $L_k$  that can be expressed by the following equations known as Rodrigues formula

$$L_k(y) = \frac{\alpha_k}{q(y)} \left[ B^k(y) q(y) \right]^{(k)} \quad (35)$$

with  $f^{(k)}(\cdot)$  the  $k$ -th derivate function of  $f(\cdot)$  and  $\alpha_k$  defined as

$$\alpha_k = \frac{(-1)^k}{\sqrt{(-1)^k k! d_k \int B^k(y) q(y) dy}}; \quad d_k = \prod_{j=0}^{k-1} (-1 + (k + j + 1) c_2). \quad (36)$$

It can be shown that  $L_k(y)$  is a polynomial of degree  $k$  and satisfies the recurrence relation

$$L_{k+1}(y) = -\frac{1}{a_k} [(b_k - y) L_k(y) + a_{k-1} L_{k-1}(y)], \quad L_0(y) = L_{-1}(y) = 1 \quad (37)$$

where

$$a_k = \frac{\alpha_k d_k}{\alpha_{k+1} d_{k+1}}, \quad b_k = k\mu_k - (k+1)\mu_{k+1}, \quad \mu_k = \frac{-a + kc_1}{-1 + 2kc_2}. \quad (38)$$

The interest of such a decomposition lies in the fact that the sequences  $L_k(y)$  are orthonormal, i.e

$$E[L_k(y) L_{k'}(y)] = \begin{cases} 0 & \text{if } k \neq k' \\ 1 & \text{if } k = k' \end{cases} \quad (39)$$

and since  $L_0(y) = 1$ , we have

$$E[L_k(y)] = 0 \quad k = 1, 2, 3, \dots \quad (40)$$

Thus, for any given variable with marginal density  $q(\cdot)$ , the above orthogonal moment conditions must hold. Generally, for distributions among the Pearson's family (Normal, Student, Gamma, exponential, beta, etc.) the polynomials  $L_k(y)$  takes simple forms and one can easily derive the above moment conditions. For the exponential distribution with parameter rate  $\alpha$ , the recurrence equations (35) are known as Laguerre polynomials and we have in that case

$$L_0(y) = 1 \quad L_1(y) = 1 - \alpha y \quad (41)$$

$$L_{k+1}(y) = \frac{1}{k+1} ((2k+1 - \alpha y) L_k(y) - k L_{k-1}(y)) \quad \forall k \geq 1. \quad (42)$$

These polynomials are orthonormal and the moment conditions  $E[L_k(y)] = 0 \quad \forall k \geq 1$  are valid (if the distribution of variable  $Y$  is an exponential one, with parameter rate  $\alpha$ ) and can be tested, individually or jointly.

**Proposition 3** *Let us consider that for model  $\mathcal{M}$ , we generate  $N$  out-of-sample ISIVaRs ( $ISIVaR_n(\alpha)$ ,  $n = 2, \dots, N+1$ ) and compute the events-hit*

count variable  $C$ . Under the null of well-specified ISIVaR model, we have

$$E [L_k(c)] = 0 \quad k = 1, \dots, p \quad (43)$$

with  $L_k(c)$  the  $k^{\text{th}}$  Laguerre polynomials,  $p$  the total number of polynomials considered. It follows that

$$E [L(c)] = 0 \quad (44)$$

where  $L(c)$  is a vector of dimension  $(p \times 1)$ . Under some regularity conditions, we know since Hansen (1982) that

$$\frac{1}{\sqrt{S}} \sum_{i=1}^S L(c_i) \xrightarrow{L} N(0, I_p) \quad (45)$$

and the conditional coverage ( $cc$ ) statistic for the accuracy of ISIVaR model is

$$J_{cc} = \left( \frac{1}{\sqrt{S}} \sum_{i=1}^S L(c_i) \right)' \left( \frac{1}{\sqrt{S}} \sum_{i=1}^S L(c_i) \right) \xrightarrow{L} \chi^2(p) \quad (46)$$

with  $S$  the length of variable  $C$ .

This test statistic is easy to compute with standard asymptotic distribution, and traditional rule of decision applies. The unconditional coverage version of our test statistic ( $J_{uc}$ ) is obtained when one considers only the first Laguerre polynomial ( $p = 1$ ). Indeed, in that case, we focus only on the mean of variable  $C$ , i.e.

$$E [L_1(c)] = 0 \Leftrightarrow E [c] = \frac{1}{\alpha} \quad (47)$$

### 3.3 Monte Carlo Study

In this section, Monte Carlo experiments are conducted to evaluate the finite-sample performance of the proposed testing procedure. More precisely, we examine the empirical size and power of the asymptotic test using sample sizes available when dealing with high-frequency returns.

To evaluate the empirical size, we directly simulate  $N$  hit-no-hit variables  $I_n$   $n = 1, \dots, N$ , using a Bernoulli distribution with rate parameter  $\alpha$ . This sample

is typically the one that must arise from a well-calibrated ISIVaR model, for a shortfall probability  $\alpha$ . We compute the events-hit-count variable  $C$  and our test statistic. The empirical size then corresponds to the rejection frequencies observed in 10 000 simulations. If the asymptotic distribution of our test is adequate, then these rejection frequencies should be close to the nominal size used to reject (or accept) the null hypothesis. Table 1 presents the empirical size of our test for various sample sizes  $N$ , number of Laguerre polynomials  $p$ , nominal shortfall probability  $\alpha$ , and a nominal size set at 5%. The conclusion from the reported results is that, the unconditional version of our test statistic is oversized. The optimal value of  $p$  when dealing with the statistic  $J_{cc}$  is 2, for which rejection rates of the test are always quite close to the nominal size. For this value, the asymptotic distribution of our test statistic is then valid with realistic sample sizes, and one can rely on the asymptotic critical values of the chi-square distribution.

