working paper
1803

Volatility, diversification and contagion

Enrique Sentana

March 2018

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 1803
March 2018

Volatility, diversification and contagion

Abstract

In this paper I describe in detail the concepts of volatility, diversification and contagion, three basic keys to understand the seemingly whimsical behaviour of financial markets. The presentation is deliberately non-technical and largely self-contained, with most required concepts defined along the way. Nevertheless, the analysis is mostly empirically oriented, with an emphasis on the methods that have been proposed to measure those concepts and a discussion of the stylised facts that the resulting measures imply. I also use those measures to study the effects of the financial crisis of 2007-2008 and the euro sovereign debt crisis of 2010-2012 on Spain.

JEL Codes: C58, G11, G32.

Keywords: Correlation, dependence, stock markets, volatility derivatives.

Enrique Sentana
CEMFI
sentana@cemfi.es

Acknowledgement

Prepared for the 25th anniversary issue of Revista de Economía Aplicada. Some of the material is based on joint work with Dante Amengual, Gabriele Fiorentini and Javier Mencía, to whom I am grateful. I am also grateful to José Gutiérrez for able research assistance, and to an anonymous referee and the audience of the CaixaForum lectures on "Financial culture for our time: Volatility, diversification and contagion" for their questions and remarks. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Economy, Industry and Competitiveness through grant ECO-2017-89689 and the Santander - CEMFI Research Chair on Finance is gratefully acknowledged.

Introduction

Life is uncertain but some risks are worth taking. Unfortunately, it is not easy to decide which ones fall in that category. We need to estimate the benefits and costs of our actions as well as the odds of success and failure. Nowhere is this perhaps more evident than in financial markets, which are an integral part of the global economy. Nevertheless, laypeople often have the impression that the prices of stocks, bonds and other financial assets, as well as the exchange rates between different currencies, go up and down without any apparent pattern, as if they reacted to the whims of the gods or the instincts and emotions of the participants in those markets.

However, in the same manner as classical science rose to the challenge of explaining the apparent motion of the planets, economics have slowly but steadily made progress in putting some order into seemingly chaotic financial markets. In fact, nowadays financial volatility is one of the best understood topics among research economists. The same is true of the relationship between different financial assets, which sometimes seem to move independently and others at unison, especially after big price falls, as if they followed the pattern of a communicable disease.

In this paper I describe in detail the concepts of volatility, diversification and contagion, three basic keys to understand the seemingly whimsical behaviour of financial markets. After introducing those concepts and briefly reviewing some of the approaches developed to measuring them, I will use several examples of the recent financial crisis of 2007-2008 and the euro sovereign debt crisis of 2010-2012 for illustrative purposes, paying special attention to the effects they had on the Spanish stock market. The approach that I take is deliberately empirical, highlighting stylised facts instead of theoretical models. Given that the academic literature on these topics is huge, I will often immodestly refer to my own work, which contains detailed references to many other more substantial contributions.

The rest of the paper consists of three main sections centred around those key concepts, followed by some conclusions and a list of references.

1 Volatility

1.1 Financial assets and their returns

Before talking about volatility, it is convenient to introduce some notation. Let D_t be the payoff received by the owner of a financial asset at the end of period t, which usually takes the form of a coupon for bonds or a dividend for stocks. A period could be a day, a week, a month, etc. depending on the context. Further, let P_t denote the price of this financial asset at the end of period t once the payoff has been made. This price is often known as the "ex coupon" or "ex dividend" price because at that point the asset no longer includes the right to perceive D_t.

Although the prices of financial assets attract all the media attention, research economists prefer to work with returns, which reflect all the gains investors really obtain. In addition, they make both intertemporal and across asset comparisons much easier. For example, the largest ever fall in the Dow Jones Industrial Average so far has been the 1,175-point wipe out that took place on February 5th, 2018. However, this represented a capital loss of 4.6% only, while the 508-point fall on the so-called "black Monday" (October 19th, 1987) represented a drop of almost 23%. Thus, we can define

$$R_t = \frac{D_t + P_t}{P_{t-1}}$$

as the one period holding return over period t on an asset bought at the end of period $t-1$. This gross return, which represents the total payoff per unit invested, should not be confused with the commonly reported net return, $R_t - 1$, which can conveniently be additively decomposed into the dividend yield D_t/P_{t-1} and the proportional capital gain $(P_t - P_{t-1})/P_{t-1}$.[1] In turn, these gross and net returns, which are arithmetic in nature, should not be confused with the geometric return, $\ln R_t$, which measures the instantaneous rate such that if it were continuously compounded between $t-1$ and t, it would yield precisely $\exp(\ln R_t) = R_t$.[2]

[1] Net returns are often reported on % terms and annualised, although the annualisation process is far from uncontroversial, as the recent discussion of the European Securities and Markets Authority (ESMA) regulation of the key information document (KID) for packaged retail investment products (PRIPs) illustrates [see Financial Conduit Authority (2018), as well as Lo (2002) for more formal arguments].

[2] Arithmetic gross and net returns conveniently aggregate across assets, as I will show in section 2, but they do not exactly aggregate over time. For example, a sequence of net returns of 10% and -10% (or vice versa) leaves a investor with 99 cents per unit invested. In contrast, geometric returns exactly aggregate over time but they only aggregate across assets for investments of infinitesimal duration [see chapter 1 in Campbell, Lo and MacKinlay (1997) for further details].

