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

January 7, 2009

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 FIVAR 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 FIVAR 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.

J.E.L. classification. C32, C13, G1.

*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 Hypotesis* (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 explicity 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 necessarly 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

[1] See 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 Fractionally Integrated VAR (FIVAR).
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 FIVAR specification with copulae densities. Finally, a simulation exercise is carried out to evaluate the ability of the bivariate FIVAR model, with different copulae specifications, to account for some sample statistics. The evidence from the estimation and simulation results highlight the ability of the FIVAR 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 forecasts 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:

$$dp_t = \mu_t dt + \sigma_t dW_t \tag{1}$$

that describes the trajectories of a semimartingale in continuous time. W_t is the Wiener process at time t, while σ_t is called *spot volatility*. The *integrated volatility*, for day t, is

defined as the integral of the spot volatility

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

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

$$[Y_t] = p \lim \sum_{j=1}^{t_j<t} (Y_{t_j} - Y_{t_{j-1}})^2 \tag{3}$$

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 *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,

$$p_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + \sum_{j=1}^{Q(s)} k(s_j) \tag{4}$$

where the mean process μ_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

$$[p]_t = plim \sum_{j=1}^{t_j<t} (p_{t_j} - p_{t_{j-1}})^2 = \int_0^t \sigma_s^2 ds + \sum_{j=1}^{Q(s)} k^2(s_j) = IV_t + \sum_{j=1}^{Q(s)} k^2(s_j) \tag{5}$$

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, ${r_{t,j}}^2$

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2 \quad t = 1, ..., T \tag{6}$$

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

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

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:

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

where θ 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:

$$\theta_t = c_\theta \cdot h_t$$

where h_t is the conditional variance obtained by the estimation of a GARCH model. The parameter c_θ 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

$$CTBPV_t = \frac{\pi}{2} \sum_{j=2}^{M} Z_1(r_{t,j}, \theta_j) Z_1(r_{t,j-1}, \theta_{j-1}) \quad t = 1, ..., T \tag{9}$$

where $Z_1(r_{t,j}, \theta_j)$ 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 θ_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

$$\begin{aligned} CTBPV_t &\to BPV \quad \text{as} \quad c_\theta \to \infty \\ TRV_t &\to RV \quad \text{as} \quad c_\theta \to \infty. \end{aligned}$$

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, $r_{i,t}$, 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 intradaiy 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.

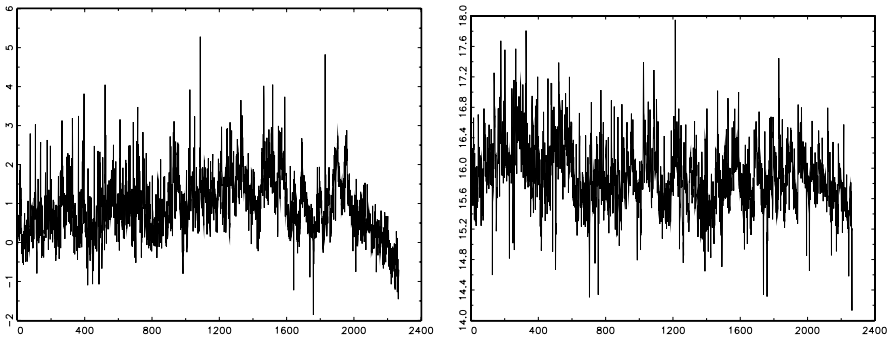

Figure 1: $\log RV_t$ and $\log V_t$

	Mean	*Std. Dev.*	*Skewness*	*Kurtosis*
r_t	-0.0002	0.0192	0.2562	4.1045
$\log RV_t$	0.000	0.7169	0.6516	1.6067
$\log BPV_t$	0.000	0.7203	-0.2677	0.6299
$\log TVR_t$	0.000	0.7025	0.4984	0.8729
$\log CTBPV_t$	0.000	0.6324	-0.0279	0.3597
$\log V_t$	0.000	0.4198	0.2235	1.1686

Table 1: Descriptive statistics of IBM returns (r_t), log-realized volatility ($\log RV_t$), log-bipower variation ($\log BPV_t$), log-threshold realized volatility ($\log TRV_t$), log-corrected threshold bipower variation ($\log CTBPV_t$), jump component (J_t) and log-volume ($\log V_t$) of IBM returns.

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:

$$U = F_{\log RV}(\log RV) \quad V = F_{\log V}(\log V)$$

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. χ 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 $Pr\{V > u | U > u\}$:

$$\overline{\chi}(u) = \frac{2 \log Pr\{U > u\}}{\log Pr\{U > u, V > u\}} - 1$$

	ρ_1	ρ_2	ρ_3	ρ_4	$BP(5)$	$BP(10)$	$BP(40)$
r_t	-0.0158	0.0016	0.0125	0.0198	5.94	12.49	54.74
$\log RV_t$	0.5638	0.4677	0.4160	0.3918	2260.1	3318.3	5756.1
$\log BPV_t$	0.5870	0.4882	0.4384	0.4111	2482.6	3673.2	6370.7
$\log TVR_t$	0.5800	0.4827	0.4287	0.4053	2411.9	3594.0	6305.8
$\log CTBPV_t$	0.5875	0.4996	0.4607	0.4351	2633.7	4006.3	7643.6
J_t	0.0628	0.0255	0.0269	0.0228	13.56	15.19	27.83
$\log V_t$	0.5638	0.4677	0.4160	0.3917	2023.6	2591.2	3684.0

Table 2: Sample autocorrelation function (ρ_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 ($\log TRV_t$), log corrected threshold bipower variation ($\log CTBPV_t$), jump component (J_t) and log-volume ($\log V_t$) of IBM returns.

	r_t	$\log BPV_t$	$\log V_t$	J_t	$\log r_t^2$	$\log RV_t$	$\log CTBPV_t$	TVR_t
r_t	1.0000	0.0024	0.0218	0.0692	0.0096	0.0218	0.0258	0.0316
$\log BPV_t$		1.0000	0.6243	0.2927	0.2469	0.9536	0.9543	0.9598
$\log V_t$			1.0000	0.2671	0.2213	0.6247	0.5876	0.6121
J_t				1.0000	0.1292	0.4466	0.2064	0.3171
$\log r_t^2$					1.0000	0.2610	0.2310	0.2554
$\log RV_t$						1.0000	0.9374	0.9855
$\log CTBPV_t$							1.0000	0.9022
$\log TVR_t$								1.0000

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 ($\log r_t^2$), log-realized volatility ($\log RV_t$), log-corrected threshold bipower variation ($\log CTBPV_t$), and log-threshold realized volatility ($\log TRV_t$)

and

$$\overline{\chi} = \lim_{u \to 1} \overline{\chi}(u)$$

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

In order to estimate $\overline{\chi}$ 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 $\log V$,

$$S = -1/\log F(\log RV) \quad T = -1/\log F(\log V) \tag{11}$$

Following Poon, Rockinger, and Tawn (2004), we calculate the Hill estimator, that is a non

parametric measure of the degree of tail dependence between volatility and volumes,

$$\widehat{\bar{\chi}} = \frac{2}{n_u}\left(\sum_{j=1}^{n_u}\log\left(\frac{z_j}{u}\right)\right) - 1 \tag{12}$$

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:

$$Var[\widehat{\bar{\chi}}] = \frac{(\widehat{\bar{\chi}}+1)}{n_u}$$

If there is evidence that $\widehat{\bar{\chi}} < 1$, $\widehat{\bar{\chi}} + 1.96\sqrt{Var[\widehat{\bar{\chi}}]} < 1$ 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

$$\widehat{\chi} = \frac{u \cdot n_u}{T} \tag{13}$$

with variance,

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

The parameter χ 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, $\hat{\bar{\chi}}$, 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.

