ON THE DEPENDENCE STRUCTURE OF REALIZED VOLATILITIES

Beatriz Vaz de Melo Mendes
Victor Bello Acciolly

On the Dependence Structure of Realized Volatilities

Beatriz Vaz de Melo Mendes

IM/COPPEAD - Federal University at Rio de Janeiro, Brazil.

Victor Bello Accioly

COPPEAD - Federal University at Rio de Janeiro, Brazil.

Abstract

Volatility plays an important role when managing risks, composing portfolios, pricing financial instruments. However it is not directly observable, being usually estimated through parametric models such as those in the GARCH family. A more natural empirical measure of daily returns variability is the so called *realized volatility*, computed from high-frequency intra day returns, an unbiased and highly efficient estimator of the return volatility. At this time point, with globalization effects driving markets' volatilities all over the world, it becomes of great interest to assess volatilities' co-movements and contagion. To this end we use pair-copulas, a powerful and flexible statistical model which allows for linear and nonlinear, possibly asymmetric forms of dependence without the restrictions posed by existing multivariate models. Given the importance of the Brazilian stock market in the Latin America, in this paper we characterize the dependence structure linking the realized volatilities of seven Brazilian stocks. The realized volatilities are computed using an 8-year sample of 5-minutes returns from 2001 through 2009. We include a more comprehensive study involving seven emerging markets, addressing the issue of contagion in a more general scenario.

Keywords: Pair-copulas; Dependence structure; Realized Volatilities; High-frequency data; Contagion; Multivariate tail dependence coefficient.

JEL subject classifications: C22, C51, G17, G32.

Resumo

Volatilidade tem um papel importante na administração de riscos, na composição e administração de carteiras, e no apreçamento de intrumentos financeiros. Contudo ela no é diretamente observável, sendo usualmente estimada através de modelos paramétricos tais como aqueles da grande família GARCH. Uma medida empírica mais natural da volatilidade de retornos financeiros diários é a *volatilidade realizada*, calculada a partir de dados de alta frequencia, um estimador eficiente e no viciado da volatilidade do retorno. Atualmente vemos os efeitos da globalização comandando os movimentos das volatilidades dos mercados de todo o mundo, sendo assim de grande interesse o entendimento desses movimentos simultâneos e do contágio resultante. Para acessar esses efeitos, neste artigo utilizamos a modelagem através de pair-copulas, um modelo estatístico flexível e poderoso que permite modelar formas de dependência lineares, no lineares, assimétricas, sem colocar restrições usualmente impostas por modelos simplificadores existentes. Dada a importância do mercado brasileiro na América Latina, neste trabalho caracterizamos a estrutura de dependência entre as 8 ações brasileiras mais comercializadas. Calculamos a volatilidade realizada utilizando uma amostra de 8 anos contendo retornos de 5-minutos de 2001 a 2009. Incluimos também uma investigação mais abrangente envolvendo 7 mercados emergentes, focando no estudo do contágio neste cenário mais geral.

Palavras-Chave: Pair-copulas; Estrutura de dependência; Volatilidade Realizada; Dados de alta frequencia; Contágio; Coeficiente de dependência multivariada de cauda.

On the Dependence Structure of Realized Volatilities

1   Introduction

Volatility plays an important role when managing risks, composing portfolios and pricing financial instruments. However it is not directly observable. The existing parametric models for estimating the latent volatility include the popular ARCH-GARCH family, stochastic volatility models and Markov-switching models. Being estimated, are subject to many sources of errors and are only valid under the specific assumptions of each approach. Ex-post squared or absolute returns may be used as a non-parametric measure of daily returns variability. However, a more natural empirical measure is the so called *realized volatility*, computed from high-frequency intra day returns, an unbiased and highly efficient estimator of the return volatility. By noting that variability of a variable during some interval may be measured as the integral of its instantaneous variability over this interval, Andersen and Bollerslev (1998a) defined the realized variance as the sum of the squared high-frequency intra day returns over this interval.

Many following research papers focused in either characterizing the systematic distributional features of the intra day return volatility process or providing applications of the new estimator. Andersen and Bollerslev (1998b) and Dacorogna et al. (1993), characterized intra day volatility pattern in foreign exchange markets, while Ederington and Lee (1993), and Fleming and Remolona (1999) did the same in bond markets. Andersen, Bollerslev, and Cai (2000) characterized the volatility in the Japanese stock market, whereas Andersen, Bollerslev, Diebold, and Ebens (2001) studied realized variance measures using 30 DJIA stocks. Improved forecasts and risk measures were computed based on the realized volatilities in Martens (2001), Moosa and Bollen (2002), Pong, Shackleton, Taylor, and Xu (2004), and Giot and Laurent (2004). All these empirical research based on high-frequency asset prices highlighted the fact that intra day return volatility processes vary substantially according to the asset and market types. This fact makes the investigation on the topic even more interesting and calls for specific data treatments.

However, all above references handle the series of realized volatility in the univariate setting. There are just a few papers addressing the problem of modeling volatilities interdependencies. For example, Andersen, Bollerslev, Diebold, and Labys

(2003) estimate the (conditional) fractionally-integrated Gaussian vector autoregression bivariate model for the logarithmic realized volatilities for the Deutschemark/Dollar and the Yen/Dollar spot exchange rates. Ané and Métais (2009) examine the unconditional distribution of the realized variance of three European stock market indexes and characterize their dependence structure using copulas. Indeed, Ané and Métais (2009) note that ... *important lessons, for both academic research and practical use, could be derived from a closer attention paid to the multivariate unconditional probabilistic features of realized volatility or variance series.* In fact, any question about the assets co-movements during crisis or about the long run simultaneous behavior of their volatilities, may only be answered through their unconditional joint distribution.

Accordingly, in this paper we focus on the unconditional *joint* parametric modeling of series of realized volatilities for seven Brazilian stocks. Although driven by similar motivations, our work differs from the Ané and Métais (2009) study in many ways: the copula model used, the data, choice of the univariate marginal model, and the risk measures computed. Instead of using copulas, we go further and parametrically characterize the dependence structure linking the series of realized volatilities using *pair-copulas*. The pair-copulas construction, being an hierarchical cascade of copulas, constitute a powerful and flexible model which allows for linear and nonlinear, possibly asymmetric forms of dependence without the restrictions posed by classical multivariate models, including high dimensional copulas. Being a factorization based on only bivariate copulas, it is easy to estimate. The bivariate copulas are free to vary and may belong to any family, from non-exchangeable to elliptical or extreme value, or archimedians, or any other, and may possess positive and different upper and lower tail dependence coefficients to model contagion during extreme events. The decomposition is then able to cover all types of dependence resulting in a tailored dependence structure.

We illustrate the modeling approach using two data sets from emerging markets, our second contribution. Most studies dealing with stock market volatility use U.S. data, and a look at the behavior of the realized volatility in emerging markets is still lacking. We firstly focus in modeling the multivariate behavior of the realized variance from Brazilian stocks due to the importance of this market in the Latin America. Using data from just one geographical area also avoids the effects of lack of synchronicity among the series, which may potentially corrupt the observed co-

movements. Our full sample consists of 8 years of 5-minutes returns from 2001 through 2009 for the seven most liquid stocks. The period analyzed include the sub-prime crisis and some local ones. Some well known characteristics of intra day volatility were found also for the Brazilian data. Then we look at contagion among volatilities from seven emerging markets' indexes.

Thirdly, we contribute to the existing efforts for modeling the univariate distribution of the realized log-volatility. The path we follow here is simple: we parametrically model using the flexible skew-t distribution.

