Long Memory and Tail dependence in Trading Volume and Volatility

Eduardo Rossi
(Università di Pavia)

Paolo Santucci de Magistris
(Università di Pavia)

# 86 (12-08)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Dicembre 2008
Long Memory and Tail dependence in Trading Volume and Volatility *

Eduardo Rossi  
University of Pavia, Italy

Paolo Santucci de Magistris  
University of Pavia, Italy

Dean Fantazzini  
M.V. Lomonosov Moscow State University, Russia

December 10, 2008

Abstract

During the last decades a wide literature has focused on the relationship volume-volatility on financial markets. This paper investigates the temporal dynamics of volatility and volumes, supposing, as in Bollerslev and Jubinski (1999), that the link has to be found in their long-run dependencies, that are supposed to be driven by the same informative process. We analyze the volume-volatility relationship using IBM stocks data. In particular, we rely on the realized volatility based on five minutes stock prices. Tail dependence analysis is carried out with two alternative estimators of the continuous part of the volatility process. The analysis shows that log-realized volatility and log-volumes are characterized by upper and lower tail dependence, where the positive tail dependence is mainly due to the jump component. We also investigate the possibility that volumes and volatility are driven by a common fractionally integrated stochastic trend, i.e. they have the same degree of long memory and are fractionally cointegrated as the Mixture Distribution Hypotesis prescribes. Moreover, we estimate a bivariate ARFIMA specification that explicitly considers the long run relationship between the two series and the tail dependence in the shocks, by parameterizing the joint density by means of different copula functions. The evidence from the model estimates, the simulation results and the forecasts comparison with HAR model highlight the ability of the bivariate ARFIMA with copula density specification to account for the common long memory pattern and tail dependence.

Keywords. Realized Volatility, Trading Volume, Long memory, Fractional Cointegration, Copula Modeling.


---

*Financial support from PRIN 2006 is gratefully acknowledged. Preliminary version.*
1 Introduction

An extensive empirical literature has focused, during the last decades, on the temporal dependencies between volumes and volatility on financial markets. The investigation of the price volatility-volume relationship has important implications in terms of microstructure of financial markets. Numerous empirical investigations find a positive and strong contemporaneous correlation between both absolute returns and volumes. One explanation for the positive price volatility-volume correlation is provided by the sequential information model, see Copeland (1975). In this model, the information is disseminated to only one trader at time and intermediate equilibria occur prior than the final equilibrium. Sequential information imply that there is a positive correlation between price volatility and trading volume in a sequential manner. In the simplest version of the Mixture Distribution Hypothesis (MDH, hereafter), see Clark (1973) and Epps and Epps (1976), price volatility and volume should be positively correlated because the joint dependence on a common underlying variable, that is interpreted as the rate of information flow. According to this theory, the dynamics of volumes and volatility are driven by a common and contemporaneous informative process and both bad news and good news are accompanied by above average volumes and volatility. However, this informative process is unobservable.

Given the leptokurtic distribution of daily returns, with respect to the normal, the MDH implies that data are generated by a conditional stochastic process with variance parameter that varies over time. In particular, MDH helps to explain the high degree of positive correlation between volumes and volatility (see Karpoff (1987)).

The literature on MDH can be classified in two groups. The first one, under the assumption of MDH, focuses on estimation of the model parameters and latent variables to evaluate the goodness of fit with respect to real data.¹ The second one concentrates on the properties of the observed series, relying on an observable (realized) measure of volatility, see Bollerslev and Jubinski (1999) and Luu and Martens (2003). MDH is deeply related to the market microstructure theory which provides a theoretical justification for the contemporaneous correlation between volumes and volatility.

Bollerslev and Jubinski (1999) version of MDH explicitly takes into account this stylized fact. In this model volume and volatility are supposed to be driven by an informative common process with long memory, while the short run dynamics are not necessarily the same. The authors interpret the MDH as a long run phenomenon in which the information arrival process has long memory properties. The low degree of persistence, found in the previous articles, is motivated by the use of very low order autoregressive type formulation. Suppose, instead, that the impact of a given day’s news will last for a random number of days. It follows that the latent aggregate information process has long memory. The MDH theory implies that the long run dependence of the latent informative process will induce the same decay of the autocorrelation functions of volume and volatility.

For all the individual shares of S&P100, they estimate the long memory parameter of volatility and volumes using the log-periodogram regression proposed by Geweke and Porter-Hudak (1984). Then, they test for a common long run hyperbolic rate of decay across the volatility and trading volume series relying on the procedure presented in Robinson (1995).

The main purpose of our paper is to model the relationship between volumes and volatility where both are supposed to be driven by an unobserved long memory process (as in

¹See the approach presented in Andersen (1996) and Liesenfeld (2001).
Bollerslev and Jubinski (1999)). To this end we use a realized measure of volatility based on the high frequency intraday squared returns (as in Luu and Martens (2003)). This is a consistent estimator of the true daily integrated volatility. The results of the Robinson (1995) test, in order to establish whether the two variables share the same order of long run dependence, are supportive of the idea that the two series share the same degree of fractional integration, as reported in table 5. This finding is supportive of the theory of MDH, at least in the version of Bollerslev and Jubinski (1999). Moreover, we investigate the possibility that volumes and realized volatility are fractionally cointegrated. In fact, if the two series were driven by the same long memory latent process, we would expect that there exists a linear combination of the two series that dampens the long run dependence. The evidence of the Nielsen and Shimotsu (2007) test for fractional cointegration does not seem to support this conclusion.

This suggests that we can model the long-run relationship between the logarithm of the realized volatility and the logarithm of volumes by a long memory bivariate model, that is a vector ARFIMA. From the univariate analysis is evident that the filtered log-volumes and log-realized volatility are characterized by leptokurtosis. Moreover the analysis of the tail-dependence of the filtered series suggests that a careful treatment of this aspect is needed. This naturally calls for a suitable choice of the joint distribution. To fully exploit the flexibility of the univariate process, we specify the multivariate distribution function as a copula distribution. These functions provide a flexible tool to model a multivariate distribution when only marginal distributions are known.

An out-of-sample forecast exercise has been carried out in order to evaluate the ability of the model to predict one-period ahead. The benchmark model is a bivariate extension of the HAR model introduced by Corsi (2003). The results are in favor of the multivariate ARFIMA specification with copula densities. Finally, a simulation exercise is carried out to evaluate the ability of the bivariate ARFIMA model, with different copula specifications, to account for some sample statistics. The evidence from the estimation and simulation results highlight the ability of the bivariate ARFIMA to account for the common long memory pattern that is observed in the data.

This paper is organized as follows. Section 2 briefly reviews the theoretical framework beside the concept of realized volatility and its decomposition. In Section 3 a brief description of the data appears. In Section 4 tail dependence analysis is carried out. Section 5 investigates the long memory property of volatility and volumes. Section 6 sets up the model for volumes and volatility. Section 7 presents the copulae functions adopted in the estimation while Section 8 reports the estimation results. Section 9 illustrates the forecast results while Section 10 describes the simulation study for the validation of the model, and Section 11 concludes.

2 Realized variation and its decomposition

In recent years, thanks to the availability of high frequency databases, the series of daily realized volatility (RV) is obtained from high frequency intraday squared returns. Suppose that the model for the variation of the price is a diffusive process:

\[ dp_t = \mu_t dt + \sigma_t dW_t \]  

that describes the trajectories of a semimartingale in continuous time. \( W_t \) is the Wiener process at time \( t \), while \( \sigma_t \) is called spot volatility. The integrated volatility, for day \( t \), is
defined as the integral of the spot volatility

\[ IV_t = \int_{t-1}^{t} \sigma(s)^2 ds \]  (2)

The integrated volatility is the realization that is directly comparable with parametric (or \textit{ex ante}) volatility measurement. Daily squared returns, as a volatility measure, constitute a poor \textit{ex post} estimator, because they overestimate the volatility. Integrated volatility is, instead, a good \textit{ex post} measure and a theoretical benchmark for other volatility estimations. A non parametric measure for integrated volatility is called \textit{realized volatility}. Barndorff-Nielsen and Shephard (2002) have demonstrated that the quadratic variation of a semimartingale that is defined as

\[ [Y_t] = \text{plim} \sum_{j=1}^{t_j \leq t} (Y_{t_j} - Y_{t_{j-1}})^2 \]  (3)

is equivalent to the integrated volatility when returns move as described in (1) and the drift element is continuous.

The sum of successively high-frequency squared returns converges to the quadratic variation of price, (see Meddahi (2002) and Andersen, Bollerslev, Diebold, and Ebens (2001)). The \textit{realized volatility} is a consistent estimator of integrated volatility as the sampling frequency increases.

However, prices sampled at high frequency are affected by the so called microstructure bias and the estimation of integrated volatility becomes imprecise. This fact has been analyzed and solved in different ways (see Ait-Sahalia (2003), Hansen and Lunde (2006) and Bandi and Russell (2003)). The simplest way to deal with this problem is sampling at lower frequencies (for example 5 minutes as in Corsi, Kretschmer, Mittnik, and Pigorsch (2005) or Bollerslev, Kretschmer, Pigorsch, and Tauchen (2005)).

More generally, assume that the price, \( p_t \), follows a continuous-time semimartingale process,

\[ p_t = \int_{0}^{t} \mu_s ds + \int_{0}^{t} \sigma_s dW_s + \sum_{j=1}^{Q(s)} k(s_j) \]  (4)

where the mean process \( \mu_t \) is continuous and of finite variation, \( \sigma_t > 0 \) denotes, as usual, the cad-lag instantaneous volatility. \( Q(t) \) is a counting process that takes value 1 if a jump occurs at \( t \), while \( k(t) \) refers to magnitude. In this case, the quadratic variation process is given by

\[ [p]_t = \text{plim} \sum_{j=1}^{t_j \leq t} (p_{t_j} - p_{t_{j-1}})^2 = \int_{0}^{t} \sigma_s^2 ds + \sum_{j=1}^{Q(s)} k^2(s_j) = IV_t + \sum_{j=1}^{Q(s)} k^2(s_j) \]  (5)

When we allow for the presence of jumps, the quadratic variation is equal to the sum of integrated volatility and jumps. As before, the quadratic variation can be estimated by the sum of the intradaily squared return, \( r_{t,j}^2 \)

\[ RV_t = \sum_{j=1}^{M} r_{t,j}^2 \quad t = 1, ..., T \]  (6)
where $M$ is the number of intraday observations. In this case, realized volatility converges to integrated volatility plus the jump component. Barndorff-Nielsen and Shephard (2004), have shown that RV allows for a direct nonparametric decomposition of the total price variation into its two separate components: a continuous part, called Bipower Variation (BPV), and a discontinuous one, the jumps. Incorporating a measure of jumps is important because, as it has been noted by Huang and Tauchen (2003), their relative contribution to the total price variability is about 7%. Moreover, the realized variance decomposition, into a continuous and jump part, could be particularly interesting, in this context, if we think at daily volumes as the sum of a noise and an informed component.