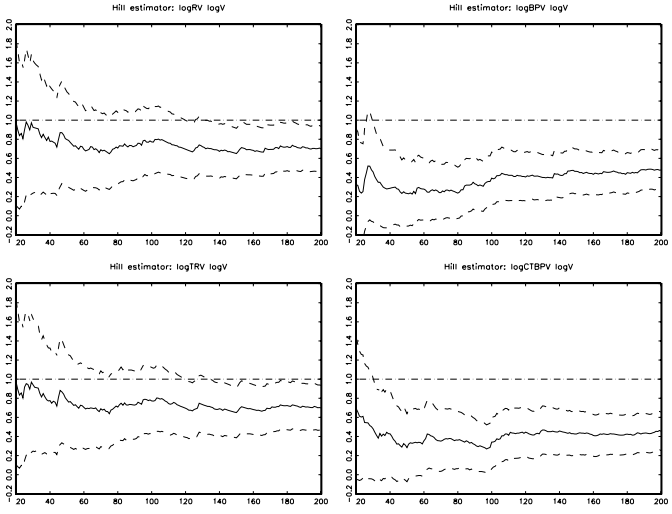

Figure 2: Hill estimator of the left tail dependence. X-axis measures $n_u = 20, ..., 200$.

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 $n_u = 57$,

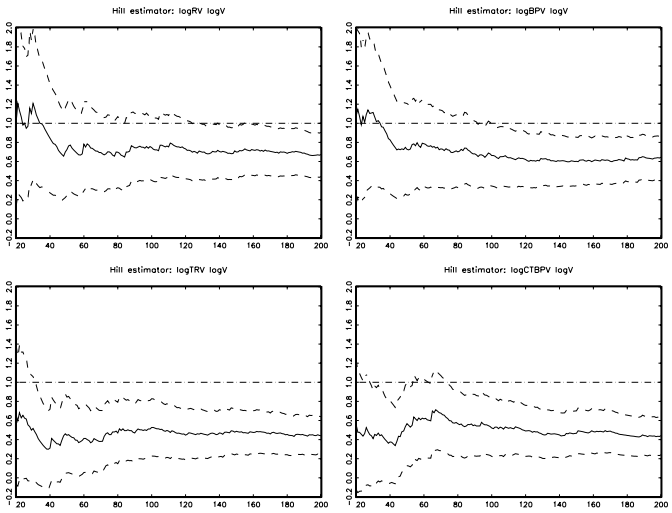

Figure 3: Hill estimator of the right tail dependence, X-axis measures $n_u = 20, ..., 200$.

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 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

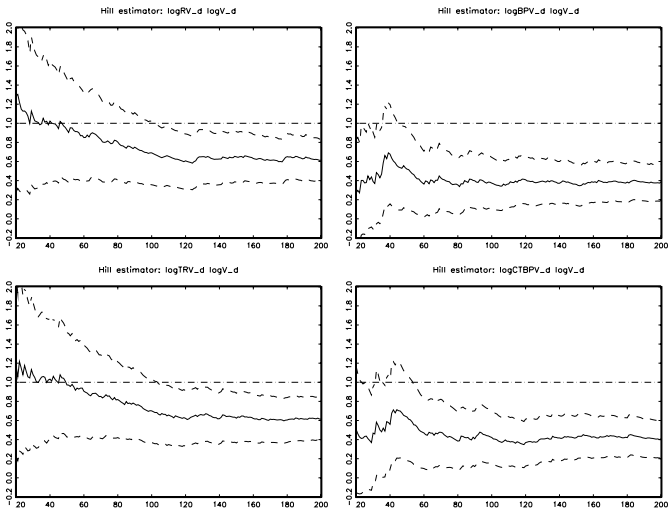

Figure 4: Hill estimator of the left tail dependence of the fractionally differenced series. X-axis measures $n_u = 20, ..., 200$.

$\log TRV$ 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:

$$(1 - L)^d y_t = \epsilon_t \quad t = 1, ..., T \tag{14}$$

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

$$(1 - l)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j - d)}{\Gamma(-d)\Gamma(j + 1)} L^j, \quad \Gamma(z) = \int_0^{\infty} t^{z-1} \mathrm{e}^{-t} dt.$$

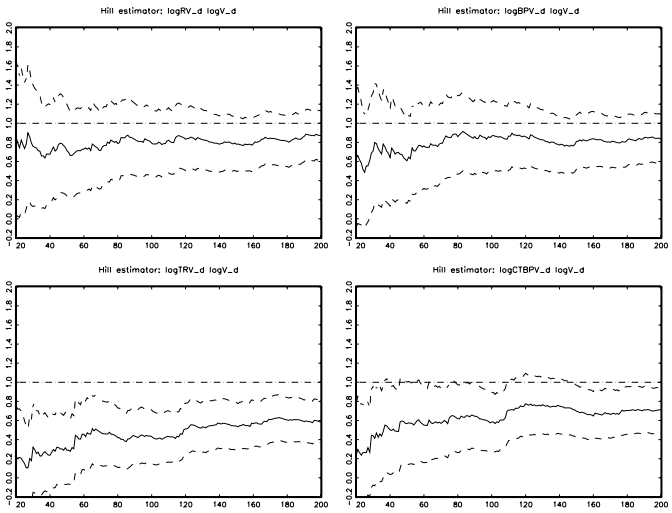

Figure 5: Hill estimator of the right tail dependence of the fractionally differenced series, X-axis measures $n_u = 20, ..., 200$.

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 antipersitent 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_{logRV} and d_{logV}, 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). The results of the estimation are presented in the table 5.[4] 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

[4]The simulation results of Hurvich, Deo, and Brodsky (1998) suggest to choose $m = T^{4/5}$ for the bandwidth.

as a χ^2_1 under the null $d_1 = d_2$, is

$$\xi = \frac{(d_1 - d_2)^2}{z'(X'X)^{-1}z \cdot f'\hat{\Omega}f} \tag{15}$$

where

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

ϵ_1 and ϵ_2 are the residuals from the log periodogram regression by Geweke and Porter-Hudak (1984) and $\lambda_j = \frac{2\pi j}{T}$ are the discrete Fourier frequencies. 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 α 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 $d_1 = \ldots = d_p = d$ it is usual to write $z_t \equiv CI(d, b)$, $b = d - d_e$. A typical situation is when $z_t = (x_t', y_t) \in I(d)$ and $e_t \in I(d_e)$ with $d > d_e \geq 0$ in the model

$$y_t = \beta' x_t + e_t. \tag{16}$$

Cointegration is commonly thought of as a stationary relation between nonstationary variables

$$d_i \geq \frac{1}{2} \quad \forall i \quad \text{and} \quad d_e < \frac{1}{2}.$$

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 $d > 0$, $d_e \geq 0$ 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 semiparametric 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 (16) relation and for arbitrary short run dynamics. Define the discrete Fourier transform of an observed vector $\{a_t,\, t = 1, \ldots, T\}$

$$w_a(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} a_t \mathrm{e}^{it\lambda}.$$