Finally, we compute measures of contagion and assess the effect of the dimensionality on the strength of the multivariate tail dependence. In the remainder of the paper we show in Section 2 how we collected and transformed the Brazilian data into volatility measures. In Section 3 we set up the selected marginal and joint models, providing a brief review on copulas and pair-copulas along with inference methodologies. Estimation results are provided. In Section 4 we assess the usefulness of the modeling strategy by computing accurate estimates of contagion measures and of the multivariate upper tail dependence coefficient. We conclude in Section 5 with suggestions for future research and discussion of the findings in this paper.

2   Realized volatility in the Brazilian stock market

High frequency data possess unique characteristics, not present in low frequency data (daily, weekly, monthly). Transactions (with variable volumes) occur in irregular time space; data show different periodic market activity patterns; there may be asynchronicity in the data; and the bid-ask bounce effect may distort inferences. The number of observations is huge, increasing the chances of many types of errors (transaction errors, recording errors). All these features call for specific data treatments and make their statistical analysis more interesting.

The high frequency data used here were provided by BOVESPA and covers the 8-years period from January, 2001 to April, 2009. We use the 7 more liquid stocks in the Brazilian market, PETR4 (Petrobras), VALE5 (Vale), TNLP4 (Telemar), USIM5 (Usiminas), BBDC4 (Bradesco), CSNA3 (Siderúrgica Nacional) and ITAU4 (Itauunibanco). According to Zivot (2005) stocks must be liquid since infrequent trading may induce negative autocorrelation. This also guarantees many quotes

per day per stock. Each data record is similar to the "Trades and Quotes" (TAQ) provided by the NYSE, and contains information on the price, volume traded, day and time of trading, and names of the trading firms.

Some specific data treatments were needed. Firstly, data were expressed using the Greenwich mean time format. This was necessary to eliminate the impacts caused by the changes in the BOVESPA closing time. Usually the BOVESPA has a continuous trading session from 10 a.m. to 5 p.m., but during summer this changes to 11 a.m. through 6 p.m. The data also required other adjustments to make prices comparable, such as taking care of bonus, splits, capital adjustments, groups, subscription, dividends, number of quotes, fusions, and so on.

We consider the 5-minutes sampling interval as the regularly spaced time for the seven hours of continuous negotiation in the Brazilian market. Thus the number of sampled observations per trading session is $m = 84$ interval/day. Let $\Delta = 1/m$ be the fraction of a trading session associated with the sampling frequency, and let $P_{i,j}$ and $Q_{i,j}$ represent a j-th price and volume for the i-th stock, during some time interval Δ.

For data alignment, a common practice denominated "before" makes use of the most recent observation, or the closest, with respect to the desired minute, and, through a linear interpolation of the log-price, obtains the average of the bid-ask values. In this work, to align the logarithmic price process to the regularly spaced time clock chosen, we compute the real value log-price $p_{i,t}$ of asset i for the interval from time t to $t + \Delta$, as

$$p_{i,t} = \ln \left(\frac{P_{i,1} \cdot Q_{i,1} + P_{i,2} \cdot Q_{i,2} + \cdots + P_{i,n} \cdot Q_{i,n}}{Q_{i,1} + Q_{i,2} + \cdots + Q_{i,n}} \right), \tag{1}$$

where n is the number of negotiations of stock i in this time interval, the so-called volume-weighted average price, VWAP. The VWAP gives rise to a smaller realized variance, since it is closest to the efficient price instead of the closing price.

According to some authors it is necessary to eliminate days with too many 5-minutes intervals with no change in the stock price (what would result in zero volatility). Since our analysis is multivariate, the same days were withdrawn from analysis for all stocks.

Then we obtain the the intra-day continuously compounded return $r_{i,t+\Delta}$, on asset i, from time t to $t + \Delta$ as

$$r_{i,t+\Delta} = p_{i,t+\Delta} - p_{i,t}. \tag{2}$$

Intra-day returns data also present seasonality (Breyman, 2003), volatility clusters, and the higher the frequency the larger the kurtosis in the data. High frequency data are typically affected by specific conditions such as periods without trading, calendar effects, or the bid-ask bounce effects. Among others market microstructure effects, equity returns usually present negative autocorrelation. An ARMA(p, q) (autoregressive moving average) model may be used to filter the series before constructing the series of realized volatility. We note that the daily return was computed by summing up all 5-minutes log-returns computed during a daily continuous session.

The realized variance $(RV_{i,t})$ for the stock i, $i = 1, \ldots, d$, $d = 7$, at day t, is defined as:

$$RV_{i,t} = \sum_{j=1}^{m} r_{i,t-1+j\Delta}^2, \ t = 1, \ldots, T \tag{3}$$

where T is the number of days in the period analyzed. We recall that the consistency of the estimator is attached to $\Delta \to 0$ ($m \to \infty$). The realized volatility $(RVOL_{i,t})$ of stock i at day t is defined as the square root of the realized variance, $RVOL_{i,t} = \sqrt{RV_{i,t}}$, and the log-realized volatility $(RLVOL_{i,t})$ is given by $RLVOL_{i,t} = \ln(RVOL_{i,t})$.

We now treat volatility as directly observed rather than latent. The seven $RLVOL$ series provided the basic statistics gathered in Table (1).

Table 1: *Basic statistics for the seven series of log-realized volatilities (RLVol).*

$RLVol_{Stock}$	mean	std	skewness	exc-kurtosis	max	min	median
$RLVol_{ITAU4}$	0.4198	0.4174	0.3430	0.6070	2.1690	-0.8068	0.4007
$RLVol_{BBDC4}$	0.4811	0.3818	0.5191	1.0857	2.1207	-1.0681	0.4505
$RLVol_{PETR4}$	0.3458	0.4065	0.7295	0.9256	2.0986	-0.6623	0.2916
$RLVol_{VALE5}$	0.3507	0.4256	0.5577	0.9643	2.1596	-0.9979	0.3118
$RLVol_{CSNA3}$	0.6307	0.4194	0.1941	0.7791	2.2620	-1.0406	0.6266
$RLVol_{USIM5}$	0.7385	0.3749	0.2938	0.4432	2.1764	-0.5653	0.7219
$RLVol_{TNLP4}$	0.5354	0.3849	0.5764	0.4223	2.0074	-0.7107	0.4869

These series, rather than the $RV_{i,t}$ or the $RVOL_{i,t}$ series, will be used to assess interdependencies and contagion in the Brazilian stock market since they are more suitable for marginal parametric modeling. Moreover, their dependence structure is the same as the one among the series of realized volatility due to the copula invariance property to increasing transformations.

3 Building up the margins and the dependence structure

To model the joint distributional behavior and characterize the types of linear and non-linear association among the $RLVOL$ series (all potentially different) we use pair-copulas. In the next subsection we provide a brief review on this subject.

3.1 Copulas and Pair-copulas

Let $X_1, \ldots, X_d$ be real-valued random variables (r.v.). Here they represent the series of log-realized volatilities and $d = 7$. The dependence between $X_1, \ldots, X_d$ is completely described by $H(x_1, \ldots, x_d)$, their joint distribution function (c.d.f.). In most applied situations, only the margins are known (estimated or fixed a priori) and the joint distribution H may be unknown or difficult to estimate.

The idea of separating H into a part which describes the dependence structure and parts which describe the marginal behavior only, has lead to the concept of copulas, introduced in statistical literature by Sklar (1959).

