Intradaily seasonality of returns distribution
A Quantile Regression approach and Intradaily VaR

Laura Coroneo¹ and David Veredas²

First draft September 2006. This version April 2007.

Abstract

We investigate intradaily seasonal patterns of the distribution of high frequency financial returns. Using quantile regression we show the expansions and shrinks of the probability law through the day for three years of 15 minutes sampled stock returns. Returns are more dispersed and less concentrated around the median at the hours near the opening and closing. We provide intradaily value at risk assessments and we show how it adapts to changes of dispersion over the day. The tests performed on the out-of-sample forecasts of the value at risk show that the model is able to provide good risk assessments.

Keywords: High frequency returns, Quantile Regression, Fourier series, Intradaily VaR.

JEL classification: C14, C22, C53, G10.

¹ECARES, Université Libre de Bruxelles and Department of Economics, University of Bologna; email: laura.coroneo@ulb.ac.be.
²ECARES, Université Libre de Bruxelles and CORE, Université catholique de Louvain; email: david.veredas@ulb.ac.be.
Corresponding address: David Veredas, ECARES, Université Libre de Bruxelles. 50 Av F.D. Roosevelt CP 114, B1050 Brussels, Belgium. Phone: +32 2 6504218. Fax: +31 2 6504475.

We are grateful to Pierre Giot, Peter R. Hansen, Alexander McNeil, Roger Koenker, Sergio Pastorello, Simone Manganelli, Maria Pacurar and Roberto Pascual and the participants at the ECARES seminar and the Zeuthen workshop for insightful remarks. Any remaining errors and inaccuracies are ours.
We investigate intradaily seasonal patterns of the probability law of high frequency financial returns. The interest of this research rests on two facts. First, intradaily data has become a major pole of interest for researchers and financial agents that practice intradaily trading. For instance, high frequency hedge fund managers open and close positions within the day. For these managers intradaily risk evaluation is an important tool to follow the market and to build optimal intradaily trading strategies. However, within the day there are significant variations in asset prices, which imply different evaluations of the tails of the return’s distribution through the day. And these variations are partly deterministic and due to the intradaily seasonality.

Second, the analysis of risk is intimately related with the analysis of probabilities and, therefore, the analysis of the conditional probability distribution. Asset returns are realizations of a random variable and their behavior is fully described by their conditional probability law. Any function, such as the density function, describing this law conveys information about the likelihood that the next realization will take a certain value. But within the day these odds depend partly on a deterministic seasonal component that makes the probability density function to expand or shrink as a function of the time of the day. This effect is illustrated in the kernel densities for returns at different hours of the day shown in Figure 1. Data are 15 minutes sampled returns for three stocks (large, medium and small caps) traded at the Spanish stock exchange. The kernel density estimates for returns at different times of the day vary significantly. Around lunch the density is more peaked and the tails are thinner while it is more dispersed at the hours near the opening and closing. Intradaily value at risk evaluations therefore depend very much on the time of the day. If an intradaily trader does not take this seasonality into account in her risk estimations, she will underestimate the expected loss at the opening and closing and overestimate it at noon.

Moreover, not only the volatility presents seasonal movements, but also the skewness and kurtosis. Table 1 shows some descriptive statistics for the data grouped according to the hour of the day. We reported the sample mean, the sample standard deviation, the skewness and the kurtosis indices for four different hours. These estimates are proxies of the intradaily behavior of the probability law. There is no evidence of an intradaily seasonal pattern in the sample mean of returns. However, there is a very clear U-shaped pattern in the sample standard deviation, as found in many former studies. In addition to this, the last two columns suggest the presence of an intradaily seasonality in the skewness and kurtosis indices. In particular, while the movements in the skewness are small in magnitude, for the kurtosis index we have large variations during the day. For all the stocks, there is a significant increase in the thickness of the tails just after 15:00, that is after the opening of the New York Stock Exchange.

1All over the analysis we use standardized returns for comparison purposes. Otherwise the scale of the plots depends on the price, perturbing the interpretation.

2The table displays the bias adjusted skewness index computed as \( \frac{n(n-1)}{n-2} m_3/s^3 \) where \( n \) is the number of days (we have one observation for each day), \( m_3 \) is the third sample central moment and \( s_2 \) is the sample standard deviation. And the bias adjusted kurtosis computed as \( \frac{n+1}{(n-1)(n-2)(n-3)} \left( (n+1) \frac{m_4}{s^4} - 3(n-1) \right) \) where \( m_4 \) is the fourth sample central moment.
The standard approach to analyze the conditional distribution function of intradaily asset returns is to fit a model for the second moment as a function of two components. One for the dynamics and another for the seasonality. If returns are Gaussian, the second moment provides information enough to describe the conditional probability law, as all the odd moments are zero and the even moments are functions of the second moment. This property of the Gaussian distribution is very appealing but, at the same time, this distribution is not able to reproduce the tail behavior present in the data. This is one of the reasons for which it is now commonly accepted that asset returns are not normally distributed. More flexible distributions, such as the Student’s t distribution, are needed. However, the drawback of these laws is that moments beyond the second are either zero – e.g. the third moment – or functions of an invariant tail index – e.g. the fourth moment. For instance, a GARCH model with a Student’s t distribution has constant kurtosis given by a function of the estimated degrees of freedom, which is not consistent with the data features. A possible solution to overcome this problem would be to fit models for different moments, similarly to Hansen (1994) or Harvey and Siddique (1999) among the others, but it is not clear the functional forms that these models should take and/or which regressors to use.