If $\{b_t,\, t = 1, \ldots, T\}$ is an another observed vector, the cross periodogram matrix between a_t and b_t is

$$I_{ab}(\lambda) = w_a(\lambda) w_b^*(\lambda) = I_{ab}^c(\lambda) + \mathrm{i} I_{ab}^q(\lambda)$$

where the asterisk is transposed complex conjugation, and c, q indicate the co- and quadrature periodogram, respectively. The discretely averaged co-periodogram

$$\widehat{F}_{ab}(k,l) = \frac{2\pi}{T} \sum_{j=k}^{l} I_{ab}^c(\lambda_j), \quad 1 \leq k \leq l \leq T-1$$

for $\lambda_j = 2\pi j/T$. Thus we obtain the frequency domain least squares (FDLS) estimator

$$\widehat{\beta}_m = \widehat{F}_{xx}^{-1}(1,m) \widehat{F}_{xy}(1,m) \tag{17}$$

of β in regression (16). If

$$\frac{1}{m} + \frac{m}{T} \to 0 \quad \text{as} \quad T \to \infty \tag{18}$$

then $\widehat{\beta}_m$ 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 18,

$$\widehat{\beta}_{im} - \beta_i = O_p\left(\left(\frac{T}{m} \right)^{d_e - d_i} \right), \quad i = 1, \ldots, p-1 \quad \text{as} \quad T \to \infty.$$

Under fractional cointegration $d_e < \min(d_i)$, so the estimator $\widehat{\beta}_m$ is consistent for β. 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

$$\log V_t = \beta_{RV} \log RV_t + e_t \tag{19}$$

and

$$\log V_t = \beta_{CTBPV} \log CTBPV_t + u_t \tag{20}$$

under the assumption that volatility and volumes have the same d.
The parameter β 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 19 and 20. The residuals from the Narrow Band Least Square estimation of equations (19) and (20) show a high estimated order of integration, that is the residuals possess long memory. Given the standard errors we can reject the null hypotheses $d_e = 0$ and $d_u = 0$. The estimates of the fractional integration parameters, $\widehat{d_e}$ and $\widehat{d_u}$, are not inferior to that of volumes. This means that is not $d_e < d$ or $d_u < d$. In particular, the 95% confidence intervals, obtained from the standard error of the parameters $\widehat{d_e}$ and $\widehat{d_u}$, contain the estimated fractional integration parameter, $\hat{d}_{\log V}$, of log-volumes. Moreover, considering different bandwidths, the estimates of β_{RV} and β_{CTBPV} turn out to be much smaller than 1. Thus we can conclude that volumes and volatility are not fractionally cointegrated. We also consider the possibility of fractional cointegration carrying out a test based analysis. Breitung and Hassler (2002) propose a trace test statistic, similar to that proposed by Johansen (1988), based on the solution of a generalized eigenvalue problem. However, they consider only the cointegration relation between non-stationary variables, such that $d > 0.5$. Instead, Robinson and Yajima (2002) discuss a semi-parametric procedure for determining the cointegration rank, focusing on stationary series. Nielsen and Shimotsu (2007) extend the analysis of Robinson and Yajima (2002), in order to consider cointegration for both stationary and non-stationary variables. In particular, they apply the exact local Whittle analysis, and estimate the rank of spectral cointegration of the dth differenced process around the origin. 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

$$\widehat{T}_0 = m(S\hat{d})' \left(S\frac{1}{4}\hat{D}^{-1}(\hat{G} \odot \hat{G})\hat{D}^{-1}S' + h(T)^2 \right)^{-1} (S\hat{d}) \tag{21}$$

where $\odot$ denotes the Hadamard product, $S = [1, -1]'$, $h(T) = \log(T)^{-k}$ for $k > 0$, $D = \text{diag}(G_{11}, G_{22})$, while $\hat{G} = \frac{1}{m}\sum_{j=1}^{m} Re(I_j)$ (see Nielsen and Shimotsu (2007) for more details). If the variables are not cointegrated, that is the cointegration rank r is zero, $\widehat{T}_0 \to \chi^2_1$, while if $r \geq 1$, $\widehat{T}_0 \to 0$. A significantly large value of $\widehat{T}_0$, with respect to χ^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 ith eigenvalue of $\hat{G}$, it is possible to apply a model selection procedure to determine r. In particular,

$$\hat{r} = \arg\min_{u=0,1} L(u) \tag{22}$$

where

$$L(u) = v(T)(2-u) - \sum_{i=1}^{2-u} \hat{\delta}_i \tag{23}$$

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

$$v(T) + \frac{1}{m_1^{1/2} v(T)} \to 0. \tag{24}$$

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 $\widehat{T}_0$ statistic takes values 1.048 and 1.9716. Since the 95% critical value of a χ^2_1 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:

$$\Phi(L)D(L)X_t = \epsilon_t \tag{25}$$

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

where $X_t = (\log RV_t, \log V_t)'$, $\Phi(L) = I_2 - \Phi_1 L - ... - \Phi_p L^p$, $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})'$, with $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = \Sigma$. This model is a Fractionally Integrated VAR (FIVAR). We assume that the ϵ_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 (φ) and the nuisance parameters (ν). 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):

$$(\epsilon_{1,t}, \epsilon_{2,t})' \sim G(\epsilon_{1,t}, \epsilon_{2,t}; \psi) = C(F_{1,t}(\epsilon_{1,t}; \delta_1), F_{2,t}(\epsilon_{2,t}; \delta_2); \gamma) \tag{26}$$

that is the joint distribution $G(.)$ of a vector of innovations ϵ_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 $u_2 = F_{1,t}(\epsilon_{2,t}; \delta_2)$, the copula probability density function is defined as

$$c(u_1, u_2; \gamma) = \frac{\partial^2 C(u_1, u_2; \gamma)}{\partial u_1 \partial u_2} \tag{27}$$

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:

$$c(\Phi(x_1), \Phi(x_2)) = \frac{1}{|R|^{1/2}} \exp\left(-\frac{1}{2}\zeta'(R^{-1} - I)\zeta\right) \tag{28}$$

   where $\zeta = \left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\right)'$ 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:

$$c(T_{\nu_c}(x_1), T_{\nu_c}(x_2)) = |R|^{-1/2} \frac{\Gamma\left(\frac{\nu_c+2}{2}\right)}{\Gamma\left(\frac{\nu_c}{2}\right)} \left[\frac{\Gamma\left(\frac{\nu_c}{2}\right)}{\Gamma\left(\frac{\nu_c+1}{2}\right)}\right]^2 \frac{\left(1 + \frac{\zeta'\Sigma^{-1}\zeta}{\nu_c}\right)^{-\frac{\nu_c+2}{2}}}{\prod_{i=1}^{2}\left(1 + \frac{\zeta_i^2}{\nu_c}\right)^{-\frac{\nu_c+1}{2}}} \tag{29}$$

   where $u_i = T_{\nu_c}(x_i)$ and $T_{\nu_c}(x_i)$ is the univariate Student's *t* cdf, $\zeta = (T_{\nu_c}^{-1}(u_1), T_{\nu_c}^{-1}(u_2))'$ is the vector of univariate inverse distribution functions, ν_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*:

$$C(u_1, u_2) = \exp\left\{-\left[(-\log u_1)^\theta + (-\log u_2)^\theta\right]^{\frac{1}{\theta}}\right\} \tag{30}$$

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

$$\begin{aligned} c(u_1, u_2) &= C(u_1, u_2) \cdot u_1^{-1} u_2^{-1} \left[(-\log u_1)^\theta + (-\log u_2)^\theta \right]^{-2+2/\theta} \\ &\quad [\log u_1 \log u_2]^{\theta-1} \times \left\{ 1 + (\theta - 1)[(-\log u_1)^\theta + (-\log u_2)^\theta]^{-\frac{1}{\theta}} \right\} \end{aligned}$$

