

# Long Memory and Tail dependence in Trading Volume and Volatility<sup>1</sup>

Eduardo Rossi<sup>2</sup>

University of Pavia e-mail: [erossi@eco.unipv.it](mailto:erossi@eco.unipv.it)

Paolo Santucci de Magistris<sup>3</sup>

University of Pavia e-mail: [paolo.santuccidemagi01@ateneopv.it](mailto:paolo.santuccidemagi01@ateneopv.it)

**Abstract:** This paper investigates the relationship between volatility and volumes, using IBM stocks data, supposing that the link has to be found in their long-run dependencies. Tail dependence analysis shows that log-realized volatility and log-volumes are characterized by upper and lower tail dependence, where the former is mainly due to the jump component. The two series share the same degree of long memory but are not fractionally cointegrated. We estimate a bivariate FIVAR specification by means of different copula functions. The estimates and the simulation comparison with a bivariate HAR model highlights the ability of the bivariate FIVAR with copula density specification to account for the common long memory pattern and tail dependence.

**Keywords:** High-frequency data, Realized Volatility, Trading Volume, Long memory, Fractional Cointegration, Copula Modeling.

## 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 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. 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.<sup>4</sup> The second one concentrates on the proper-

---

<sup>1</sup>The present paper is financially supported by PRIN 2006.

<sup>2</sup>Address of correspondence: Dipartimento di economia politica e met. quant. Via San Felice, 5. 27100 Pavia. ITALY

<sup>3</sup>Address of correspondence: Dipartimento di economia politica e met. quant. Via San Felice, 5. 27100 Pavia. ITALY

<sup>4</sup>See the approach presented in Andersen (1996) and Liesenfeld (2001).

ties 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. 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 previous studies, 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.

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 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. 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).

The analysis of the tail-dependence of the filtered series suggests that a careful treatment of the extreme dependence 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.

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.

## 2. Tail Dependence

- 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, 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 *et al.* (2008), the bipower variation underestimates the jump component, in particular in case of two consecutive jumps in the intradaily returns.
- 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.

### 3. The MDH as a Long Memory Relationship

The test for the null hypothesis of equality of long memory parameters in the log series can not be rejected, using the Robinson (1995) test, based on the estimates of the fractional integration obtained with the Geweke and Porter-Hudak (1984) estimator. Nevertheless, the equality of the fractional integration orders does not imply the validity of the MDH theory. In fact, if the MDH theory were verified, there should exist a common stochastic trend, that is the latent information arrival process with long memory, that drives the dynamics of both 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 volumes and volatility. The fractional cointegration analysis rejects the presence of a common fractional stochastic trend.

### 4. 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:

$$\Phi(L)D(L)X_t = \epsilon_t \quad (1)$$

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

where  $X_t = (\log RV_t, \log V_t)'$ ,  $\Phi(L) = I_2 - \Phi_1 L - \dots - \Phi_p L^p$ ,  $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})'$ , with  $E(\epsilon_t) = 0$  and  $Var(\epsilon_t) = \Sigma$ . This model is a Fractionally Integrated VAR (FIVAR). We assume that the  $\epsilon_t$  have a joint distribution  $\epsilon_t \sim G(\epsilon_t; \psi)$ , with  $G(\cdot)$  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 specify the joint multivariate density by means of a copula function density.

## 5. Conclusions

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 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. 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 extreme dependence. 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. The evidence from the simulation exercise highlights the ability to capture the essential features of the data.

## References

- Andersen T.G. (1996) Return volatility and trading volume: An information flow interpretation of stochastic volatility, *Journal of Finance*, 51, 169–203.
- Bollerslev T. and Jubinski D. (1999) Equity trading volume and volatility: Latent information arrivals and common long-run dependencies, *Journal of Business & Economic Statistics*, Vol. 17, 9–21.
- Clark P. (1973) A subordinated stochastic process model with finite variance for speculative prices, *Econometrica*, 41, 135–155.
- Copeland T. (1975) A model of asset trading under the assumption of sequential information arrival, *Journal of Finance*, September, 1149–1168.
- Corsi F., Pirino D. and Renò R. (2008) Volatility forecasting: the jumps do matter, Technical report, Quaderni del Dipartimento Di Economia Politica, Università Degli Studi Di Siena.
- Epps T. and Epps M. (1976) The stochastic dependence of security price changes and transaction volumes: Implications for the mixture distribution hypothesis, *Econometrica*, 44, 305–321.
- Geweke J. and Porter-Hudak S. (1984) The estimation and application of long memory time series models, *Journal of Time Series Analysis*, 4, 221–238.
- 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 Martens M. (2003) Testing the mixture-of-distributions hypothesis using “realized” volatility, *The Journal of Future Markets*, 7, 661–679.
- Nielsen M. and Shimotsu K. (2007) Determining the cointegration rank in nonstationary fractional system by the exact local whittle approach, *Journal of Econometrics*, 141, 574–596.
- Robinson P.M. (1995) Log-periodogram regression of time series with long range dependence, *The Annals of Statistics*, 23, 1048–1072.