In order to evaluate empirical power, we simulate a sample of size  $N$  of Bernoulli trials with rate parameter  $\alpha + \kappa$ , where  $\kappa$  is drawn randomly in an uniform distribution on the interval  $[0, 0.1]$ . This reflects a situation where one uses an ISIVaR model, which underestimates the latent level of risk for price events, leading to an excessive number of hits<sup>7</sup>. With this sample of Bernoulli trials (*i.e.* hit-no-hit variable), we compute the events-hit-count variable  $C$  and apply our test statistic. The power is equal to the rejections rate for 10 000 simulations, with a given nominal size. Table 2 presents the results (for nominal size set at 5%). For the  $J_{cc}$  statistic, we notice that, with given values of  $\alpha$  and size  $N$ , the powers decrease when the number of Laguerre polynomials increases. The optimal number of Laguerre polynomials for this version of our test statistic is then two, a compromise between accurate size and high level of power and one can see that the obtained values are very clear-cut. In-

---

<sup>7</sup> We also consider the converse and obtain similar results. In this case, the parameter rate of the bernoulli distribution is  $\alpha - \kappa > 0$ .

deed, with the smaller sample size used ( $N = 1000$  events) the power is about 0.8978 for a shortfall probability  $\alpha$  equal to 1% and 0.8042 when  $\alpha = 5\%$ , and converges quickly towards one, when the sample size increases.

## 4 Empirical Applications

In this last section, we empirically assess the relevance of our methodology to compute irregularly spaced intraday market risk, using tick by tick data for two stocks traded at the NYSE, *i.e.* IBM and EXXON. The data was extracted from the Trade and Quotes (TAQ) database and include for each stock, information on every single trade and quote over the period February-April 97. The database consists of two parts: the *trade database* that summarizes the trading process, contains the date and time stamp ( $t_i$ ) for the  $i^{th}$  trade, with additional marks, such as transaction prices ( $tp_{t_i}$ ), volume ( $v_{t_i}$ ). The *quote database* is about the quoting process and reports the date and time ( $t_j$ ) occurrence of the  $j^{th}$  quote, along with the bid ( $b_j$ ) and ask ( $a_j$ ) prices, etc.

Because, all trades and quotes are not valid, we work only with regular ones, by deleting trades and quotes recorded outside the range of market opening (9:30 am -16:00 pm). We also screen trades by removing negative trades prices or volumes. Finally, quotes are also screened by deleting zero bid and ask prices. After merging both databases, we retain the following marked point process  $(t_i, a_i, b_i, v_i)$ , where  $b_i$  and  $a_i$  are the bid and ask prices prevailing when the  $i^{th}$  trade occurs, and define the prices process at the mid-point of the bid and ask prices.

Let us recall that our objective in this paper is to rely on ACD model for price durations in order to forecast price events conditional volatility and then ISIVaR. Thus, we compute price durations (as explained above) by using thresholds  $c = 1/8\$$  for IBM data (see Engle and Russell (1998)) and  $c =$

1/16\$ for EXXON. Table 3 reports price durations statistics. The average time needed for the price to have a significant change is about four minutes, with a minimum of two or three seconds for both stocks. As usually reported in empirical applications, price durations exhibit overdispersion *i.e.* standard deviation higher than mean.

#### 4.1 Seasonal Adjustment and EACD model

In this subsection, we calibrate an EACD model for price durations to compute the conditional volatility for price events returns. We thus divide the original sample into two parts, one for estimation and the last for *out-of-sample* ISI-VaR forecast. The estimation sample covers the period 02/03/07-03/06/07, which leads to a total of 1992 (resp. 1942) price durations for IBM (resp. EXXON). However, it is generally reported that durations exhibit significant diurnal patterns that must be first removed from raw durations before estimating ACD models. Indeed, in empirical applications, it is usually shown that trading activity is not constant over the course of the day and present a typical pattern, *i.e.* shorter durations at the beginning and close of the day, and longer durations in the middle of the day. This time-of-day component of durations is by nature almost perfectly predictable and constitutes the deterministic part of durations data. Many procedures have been proposed in the literature to estimate the seasonal component (see Pacurar (2006) for a review).

In this paper, we follow Bauwens and Giot (2000) by taking the deterministic component as the expected price duration conditioned on time-of-day, but also the day-of-week, where the expectation is computed by averaging the durations over seven non-overlapping intervals. These intervals are delimited by eight nodes set on each hour with an additional node in the last half hour of the trading day. Cubic splines are finally used to smooth the time-of-day

effect and then to extrapolate the latter for any time along the day. The Figure 1 displays the estimated seasonal components. The overall conclusion is that, it exists a day-of-week effect, since the time-of-day component of Monday is different from the time-of-day of Tuesday and so on. One can also notice the well-documented inverted-U shape (see for example, Engle and Russell (1998), Bauwens and Giot (2000), etc.). To remove the seasonal component, we divide the raw data of observed price durations by the time-of-day effect, and run an EACD(2,2) model to the stochastic component. The results are given in table 4, where all estimated coefficients are significant.

