HAL open science

Forecasting the intra-day effective bid ask spread by combining density forecasts

Malick Fall, Waël Louhichi, Jean-Laurent Viviani

To cite this version:

Malick Fall, Waël Louhichi, Jean-Laurent Viviani. Forecasting the intra-day effective bid ask spread by combining density forecasts. Applied Economics, 2021, 53 (50), pp.5772-5792. 10.1080/00036846.2021.1929821 . hal-03257268

HAL Id: hal-03257268

https://hal.science/hal-03257268v1

Submitted on 10 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Forecasting the intra-day effective bid ask spread by combining density forecasts.

Malick Fall[a], Waël Louhichi[b], Jean-Laurent Viviani[a]

[a] *Université de Rennes 1 - CREM*
[b] *ESSCA School of Management*

Abstract

The bid-ask spread refers to the tightness dimension of liquidity and can be used as a proxy for transaction costs. Despite the importance of the bid-ask spread in the financial literature, few studies have investigated its forecastability. We propose a new methodology to predict the bid ask spread by combining density forecasts of two types of models: Multiplicative Errors Models and ARMA-GARCH models. Our method is employed to predict the *effective* intra-day bid-ask spread series of all shares pertaining to the CAC40 index. Using a one-step-ahead out-of-sample framework, we resort on the Model Confidence Set procedure of Hansen et al. (2004) to classify models and we found that the proposed model appears to beat all the benchmark specifications.

Keywords: Effective bid-ask spread ; High-Frequency ; Multiplicative Errors Models ; Forecasting

To cite this article: Malick Fall, Waël Louhichi & Jean Laurent Viviani (2021) Forecasting the intra-day effective bid ask spread by combining density forecasts, Applied Economics, DOI: 10.1080/00036846.2021.1929821

1. Introduction

The main objective of this paper is to propose a new forecasting framework to predict the *intra-day* effective bid-ask spread dynamics. In the market microstructure theory, the bid-ask spread is explained by two principal theories: the inventory control models and the asymmetric information models. According to the first theory, the spread remunerates the market maker for the costs caused by holding a non-diversified portfolio (Garman, 1976; Ho and Stoll, 1981). However, the asymmetric information paradigm argues that the bid-ask spread compensates for the adverse selection risk related to the existence of informed traders (Copeland and Galai, 1983; Milgrom and Glosten, 1985).

The need to investigate the bid-ask price is principally motivated by its predominant role in the liquidity literature (Amihud, 2019). In fact, the bid-ask spread refers to the tightness dimension of liquidity (Kyle, 1985) and is considered as an immediate cost of transaction for small size trades (Demsetz, 1968; Roll, 1984). In those market models with friction, liquidity is time-varying and induces cost for investors. Henceforth investors require a liquidity premium as compensation (Amihud and Mendelson, 1986; Brunnermeier and Pedersen, 2008). The main empirical findings of this literature are summarized below: illiquidity has a positive effect on expected stocks returns (Amihud and Mendelson, 1986; Brennan and Subrahmanyam, 1996; Datar et al., 1998; Acharya and Pedersen, 2005), fluctuations in the aggregated market liquidity conditions impact expected returns on stocks (Pastor and Stambaugh, 2003), liquidity and volatility in assets are negatively correlated (Acharya and Pedersen, 2005), there exists a "flight to liquidity" effect: less liquid stocks also have more volatility in liquidity innovations (Pastor and Stambaugh, 2003; Acharya and Pedersen, 2005), liquidity can be used to predict future returns (Pastor and Stambaugh, 2003; Acharya and Pedersen, 2005), there is co-movement between the individual liquidity and the market liquidity (Pastor and Stambaugh, 2003), there exists a common factor in liquidity (Chordia et al., 2001; Hasbrouck and Seppi, 2001) and less liquid stocks tend to get higher autocorrelations in returns than high liquid stocks (Avramov et al., 2006).

Despite the availability of high frequency data and the importance of the bid-ask spread in the financial literature, little attention has been paid to its forecastability. In this article, we investigate the empirical statistical property of univariate intra-day bid-ask spread time-series by extending the literature to its forecastability analysis. Accordingly, we aim to provide a comprehensive comparison between the forecasting power of two competitive types of models, namely ARMA-GARCH[1] (ARMA hereafter) and Multiplicative Error Models (MEM hereafter), when applied to non-negative series. MEM models have been introduced by Engle (2002) and are specifically designed for non-negative processes. This non-negativity property makes them natural candidates for bid-ask spread modelling. In contrast, ARMA models, *may* produce negative values but allow for more flexibility by allowing a time-varying conditional variance process. A natural question that arises is what type of model researchers and practitioners should favour to model the bid ask spread? Most studies have focused on the predictive power of ARMA models (Giot and Laurent (2004); Hansen and Lunde (2005); Wilhelmsson (2013)) and on MEM models (Bauwens et al., 2004; Allen et al., 2009) separately. However, we found no papers comparing these two types of models extensively. By testing a large number of specifications and combining density forecasts of both types of models, we intend to fill this gap. This allows us to determine whether omitting the non-negativity property, in case of ARMA modelling, is a serious concern.

Our paper differs in several ways from Gross-Klussmann and Hautsch (2013), who provide a methodology to predict the *quoted* bid-ask spreads (a discrete positive process) for four stocks from the Russel index. In fact, we propose to forecast the (relative) *effective* bid-ask spread which corresponds to the spread effectively incurred by a trader (in contrast to the quoted spread which is a hypothetical cost of trading (Holden et al., 2014) for all shares pertaining to the CAC40 Index, which is the main benchmark for Euronext Paris. This choice, to favour the relative effective spread instead of the quoted spread, is motivated by several reasons. First, regulation institutions stress the importance of measuring *"the spreads actually paid by investors"* (see the Rule 11Ac1-5 of the SEC), but trades sometimes occur inside or outside the Best Bid and Offer (BBO). This is typically the case when a large market buy exhausts the quantity available at the best offer. In this case the quoted spread does not adequately represent liquidity cost and risk models may suffer from it. Secondly, by taking the effective spread, we also circumvent the discrete nature of spread series (multiple ticks) and this allows the use of models designed for continuous variables whose properties are far better known than for discrete positive models. Additionally, discrete models as the LMACP proposed by Gross-Klussmann and Hautsch (2013), do not consider the case when the tick size varies according to the current price, which is effectively the case for our data from Euronext. Finally, we analyse the *relative* spread (i.e. by normalising the spread by the mid-price) instead of its raw value in order to ease its integration in risk models and to facilitate cross-stock comparison.

The aim of this paper is to propose a methodology to better forecast the *effective* bid-ask spread. We compare the prediction power of our method to a benchmark of models proposed in the literature. Model comparison and combination are achieved within a density forecast framework, opposite to point forecast, in order to fully capture forecast uncertainty. This is combined with the Model Confidence Set (MCS) approach of Hansen et al. (2004) to rank models. Forecasts are made for all the shares pertaining to the CAC40 Index[2] for 510 out-of-sample density forecast corresponding to an entire trading week at an observation frequency of 5 minutes. Through the use of truncated scoring rules, we investigate the forecast accuracy in a specific region of interest to help risk managers discriminate among models. We also demonstrate the potential trading profit derived from a correct forecasting model. Our main result shows that no individual model systematically outperforms other models proposed in the literature. However, our methodology based on the combining density forecasts of ARMA and MEM models, outperforms all the benchmark specifications.

The paper is organized as follows: the materiel and data are presented in section 2. The competitive models and the forecasting evaluation procedure are presented in sections 3 and 4. Sections 5 and 6 report the empirical results and conclude.

[1] Autoregressive Moving Average - Generalized Autoregressive Conditional Heteroscedastic models.

[2] These stocks are actively traded, henceforth a trade-based liquidity metric as the effective spread appears relevant.

2. Materials and data

Estimation and forecasts of the bid-ask spread are implemented using Matlab and OX programming language developed by Jurgen A. Doornik (Doornik and Ooms, 2006) with G@rch (Laurent and Peters, 2002) package. A new OX package, devoted to the estimation of MEM models has been developed by the authors for this study and is freely available, as well as the whole material and code of this paper, from the author upon request.

2.1. Constructing the Bid-Ask Spread

The relative effective spread measures trading cost using the prices *actually* obtained by investors and is defined, for a trade i, as:

$$ES_i = 2 \times D_i \times (p_i - MQ_{i-})/MQ_{i-} \tag{1}$$

where D_i is the order direction indicator (+1 for buyer-initiated and -1 for seller initiated trades), p_i is the trade price and $MQ_{i-} = 0.5 \times (Ask_{i-} + Bid_{i-})$ is the midquote *prevailing* just before the transaction. This positive metric measures the "slippage", in other words the deviation of the actual execution price from the midprice. The analysis is conducted at a frequency of 5 minutes by weighting the effective spread by the number of shares traded over this interval. The weighted (relative) effective spread is defined over an interval u as:

$$\text{Weighted } ES_u = \frac{\sum_{i=1}^{N} (ES_i \times size_i)}{\sum_{i=1}^{N} size_i} \tag{2}$$

where N is the number of trades during the interval u, $size_i$ refers to the number of shares traded for the transaction i and u is an interval of 5 minutes. The whole empirical part is based on this weighted effective spread and we will refer to it simply by "bid-ask spread "in the remainder of the document.

2.2. Data

Tick-by-tick data is collected from the Bloomberg database via its Application Programming Interface (API). We focus on two months of tick-by-tick data for all the shares pertaining to the CAC 40 Index traded on the limit order market of Euronext in January 2019. This index contains the 40 largest equities listed in France, measured by free-float market-capitalization and liquidity. These stocks are subject to different trading rules, we restrict our analysis to a homogeneous sample of stocks by selecting stocks belonging to trading groups F1 and F2 as defined by EuroNext. This discards 2 stocks from the sample leaving 38 stocks. The data span from 01/02/2019 to 31/03/2019 (41 days of trading) and contains tick by tick trades and quotes data (recorded chronologically) for all stocks of the sample. We restrict our analysis to the continuous trading hours period, between 9 a.m. and 5h30 p.m. (local time). Due to the 5 minutes aggregation, each day contains 102 weighted bid-ask spread points giving 38 time series, each containing 4182 (= 41 × 102) data points. Bid-ask spread series are computed as indicated in equation (1) except that trades with the same timestamp at sub-second resolution have been merged. Trade size direction has been computed via the algorithm proposed by Lee and Ready (1991). Table A.1 summarises descriptive statistics of the weighted bid-ask spread series. A series of stationarity tests (unreported) has been conducted and all tests give strong results in favour of the hypothesis of stationarity of the bid-ask spread.

Prior to estimation, the cleaning of the dataset has been carried by applying sequentially 8 filters presented in Appendix C. Unlike the standard dataset used in the literature (such has the TAQ), Bloomberg records *chronologically* both trades and quotes data in the same data file. This facilitates the analysis since it is not necessary to merge two datasets (trades and quotes) and the traditional "reporting delay"issue [3] is not faced.

[3] see footnote 10 of Lee and Ready (1991)

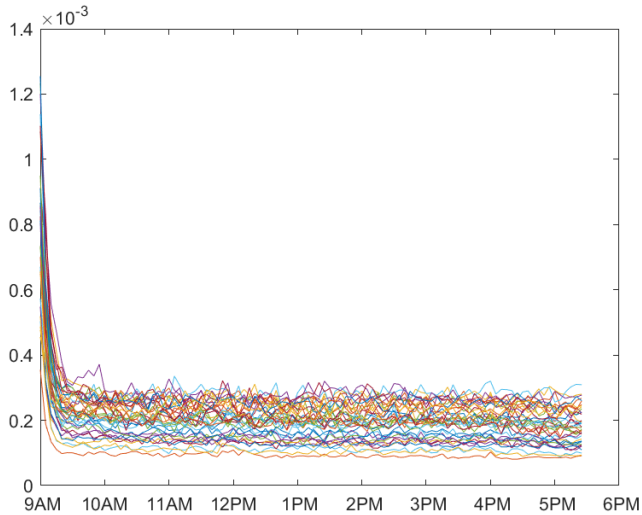

Figure 1: Seasonal effect. This graphic displays the evolution of the seasonal effect $\phi(\tau)$ (equation 3) over a trading day for all series.