Since our interest is the analysis of the seasonality of the conditional distribution, a natural alternative is to model directly the conditional probability. Among all the functions that characterize the conditional probability (density, cumulative, characteristic, Laplace, hazard, etc), the conditional quantiles are the better suited due to the existence of quantile regression, introduced by the seminal work of Koenker and Basset (1978). Indeed quantile regression has a number of useful features. First, quantile regression is one of the possible ways to characterize the conditional probability law and, since there is a one to one relation with all the other possible characterizations, we can, indirectly, analyze the effect of the time of the day on the density function of asset returns. Second, quantile regression does not assume the existence of any moment. In fact, it does not assume anything about the moments. Often it happens that the tails of returns are so thick that some important moments do not exist. For instance, Table 2 shows the estimated parameters of a GARCH(1,1) with Student’s t distribution. The estimated degrees of freedom for the three stocks are very low. So low that, according to the model, kurtosis does not exist for any of them and variance does not even exist for one of them.

Third, quantile regression is robust in the sense that the estimated coefficients are not sensitive to outliers on the dependent variable. This is particularly useful in the analysis of high frequency financial returns since often we do find outliers or, at least, observations that are remarkably different to the rest of the process. For instance, Figure 2 shows the actual returns for the three stocks we analyze. For all there is at least one observation that is unusually high. Fourth, quantile regression is a distribution free model. This is a very compelling feature. It does not rely on any distribution specification but, ironically, it is an estimate of the conditional probability distribution. As noted earlier, and shown in Table 2, assuming a parametric distribution for intradaily asset returns entails a series of problems that sometimes, e.g. infinite variance, are difficult to overcome.

The use of quantile regression in asset returns is not new. One of the first to use it are Engle and Manganelli (2004) who introduce the CAViaR (Conditional Autoregressive Value at Risk). CAViaR extends the traditional linear quantile regression to a nonlinear framework and
develop a new test of model adequacy, the Dynamic Quantile (DQ) test, using the criterion that each period the probability of exceeding the VaR must be independent of all the past information. Gourieroux and Jasiak (2005) introduce a new dynamic quantile model univariate series and panel data as well as the Quantile Factor Model. Less related, Bouye and Salmon (2003) use quantile regression in a copula context, that is they deduce the form of the non linear conditional quantile regression implied by the copula. As for intraday VaR, Giot (2005) quantify intraday VaR (15 and 30 minute returns) using normal GARCH, Student GARCH, RiskMetrics and Log-ACD models. He shows that Student GARCH model performs best. Last, Dionnea et al. (2005) investigate the use of tick-by-tick data for market risk measurement and propose an intraday Value at Risk at different horizons based on irregularly time-spaced high-frequency data by using an intraday Monte Carlo simulation.

Using quote midpoints of three stocks traded at the Spanish stock exchange from January 2001 to December 2003, we show that indeed the conditional probability distribution depends on the time of the day. At the opening and closing the density flattens and the tails become thicker, while in the middle of the day returns concentrate around the median and the tails are thinner. Results are intuitive, in the sense that they confirm the general perception that in the opening and closing the probabilities of finding large price fluctuations are higher than at lunch. Results, in terms of quantiles, permit straightforward intraday risk evaluations, such as value at risk. We show the intraday variation of the maximum expected loss at 2.5%, 1% and 0.5% confidence levels. The maximum expected loss is maximal at the opening and closing and minimal at lunch time. Failure rates tests, based on Kupiec (1995) and Christoersen (1998) confirm that the model is able to provide good forecasts of the maximum expected loss.

The structure of the paper is as follows. Section 2 introduces the data and the market structure. Section 3 briefs quantile regression, the model that is used for estimation and how to interpret results in term of density functions. Section 4 shows the estimation results for the proposed model. Section 5 contains intraday value at risk forecast and evaluation. Section 6 concludes.

2 Market and Data

Data come from the Spanish Stock Exchange (SSE), the 9th world largest stock exchange in terms of capitalization (the 3th among continental European markets), and the 7th in terms of total value of share trading (the 3th in continental Europe) according to the World Federation of Exchanges. The Spanish stock exchange interconnection system is the electronic platform that connects, since 1995, the four exchanges that compose the SSE (Barcelona, Bilbao, Madrid, and Valencia). This system holds all the Spanish stocks that achieve pre-determined minimum levels of trading frequency and liquidity. Every order submitted to the system is electronically routed to a centralized limit order book (LOB) to proceed with its immediate execution or storage. The matching of orders is, therefore, computerized. The LOB on the brokers’ screens is updated each time there is a cancelation, execution, modification or new submission. The SSE is organized as an order-driven market with a daily continuous trading session from 9:00 a.m. to 5:30 p.m. and two call auctions that determine the opening and closing prices.

During the continuous trading session, a trade takes place if an only if an order hits the quotes. Pre-arranged trades are not allowed during the continuous session, and price-improvements are impossible. There are no market makers and there is no floor trading. The market is governed by a strict price-time priority rule, but an order may lose priority if modified. Stocks are quoted in euros. The minimum price variation (tick) equals 0.01 for prices
below 50 and 0.05 for prices above 50. The minimum trade size is one share. There are three basic types of orders: market, limit, and market-to-limit. Market orders are executed against the best prices on the opposite side of the book. Any excess that cannot be executed at the best bid or ask quote is executed at less favorable prices by walking down (up) the book until the order is fulfilled. Market-to-limit orders do not specify a limit price but are limited to the best opposite-side price on the book at the time of entry. Any excess that cannot be executed is converted into a limit order at that price. Finally, limit orders are to be executed at the limit price or better. Any unexecuted part of the order is stored in front of the book at the limit price. By default, orders expire at the end of the session.