Copulas may be defined as follows: Let $X_1, \ldots, X_d$ be continuous r.v. with joint distribution function $H(x_1, \ldots, x_d)$, joint density function h, marginal c.d.f.s $F_1, \ldots, F_d$, and marginal densities f_i, $i = 1, \cdots, d$. For every $(x_1, \ldots, x_d) \in [-\infty, \infty]^d$ consider the point in $[0, 1]^{d+1}$ with coordinates $(F_1(x_1), \ldots, F_d(x_d), H(x_1, \ldots, x_d))$. This mapping from $[0, 1]^d$ to $[0, 1]$ is a d-*dimensional copula*.

The Sklar's Theorem may be stated as follows: Let H be a d-dimensional c.d.f. with marginal c.d.f.s $F_1, \ldots, F_d$. Then there exists an d-dimensional copula C such that for all $(x_1, \ldots, x_d) \in [-\infty, \infty]^d$,

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{4}$$

Conversely, if C is a d-dimensional copula and $F_1, \ldots, F_d$ are c.d.f.s, the function H defined by (4) is a d-dimensional distribution function with margins $F_1, \ldots, F_d$. Furthermore, if all marginal c.d.f.s are continuous, C is unique.

Given a joint c.d.f. H with continuous margins $F_1, \ldots, F_d$, as in Sklar's Theorem, it is easy to construct the corresponding copula:

$$C(u_1, \ldots, u_d) = H(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)), \tag{5}$$

where F_i^{-1} is the generalized inverse of F_i.

We assume C being absolutely continuous, and by taking partial derivatives of (4) one obtains

$$h(x_1, \cdots, x_d) = c(F_1(x_1), \cdots, F_d(x_d)) f_1(x_1) \cdots f_d(x_d) \qquad (6)$$

where c represents the copula density. This expression will prove useful later for parameter estimation.

The copula C summarizes the dependence structure of H independently of the specification of the marginal distributions. It is invariant under strictly increasing transformations of the marginal variables being therefore very convenient for studying the dependence structure of the log-realized volatilities.

However, the copula approach has also its limitations. On the one hand, most of the available softwares deal only with the bivariate case. On the other hand and most importantly, high-dimensional parametric copula families usually have a common parameter for the dependence structure and marginal distributions, which would restrict all pairs to possess the same type or strength of dependence. A solution may be a pair-copula decomposition.

The decomposition of a multivariate distribution into a cascade of bivariate copulas was originally proposed by Joe (1996), and later discussed in detail by Bedford and Cooke (2001, 2002), Kurowicka and Cooke (2006) and Aas, Czado, Frigessi, and Bakken (2007). The pair-copulas hierarchical construction, being a collection of potentially different bivariate copulas, is flexible and very appealing. The variables are sequentially incorporated into the conditioning sets as one moves from the first modeling level (tree 1) to the other levels (up to tree $d - 1$). The composing bivariate copulas may vary freely, being chosen from any parametric family and parameter values. Therefore, all types and strengths of dependence can be covered. Pair-copulas are easy to estimate and simulate, making them very appropriate for modeling large dimensional data sets.

To derive a pair-copula factorization, first note that any multivariate density function may be uniquely decomposed as

$$h(x_1, ..., x_d) = f_d(x_d) \cdot f(x_{d-1} | x_d) \cdot f(x_{d-2} | x_{d-1}, x_d) \cdots f(x_1 | x_2, ..., x_d). \qquad (7)$$

The conditional densities in Equation (7) may be written as functions of the corresponding copula densities. That is, for every j

$$f(x \mid v_1, v_2, \cdots, v_d) = c_{xv_j | \mathbf{v}_{-j}}(F(x \mid \mathbf{v}_{-j}), F(v_j \mid \mathbf{v}_{-j})) \cdot f(x \mid \mathbf{v}_{-j}), \qquad (8)$$

where $\mathbf{v}_{-j}$ denotes the d-dimensional vector $\mathbf{v}$ excluding the jth component. Note that $c_{xv_j|\mathbf{v}_{-j}}(\cdot, \cdot)$ is a *bivariate* marginal copula density. For example, when $d = 3$,

$$f(x_1|x_2, x_3) = c_{13|2}(F(x_1|x_2), F(x_3|x_2)) \cdot f(x_1|x_2)$$

and

$$f(x_2|x_3) = c_{23}(F(x_2), F(x_3)) \cdot f(x_2).$$

Expressing all conditional densities in Equation (7) by means of Equation (8), we derive a decomposition for $h(x_1, \cdots, x_d)$ that consists of only univariate marginal distributions and bivariate copulas. This also provides the *pair-copula decomposition* for the d-dimensional copula c, a factorization of a d-dimensional copula based only in bivariate copulas. This is a very flexible and natural way of constructing a higher dimensional copula.

The conditional c.d.f.s necessary for pair-copulas construction are given (Joe, 1996) by

$$F(x \mid \mathbf{v}) = \frac{\partial C_{x,v_j|\mathbf{v}_{-j}}(F(x \mid \mathbf{v}_{-j}), F(v_j \mid \mathbf{v}_{-j}))}{\partial F(v_j \mid \mathbf{v}_{-j})}.$$

For the special case (unconditional) when v is univariate, and x and v are standard uniform, we have

$$F(x \mid v) = \frac{\partial C_{xv}(x, v, \Theta)}{\partial v}$$

where Θ is the set of copula parameters.

For large d, the number of possible pair-copula constructions is very large. As shown in Bedford and Cooke (2001), there are 240 different decompositions when $d = 5$. These authors introduce a systematic way to obtain the decompositions, which involves graphical models, the so called *regular vines*. The graphical models also aid in understanding the conditional specifications made for the joint distribution. Special cases are the hierarchical canonical vines (C-vines) and the D-vines. Each of these graphical models provides a specific way of decomposing the density $h(x_1, \cdots, x_d)$. For example, for a D-vine, $h()$ is equal to

$$\prod_{k=1}^{d} f(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\dots,i+j-1}(F(x_i|x_{i+1}, \dots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \dots, x_{i+j-1})).$$

In a D-vine, there are $d - 1$ hierarchical nested *trees* with increasing conditioning sets, and there are $d(d - 1)/2$ bivariate copulas. For a detailed description, see

Aas, Czado, Frigessi, and Bakken (2007). Tree T_j of the D-vine has $d - j$ bivariate copulas, $j = 1, \cdots, (d - 1)$. Those in tree 1 are unconditional, and all others are conditional.

It is not essential that all the bivariate copulas involved belong to the same family. This feature is exactly what we are searching for, since our objective is to construct and estimate a multivariate distribution that best represents the data at hand, which might be comprised by completely different margins (symmetric, asymmetric, with different dynamic structures, and so on) and, more importantly, could be pair-wise joined by more complex dependence structures possessing linear and/or non-linear forms of dependence, including tail dependence, or, otherwise, could be joined independently.

For example, one may combine the following types of (bivariate) copulas: Gaussian (no tail dependence, elliptical); t-student (equal lower and upper tail dependence, elliptical); Clayton (lower tail dependence, Archimedean); Gumbel (upper tail dependence, Archimedean/EV); Tawn (non-exchangeable with different upper and lower tail dependence); BB7 (different lower and upper tail dependence, Archimedean), and others. See Joe (1997) for a copula catalogue. We have considered 12 families.

Given a d-dimensional data set and a a set of parametric copula families, equation (6) suggests that estimation may be carried on in two steps: first the univariate fits, and then the copula fit. When both steps apply the maximum likelihood method, the methodology is refereed to as the IFM method (Joe, 1997). This method yields consistent and efficient estimates. In the case of pair-copulas, maximum likelihood estimators depend on (i) the choice of factorization and (ii) the choice of pair-copula families. Algorithm implementation is straightforward. For smaller dimensions, we may compute the log-likelihood of all possible decompositions. For $d >= 5$ a specific decomposition may be chosen. One possibility is to look for the pairs of variables having the stronger tail dependence, and let those be in Tree 1 and determine the decomposition to estimate. To this end, a t-copula may be fitted to all pairs and pairs would be ranked according to the smallest number of degrees of freedom.