The intra-day periodicity in high frequency data is a well-known fact (Andersen and Bollerslev (1997)), and this the reason why prior to analysis, the deterministic diurnal effect has been removed. This is achieved by computing the adjusted bid-ask spread $y_{i,t}$ as:

$$y_{i,t} = \frac{Y_{i,t}}{\phi(t)}$$

where $Y_{i,t}$ is the original bid ask spread value and $\phi(t)$ is the seasonal effect at time t. We follow Bauwens and Veredas (2004) by modelling the deterministic diurnal effect by a non-parametric regression of the observed proxy value on the time of the day. It is given by :

$$\phi(\tau) = \frac{\sum_{i=1}^{N} K((\tau - t_i)/h) Y_{i,t}}{\sum_{i=1}^{N} K((\tau - t_i)/h)} \tag{3}$$

where τ is the number of cumulative seconds from midnight every day. The bandwidth is fixed as in Bauwens and Veredas (2004) and the kernel chosen (K)is the quartic. Figure 1 displays the seasonal effect for all series. We observe that the bid-ask spread exhibits a strong periodicity dynamic, with an important decrease in the first 10 minutes of trading and stabilises after. This is explained by the greater asymmetry of information there is at the opening. Forecasts are performed on the adjusted $y_{i,t}$ series and the non-adjusted forecast can be recovered by post-multiplying forecasts by the *deterministic* seasonal factor $\phi(\tau_i)$.

To make the model more meaningful, we show the different steps of the bid ask spread construction for a single stock (ticker ACAFP - Credit Agricole SA): Figure 2 shows the bid-ask spread before and after the 5-minute aggregation and Figure 3 displays the same data before and after the periodicity adjustment. We observe that the periodicity adjustment lower the number of jumps and shift the series to a unit mean process. Forecasts are based on periodicity adjusted series, that is on the lower plot of Figure 3. Figure 4 displays the autocorrelation function of the original bid-ask spread and Figure 5 displays the autocorrelation function (ACF) of the weighted bid-ask spread series before and after the periodicity adjustment. We clearly see that the periodicity adjustment served its purpose by removing the strong periodicity pattern observed in the ACF.

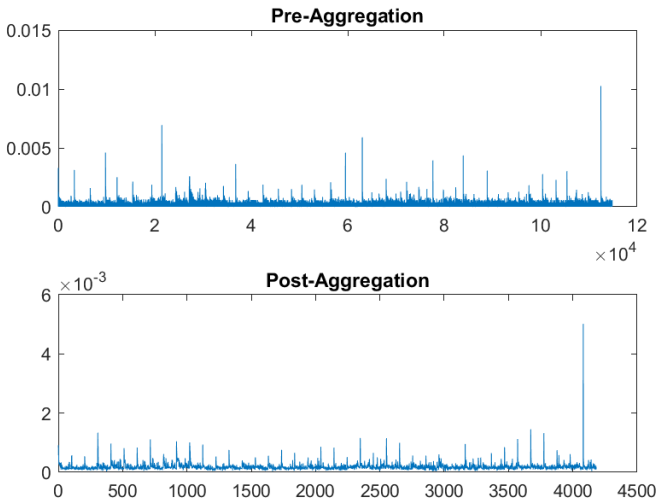

Figure 2: Bid-ask spread time-series. This graphic displays the raw bid ask spread time-series (upper panel) and 5 minutes weighted bid ask spread time-series (lower panel) of Credit Agricole SA company.

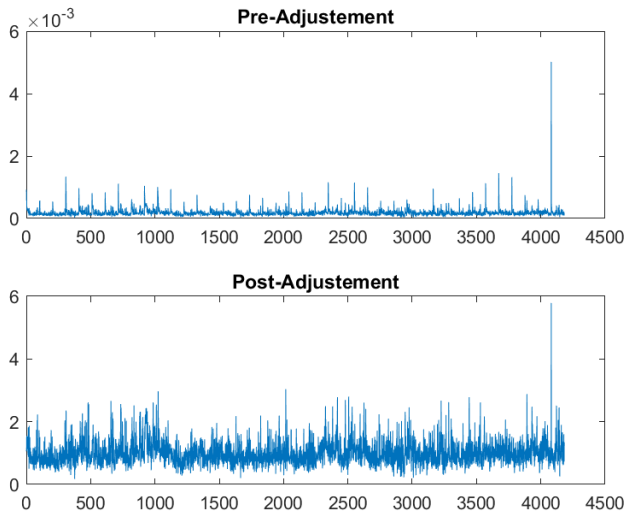

Figure 3: Periodicity adjustment. This graphic displays the unadjusted weighted bid-ask spread time series (upper panel) and the diurnally adjusted weighted bid-ask spread time series (lower panel) of Credit Agricole SA company.

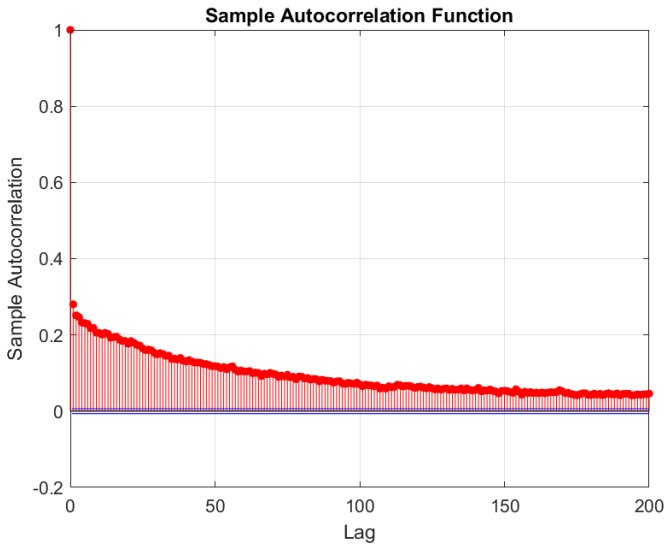

Figure 4: ACF. This graphic displays the autocorrelation function of the raw bid-ask spread time-series of Credit Agricole SA company.

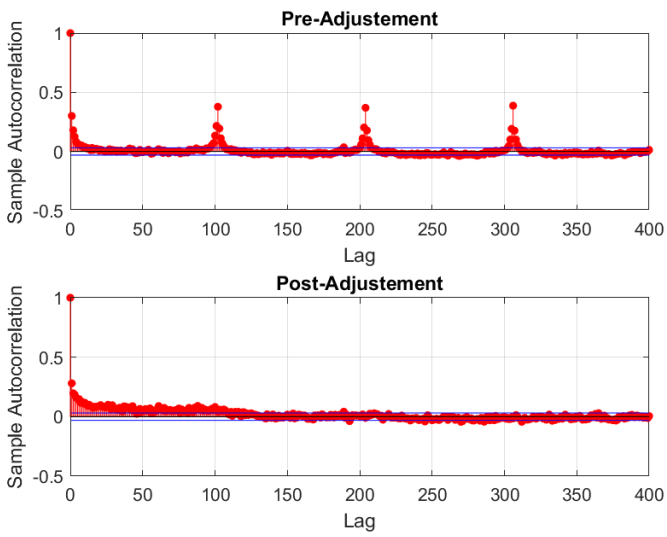

Figure 5: Autocorrelation. This graphic displays the autocorrelation function of the weighted bid-ask spread series of Credit Agricole SA company before (upper panel) and after (lower panel) the periodicity adjustment.

3. Forecasting Models

Our purpose is to predict the bid-ask spread based on the intrinsic time series properties of spread series only; we therefore *do not* include any explanatory variables and all models are performed in a univariate setting. In this section we provide an overview of the types of model considered, MEM and ARMA models.

3.1. ARMA-GARCH type of models

The conditional mean process is modelled via the standard ARMA model. It is defined as:

$$\varnothing\left(L\right)\left(1-L\right)y_t = w + \theta\left(L\right)\epsilon_t, \tag{4}$$

where y_t is the bid-ask spread, L is the traditional lag operator and ϵ_t is white noise residual. $\varnothing(L)$ and $\theta\left(L\right)$ are polynomials of order p and q, where p and q are non-negative integers, in the lag operator L with all their roots lying outside the unit circle. The error term is defined as usual by:

$$\epsilon_t = \sigma_t \eta_t \tag{5}$$

where η_t and σ_t are, respectively, the standardised innovations and conditional standard deviation process.

Regarding the conditional variance process, we tested four well-known models presented in Table 1. Each variance specification aims to capture different features of bid-ask spread series. The GJR model of Glosten et al. (1993) which allows an asymmetric relation between bid-ask spread innovations and volatility changes is considered. Indeed, since illiquidity series are bounded by zero, it is reasonable to assume that shocks have different impacts on the variance process depending on their sign. The Exponential-GARCH (EGARCH - Nelson (1991)) is also considered because, while controlling this asymmetric effect, it also allows the magnitude of shocks to be taken into account. Another feature of bid-ask spread series is the existence of jumps in the series as depicted in Figure 3 and documented by Boudt and Petitjean (2013). To account for this, we implement the BETA-T-GARCH model (a special case of the Generalised Autoregressive Score model (Harvey, 2013)) developed by Harvey and Chakravarty (2008). This model has the advantage of lowering the effect of jumps on the conditional variance process. A peak in the bid-ask spread may be caused by a single very large trade but does not necessarily imply a long-lasting effect. Finally, the standard GARCH model of Bollerslev (1986) is also considered.

The choice of the standardised innovation distribution is pivotal, the aforementioned zero-bounded feature of illiquidity series is likely to play a large role and we expect the bid-ask spread innovations to differ widely from the normal distribution. To evaluate this effect, the standardised innovations are specified to be Skewed Student distributed or Student-t distributed. The Skewed Student distribution initially proposed by Fernandez and Steel (1998) intend to capture both fat tails and asymmetry in the error term conditional density. This distribution has four parameters: location, dispersion, skewness and tail thickness parameters and allows great flexibility in the error term conditional density.[4] To summarise, for each series, we consider a conditional mean processes (ARMA) and four conditional variance processes (GARCH, EGARCH, GJR, BETA-T-GARCH) with Student-t or Skewed Student errors. This modelling framework leads us to estimate 8 different ARMA-GARCH types of model for each of the 38 series. Finally, we also consider four basic benchmark models: an AR(1) model and an ARMA(p,q) model (both without conditional variance process), with Normal or Skewed Student errors.

Estimation is performed by approximate quasi maximum likelihood (Sowell, 1992). The estimation process used a sequential quadratic algorithm implemented in the G@rch package. We relied on the classical BIC Criteria to select orders parameters of the mean process. As an illustration and regarding the ARMA-GARCH,we fit for each series all permutations of the ARMA(p,q)-GARCH(1,1) model with p,q = [0:2] (9 permutations). We then select the most parsimonious one according to BIC.

[4] See Giot and Laurent (2004); Lambert and Laurent (2002) for details.

Table 1: Conditional Variance processes. This table displays the four conditional variance model specifications. GARCH refers to the Generalized Autoregressive Conditional Heteroskedasticity model of Engle (1982), EGARCH to the Exponential GARCH model of Nelson (1991), GJR to the Glosten-Jagannathan-Runkle GARCH model of Glosten et al. (1993) and the BETA-T-GARCH model is proposed by (Harvey and Chakravarty, 2008). All models are estimated with one lag and by the quasi-maximum likelihood method with Skewed-Student errors. Parameters obtained via the QMLE method are : $w, \alpha, \beta, \lambda, \phi_1, \lambda_1, \lambda_2$. I_t is a dummy variable that takes the value 1 when ϵ_t is negative and 0 otherwise.