The degree of upper tail dependence for the Gumbel copula is equal to $2 - 2^{\frac{1}{\theta}}$. 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):

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

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

$$C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta} \tag{31}$$

whereas the density is given by

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

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 δ_i, $i = 1, 2$ are the parameters of the marginal distribution F_i and γ is the vector of the copula parameters. It follows from (26) that the log-likelihood function for the joint conditional distribution $H_t(\,\cdot\,; \theta)$ is given by

$$l(\Theta) = \sum_{t=1}^{T} \log(c(F_1(x_{1,t}; \delta_1), F_2(x_{2,t}; \delta_2); \gamma)) + \sum_{t=1}^{T} \sum_{i=1}^{2} \log f_i(x_{i,t}; \delta_{i,t}). \tag{32}$$

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 (32) 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 δ_i, $i = 1, \ldots, n$ of the marginal distributions F_i using the ML method:
$$\hat{\delta}_i = \arg\max l^i(\delta_i) = \arg\max \sum_{t=1}^{T} \log f_i(x_{i,t}; \delta_i), \tag{33}$$
where l^i is the log-likelihood function of the marginal distribution F_i;

2. Estimate the copula parameters γ, given the estimations performed in step 1):
$$\hat{\gamma} = \arg\max l^c(\gamma) = \arg\max \sum_{t=1}^{T} \log(c(F_1(x_{1,t}; \hat{\delta}_1), F_2(x_{n,t}; \hat{\delta}_n); \gamma)), \tag{34}$$
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. Therefore, 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 conditional maximum likelihood technique that considers the infinite AR representation of a long memory process (see Beran (1994))[5].
Following Palma (2007), both series are filtered by long memory considering the presample value equal to zero, so that the infinite autoregressive representation is expanded to

$$(1 - L)^{d_i} X_{i,t} = \sum_{j=0}^{t-1} \pi_j X_{i,t-j} \tag{35}$$

where $\pi_0 = 1$, $\pi_1 = -d$, $\pi_j = \frac{1}{j}\pi_{j-1}(j - 1 - d)$ for $j \geq 2$.
We choose as marginal distributions for ϵ_i the skew-*t* distribution (Azzalini and Capitanio (2003)). This corresponds to the transformation

$$\epsilon_i = \mu_i + Z_i/\sqrt{V}$$

where $Z \sim SN(0, \sigma_i^2, \alpha_i)$ and $V \sim \chi^2_\nu/\nu$. The density of ϵ_i is given by

$$\frac{2}{\sigma_i} t_{\nu_i}\left(\frac{\epsilon_i - \mu_i}{\sigma_i}; 1, \nu_i\right) T_{\nu_i}\left(\alpha_i \frac{\epsilon_i - \mu_i}{\sigma_i} \sqrt{\frac{\nu_i + 1}{\left(\frac{\epsilon_i - \mu_i}{\sigma_i}\right)^2 + \nu_i}}, 1, \nu_i + 1\right).$$

The log-likelihood functions for each model are:

[5] A preliminary analysis conducted on the filtered series suggest that the optimal lag choice should be 1 in the VAR specification.

- NORMAL COPULA (*NCOP*):

$$\begin{aligned} l_t(\Theta) &= \sum_{i=1}^{2} \log\left(\frac{2}{\sigma_i} t_{\nu_i}\left(\frac{\epsilon_i-\mu_i}{\sigma_i};1,\nu_i\right) + T_{\nu_i}\left(\alpha_i \frac{\epsilon_i-\mu_i}{\sigma_i}\sqrt{\frac{\nu_i+1}{\left(\frac{\epsilon_i-\mu_i}{\sigma_i}\right)^2+\nu_i}},1,\nu_i+1\right)\right) \\ &+ \left(-\frac{1}{2}(1-\rho^2)^{-1}(\epsilon_{1,t}^2+\epsilon_{2,t}^2-2\rho\epsilon_{1,t}\epsilon_{2,t})\cdot\frac{1}{2}(\epsilon_{1,t}^2+\epsilon_{2,t}^2)\right) \end{aligned}$$

where $t_{\nu_i}(.,1,\nu)$ and $T_{\nu_i}(.,1,\nu)$ are respectively the pdf and cdf of the t distribution with ν degrees of freedom and scale equal to 1. the parameter α measures the degree of skewness, while μ and σ are the location and scale parameters respectively.

- CLAYTON COPULA (*CCOP*):

$$\begin{aligned} l_t(\Theta) &= \sum_{i=1}^{2} \log\left(\frac{2}{\sigma_i} t_{\nu_i}\left(\frac{\epsilon_i-\mu_i}{\sigma_i};1,\nu_i\right) + T_{\nu_i}\left(\alpha_i \frac{\epsilon_i-\mu_i}{\sigma_i}\sqrt{\frac{\nu_i+1}{\left(\frac{\epsilon_i-\mu_i}{\sigma_i}\right)^2+\nu_i}},1,\nu_i+1\right)\right) \\ &+ \log\left((1+\theta)(u_{1,t}u_{2,t})^{-\theta-1}(u_{1,t}^{-\theta}+u_{2,t}^{-\theta}-1)^{-(2+\theta^{-1})}\right) \end{aligned}$$

- T-COPULA (*TCOP*):

$$\begin{aligned} l_t(\Theta) &= \sum_{i=1}^{2} \log\left(\frac{2}{\sigma_i} t_{\nu_i}\left(\frac{\epsilon_i-\mu_i}{\sigma_i};1,\nu_i\right) + T_{\nu_i}\left(\alpha_i \frac{\epsilon_i-\mu_i}{\sigma_i}\sqrt{\frac{\nu_i+1}{\left(\frac{\epsilon_i-\mu_i}{\sigma_i}\right)^2+\nu_i}},1,\nu_i+1\right)\right) \\ &+ \log\left(\frac{\Gamma\left(\frac{\nu_c+2}{2}\right)\Gamma\left(\frac{\nu_c}{2}\right)}{\Gamma\left(\frac{\nu_c+1}{2}\right)^2}\cdot|R|^{-\frac{1}{2}}\left(1+\frac{\zeta_t' R^{-1}\zeta_t}{\nu_c}\right)^{-\frac{\nu_c+2}{2}}\right) \\ &+ \sum_{i=1}^{2}\log\left(1+\frac{\zeta_{it}^2}{\nu_c}\right)^{\left(\frac{\nu_c+1}{2}\right)} \end{aligned}$$

- GUMBEL COPULA (*GCOP*):

$$\begin{aligned} l_t(\Theta) &= \sum_{i=1}^{2} \log\left(\frac{2}{\sigma_i} t_{\nu_i}\left(\frac{\epsilon_i-\mu_i}{\sigma_i};1,\nu_i\right) + T_{\nu_i}\left(\alpha_i \frac{\epsilon_i-\mu_i}{\sigma_i}\sqrt{\frac{\nu_i+1}{\left(\frac{\epsilon_i-\mu_i}{\sigma_i}\right)^2+\nu_i}},1,\nu_i+1\right)\right) \\ &+ \log\left(C(u_1,u_2)\cdot u_1^{-1}u_2^{-1}\left[(-\log u_1)^\theta+(-\log u_2)^\theta\right]^{-2+2/\theta}\left[\log u_1 \log u_2\right]^{\theta-1}\right) \\ &+ \log\left\{1+(\theta-1)[(-\log u_1)^\theta+(-\log u_2)^\theta]^{-\frac{1}{\theta}}\right\} \end{aligned}$$