The official market index of the SSE is the IBEX-35, which includes the 35 most liquid and active stocks of the exchange, weighted by market capitalization. Its composition is regularly revised every semester. Our initial sample is formed by the 35 index constituents from January 2001 to December 2003. The data used in this study consists of 15 minutes sampled quote midpoints during 3 years, from January 2001 to December 2003, of the 35 companies listed in the IBEX-35. For each stock we have 34 intradaily observations for a total of 25,430 observations. For simplicity, we will report the analysis only on 3 of the 35 companies but the results are valid for all of them and they are available upon request. Among the 35 companies of the IBEX-35, we report the results for Telefonica (TEF), Endesa (ELE) and Aciona (ANA) that are, respectively, a big, medium and small company, weighting, approximately, 20%, 6% and 0.8% in the index.

3 Quantile Regression as Density Regression

The probability law of a random variable $r_t$ can be characterized by means of different functions. Some, like the density or the cumulative functions, are common. Others, like the quantile function, the hazard function or the characteristic function are less used. Yet, any can be written as a function of the others and hence the knowledge of one implies the knowledge of the others. The quantile function is particularly compelling in the context of conditional distributions. This is due to the existence of a solid theory on quantile regression (see Koenker, 2005, for a survey). Let $Q_{r_t}(\tau)$ be the $\tau$-th quantile of $r_t$. It is well known that

$$f(r_t) = \frac{\partial}{\partial r_t} F(r_t) \quad \text{and} \quad Q_{r_t}(\tau) = F^{-1}(\tau) = \inf\{r_t : F(r_t) \geq \tau\}, \quad \tau \in (0,1),$$

where $f(r_t)$ is the probability density function, pdf hereafter, and $F(r_t)$ is the cumulative distribution function, cdf hereafter. We can easily pass from the pdf to the quantile function. Top row of Figure 3 shows this idea. The density is symmetric around the mean, which implies that the quantiles are also symmetric around the median (that equals the mean). The density is centered at zero, and hence the quantile function at the median, $Q_{r_t}(0.5)$, is zero.

One may question what happens with the pdf and the quantile function if there is a location-scale shift. Second to fourth rows of Figure 3 illustrate these cases. The second row shows a positive location shift in the density, which produces a parallel upward shift of the quantile function. Or, inversely, if the quantile function shifts, the density shifts the location. It is worth noticing that after the shift the quantile function at the median, $Q_{r_t}(0.5)$, is not zero anymore as the mean in the pdf is not zero anymore either.

[FIGURE 3 ABOUT HERE]
The third row shows the effect of a positive scale shift in the density. This shift produces an expansion of the quantile function, or, inversely, an expansion of the quantile function implies a positive scale shift in the density. The expansion in the quantile function implies an increase in the dispersion of the quantiles. This happens when we compare the probability law at, for instance, lunch and the closing, as already noted in Figure 1. By contrast, the dispersion of the observations decreases if we compare the probability law the opening and at lunch which means a contraction of the quantiles. Finally, last row illustrates a positive location shift and a scale shift in the density, implying an asymmetric shift—a mix of shift and expansion effects in the quantile function. More complex shifts are possible. For instance a one-sided expansion in the quantile implies an increase of the dispersion in only one side of the density, creating skewness. Fat tails can be also created in the density if the quantile are stretched only at the highest and lowest values, say 1% and/or 99% quantiles. In sum, either a location shift or a scale shift, or both, in the pdf has a clear representation in terms of quantiles, as both functions contain the same information about the random variable of interest.

The understanding of the effect of these shifts and how the quantile and the density function are affected by them is important in a conditional context. In fact, the movements in the densities of Figure 1 are produced by the intraday seasonality. It is therefore meaningful to model how the probability distribution evolves conditional to the time of the day. Quantile regression (QR henceforth), introduced by Koenker and Bassett (1978), is the appropriate tool. The problem of finding the $\tau$-th unconditional quantile can be expressed as the solution of a simple linear optimization problem. Generalizing these results to the case in which the quantiles are linear functions of some explanatory variables leads to the QR method. The fundamental difference of QR with respect to mean regression is that the latter considers the effect of the regressor on the mean of the regressand while QR considers the effect of the regressor on the specific $\tau$-th quantile of the regressand. Hence, for a sufficiently narrow grid of $\tau$, the QR method can fully describe the quantile function. The basic QR model is

$$Q_{\tau_t}(x_t) = \omega(\tau) + \sum_{j=1}^{J} \beta_j(\tau)x_{jt}, \quad \tau \in (0,1),$$

where the intercept $\omega(\tau)$ and the slope parameters $\beta_j(\tau)$ are functions of $\tau$. While in the mean regression model there is a unique parameter $\beta_j$ that describes the effect that $x_{jt}$ has on the conditional mean of $r_t$, in QR for each $\tau \in (0,1)$ there is a parameter $\beta_j(\tau)$ that describes the effect of $x_{jt}$ on the $\tau$-th conditional quantile of $r_t$. In other words, QR measures the effect of the regressors on each quantile of the conditional distribution of the dependent variable. In this way it allows to analyze how a shock in the regressors affects the different quantiles and hence the pdf of returns.

The set of regressors $x_{jt}$ is divided in two parts. One accounts for the intraday seasonality, the main object of interest, and the second controls for the dynamics. As for the seasonality, we model it using a Fourier series of order 3:

$$seas_d(\tau) = \sum_{j=1}^{3} \alpha_j(\tau) \cos \left(2\pi j \frac{d}{34}\right) + \gamma_j(\tau) \sin \left(2\pi j \frac{d}{34}\right),$$

where 34 is the number of intraday time intervals (for the 15 minutes sampled returns) and $d$ denotes the time of the day in ordinal sense (i.e. the sequence 1, 2, ..., 34). Fourier series are convenient expressions for seasonality as the combination of cosines and sines is flexible.