Model	Variance process
GARCH(1,1)	$\sigma_t^2 = w + \phi_1 \sigma_{t-1}^2 + \alpha_1 \epsilon_{t-1}^2$
EGARCH(1,1)	$\log(\sigma_t^2) = w + \beta \log(\sigma_{t-1}^2) + \alpha \left(\lambda_1 \epsilon_{t-1} + \lambda_2[\lvert\epsilon_{t-1}\rvert - E\lvert\epsilon_{t-1}\rvert]\right)$
BETA-T-GARCH(1,1)	$\sigma_t^2 = w + \alpha_1 u_{t-1} \sigma_{t-1}^2 + \phi_1 \sigma_{t-1}^2$
GJR(1,1)	$\sigma_t^2 = w + \alpha \epsilon_{t-1}^2 + \lambda I_{t-1} \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$

3.2. Multiplicative Error Models (MEM)

MEM models are primarily used to model series with irregularly spaced data or strictly positive series. They therefore appear as natural candidates to forecast the bid-ask spread (Gong et al., 2018). The most popular MEM type model is the Autoregressive Conditional Duration (ACD) model of Engle and Russell (1998) whose primary objective was the modelling of time between events but may be applied to all non-negative processes. This model has led to the developments of a plethora of extensions (see Pacurar (2008)). Formally, a non-negative discrete time process is said to follow a MEM if it can be expressed as (Brownlees et al., 2012; Hautsch, 2012):

$$y_i = \Phi_i \epsilon_i \tag{6}$$

where ϵ_i is a unit-mean non-negative I.I.D random variable :

$$\epsilon_i \sim D^+, E(\epsilon_i) = 1 \tag{7}$$

$$\epsilon_i = \frac{y_i}{\Phi_i} \sim i.i.d \qquad \rightarrow f(\epsilon_i | y_{i-1}, \ldots, y_1; \theta) = f(\epsilon_i; \theta) \tag{8}$$

The conditional mean Φ_i is a positive quantity that evolves deterministically according to a parameter vector θ. The following condition holds :

$$E(y_i | y_{i-1}, \ldots, y_1; \theta) = E(\Phi_i | u_{i-1}, \ldots, \Phi_1; \theta) = \Phi_i,$$
$$V(y_i | y_{i-1}, \ldots, y_1) = \sigma^2 \Phi_i^2$$

In this model, the variance of y_i is proportional to the square of the variance of the error term and *is not* time-varying. This appears as a plausible disadvantage with respect to the ARMA-GARCH type of model. Several models have been proposed based on different specifications of 1) the distribution D^+ of the error term and 2) the specification of the conditional mean process Φ_i. The conditional density of y_i is simply given by:

$$g(y_i) = \Phi_i^{-1} D(y_i \Phi_i^{-1})$$

The most widely used model, proposed by Engle and Russell (1998), is the ACD(m, q) where the m and q refers to the orders of the lags. This model postulates that Φ_i evolves according to the following Linear ARMA-type process:

$$\Phi_i = \omega + \sum_{j=1}^{m} \alpha_j y_{i-j} + \sum_{j=1}^{q} \beta_j \Phi_{i-j}$$

where the unknown parameters are $\theta = (\omega, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_q)$. Its main disadvantage is that constraints on parameters are needed to ensure positivity of y_i and that it imposes a linear relation between the conditional mean process and y_i. To overcome these problems, Bauwens et al. (2008) introduce the LOG-ACD model (two versions LOG-ACD$_1$ and LOG-ACD$_2$). They differ from the standard ACD model by

modelling the logarithm of the conditional expectation. Additionally the LOG-ACD$_1$ specification allows for non-linear effects of small and large spread value by introducing an asymmetry effect: if y_i is superior to the error mean ($y_i > 1$) it has a positive and marginally decreasing effect on the log of the expected spread series while small spread values have a negative effect. However, the use of the logarithmic function implies a possible over-adjustment for small values. The EXACD model of Dufour and Engle (2000) overcome this limitation by explicitly adding an asymmetry parameter. In this model, the impact news function has a piece-wise linear specification, if $\epsilon_{i-j} > 1$ (slope impact: $\alpha_j + \delta_j$ and intercept: $w - \delta_j$) or if $\epsilon_{i-j} < 1$ (slope: $\alpha_j - \delta_j$ and intercept: $w + \delta_j$). We also evaluate the BCACD model, proposed by the same authors, and which include the ACD and LOGACD$_1$ as special cases. To increase the flexibility Fernandes and Grammig (2006) propose the Power ACD (PACD) model by applying a Box–Cox transformation to the conditional mean process through a shape parameter λ. This model adopts for limiting behaviour the LOG-ACD$_2$ ($\lambda \to 0$) and the linear ACD model ($\lambda \to 1$). Finally, two long-memory models are also considered: the Fractionally Integrated ACD (FIACD) model of Jasiak (1998) and the Long Memory ACD (LMACD) model proposed by Karanasos (2003). Both models allow for a hyperbolic decay rate of autocorrelations and have a similar conditional mean process, and only differ in the way they treat the constant term [5].

The previous eight MEMs are considered, in the empirical part, as competitive models and their conditional mean formulations are formally defined in Table 2.

Table 2: Conditional mean process of Multiplicative Errors Models (MEM). y refers to the *adjusted* bid ask spread series. The following conditions hold $E(y_i) = \Phi_i$ and $E(y_i) = \exp(\psi_i)$. The polynomial $\phi(L)$ in the FIACD and LMACD model is defined as $\phi(L) = [1 - \alpha L - \beta L](1 - L)^{-1}$. L is standard Lag operator. Other polynomials are given by $\alpha(L) = \sum_{i=1}^{q} \alpha_i L^i$; $\beta(L) = \sum_{i=1}^{p} \beta_i L^i$. The parameters : $\{\alpha, \beta, \omega, \phi_1, \delta, \lambda\}$ are all estimated via MLE.

ACD	$\Phi_i = \omega + \alpha(L)y_i + \beta(L)\Phi_i$
FIACD	$\Phi_i = \omega[1 - \beta(L)]^{-1} + \left[1 - [1 - \beta(L)]^{-1}\phi(L)(1 - L)^d\right] y_i$
LMACD	$\Phi_i = \omega + \left[1 - (1 - \beta L)^{-1}\phi(L)(1 - L)^d\right] (y_i - w)$
PACD	$\Phi_i^\lambda = \omega + \alpha(L)y_i^\lambda + \beta(L)\Phi_i^\lambda$
LOG-ACD$_1$	$\psi_i = \omega + \alpha(L)\log(\epsilon_i) + \beta(L)\psi_i$
LOG-ACD$_2$	$\psi_i = \omega + \alpha(L)\epsilon_i + \beta(L)\psi_i$
EXACD	$\psi_i = \omega + \alpha(L)\left[\epsilon_i + \delta_j\lvert\epsilon_i - 1\rvert\right] + \beta(L)\psi_i$
BCACD	$\psi_i = \omega + \alpha(L)\epsilon_i^v + \beta(L)\psi_i$

As with the ARMA model, the (multiplicative) error term D^+ in equation 7 has a major impact on the goodness of fit. To allow for great flexibility, we consider four different distributions: Exponential, Weibull, Generalized Gamma and Burr. These distributions are formally presented in Appendix B. The Exponential distribution initially proposed by Engle and Russell (1998) is the simplest distribution specification and implies a constant hazard function. The Weibull, Generalized Gamma and Burr distributions are more flexible and allow the hazard function conditional on past information to be either increasing, decreasing or constant with respect to the bid-ask spread. The Burr distribution, first applied to MEM by Grammig and Maurer (2000), is the most flexible one and admits the three other distributions as a special case.

Estimation. The orders of all MEM models are selected based on the Bayesian information criterion (BIC). The eight MEM models coupled with the four error distributions lead us to test, for each series, 32 competitive MEM models. Each model is estimated by using the conditional likelihood functions based on the crucial assumption of IID innovations. The general log-Likelihood is:

$$L(\theta) = -\sum_{i=1}^{n}\left[\log D\left(\frac{x_i}{\Phi_i};\theta_\epsilon\right) + \log\Phi_i\right]$$

and error distribution specific log-likelihood functions are given in Appendix B .

[5] We follow the procedure proposed by Mccarthy et al. (2003) to fractionnaly difference time series data.

3.3. Forecasting Procedure

For each bid-ask spread series, we apply the 32 MEM models and 12 ARMA-GARCH models to obtain 44 univariate predicted time-series. We produce 510 one step-ahead density forecasts at 5 minute frequency in an out-of sample framework. Each model is re-estimated at every 5 predictions corresponding to a 25 minutes interval.

4. Measuring Forecast Accuracy: the Model Confidence Set Approach

We employed the "Model Confidence Set"(MCS) procedure of Hansen et al. (2004) to measure forecast accuracy. This procedure, based on White (2000) previous research, is particularly useful to compare a large number of models. It starts with a collection of models and determines a set of models that will contain the best model with a given level of confidence, where best is defined in terms of a criterion that is user-specified.

The MCS procedure can be briefly presented in three steps. First, a collection of competing models M^o, a criterion L and a given confidence level α are selected. Second, two tests are applied recursively on the competing models: an equivalence test and an elimination rule. The equivalence test is used to evaluate whether competing models forecast equivalently and the elimination rule discard a model found to be significantly inferior to the other models. Both tests are applied at *the same* confidence level α. At every step of the recursion the number of competing models (survivors) decreases and the algorithm stops when all survivors are equivalent. Finally the "surviving" models are referred to as the "Set of Superior Models" (SSM)[6] and correspond to a set which is guaranteed to contain the best model with a certain level of confidence.

The MCS procedure has the advantage of limiting the "data snooping" bias, by allowing the number of models in the SSM to depend on the information contained in the data [7], thus our results are less dependent on our particular data sample than if we had used more a common ranking procedure.

The MCS methodology is formally presented below. We consider a set M^o containing all the models $(i = 0, \ldots, m)$. For each model the same loss function is computed and denoted $L_{i,t}$ where t refers to a forecast time point. The relative performance variables $d_{ij,t}$ are defined as:

$$d_{ij,t} \equiv L_{i,t} - L_{j,t}, \qquad \forall i, j \in M^o \tag{9}$$

and the variable μ_{ij} as :

$$\mu_{ij,t} \equiv E(d_{ij,t})$$

Model i is preferred if $E(d_{ij,t} < 0) \Leftrightarrow \mu_{ij} < 0$. In this case the expected loss of model i is less than other model losses. The Set of Superior Models (SSM) is defined by :

$$SSM \equiv i \in M^o : \mu_{ij} < 0, \qquad \forall j \in M^o.$$

The equivalence test, which is applied recursively, evaluates the null hypothesis of similar power among models: $H_{0,M} : \mu_{ij} = 0, \forall i, j \in M$ where M is a subset of M^o against the alternative $H_{A,M} : \mu_{ij} \neq 0$ for some $i, j \in M$. We employed statistic proposed by Hansen et al. (2004) associated with this test, given by:

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{\text{var}}(\bar{d}_{ij})}}, \qquad \forall i, j \in M$$

where $\bar{d}_{ij}$ measures the relative sample loss between the i-th and j-th models: $\bar{d}_{ij} = n^{-1} \sum_{t=1}^{n} d_{ij,t}$ and $\widehat{\text{var}}(\bar{d}_{ij})$ is a bootstrapped estimate of $\text{var}(\bar{d}_{ij})$. The higher the estimate of $\text{var}(\bar{d}_{ij})$, the larger the number

[6] Originally this final set was called the "Model Confidence Set"by Hansen et al. (2004), we change this name to avoid confusion with the procedure which has the same terminology.

[7] Put in simple terms: a dataset with few variations in the data will lead to a large SSM and a dataset with large variation will lead to a smaller number of models in the SSM.

of models in final steps. We employed the range statistic $T_{R,M} \equiv \max |t_{ij}|, i, j \in M$, proposed by the same authors to test the null hypothesis of equivalence between models.
The asymptotic behaviour of this statistic is computed via the moving-block bootstrap method, in which block lengths are determined by the (corrected) algorithm proposed in Politis and White (2004). If H_0 is rejected, an elimination rule is applied to identify the model to be removed. The elimination rule is simply defined by removing the model that contributes most to the test statistic.