The BPV is defined as

$$BPV_t = \frac{\pi}{2} \sum_{j=2}^{M} |r_{t,j}||r_{t,j-1}| \quad t = 1, ..., T$$ (7)

and converges to $IV_t$ as $M$ diverges.

Mancini (2007) propose an alternative method for identifying the continuous part of realized volatility based on the following truncation:

$$TRV_t = \sum_{j=1}^{M} r_{t,j}^2 \cdot I(|r_{t,j}| < \theta)$$ (8)

where $\theta$ is a threshold function. This method will throw out more returns as jumps during a high volatility period than during a low volatility period. In a recent paper, Mancini and Renò (2006) recur to a time varying threshold:

$$\theta_t = c_0 \cdot h_t$$

where $h_t$ is the conditional variance obtained by the estimation of a GARCH model. The parameter $c_0$ is set equal to 9, meaning that the estimator cuts observations whose variations are three conditional standard deviations away from zero and it is more accurate in detecting jumps when the diffusive variance is high, that is when a large movement could be more likely due to the diffusive component instead of a jump.

Corsi, Pirino, and Renò (2008) show that the apparent puzzle found in Andersen, Bollerslev, and Diebold (2005), where the jumps seem to not have forecast ability for the future volatility is due to a measurement bias, introduced by the bipower variation in finite samples. In fact, suppose $r_{t,j}$ contains a jump. In the case of bipower variation, it will multiply two adjacent returns, $r_{t,j-1}$ and $r_{t,j+1}$. Asymptotically, both these returns will vanish and bipower variation will converge to integrated continuous volatility. But when $M$ is finite, these returns will not vanish, causing a positive bias which will be larger as $r_{t,j}$ increases. This consideration suggests that the bias of multipower variation will be extremely large in case of consecutive jumps. This causes a positive bias when the bipower variation is used to account for the continuous part of volatility, in particular when two jumps occur in the same daily trajectory.

Corsi, Pirino, and Renò (2008) provide an alternative estimator of the continuous part of volatility, the Corrected Threshold Bipower Variation, hence after $CTBPV$, that is

$$CTBPV_t = \frac{\pi}{2} \sum_{j=2}^{M} Z_1(r_{t,j}, \theta_j)Z_1(r_{t,j-1}, \theta_{j-1}) \quad t = 1, ..., T$$ (9)
where \( Z_t(r_{t,j}, \theta_j) \) is a special function equal to \( |r_{t,j}| \) when \( r_{t,j} < \theta_j \), and equal to \( 1.094 \sqrt{\theta_j} \) when \( r_{t,j} \geq \theta_j \), and \( \theta_j \) is the threshold that is a multiple of local variance, \( \hat{V}_j \), that is chosen according to an iterative procedure, so that

\[
\theta_j = c_\theta \cdot \hat{V}_j
\]

Even if \( CTBPV_t \) and \( TRV_t \) converge to \( IV_t \) when \( \delta \to 0 \), for \( \delta > 0 \) we have that

\[
CTBPV_t \to BPV \quad \text{as} \quad c_\theta \to \infty
\]

\[
TRV_t \to RV \quad \text{as} \quad c_\theta \to \infty
\]

The residual jump component is then calculated as the difference between the realized volatility and the \( CTBPV \)

\[
J_t = RV_t - CTBPV_t \quad t = 1, ..., T \tag{10}
\]

3 Data

Our data set consists of 5-minutes IBM transaction prices from January 1, 1995 through December 31, 2003. Returns, \( r_{t,i} \), over five minutes interval are then calculated, and realized volatility is obtained as the sum of 81 intraday squared returns over five minutes intervals. Daily volumes are given by the sum of intraday volumes.\(^2\)

The series consists of 2267 daily observations. \( BPV_t \) and \( J_t \) are obtained as in formulas (7) and (10). The logarithm of volumes and realized volatility are presented in figure 1. This means that periods with volatility (or volumes) above the mean are followed by periods of volatility (or volumes) below the mean. There is no graphical evidence of the presence of a strong time trend. However, we fit a quadratic trend and consider for the subsequent analysis the detrended series.\(^3\) Some descriptive statistics of the sample are presented in the table 1.

The Box-Pierce portmanteau test statistic (in table 2) shows that volatility and volumes have high degrees of autocorrelation, while returns and \( J_t \) are much less persistent.

4 Tail Dependence

Once the series of daily volatility and volumes are obtained from intraday data, it would interesting to investigate the kind of dependence between the two series. In table 3 we report the estimated contemporaneous correlations. Notice the high correlation between realized measures of volatility and volumes. Other proxies of volatility, such as the squared daily returns, have a very low correlation with the log-volumes. This difference highlights the crucial role of the volatility measurement for the analysis of the volatility-volume relationship. This reinforces the idea of using an high frequency based estimator for the analysis of the dependence between volatility and volumes.

\(^2\)The total number of observations is 183627. The raw data are the tick-by-tick prices and volumes on IBM relative to the open market (from 9:30 am to 4:15 pm). Using the method of previous tick, the series of prices over a grid of five minutes have been created, as well the volumes, as the sum of the number of transactions since the last interval. Week-end and festivity are excluded from the database to avoid seasonality effects.

\(^3\)In the rest of the paper, we refer to \( \log RV_t \), \( \log BPV_t \), \( \log CTBPV_t \) and \( \log V_t \) as the detrended versions of the corresponding measures. These are obtained simply regressing log-volatilities and log-volumes on a constant, a time trend and a squared time trend.
The Pearson correlation measure only applies to observations that are not far out in the tails. The MDH does not provide an explanation for possible positive or negative upper/lower tail dependence. Nevertheless, the exploration of the extremal dependence structure, between volume and volatility, becomes fundamental for identifying and modeling their joint-tail dependence. In order to characterize tail dependence it is helpful to remove the influence of marginal aspects first by transforming the original variables into a common marginal distribution:

\[ U = F_{\text{log } RV}(\text{log } RV) \quad V = F_{\text{log } V}(\text{log } V) \]

As the variables \( U \) and \( V \) are defined on a common scale events of the form \( \{U > u\} \) and \( \{V > u\} \) correspond, for large values of \( u \), to equally extreme events for each variable. By defining the limit

\[ \chi = \lim_{u \to 1} \Pr\{V > u | U > u\} \]

where \( 0 \leq \chi \leq 1 \), we say that when \( \chi > 0 \) the variables are asymptotically dependent while when \( \chi = 0 \) they are asymptotically independent. \( \chi \) measures the degree of dependence that is persistent in the limit. However, \( \chi(u) \equiv \Pr\{V > u | U > u\} \) has a lower power to detect the asymptotic independence. Coles, Heffernan, and Tawn (1999) propose a dependency measure based on \( \Pr\{V > u | U > u\} \):

\[ \overline{\chi}(u) = \frac{2 \log \Pr\{U > u\}}{\log \Pr\{U > u, V > u\}} - 1 \]
Table 2: Sample autocorrelation function (\( \rho_j, j = 1, \ldots, 4 \)). Box-Pierce Portmanteau test statistic for 5, 10 and 40 lags of log-realized volatility (log \( RV_t \)), log-bipower variation (log \( BPV_t \)), log-threshold realized volatility (log \( TVR_t \)), log corrected threshold bipower variation (log \( CTBPV_t \)), jump component (\( J_t \)) and log-volume (log \( V_t \)) of IBM returns.

Table 3: Correlation Matrix. The table reports correlation estimates of returns (\( r_t \)), log-bipower variation (log \( BPV_t \)), log-volumes (log \( V_t \)), log-relative jump (\( J_t \)), log-squared returns (log \( r^2_t \)), log-realized volatility (log \( RV_t \)), log-corrected threshold bipower variation (log \( CTBPV_t \)), and log-threshold realized volatility (log \( TVR_t \))

\[
\chi = \lim_{u \to 1} \chi(u)
\]

\( \chi > 0 \) when \((U, V)\) are positively associated in the extremes, \( \chi = 0 \) when are exactly independent, and \( \chi < 0 \) when are negatively associated. The pair of dependence \((\chi, \chi)\) measures together provides all the necessary information to characterize the form and degree of extremal dependence. For asymptotically dependent variables, we have \( \chi = 1 \) with the degree of dependence given by \( \chi > 0 \). For asymptotically independent variables we have \( \chi = 0 \) with the degree of dependence given by \( \chi \). It is important to test first \( \chi = 1 \) before drawing conclusions about asymptotic dependence based on estimates of \( \chi \).

In order to estimate \( \chi \) it is convenient to transform log \( RV \) and log \( V \) via the Fréchet marginals. Let \( S \) and \( T \) be the unit Fréchet marginals of log \( RV \) and log \( V \),

\[
S = -1 / \log F(\log RV) \quad T = -1 / \log F(\log V)
\] (11)

Following Poon, Rockinger, and Tawn (2004), we calculate the Hill estimator, that is a non
A parametric measure of the degree of tail dependence between volatility and volumes,

\[ \hat{\chi} = \frac{2}{n_u} \left( \sum_{j=1}^{n_u} \log \left( \frac{z_j}{u} \right) \right) - 1 \]  \hspace{1cm} (12)

where \( n_u \) is the number of observation over the threshold \( u \), \( z_j \) is the \( j \)-th order statistic from \( Z = \min(S, T) \). where \( F(\cdot) \) is the univariate empirical distribution function. The variance is given by:

\[ \text{Var}[\hat{\chi}] = \frac{(\hat{\chi} + 1)}{n_u} \]

If there is evidence that \( \hat{\chi} < 1 \), \( \hat{\chi} + 1.96 \sqrt{\text{Var}[\hat{\chi}]} < 1 \) then we can infer that the variables are asymptotically independent. Only if there is no significant evidence to reject \( \bar{\chi} = 1 \), we can estimate the degree of tail dependence that is

\[ \hat{\chi} = \frac{u \cdot n_u}{T} \]  \hspace{1cm} (13)

with variance,

\[ \text{Var}[\hat{\chi}] = \frac{u^2 n_u (T - n_u)}{T^3} \]

The parameter \( \chi \) measures the degree of upper tail dependence, that is the probability of observing a large value of volatility given a large realization of volumes. The analysis of the lower tail dependence is symmetric to the right tail, since the data are multiplied by \(-1\). Figures 2 and 3 report the calculated degree of tail dependence, \( \hat{\chi} \), between the raw series of volatility (including bipower variation, threshold realized volatility and corrected threshold-bipower variation) and \( \log V \), for different choices of the threshold \( u \).