---

3These type of shifts are of particular relevance for this article.
4We tried higher orders of the Fourier series but results do not change substantially.
enough to capture virtually any seasonal pattern. The cosine component of the first Fourier series reaches the maximum at the opening and at the closing, the hours of the day in which the dispersion is higher, and has the minimum at lunch time, the time of the day in which the dispersion is minimal. We may therefore expect this cosine term to capture most of the seasonal pattern.

To control for the dynamics, we follow Koenker (2006) choosing one lag of the absolute value of returns: \( \beta(t)|r_{t-1}| \). More lags or other functions of \( r_t \) to capture the dynamics as, for instance, square returns are possible. However, in a robust setting, the choice of absolute values is more sensible.\(^5\) Putting all the elements together the model we estimate is

\[
Q_{r_t}(\tau|d, |r_{t-1}|) = \omega(\tau) + \beta(\tau)|r_{t-1}| + \sum_{j=1}^{3} \alpha_j(\tau) \cos \left( 2\pi \frac{j \, d}{D} \right) + \gamma_j(\tau) \sin \left( 2\pi \frac{j \, d}{D} \right). \tag{2}
\]

Estimation has been implemented in GAUSS using a modified version of the library Qreg.\(^6\) Parameters are estimated using the interior point method, as described by Portnoy and Koenker (1997). The chosen grid of quantiles is (0.05, 0.10, ..., 0.95) and the limiting covariance matrix has been computed in GAUSS using the procedure described in Appendix A.

4 Estimation Results

Parameters in equation (2) depend on the quantile considered, \( \tau \). There are as many parameters as quantiles times the number of explanatory variables plus those in the intercept \( \omega(\tau) \). Because this number may become large, in our case is 168 per stock, we follow the literature, see for instance Koenker (2006), and we present all the results graphically. This presentation nicely dovetails with Figure 3 as the interpretation of density movements in terms of quantiles applies. Figure 4 shows the estimated parameters for model (2) for TEF, ELE and ANA respectively. Every point is an estimated parameter for a different quantile. We also plot the 5% point-wise confidence intervals.

Top left plots of each panel show the intercept parameters, \( \hat{\omega}(\tau) \) while the coefficients for past absolute returns, \( \hat{\beta}(\tau) \), are in the top right plots. For all the stocks, the magnitude of the lagged value of the return is an important source of variation. But it affects differently the different quantiles of the distribution. The median is unaffected by a shock in \( |r_{t-1}| \). Following the logic of Figure 3, there is no location shift and hence the median remains unchanged for any value of \( |r_{t-1}| \). It changes however any quantile beyond and below 50%. For a given past absolute return, the effect on the extreme quantiles is larger than for the quantiles near the median. Exemplifying, if return at \( t - 1 \) was zero, the density, conditional to the time of the day, remains unchanged. If, by contrast, at \( t - 1 \) there is a large movement in returns, the density becomes more sparse around the median, that remains unchanged, increasing the probabilities of finding a large price variation the next period. If return at \( t - 1 \) is small, the density shrinks, decreasing the probabilities of finding large price variations.

\(^5\)We have also tried with more lags of absolute returns and results, available upon request, don’t change qualitatively.

The remaining six plots show the estimated values of the parameters for the Fourier series. The second line refers to the estimated parameters of the first Fourier series, the third one to the estimated parameters of the second Fourier series and the last one to the third Fourier series. The coefficients for the cosine terms, the alphas, are larger, for all stocks, than the sinus ones, the gammas. This is due to the fact that the cosine series peak at the opening and the closing, the times of the day at which trading activity is more intense and return dispersion is bigger. The coefficients for the three cosine terms are virtually identical, in shape, for all the stocks. The first one, \(\hat{\alpha}_1(\tau)\), has the largest magnitude and this is easily explained by its shape: it has two peaks -in the opening and the closing- and the minimum in the middle of the day. The other two cosine series also have two maxima -at the opening and at the closing- but they are not the only one. For this reason the magnitude of \(\hat{\alpha}_2(\tau)\) and \(\hat{\alpha}_3(\tau)\) is smaller than the one of \(\hat{\alpha}_1(\tau)\). On the other side, the coefficients of the three sinus terms are smaller and display differences among the stocks. These results confirm the intuition that the cosine series are more suitable to describe the behavior of the dispersion during the day than the sinus ones. None of the coefficients is different from zero for \(\tau = 0.5\), meaning that the median is not affected by past observation nor the time of the day. In other words, no profit strategies based on the time of the day are found. Consequently, since also the estimated coefficient of \(|\tau_{t-1}|\) for \(\tau = 0.5\) is zero, the conditional median is equal to the unconditional one, that is zero.

This discussion on the cosines and sines of the Fourier terms is rather technical but the conclusion is clear: the density of returns does depend on the time of the day. To grasp further insights, Figure 5 shows the estimated seasonal component, \(seas_d(\tau)\), computed as in (1). The plots read as follows. Each line is the seasonal component for a specific hour of the day for different quantiles. The estimated seasonal components displays different shapes within the day and some conclusions can be drawn. First, the shape and the magnitude of the seasonal component is fairly similar for all the stocks. In particular, there is no seasonal behavior at the median. But there is beyond it and becomes more remarkable as we approach the extreme quantiles. Second, the seasonal component is clearly different at the opening and the closing, with values that are negative for taus smaller than 0.5 and positive for taus bigger than 0.5. Third, the seasonal component at 13:00 and 14:00 displays exactly the opposite behavior with respect to the one at the opening and closing, but with a smaller magnitude.