The MCS procedure yields p-values for each of the models. For a given model $i \in M_0$, the MCS-p-value, $\hat{p}_i$, is the threshold at which $i \in SS\hat{M}_{1-\alpha}$, if and only if $\hat{p}_i \geq \alpha$. Thus a model with a high MCS-p-value is more likely to be one of the "best"models (i.e. to belong to the SSM). This p-value is the main indicator of the forecast accuracy of a model in the sequel.

MCS and Density Forecasts. In order to fully manage uncertainty, we focus on density forecast evaluation instead of point forecasts. By observing the whole future probability distribution of the forecast instead of its central tendency (point forecast), density forecast provides the complete description of the uncertainty associated with a forecast. We follow Wilhelmsson (2013) by applying the MCS procedure to density forecasts. Thus we use a density forecasts metrics as a loss function. Following Mitchell and Hall (2005); Amisano and Giacomini (2007), we select the logarithmic score rule as a loss metric. In terms of density forecast, loss functions are defined as the difference between the density forecast and the true (unknown) density. Scoring rules relate the density forecast to the actual realisation (and therefore to the unknown density) and are designed to assign a high score when the accuracy is good. The proper logarithmic score rules is defined as,

$$S(f, y_{t+1}) = \log \hat{f}_t(y_{t+1})$$

where y_{t+1} is the *realised* value at time $t+1$ and $\hat{f}_t(.)$ the *predictive* density forecast. It assigns a high (low) score if the observed value falls within a region with high (low) predictive density.

This methodology involves computing for each forecast its associated predictive density forecast $\hat{f}(.)$. For both MEM and ARMA models, these values are easily obtained since they are part of the likelihood function which is maximised during estimation [8]. As an illustration, the density forecast of a GARCH model with Gaussian errors is given by $\hat{f}_t(y_{t+1}) = f_{t,N}(y_{t+1})$ where $f_{t,N}$ is the pdf of a $N(\mu_{t+1|t}, \sigma_{t+1|t})$ distribution where $\mu_{t+1|t}$ and $\sigma_{t+1|t}$ refers respectively to the forecasted conditional mean and conditional variances values.

As with MEM models, the predictive likelihood corresponds to the conditional density of y_t given the past information (see Bauwens et al. (2004)) and is formally given by:

$$\hat{f}_t(y_{t+1}) = D_t(y_{t+1}\Psi^{-1}_{t+1|t})\Psi^{-1}_{t+1|t}$$

where $D_t(.)$ corresponds to the estimated error distribution at time t. Finally, since the MCS procedure expects a loss function, we use the negative of these scoring rules.

Tail Evaluations. Density forecasts place equal weight over all the density, in other words uncertainty is treated as a block. However, risk managers generally focus on tail events, so it seems worthwhile addressing the question of density forecast accuracy for a specific region of interest. Diks et al. (2011) provides a tool tailored to this need: the proper censored likelihood (CSL) scoring rule formally defined as:

$$S^{csl}(f, y_{t+1}) = I(y_t \in A_t) \ln \hat{f}_t(y_t) + I(y_t \in A_t^C) \ln \left(\int_{A_t^C} \hat{f}_t(y_t) dy \right) \tag{10}$$

where A_t is the region of interest and A_t^c is the complement of A_t. $I(.)$ is an indicator function that takes the value 1 if the argument is true. The first part of this scoring rule focuses on the behaviour of the density forecast in the region of interest and the second part is required to avoid favouring density with a high

[8] Only for estimation based on out-of-sample forecasts.

mass in the region of interest even if these densities are incorrect. In addition to the full distribution, two regions are considered: the 10 % and 5 % Tails regions. The 10 % Tail region refers to the upper and lower 10 % region of the distribution while the 5 % Tails region is defined similarly. Thus the 10 % Tail region corresponds to A_t defined as $y_t > q^{0.9}$ and $y_t < q^{0.1}$ with q^k the k -th quantile of the *empirical* CDF of the the out-of-sample values $y_1, \ldots, y_t$. Our methodology has the advantage of observing the same number of points in the tails for all series. By doing so, we can put greater weight on particular regions (tails) of the distribution of the variable and compare models according to their accuracy with respect to different scenarios.

Combining density forecasts. We investigate the benefit of combining density forecasts by resorting to the combination scheme proposed by Opschoor et al. (2015). The intuition of using a combination of models is that models can perform unequally over time depending on the evolution of the true data generating process (DGP) of the series. As an illustration, we may think that a large trade of stocks split in several orders can lead to jumps in the series, which in turn allows the BETA-T-GARCH model to be the optimum specification. However as soon as all new stocks are absorbed by the market, jumps may disappear and another (smoother) model may better describe the DGP. Opschoor et al. (2015) extend the CSL score (equation 10) to a combined version of n available predictive density $\hat{f}_{it}, i = 1, \ldots, n$ defined by:

$$S^{csl}(f, y_{t+1}) = \sum_{t=1}^{T} \log \left[\sum_{i=1}^{n} w_{t,i} \left(I(y_t \in A_t) \hat{f}_{it}(y_t) + I(y_t \in A_t^C) \int_{A_t^C} \hat{f}_{it}(y_t) dy \right) \right] \tag{11}$$

We propose an innovative and simple method to find the optimum weights $w_{t,i}$ inspired by portfolio optimisation techniques. We allow the weights to be time-varying by maximising the score every 5 steps for the last 65 density points in a moving-average window way while taking into account forecast uncertainty measured by the covariance matrix of the scoring rules. Specifically, we employ the following optimisation algorithm to determine the weights:

$$\begin{gathered} \text{maximize } w'P - \lambda w' \Sigma w \\ \text{sc } w^{'} e = 1 \\ w_i > 0 \\ \sum b_i \leq K \end{gathered}$$

where P is the matrix of the individual (past) predictive densities, w is the (N x 1) vector of weights, Σ is the covariance matrix of the scoring rules, e is a (N x 1) vector of 1, b_i is a binary value taking the value one if the weight of the i series is strictly positive. The first condition imposes that the sum of weight adds up to one, the second imposes that all weights are positive and the last one imposes a maximum number of selected models. Estimation is performed via second-order cone programming (SOCP). We arbitrarily fix K to be 3 so that, at maximum, only 3 models are combined. The λ parameter is fixed to 0.25. We do not evaluate different values of these parameters as this will go beyond the aim of this paper. The particular chosen parameters are sufficient to show the benefit of combining model. At result, some combined densities may be composed only of ARMA models, only MEM models or a mixture of both types.

5. Empirical Results

Table 3 summarises the composition of the Set of Superior Models (SSM) for the three regions under study (Full, 10 % and 5% Tail distribution). Table A.2 displays summarise statistics of the predictive log-likelihood.

As major observation, we found that combined models systematically outperform other models for alls regions of interest. The maximum percentage of times the combined model is selected as part of the SSM at 25 % confidence interval is 89.47 % , 94.74% and 94.74 % for respectively the full, 10 % tail and 5 % tail region. Thus, the combined model has good forecast accuracy for at least 34 series over the total of 38 series

(89.47%) and this is the higher ratio of all the studied models. This indicates that by combining models we can produce more accurate forecasts. As robustness test, results are display at 10% confidence interval in A.3 and confirm this observation. This finding demonstrates that it is possible to take advantage of the various models accordingly their power to describe series at time t. It worth noting that the study of the evolution of the weights $w_{t,i}$ (not investigated in this paper) may provide valuable comments on the various phases of the process. Nevertheless, the added forecasting accuracy of combined models is not found to be extreme in the sense that the following respective best models have comparable values (0.8664, 0.8421 and 0.8684). A second important observation is that ARMA models seem more suited to predict the bid-ask spread than MEM models. Indeed, the top 5 models for all regions do not contain any MEM model. Caution should be taken on this remark as we only compare *individuals* models here.

Regarding MEM models, the choice of the error term is found to be pivotal. MEM with the Exponential or Weibull distribution are rarely selected as part of the SSM. Inversely, the Gamma and, most of all, Burr distribution models increase the probability of MEM accurate forecast. The Burr distribution appears systematically to outperform other error distributions. This finding indicates that a highly flexible distribution, such as the Burr distribution, is required to produce accurate forecasts and our results suggest practitioners should favour it. Regarding the conditional mean specification, the LMACD and FIACD models are the best models, indicating that the process may have a long memory feature.

Regarding the ARMA type of models, the EGARCH, GJR and BETA-T-GARCH (GAS) conditional variance specifications are the best in class models. The EGARCH and GJR models allow for an asymmetric effect of shocks that, henceforth, seems to be present in the series. We also clearly find that there is a benefit in modelling the conditional variance process since benchmark models which do not possess this feature produce poor forecasts. The error term distribution is also important, the Skewed Student distribution is required to direct errors towards the non-bounded side of bid-ask spread series. The BETA-T-GARCH model performs also well, and we deduced that the observed peaks in bid-ask series may be a concern for the estimation of the conditional variance process.

Regarding parameters estimates and due to the large number of models estimated, we restrict the analysis to the first-ranked model (EGARCH-S) accordingly the MCS results. Thus parameters of The EGARCH-S are presented in Table 4. Parameters are those obtained at the last point of the out-of-sample period. Estimates of the parameters of the error term indicate that bid-ask spread innovations have fat tails and are asymmetric toward the right-hand side of the distribution. Skewness and kurtosis parameters of the Skewed Student distribution are significant and positive. The flexibility of the error term distribution overcomes the zero-limit bound by allowing a strong correction of the innovations toward the non-bounded side. Kurtosis values indicate that bid-ask spread is subject to extreme events. Estimates of the sign (λ_1) and magnitude (λ_2) parameters are positive and significant for most of the series. This indicates that the future conditional variances will be more impacted proportionality as a result of a positive shock than for a negative shock of the same absolute magnitude. Bad news (positive shock on the bid-ask spread) has a bigger impact on the volatility of the bid ask spread than a good ones. We believe this was expected as the bid-ask spread is bounded at zero. Second, coefficients α are negative and significant, thus positive shock tends to decrease the volatility of the series indicating a fast level-reverting effect. This effect is confirmed by coefficients β that are close to one, indicating strong persistence in the variance.

Financial application. The previous analysis was focused mainly on comparing *individual* models *among* the two types of model considered [9] and to provide a combination method. We show that by combining models the overall forecast accuracy can be improved. This leads us, in this section and as a financial application, to consider the potential trading gain (or cost saving) arising from the various *types* of models. That is to say that for *each* series, we select the best ARMA model, the best MEM model and the combined model where best is expressed in terms of forecasting power evaluated via the MCS procedure.

We construct a fictive trading scenario in order to evaluate benefit of each type of model. We assume a basic agent who trades on a regular schedule every 10 minutes. We assume that for each 10-minute interval,

[9] for instance, the FIACD (MEM type) versus the ARMA-GJR (ARMA type) model

Table 3: MCS Summary results. This table reports for each econometric model, the percentage of time this model belongs to the Set of Superior Model $SSM^*_{1-0.25}$. Interpretation: The first value 89.47% indicates that at the 25% threshold, 89.47% of the 38 series predicted with the combined model belongs to the Set of Superior Model. A high percentage is a synonym of better forecasts.