3.2 Estimation of marginal and joint models

The basic statistics in Table (1) suggest to unconditionally model the univariate series of log-realized volatilities using an asymmetric distribution also able to ac-

commodate the observed high kurtosis. We chose the skew-t density of Hansen (1994), which has a closed form and for which the implementation of the maximum likelihood method is an easy task because there are only 4 parameters to estimate (μ, λ, ν, σ). The parameter μ and σ equal, respectively, to the population mean and standard deviation, which exists if $\nu > 2$. When the skewness parameter λ is zero the symmetric case is recovered.

Figure (1) shows the good skew-t fit to the log-realized volatility of PETR4. In fact, the Kolmogorov test accepted the good quality of all seven skew-t fits which provided highly significant parameters estimates. We recall that under the true distribution, $U = F_i(X_i)$ is uniformly distributed on $[0, 1]$, but this may not be true under the estimated marginal distribution, $U = \widehat{F}_i(X_i)$. When applying the IFM method it is crucial that the estimated marginal distributions are indistinguishable from the true marginal distributions. Marginal fits should be carefully checked since a poor fit will result in any copula model being mis-specified. Evaluation whether the transformed series are $i.i.d.$ may be accomplished by visual assessment using the autocorrelation function. Diebold et al. (1998) suggest that the autocorrelations should be computed and inspected for the first four moments. The hypothesis that the transformed series are standard uniform may be tested via the Kolmogorov-Smirnov test. All these tools were used to assess the good quality of transformed series.

Among some well documented empirical facts on the distribution of realized volatilities (see Andersen, Bollerslev, Diebold, and Labys (2000), (2001) and (2003)), one states that although the distributions of realized volatilities are clearly asymmetric, the distributions of the logarithms of realized volatilities are approximately Gaussian. This was not the case for the stocks analyzed, as confirmed by the univariate fits and illustrated in Figure (1). In addition, all stocks daily returns standardized by their realized volatilities failed to accept the null hypothesis of both the Shapiro-Wilk and Jarque Bera tests at the 1% level, contradicting another stylized fact.

We now choose a particular pair-copula decomposition to estimate. The support set for the empirical pair-copula are the pseudo standard uniform values obtained from the skew-t fits. To help choosing the leading pairs in Tree 1 of the D-vine, we inspected the scatter plots of pairs of the transformed data, computed their linear correlation coefficients, and fitted simple appropriate bivariate copulas such
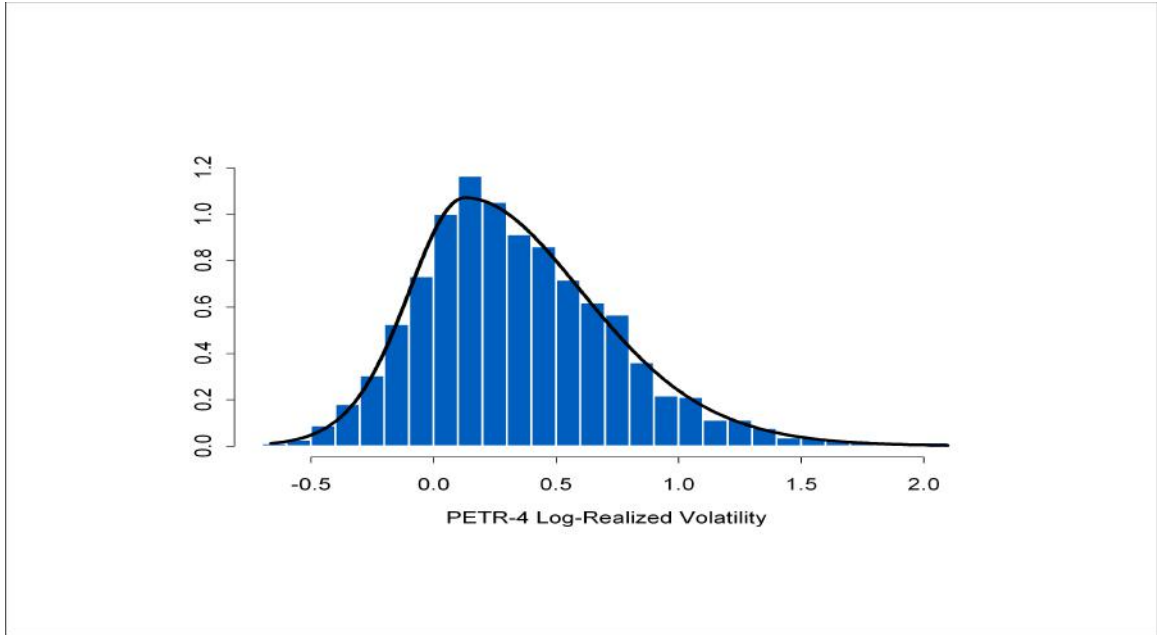
Figure 1: *Skew-t fit to the log-realized volatility of PETR4.*

as the Gumbel and the t-copula, looking for those pairs showing higher correlations, stronger tail dependence and smaller degrees of freedom. These numbers along with the information gathered from the plots, indicated the D-vine decomposition depicted in Figure (2). In this Figure, the notations "I", "B", "P", "V", "C", "U", and "T", refer to the log-realized volatilities of stocks ITAU4, BBDC4, PETR4, VALE5, CSNA3, USIM5, and TNLP4, respectively. For the sake of simplicity, we write I & P |B, instead of I|B & P|B, for example.

Estimation of the composing bivariate copulas was carried on by maximum likelihood and best fits were chosen based on the AIC values. Most important copulas were considered (Gaussian, Frank, Kimeldorf Sampson (Clayton), Gumbel, Galambos, Husler Reiss, Tawn, BB1, BB2, BB3, BB4, BB5, BB6, BB7, Normal Mixture, Joe, t-student, and Survival Clayton). Figure (2) also shows the final fits and corresponding values of the lower and upper tail dependence coefficients (λ_L, λ_U). All parameters estimates were highly significant.

All copulas in Tree 1 have positive upper tail dependence coefficients and, according to Joe, Li, and Nikoloulopoulos (2008), this is a sufficient condition for the fitted D-vine to possess multivariate upper tail dependence. The large values of the

13

Figure 2: *Pair-copulas decomposition for the 7 series of log-realized volatilities. Figure shows the copula families fitted along with the values of the lower and upper tail dependence coefficients* (λ_L, λ_U).

I & B	B & P	P & V	V & C	C & U	U & T
tawn copula	tawn copula	tawn copula	tawn copula	tawn copula	gumbel copula
(0.000,0.555)	(0.000,0.541)	(0.000,0.505)	(0.000,0.446)	(0.000,0.440)	(0.000,0.339)

I & P | B	B & V | P	P & C | V	V & U | C	C & T | U
t_{17} copula	t_8 copula	t_7 copula	tawn copula	tawn copula
(0.006,0.006)	(0.057,0.057)	(0.063,0.063)	(0.000,0.147)	(0.000,0.230)

I & V | B P	B & C | P V	P & P | V C	V & T | C U
t_{15} copula	tawn copula	gumbel copula	t_{22} copula
(0.003,0.003)	(0.000,0.158)	(0.000,0.164)	(0.002,0.002)

I & C | B P V	B & P | P V C	P & T | V C U
t_{10} copula	gumbel copula	t_{28} copula
(0.010,0.010)	(0.000,0.156)	(0.001,0.001)

I & P | B P V C	B & T | P V C U
t_{19} copula	frank copula
(0.0003,0.0003)	(0.000,0.000)

I & T | B P V C U
clayton copula
(0.058,0.000)

λ_U coefficients in Tree 1 tell us that high realized volatilities in the Brazilian stock market tend to occur simultaneously. Although all values of the λ_L coefficients in Tree 1 are approximately zero, the Tawn fit was necessary to model the asymmetry of the data, a feature captured only by a non-exchangeable copula (note that the second best fit was based on a Gumbel copula).