![Figure 2: Hill estimator of the left tail dependence. X-axis measures \( n_u = 20, ..., 200 \).](image)

We have repeated the tail dependence analysis on the series filtered by the long memory component, where the parameter \( d \) has been estimated with the exact Whittle estimator, see figure 4 and 5. Choosing a threshold \( u \) equal to 2.5% of observations, so that \( n_u = 57 \),
the log $V$ and log $RV$ show left tail dependence, the same is true for the filtered series. For what concerns the right tail dependence, log $RV$ and log $BPV$ show positive dependence with log $V$, in particular when the series are fractionally differenced. From the standard error expression is clear that decreasing $u$ we increase the estimate uncertainty. The estimated degree of right tail dependence between log $BPV$ and log $V$, $\hat{\chi}$, is positive, $\hat{\chi} = 0.3306$, with 2.5% of observations on the right tail. log $RV$ present a significant level of asymptotic upper and lower tail dependence with respect to log-volumes. The estimated $\hat{\chi}$ is equal to 0.3622 and 0.2904, with 2.5% of observations respectively on right and left tail, when considering the fractionally differenced series. Interestingly, we don’t find the same evidence for the log $CTBPV$. In fact, even if for $n_u = 57$ it shows asymptotic right tail dependence, the confidence band for the Hill estimator between log $CTBPV$ and log $V$ does not contain the value 1 in most cases. Moreover, the log $TRV$ does not show right tail dependence, while behaves exactly as the log $RV$ on the left tail. The $\hat{\chi}$ in this case is positive and equal to 0.3245 and to 0.2904 when the series is fractionally differenced.

Figures 2 and 3 highlight three important features that characterize the relationship between volatility and volumes.

- First, log-realized volatility and log-volumes display positive upper and lower tail dependence. This means that, given an extreme positive value of volumes, there is a positive probability ($0.3989$) to observe very high volatility the same day. There is also evidence of positive left tail dependence, i.e. when the trades are few, volatility and volumes are asymptotically positively correlated.

- Second, the positive upper tail dependence is mainly due to the contribution of jumps to the realized volatility. In fact, log $CTBPV$, that is the continuous part of realized volatility, and log $V$ are asymptotically independent. This highlights the importance of a good estimation of the jump component of realized volatility. As noted by Corsi, Pirino, and Renò (2008), the bipower variation underestimates the jump component, in particular in case of two consecutive jumps in the intradaily returns. Moreover, the
log TRV well describes the continuous component of realized volatility when positive jumps occur, while seems to be unable to account for jumps with negative sign, that determine the level of left tail dependence.

- Third, the positive lower and upper tail dependence is not due to the long memory component. The positive tail dependence is still present after the fractional differencing.

5 The MDH as a Long Memory Relationship

There is accordance in literature (Andersen, Bollerslev, Diebold, and Labys (2003), Corsi, Kretschmer, Mittnik, and Pigorsch (2005) and Bollerslev, Kretschmer, Pigorsch, and Tauchen (2005)) on some stylized facts:

- the distribution of realized volatility is asymmetric and leptokurtic, but the density of logarithm of the series is close to the Normal.

- both volatility and volumes seems to be fractionally integrated. This means that the effect of a shock decays slowly. This fact is in contrast with an ARMA representation (which implies an exponential decay) or a unit root process.

A suitable model for this observed feature can be an ARFIMA$(0, d, 0)$ process where $d$ is the long memory parameter:

$$ (1 - L)^d y_t = \epsilon_t \quad t = 1, ..., T $$

in this way $y_t$ is a fractionally integrated process of order $d$ and $\epsilon_t$ is an i.i.d. $(0, \sigma^2)$ sequence, and $(1 - l)^d$ is defined by its binomial expansion

$$ (1 - l)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j - d)}{\Gamma(-d)\Gamma(j + 1)} L^j, \quad \Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt. $$
If $|d| \in (0, 1/2)$ the process is stationary. In particular, if $d \in (0, 1/2)$, it presents long memory; instead, if $d \in (-1/2, 0)$ the process is antipersistent with short memory.

As noted by Bollerslev and Jubinski (1999), the high contemporaneous correlation of volatility and volumes could be explained by a modified version of MDH theory, accounting for a common latent stochastic component with long memory that drives their dynamics over time. This hypothesis can be easily verified, estimating the fractional integration parameters for both series, $d_{\log RV}$ and $d_{\log V}$, and then checking if these are statistically equal. Lieberman and Phillips (2006) provides an analytical explanation for the evidence of long memory in the series of realized volatility. In fact, the autocovariance structure of the realized volatility estimator depends on those of the intraday returns. Then, even if the intraday increments are short memory, the sampling scheme renders the $RV$ to be long memory. This suggest that the latent information arrival process, that is approximated by the realized volatility, should also have long memory, since it is the sum of independent intraday information arrivals.

Given this analytical result, the MDH theory can be tested by comparing the degree of fractional integration of the two series. As already noted by Bollerslev and Jubinski (1999) and Luu and Martens (2003), the MDH theory prescribes that the fractional integration order, $d$, of volumes and volatility is the same, since they are both driven by the same long memory process.

We estimate $d$ with the semiparametric estimator of Geweke and Porter-Hudak (1984). Let $I_i(\omega_j)$ denote the sample periodogram for series $i$ at the $j$-th Fourier frequency, $\lambda_j = 2\pi j/T$, where $T$ is the number of observations in the sample. The estimator of $d_i$ is then based on the regression

$$\log[I_i(\lambda_j)] = \alpha - 2d_i \log(\lambda_j) + \epsilon_i \quad (15)$$

where $i = 1, 2$ for volatility and volume, respectively, and $j = 1, ..., m$ where $m$ is such that

$$\frac{1}{m} + \frac{T}{m} \to 0 \quad (16)$$
as $T$ goes to infinity. The results of the estimation are presented in the Table 5. The two series are fractionally integrated and stationary, namely, the estimated $d$ is between 0 and 0.5 for both series. Moreover, we investigate the possibility that the two series exhibit the same degree of fractional integration. Robinson (1995) provides a framework for testing the equality in the fractional integration order. We implement the test to assess if volume and volatility shares the same long run dependence. The test statistic, that is distributed asymptotically as a $\chi^2$ under the null $d_1 = d_2$, is

$$
\xi = \frac{(d_1 - d_2)^2}{z'(X'X)^{-1}z \cdot f'\Omega f}
$$

where

$$
z = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & -2\ln \lambda_1 \\ \vdots & \vdots \\ 1 & -2\ln \lambda_m \end{bmatrix}, \quad f = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \Omega = \begin{bmatrix} \text{Var}(\epsilon_1) & \text{Cov}(\epsilon_1, \epsilon_2) \\ \text{Cov}(\epsilon_1, \epsilon_2) & \text{Var}(\epsilon_1) \end{bmatrix},
$$

Table 5 reports the GPH estimates of the fractional integration parameter $d$ for each series, and for different choices of the bandwidth parameter $m$. All $d$ estimates are below $1/2$, this suggests that both series are covariance-stationary long memory series. We find evidence that the two series share the same degree of fractional integration (as in Luu and Martens (2003)). The Robinson test accept the null hypothesis, i.e. $d\log RV = d\log V$. This finding is supportive of the theory of MDH, at least in the version of Bollerslev and Jubinski (1999). The univariate analysis suggests that both volatility and volume are characterized by long memory and that they are, as expected, strictly connected. Nevertheless, this finding it is not sufficient to guarantee the validity of the MDH theory. In fact, if the MDH theory were verified, there should exist a common stochastic trend, that is the information arrival process, with long memory, that drives the dynamics of volatility and volumes through time. Hence, the analysis of the validity of the MDH should be carried out investigating the degree of fractional cointegration of volume and volatility.

### 5.1 Fractional cointegration analysis

According to the definition in Granger (1986), two (or more) $I(d)$ series are fractionally cointegrated if there exists a linear combination that is $I(d_e)$, with $d_e < d$. Thus the errors are of lower order of fractional integration than the levels. This means that the series share fractionally integrated stochastic trends of different orders ($I(d)$ and $I(d_e)$), and a linear combination eliminates the largest. More precisely, suppose that $z_t$ is a vector $(p \times 1)$ of observables, where the $i$-th element $z_{it} \sim I(d_i)$, with $d_i > 0$, $i = 1, \ldots, p$, we say that they are fractionally cointegrated if there exists a vector $\alpha \neq 0$

$$
e_t = \alpha'z_t \equiv I(d_e) \quad 0 \leq d_e < \min_{1 \leq i \leq p} d_i
$$

This is possible if and only if $d_i = d_j$, some $i \neq j$; a necessary condition for $\alpha$ to be a cointegrating vector is that its $i$-th component be equal to zero if $d_i > d_j$ for all $i \neq j$. When