The performance of ACD models in capturing the latent structure of price durations can be assessed by looking at the residuals  $\hat{v}_i = x_i/\hat{\psi}_i$  where  $\hat{\psi}_i$  denoted the expected duration for the price event number  $i$  and given by equation (6). ACD models fit well data, if the series of residuals is a white noise, and this can be tested using Ljung-Box statistics. Here, we notice that the EACD(2,2) model successfully removes the autocorrelation structure in the original adjusted durations: for both stock, the residuals are not significantly autocorrelated at order 15 (see table 4).

Recall that the interest of using an ACD model applied to price durations is to infer price changes volatility from the estimated function of conditional intensity (see equations 15-16). The relation between intensity and volatility implies an inverse relationship between price durations and volatility. Since ACD models are well known to model with accuracy the clustering of durations, it can therefore also seize volatility clustering, and can be considered as an alternative to GARCH model when dealing with irregularly spacing data<sup>8</sup>. Figures 2 and 3 give the estimated conditional volatility respectively for IBM

---

<sup>8</sup> Note that another approach to model conditional volatility for irregularly spaced series of intraday returns is that of Ghysels and Jasiak (1997). They proposed a class of ARCH models for series sampled at unequal time intervals, by combining ACD models and results from the temporal aggregation for GARCH models discussed by Drost and Nijman (1993).

and EXXON stocks where volatility clustering is apparent.

#### 4.2 Backtesting ISIVaR Models

In this section, we generate *out-of-sample* ISIVaR using our algorithm and a fixed forecasting scheme. A fixed forecasting scheme consists in estimating the parameters only once with the estimation sample of size  $N_c$  and then using these estimates to produce all the forecasts for the *out-of-sample* period ranging from  $N_c + 1$  to  $N$ . We rely on the above estimation sample, *i.e.* the period from 02/03/97 to 03/06/97. We consider two *out-of-sample* sizes of different lengths, the first period from 03/07/97 to 03/31/97, and the second from 03/07/97 to 04/30/97. This leads to two different backtesting exercises. Figure 4 and 5 compare for the first period, the 1% *out-of-sample* ISIVaR and the corresponding returns of price events for both stocks. We can observe that the clustering of ISIVaR forecasts track quite well the evolution of price events returns. Besides, our ISIVaR also allows measuring liquidity risk through the expected conditional duration  $\hat{\psi}_i$  or equivalently through a Time at Risk (TaR) measure (equation 2). Figures 6 and 7 display the 1% TaR measures for IBM and EXXON stocks. Special care need to be exercised when interpreting these results since the reported values corresponds to the TaR based on the deseasonalized price event durations.

We now examine the statistical performance of our methodology to compute ISIVaR. We first compare observed events returns and *out-of-sample* ISIVaR forecasts to generate the hit-no-hit variable  $I_n$  and then apply our test statistics for validation purpose. Table 5 presents the results for both stocks, and for various shortfall probability where the values in brackets are p-values. Focusing on the unconditional version of our test statistics ( $J_{uc}$ ), the results show that our ISIVaR model performs well, meaning that the proportion of hits is not statistically different from  $\alpha$ , the shortfall probability. Indeed, the only

one exception is the 5% ISIVaR for EXXON stock over the second backtesting period. Concerning the conditional coverage test ( $J_{cc}$ ) the ISIVaR model performs well for both stocks at shortfall probability  $\alpha = 1\%$  or  $\alpha = 2.5\%$ . However, at 5%, the conditional coverage property is not reached, due to the violation of the independence assumption.

These shortcomings are close to the results of Dionne et al. (2006), even if their methodology relates rather to the estimation of Intraday-VaR (IVaR) at fixed-time horizon. Indeed, from their results, it comes out primarily that for high level of shortfall probability, IVaR model tends to be rejected when the fixed forecast horizon is very short (15 minutes in their applications). Since the average values of price events durations are around 4 minutes for both stocks in the present paper, the forecast horizon is then short and our results converge towards this observation that remains however empirical.

To conclude, the semi-parametric forecasting method proposed to compute ISIVaR gives quite satisfactory results. The few cases where our model fails to fit with accuracy the latent level of market events risk could be attributed to the estimation of parameter  $q$ , the quantile of the standardized series of returns  $\widehat{\varepsilon}_{t_i}$ . Indeed, the roughly Historical Simulation (HS) used to compute the quantile suffers from its logical drawbacks largely studied by Boudoukh et al. (1998) and Pritsker (2001). Firstly, the assumption that the standardized residuals are *i.i.d.* is still required, and the task is actually more complicated by the irregularly spacing nature of residuals and the discreteness of tick data. Needless to say that an appropriate method for the estimation of quantiles for irregularly spacing data will constitute an improvement over our ISIVaR forecasting.

## 5 Conclusion

Risk modelling and evaluation has emerged over the last several years as a key component in the management of financial institutions. The official horizon for assessing market risk models and to determine regulatory capital requirements is 10 days, as laid out by the Basel Committee on Banking Supervision (BCBS) in 1996. For internal purpose however, banks routinely compute 1-day VaR using daily prices. However, with the recent evolutions in the organization of trading process in financial markets, high frequency or tick by tick data are available and forecasting risk at very high frequency is nowadays possible.

In this paper, we have introduced a general setup for computing Intraday-VaR by explicitly making use of the irregularly spacing nature of high frequency data. Instead of predicting VaR at fixed-time horizon (10 minutes, 15 minutes, etc.), our methodology is attractive in the sense that the forecast horizon is stochastic and is related to the trading intensity. Risk is forecast in events time, such that the traditional dimension of risk, usually summarized by only the level of VaR is coupled with the expected market events durations, giving a framework for a real-world monitoring of risk exposure. Indeed, with our methodology, we provide two simultaneous information, meaning the expected duration for the price to have a significant change and the corresponding level of risk.