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, ϕ_{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[6].
The parameter α_{rv} and α_v capture the positive skewness of the two series, in particular of realized volatility. 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 0.3477 obtained with the Clayton copula up to 0.4316 with the Normal 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 (0.4956), whereas the t-copula presents a lower value (0.1456). Clayton copula shows a strong positive lower tail dependence equal to 0.5289. 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 α_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-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

$$\begin{aligned}
\log V_t &= \omega_1 + \delta_{11}\log RV_{t-1} + \delta_{12}\log RV^W_{t-1} + \delta_{13}\log RV^M_{t-1} + \psi_{11}\log V_{t-1} \\
&\quad + \psi_{12}\log V^W_{t-1} + \psi_{13}\log V^M_{t-1} + \eta_{1t} \\
\log RV_t &= \omega_2 + \delta_{21}\log RV_{t-1} + \delta_{22}\log RV^W_{t-1} + \delta_{23}\log RV^M_{t-1} + \psi_{21}\log V_{t-1} \\
&\quad + \psi_{22}\log V^W_{t-1} + \psi_{23}\log V^M_{t-1} + \eta_{2t}
\end{aligned}$$

where $(\eta_{1t}, \eta_{2t})'$ is distributed as a bivariate normal with zero mean and variance and covariance matrix, Γ, 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:

[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.

- *Mean Squared Error*,

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(X_{t+i} - \hat{X}_{t+i|t+i-1})^2 \tag{36}$$

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

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_{t+i} - \hat{X}_{t+i|t+i-1})^2} \tag{37}$$

- *Mean Absolute Error*,

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|X_{t+i} - \hat{X}_{t+i|t+i-1}| \tag{38}$$

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 FIVAR with respect to the benchmark HAR model.

We also implement a Diebold and Mariano (1995) test, to directly compare the forecasting ability of the FIVAR 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

$$DM = \frac{1}{N}\frac{l_i - l_*}{s_i} \approx N(0,1) \tag{39}$$

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

As shown in table 12, the Diebold-Mariano test highlights a forecasting superiority of the FIVAR 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, $\hat{X}_{t|t-1}$

$$X_{t+h} = \alpha + \beta \hat{X}_{t+h|t+h-1} + v_t \tag{40}$$

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

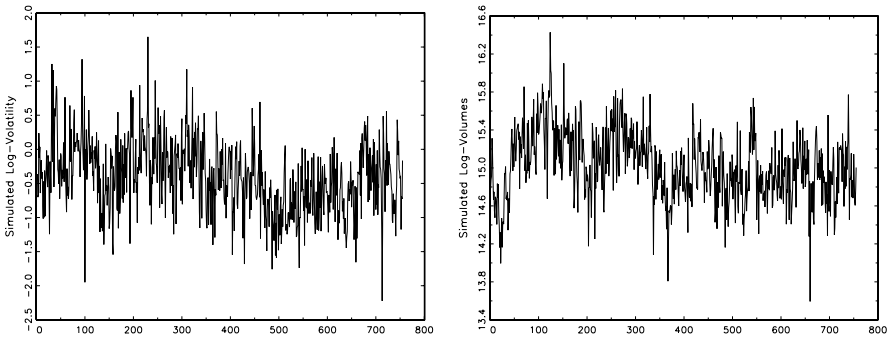

Figure 6: Simulated paths of log-Volatility and log-Volumes.

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. Moreover, for all copula models, the simulated

confidence intervals include the sample skewness. The same is not 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 sample autocorrelations and the corresponding simulated 95% confidence bands. Our bivariate long memory models, for $\log RV_t$ and $\log V_t$, 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 sample 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 recent 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 variation 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 provide 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 integration 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 estimate 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 multivariate 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

AIT-SAHALIA, Y E A. MYKLAND, P. (2003): "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," NBER Working Papers 9611, National Bureau of Economic Research, Inc.

ANDERSEN, T. G. (1996): "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance*, 51, 169–203.

ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2005): "Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," Discussion paper.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND H. EBENS (2001): "The Distribution of Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

AZZALINI, A., AND A. CAPITANIO (2003): "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution.," *J.R. Statist. Soc. B*, 65.

BANDI, F., AND J. RUSSELL (2003): "Separating market microstructure noise from volatility," Discussion paper, Working paper, GSB - University of Chicago.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2002): "Econometric analysis of realized 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 with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

BERAN, J. (1994): *Statistics for Long-Memory Processes*. Chapman & Hall.

BOLLERSLEV, T., AND D. JUBINSKI (1999): "Equity Trading Volume and Volatility: Latent Information Arrivals and Common Long-Run Dependencies," *Journal of Business & Economic Statistics*, Vol. 17, 9–21.

BOLLERSLEV, T., U. KRETSCHMER, C. PIGORSCH, AND G. E. TAUCHEN (2005): "A Discrete-Time Model for Daily S&P 500 Returns and Realized Variations: Jumps and Leverage Effects," Discussion paper, Duke University.

BREITUNG, J., AND U. HASSLER (2002): "Inference on the cointegration rank in fractionally integrated processes," *Journal of Econometrics*, 110(2), 167–185.

CHERUBINI, U., E. LUCIANO, AND W. VECCHIATO (2005): *Copula Methods in Finance*. Wiley & Sons Ltd.

CLARK, P. (1973): "A subodinated Stochastic Process Model with Finite Variance For Speculative Prices," *Econometrica*, 41, 135–155.

COLES, S., J. HEFFERNAN, AND J. TAWN (1999): "Dependence measures for extreme value analyses," *Extremes*, 3, 5–38.

COPELAND, T. (1975): "A Model of Asset Trading Under the Assumption of Sequential Information Arrival," *Journal of Finance*, September, 1149–1168.

CORSI, F. (2003): "A Simple Long Memory Model of Realized Volatility," Discussion paper, Manuscript, University of Southern Switzerland.

CORSI, F., U. KRETSCHMER, S. MITTNIK, AND C. PIGORSCH (2005): "The Volatility of Realized Volatility," CFS Working Paper Series 2005/33, Center for Financial Studies, available at http://ideas.repec.org/p/cfs/cfswop/wp200533.html.

CORSI, F., D. PIRINO, AND R. RENÒ (2008): "Volatility forecasting: the jumps do matter," Discussion paper, Quaderni del Dipartimento Di Economia Politica, Università Degli Studi Di Siena.

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13(3), 253–63.

EMBRECHTS, P., F. LINDSKOG, AND A. MCNEIL (2003): "Modelling dependence with copulas and applications to risk management," in *Handbook of heavy tailed distributions in finance*, ed. by S. Rachev. Elsevier/North-Holland, Amsterdam.

EPPS, T., AND M. EPPS (1976): "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture Distribution Hypotesis," *Econometrica*, 44, 305–321.

FANG, K. T., S. KOTZ, AND K. W. NG (1990): *Symmetric Multivariate and Related Distribution*. Chapman & Hall, London.

FANTAZZINI, D. (2008): "The Effects of Misspecified Marginals and Copulas on Computing the Value at Risk: A Monte Carlo Study," *Computational Statistics and Data Analysis*, Article in Press.

GEWEKE, J., AND S. PORTER-HUDAK (1984): "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221–238.