Full region		10% Tail		5% Tail	
Model	$MCS_{75\%}$	Model	$MCS_{75\%}$	Model	$MCS_{75\%}$
Combined	89,47%	Combined	94,74%	Combined	94,74%
EGARCH-S	86,84%	EGARCH-S	84,21%	EGARCH-S	86,84%
GAS-S	86,84%	GAS-S	84,21%	GAS-S	86,84%
GJR-S	84,21%	GARCH-S	78,95%	GJR-S	81,58%
GARCH-S	76,32%	GJR-S	73,68%	GARCH-S	73,68%
FIACD BURR	76,32%	LMACD BURR	68,42%	LMACD BURR	73,68%
LMACD BURR	76,32%	FIACD BURR	65,79%	FIACD BURR	68,42%
LOGACD2 BURR	55,26%	ARMA-S	57,89%	ARMA-S	60,53%
ARMA	52,63%	ACD BURR	52,63%	LOGACD2 BURR	57,89%
ACD BURR	52,63%	LOGACD BURR	52,63%	EXACD BURR	57,89%
BCACD BURR	52,63%	AR1-S	50,00%	PACD BURR	57,89%
ARMA-S	50,00%	LOGACD2 BURR	50,00%	BCACD BURR	57,89%
PACD BURR	50,00%	ARMA	47,37%	ACD BURR	55,26%
GARCH-T	47,37%	PACD BURR	47,37%	AR1-S	50,00%
EGARCH-T	47,37%	EXACD BURR	44,74%	LOGACD BURR	50,00%
EXACD BURR	47,37%	BCACD BURR	42,11%	ARMA	42,11%
GAS-T	44,74%	ACD WEIBULL	36,84%	FIACD WEIBULL	42,11%
GJR-T	44,74%	FIACD GAMMA	36,84%	LMACD WEIBULL	39,47%
LOGACD BURR	44,74%	LMACD GAMMA	36,84%	ACD GAMMA	36,84%
AR1-S	42,11%	AR1	34,21%	LOGACD2 GAMMA	36,84%
AR1	39,47%	LOGACD2 WEIBULL	34,21%	LOGACD2 WEIBULL	36,84%
FIACD GAMMA	36,84%	BCACD WEIBULL	34,21%	PACD WEIBULL	36,84%
EXACD WEIBULL	34,21%	FIACD WEIBULL	34,21%	BCACD WEIBULL	36,84%
FIACD WEIBULL	34,21%	EXACD WEIBULL	31,58%	LMACD GAMMA	36,84%
ACD GAMMA	31,58%	PACD WEIBULL	31,58%	AR1	34,21%
EXACD GAMMA	31,58%	LMACD WEIBULL	31,58%	ACD WEIBULL	34,21%
PACD WEIBULL	31,58%	ACD GAMMA	28,95%	EXACD WEIBULL	34,21%
BCACD WEIBULL	31,58%	LOGACD2 GAMMA	28,95%	BCACD GAMMA	34,21%
LMACD WEIBULL	31,58%	EGARCH-T	26,32%	FIACD GAMMA	34,21%
ACD WEIBULL	28,95%	LOGACD WEIBULL	26,32%	LOGACD WEIBULL	31,58%
LOGACD2 GAMMA	28,95%	GAS-T	23,68%	EXACD GAMMA	31,58%
LMACD GAMMA	28,95%	PACD GAMMA	23,68%	PACD GAMMA	31,58%
LOGACD2 WEIBULL	26,32%	BCACD GAMMA	23,68%	LOGACD GAMMA	28,95%
PACD GAMMA	26,32%	GARCH-T	21,05%	EGARCH-T	18,42%
BCACD GAMMA	26,32%	GJR-T	18,42%	GAS-T	18,42%
LOGACD WEIBULL	23,68%	LOGACD GAMMA	18,42%	GJR-T	18,42%
LOGACD GAMMA	21,05%	EXACD GAMMA	13,16%	GARCH-T	15,79%
ACD EXP	0,00%	ACD EXP	0,00%	ACD EXP	0,00%
LOGACD EXP	0,00%	LOGACD EXP	0,00%	LOGACD EXP	0,00%
LOGACD2 EXP	0,00%	LOGACD2 EXP	0,00%	LOGACD2 EXP	0,00%
EXACD EXP	0,00%	EXACD EXP	0,00%	EXACD EXP	0,00%
PACD EXP	0,00%	PACD EXP	0,00%	PACD EXP	0,00%
BCACD EXP	0,00%	BCACD EXP	0,00%	BCACD EXP	0,00%
FIACD EXP	0,00%	FIACD EXP	0,00%	FIACD EXP	0,00%
LMACD EXP	0,00%	LMACD EXP	0,00%	LMACD EXP	0,00%

Table 4: EGARCH-S parameters estimates.
Skew and Kurtosis refer to parameters of the Skew t distribution. W and V are ,respectively , the intercept of the conditional mean and conditional variance processes. λ_1 is the sign effect and λ_2 is magnitude effect, α and β are standard parameters of the GARCH process. Stars indicate rejection of the null hypothesis at ** 0.01; *0.05 confidence level. Standards deviations are in parentheses. Column ID refers to the identifier of stocks such as presented in Table A.1. ARMA coefficients are not reported.

ID	W	V	α	β	λ_1	λ_2	Skew	Kurt
1	0.99**(0.02)	-1.73**(0.22)	-0.53**(0.09)	0.91**(0.02)	0.14**(0.03)	0.12**(0.04)	0.39**(0.03)	6.07**(0.67)
2	0.99**(0.02)	-0.96**(0.20)	-0.56**(0.11)	0.95**(0.02)	0.04(0.02)	0.16**(0.04)	0.33**(0.03)	7.08**(1.23)
4	0.97**(0.02)	-1.43**(0.29)	-0.56**(0.11)	0.96**(0.01)	0.08**(0.02)	0.12**(0.03)	0.34**(0.03)	5.39**(0.52)
5	0.97**(0.03)	-1.49**(0.26)	-0.52**(0.09)	0.94**(0.01)	0.08**(0.02)	0.12**(0.03)	0.31**(0.02)	6.87**(0.97)
6	0.98**(0.01)	-1.57**(0.22)	-0.31(0.20)	0.88**(0.06)	0.12**(0.03)	0.15**(0.03)	0.35**(0.03)	5.37**(0.49)
7	1.00**(0.02)	0.95(0.49)	-0.38**(0.10)	0.94**(0.01)	0.08**(0.02)	0.16**(0.02)	0.71**(0.03)	3.59**(0.26)
8	0.98**(0.02)	-1.33**(0.24)	-0.56**(0.14)	0.87**(0.05)	0.05*(0.03)	0.22**(0.04)	0.31**(0.03)	7.81**(0.93)
9	1.00**(0.11)	0.66(3.98)	-0.39**(0.11)	0.98**(0.00)	0.05(0.02)	0.19**(0.03)	0.25**(0.03)	13.59**(1.12)
10	0.99**(0.01)	-1.65**(0.21)	-0.26(0.14)	0.86**(0.04)	0.02(0.02)	0.13**(0.03)	0.31**(0.03)	7.45**(1.04)
11	0.99**(0.01)	-1.92**(0.21)	-0.23(0.27)	0.83**(0.04)	0.06*(0.03)	0.10**(0.04)	0.24**(0.02)	7.50**(1.13)
12	0.94**(0.01)	-1.24**(0.34)	-0.54**(0.12)	0.92**(0.02)	0.03(0.02)	0.20**(0.04)	0.22**(0.03)	11.67**(1.31)
13	-1.12**(0.01)	1.88**(0.04)	-0.62(0.43)	0.99(0.63)	-0.02(0.02)	0.22**(0.03)	0.17**(0.03)	8.72(6.80)
14	0.98**(0.03)	-0.51(0.32)	-0.43**(0.10)	0.93**(0.02)	0.05(0.03)	0.21**(0.03)	0.36**(0.03)	9.08**(1.08)
15	1.00**(0.01)	-1.19**(0.16)	-0.15(0.20)	0.90**(0.04)	0.05**(0.02)	0.14**(0.03)	0.25**(0.03)	15.76**(3.86)
16	0.95**(0.05)	-1.17*(0.57)	-0.38**(0.11)	0.98**(0.00)	0.12**(0.02)	0.07*(0.03)	0.36**(0.03)	5.96**(0.62)
17	0.99**(0.01)	-1.43**(0.31)	-0.39*(0.17)	0.93**(0.04)	0.08**(0.03)	0.12**(0.03)	0.32**(0.03)	7.69**(0.88)
18	0.96**(0.03)	-0.20(0.55)	-0.16(0.13)	0.96**(0.01)	0.08**(0.02)	0.14**(0.03)	0.32**(0.03)	8.47**(0.86)
19	0.97**(0.02)	-2.04**(0.14)	-0.02(0.20)	0.68**(0.09)	0.04(0.03)	0.10**(0.03)	0.38**(0.03)	5.27**(0.55)
20	0.98**(0.02)	-0.09(0.34)	-0.68**(0.05)	0.98**(0.01)	0.16**(0.03)	0.15**(0.03)	0.39**(0.03)	5.00**(0.52)
21	0.99**(0.02)	-0.75**(0.29)	-0.22(0.18)	0.91**(0.04)	0.02(0.02)	0.16**(0.04)	0.29**(0.02)	17.83**(6.51)
22	0.98**(0.01)	-2.24**(0.06)	0.74**(0.16)	-0.44**(0.10)	0.09**(0.03)	0.07(0.04)	0.31**(0.03)	4.87**(0.44)
23	1.04**(0.06)	-0.79(1.37)	-0.11(0.18)	0.90**(0.01)	0.02(0.03)	0.16**(0.03)	0.33**(0.03)	11.51**(1.12)
24	0.98**(0.02)	-1.91**(0.09)	-0.15(28.56)	0.82**(0.05)	0.03**(0.01)	0.08(0.07)	0.14**(0.03)	15.87**(5.51)
25	1.00**(0.01)	-1.70**(0.14)	-0.15(0.36)	0.93**(0.02)	-0.02(0.01)	0.10**(0.04)	0.14**(0.03)	18.87**(5.34)
27	0.99**(0.01)	-1.65**(0.05)	0.62**(0.08)	-0.82**(0.03)	0.03(0.02)	0.18**(0.03)	0.28**(0.03)	10.63**(1.91)
29	1.00**(0.01)	-1.90**(0.12)	0.11(0.33)	0.61**(0.10)	0.12**(0.03)	0.10**(0.03)	0.42**(0.03)	6.17**(0.60)
30	1.00**(0.02)	-1.47**(0.17)	-0.47**(0.12)	0.93**(0.02)	-0.01(0.02)	0.17**(0.05)	0.21**(0.03)	22.55**(8.31)
31	0.97**(0.02)	-1.62**(0.22)	-0.45**(0.15)	0.88**(0.02)	0.07*(0.03)	0.14**(0.04)	0.41**(0.03)	5.21**(0.43)
32	0.98**(0.02)	-1.44**(0.29)	-0.47**(0.12)	0.96**(0.01)	0.08**(0.03)	0.11**(0.03)	0.33**(0.03)	7.31**(0.80)
33	0.98**(0.01)	-0.85**(0.25)	-0.10(0.26)	0.94**(0.01)	0.01(0.01)	0.14**(0.04)	0.25**(0.02)	11.79**(2.17)
34	1.00**(0.01)	-1.93**(0.12)	-0.49**(0.15)	0.91**(0.03)	0.05*(0.02)	0.16**(0.03)	0.18**(0.03)	9.62**(1.61)
35	0.97**(0.02)	0.46(0.55)	-0.62**(0.06)	0.98**(0.01)	0.10**(0.02)	0.20**(0.03)	0.33**(0.02)	8.73**(1.11)
36	0.99**(0.01)	-1.38**(0.23)	-0.24(0.25)	0.88**(0.04)	0.02(0.02)	0.12**(0.03)	0.31**(0.03)	10.18**(2.36)
37	0.93**(0.02)	1.39(1.04)	-0.53**(0.13)	0.99**(0.00)	0.07**(0.02)	0.11**(0.03)	0.35**(0.03)	9.28**(1.18)
38	0.98**(0.02)	-0.23(0.48)	-0.68**(0.15)	0.99**(0.02)	0.05*(0.02)	0.13**(0.04)	0.25**(0.03)	19.22**(3.90)

he may choose to trade in the middle of the interval (5 minutes) or at the end of it. One simple alternative trading strategy is considered and consists at time t in producing two one-step ahead forecasts for times $t+5$ and $t+10$ on two-time grids [10] and in selecting the optimum trade timing over both grids based on the following rule :

$$\text{Trade t+5 if : } \hat{f}_{t,5}^{-1}(0.55) \leq \hat{f}_{t,10}^{-1}(\delta)$$
$$\text{Trade t+10 if : } \hat{f}_{t,5}^{-1}(0.55) > \hat{f}_{t,10}^{-1}(\delta)$$

where $\hat{f}_{t,5}^{-1}(.)$ and $\hat{f}_{t,10}^{-1}(.)$ are the inverse quantile functions of, respectively, the forecasted densities at time t for time periods $t+5$ and $t+10$ [11]. The parameter δ represents how safe the strategy is, and is defined over the interval $[0.04 : 0.01 : 0.6]$. A high value δ indicates a safer strategy and vice versa. For the 5-minute grid we only consider the first predicted density since the agent has to make a decision at time t [12].
This strategy is iterated during 255 five-minute intervals and the profits is computed relative to a benchmark strategy. The benchmark strategy consists to randomly trade at the middle or at the end of the 10-minute interval and is computed via Monte Carlo simulations (10 000 replications). The results are presented re-expressed in terms of cost saving (in percentage of the stocks price). Table 5 display results.