---

4The simulation results of Hurvich, Deo, and Brodsky (1998) suggest to choose $m = T^{1/5}$ for the bandwidth.
\[ d_1 = \ldots = d_p = d \text{ it is usual to write } z_t \equiv CI(d, b), b = d - d_e. \text{ A typical situation is when } \\
\] \[ z_t = (x'_t, y_t) \in I(d) \text{ and } e_t \in I(d_e) \text{ with } d > d_e \geq 0 \text{ in the model} \\
\]
\[ y_t = \beta'x_t + e_t. \tag{18} \]
Cointegration is commonly thought of as a stationary relation between nonstationary variables
\[ d_i \geq \frac{1}{2} \forall i \text{ and } d_e < \frac{1}{2}. \]
But another possibility is represented by \( 0 \leq d_i < \frac{1}{2} \forall i \) when \( z_t \) and \( e_t \) are stationary. Thus the case where \( d > 0, d_e > 0 \) and \( d + d_e \leq 1/2 \) is called stationary fractional cointegration.
The main drawback of fully specified parametric models is that they provide inconsistent estimators of the long-run parameters if the model is not correctly specified. Robinson (1994) shows that conventional estimators, and in particular OLS are inconsistent when the errors are fractionally integrated. He introduces narrow-band least squares, a semi-parametric method, and proves it is consistent even in situations where the error term is correlated with the regressors. Robinson and Marinucci (2003) and Marinucci and Robinson (2001) show that these semiparametric estimators are consistent for general orders of fractional integration \( d \) for the individual series and \( d_e \) for the errors in the cointegrating (18) relation and for arbitrary short run dynamics. Define the discrete Fourier transform of an observed vector \( \{a_t, t = 1, \ldots, T\} \)
\[ w_a(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} a_t e^{it\lambda}. \]
If \( \{b_t, t = 1, \ldots, T\} \) is an another observed vector, the cross periodogram matrix between \( a_t \) and \( b_t \) is
\[ I_{ab}(\lambda) = w_a(\lambda)w_b^*(\lambda) = I_{ab}^c(\lambda) + iI_{ab}^q(\lambda) \]
where the asterisk is transposed complex conjugation, and \( c, q \) indicate the co- and quadrature periodogram, respectively. The discretely averaged co-periodogram
\[ \hat{F}_{ab}(k, l) = \frac{2\pi}{T} \sum_{j=k}^{l} I_{ab}^c(\lambda_j), \quad 1 \leq k \leq l \leq T - 1 \]
for \( \lambda_j = 2\pi j/T \). Thus we obtain the frequency domain least squares (FDLS) estimator
\[ \hat{\beta}_m = \hat{F}_{xx}^{-1}(1, m) \hat{F}_{xy}(1, m) \tag{19} \]
of \( \beta \) in regression (18). If
\[ \frac{1}{m} + \frac{m}{T} \rightarrow 0 \quad \text{as} \quad T \rightarrow \infty \tag{20} \]
then \( \hat{\beta}_m \) is called a narrow-band FDLS estimator, since it uses only a degenerating band of frequencies around the origin. Robinson and Marinucci (2003) show, under some assumptions and 20,
\[ \hat{\beta}_{im} - \beta_i = O_p \left( \left( \frac{T}{m} \right)^{-d_e - d_i} \right), \quad i = 1, \ldots, p - 1 \quad \text{as} \quad T \rightarrow \infty. \]
Under fractional cointegration \( d_e < \min(d_i) \), so the estimator \( \hat{\beta}_m \) is consistent for \( \beta \). Moreover, if the integration order of the raw data series is common, i.e. \( d_i = d \) for all \( i = 1, \ldots, p \) the stochastic order of magnitude of the estimator varies with the strength of the cointegrating relation \( b = d - d_e \).

MDH prescribes full cointegration, in the sense that \( b = d \). A necessary condition for fractional cointegration is that the two largest orders of integration are equal.

A simple way to test the hypothesis of fractional cointegration is examining the degree of persistence of the residuals of the following regressions

\[
\log V_t = \beta_{RV} \log RV_t + \epsilon_t \quad (21)
\]

and

\[
\log V_t = \beta_{CTBPV} \log CTBPV_t + u_t \quad (22)
\]

under the assumption that volatility and volumes have the same \( d \).

The parameter \( \beta \) is estimated by the Frequency Domain Least Squares. The results are reported in table 6. The first part of the table shows the results for the regression 21 and 22. The residuals from the Narrow Band Least Square estimation of equations (21) and (22) show a high estimated order of integration, that is the residuals possess long memory. Given the standard errors we can reject the null hypotheses \( (22) \) show a high estimated order of integration, that is the residuals possess long memory. Since the presence or absence of cointegration is not known when the fractional integration order is estimated, they propose, as in Robinson and Yajima (2002), a test statistic for the equality of integration orders that is informative in both circumstances, in the bivariate case.

\[
\hat{T}_0 = m(Sd^\prime)^2 \left( S^{-\frac{1}{4}} \hat{\mathcal{D}}^{-1} (\hat{G} \circ \hat{G}) \hat{\mathcal{D}}^{-1} S^\prime + h(T)^2 \right)^{-1} (Sd) \quad (23)
\]

where \( \circ \) denotes the Hadamard product, \( S = [1, -1] \), \( h(T) = \log(T)^{-k} \) for \( k > 0 \), \( D = \text{diag}(G_{11}, G_{22}) \), while \( \hat{G} = \frac{1}{m} \sum_{j=1}^m \text{Re}(I_j) \) (see Nielsen and Shimotsu (2007) for more details). If the variables are not cointegrated, that is the cointegration rank \( r \) is zero, \( \hat{T}_0 \to \chi^2_1 \), while if \( r \geq 1 \), \( \hat{T}_0 \to 0 \). A significantly large value of \( \hat{T}_0 \), with respect to \( \chi^2_1 \), can be taken as an evidence against the equality of the integration orders.

Moreover, the estimation of the cointegration rank \( r \) is obtained by calculating the eigenvalues of the estimated matrix \( \hat{G} \). The estimator \( \hat{G} \) uses a new bandwidth parameter \( m_1 \).
Let $\hat{\delta}_i$, the $i$th eigenvalue of $\hat{G}$, it is possible to apply a model selection procedure to determine $r$. In particular,

$$\hat{r} = \arg \min_{u=0,1} L(u)$$

(24)

where

$$L(u) = v(T)(2-u) - \sum_{i=1}^{2-u} \hat{\delta}_i$$

(25)

for some $v(T) > 0$ such that

$$v(T) + \frac{1}{m_1^{1/2} v(T)} \rightarrow 0.$$  

(26)

Table 7 shows the results of the Nielsen and Shimotsu (2007) fractional cointegration analysis, with two different choices for the bandwidths, $m$ used in the estimation of $d$‘s in the exact local Whittle estimation, and $m_1$ used in the estimation of $L(u)$. The values of the fractional orders are close to that obtained with the GPH procedure. The $T_0$ statistic takes values 1.048 and 1.9716. Since the 95% critical value of a $\chi^2$ is 3.841 we do not reject the null of equality of the integration orders in both cases. The analysis of the cointegration rank confirms the absence of cointegration, in fact $\hat{r} = 0$ in all cases. This finding reinforces our belief against the idea of MDH theory as a long memory relationship.

6 The Model

Given the results of the fractional cointegration and tail dependence analysis, it is interesting to study the long run dependence of the two series and their interdependencies in a multivariate framework defined as a system of two equations:

$$\Phi(L)D(L)X_t = \epsilon_t$$

(27)

$$D(L) = \begin{bmatrix} (1-L)^{d_1} & 0 \\ 0 & (1-L)^{d_2} \end{bmatrix}$$

where $X_t = (log RV_t, log V_t)'$, $\Phi(L) = I_2 - \Phi_1 L - \ldots - \Phi_p L^p$, $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$, with $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = \Sigma$. This model is a Fractionally Integrated VAR (FIVAR). We assume that the $\epsilon_t$ have a joint distribution $\epsilon_t \sim G(\epsilon_t; \psi)$, with $G(.)$ continuous density function. The vector $\psi = (\varphi, \nu)$ contains the parameters of the conditional mean, variances and covariance ($\varphi$) and the nuisance parameters ($\nu$). We can specify the joint multivariate density by means of a copula function density. The copula theory provides an easy way to deal with the (otherwise) complex multivariate modeling. The essential idea of the copula approach is that a joint distribution can be factorized into the marginals and a dependence function called copula. The joint distribution $G(\epsilon_{1,t}, \epsilon_{2,t}; \psi)$ can be expressed as follows, thanks to the so-called Sklar’s theorem (1959):

$$(\epsilon_{1,t}, \epsilon_{2,t})' \sim G(\epsilon_{1,t}, \epsilon_{2,t}; \psi) = C(F_{1,t}(\epsilon_{1,t}; \delta_1), F_{2,t}(\epsilon_{2,t}; \delta_2); \gamma)$$

(28)

that is the joint distribution $G(.)$ of a vector of innovations $\epsilon_t$ is the copula $C(\cdot; \gamma)$ of the cumulative distribution functions of the innovations marginals $F_{1,t}(\epsilon_{1,t}; \delta_1)$ $F_{2,t}(\epsilon_{2,t}; \delta_2)$, where $\gamma, \delta_1, \delta_2$ are the copula and marginals parameters, respectively. Setting $u_1 = F_{1,t}(\epsilon_{1,t}; \delta_1)$ and $u_2 = F_{2,t}(\epsilon_{2,t}; \delta_2)$, the copula probability density function is defined as

$$c(u_1, u_2; \gamma) = \frac{\partial^2 C(u_1, u_2; \gamma)}{\partial u_1 \partial u_2}$$

(29)
7 Copula Modeling

The copula couples the marginal distributions together in order to form a joint distribution. The dependence relationship is entirely determined by the copula, while scaling and shape (mean, standard deviation, skewness, and kurtosis) are determined by the marginals (see Sklar (1959), Joe (1997) and Nelsen (1999). Cherubini, Luciano, and Vecchiato (2005) provide a detailed discussion of copula techniques for financial applications. Copulae can therefore be used to obtain more realistic multivariate densities than the traditional joint normal one, which is simply the product of a normal copula and normal marginals; marginals can be entirely general, e.g. Skewed Student’s $t$ marginals, while the normal dependence relation can be preserved using a normal copula.

7.1 Elliptical Copulae

The class of elliptical distributions provides useful examples of multivariate distributions because they share many of the tractable properties of the multivariate normal distribution. Furthermore, they allow to model multivariate extreme events and forms of non-normal dependencies. Elliptical copulae are simply the copulae of elliptical distributions (see Fang, Kotz, and Ng (1990) for a detailed treatment of elliptical distributions).

We present two copulae belonging to the elliptical family that will be later used in the empirical applications: the Gaussian and Student’s $t$-Copula.

1. The probability density function of the Gaussian copula is:

$$c(\Phi(x_1), \Phi(x_2)) = \frac{1}{|R|^{1/2}} \exp \left( -\frac{1}{2} \zeta^\prime (R^{-1} - I) \zeta \right)$$

where $\zeta = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n))^\prime$ is the vector of univariate normal inverse distribution functions, $u_i = \Phi(x_i)$, while $R$ is the correlation matrix.

2. On the other hand, the copula of the multivariate Student’s $t$ distribution is the Student’s $t$-Copula, and its density function is:

$$c(T_{\nu_c}(x_1), T_{\nu_c}(x_2)) = |R|^{-1/2} \left[ \frac{\Gamma \left( \frac{\nu_c+2}{2} \right)}{\Gamma \left( \frac{\nu_c}{2} \right)} \right] \left[ \frac{\Gamma \left( \frac{\nu_c+1}{2} \right)}{\Gamma \left( \frac{\nu_c}{2} \right)} \right] \left( 1 + \zeta^\prime \Sigma^{-1} \zeta \right)^{-\frac{\nu_c+2}{2}} \prod_{i=1}^n \left( 1 + \frac{\zeta^2_i}{\nu_c} \right)^{-\frac{\nu_c+1}{2}}$$

where $u_i = T_{\nu_c}(x_i)$ and $T_{\nu_c}(x_i)$ is the univariate Student’s $t$ cdf, $\zeta = (T_{\nu_c}^{-1}(u_1), T_{\nu_c}^{-1}(u_2))^\prime$ is the vector of univariate inverse distribution functions, $\nu_c$ are the degrees of freedom, and $R$ is the correlation matrix.