Technically, our model we named Irregularly Spaced Intraday VaR (ISIVaR) makes use of the relation between instantaneous volatility and price change durations intensity to compute volatility for price events. Empirical quantile of the standardized residuals as then multiply by the square root of the forecast volatility to derive semi-parametric VaR for the next price change. In this line, our methodology can be viewed as an improvement over Giot (2005) Log-ACD intraday-VaR, where normality is assumed, throwing out fat tailedness of high frequency data. Needless to say that active market participants such as traders

and financial institutions like banks involving with frequent margins setting are natural recipients of the proposed method.

We also proposed a test for the predictive abilities of such unequally spaced VaR models, by first noting that the traditional conditional coverage criterion remains valid even with irregularly spaced VaR model, but with a hit-no-hit variable defined in event time. This last assumption is tested using the GMM distributional assumption testing of Bontemps and Meddahi (2006). Through monte carlo replications, we show that the proposed test has reasonable properties at finite distance. Applications to IBM and EXXON stocks traded at the NYSE reveal that the ISIVaR model are volatile and track well the evolution of price change returns. *Out-of-sample* evaluations are also conducted and give satisfactory results.

Let us finish by indicating that the suggested method to compute ISIVaR is not the only one, and one can for example rely on the ACD-GARCH model of Ghysels and Jasiak (1997) to compute volatility. Relative performance of both models can be assessed in VaR framework, by using for example the model-free quantiles comparison test of Giacomini and Komunjer (2005). Beyond both methods, it is clear that the challenge when one wants to forecast intraday VaR (in price events time) is to develop a method for the estimation of quantiles for data sampled at unequal time intervals. We are exploring this general issue as a direction for further research.

## A References

- BAUWENS, L. AND P. GIOT (2000), "The Logarithmic ACD Model: an Application to the Bid-Ask Quote Process of Three NYSE Stocks", *Annales d'Économie et de Statistique*, 60, pp. 117-149.
- BERKOWITZ, J. (2001), "Testing Density Forecasts With Applications to Risk Management", *Journal of Business and Economic Statistics*, 19, pp. 465-474.
- BERKOWITZ, J. AND J. O'BRIEN (2002), "How Accurate are the Value-at-Risk Models at Commercial Banks", *Journal of Finance*, 57, pp. 1093-1111.
- BONTEMPS, C. AND N. MEDDAHI (2005), "Testing Normality: A GMM Approach", *Journal of Econometrics*, 124, pp. 149-186.
- BONTEMPS, C. AND N. MEDDAHI (2006), "Testing Distributional Assumptions: A GMM Approach", Working Paper.
- BOUDOUGH, J., RICHARDSON M. AND R. WHITELAW (1998), "The Best of Both Worlds," *Risk* 11 (May), 64-67.
- CHRISTOFFERSEN, P. F. (1998), "Evaluating Interval Forecasts", *International Economic Review*, 39, pp. 841-862.
- CHRISTOFFERSEN, P. F. AND D. PELLETIER (2004), "Backtesting Value-at-Risk: A Duration-Based Approach", *Journal of Financial Econometrics*, 2, 1, pp. 84-108.
- DIONNE, G., DUCHESNE P. AND M. PACURAR (2006), "Intraday Value at Risk (IVaR) Using Tick-by-Tick Data with Application to the Toronto Stock Exchange", Working Paper.
- EASLEY, D. AND M. O'HARA (1992), "Time and the Process of Security Price Adjustment", *Journal of Finance*, 47, pp. 577-606.
- ENGLE, R. F. (2000), "The Econometrics of Ultra-High Frequency Data", *Econometrica*, 68, pp. 1-22.
- ENGLE, R. F. AND S. MANGANELLI (2001), "Value at Risk Models in Finance", *European Central Bank Working Paper* 75.
- ENGLE, R. F. AND S. MANGANELLI (2004), "CAViaR: Conditional Autoregressive Value-at-Risk by Regression Quantiles", *Journal of Business and Economic Statistics*, 22, pp. 367-381.
- ENGLE, R. F., AND J. R. RUSSELL (1998), "Autoregressive Conditional Du-

ration: a New Model for Irregularly Spaced Transaction Data", *Econometrica*, 66, pp. 1127-1162.

GHYSELS, E. AND J. JASIAK (1997), "GARCH for irregularly spaced financial data: the ACD-GARCH model", *Studies in Nonlinear Dynamics & Econometrics*, 2, 4.

GHYSELS, E., GOURIEROUX C. AND J. JASIAK (2004), "Stochastic Volatility Duration Models", *Journal of Econometrics*, 119, pp. 413-433.

GIACOMINI, R. AND I. KOMUNJER (2005), "Evaluation and Combination of Conditional Quantile Forecasts", *Journal of Business & Economic Statistics*, 23, 416-431.

GIOT, P. (2005), "Market risk models for intraday data", *European Journal of Finance*, 2005, pp. 309-324.

GRAMMIG, J. AND K. O. MAURER (2000), "Non-Monotonic Hazard Functions and the Autoregressive Conditional Duration Model", *Econometrics Journal*, 3, pp. 16-38.

LE FOL, G. AND L. MERCIER (1998), "Time deformation: definition and comparisons", *Journal of Computational Intelligence in Finance*, September/October, pp. 19-33.