It is clear now that such an accurate and tailored modeling of the data dependence structure, in particular its complex pattern of dependence in the tails and the non-exchangeable characteristic, would not be possible using any multivariate family of distribution or even a d-dimensional copula.

4   Contagion measures

4.1   Contagion in the Brazilian stock market

The multivariate model fitted to the log-realized volatilities of Brazilian stocks may now be used to help understanding co-movements and contagion in this stock mar-

ket. In the bivariate case, asymmetric contagion may be assessed by computing and comparing the probabilities $\Pr\left\{U_i > u \mid U_j > v\right\}$ and $\Pr\left\{U_j > u \mid U_i > v\right\}$, for meaningful u and v values. In the case the first probability is larger than the second one, we may say that volatility of stock i is more influenced by volatility of stock j than the other way around.

We propose to compute similar conditional probabilities based on fitted D-vines and for some $(d-1)$-dimensional conditioning set $\mathcal{C}$. We consider

$$\Pr\left\{U_i > u \mid \mathcal{C} > \mathbf{v}_{-i}\right\} = \frac{\Pr\left\{U_i > u, \mathcal{C} > \mathbf{v}_{-i}\right\}}{\Pr\left\{\mathcal{C} > \mathbf{v}_{-i}\right\}} \tag{9}$$

where U_i, $i = 1, \cdots, d$, $d = 7$, represents the uniform$(0,1)$ transformed log-realized volatility of stock i, $\mathcal{C}$ represents the set of the remaining six standard uniform transformed log-realized volatilities, and $\mathbf{v}_{-i}$ represents a value in $[0,1]^{d-1}$. We compute these probabilities by simulating from the fitted pair-copula.
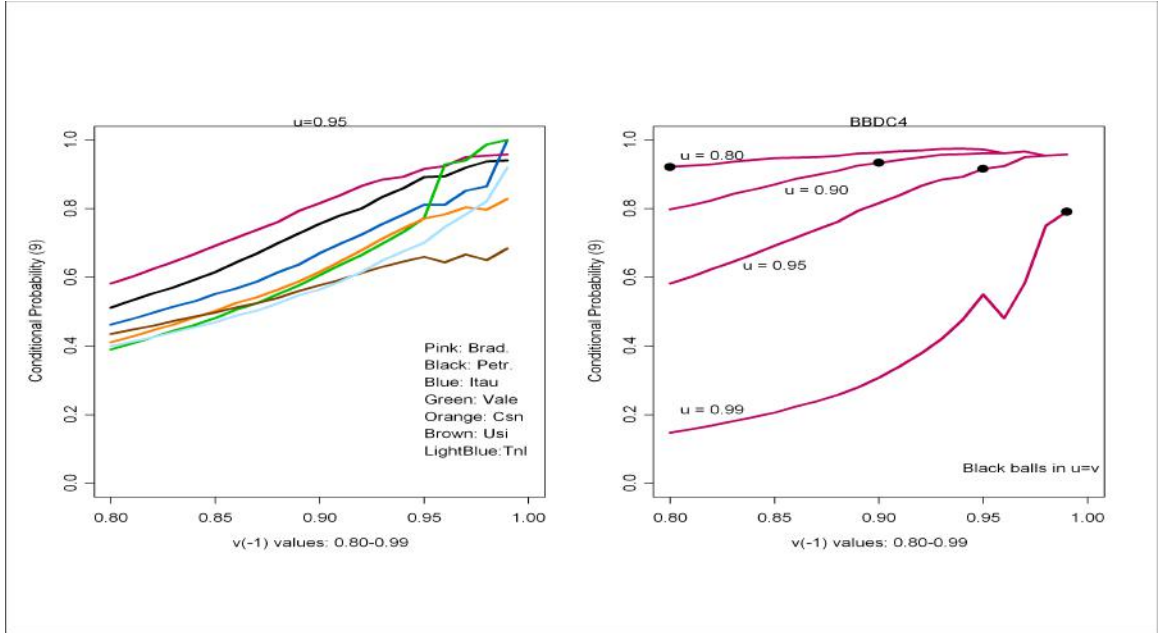

Figure 3: *Conditional probabilities (9) measuring contagion for varying levels $\mathbf{v}_{-i}$ of the conditioning set. Plot on the left hand side has u fixed equal to 0.95. Plot on the right hand side has i fixed to be the realized volatility of BBDC4, and u varying in $\{0.80, 0.90, 0.95, 0.99\}$.*

Figure 3 shows the conditional probabilities (9) for fixed u and for increasing values for $\mathbf{v}_{-i}$, varying from $(0.80, \cdots, 0.80)$ to $(0.99, \cdots, 0.99)$. Four values are

chosen for u: $0.80, 0.90, 0.95, 0.99$. The plot on the left hand side has u fixed equal to 0.95 for all seven realized volatilities. We observe that the stock BBDC4 (pink line) seems to be the one whose volatility is most influenced by the others (followed by PETR4, black line). The spill-overs observed for the seven log-realized volatilities are very strong. Note that under independence the (unconditional) probability equals to 0.05 $(1 - u)$. Plot on the right hand side has i fixed to be the realized volatility of BBDC4, and u varying in $\{0.80, 0.90, 0.95, 0.99\}$. The solid dot enhances the situation where u and elements of $\mathbf{v}_{-i}$ coincide.

The increasing feature of Figure 3, apart from the high variability observed as all $\mathbf{v}_{-i}$ values approach 1, reflects the positive interdependence among the Brazilian stocks' volatilities. As long as most important stocks experiment high turbulence, each one in turn also increase its conditional probability to be extreme. In other words, the results indicate that one may expect high volatility from any Brazilian liquid stock whenever the set of remaining liquid stocks present high variability. In summary, whatever is causing the phenomenon, clusters of high volatility seems to occur simultaneously for Brazilian liquid stocks.

Alternatively, one may be interested in computing the conditional probability

$$\Pr\left\{\mathcal{C} > \mathbf{v}_{-i} \mid U_i > u\right\} = \frac{\Pr\left\{U_i > u, \mathcal{C} > \mathbf{v}_{-i}\right\}}{\Pr\{U_i > u\}}, \tag{10}$$

which would indicate the influence of each realized volatility on the behavior of the remaining set. Actually, this leads to the concept of multivariate tail dependence coefficient if we let $(\mathbf{v}_{-i}, u) = (u, \cdots, u, u)$ and $u \to 1$.