The Student’s $t$-copula generates symmetric tail dependence, i.e. lower and upper tail dependence are equal, while the normal copula generates zero tail dependence, instead.

7.2 Archimedean Copulae

An alternative to the elliptical copulae is the class of Archimedean copulae. Archimedean copulae provide analytical tractability and a large spectrum of different dependence measures. They present several advantages: the ease with which they can be constructed, the
large number of parametric families of copulae belonging to this class, the great variety of different dependence structures (see Embrechts, Lindskog, and McNeil (2003) and Joe (1997)).

Among the different Archimedean copulae, we will make use of the Gumbel copula:

$$C(u_1, u_2) = \exp \left\{ - \left[ (-\log u_1)^\theta + (-\log u_2)^\theta \right]^{\frac{1}{\theta}} \right\}$$  \hspace{1cm} (32)

where $\theta > 1$ is the copula parameter, whereas the density is given by

$$c(u_1, u_2) = C(u_1, u_2) \cdot u_1^{\theta - 1} u_2^{\theta - 1} \left[ (-\log u_1)^\theta + (-\log u_2)^\theta \right]^{-2+2/\theta}$$

$$\times \left\{ (\log u_1 \log u_2)^{\theta - 1} \times \{ 1 + (\theta - 1)(-\log u_1)^\theta + (-\log u_2)^\theta \}^{-\frac{1}{\theta}} \right\}$$

The degree of upper tail dependence for the Gumbel copula is equal to $2 - \frac{2}{\theta}$. This is a measure of dependence between random variables in the extreme upper joint tails. Broadly speaking, we can say that the upper tail dependence measures the probability of an extremely large positive realization in one covariate, given that we have observed a large positive realization in another.

We also use the Clayton (or Cook Johnson) copula, which corresponds to copula B4 in Joe (1997):

$$C(u_1, u_2) = \max \left[ \left( \sum_{i=1}^{2} u_i^{-\theta} - 1 \right)^{-1/\theta}, 0 \right]$$

when the copula parameter $\theta > 0$ the copula simplifies to

$$C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$$  \hspace{1cm} (33)

whereas the density is given by

$$c(u_1, u_2) = (1 + \theta)(u_1 u_2)^{-\theta - 1} \left( \sum_{i=1}^{2} u_i^{-\theta} - 1 \right)^{-\frac{1}{2} - \frac{1}{\theta} - 2}$$

It has positive lower tail dependence. This is a measure of dependence between random variables in the extreme lower joint tails. The Clayton copula implies a degree of tail dependence equal to $2(-1/\theta)$. See Joe (1997) and Cherubini, Luciano, and Vecchiato (2005) for more details.

### 7.3 Copula and Marginals Estimation

Let $\Theta = (\delta_1, \delta_2; \gamma)$ be the parameters vector to be estimated, where $\delta_i$, $i = 1, 2$ are the parameters of the marginal distribution $F_i$ and $\gamma$ is the vector of the copula parameters. It follows from (28) that the log-likelihood function for the joint conditional distribution $H_t(\cdot; \theta)$ is given by

$$l(\Theta) = \sum_{t=1}^{T} \log(c(F_1(x_{1,t}; \delta_1), F_2(x_{2,t}; \delta_2); \gamma)) + \sum_{t=1}^{T} \sum_{i=1}^{2} \log f_i(x_{i,t}; \delta_i, t).$$  \hspace{1cm} (34)

where $c$ is the copula density, whereas $f_i$ are the marginals densities. Hence, the log-likelihood of the joint distribution is just the sum of the log-likelihoods of the margins and
the log-likelihood of the copula. Standard ML estimates may be obtained by maximizing
the above expression with respect to the parameters \((\delta_1, \ldots, \delta_n; \gamma)\). In practice this can
involve a large numerical optimization problem with many parameters which can be difficult
to solve. However, given the partitioning of the parameter vector into separate parameters
for each margin and parameters for the copula, one may use (34) to break up the optimization
problem into several small optimizations, each with fewer parameters. This multi-step
procedure is known as the method of Inference Functions for Margins (IFM), see Joe and
Xu (1996) and Joe (1997). According to the IFM method, the parameters of the marginal
distributions are estimated separately from the parameters of the copula. In other words,
the estimation process is divided into the following two steps:

1. Estimate the parameters \(\delta_i, i = 1, \ldots, n\) of the marginal distributions \(F_i\) using the ML
   method:
   \[
   \hat{\delta}_i = \arg \max l^i(\delta_i) = \arg \max \sum_{t=1}^{T} \log f_i(x_{i,t}; \delta_i),
   \]
   where \(l^i\) is the log-likelihood function of the marginal distribution \(F_i\);

2. Estimate the copula parameters \(\gamma\), given the estimations performed in step 1):
   \[
   \hat{\gamma} = \arg \max l^c(\gamma) = \arg \max \sum_{t=1}^{T} \log (c(F_1(x_1,t; \hat{\delta}_1), F_2(x_n,t; \hat{\delta}_n); \gamma)),
   \]
   where \(l^c\) is the log-likelihood function of the copula.

Joe (1997) compares the efficiency of the IFM method relative to full maximum likelihood
for a number of multivariate models and finds the IFM method to be highly efficient. There-
fore, we think it is safe to use the IFM method and benefit from the huge reduction in
complexity it implies for the numerical optimization. The models are estimated with a con-
ditional maximum likelihood technique that considers the infinite AR representation of a
long memory process (see Beran (1994))\(^5\).

The log-likelihood functions for each model are:

- **NORMAL COPULA (NCOP):**
  \[
  l_t(\Theta) = \sum_{i=1}^{2} \log \left( \frac{\Gamma\left(\frac{\nu_i+1}{2}\right)}{\sqrt{\nu_i \pi} \Gamma\left(\frac{\nu_i}{2}\right)} \left( 1 + \frac{\epsilon_{1,t}^2}{\nu_i} \right)^{-\left(\frac{\nu_i+1}{2}\right)} \right) + \log \left( (1 - \rho^2)^{-0.5} \right) \\
  + \left( -\frac{1}{2} (1 - \rho^2)^{-1} (\epsilon_{1,t}^2 + \epsilon_{2,t}^2 - 2 \rho \epsilon_{1,t} \epsilon_{2,t}) \cdot \frac{1}{2} (\epsilon_{1,t}^2 + \epsilon_{2,t}^2) \right)
  \]

- **CLAYTON COPULA (CCOP):**
  \[
  l_t(\Theta) = \sum_{i=1}^{2} \log \left( \frac{\Gamma\left(\frac{\nu_i+1}{2}\right)}{\sqrt{\nu_i \pi} \Gamma\left(\frac{\nu_i}{2}\right)} \left( 1 + \frac{\epsilon_{1,t}^2}{\nu_i} \right)^{-\left(\frac{\nu_i+1}{2}\right)} \right) \\
  + \log \left( (1 + \theta)(u_{1,t} u_{2,t})^{-\theta-1} (u_{1,t}^{\theta} + u_{2,t}^{\theta} - 1)^{-2+\theta^{-1}} \right)
  \]

\(^5\)A preliminary analysis conducted on the filtered series suggest that the optimal lag choice should be 1 in
the VAR specification.
• **T-COPULA (TCOP):**

\[
l_t(\Theta) = \sum_{i=1}^{2} \log \left( \frac{\Gamma(\nu_i + 1)}{\sqrt{\nu_i \pi \Gamma(\nu_i / 2)}} \left( 1 + \frac{\epsilon_{i,t}^2}{\nu_i} \right)^{-\frac{\nu_i + 1}{2}} \right) \\
+ \log \left( \frac{\Gamma(\nu_{c} + 2)}{\Gamma(\nu_c / 2)^2} \right) \cdot |R|^{-\frac{1}{2}} \left( 1 + \frac{\zeta_t^2 R_{-1}^{c} \zeta_t}{\nu_c} \right)^{-\frac{\nu_{c} + 2}{2}} \\
+ \sum_{i=1}^{2} \log \left( 1 + \frac{\epsilon_{i,t}^2}{\nu_c} \right)^{(\frac{\nu_i + 1}{2})} 
\]

• **GUMBEL COPULA (GCOP):**

\[
l_t(\Theta) = \sum_{i=1}^{2} \log \left( \frac{\Gamma(\nu_i + 1)}{\sqrt{\nu_i \pi \Gamma(\nu_i / 2)}} \left( 1 + \frac{\epsilon_{i,t}^2}{\nu_i} \right)^{-\frac{\nu_i + 1}{2}} \right) \\
+ \log \left( C(u_1, u_2) \cdot u_1^{-1} u_2^{-1} \left[ (-\log u_1)^{\theta} + (-\log u_2)^{\theta} \right]^{-2+2/\theta} \left[ \log u_1 \log u_2 \right]^{\theta-1} \right) \\
+ \log \left\{ 1 + (\theta - 1) \left[ (-\log u_1)^{\theta} + (-\log u_2)^{\theta} \right]^{-\frac{1}{\theta}} \right\} 
\]

## 8 Estimation Results

The maximum likelihood estimates of \(d_{\logRV}\) and \(d_{\logV}\) are always close to the semi-parametric estimates obtained with the Geweke and Porter-Hudak (1984) estimator. The \(R^2\) are about 30% for both log-volumes and log-volatility. The estimated parameters, \(\phi_{ij}\), turns out to be statistically significant in the equation of the realized volatility, meaning that lagged filtered log-volumes give some information on the actual filtered realized log-volatility. This indicates that, once the long memory of the series is accounted for, volumes leads volatility. However, this finding contrasts the results in Luu and Martens (2003) that ascertain, in a VAR framework, a bidirectional Granger causality from realized volatility to volumes and in the other way round.

The copula estimates show a positive dependence: if we compute a common dependence measure -such as the Kendall’s tau - by using the parameters’ estimates, it ranges from 0.43 obtained with the \(t\)-copula up to 0.50 with the Gumbel copula. Instead the three copulae differ on the degree of tail dependence, that is dependence in the extremes: the Gumbel estimates are characterized by a strong upper tail dependence (0.5864), whereas the \(t\)-copula presents a lower value (0.3916). Clayton copula shows a strong positive lower tail dependence equal to 0.6746. The tail dependence coefficient is zero for the normal copula by construction. In a recent large scale simulation study, Fantazzini (2008) found that if the true marginals show positive skewness, then using symmetric marginals causes the Clayton parameter \(\alpha_c\) to be positively biased, thus overestimating the tail dependence coefficient.