LUNDE, A. (1999), "A Generalized Gamma Autoregressive Conditional Duration Model", Working Paper, Aalborg University.

NELSON, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, 59, pp. 347-370.

PACURAR, M. (2006), "Autoregressive Conditional Duration (ACD) Models in Finance: A Survey of the Theoretical and Empirical Literature", Working Paper, Dalhousie University.

PRITSKER, M. (2001), "The Hidden Dangers of Historical Simulation", No 2001-27, Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System (U.S.)

Table 1. Empirical Size for  $J$  Statistics

Backtesting $\alpha = 1\%$ ISIVaR					
Sample Size	$J_{uc}$	$J_{cc}$			
	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
$N = 1000$	0.0834	0.0358	0.0310	0.0276	0.0289
$N = 1500$	0.0945	0.0440	0.0344	0.0316	0.0345
$N = 2000$	0.0990	0.0427	0.0346	0.0394	0.0340
$N = 2500$	0.1006	0.0476	0.0373	0.0379	0.0342
$N = 3000$	0.1083	0.0495	0.0412	0.0383	0.0349
$N = 4000$	0.1182	0.0506	0.0395	0.0383	0.0350
$N = 5000$	0.1167	0.0507	0.0428	0.0404	0.0397
Backtesting $\alpha = 5\%$ ISIVaR					
Sample Size	$J_{uc}$	$J_{cc}$			
	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
$N = 1000$	0.1004	0.0421	0.0357	0.0327	0.0290
$N = 1500$	0.0998	0.0399	0.0362	0.0324	0.0280
$N = 2000$	0.1098	0.0395	0.0376	0.0359	0.0329
$N = 2500$	0.1206	0.0354	0.0394	0.0343	0.0328
$N = 3000$	0.1210	0.0410	0.0404	0.0372	0.0351
$N = 4000$	0.1370	0.0390	0.0446	0.0415	0.0307
$N = 5000$	0.1430	0.0406	0.0469	0.0456	0.0370

Notes:  $p$  denotes the number of Laguerre polynomials used.  $J_{uc}$  (for  $p = 1$ ) denotes the  $J$  test of the unconditionnal coverage null hypothesis and  $J_{cc}$  denotes  $J$  test of the conditionnal coverage null hypothesis. For each experiment, the  $N$  hit-no-hit variables are simulated under the null according to  $N$  bernouilli trials with a rate parameter equal to the coverage rate of VaR (1% or 5%). The frequencies of rejections of  $J$  tests are reported for 10 000 replications and correspond to empirical sizes. The nominal size of test is set at 5%.

Table 2: Empirical Power for  $J$  Statistics

Backtesting $\alpha = 1\%$ ISIVaR					
Sample Size	$J_{uc}$	$J_{cc}$			
	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
$N = 1000$	0.9286	0.8978	0.8886	0.8781	0.8643
$N = 1500$	0.9448	0.9254	0.9112	0.9069	0.8941
$N = 2000$	0.9537	0.9352	0.9240	0.9189	0.9128
$N = 2500$	0.9578	0.9408	0.9342	0.9278	0.9235
$N = 3000$	0.9596	0.9506	0.9416	0.9366	0.9277
$N = 4000$	0.9681	0.9589	0.9509	0.9407	0.9412
$N = 5000$	0.9723	0.9617	0.9574	0.9513	0.9444
Backtesting $\alpha = 5\%$ ISIVaR					
Sample Size	$J_{uc}$	$J_{cc}$			
	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
$N = 1000$	0.8553	0.8042	0.7786	0.7518	0.7410
$N = 1500$	0.8883	0.8419	0.8263	0.8055	0.7962
$N = 2000$	0.9031	0.8683	0.8440	0.8352	0.8265
$N = 2500$	0.9083	0.8843	0.8733	0.8580	0.8475
$N = 3000$	0.9245	0.8936	0.8831	0.8657	0.8600
$N = 4000$	0.9339	0.9145	0.8987	0.8970	0.8856
$N = 5000$	0.9453	0.9239	0.9134	0.9066	0.8967

Notes:  $p$  denotes the number of Laguerre polynomials used.  $J_{uc}$  (for  $p = 1$ ) denotes the  $J$  test of the unconditional coverage null hypothesis and  $J_{cc}$  denotes  $J$  test of the conditionnal coverage null hypothesis. For each replication, we simulate  $N$  bernoulli trials with rate parameter  $\alpha + \kappa$ , where  $\kappa \in [0, 0.1]$ . We then are in a situation, where an ISIVaR model leads to an excessive number of hits, violating the unconditional coverage property. For each couple  $(\alpha, N)$ , we obtain variable  $C$  and apply our test for a nominal size set at 5%. The frequencies of rejections are reported for 10 000 replications and correspond to powers.

Table 3: Price Durations Statistics<sup>9</sup>

	IBM stock	EXXON stock
Number of bid-ask quotes	5638	5125
Mean	251.25	275.81
Minimum	3	2
Maximum	4471	5778
Standard deviation	328.34	366.13

Notes: Price durations for IBM and EXXON stocks obtained by filtering original bid and ask quotes (corresponding to a trade) using thresholds  $c = \frac{1}{8}\$$  for IBM and  $\frac{1}{16}\$$  for EXXON. The period is February-April 1997.