The tail dependence coefficient is an asymptotic measure that reveals the strength of dependence at extreme levels. It is very important for predicting the degree of association among markets, or stocks, or volatilities, during crisis. Similarly to the definition in the bivariate case, the multivariate upper tail dependence coefficient Λ_U (the definition of the lower tail dependence coefficient Λ_L is analogous) is the conditional tail probability that components go to extreme values at the same rate:

$$\Lambda_U = \lim_{u \to 1} \Pr\left\{(U_1, \cdots, U_{i-1}, U_{i+1}, \cdots, U_d) > (u, \cdots, u) \mid U_i > u\right\}. \tag{11}$$

Therefore, as $u \to 1$, it does not matter on which variable U_i we are conditioning, in the limit the probability converges to a fixed value $\in [0, 1]$ (if the limit exists). If this limit equals 0 then we say that the set of d variables are asymptotically independent.
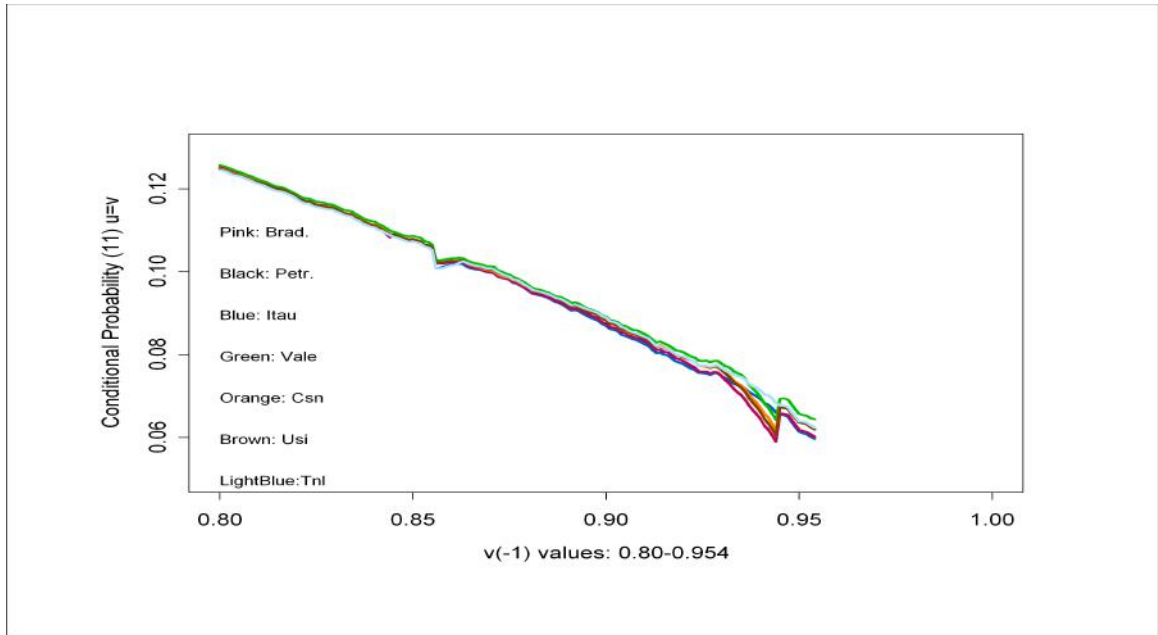
Figure 4: *Multivariate upper tail dependence coefficient: conditional probability (11).*

This measure have been considered by other authors. Joe, Li, and Nikoloulopou-los (2008) developed various multivariate tail dependence functions and determined tail dependence properties for some specific vine copulas. The authors analyze how the tail dependence of a vine copula is built up from that of lower-dimensional margins and how it is affected by the bivariate tail dependence of basic linking copulas. According to Joe, Li, and Nikoloulopoulos (2008) the theoretical development of tail dependence functions for pair-copulas is still an open issue. Estimation is also difficult, as it is any asymptotic measure, see Klüppelberg, Kuhn and Peng, 2008.

The probability in (11) is actually the d-dimensional copula built up by the bivariate copulas. As we already commented, for the seven Brazilian log-realized volatilities, the fitted vine copula is tail dependent since all the bivariate baseline copulas are tail dependent.

We obtain the limit (11) for the 7 Brazilian realized volatilities using simula-tions. Of course, as u approaches 1 the variability increases. It is very difficult to empirically estimate an asymptotic measure, and to get a feeling where to stop when going up to the extreme quantiles, we have simulated this probability in the bivariate case for copulas for which we do know the true theoretical value, actually,
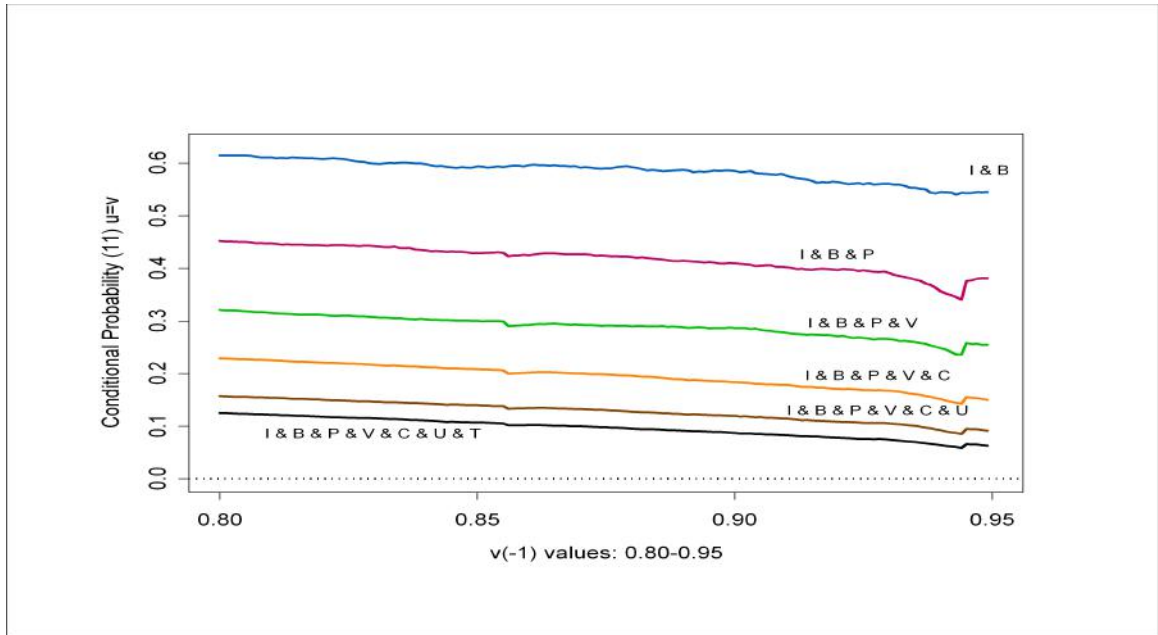
17

Figure 5: *Multivariate upper tail dependence coefficient for dimensions 2 up to 7.*

for the copulas in Tree 1 (see one example in Figure 5). The simulations indicated that beyond $u = 0.96$ the estimation error is very large.

Figure 4 shows (11) computed by conditioning in each U_i, $i = 1, 2, \cdots, 7$. All lines practically coincide, and it shows that although all upper tail dependence coefficients in Tree 1 are very high, the 7-dimensional upper tail dependence of the vine copula may be low. Here it seems to converge to a value around 0.061. Similar situations are usually referred to as "the curse of dimensionality", and we illustrate it in Figure 5. In this figure we show the convergence of the upper tail dependence coefficient for varying vine copulas with increasing dimensionality, that is, from 2 up to 7 (we use the order defined in Figure 2). Note that the limiting simulated value in the case of the bivariate copula linking "I and B" (first pair in Tree 1) it is very close to the maximum likelihood estimate 0.555 given in Figure 2.