To better see how the conditional distribution of returns moves though the day, Figure 6 plots the conditional quantiles of the 15 minutes returns for different hours of the day. Rewriting equation (2) conditional to a particular value of past absolute return and on different hours of the day, we have

\[
Q_{\tau_t}(d, |r_{t-1}|) = \omega(\tau) + \beta(\tau)|r_{t-1}| + seas_d(\tau).
\]

The choice of the conditioning value of \(|r_{t-1}|\) has a quantitative but not qualitative effect. For a given \(\tau\), \(\beta(\tau)|r_{t-1}|\) is constant, while the term \(seas_d(\tau)\) changes according to the hour of the day (as we saw in Figure 5). The only effect that the chosen value of \(|r_{t-1}|\) has is to shift all the conditional quantiles at the same \(\tau\) by the same amount. The figure reads as follows: the closer the line is to the horizontal zero line, the more concentrated is the density around the median. And the further it is, the more dispersed it is. The time of the day at which we have the largest seasonal effect is at 17:00, the closure of the market, followed by the effect at 9:30, the opening. At these hours the conditional density becomes more dispersed. In the opposite direction, for all the companies, are the seasonal effects at 13:00 and 14:00. They decrease (in absolute value) the conditional quantiles, decreasing the dispersion. This effect can be associated to a reduced trading activity during the lunch break.
5 Intradaily Value at Risk

As shown in Section 3, there is a one to one relation among the quantile and density functions. This is particularly appealing in the construction of risk measures, which are intimately related with the analysis of the tails of the density function. Using the results of Figure 6 and equation (4) in the Appendix, we can compute the conditional density at different quantiles. Figure 7 shows the tails of these densities. Each point of the conditional density is derived from its relative conditional quantile. As expected, the density mass at the extremes is way larger around the opening and closing than around lunch. This seasonal tail behavior has to be taken into account in the computation of intradaily risk measures, such as VaR.

Value at Risk was developed to provide a single number that could summarize the information about the risk in a portfolio. Over the last ten years, this technique has been increasingly used by banks and regulators all over the world as a way to estimate possible losses related to the trading of financial assets, i.e. as a tool designed to quantify and forecast market risk. In particular, the goal of VaR is to assess the possible loss that can be incurred by a trader or bank, for a given portfolio of assets, over a given time period and for a certain confidence level. The time period and the confidence level are the two major parameters that should be chosen in a way appropriate to the overall goal of risk measurement. When the primary goal is to satisfy external regulatory requirements, such as bank capital requirements of the Basel II Capital Accord, the confidence level is typically small, 1%, and the time horizon is long (usually a 10 day period). However for an internal risk management model, used by a company to control the risk exposure, the typical confidence level is even smaller and the time horizon shorter. In particular, for active market participants such as high frequency traders, floor traders or market makers, the time horizon of their returns is shorter and the corresponding trading risk must be assessed on such short time intervals. Therefore a VaR model that characterizes the market risk on an intradaily basis is useful for market participants (such as intradaily traders and market makers) involved in frequent intradaily trades. Intradaily traders can use intradaily VaR to adapt the composition of their portfolio. For instance, taking short or long positions in different assets when their risk exposure reaches a given threshold.

The VaR at a confidence level of $\tau$ for a given portfolio is the loss at the $\tau$ percent probability level, which can simply be defined as the $\tau$ empirical quantile of the conditional distribution of returns:

$$Pr[r_t < \text{VaR}_t(\tau|\mathcal{S}_{t-1})] = \tau \iff \text{VaR}_t(\tau|\mathcal{S}_{t-1}) = Q_{r_t}(\tau|\mathcal{S}_{t-1}).$$

From an empirical point of view, the computation of the $\text{VaR}_t(\tau|\mathcal{S}_{t-1})$ of a portfolio of assets requires the computation of the empirical quantile at level $\tau$ of the distribution of the future returns of the portfolio given the information set available at time $t-1$, $\mathcal{S}_{t-1}$. Engle and Manganelli (1999) introduced nonlinear QR as a method for computing VaR. The originality of our model relies on two points: the use of high frequency data to forecast VaR at intradaily time horizon and the use of the Fourier series to model the intradaily seasonality of returns in a quantile regression framework. Our model defines the information set available up to time $t-1$, $\mathcal{S}_{t-1}$, as including the lagged absolute value of returns, $|r_{t-1}|$, and the three deterministic Fourier series, that are indexed by the time of the day $d$

$$\text{VaR}_t(\tau|d, |r_{t-1}|) = Q_{r_t}(\tau|d, |r_{t-1}|).$$

The one step ahead out-of-sample VaR forecast is conducted using a rolling window scheme, a method popular among practitioners since Fama and MacBeth (1973) and Gonedes (1973).

---

7A full picture of the density is possible but not relevant as the financial interest lies on the tails and not around the median. And, moreover, it has been shown earlier that nothing interesting happens around the median.
The use of rolling windows is justified by parameter instability, which can distort the out-of-sample forecast. The window size is adapted to the liquidity of the stock. For the most liquid stock, TEF, we use a rolling window of 2000 observations, for ELE a window of 2500 observations and for ANA the less liquid stock a bigger window of 3000 observations. This choice is motivated by the fact that in the same time span we can have a different number of transactions, while for the most liquid stocks (like TEF) in 2000 observations of 15 minutes returns we have enough information due to the high number of transactions, for the less liquid stocks (like ANA) this time interval is too short because it includes a fewer number of transactions. This lead to 23.430, 22.930 and 22.430 one-step ahead forecasts for TEF, ELE and ANA respectively.