The weights used to combine the densities correspond to those found by maximising the Full scoring rule. Due to the fact that a mixture of density does not correspond to a simple weighted of its component, we employ a Monte Carlo procedure to infer the combined density. More specifically, the densities of the combined models are estimated by simulating the error terms ϵ_i in equation 6 and η_t in equation 5 and by using the required estimated parameters at time t. The three simulated series are then combined by applying the weights that are optimum at time t as described in section 4.

The results show potential profits (in bold). The combined model appears as the best type of model, producing higher positive profit and for a wider range of δ values than other models. The maximum average cost savings arising are 0.089%, 0.075 % and 0.022% for, respectively, the Combined, MEM and ARMA types of model. Surprisingly, ARMA models produce less profitable opportunities. This may be explained by the fact that we no longer look for a single individual best model, but select the best MEM models among all MEM models available. This result tends to indicate that even-those MEM models, when considered individually, do not beat ARMA models, the result is opposite if when consider them globally (i.e when we pick the most relevant MEM model for each series). Overall, we find that by combining density forecasts we manage to increase forecast accuracy, leading to the most profitable trading strategy.

[10] that is to say that we aggregate the series over two time scales: 5 and 10 minutes.

[11] To ease computation, when evaluating the combined density forecast type of model, we only employ a combination of models for the 5 min grid forecast, 10 min forecasts are based on GARCH density forecasts. This is not a concern since we expect the trading profit to increase if we had use the full combined scheme, thus keeping the direction of our results.

[12] Between time t and $t+10$ we observe 2 one-step-ahead forecasts on the 5-minute grid; only the first point is considered.

Table 5: Profit of trading strategy. This table reports summarise statistics of the trading strategy. Percentage refers to the average cost saving in the percentage of the stock price across all stocks. δ is the risk parameter. The line labelled Maximum corresponds to the highest profit. Bold figures indicate benefit.

δ	MEM	ARMA	COMBINED
0,4	-0,010%	-0,007%	-0,001%
0,41	-0,009%	-0,005%	**0,002%**
0,42	-0,007%	-0,005%	**0,006%**
0,43	-0,003%	-0,003%	**0,009%**
0,44	-0,003%	-0,003%	**0,013%**
0,45	**0,000%**	**0,003%**	**0,022%**
0,46	**0,004%**	**0,003%**	**0,023%**
0,47	**0,011%**	**0,010%**	**0,024%**
0,48	**0,013%**	**0,014%**	**0,035%**
0,49	**0,023%**	**0,011%**	**0,043%**
0,5	**0,030%**	**0,013%**	**0,056%**
0,51	**0,044%**	**0,016%**	**0,069%**
0,52	**0,047%**	**0,022%**	**0,082%**
0,53	**0,059%**	**0,022%**	**0,084%**
0,54	**0,075%**	**0,012%**	**0,080%**
0,55	**0,067%**	**0,004%**	**0,089%**
0,56	**0,065%**	**0,007%**	**0,078%**
0,57	**0,054%**	**0,004%**	**0,082%**
0,58	**0,043%**	**0,005%**	**0,065%**
0,59	**0,036%**	**0,008%**	**0,052%**
0,6	**0,032%**	**0,016%**	**0,044%**
Maximum	**0,075%**	**0,022%**	**0,089%**

6. Conclusion

The aim of this paper is to propose a methodology to better forecast the *effective* bid-ask spread. Despite the importance of the spread as a determinant of market liquidity (Amihud, 2019), few studies have investigated the forecasting of the bid-ask spread (Gross-Klussmann and Hautsch, 2013). In this paper we contribute to this literature in several ways. First, we provide a methodology to predict the effective bid-ask spreads. Second and as far as we know, our paper is the first to compare the forecasting power of MEM and ARMA types of model. This comparison is applied to bid-ask spread series, which are of critical importance for liquidity risk and portfolio management purposes. Third we propose a new way to combine their density forecasts.

As main results we found that there is no evidence of a clear superiority of a single model originating from the ARMA and MEM models. We show the paramount importance of a correct specification of the error term conditional density; we recommend the Skew-Student distribution for ARMA models and the BURR distribution for MEM models when modelling the bid-ask spread. For MEM models, long-memory models (FIACD, LMACD) are recommended and EGARCH and GAS specifications are recommended for ARMA models. Interestingly, we show that when we combine density forecasts of both types of models, the prediction accuracy can be improved and our methodology outperforms both ARMA and MEM models. Moreover, the Model Confidence Set approach allows us to generalise our results because they are not tied to our particular dataset.

We believe that accurate forecasting of the bid-ask spread dynamics can help in at least three fields. For risk management purposes, the behaviour of models for tail events is of crucial importance and an appropriate model can considerably reduce the risk. For high-frequency research purposes, models often need to forecast non-negativity processes and our results can serve as additional information to decide which model to use. Finally, regarding trading, an accurate prediction of future spread values can obviously lower the cost of transactions, a simple trading strategy is implemented to illustrate it.

As possible further research, it could be worthwhile extending the whole analysis to multivariate econometric models (Vector MEM and Multivariate GARCH). However this will raise difficult, but interesting, issues to apply a combination framework to these models.

Appendix A. Tables

Table A.1: Summarize statistics - Weighted Bid Ask Spread. This table reports summarise statistics for each series, series are aggregated at the 5-minute frequency but have not been diurnally adjusted. Min, Max, Var, Mean, and Med are respectively the mean, minimum, maximum, variance and median of the weighted bid ask spread. For convenience data has been pre-multiplied by a factor of 100 before to generate this table. LB10 and LB20 are respectively the Ljung Box test of the residuals at lag 10 and 20. Nb Trades and Avg Size correspond to the total number of trades and the average trade size over the whole period. To compute the total number of trades, a maximum resolution of one second has been adopted, it means that if two trades occurs within the same second, they are counted as a single trade.

Stock	Min	Max	Var	Mean	Med	LB10	LB20	Nb	Nb Trades	Avg Size
ACAFP	3,00E-03	0,501673	1,62E-06	0,018	0,016	671,328	691,500	4182	603265	1626,2
ACFP	3,84E-14	0,544844	2,45E-06	0,021	0,019	274,948	277,121	4182	273895	559,1
AIFP	2,67E-14	0,134278	1,38E-06	0,027	0,027	193,889	200,172	4182	317929	433,5
AIRFP	6,50E-04	0,843522	2,37E-06	0,014	0,013	123,107	134,044	4182	799579	335,9
ATOFP	6,74E-04	0,365706	2,84E-06	0,022	0,019	688,222	699,189	4182	384816	158,6
BNFP	4,33E-04	0,184066	4,97E-07	0,012	0,011	650,036	696,072	4182	467098	382,2
BNPFP	2,98E-03	0,58024	2,23E-06	0,015	0,012	193,817	200,562	4182	1220256	655,6
CAFP	3,95E-14	0,24976	1,84E-06	0,024	0,022	638,972	666,133	4182	320060	1123,7
CAPFP	2,65E-14	0,26725	2,35E-06	0,027	0,025	1000,444	1429,178	4182	396597	242,2
CSFP	6,43E-04	0,282281	5,99E-07	0,014	0,013	163,347	167,826	4182	598071	1662,1
DGFP	3,40E-14	0,198375	6,49E-07	0,016	0,015	245,949	250,210	4182	442934	367,0
DSYFP	2,22E-14	0,289573	2,24E-06	0,026	0,024	398,721	443,257	4182	242089	177,8
ELFP	2,56E-14	0,24378	2,07E-06	0,029	0,028	1074,129	1840,951	4182	365488	395,4
ENFP	2,23E-14	0,239752	2,38E-06	0,024	0,022	797,984	836,307	4182	234055	466,7
ENGIFP	2,52E-14	0,621779	2,52E-06	0,021	0,020	173,038	173,945	4182	384053	2417,3
FPFP	1,06E-03	0,068635	2,40E-07	0,010	0,009	2215,293	3430,902	4182	1207160	906,2
FRFP	1,66E-03	0,287461	2,50E-06	0,027	0,025	708,006	738,779	4182	504396	571,2
FTIFP	3,37E-14	0,518366	3,99E-06	0,029	0,026	1049,420	1143,630	4182	385139	747,6
GLEFP	3,70E-03	0,195439	1,08E-06	0,016	0,014	739,347	782,459	4182	1043807	903,1
KERFP	2,23E-14	0,86551	3,77E-06	0,018	0,015	222,325	240,656	4182	537126	70,0
LRFP	2,35E-14	0,278447	1,91E-06	0,022	0,020	342,317	361,733	4182	186886	315,1
MCFP	1,69E-03	1,218281	4,10E-06	0,014	0,013	51,864	52,760	4182	773407	132,0
MLFP	5,33E-14	0,335191	2,25E-06	0,025	0,023	589,508	771,489	4182	430129	235,5
ORAFP	2,47E-14	0,091325	7,83E-07	0,022	0,021	202,340	250,280	4182	450608	2778,7
ORFP	5,03E-14	0,169109	1,29E-06	0,027	0,026	152,543	160,980	4182	383392	220,2
PUBFP	2,65E-14	0,399206	2,49E-06	0,021	0,018	614,536	735,145	4182	310732	350,3
RIFP	3,63E-14	0,172438	1,21E-06	0,021	0,020	263,332	277,895	4182	238190	186,8
RMSFP	4,01E-14	0,580136	3,62E-06	0,028	0,025	537,362	617,606	4182	215111	34,2
RNOFP	2,40E-03	0,19366	1,31E-06	0,017	0,015	955,653	964,329	4182	585261	275,8
SAFFP	4,06E-14	0,168307	1,28E-06	0,025	0,024	171,896	188,753	4182	369141	388,7
SANFP	3,59E-04	0,301519	5,85E-07	0,011	0,010	330,081	349,059	4182	695140	411,3
SGOFP	6,48E-04	0,353211	1,32E-06	0,016	0,014	642,410	657,539	4182	514615	468,1
STMFP	2,35E-14	0,149514	1,20E-06	0,021	0,020	422,523	432,141	4182	419612	1339,5
SUFP	3,44E-04	0,134912	7,36E-07	0,020	0,019	327,286	369,827	4182	489646	483,2
SWFP	2,81E-14	0,424001	3,08E-06	0,022	0,019	1175,140	1253,178	4182	201714	145,3
UGFP	3,40E-14	1,198239	5,67E-06	0,027	0,025	109,227	111,661	4182	390495	1008,8
VIEFP	3,55E-14	0,19804	1,45E-06	0,020	0,018	1146,756	1507,437	4182	248470	957,9
VIVFP	2,93E-14	0,138523	1,60E-06	0,025	0,024	278,791	324,830	4182	292705	1894,7

Table A.2: Predictive log-likelihood summary statistics. This table reports summarise statistics of the negative predictive log-likelihood per model. Q1, Q2 and Q3 are respectively the first, second and third quartile. NB is the number of series (for some series and model, the estimation does not converge). A highly negative predictive log-likelihood indicates a poor performance as it represents a loss metric. Interpretation: the average negative predictive log-likelihood across the 37 stocks for the ARMA-GARCH model with Skewed Student error is 0.3814.