4.2 Contagion in Emerging markets

At this time point, with globalization effects driving markets' volatilities all over the world, it becomes of great interest to assess volatilities' co-movements and contagion in a more general scenario. Following a referee's suggestion, in this sub-

18

section we complement the analysis made for the Brazilian intra-day stocks using data from 7 emerging markets. Daily prices for 7 indexes were obtained from the www.msci.com site: the MXRU Index (Russia), MXIN Index (India), MXZA Index (South Africa), MXMX Index (Mexico), MXID Index (Indonesia), MXCN Index (China), and MXBR Index (Brazil). Data cover the period from 01/02/2006 through 10/05/2011 and length of contemporaneous series is 1275.
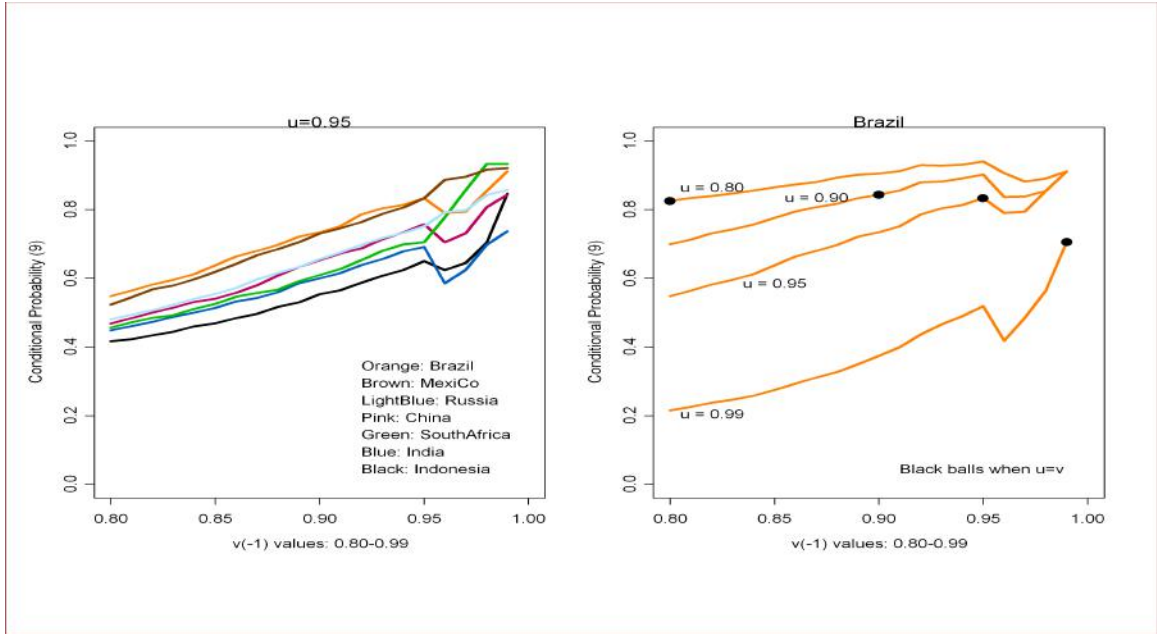

Figure 6: *Conditional probabilities (9) measuring contagion for varying levels* $\mathbf{v}_{-i}$ *of the conditioning set. Plot on the left hand side has u fixed equal to* 0.95*. Plot on the right hand side has i fixed to be the realized volatility of Brazil, and u varying in* $\{0.80, 0.90, 0.95, 0.99\}$*.*

As suggested by the referee, we use the absolute returns as a naive measure of volatility. Due to the strictly decreasing behavior and strong right asymmetry shown by the corresponding empirical densities, the marginal distributions were estimated for the logarithm of the absolute returns. As before, excellent skew-t fits were obtained for all series and used to transform the data. The same inference steps and statistical tests carried on for the previous illustration were applied here.

The order of variables in the leading Tree 1 is MXIN, MXCN, MXID, MXZA, MXBR, MXMX, MXRU (India, China, Indonesia, South Africa, Brazil, Mexico, and Russia). This particular choice resulted in the highest correlations for the unconditional copulas. The copula families providing the best fits were Gumbel

(13), BB7 (3), Tawn (3), and t-student (2). All copulas fitted in Tree 1 possess upper tail dependence. The asymmetric Tawn copula links the volatilities of Mexico and Russia, and the other 5 unconditional fits were based on the Gumbel copula.
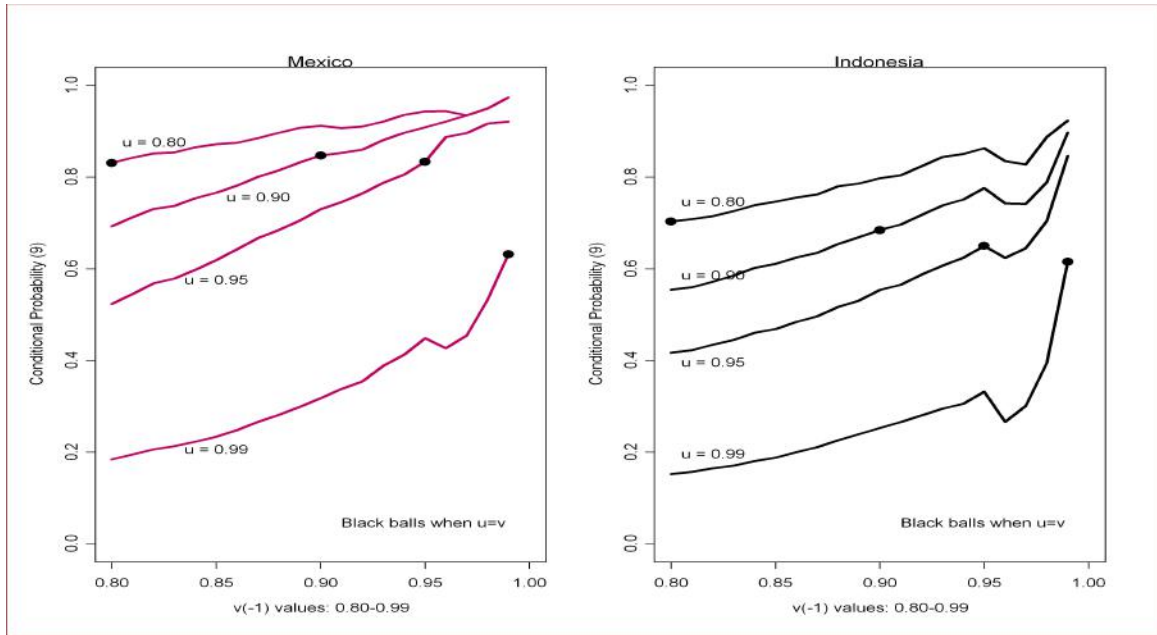

Figure 7: *Conditional probabilities (9) measuring contagion : The realized volatility of Mexico (left) and Indonesia (right) and u varying in $\{0.80, 0.90, 0.95, 0.99\}$, and for varying levels $\mathbf{v}_{-i}$ of the conditioning set.*

Likewise Figure 3, Figure 6 shows the conditional probabilities (9) for fixed u and for $\mathbf{v}_{-i}$ varying from 0.80 to 0.99 for each variable. We observe that the Latin America indexes are those seeming to be most influenced by the others. As expected, contagion is very strong. Plot on the right hand side has i fixed to be the volatility of Brazil, and u varying in $\{0.80, 0.90, 0.95, 0.99\}$. This picture when i represents the volatility of Mexico and Indonesia is provided in Figure 7.

All plots confirm the positive interdependence among the Emerging markets' volatilities. Contagion is evident at all levels. It is interesting to note that although all emerging markets move together, we observe stronger links between Brazil and Mexico, and note that Indonesia is the one suffering less influence by the others. The seven Emerging markets log-absolute returns are tail dependent since all the bivariate baseline copulas are tail dependent. The value of the multivariate upper tail dependence coefficient Λ_U was obtained by simulations and converged to a value

around 0.04.