Table 4: Results of EACD(2,2) Model<sup>10</sup>

	IBM		EXXON	
	Estimates	t-Statistics	Estimates	t-Statistics
$w$	0.0375	1.4646	0.1820	3.4251
$\alpha_1$	0.2495	5.9476	0.1065	4.9797
$\alpha_2$	-0.1832	-3.6270	0.1164	5.6246
$\beta_1$	1.1544	6.1368	-0.1457	-3.1663
$\beta_2$	-0.2569	-1.8517	0.7431	15.893
$Q(20)$	340.59		176.09	
$Q_v(20)$	12.103		7.4272	

Notes: The estimation sample covers the period 02/03/07-03/06/07, with a total of 1992 price durations for IBM and 1942 for EXXON.  $Q(15)$  is the Ljung-Box Q-statistics associated to the seasonally adjusted price durations.  $Q_v(15)$  is the same statistics for the fitted series of residuals from the EACD(2,2) model.

Table 5: Backtesting Results

IBM						
	Period 1			Period 2		
	%Hits	$J_{uc}$	$J_{cc}$	%Hits	$J_{uc}$	$J_{cc}$
$\alpha = 1\%$	0.97	0.011	0.199	1.21	1.624	2.784
		(0.916)	(0.905)		(0.202)	(0.248)
$\alpha = 2.5\%$	1.93	2.150	4.363	2.08	2.591	14.731
		(0.142)	(0.112)		(0.107)	(<0.001)
$\alpha = 5\%$	4.34	1.329	12.631	4.53	1.590	33.683
		(0.249)	(0.002)		(0.207)	(<0.001)
EXXON						
	Period 1			Period 2		
	%Hits	$J_{uc}$	$J_{cc}$	%Hits	$J_{uc}$	$J_{cc}$
$\alpha = 1\%$	0.72	0.885	3.921	0.69	3.382	3.462
		(0.346)	(0.140)		(0.065)	(0.177)
$\alpha = 2.5\%$	2.10	0.803	2.281	2.42	0.030	3.625
		(0.370)	(0.319)		(0.860)	(0.163)
$\alpha = 5\%$	4.98	0.006	0.800	6.22	8.000	14.628
		(0.936)	(0.670)		(0.004)	(<0.001)

Notes: For each period, we backtest ISIVaR model using various shortfall probability  $\alpha$  and the two versions of our test statistics.  $J_{uc}$  (for  $p = 1$ ) denotes the  $J$  test of the unconditional coverage null hypothesis and  $J_{cc}$  denotes  $J$  test of the conditionnal coverage null hypothesis. %Hits is the proportion of hits. The length of the first period is respectively 1658 and 1527 for IBM and EXXON (respectively 3646 and 3183 for the second period).

Figure 1: Estimated Seasonal Time-of-Day Deterministic Components.