	Average	Max	Min	Std	Q1	Q2	Q3	NB
GAS-S	0,3636	0,7884	-0,0954	0,1815	0,2653	0,3524	0,4772	36
GAS-T	0,3768	0,8140	-0,0604	0,1757	0,2810	0,3526	0,4927	36
EGARCH-T	0,3772	0,8106	-0,0616	0,1751	0,2807	0,3537	0,4899	36
GJR-S	0,3685	0,8143	-0,0834	0,1787	0,2668	0,3630	0,4741	37
EGARCH-S	0,3680	0,7806	-0,0966	0,1766	0,2612	0,3652	0,4781	35
GARCH-S	0,3847	0,8090	0,0977	0,1608	0,2714	0,3654	0,4815	35
ARMA-S	0,3814	0,8618	-0,0481	0,1719	0,2611	0,3675	0,4817	37
GARCH-T	0,3942	0,8390	-0,0526	0,1725	0,2978	0,3728	0,4962	36
FIACD BURR	0,3725	0,9211	-0,1009	0,2043	0,2517	0,3736	0,4902	33
BCACD BURR	0,3772	0,9520	-0,1077	0,2106	0,2560	0,3738	0,4963	32
PACD BURR	0,3783	0,9510	-0,1077	0,2084	0,2573	0,3740	0,4952	33
ACD BURR	0,3785	0,9400	-0,1012	0,2058	0,2561	0,3743	0,4947	33
LOGACD2 BURR	0,3789	0,9398	-0,0991	0,2056	0,2567	0,3743	0,4945	33
EXACD BURR	0,3781	0,9497	-0,1077	0,2090	0,2607	0,3743	0,4616	30
LMACD BURR	0,3728	0,9211	-0,0989	0,2041	0,2548	0,3746	0,4902	33
GJR-T	0,4007	0,8731	-0,0529	0,1958	0,2896	0,3837	0,5150	37
LOGACD BURR	0,3963	1,0812	-0,1081	0,2272	0,2564	0,3889	0,5180	33
AR1-S	0,4024	0,9429	0,0559	0,1670	0,2916	0,3922	0,4862	37
EXACD GAMMA	0,4460	0,8514	0,1381	0,1699	0,2970	0,4406	0,5666	33
LMACD GAMMA	0,4607	1,1513	0,1268	0,2089	0,2940	0,4464	0,5626	34
FIACD GAMMA	0,4609	1,1513	0,1268	0,2090	0,2940	0,4515	0,5659	34
LOGACD2 GAMMA	0,4676	1,1749	0,1379	0,2090	0,3014	0,4523	0,5807	34
ACD GAMMA	0,4675	1,1745	0,1378	0,2093	0,2977	0,4528	0,5786	34
PACD GAMMA	0,4745	1,1781	0,1377	0,2104	0,3096	0,4612	0,5786	33
BCACD GAMMA	0,4741	1,2023	0,1377	0,2122	0,3094	0,4617	0,5791	33
ACD WEIBULL	0,5583	2,7279	0,1079	0,4044	0,3752	0,4728	0,6090	37
PACD WEIBULL	0,5558	2,7072	0,1088	0,4019	0,3745	0,4731	0,6079	37
BCACD WEIBULL	0,5568	2,7070	0,1099	0,4011	0,3784	0,4736	0,6084	37
LOGACD2 WEIBULL	0,5558	2,7447	0,1097	0,4017	0,3773	0,4736	0,6041	38
LMACD WEIBULL	0,5407	2,5291	0,1017	0,3755	0,3716	0,4748	0,5874	37
FIACD WEIBULL	0,5463	2,6276	0,1017	0,3858	0,3725	0,4769	0,5917	38
LOGACD GAMMA	0,5679	2,9405	0,1373	0,4914	0,3097	0,4775	0,6124	33
EXACD WEIBULL	0,5601	2,6921	0,1095	0,4042	0,3777	0,4776	0,6128	36
ARMA	0,5155	1,7586	0,1184	0,2699	0,3676	0,4855	0,5421	38
AR1	0,5360	1,7493	0,1855	0,2673	0,3717	0,4993	0,5993	38
LOGACD WEIBULL	0,8300	7,9818	0,1046	1,3327	0,3798	0,5053	0,6359	38
PACD EXP	0,9946	1,2314	0,4927	0,1119	0,9645	1,0095	1,0579	38
ACD EXP	0,9944	1,2302	0,4923	0,1119	0,9646	1,0095	1,0581	38
FIACD EXP	0,9934	1,2267	0,4891	0,1122	0,9635	1,0095	1,0578	38
BCACD EXP	0,9946	1,2316	0,4929	0,1119	0,9646	1,0096	1,0580	38
LOGACD2 EXP	0,9945	1,2303	0,4929	0,1118	0,9646	1,0096	1,0583	38
EXACD EXP	0,9945	1,2312	0,4936	0,1119	0,9639	1,0096	1,0581	38
LMACD EXP	0,9934	1,2267	0,4891	0,1122	0,9638	1,0099	1,0578	38
LOGACD EXP	0,9995	1,2984	0,4954	0,1156	0,9677	1,0142	1,0599	38

Table A.3: MCS Summary results. This table reports for each econometric model, the percentage of time this model belongs to the Set of Superior Model $SSM^*_{1-0.1}$.

Full region		10% Tail		5% Tail	
Model	$MCS_{90\%}$	Model	$MCS_{90\%}$	Model	$MCS_{90\%}$
Combined	97,37%	Combined	100,00%	Combined	97,37%
GAS-S	92,11%	EGARCH-S	92,11%	GAS-S	92,11%
GJR-S	92,11%	GAS-S	92,11%	EGARCH-S	89,47%
EGARCH-S	89,47%	GJR-S	92,11%	GJR-S	89,47%
GARCH-S	78,95%	GARCH-S	86,84%	GARCH-S	84,21%
FIACD BURR	78,95%	ARMA-S	76,32%	ARMA-S	76,32%
LMACD BURR	78,95%	LMACD BURR	76,32%	FIACD BURR	73,68%
ARMA-S	71,05%	FIACD BURR	73,68%	LMACD BURR	73,68%
BCACD BURR	65,79%	ACD BURR	60,53%	ACD BURR	71,05%
ACD BURR	63,16%	LOGACD2 BURR	60,53%	PACD BURR	71,05%
LOGACD2 BURR	63,16%	AR1-S	57,89%	BCACD BURR	71,05%
LOGACD BURR	60,53%	EXACD BURR	57,89%	LOGACD2 BURR	68,42%
EXACD BURR	60,53%	PACD BURR	57,89%	EXACD BURR	65,79%
PACD BURR	60,53%	LOGACD BURR	55,26%	AR1-S	60,53%
EGARCH-T	57,89%	BCACD BURR	55,26%	LOGACD BURR	57,89%
GAS-T	55,26%	FIACD GAMMA	50,00%	FIACD WEIBULL	55,26%
ARMA	55,26%	ARMA	47,37%	LMACD WEIBULL	52,63%
GARCH-T	52,63%	FIACD WEIBULL	47,37%	ARMA	50,00%
GJR-T	52,63%	LMACD GAMMA	47,37%	LMACD GAMMA	50,00%
AR1-S	52,63%	EGARCH-T	42,11%	ACD WEIBULL	47,37%
AR1	44,74%	AR1	42,11%	LOGACD2 WEIBULL	47,37%
FIACD GAMMA	44,74%	ACD GAMMA	42,11%	PACD WEIBULL	47,37%
LOGACD2 WEIBULL	42,11%	ACD WEIBULL	42,11%	BCACD WEIBULL	47,37%
FIACD WEIBULL	42,11%	PACD WEIBULL	42,11%	FIACD GAMMA	47,37%
LMACD GAMMA	42,11%	BCACD WEIBULL	42,11%	AR1	44,74%
EXACD WEIBULL	39,47%	LMACD WEIBULL	42,11%	EXACD WEIBULL	44,74%
BCACD WEIBULL	39,47%	LOGACD2 GAMMA	39,47%	LOGACD2 GAMMA	42,11%
LMACD WEIBULL	39,47%	LOGACD2 WEIBULL	39,47%	EXACD GAMMA	42,11%
ACD WEIBULL	36,84%	EXACD WEIBULL	39,47%	ACD GAMMA	39,47%
PACD WEIBULL	36,84%	GARCH-T	36,84%	PACD GAMMA	39,47%
ACD GAMMA	34,21%	GAS-T	36,84%	BCACD GAMMA	39,47%
LOGACD2 GAMMA	34,21%	GJR-T	36,84%	LOGACD GAMMA	34,21%
BCACD GAMMA	34,21%	EXACD GAMMA	36,84%	LOGACD WEIBULL	34,21%
LOGACD WEIBULL	31,58%	PACD GAMMA	36,84%	EGARCH-T	31,58%
EXACD GAMMA	31,58%	BCACD GAMMA	36,84%	GAS-T	31,58%
PACD GAMMA	28,95%	LOGACD WEIBULL	31,58%	GJR-T	31,58%
LOGACD GAMMA	23,68%	LOGACD GAMMA	23,68%	GARCH-T	28,95%
ACD EXP	0,00%	ACD EXP	0,00%	ACD EXP	0,00%
LOGACD EXP	0,00%	LOGACD EXP	0,00%	LOGACD EXP	0,00%
LOGACD2 EXP	0,00%	LOGACD2 EXP	0,00%	LOGACD2 EXP	0,00%
EXACD EXP	0,00%	EXACD EXP	0,00%	EXACD EXP	0,00%
PACD EXP	0,00%	PACD EXP	0,00%	PACD EXP	0,00%
BCACD EXP	0,00%	BCACD EXP	0,00%	BCACD EXP	0,00%
FIACD EXP	0,00%	FIACD EXP	0,00%	FIACD EXP	0,00%
LMACD EXP	0,00%	LMACD EXP	0,00%	LMACD EXP	0,00%

Appendix B. Error Terms Distribution and Log-Likelihood Function

Four errors terms distribution are considered. Each of one is parametrised such as it has a unit mean. This implies for the Weibull, Burr and Generalized Gamma distribution to let the scale parameter of the distribution to be a function of other parameters and it has the effect to cancel out one parameter in the likelihood functions. For the Exponential distribution λ is fixed to one to ensure a unit mean. The probability density function $f(x)$, survival function $S(x)$, expected value $E(x)$ and the associated log-likelihood function $\log L(\theta)$ are formally given below.