5 Concluding remarks

We have characterized the interdependencies and contagion among volatilities in Emerging markets. Series of realized volatility were obtained from high-frequency intraday stock prices. The realized (integrated) volatility being the square root of the sum of squared high frequency returns over a certain period is considered an observable variable which evolves stochastically over the trading day. It is not equal to the volatility estimated through discrete-time GARCH models which only utilizes the past daily return observations, and it is a consistent and efficient model-free estimator of the daily return variability.

In a first exercise, the realized volatilities from the seven most liquid stocks in the Brazilian equity market were obtained from high-frequency intraday stock prices. Some particular characteristics of the Brazilian market required special treatment of the high frequency data before constructing the series of realized volatilities. We found that the marginal behavior of the series did not confirm some well known stylized facts. For instance, the distributions of the logarithms of the realized volatilities were not approximately Gaussian. The main factor contradicting the normality hypothesis was the right asymmetry. Actually, very good marginal fits based on the asymmetric t-distribution were obtained for all seven series. In addition, the stocks daily returns standardized by their realized volatilities failed to accept the normality hypothesis at the 1% level, another empirical fact described in the literature. An observed characteristic shared by other data sets was the presence of short and long range dynamics in the series of realized volatilities.

The multivariate modeling of the realized volatilities was based on the pair-copulas construction. The flexibility of the process was outstanding. It allowed for completely different marginal distributions joined by a 7-dimensional copula built up by completely different bivariate copulas possessing complex dependence structures, asymmetry and non-exchangeability, able to model linear and/or non-linear forms of dependence, as well as upper and lower tail dependence. We note that asymmetries in information transmission within the Brazilian market, and more generally among emerging markets, are likely to induce asymmetries in the dependence of stock's variances, and this type of behavior can only be captured by exchangeable copulas.

The methodology was then applied in a more general context, where contagion among emerging markets was investigated. Volatilities of indexes from seven emerging markets were analyzed confirming the high level of interdependencies among the corresponding markets.

The estimated pair-copula model provided measures of shocks transmission and contagion among volatilities. We found that high volatility may be expected from any emerging index or liquid stock whenever the set of remaining indexes present high variability. In summary, whatever is the cause of the phenomenon, clusters of high volatility seems to occur simultaneously within the emerging markets. Using simulations we computed the multivariate upper tail dependence coefficient associated with the fitted vine. As expected, all upper tail dependence coefficients between pairs of volatilities were very high, and the 7-dimensional upper tail dependence of the vine copula converged to a positive lower value.

Many other applications may follow such a tailored unconditional fit, and we expect to see them in future papers.

References

AAS, K. et al. Pair-copula constructions of multiple dependence. **Insurance:** Mathematics and Economics, v. **44**, n. 2, p.182-198, 2007.

ANDERSEN, T. G. et al. The distribution of realized stock return volatility. **Journal of Financial Economics**, v. 61, n. 1, p. 43-76, 2001.

ANDERSEN, T. G.; BOLLERSLEV, T. Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. **Journal of Finance**, v. 52, n. 3, p. 975-1005, 1997a.

_____. Intraday periodicity and volatility persistence in financial markets. **Journal of Empirical Finance**, v. 4, n. 2-3, p. 115-158, 1997b.

_____. Answering the skeptics: Yes standard volatility models do provide accurate forecasts. **International Economic Review**, v. 39, n. 4, p. 885-905, 1998a.

_____. Deutschemark-dollar volatility: intraday activity patterns, macroeconomic announcements, and longer-run dependencies. **Journal of Finance**, v. 53, n. 1, p. 219-265, 1998b.

ANDERSEN, T.G.; BOLLERSLEV, T.; LANGE, S. Forecasting financial market volatility: sample frequency vis-a'-vis forecast horizon. **Journal of Empirical Finance** , v. 6,

n. 5, p. 457- 477, 1999.

ANDERSEN, T. G.; BOLLERSLEV, T.; CAI, J. Intraday and interday volatility in the Japanese stock market. **Journal of International Financial Markets**, Institutions and Money, v. 10, n. 2, p. 107-130, 2000.

ANDERSEN, T. G. et al. Great realizations. **Risk Magazine**, v. 13, p. 105-108, 2000.

ANDERSEN, T. G. The distribution of realized exchange rate volatility. **Journal of the American Statistical Association**, v. 96, n. 453, p. 42-55, 2001.

ANDERSEN, T. G. Modeling and forecasting realized volatility. **Econometrica**, v. 71, n. 2, p. 579-625, 2003.

AN, T.; MTAIS, C. The distribution of realized variances: marginal behaviors, asymmetric dependence and contagion effects. **International Review of Financial Analysis**, v. 18, n. 3, p. 134-150, 2009.

KUROWICKA, D.; COOKE, R. M. Conditional, partial, and rank correlation for the elliptical copula; Dependence modeling in uncertainty analysis. Proceedings ESREL, 2001. In: Zio, M.; Demichela, N.; Piccinini, (Ed.). Safety and Reliability, **Proceedings of ESREL 2001**, an International Conference on Safety and Reliability, September 16-20, 2001, Torino, Italy.", Politecnico di Torino, 2001.

COOKE, R. M.; BEDFORD, T.J. **Reliability methods as management tools**: dependence modeling and partial mission success. ESREL95, 153160. London, 1995.

BEDFORD, T.; COOKE, R. M. Probability density decomposition for conditionally dependent random variables modeled by vines. **Annals of Mathematics and Artificial Intelligence**, v. 32, n. 1-4, p. 245-268, 2001.

———. Vines - a new graphical model for dependent random variables. **Annals of Statistics**, v. 30, n. 4, p. 1031-1068, 2002.

DIEBOLD, F. X.; GUNTHER, T.; TAY, A. S. Evaluating density forecasts, with applications to financial risk management, **International Economic Review**, v. 39, n. 4, p.863-883, 1998.

HANSEN, B. Autoregressive conditional density estimation. **International Economic Review**, v. 35, n.3, p. 705-730.1994.

JOE, H. **Multivariate models and dependence concepts**. Boca Raton, Flrida: Chapman & Hall, 1997. 399 p.

JOE, H.; LI, H.; NIKOLOULOPOULOS, A. K. Tail dependence functions and vine copulas. **Journal of Multivariate Analysis**, v. 101, n. 1, p. 252-270, 2010.

KLPPELBERG, C.; KUHN, G.; PENG, L. Semi-parametric models for the multivariate tail dependence function the asymptotically dependent. **Scandinavian Journal of Statistics**, v. 35, n. 4, p. 701-718, 2008.

KUROWICKA, D.; COOKE, R. M. Completion problem with partial correlation vines. **Linear Algebra and its Applications**, v. 418, n. 1, p. 188-200, 2006.

MISIEWICZ, J.; KUROWICKA, D.; COOKE, R. M. Elliptical copulae. In: SCHUELLER, G. I. ; SPANOS, P. D. (Eds.) **Monte Carlo Simulation**: Proceedings of the International Conference on Monte Carlo Simulation. Rotterdam: Balkema, 2001.

TSAY, R. S. **Analysis of Financial Time Series**. 2nd ed. New Jersey: J. Wiley, 2005. (Wiley Series in Probability and Statistics).

YAN, B.; ZIVOT, E. **Analysis of High-Frequency Financial Data with S-PLUS**, [Working Papers], n. 03, University of Washington, Department of Economics, 2003.

ZIVOT, E. **Analysis of high frequency financial data**: models, methods and software. Part II:Modeling and forecasting realized variance measures. Working paper, n. 3, University of Washington, Department of Economics, 2005.