These results are in accordance with the findings of the preliminary non parametric analysis which highlights positive upper and lower tail dependence. In particular, the tail dependence value associated with the \(t\)-copula model is very close to the one obtained with the Hill’s estimator. Besides, Kole, Koedijk, and Verbeek (2007), by using a new goodness-

---

6 The Granger causality test, given a VAR(1) model for our series, results in the acceptance of causality in both directions at 5% of significance. This results is robust to different choices of the lags of the VAR.
of-fit testing procedure, found that the Gaussian copula underestimates the probability of joint extreme downward movements, while the survival Gumbel copula overestimates this risk, and they provide evidence in favor of the Student’s t-copula.

9 Forecasts

An out-of-sample forecast exercise has been carried out in order to evaluate the ability of the model to predict one-period ahead. A rolling window of 2167 observations has been used for the parameter estimation and 100 for the one-period ahead forecast. As a benchmark provision, we adopt an extension of the bivariate HAR model, introduced by Corsi (2003). This simple model emphasizes the idea of heterogeneity among different financial investors on the financial markets. For this reason, Corsi (2003) suggests that the present volatility depends on the past daily, weekly and monthly realizations. We also include the volumes, so the extended bivariate HAR model is

\[
\log V_t = \omega_1 + \delta_1 \log RV_{t-1} + \delta_2 \log RV_{W_{t-1}} + \delta_3 \log RV_{M_{t-1}} + \psi_{11} \log V_{t-1} + \psi_{12} \log V_{W_{t-1}} + \psi_{13} \log V_{M_{t-1}} + \eta_{1t}
\]

\[
\log RV_t = \omega_2 + \delta_1 \log RV_{t-1} + \delta_2 \log RV_{W_{t-1}} + \delta_3 \log RV_{M_{t-1}} + \psi_{21} \log V_{t-1} + \psi_{22} \log V_{W_{t-1}} + \psi_{23} \log V_{M_{t-1}} + \eta_{2t}
\]

where \((\eta_{1t}, \eta_{2t})'\) is distributed as a bivariate normal with zero mean and variance and covariance matrix, \(\Gamma\), while \(\log RV_{W_{t-1}} = \frac{1}{5} \sum_{i=1}^{5} \log RV_{t-i}\) and \(\log RV_{M_{t-1}} = \frac{1}{22} \sum_{i=1}^{22} \log RV_{t-i}\), analogously for \(\log V_{W_{t-1}}\) and \(\log V_{M_{t-1}}\). The model is estimated by maximum likelihood. Table 10 reports the estimation results for the HAR model:

First, we compute the following loss functions:

- **Mean Squared Error**,

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (X_{t+i} - \hat{X}_{t+i|t+i-1})^2
\]  

(37)

- **Root Mean Squared Error, RMSE**,

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X_{t+i} - \hat{X}_{t+i|t+i-1})^2}
\]  

(38)

- **Mean Absolute Error**,

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |X_{t+i} - \hat{X}_{t+i|t+i-1}|
\]  

(39)

where \(\hat{X}_{t+i|t+i-1}\) is the one-period ahead model forecast and \(N\) is equal to 100. Table 11 reports the above statistics. We notice a mild forecasting superiority for the volatility of the bivariate ARFIMA with respect to the benchmark HAR model.

We also implement a Diebold and Mariano (1995) test, to directly compare the forecasting ability of the bivariate ARFIMA with respect to the HAR model. The Diebold-Mariano
test, in fact, is a statistic based on the difference between the loss functions of two alternative model forecasts. In the case of one-step ahead forecast, the Diebold-Mariano test reduces, under the null $l_i = l_*$, to

$$DM = \frac{1}{N} \frac{l_i - l_*}{s_i} \approx N(0, 1)$$

where $l_i$ is the loss function of the forecast relative to the $i$-th model, while $l_*$ is the loss function relative to the benchmark model; $s_i$ is the variance of the difference between the loss functions of the two competing models. The alternative hypothesis is $l_i \neq l_*$. Following Patton and Sheppard (2007), we select six alternative loss functions that weight differently the forecast errors of the respective models. These are

- **MSE-LOG** = $(\log X_{t+i} - \log \hat{X}_{t+i|t+i-1})^2$,
- **MAE-LOG** = $|\log X_{t+i} - \log \hat{X}_{t+i|t+i-1}|$;
- **MSE-SD** = $(\sqrt{X_{t+i}} - \sqrt{\hat{X}_{t+i|t+i-1}})^2$;
- **MAE-SD** = $|\sqrt{X_{t+i}} - \sqrt{\hat{X}_{t+i|t+i-1}}|$;
- **MSE-PROP** = $(\frac{\hat{X}_{t+i|t+i-1}}{X_{t+i}} - 1)^2$;
- **MAE-PROP** = $|\frac{\hat{X}_{t+i|t+i-1}}{X_{t+i}} - 1|$;

As shown in table 12, the Diebold-Mariano test highlights a forecasting superiority of the bivariate ARFIMA model, in fact the signs are always negative, and sometimes they are statistically different from 0. The difference is particularly significant when the proportional loss function is used.

We also test the the joint null hypothesis $\alpha = 0$ and $\beta = 1$ in a Mincer and Zarnowitz (1969) regression setup, where, $X_t$ is regressed on a constant and on the model forecast, $\hat{X}_{t|t-1}$

$$X_{t+h} = \alpha + \beta \hat{X}_{t+h|t+h-1} + v_t$$ (41)

As shown in table 13, the null hypothesis is $\alpha = 0 \cap \beta = 1$ cannot be rejected in all the cases under exam with the exception of the HAR model.

Tables 11, 12 and 13 illustrate the importance of accounting for the long memory property of volatility and volumes, in particular using a dynamic model that allows for fractional integration.

### 10 Model Simulations

In the previous section, we discuss the model estimation results in terms of goodness of fit and their interpretations in the copula framework. Now, through the use of simulations, we consider, for the different model specifications, the ability to account for the sample characteristics of the observed data (see table 14 and 15). According to the different copula specifications, we generate the model innovations from the corresponding bivariate distribution; 4267 observations from the estimated system are generated by our Monte Carlo
exercise, keeping only the last 2267 observations corresponding to the sample size of our
data. The first 2000 simulated observations serve as a burn-in period. Then we repeat

Figure 6: Simulated paths of log-Volatility and log-Volumes.

this simulation 1000 times, in order to obtain 1000 daily sample paths for the logarithmic
volumes and volatility. Figure 6 displays two simulated path for the logarithmic volumes
and volatility from the Normal copula model. The similarity with the observed series is
notable.

From the 1000 simulated path, we calculate the model-implied sample distribution for the
respective descriptive statistics. Table 14 and 15 report the descriptive statistics of log RV
and log V, respectively, and the 95% simulated confidence intervals. We also report actual
quantiles and simulated confidence interval.

For what concerns the log-realized volatility, nearly all of the sample statistics, including
all of the reported 0.01 to 0.99 quantiles, lie within the simulated confidence bands obtained
with FIVAR with copula densities. However, for all copula models the simulated confidence
intervals do not include the sample skewness. The same is true for the HAR model. Notice
that the confidence interval from the simulation of a bivariate HAR contain neither the
upper nor the lower empirical quantiles. The results are better in the case of log V.

Moreover, we explore the dynamic implications of the models, in terms of ability to account
for the hyperbolic rate of decay of the autocorrelation functions. Figure 7 shows the sam-
ple autocorrelations and the corresponding simulated 95% confidence bands. Our bivariate
long memory models, for log RV and log V, reproduce the highly significant and very slowly
decaying sample autocorrelations over longer multi-month lags.

These results show how our bivariate FIVAR well describes the dynamics of both volumes
and volatility. In fact, the long memory bivariate model is able to reproduce both the sam-
ple statistics and the long run dynamics of the observed data in particular when the joint
distribution is described by the copula.

11 Conclusions

This paper has focused on the relation between volatility and volumes. Thanks to the re-
cent developments on high-frequency based realized volatility, the former can be estimated
rather precisely from the high frequency returns. We disentangle the realized volatility in
a continuous and jump component, showing that volumes are highly correlated with the
continuous part of volatility and that jumps are much less persistent than bipower varia-
tion and volumes. We also show that there exist a strong upper and lower tail dependence
between the volatility and volumes that is due to the presence of jumps. We don’t pro-
vide a specific model for jumps, but we investigate the long memory property of realized
volatility and volumes, showing that the two series have the same degree of fractional in-
tegration but they do not appear to be fractionally cointegrated, in the sense that a linear
combination of them does not reduce the degree of fractional integration. This finding is
not supportive of the presence of a common stochastic long memory informative process for
both volumes and volatility as in the MDH version of Bollerslev and Jubinski (1999).
Given the result of fractional integration and cointegration analysis, we propose and es-
timate a bivariate model for the volumes and realized volatility, which takes into account
their long memory pattern and their dependence. Different hypothesis for the joint mult-
variate density are investigated. In particular, we adopt different copulae for the joint
distribution of the logarithm of the realized volatility and volume. The whole system is
estimated with an efficient full information maximum likelihood technique with different
degrees of tail dependence. The evidence from the forecasting and simulation exercise
highlights the predictive ability of our bivariate model with respect to other competitive
models. Moreover, our model well account for the long run dynamics of both volatility and
volumes and their sample distribution.

References

Process in the Presence of Market Microstructure Noise,” NBER Working Papers 9611,
National Bureau of Economic Research, Inc.


Jump Components in the Measurement, Modeling and Forecasting of Return Volatility,”
Discussion paper.


BANDI, F., AND J. RUSSELL (2003): “Separating market microstructure noise from volatil-

volatility and its use in estimating stochastic volatility models,” Journal of the Royal
Statistical Society, Series B, 64, 253–280.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2004): “Power and Bipower Variation


BOLLERSLEV, T., AND D. JUBINSKI (1999): “Equity Trading Volume and Volatility: Lat-
tent Information Arrivals and Common Long-Run Dependencies,” Journal of Business


Table 4: Tail Dependence Analysis. This table reports the degree of positive and negative tail dependence, measured by the Hill estimator, of three different estimators of log-volatility, $\log RV$, $\log BPV$, and $\log CTBPV$, with log-volumes, $\log V_t$, for unfiltered and filtered series $(1 - L)^d y_t$. The parameter $d$ has been estimated exact local Whittle estimator with a bandwidth equal to 200. The table reports also the Pearson’s $\rho$. The threshold, $u$, has been chosen in order to leave on the right (left) the 2.5% of the observations.