GRANGER, C. W. J. (1986): "Developments in the Study of Cointegrated Economic Variables," *Oxford Bulletin of Economics and Statistics*, 48(3), 213–28.

HANSEN, P. R., AND A. LUNDE (2006): "Realized Variance and Market Microstructure Noise," Discussion paper, American Statistical Association Journal of Business & Economic Statistics.

HUANG, X., AND G. TAUCHEN (2003): "The Relative Contribution of Jumps to Total Price Variance," Discussion paper, Department of Economics, Duke University,.

HURVICH, C. M., R. DEO, AND J. BRODSKY (1998): "The mean squared error of Geweke and Porter Hudak's estimator of the memory parameter of a long memory time series," *Journal of Time Series Analysis*, 19, 19–46.

JOE, H. (1997): *Multivariate Models and Dependence Concepts*. Chapman & Hall, New York.

JOE, H., AND J. XU (1996): "The estimation method of inference functions for margins for multivariate models," Discussion paper, Technical Report no. 166, Department of Statistics, University of British Columbia.

JOHANSEN, S. (1988): "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

KARPOFF, J. M. (1987): "The Relation between Price Changes and Trading Volume: A Survey," *Journal of Financial and Quantitative Analysis*, 22, 109–123.

KOLE, KOEDIJK, AND VERBEEK (2007): "Selecting copulas for risk management," *Journal of Banking and Finance*, 31, 2405–2423.

LIEBERMAN, O., AND P. C. PHILLIPS (2006): "Refined Inference on Long-Memory In Realized Volatility," Discussion paper, Cowles Fundation, Yale University.

LIESENFELD, R. (2001): "A generalized bivariate mixture model for stock price volatility and trading volume," *Journal of Econometrics*, 104, 141–178.

LUU, J. C., AND M. MARTENS (2003): "Testing the mixture-of-distrubutions hypotesis using "realized" volatility," *The Journal of Future Markets*, 7, 661–679.

MANCINI, C. (2007): "Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps,," *Scandinavian Journal of Statistics*, Forthcoming.

MANCINI, C., AND R. RENÒ (2006): "Threshold estimation of jump-diffusion models and interest rate modeling," Discussion paper, Available at SSRN: http://ssrn.com/abstract=1158439.

MARINUCCI, D., AND P. M. ROBINSON (2001): "Semiparametric fractional cointegration analysis," *Journal of Econometrics*, 105, 225–247.

MEDDAHI, N. (2002): "A theoretical comparison between integrated and realized volatility," *Journal of Applied Econometrics*, 17(5), 479–508.

MINCER, J. A., AND V. ZARNOWITZ (1969): "The Evaluation of Economic Forecasts," Nber working papers, National Bureau of Economic Research, Inc.

NELSEN, R. (1999): *An Introduction to Copula*. Springer-Verlag.

NIELSEN, M. ., AND K. SHIMOTSU (2007): "Determining the cointegration rank in nonstationary fractional system by the exact local Whittle approach," *Journal of Econometrics*, 141, 574–596.

PALMA, W. (2007): *Long-Memory Time Series: Theory and Methods*. Publisher: Wiley-Interscience, Wiley Series in Probability and Statistics.

PATTON, A. J., AND K. SHEPPARD (2007): "Evaluating Volatility and Correlation Forecasts," Discussion paper, Handbook of Financial Time Series.

POON, S., M. ROCKINGER, AND J. TAWN (2004): "Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications," *Review of Financial Studies*, 17(2), 581–610.

ROBINSON, P. M. (1994): "Rates of convergence and optimal spectral bandwidth for long range dependence," *Probability Theory and Related Fields*, 99, 443–473.

——— (1995): "Log-periodogram regression of time series with long range dependence," *The Annals of Statistics*, 23, 1048–1072.

ROBINSON, P. M., AND D. MARINUCCI (2003): "Semiparametric frequency domain analysis of fractional cointegration," in *Time Series with Long Memory*, ed. by P. M. Robinson, pp. 334–373. Oxford University Press.

ROBINSON, P. M., AND Y. YAJIMA (2002): "Determination of cointegrating rank in fractional systems," *Journal of Econometrics*, 106, 217–241.

SKLAR (1959): "Fonctions de répartition à n dimensions et leurs marges," Discussion paper, Publications de l'Institut de Statistique de L'Université de Paris.

	Unfiltered					Filtered				
	Right Tail					Right Tail				
	ρ	$\bar{\chi}$	$s.e.$	χ	$s.e.$	ρ	$\bar{\chi}$	$s.e.$	χ	$s.e.$
$\log RV$	0.6247	0.6700	0.2212	0.3622	0.04737	0.6143	0.7136	0.2269	0.4000	0.0523
$\log BPV$	0.6243	0.7655	0.2338	0.3306	0.0432	0.6015	0.7684	0.2342	0.3858	0.0505
$\log TVR$	0.6121	0.3682	0.1812	–	–	0.4203	0.6037	0.1862	–	–
$\log CTBPV$	0.5876	0.6253	0.2153	0.2810	0.03675	0.5682	0.6069	0.2128	0.3206	0.0419
	Left Tail					Left Tail				
$\log RV$	0.6247	0.6961	0.2246	0.3245	0.0424	0.6143	0.9023	0.2519	0.2904	0.0380
$\log BPV$	0.6243	0.2960	0.1716	–	–	0.6015	0.4181	0.1878	–	–
$\log TRV$	0.6121	0.6954	0.2245	0.3245	0.0424	0.6037	0.9280	0.2553	0.2904	0.0380
$\log CTBPV$	0.5876	0.3395	0.1776	–	–	0.5682	0.5048	0.1993	–	–

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 ρ. The threshold, u, has been chosen in order to leave on the right (left) the 2.5% of the observations.

	$m = T^{4/5} = 483$	$m = T^{0.7} = 223$	$m = T^{2/3} = 103$
d_{logRV}	0.3707	0.3920	0.4476
d_{logV}	0.3732	0.3520	0.3726
ξ	0.0097	2.3941	8.5162
p-value	0.9214	0.1217	0.0035

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

$\log V_t = \beta_{RV} \log RV_t + e_t$		
Bandwidth	$\hat{\beta}_m$	$\widehat{d_e}$
$m = T - 1$	0.3658	0.3754 (0.0315)
$m = 20$	0.2697	0.3722 (0.0315)
$m = 15$	0.2412	0.3683 (0.0315)
$m = 9$	0.2362	0.3697 (0.0315)
$m = 6$	0.2165	0.3725 (0.0315)
$\log V_t = \beta_{CTBPV} \log CTBPV_t + u_t$		
Bandwidth	$\hat{\beta}_m$	$\widehat{d_u}$
$m = T - 1$	0.3901	0.3912 (0.0315)
$m = 20$	0.2829	0.3805 (0.0315)
$m = 15$	0.2604	0.3767 (0.0315)
$m = 9$	0.2496	0.3747 (0.0315)
$m = 6$	0.2196	0.3686 (0.0315)

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

Panel A		
	$m = T^{0.5} = 48$	$m = T^{0.6} = 103$
$d_{\log RV}$	0.4314	0.4391
	(0.0721)	(0.0492)
$d_{\log V}$	0.3317	0.3476
	(0.0721)	(0.0492)
$\widehat{T}_0$	1.048	1.9716
Panel B		
	$m_1 = T^{0.4} = 22$	$m_1 = T^{0.5} = 48$
δ_1	0.0152	0.0106
δ_2	0.0613	0.0557
$L(u)$	$v(T) = m_1^{-0.45}$	$v(T) = m_1^{-0.35}$
$m = 48$,$m_1 = 22$		
$L(0)$	-1.4918	-1.3109
$L(1)$	-1.3287	-1.2383
$\hat{r}$	0	0
$L(u)$	$v(T) = m_1^{-0.45}$	$v(T) = m_1^{-0.35}$
$m = 103$,$m_1 = 48$		
$L(0)$	-1.6463	-1.4802
$L(1)$	-1.4737	-1.3906
$\hat{r}$	0	0

Table 7: Panel A: Fractional integration estimation with exact local Whittle estimator (standard error in parenthesis). The $\widehat{T}_0$ test statistic is calculated with $h(T) = \log(T)$. Panel B: Fractional cointegration estimation. The table reports the estimated eigenvalues (δ_i) and the value of the function $L(u)$ for different choices of m and m_1.