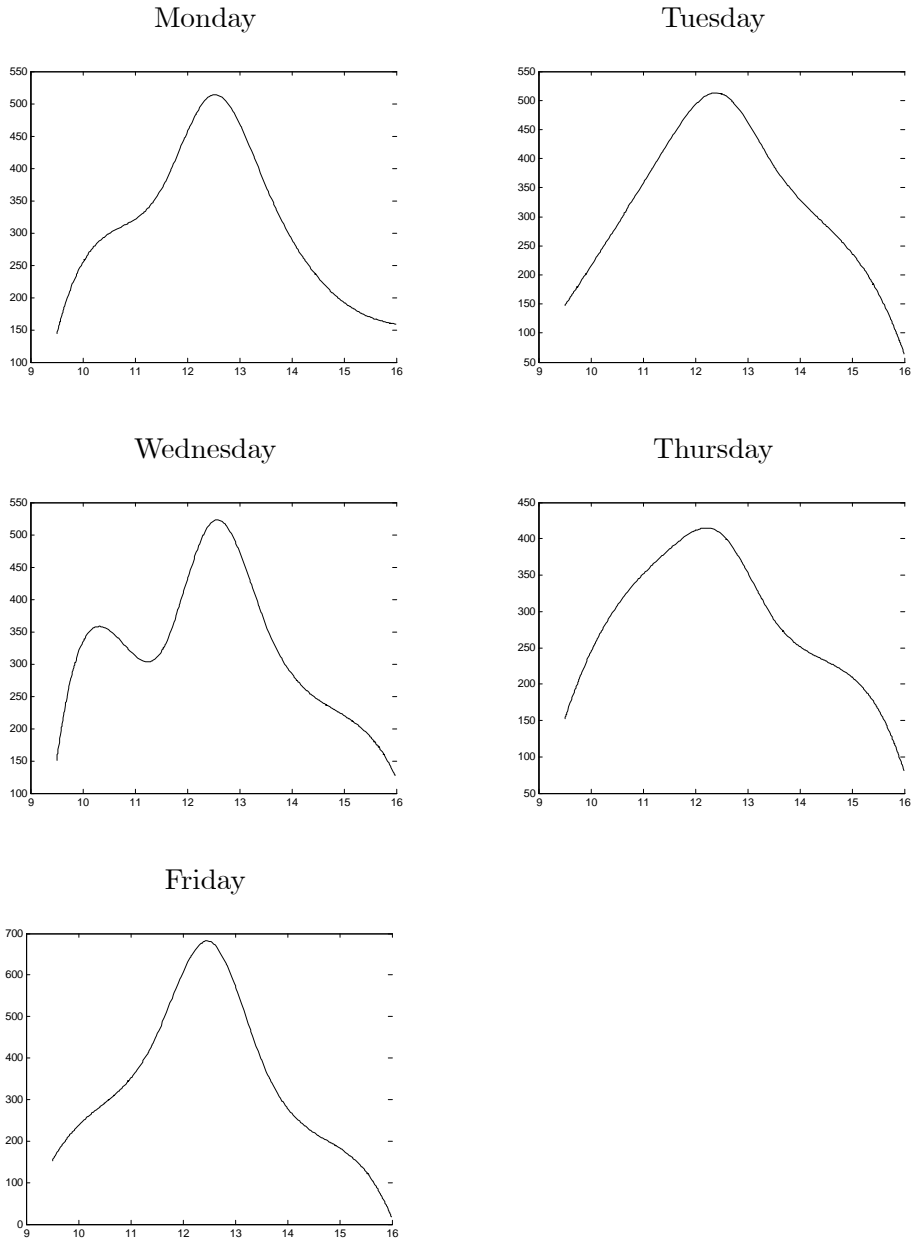


Figure 2. Conditional Volatility for Price Events (IBM)

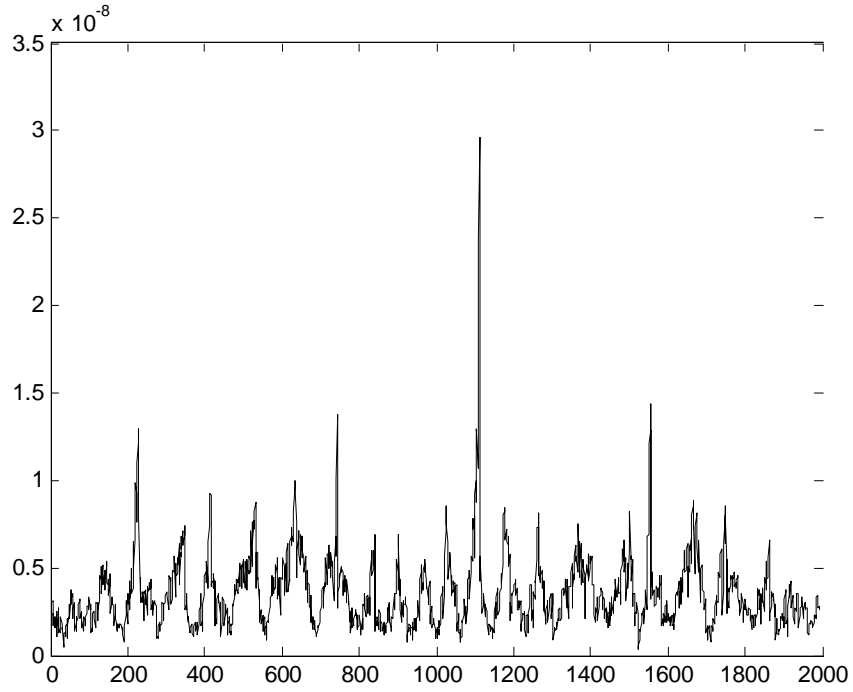


Figure 3. Conditional Volatility for Price Events (EXXON)

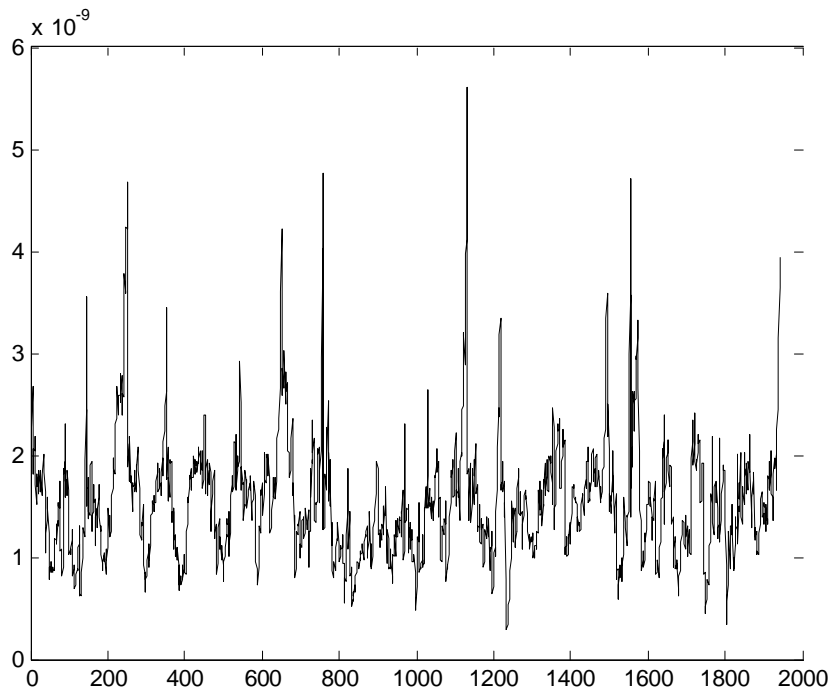


Figure 4. Out-of-sample forecasts of 1% ISIVaR and observed returns (IBM)