<table>
<thead>
<tr>
<th></th>
<th>Unfiltered</th>
<th></th>
<th>Filtered</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Right Tail</td>
<td></td>
<td>Right Tail</td>
<td></td>
</tr>
<tr>
<td>$\log RV$</td>
<td>$\rho$</td>
<td>$\tilde{\chi}$</td>
<td>$\chi$</td>
<td>$\tilde{\chi}$</td>
</tr>
<tr>
<td></td>
<td>$0.6247$</td>
<td>$0.6700$</td>
<td>$0.2212$</td>
<td>$0.3622$</td>
</tr>
<tr>
<td></td>
<td>$0.6243$</td>
<td>$0.7655$</td>
<td>$0.2338$</td>
<td>$0.3306$</td>
</tr>
<tr>
<td></td>
<td>$0.6121$</td>
<td>$0.3682$</td>
<td>$0.1812$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$0.5876$</td>
<td>$0.6253$</td>
<td>$0.2153$</td>
<td>$0.2810$</td>
</tr>
<tr>
<td>$\log BPV$</td>
<td>$\rho$</td>
<td>$\tilde{\chi}$</td>
<td>$\chi$</td>
<td>$\tilde{\chi}$</td>
</tr>
<tr>
<td></td>
<td>$0.6247$</td>
<td>$0.6961$</td>
<td>$0.2246$</td>
<td>$0.3245$</td>
</tr>
<tr>
<td></td>
<td>$0.6243$</td>
<td>$0.2960$</td>
<td>$0.1716$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$0.6121$</td>
<td>$0.6954$</td>
<td>$0.2245$</td>
<td>$0.3245$</td>
</tr>
<tr>
<td>$\log TVR$</td>
<td>$\rho$</td>
<td>$\tilde{\chi}$</td>
<td>$\chi$</td>
<td>$\tilde{\chi}$</td>
</tr>
<tr>
<td></td>
<td>$0.6143$</td>
<td>$0.7136$</td>
<td>$0.2269$</td>
<td>$0.3622$</td>
</tr>
<tr>
<td></td>
<td>$0.6015$</td>
<td>$0.7684$</td>
<td>$0.2342$</td>
<td>$0.3858$</td>
</tr>
<tr>
<td></td>
<td>$0.4203$</td>
<td>$0.6037$</td>
<td>$0.1862$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\log CTBPV$</td>
<td>$\rho$</td>
<td>$\tilde{\chi}$</td>
<td>$\chi$</td>
<td>$\tilde{\chi}$</td>
</tr>
<tr>
<td></td>
<td>$0.6143$</td>
<td>$0.9023$</td>
<td>$0.2519$</td>
<td>$0.2904$</td>
</tr>
<tr>
<td></td>
<td>$0.6015$</td>
<td>$0.4181$</td>
<td>$0.1878$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$0.6037$</td>
<td>$0.9280$</td>
<td>$0.2553$</td>
<td>$0.2904$</td>
</tr>
</tbody>
</table>

Left Tail

<table>
<thead>
<tr>
<th></th>
<th>Unfiltered</th>
<th></th>
<th>Filtered</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left Tail</td>
<td></td>
<td>Left Tail</td>
<td></td>
</tr>
<tr>
<td>$\log RV$</td>
<td>$\rho$</td>
<td>$\tilde{\chi}$</td>
<td>$\chi$</td>
<td>$\tilde{\chi}$</td>
</tr>
<tr>
<td></td>
<td>$0.6247$</td>
<td>$0.6961$</td>
<td>$0.2246$</td>
<td>$0.3245$</td>
</tr>
<tr>
<td></td>
<td>$0.6243$</td>
<td>$0.2960$</td>
<td>$0.1716$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$0.6121$</td>
<td>$0.6954$</td>
<td>$0.2245$</td>
<td>$0.3245$</td>
</tr>
<tr>
<td>$\log BPV$</td>
<td>$\rho$</td>
<td>$\tilde{\chi}$</td>
<td>$\chi$</td>
<td>$\tilde{\chi}$</td>
</tr>
<tr>
<td></td>
<td>$0.6143$</td>
<td>$0.9023$</td>
<td>$0.2519$</td>
<td>$0.2904$</td>
</tr>
<tr>
<td></td>
<td>$0.6015$</td>
<td>$0.4181$</td>
<td>$0.1878$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>$0.6037$</td>
<td>$0.9280$</td>
<td>$0.2553$</td>
<td>$0.2904$</td>
</tr>
</tbody>
</table>
\[ m = T^{4/5} = 483 \quad m = T^{0.7} = 223 \quad m = T^{2/3} = 103 \]

| \( d_{\log RV} \) | 0.3707 | 0.3920 | 0.4476 |
| \( d_{\log V} \) | 0.3732 | 0.3520 | 0.3726 |
| \( \xi \) | 0.0097 | 2.3941 | 8.5162 |
| \( p\text{-value} \) | 0.9214 | 0.1217 | 0.0035 |

Table 5: Fractional integration estimation. \( \xi \) is the Robinson test statistic for \( H_0: d_{\log RV} = d_{\log V} \)

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>( \hat{\beta}_m )</th>
<th>( d_e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m = T - 1 )</td>
<td>0.3658</td>
<td>0.3754 (0.0315)</td>
</tr>
<tr>
<td>( m = 20 )</td>
<td>0.2697</td>
<td>0.3722 (0.0315)</td>
</tr>
<tr>
<td>( m = 15 )</td>
<td>0.2412</td>
<td>0.3683 (0.0315)</td>
</tr>
<tr>
<td>( m = 9 )</td>
<td>0.2362</td>
<td>0.3697 (0.0315)</td>
</tr>
<tr>
<td>( m = 6 )</td>
<td>0.2165</td>
<td>0.3725 (0.0315)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>( \hat{\beta}_m )</th>
<th>( d_u )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m = T - 1 )</td>
<td>0.3901</td>
<td>0.3912 (0.0315)</td>
</tr>
<tr>
<td>( m = 20 )</td>
<td>0.2829</td>
<td>0.3805 (0.0315)</td>
</tr>
<tr>
<td>( m = 15 )</td>
<td>0.2604</td>
<td>0.3767 (0.0315)</td>
</tr>
<tr>
<td>( m = 9 )</td>
<td>0.2496</td>
<td>0.3747 (0.0315)</td>
</tr>
<tr>
<td>( m = 6 )</td>
<td>0.2196</td>
<td>0.3686 (0.0315)</td>
</tr>
</tbody>
</table>

Table 6: Fractional Cointegration Analysis: the log-volumes \( \log V_t \) are regressed respectively on \( \log RV_t \) and \( \log CTBPV_t \), the bandwidth for the calculation of the fractional orders with GPH is \( T^{0.8} = 483 \). Standard errors in parentheses
<table>
<thead>
<tr>
<th></th>
<th>$m = T^{0.5} = 48$</th>
<th>$m = T^{0.6} = 103$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{\log RV}$</td>
<td>0.4314</td>
<td>0.4391</td>
</tr>
<tr>
<td></td>
<td>(0.0721)</td>
<td>(0.0492)</td>
</tr>
<tr>
<td>$d_{\log V}$</td>
<td>0.3317</td>
<td>0.3476</td>
</tr>
<tr>
<td></td>
<td>(0.0721)</td>
<td>(0.0492)</td>
</tr>
<tr>
<td>$T_0$</td>
<td>1.048</td>
<td>1.9716</td>
</tr>
</tbody>
</table>

Panel B

<table>
<thead>
<tr>
<th></th>
<th>$m_1 = T^{0.4} = 22$</th>
<th>$m_1 = T^{0.5} = 48$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_1$</td>
<td>0.0152</td>
<td>0.0106</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.0613</td>
<td>0.0557</td>
</tr>
<tr>
<td>$L(u)$</td>
<td>$v(T) = m_1^{-0.45}$</td>
<td>$v(T) = m_1^{-0.35}$</td>
</tr>
<tr>
<td>$m = 48, m_1 = 22$</td>
<td>$-1.4918$</td>
<td>$-1.3109$</td>
</tr>
<tr>
<td>$L(0)$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$L(1)$</td>
<td>$-1.3287$</td>
<td>$-1.2383$</td>
</tr>
<tr>
<td>$\hat{r}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$L(u)$</td>
<td>$v(T) = m_1^{-0.45}$</td>
<td>$v(T) = m_1^{-0.35}$</td>
</tr>
<tr>
<td>$m = 103, m_1 = 48$</td>
<td>$-1.6463$</td>
<td>$-1.4802$</td>
</tr>
<tr>
<td>$L(0)$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$L(1)$</td>
<td>$-1.4737$</td>
<td>$-1.3906$</td>
</tr>
<tr>
<td>$\hat{r}$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 7: Panel A: Fractional integration estimation with exact local Whittle estimator (standard error in parenthesis). The $T_0$ test statistic is calculated with $h(T) = \log(T)$. Panel B: Fractional cointegration estimation. The table reports the estimated eigenvalues ($\delta_i$) and the value of the function $L(u)$ for different choices of $m$ and $m_1$. 
<table>
<thead>
<tr>
<th></th>
<th>NORM</th>
<th>NCOP</th>
<th>TCOP</th>
<th>GCOP</th>
<th>CCOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{\log RV}$</td>
<td>0.4044^a</td>
<td>0.4014^a</td>
<td>0.4005^a</td>
<td>0.4081^a</td>
<td>0.3869^a</td>
</tr>
<tr>
<td>$d_{\log V}$</td>
<td>0.3765^a</td>
<td>0.3727^a</td>
<td>0.3788^a</td>
<td>0.3703^a</td>
<td>0.3888^a</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>$-0.0965^a$</td>
<td>$-0.1081^a$</td>
<td>$-0.0917^a$</td>
<td>$-0.1003^a$</td>
<td>$-0.0733^a$</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>0.1845^a</td>
<td>0.1597^a</td>
<td>0.1693^a</td>
<td>0.1858^a</td>
<td>0.1323^a</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.0249</td>
<td>0.0115</td>
<td>0.0110</td>
<td>0.0043</td>
<td>0.0217</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>0.1138^a</td>
<td>0.1034^a</td>
<td>0.1074^a</td>
<td>0.1175^a</td>
<td>0.0873^a</td>
</tr>
<tr>
<td>$\theta$</td>
<td>$-\alpha_c$</td>
<td>$-\nu$</td>
<td>$\rho$</td>
<td>$\nu_{rv}$</td>
<td>$\nu_{\nu}$</td>
</tr>
<tr>
<td>$P(Q_{15}^L &gt; q)$</td>
<td>0.3059</td>
<td>0.2566</td>
<td>0.2473</td>
<td>0.1814</td>
<td>0.1938</td>
</tr>
<tr>
<td>$P(Q_{15}^{LM} &gt; q)$</td>
<td>0.2977</td>
<td>0.2688</td>
<td>0.2780</td>
<td>0.2007</td>
<td>0.2285</td>
</tr>
</tbody>
</table>

Table 8: System Estimates with different copulae densities. $a,b$ and $c$ stands for 1%, 5% and 10% significance level of the corresponding $t$-ratio test. $P(Q_{15}^L > q)$ and $P(Q_{15}^{LM} > q)$ are the $p$-values of, respectively, the Portmanteau test by Lutkepohl (2005) and the Breush Godfrey LM-test for autocorrelation of the residuals.
Table 9: Kendall’s Tau and Tail Dependence.