Figure 8 displays the last 500 observations of the 15 minutes sampled returns for TEF, ELE and ANA with the relative VaR forecasts at the confidence levels of 2.5%, 1% and 0.5%. The estimated VaRs show clearly the effects of the two components that we used to model the conditional quantiles. The seasonal component is responsible of the deterministic daily oscillations, while the dynamic one is amplifying or reducing the oscillations to take into account the dispersion clustering. Moreover, as the confidence level of the VaR decreases, the dynamic component becomes more relevant.

At first sight, it looks that the estimated VaR for the three confidence levels and for all the stocks are close enough to the data, i.e. we are not overestimating the risk, and that the number of times that we are below the estimated VaR is not too big. As a check we computed the failure rates. That is the percentage of times that the observations are below the VaR. If the VaR is well specified, then the empirical failure rates, denoted by \( \hat{f} \), should be close enough to the confidence level. Table 3 reports the empirical failure rates for the three stocks and for the confidence levels of 2.5%, 1% and 0.5%. The values that are in parenthesis refer to the confidence intervals computed according to the Kupiec (1995) test. The null hypothesis of the test is that the empirical failure rate, \( \hat{f} \), is equal to the confidence level of the VaR, \( \tau \). The 5% confidence interval for \( \tau \) is given by \( \hat{f} \pm 1.96 \sqrt{\frac{\hat{f}(1-\hat{f})}{N}} \), where \( N \) is the total number of observations that we are evaluating, that is 23,430, 22,930 and 22,430 for TEF, ELE and ANA respectively. For all the stocks and all the confidence levels, the confidence interval contains the theoretical confidence levels of 2.5%, 1% and 0.5% respectively, therefore we accept the null hypothesis that the empirical failure rates are equal to the theoretical ones for all the confidence levels of the VaR and for all the stocks.

A test that is equivalent the Kupiec’s is the likelihood ratio test of unconditional coverage developed by Christoersen (1998). This test is based on a hit variable, that takes value 1 if there is a success, that is if the realized return is bigger than the expected VaR, and 0 otherwise, and therefore distributed according to a binomial distribution. The test is

\[
LR_{ac} = -2 \log \left( \frac{(1-\tau)^{n_0} \tau^{n_1}}{(1-\hat{f})^{n_0} \hat{f}^{n_1}} \right) \sim \chi^2, \]

where \( n_0 \) is the number of failures and \( n_1 \) the number of successes. The first panel of Table 4 reports the values of the test with the relative p-values. The conclusion are similar to Kupiec’s test. The model is able to predict well the VaR for all the stocks and for all the confidence levels considered.

82500 observation of 15 minutes return correspond to 58 days (three months), while 2500 observation of 15 minutes observation cover a time span of 73 days (four months) and, finally, 3000 observations are equivalent to 88 days (five months).

9The choice of the optimal window, although relevant in this literature, is out of the scope of this paper.

10These are reasonable confidence levels for intradaily market risk evaluations as Basel threshold is 1%.
However, a drawback of the Kupiec and the likelihood ratio test of Christoffersen is that they just count the number of successes and of failures, testing only the equality between the VaR violations and the confidence level. In a risk management framework, it is also important that the VaR violations are not correlated in time. The likelihood ratio test of independence, Christoffersen (1998), examines serial independence of VaR estimates. As the previous likelihood ratio test, this test is built starting from a hit variable that takes values according to

\[ I_t = \begin{cases} 
1, & \text{if } r_t > \text{VaR}_t(\tau|d_s|r_{t-1}); \\
0, & \text{otherwise.} 
\end{cases} \]

The likelihood ratio test of independence tests the null of independence against a the alternative of a first order Markov process of the violations. If we denote with \( n_{ij} \) the number of observation of \( I \) with value \( i \) followed by \( j \) then we can express the likelihood ratio test of independence as

\[ LR_{\text{ind}} = -2\log \left( \frac{(1 - \hat{f})^{n_{00} + n_{11}} \hat{f}^{n_{01} + n_{10}}}{(1 - \hat{f}_{01})^{n_{00}} \hat{f}_{01}^{n_{01}} (1 - \hat{f}_{11})^{n_{10}} \hat{f}_{11}^{n_{11}}} \right) \sim \chi^2_1, \]

where \( \hat{f}_{01} \) is the percentage of successes after a failure and \( \hat{f}_{11} \) is the percentage of successes after a success. The null of the test is that \( \hat{f}_{01} = \hat{f}_{11} = \hat{f} \). Third panel of Table 4 reports the value of the test with the relative p-values. The null of independence of the violations is accepted for all the stocks and all the confidence levels of the VaR. Finally, as a more powerful tool we performed the joint likelihood ratio test of independence and coverage. The Christoffersen’s likelihood ratio test of conditional coverage

\[ LR_{\text{cc}} = LR_{\text{uc}} + LR_{\text{ind}} \sim \chi^2_2, \]

in which the null of the unconditional coverage is tested against the alternative of the independence test. Bottom panel of Table 4 reports the results. For all the confidence levels we accept the null of conditional coverage confirming that the model is well specified.

The tests results show the ability of the model to provide good out-of-sample forecasts of the intradaily VaR confirming the importance of well specifying the intradaily seasonality. This component, as shown in Figure 8, seems to have a crucial role in the determination of the intradaily VaR.

6 Conclusions

We investigate intradaily seasonal patterns on the probability law of high frequency financial returns. Within the day there are significant variations in asset prices, which imply different evaluations of the tails of the return’s distribution through the day. And these variations are partly deterministic and due to the intradaily seasonality. As returns are realizations of a random variable and as such their behavior is fully described by their conditional probability law. To analyze the intradaily behavior of the probability law we use quantile regression, where the regressors are Fourier series that capture the time of the day and past absolute returns.