- Exponential(λ)

$$
\begin{aligned}
f(x) &= \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \\
S(x) &= e^{-\frac{x}{\lambda}} \\
E(x) &= \lambda \\
\log L(\theta) &= \sum_{i=t_0}^{T} \left[-\log(\Phi_i) - \frac{x_i}{\Phi_i} \right]
\end{aligned}
$$

- Weibull(γ, β) with restrictions $x > 0, \gamma > 0, \beta > 0$

$$
\begin{aligned}
f(x) &= \gamma \beta^{-\gamma} (x)^{\gamma-1} e^{-(x/\beta)^{\gamma}} \\
S(x) &= e^{[-(x/\beta)^{\gamma}]} \\
E(x) &= \beta \Gamma \left(1 + \frac{1}{\gamma} \right) \\
\log L(\theta) &= \sum_{i=t_0}^{T} \gamma \ln \left[\Gamma \left(1 + \frac{1}{\gamma} \right) \right] + \ln \left(\frac{\gamma}{x_i} \right) + \gamma \log \left(\frac{x_i}{\Phi_i} \right) - \left(\frac{\Gamma(1 + 1/\gamma) x_i}{\Phi_i} \right)^{\gamma}
\end{aligned}
$$

- Burr(u, k, σ^2) Stacy (1962) with restrictions $0 < \sigma^2 < k$

$$
\begin{aligned}
f(x) &= \frac{\mu k x^{k-1}}{(1 + \sigma^2 \mu x^k)^{\left[\frac{1}{\sigma^2} + 1\right]}} \\
S(x) &= (1 + \sigma^2 \mu x^k)^{-\frac{1}{\sigma^2}} \\
E(x) &= \mu^{-(1/k)} \frac{\Gamma(1 + k^{-1}) \Gamma(\frac{1}{\sigma^2} - k^{-1})}{\sigma^{2(1+k^{-1})} \Gamma(\frac{1}{\sigma^2} + 1)} \\
logL(\theta) &= \sum_{i=t_0}^{T} \left[\log(\mu) + \ln k - k \times \ln(\Phi_i) + (k-1) \ln x_i - (\frac{1}{\sigma^2} + 1) \ln(1 + \sigma^2 \mu \Phi_i^{-k} x_i^k) \right]
\end{aligned}
$$

- Generalized Gamma where β, α and k are respectively, the location parameter and two shape param-

eters with restrictions $x > 0, k > 0, \alpha > 0, \beta > 0$

$$f(x) = \frac{\alpha x^{k\alpha-1}}{\beta^{k\alpha}\Gamma(k)} \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right]$$

$$S(x) = 1 - \frac{\gamma(k, (x/\beta)^{\alpha})}{\Gamma(k)}$$

$$E(x) = \beta\frac{\Gamma(k + \alpha^{-1})}{\Gamma(k)}$$

$$\log L(x|\theta, x_{i,0}) = \sum_{i=t_0}^{T} \ln\left(\frac{\alpha}{\Gamma(k)}\right) + (k\alpha - 1)\ln(x_i) - (k\alpha)\log(\Phi_i\lambda) - \left(\frac{x_i}{\Phi_i\lambda}\right)^{\alpha}$$

where $\gamma(a, z)$ is the lower incomplete gamma function,

$$\gamma(a, z) = \int_o^z t^{a-1}e^{-t}dt$$

Appendix C. Screening - Filters

The following eight filters are applied to the high frequency data in order to obtain a reliable sample. These filters are recommended by Hautsch (2012) :

1. Delete observations which are directly indicated to be incorrect, delayed or subsequently corrected.
2. Delete entries outside the regular trading hours.
3. Delete all entries with negative spreads.
4. Delete entries whenever the price is outside the interval [Bid - 2 spread , Ask + 2 Spread].
5. Delete all entries with the spread being greater or equal than 50 times the median spread of that day.
6. Delete all entries with the price being greater or equal than 5 times the median mid-quote of that day.
7. Delete all entries with the mid-quote being greater or equal than 10 times the mean absolute deviation from the local median mid-quote.
8. Delete all entries with the price being greater or equal than 10 times the mean absolute deviation from the local median mid-quote.

References

Acharya, V., Pedersen, L., 2005. Asset pricing with liquidity risk. Journal of Financial Economics 77, 375–410. doi:10.1016/j.jfineco.2004.06.007.

Allen, D., Lazarov, Z., McAleer, M., Peiris, S., 2009. Comparison of alternative ACD models via density and interval forecasts: Evidence from the Australian stock market. Mathematics and Computers in Simulation 79, 2535–2555. doi:10.1016/j.matcom.2008.12.014.

Amihud, Y., 2019. Illiquidity and stock returns: A revisit. Critical Finance Review 8, 203–221.

Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. Journal of financial Economics 17.

Amisano, G., Giacomini, R., 2007. Comparing Density Forecasts via Weighted Likelihood Ratio Tests. Journal of Business and Economic Statistics 25, 177–190. doi:doi:10.1198/073500106000000332.

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4, 115–158. doi:10.1016/S0927-5398(97)00004-2.

Avramov, D., Chordia, T., Goyal, A., 2006. Liquidity and Autocorrelations in Individual stock returns. The Journal of Finance LXI.

Bauwens, L., Galli, F., Giot, P., 2008. The Moments of Log-ACD Models. Qualitative and Quantitative Analysis in Social Sciences 2, 1–28. doi:10.2139/ssrn.375180.

Bauwens, L., Giot, P., Grammig, J., Veredas, D., 2004. A comparison of financial duration models via density forecasts. International Journal of Forecasting 20, 589–609. doi:10.1016/j.ijforecast.2003.09.014.

Bauwens, L., Veredas, D., 2004. The stochastic conditional duration model: A latent variable model for the analysis of financial durations. Journal of Econometrics 119, 381–412. doi:10.1016/S0304-4076(03)00201-X.

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. Journal of econometrics 31, 307–327.

Boudt, K., Petitjean, M., 2013. Intraday liquidity dynamics and news releases around price jumps: Evidence from the DJIA stocks. Journal of Financial Markets , 1–29doi:10.1016/j.finmar.2013.05.004.

Brennan, M.J., Subrahmanyam, A., 1996. Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. Journal of Financial Economics 41, 441–464.

Brownlees, C.T., Cipollini, F., Gallo, G.M., 2012. Multiplicative Error Models, in: Handbook of Volatility Models and Their Applications, pp. 223–247. doi:10.1002/9781118272039.ch9.

Brunnermeier, M.K., Pedersen, L.H., 2008. Market Liquidity and Funding Liquidity. Review of Financial Studies 22, 2201–2238. doi:10.1093/rfs/hhn098.

Chordia, T., Roll, R., Subrahmanyam, A., 2001. Market liquidity and trading activity. The Journal of Finance LVI.

Copeland, T.E., Galai, D., 1983. Information Effects on the Bid-Ask Spread. The Journal of Finance 38, 1457–1469. doi:10.1111/j.1540-6261.1983.tb03834.x.

Datar, V.T., Y. Naik, N., Radcliffe, R., 1998. Liquidity and stock returns: An alternative test. Journal of Financial Markets 1, 203–219. doi:10.1016/S1386-4181(97)00004-9.

Demsetz, H., 1968. The Cost of Transacting. The Quarterly Journal of Economics 82, 33–53.

Diks, C., Panchenko, V., van Dijk, D., 2011. Likelihood-based scoring rules for comparing density forecasts in tails. Journal of Econometrics 163, 215–230. doi:10.1016/j.jeconom.2011.04.001.

Doornik, J.A., Ooms, M., 2006. Introduction to Ox. Technical Report. Ox.

Dufour, A., Engle, R.F., 2000. The ACD model: predictability of the time between consecutive trades. University of Reading and University of California at San Diego 44, 59. doi:10.1007/s00382-008-0398-2.

Engle, R., 2002. New frontiers for ARCH models. Journal of Applied Econometrics 17, 425–446. doi:10.1002/jae.683.

Engle, R., Russell, J., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.

Engle, R.F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica 50, 987–1007.

Fernandes, M., Grammig, J., 2006. A family of autoregressive conditional duration models. Journal of Econometrics 130, 1–23. doi:10.1016/j.jeconom.2004.08.016.

Fernandez, C., Steel, M., 1998. On Bayesian modeling of fat tails and skewness. Journal of the American Statistical ... 93, 359–371.

Garman, M.B., 1976. Market microstructure. Journal of Financial Economics 3, 257–275.

Giot, P., Laurent, S., 2004. Modelling daily Value-at-Risk using realized volatility and ARCH type models. Journal of Empirical Finance 11, 379–398. doi:10.1016/j.jempfin.2003.04.003.

Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. Journal of Finance 48, 1779–1801. doi:10.2307/2329067.

Gong, Y., Chen, Q., Liang, J., 2018. A mixed data sampling copula model for the return-liquidity dependence in stock index futures markets. Economic Modelling 68, 586–598. doi:10.1016/j.econmod.2017.03.023.

Grammig, J., Maurer, K.O., 2000. Non-monotonic hazard functions and the autoregressive conditional duration model. Econometrics Journal 3, 16–38. doi:10.1111/1368-423X.00037.

Gross-Klussmann, A., Hautsch, N., 2013. Predicting bid-ask spreads using long-memory autoregressive conditional poisson models. Journal of Forecasting 32, 724–742. doi:10.1002/for.2267.

Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? Journal of Applied Econometrics 20, 873–889. doi:10.1002/jae.800.

Hansen, P.R., Lunde, A., Nason, J.M., 2004. The Model Confidence Set. Ssrn 79, 453–497. doi:10.2139/ssrn.522382, arXiv:1410.8504.

Harvey, A., Chakravarty, T., 2008. Beta-t-(E) garch. Technical Report September. Faculty of Economics, Cambridge University.
Harvey, A.C., 2013. Dynamic Models for Volatility and Heavy Tails. Cambridge University Press, Cambridge. doi:10.1017/CBO9781139540933.
Hasbrouck, J., Seppi, D., 2001. Common factors in prices, order flows, and liquidity. Journal of financial Economics 59, 383–411.
Hautsch, N., 2012. Econometrics of Financial-High Frequency Data. Springer.
Ho, T., Stoll, H.R., 1981. Optimal dealer pricing under transactions and return uncertainty. Journal of Financial Economics 9, 47–73.
Holden, C.W., Jacobsen, S., Subrahmanyam, A., 2014. The Empirical Analysis of Liquidity. Foundations and Trends in Finance 8, 263–365. doi:10.1561/0500000044.
Jasiak, J., 1998. Persistence in Intertrade Durations. Finance 19, 166–195. doi:10.2139/ssrn.162008.
Karanasos, M., 2003. The Statistical Properties of Long-Memory and Exponential Acd Models. unpublished paper .
Kyle, A.S., 1985. Continuous Auctions and Insider Trading. Econometrica 53, 1315–1336.
Lambert, P., Laurent, S., 2002. Modelling skewness dynamics in series of financial data using skewed location-scale distributions. Technical Report. Université catholique de Louvain - CREST.
Laurent, S., Peters, J.P., 2002. G@rch 2.2: An OX Package for Estimating and Forecasting Various ARCH Models. Journal of Economic Surveys 16, 447–485.
Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. The Journal of Finance .
Mccarthy, J., Disario, R., Saraoglu, H., 2003. A Recursive Algorithm For Fractionally Differencing Long Data Series. Journal of Modern Applied Statistical Methods 2, 272–278.
Milgrom, P.R., Glosten, L.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.
Mitchell, J., Hall, S.G., 2005. Evaluating, comparing and combining density forecasts using the KLIC with an application to the bank of england and NIESR fan charts of inflation. Oxford Bulletin of Economics and Statistics 67, 995–1033. doi:10.1111/j.1468-0084.2005.00149.x.
Nelson, D.B., 1991. Conditional Heteroskedasticity in Asset Returns : A New Approach. Econometrica 59, 347–370.
Opschoor, A., Dijk, D.V., Wel, M.V.D., 2015. Combining Density Forecasts using Censored Likelihood Scoring Rules. unpublished paper .
Pacurar, M., 2008. Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. Journal of Economic Surveys 1.
Pastor, L., Stambaugh, R.F., 2003. Liquidity Risk and Expected Stock Returns. The Journal of Finance 111, 642.
Politis, D.N., White, H., 2004. Automatic Block-Length Selection for the Dependent Bootstrap. Econometric Reviews 23, 53–70. doi:10.1081/ETC-120028836.
Roll, R., 1984. A simple implicit measure of the effective bid ask spread in an efficient market. The Journal of Finance XXXIX.
Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53, 165–188. doi:10.1016/0304-4076(92)90084-5.
Stacy, E.W., 1962. A Generalization of the Gamma Distribution. The Annals of Mathematical Statistics 33, 1187–1192. doi:10.1214/aoms/1177704481.
White, H., 2000. A Reality Check for Data Snooping. Econometrica 68, 1097–1126. doi:10.1111/1468-0262.00152.
Wilhelmsson, A., 2013. Density Forecasting with Time-Varying Higher Moments: A Model Confidence Set Approach. Journal of Forecasting 32, 19–31. doi:10.1002/for.