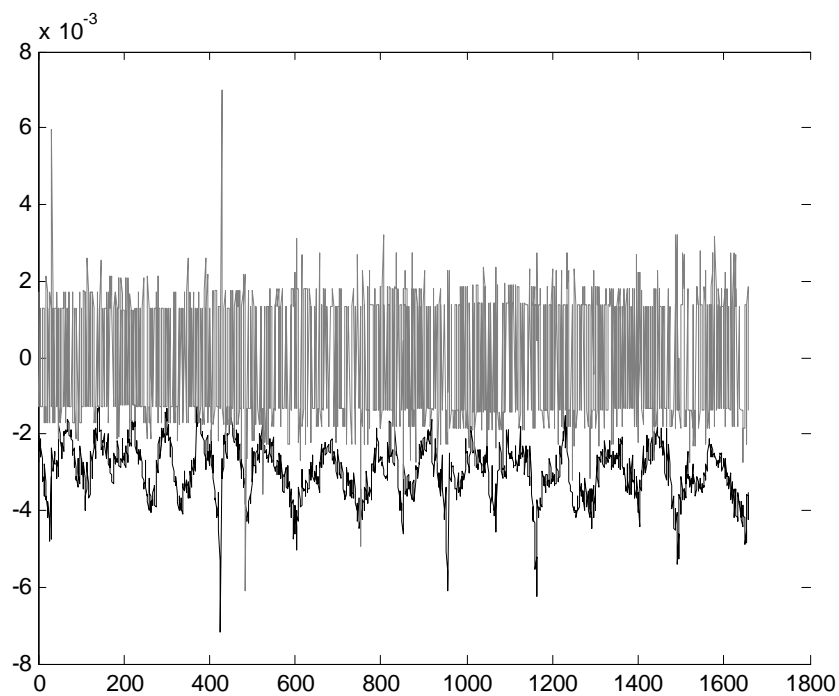


Figure 5. Out-of-sample forecasts of 1% ISIVaR and observed returns (EXXON)

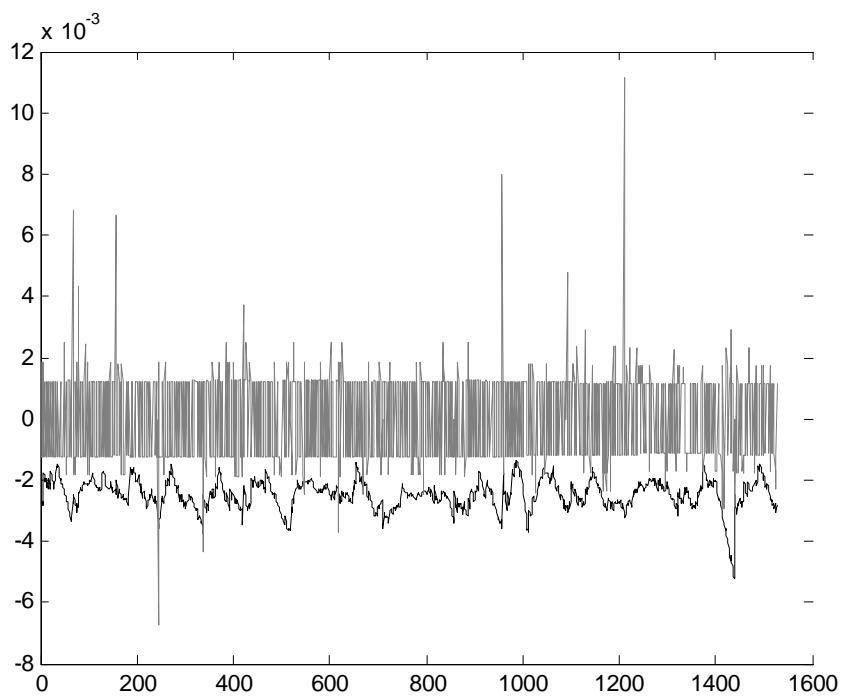


Figure 6. Out-of-sample forecasts 1% Time at Risk(IBM)

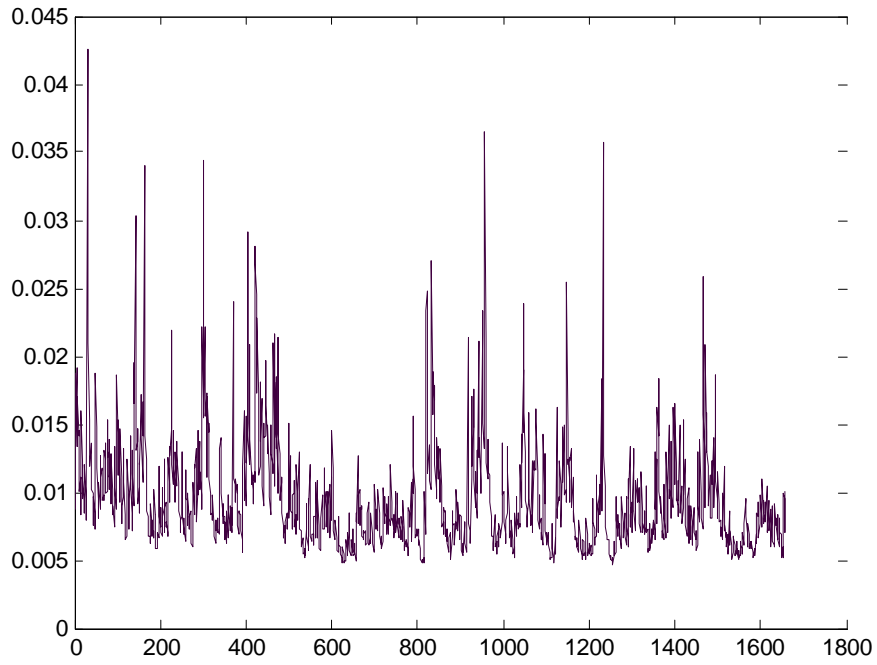


Figure 7. Out-of-sample forecasts 1% Time at Risk (EXXON)