Using quote midpoints of three stocks traded at the Spanish stock exchange from January 2001 to December 2003, we show that indeed the conditional probability distribution depends on the time of the day. At the opening and closing the density flattens and the tails become thicker, while in the middle of the day returns concentrate around the median and the tails are thinner. Results are intuitive, in the sense that they confirm the general perception that in the opening and closing the probabilities of finding large price fluctuations are higher than
at lunch. Results, in terms of quantiles, permit straightforward intraday risk evaluations, such as value at risk. We show the intraday variation of the maximum expected loss at 2.5%, 1% and 0.5% confidence levels. The maxima expected losses are, as expected, maximal at the opening and closing and minimal at lunch time. Moreover the test performed on the out-of-sample forecasts of the value at risk show that the model is able to provide good risk assessments.

Appendix

In this appendix we describe the estimation procedure that we followed for the estimation of the asymptotic covariance matrix of the QR estimates. We follow Koenker (2005). Consider the basic model presented in equation (3). The asymptotic distribution of the QR estimator in a non-iid setting

$$
\sqrt{T}(\hat{\beta}(\tau) - \beta(\tau)) \sim N(0, \tau(1 - \tau)H_T^{-1}J_TH_T^{-1})
$$

where

$$
J_T(\tau) = T^{-1}\sum_{t=1}^{T} x_t x'_t
$$

and

$$
H_T(\tau) = \lim_{T \to \infty} T^{-1}\sum_{t=1}^{n} x_t x'_t f_t(\xi_t(\tau))
$$

(3)

and $f_t(\xi_t(\tau))$ denotes the conditional density of the $r_t$ evaluated at the $\tau$-th percent conditional quantile. The asymptotic covariance among estimates at different quantiles has blocks

$$
\text{Cov}(\sqrt{T}(\hat{\beta}(\tau_l) - \beta(\tau_l)), \sqrt{T}(\hat{\beta}(\tau_s) - \beta(\tau_s))) = [\tau_l \wedge \tau_s - \tau_l \tau_s]H_T(\tau_l)^{-1}J_TH_T(\tau_s)^{-1}
$$

The conditional density $f_t(\xi_t(\tau))$ in (3) is estimated using the Hendricks and Koenker (1991) sandwich form. This estimation procedure requires at first to compute the optimal bandwidth for each $\tau$, $h_T$. To do it we used the optimal bandwidth suggested by Bofinger(1975)

$$
h_T = T^{1/5} \left( \frac{4.5\phi^4(\Phi^{-1}(\tau))}{(2\Phi^{-1}(\tau)^2 + 1)^2} \right)^{1/5}
$$

where $T$ is the sample size, $\phi$ is the normal pdf and $\Phi^{-1}$ is the normal quantile function, i.e. the inverse of the normal cdf. Last, we re-perform the QR estimation for the grids $\tau - h_n$ and $\tau + h_n$.

As we showed in Figure 3, the cdf can be obtained inverting the quantile function and, once that we have the cdf, we can recover the density function differentiating. Following this intuition, Hendricks and Bofinger suggest to estimate the conditional density function as

$$
\hat{f}_t = \max\{0, 2h_T/(x'_t \hat{\beta}(\tau + h_T) - x'_t \hat{\beta}(\tau - h_T) - \epsilon)\}
$$

(4)

where $\hat{\beta}(\tau + h_n)$ and $\hat{\beta}(\tau - h_n)$ are the estimated parameters at $\tau - h_n$ and $\tau + h_n$ and $\epsilon$ is a small tolerance parameter that we fixed to 0.01 to avoid dividing by zero.
References


Table 1: Descriptive statistics at different hours of the day

<table>
<thead>
<tr>
<th>Time</th>
<th>TEF Mean</th>
<th>S. Dev</th>
<th>Skew</th>
<th>Kurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30</td>
<td>0.000</td>
<td>0.065</td>
<td>-0.283</td>
<td>7.053</td>
</tr>
<tr>
<td>12:00</td>
<td>0.002</td>
<td>0.035</td>
<td>-0.196</td>
<td>5.945</td>
</tr>
<tr>
<td>15:15</td>
<td>0.000</td>
<td>0.035</td>
<td>-0.274</td>
<td>7.296</td>
</tr>
<tr>
<td>17:15</td>
<td>-0.002</td>
<td>0.048</td>
<td>-0.559</td>
<td>6.062</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>ELE Mean</th>
<th>S. Dev</th>
<th>Skew</th>
<th>Kurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30</td>
<td>-0.004</td>
<td>0.063</td>
<td>-0.076</td>
<td>6.512</td>
</tr>
<tr>
<td>12:00</td>
<td>0.001</td>
<td>0.039</td>
<td>0.353</td>
<td>9.017</td>
</tr>
<tr>
<td>15:15</td>
<td>-0.001</td>
<td>0.033</td>
<td>-1.075</td>
<td>12.684</td>
</tr>
<tr>
<td>17:15</td>
<td>0.002</td>
<td>0.051</td>
<td>0.183</td>
<td>6.066</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>ANA Mean</th>
<th>S. Dev</th>
<th>Skew</th>
<th>Kurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30</td>
<td>-0.008</td>
<td>0.136</td>
<td>-0.604</td>
<td>8.077</td>
</tr>
<tr>
<td>12:00</td>
<td>0.005</td>
<td>0.080</td>
<td>-0.249</td>
<td>8.368</td>
</tr>
<tr>
<td>15:15</td>
<td>-0.003</td>
<td>0.084</td>
<td>-2.338</td>
<td>34.049</td>
</tr>
<tr>
<td>17:15</td>
<td>-0.002</td>
<td>0.111</td>
<td>0.253</td>
<td>6.290</td>
</tr>
</tbody>
</table>

The first column reports the time of the day to which the statistics refer. The second displays the sample mean of all the observation at the selected time of the day. The third the sample standard deviation. The fourth the bias corrected skewness and the last one shows the bias adjusted kurtosis.