	$NORM$	$NCOP$	$TCOP$	$GCOP$	$CCOP$
$d_{\log RV}$	0.4044^a	0.3932^a	0.3911^a	0.3966^a	0.3895^a
$d_{\log V}$	0.3765^a	0.3740^a	0.3825^a	0.3782^a	0.3822^a
ϕ_{11}	-0.0965^a	-0.1195^a	-0.0981^a	-0.1097^a	-0.0799^a
ϕ_{12}	0.1845^a	0.1575^a	0.1581^a	0.1731^a	0.1282^a
ϕ_{21}	0.0249	0.0037	0.0064	0.0043	0.0199
ϕ_{22}	0.1138^a	0.1014^a	0.1006^a	0.1175^a	0.0864^a
θ	$--$	$--$	$--$	1.6975^a	–
α_c	$--$	$--$	$--$	$--$	1.0661^a
ν	$--$	$--$	9.7145^a	$--$	$--$
ρ	0.6265	0.6273^a	0.6263	$--$	$--$
ν_{rv}	$--$	5.5261^a	5.6359^a	6.9277^a	5.8399^a
ν_v	$--$	4.5594^a	4.5979^a	5.1350^a	5.4566^a
μ_{rv}	$--$	-0.3543	-0.3202	-0.4287^a	-0.7027
μ_v	$--$	-0.1651	-0.1322	-0.2002^a	-0.0481
σ_{rv}	$--$	0.5273	0.5168	0.5947	0.4952
σ_v	$--$	0.2763	0.2685	0.3051	0.2675
α_{rv}	$--$	1.0071	0.8793	1.3595^a	0.1987
α_v	$--$	0.7802	0.5781	1.0068^a	0.5900
$P(Q_{15}^P > q)$	0.3059	0.1759	0.2042	0.1959	0.1469
$P(Q_{15}^{LM} > q)$	0.2977	0.1808	0.2274	0.2111	0.1794

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. $P(Q_{15}^P > q)$ 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

Copula	Kendall's Tau		Tail Dependence	
NORM	$\frac{2}{\pi} \cdot \arcsin(\rho)$	0.4316	0	0
TCOP	$\frac{2}{\pi} \cdot \arcsin(\rho)$	0.4308	$2t_{\nu+1}\left(\frac{-\sqrt{\nu+1}\sqrt{1-\rho}}{\sqrt{1+\rho}}\right)$	0.1456
GUMB	$1 - 1/\theta$	0.4108	$2 - 2^{1/\theta}$	0.4956
CLAY	$\alpha_c/(\alpha_c + 2)$	0.3477	$2^{(-1/\alpha_c)}$	0.5219

Table 9: Kendall's Tau and Tail Dependence.

	$\log RV$		$\log V$
ω_1	0.0004	ω_2	-0.0028
δ_{11}	0.4576^a	δ_{21}	0.2337^a
δ_{12}	0.1460^a	δ_{22}	-0.2088^b
δ_{13}	0.1928^a	δ_{23}	-0.1398
ψ_{11}	0.0177	ψ_{21}	0.2227^a
ψ_{12}	0.0237	ψ_{22}	0.4059^a
ψ_{13}	-0.0292	ψ_{23}	0.2422^a
$Q_{\hat{\eta}}(10)$	4.8217		3.9089
	(0.4380)		(0.5627)

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, $Q_{\hat{\eta}}(10)$, for ten lags of both equation residuals, and the corresponding $p - values$ in parentheses.

	Realized Volatility			Volumes		
	MSE	RMSE	MAE	MSE	RMSE	MAE
$NCOP$	0.0648	0.2545	0.1978	0.0577	0.2400	0.1639
$TCOP$	0.0649	0.2547	0.1990	0.0576	0.2399	0.1639
$GCOP$	0.0650	0.2549	0.1978	0.0576	0.2400	0.1639
$CCOP$	0.0648	0.2545	0.1978	0.0577	0.2402	0.1639
HAR	0.0660	0.2569	0.2079	0.0574	0.2395	0.1650

Table 11: Forecast Statistics.

	Realized Volatility					
	MSE-LOG	MAE-LOG	MSE-SD	MAE-SD	MSE-PROP	MAE-PROP
NCOP-HAR	-1.1790	-1.0089	-1.0416	-0.9728	-1.5506	-1.4853
TCOP-HAR	-1.1709	-0.8229	-0.9836	-0.7931	-1.6807^{c}	-1.4612
GCOP-HAR	-1.2645	-0.8789	-1.0350	-0.8606	-1.8645^{c}	-1.6329^{c}
CCOP-HAR	-0.8830	-0.7331	-0.8203	-0.6839	-1.1581	-1.0339
	Volumes					
	MSE-LOG	MAE-LOG	MSE-SD	MAE-SD	MSE-PROP	MAE-PROP
NCOP-HAR	-0.3948	-1.0856	-0.0829	-0.8925	-0.2589	-1.3892
TCOP-HAR	-0.7578	-1.2491	-0.2507	-1.0545	-1.6740^{c}	-1.6740^{c}
GCOP-HAR	-0.6569	-1.2822	-0.2519	-1.0944	-0.9185	-1.7643^{c}
CCOP-HAR	-0.6851	-1.0158	-0.1154	-0.8182	-2.3621^{b}	-2.0019^{b}

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

	Realized Volatility			Volumes		
	α	β	$\alpha = 0 \cap \beta = 1$	α	β	$\alpha = 0 \cap \beta = 1$
HAR	0.1149	0.7648	2.2680^{b}	-0.0181	1.0001	0.2780
NCOP	0.0950	0.7978	1.4827	-0.0043	0.9890	0.1558
TCOP	0.1023	0.7870	1.5528	0.0074	0.9767	0.1303
GCOP	0.1049	0.7845	1.5279	0.0016	0.9824	0.1494
CCOP	0.0949	0.7950	1.60695	0.0141	0.9895	0.1085

Table 13: Mincer-Zarnowitz Regression. *a*,*b* and *c* stands for 1%, 5% and 10% significance level of the test.