<table>
<thead>
<tr>
<th>Copula</th>
<th>Kendall’s Tau</th>
<th>Tail Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORM</td>
<td>$\frac{\pi}{2} \cdot \arcsin(\rho)$</td>
<td>0.4311</td>
</tr>
<tr>
<td>TCOP</td>
<td>$\frac{\pi}{2} \cdot \arcsin(\rho)$</td>
<td>0.4298</td>
</tr>
<tr>
<td>GUMB</td>
<td>$1 - \theta$</td>
<td>0.4998</td>
</tr>
<tr>
<td>CLAY</td>
<td>$\alpha_c/(\alpha_c + 2)$</td>
<td>0.4627</td>
</tr>
</tbody>
</table>

Table 10: System Estimation for HAR model: $a, b$ and $c$ stands for 1%, 5% and 10% significance level of the corresponding t-ratio test. Bottom lines reports the Ljung-Box test statistic, $Q_{\eta}(10)$, for ten lags of both equation residuals, and the corresponding $p - values$ in parentheses.

<table>
<thead>
<tr>
<th>Copula</th>
<th>Realized Volatility</th>
<th>Volumes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>RMSE</td>
</tr>
<tr>
<td>NCOPE</td>
<td>0.0638</td>
<td>0.2526</td>
</tr>
<tr>
<td>TCOPE</td>
<td>0.0651</td>
<td>0.2551</td>
</tr>
<tr>
<td>GCOPE</td>
<td>0.0651</td>
<td>0.2551</td>
</tr>
<tr>
<td>CCPE</td>
<td>0.0651</td>
<td>0.2551</td>
</tr>
<tr>
<td>HAR</td>
<td>0.0660</td>
<td>0.2569</td>
</tr>
</tbody>
</table>

Table 11: Forecast Statistics.
<table>
<thead>
<tr>
<th></th>
<th>Realized Volatility</th>
<th>Volumes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE-LOG  MAE-LOG</td>
<td>MSE-SD MAE-SD</td>
</tr>
<tr>
<td>NCOP-HAR</td>
<td>-1.3353 -0.8938</td>
<td>-1.0704 -0.8762</td>
</tr>
<tr>
<td>TCOP-HAR</td>
<td>-1.3050 -0.6133</td>
<td>-0.9708 -0.6110</td>
</tr>
<tr>
<td>GCOP-HAR</td>
<td>-1.3552 -0.5552</td>
<td>-0.9698 -0.5849</td>
</tr>
<tr>
<td>CCOP-HAR</td>
<td>-0.9962 -0.6345</td>
<td>-0.8382 -0.5929</td>
</tr>
</tbody>
</table>

|                  | MSE-LOG  MAE-LOG    | MSE-SD MAE-SD   | MSE-PROP MAE-PROP |
| NCOP-HAR         | -0.5067 -0.9550     | -0.1125 -0.7530 | -0.7870 -1.4623   |
| TCOP-HAR         | -0.8561 -1.1618     | -0.3052 -0.9674 | -2.1527 -2.0041   |
| GCOP-HAR         | -0.7130 -1.1646     | -0.3075 -0.9774 | -1.0727 -1.6744   |
| CCOP-HAR         | -0.7188 -0.9813     | -0.1359 -0.7870 | -2.5329 -2.0221   |

Table 12: Diebold-Mariano Test. \(a, b\) and \(c\) stands for 1%, 5% and 10% significance level of the test.

<table>
<thead>
<tr>
<th></th>
<th>Realized Volatility</th>
<th>Volumes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\alpha) (\beta)</td>
<td>(\alpha = 0 \cap \beta = 1)</td>
</tr>
<tr>
<td>HAR</td>
<td>0.1149 0.7648</td>
<td>2.2680</td>
</tr>
<tr>
<td>NCOP</td>
<td>0.1022 0.8050</td>
<td>0.9224(^a)</td>
</tr>
<tr>
<td>TCOP</td>
<td>0.1094 0.7935</td>
<td>1.0240(^a)</td>
</tr>
<tr>
<td>GCOP</td>
<td>0.1150 0.7890</td>
<td>1.0147(^a)</td>
</tr>
<tr>
<td>CCOP</td>
<td>0.0944 0.8102</td>
<td>1.0186(^a)</td>
</tr>
</tbody>
</table>

Table 13: Mincer-Zarnowitz Regression. \(a, b\) and \(c\) stands for 1%, 5% and 10% significance level of the test.
\begin{table}[h!]
\centering
\begin{tabular}{l|c|cc|cc|cc|cc|cc}
\hline
 & \textbf{log RV} & \textbf{Clayton} & \textbf{Gumbel} & \textbf{Normal} & \textbf{T-copula} & \textbf{HAR} \\
\hline
\textbf{Statistics} & \textbf{log RV} & 95\% Intervals & 95\% Intervals & 95\% Intervals & 95\% Intervals & 95\% Intervals \\
\hline
Mean & 0 & -0.1441 & 0.1551 & -0.1537 & 0.1534 & -0.1462 & 0.1593 & -0.1591 & 0.1510 & -0.1472 & 0.1055 \\
Std.Dev & 0.7169 & 0.6507 & 0.7594 & 0.6844 & 0.7993 & 0.6774 & 0.7869 & 0.6832 & 0.7944 & 0.6701 & 0.7446 \\
Skewness & 0.6516 & -0.3628 & 0.3487 & -0.5568 & 0.3080 & -0.3281 & 0.3143 & -0.3070 & 0.2798 & -0.1315 & 0.1280 \\
Excess kurtosis & 1.6067 & 0.7017 & 5.3135 & 0.8518 & 70.670 & 0.6323 & 4.1845 & 0.5715 & 3.4684 & -0.1900 & 0.2074 \\
$\theta_{0.01}$ & -1.4604 & -1.9675 & -1.4791 & -21.266 & -15.979 & -2.0490 & -1.5460 & -2.0675 & -1.5656 & -0.7160 & -0.6161 \\
$\theta_{0.05}$ & -1.0587 & -1.2892 & -0.9353 & -13.756 & -0.9808 & -1.3387 & -0.9701 & -1.3704 & -0.9933 & -0.6053 & -0.4727 \\
$\theta_{0.10}$ & -0.8413 & -1.0010 & -0.6821 & -10.732 & -0.7145 & -1.0455 & -0.7074 & -1.0766 & -0.7277 & -1.0438 & -0.7053 \\
$\theta_{0.50}$ & -0.054 & -0.1445 & 0.1450 & -0.1441 & 0.1579 & -0.1459 & 0.1561 & -0.1526 & 0.1493 & -0.0572 & 0.0568 \\
$\theta_{0.90}$ & 0.9332 & 0.6774 & 1.0236 & 0.7154 & 10.609 & 0.7172 & 1.0680 & 0.7143 & 1.0636 & 0.4715 & 0.3409 \\
$\theta_{0.95}$ & 1.2805 & 0.9387 & 1.3106 & 0.9763 & 13.592 & 0.9784 & 1.3618 & 0.9830 & 1.3694 & 0.6169 & 0.7589 \\
$\theta_{0.99}$ & 1.9777 & 1.4822 & 1.9966 & 15.618 & 20.646 & 1.5454 & 2.0560 & 1.5461 & 2.0682 & 0.8856 & 1.0704 \\
\hline
\end{tabular}
\caption{Simulation intervals for log RV statistics}
\end{table}
<table>
<thead>
<tr>
<th>Statistics</th>
<th>Mean</th>
<th>Std.Dev</th>
<th>Skewness</th>
<th>Excess kurtosis</th>
<th>$q_{0.01}$</th>
<th>$q_{0.05}$</th>
<th>$q_{0.10}$</th>
<th>$q_{0.50}$</th>
<th>$q_{0.90}$</th>
<th>$q_{0.95}$</th>
<th>$q_{0.99}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0.4198</td>
<td>0.2235</td>
<td>1.1686</td>
<td>-0.9953</td>
<td>-0.6438</td>
<td>-0.488</td>
<td>-0.0153</td>
<td>0.527</td>
<td>0.716</td>
<td>1.106</td>
</tr>
<tr>
<td></td>
<td>log $V_t$ 95% Intervals</td>
<td>log $V_t$ 95% Intervals</td>
<td>log $V_t$ 95% Intervals</td>
<td>log $V_t$ 95% Intervals</td>
<td>log $V_t$ 95% Intervals</td>
<td>log $V_t$ 95% Intervals</td>
<td>log $V_t$ 95% Intervals</td>
<td>log $V_t$ 95% Intervals</td>
<td>log $V_t$ 95% Intervals</td>
<td>log $V_t$ 95% Intervals</td>
<td>log $V_t$ 95% Intervals</td>
</tr>
<tr>
<td>Clayton</td>
<td>-0.0988</td>
<td>0.1040</td>
<td>-0.0858</td>
<td>0.0865</td>
<td>-0.0861</td>
<td>0.0904</td>
<td>-0.0924</td>
<td>0.0913</td>
<td>-0.0583</td>
<td>0.0548</td>
<td></td>
</tr>
<tr>
<td>Gumbel</td>
<td>0.4068</td>
<td>0.4818</td>
<td>0.4000</td>
<td>0.4644</td>
<td>0.3953</td>
<td>0.4620</td>
<td>0.4018</td>
<td>0.4687</td>
<td>0.3995</td>
<td>0.4411</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>-0.3965</td>
<td>0.3724</td>
<td>-0.3400</td>
<td>0.3116</td>
<td>-0.3208</td>
<td>0.2827</td>
<td>-0.2977</td>
<td>0.2924</td>
<td>-0.1182</td>
<td>0.1253</td>
<td></td>
</tr>
<tr>
<td>T-copula</td>
<td>0.5295</td>
<td>0.4818</td>
<td>0.5311</td>
<td>0.5125</td>
<td>0.5254</td>
<td>0.3934</td>
<td>0.4255</td>
<td>2.9749</td>
<td>0.1996</td>
<td>0.2020</td>
<td></td>
</tr>
<tr>
<td>HAR</td>
<td>48.418</td>
<td>0.5295</td>
<td>48.418</td>
<td>0.5295</td>
<td>48.418</td>
<td>0.5295</td>
<td>48.418</td>
<td>0.5295</td>
<td>48.418</td>
<td>0.5295</td>
<td></td>
</tr>
</tbody>
</table>

Table 15: Simulation intervals for $\log V$ statistics
Figure 7: Simulated ACF confidence intervals of volatility and volumes.