Table 2: GARCH(1,1) estimates with Students t-distribution

<table>
<thead>
<tr>
<th></th>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>d.f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEF</td>
<td>0.048</td>
<td>0.270</td>
<td>0.730</td>
<td>3.48</td>
</tr>
<tr>
<td>ELE</td>
<td>0.097</td>
<td>0.354</td>
<td>0.646</td>
<td>3.35</td>
</tr>
<tr>
<td>ANA</td>
<td>0.164</td>
<td>0.424</td>
<td>0.576</td>
<td>2.49</td>
</tr>
</tbody>
</table>

Student’s t GARCH(1,1) $h_t = \omega + \alpha r^2_{t-1} + \beta h_{t-1}$ estimates. d.f. stands for degrees of freedom.
Table 3: Kupiec test on the VaR forecasts

<table>
<thead>
<tr>
<th></th>
<th>VaR(2.5%)</th>
<th>VaR(1%)</th>
<th>VaR(0.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEF</td>
<td>0.0237 (0.0217 0.0256)</td>
<td>0.0100 (0.0088 0.0113)</td>
<td>0.0054 (0.0044 0.0063)</td>
</tr>
<tr>
<td>ELE</td>
<td>0.0233 (0.0213 0.0252)</td>
<td>0.0104 (0.0091 0.0117)</td>
<td>0.0058 (0.0048 0.0067)</td>
</tr>
<tr>
<td>ANA</td>
<td>0.0254 (0.0234 0.0275)</td>
<td>0.0106 (0.0093 0.0120)</td>
<td>0.0058 (0.0048 0.0068)</td>
</tr>
</tbody>
</table>

Empirical failure rates of the VaR forecasts at the confidence levels of 2.5% 1% and 0.5%, with the the confidence intervals in parenthesis.

Table 4: Christoersen’s likelihood ratio test on the VaR forecasts

<table>
<thead>
<tr>
<th></th>
<th>LR test of unconditional coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VaR(2.5%)</td>
</tr>
<tr>
<td>TEF</td>
<td>1.68 (0.19)</td>
</tr>
<tr>
<td>ELE</td>
<td>2.81 (0.10)</td>
</tr>
<tr>
<td>ANA</td>
<td>0.16 (0.69)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LR test of independence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VaR(2.5%)</td>
</tr>
<tr>
<td>TEF</td>
<td>1.06 (0.30)</td>
</tr>
<tr>
<td>ELE</td>
<td>0.96 (0.33)</td>
</tr>
<tr>
<td>ANA</td>
<td>0.50 (0.48)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LR joint test of coverage and independence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VaR(2.5%)</td>
</tr>
<tr>
<td>TEF</td>
<td>2.74 (0.25)</td>
</tr>
<tr>
<td>ELE</td>
<td>3.77 (0.15)</td>
</tr>
<tr>
<td>ANA</td>
<td>0.66 (0.72)</td>
</tr>
</tbody>
</table>

Christoffersen’s likelihood ratio test for the the VaR forecasts at the confidence levels of 2.5% 1% and 0.5%. The first panel presents results for the Christoffersen’s likelihood ratio test of unconditional coverage, $LR_{uc}$ with p-values in parenthesis. The second panel presents results for Christoffersen’s likelihood ratio test of independence, $LR_{iid}$ with p-values in parenthesis. Last panel presents results for Christoffersen’s joint likelihood ratio test of coverage and independence, $LR_{iac}$ with p-values in parenthesis.
Figure 1: Kernel estimates at different hours of the day.

Nonparametric density estimate of the 15 minutes returns at different hours of the day. For each day we included the observation at the selected hour, therefore each sample contains a number of observation equal to the number of days. The estimate is based on a Gaussian kernel with optimal bandwidth.
Standardized 15 minutes returns. The sample period runs from January 2001 to December 2003. For each stock we have 34 intradaily observations for a total of 25,400.
Figure 3: Changes in location and scale in the pdf through the quantile function.

Top row shows the pdf, cdf and quantile function of a standardized normal. For the other three rows, the continuous line indicates the pdf, cdf and quantile function of the standardized normal. The dashed line in the second row refers to the pdf, cdf and quantile function of a normal with mean 1 and variance 1. In the third row the dashed line indicates a normal with mean 0 and variance 1.5 and in the last row a normal with mean 1 and variance 1.5.
The figure displays the estimated parameters of equation (2). The continuous line indicates the estimated parameters for each \( \tau \) quantile. The dashed one refers to the 5% point-wise confidence intervals.
Figure 5: Seasonal component

Estimated seasonal component $s_\hat{c}a_s(\tau)$, as presented in equation (1), for different times of the day.
Conditional quantiles of $r_t$ given $|r_{t-1}|$ equal to its 50 percent empirical quantile, $Q_{r_t}(\tau|t, |r_{t-1}| = Q_{|r_{t-1}|}(0.50))$, and for different times of the day.
Left and right tail of the conditional densities of $r_t$ given $|r_{t-1}|$ equal to its 50 percent empirical quantile for different times of the day. The conditional density is computed using equation (4) in the Appendix.
Last 500 standardized 15 minutes returns and the relative out of sample Value at Risk forecast at confidence levels 2.5%, 1% and 0.5%.