	$\log RV$									
		Clayton		Gumbel		Normal		T-copula		HAR	
Statistics	$\log RV_t$	95% Intervals		95% Intervals		95% Intervals		95% Intervals		95% Intervals	
Mean	0	-0.0867	0.2169	-0.0674	0.2513	-0.2024	0.1736	-0.1423	0.1912	-0.1472	0.1055
Std.Dev	0.7169	0.6752	0.7822	0.6784	0.7875	0.6458	0.7811	0.6718	0.7925	0.6701	0.7446
Skewness	0.6516	0.0400	0.7097	0.3172	0.8600	0.1369	1.0610	0.1638	0.9705	-0.1315	0.1280
Excess kurtosis	1.6067	0.6820	4.7980	0.6103	4.3392	0.6331	7.4001	0.6956	6.2785	-0.1900	0.2074
$q_{0.01}$	-1.4604	-1.8351	-1.3772	-1.6753	-1.2619	-1.8915	-1.3451	-1.8484	-1.3695	-0.7160	-0.6161
$q_{0.05}$	-1.0587	-1.2406	-0.8876	-1.1883	-0.8403	-1.3076	-0.9013	-1.2846	-0.8963	-0.6053	-0.4727
$q_{0.10}$	-0.8413	-0.9716	-0.6545	-0.9483	-0.6184	-1.0522	-0.6842	-1.0330	-0.6705	-1.0438	-0.7053
$q_{0.50}$	-0.054	-0.1087	0.1944	-0.1093	0.2021	-0.2113	0.1291	-0.1759	0.1552	-0.0572	0.0568
$q_{0.90}$	0.9332	0.7876	1.1431	0.8283	1.2015	0.6562	1.0890	0.7273	1.1254	0.4715	0.3409
$q_{0.95}$	1.2805	1.0746	1.4036	1.1391	1.5529	0.9539	1.4286	1.0256	1.4634	0.6169	0.7589
$q_{0.99}$	1.9777	1.6975	2.2756	1.8038	2.4019	1.6368	2.3112	1.7110	2.3073	0.8856	1.0704

Table 14: Simulation intervals for $\log RV$ statistics

	$\log V$									
		Clayton		Gumbel		Normal		T-copula		HAR	
Statistics	$\log V_t$	95% Intervals		95% Intervals		95% Intervals		95% Intervals		95% Intervals	
Mean	0	-0.0932	0.1135	-0.0512	0.1411	-0.0795	0.1562	-0.0940	0.1220	-0.0583	0.0548
Std.Dev	0.4198	0.4235	0.4990	0.4116	0.4831	0.4193	0.5124	0.4307	0.5256	0.3995	0.4411
Skewness	0.2235	-0.2806	0.3986	0.2044	0.8964	-0.0324	1.2914	-0.0152	1.1953	-0.1182	0.1253
Excess kurtosis	1.1686	0.3700	3.7918	0.6194	5.3000	0.8348	10.255	0.8003	10.830	0.1996	0.2020
$q_{0.01}$	-0.9953	-1.2710	-0.9349	-1.0738	-0.7886	-1.2032	-0.8214	-1.2369	-0.9002	-1.8489	-1.5002
$q_{0.05}$	-0.6438	-0.8606	-0.6026	-0.7532	-0.5192	-0.8045	-0.5301	-0.8421	-0.5876	-1.3288	-1.0475
$q_{0.10}$	-0.488	-0.6720	-0.4435	-0.5941	-0.3877	-0.6319	-0.3873	-0.6604	-0.4348	-1.0606	-0.7972
$q_{0.50}$	-0.0153	-0.0997	0.1081	-0.0755	0.1155	-0.0900	0.1284	-0.1050	0.1033	-0.1526	0.1064
$q_{0.90}$	0.527	0.4611	0.6993	0.4888	0.7125	0.4715	0.7378	0.4608	0.7163	0.7498	1.0218
$q_{0.95}$	0.716	0.6320	0.8868	0.6686	0.9181	0.6556	0.9711	0.6499	0.9354	0.9959	1.2900
$q_{0.99}$	1.106	0.9826	1.3390	1.0694	1.4250	1.0899	1.5681	1.0758	1.5157	1.4528	1.8123

Table 15: Simulation intervals for $\log V$ statistics

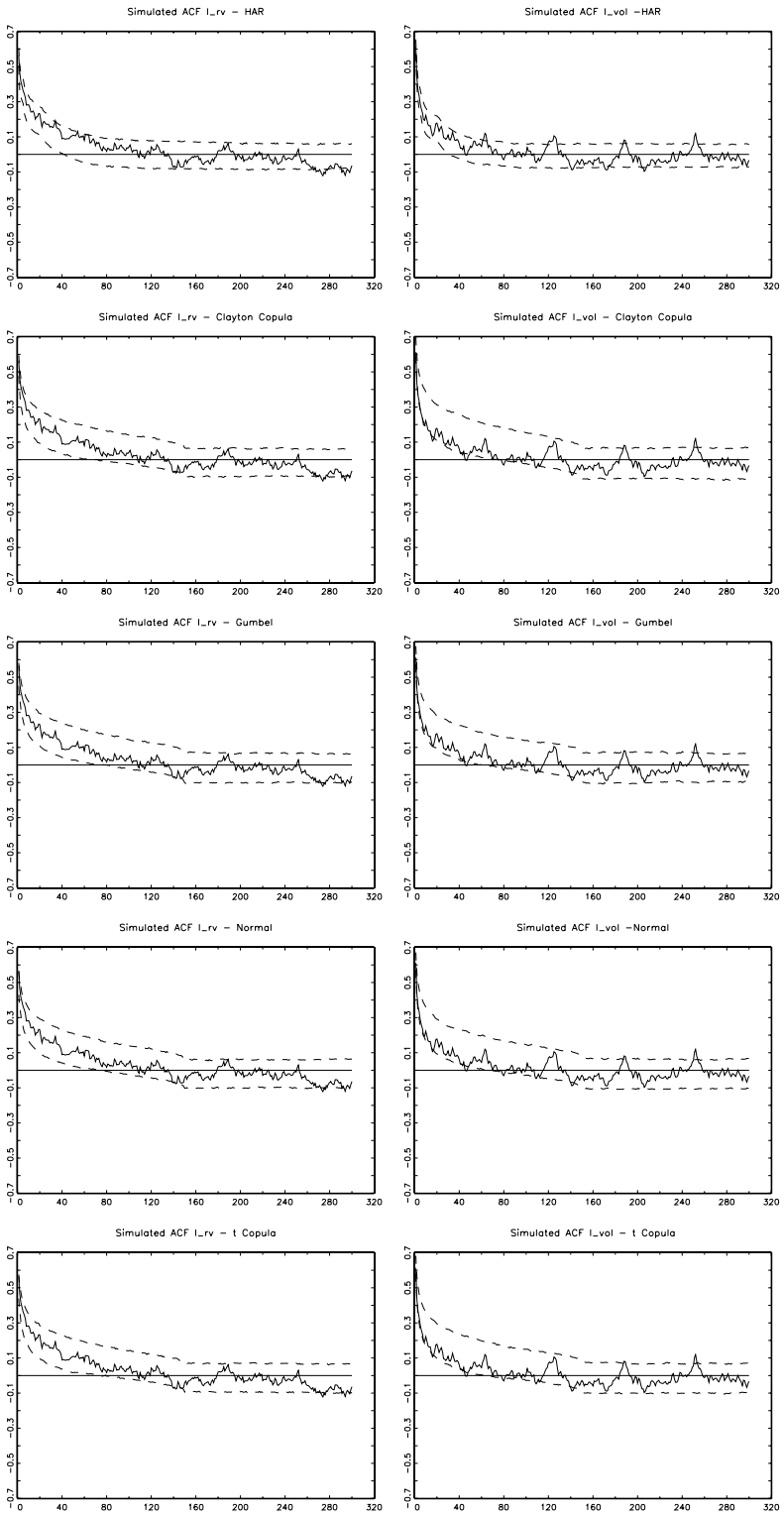

Figure 7: Simulated ACF confidence intervals of volatility and volumes.