In what follows, I will assume that there is a (conditionally) safe asset, whose returns over period t, say R_{st}, are known with certainty at time $t-1$, although they can change from period to period. The interbank loans underlying the Euro OverNight Index Average (EONIA) rate, and short term Treasury bills would be two obvious examples. The existence of a safe asset, though, is not a trivial assumption for at least two reasons. First, investors often care about the value of their investments in real terms, that is, in terms of the amount of goods they will be able to buy in the future. Although inflation has recently been very low and predictable in the eurozone and most other developed countries, there are other countries and historical periods in which this is certainly not the case.[3] Second, the issuers of these supposedly safe assets, which are often sovereign states or large financial institutions, might default on their obligations, or at least modify the terms of their repayments, as we have seen in recent financial crises.

In any case, in what follows I will denote by $r_t = R_t - R_{st}$ the returns on a risky asset in excess of the returns of the safe asset. Interestingly, those excess returns are the payoffs at time t to an investor who buys the risky asset on credit at the end of period $t-1$ using funds she has borrowed at the safe rate. As a result, r_t cannot be understood as a payoff per unit invested because there are no net outflows at time $t-1$. In fact, if I further assume that there are constant returns to scale in the investment technology, so that the returns of buying 100 units of an asset are 10 times the returns of buying 10 units,[4] r_t could be scaled up or down as much as desired because kr_t would simply represent the payoffs at time t to an investor who buys k units of the risk asset on credit at the end of period $t-1$. The multiplicative factor k is usually referred to as the degree of leverage of the position. Although it may seem that only large institutional investors can benefit from leverage, financial forward and futures contracts allow retail investors to modify k rather easily, even to make it negative by taking the opposite side of the contract.[5]

[3] Inflation-linked bonds provide protection about future unexpected falls in the purchasing power of the currency in which they are denominated, but because of lags in the reporting of price indices, they are not 100% perfect.

[4] This is a reasonably good approximation for medium size investments but not for very small or very large ones due to pure transaction costs, market microstructure effects, and slippage (the usually negative impact of investors' own actions on prices), which are closely related to liquidity, especially during fire sales.

[5] A forward is a contract in which the parties agree to exchange the underlying asset at a certain future date T (the delivery date) at a certain price K specifed at the contract date t. The forward price K is determined by the condition that its initial value be zero, so payments between the parties only take place at the expiration date. When $\tau = T - t = 1$, those payoffs can be exactly replicated by a

If investors only care about the payoffs they will receive in the future regardless of the circumstances in which they accrue,[6] the probability distribution of returns contains all the relevant information for making investment decisions. Figure 1 depicts the density functions of three hypothetical return distributions: a normal, a symmetric Student t with 10 degrees of freedom and an asymmetric Student t with 10 degrees of freedom too but skewness parameter -1.5 [see Mencía and Sentana (2012) for details]. These functions have a very intuitive interpretation: the area under a density function between points a and b (say -1 and 0) measures the probability that the realised returns will take some value larger than a but smaller than b.

Unfortunately, most investors, including professional ones, suffer an information overload when looking at these densities. For that reason, they often focus on a few summary numbers. Specifically, many only pay attention to the mean (or expected value) and standard deviation of the return distribution. As is well known, the expected return is defined as the sum of all possible returns one might obtain weighted by the probability of obtaining them. More formally,

$$\nu = E(R) = \int_{R_{\min}}^{R_{\max}} u dF(u),$$

where $\int$ means the "sum" between the smallest and largest possible returns, $R_{\min}$ and $R_{\max}$, and $dF(u)$ the probability of R taking a value arbitrarily close to u. It is the most common measure of the centre (or location) of a return distribution.

On the other hand, the standard deviation is defined as the square root of the variance, which in turn is the sum of the square differences between the returns and their expected values, weighted again by the probability of observing them. More formally,

$$\sigma^2 = V(R) = \int_{R_{\min}}^{R_{\max}} (u-\nu)^2 dF(u). \tag{1}$$

The standard deviation is the most common measure of return dispersion (or scale), the square root guaranteeing that it has the same units as the returns and their mean. In fact, it is often reported as the sole measure of risk of a financial asset, even though

portfolio which buys one unit of the underlying risky asset by selling P_t units of the safe one. Therefore, absence of arbitrage requires that $K = P_t R_{st}$.

[6] Socially responsible investors, who avoid certain industries or commercial practices, as well as those who might refuse to benefit from shortages resulting from natural catastrophes, violate this assumption. Similarly, buyers of unit linked products with a life insurance component covering their own death most likely treat future payoffs differently depending on whether they or their heirs receive them.

mean and standard deviation only provide a partial characterisation of a probability distribution, except in some special but empirically unrealistic cases, such as when returns are Gaussian. This point is clearly illustrated by the markedly different shapes of the densities in Figure 1, even though all three underlying random variables have zero mean and unit standard deviation by construction.

Interestingly, the mean excess return, $E(r)$, which I will henceforth denote by μ, is trivially $\nu - R_s$, while $V(r) = V(R)$ precisely because the safe asset is riskless. A popular descriptive measure of the excess returns on a risky asset is its Sharpe ratio, which is the ratio of its mean μ to its standard deviation σ. It was introduced by Sharpe (1966) to measure the performance of an investment irrespective of its leverage.

The unconditional standard deviation of a financial return series relates to its dispersion over the long run. Volatility, in contrast, is usually defined as the conditional standard deviation

$$\sigma_t^2 = Var(r_t|I_{t-1}),$$

where I_{t-1} denotes the information available to market participants at time $t-1$, when they make their investment decisions. In principle, one would expect volatility to change over time as new information is observed. For example, a priori volatility should be larger at the outset of an economic downturn than half way through an expansion. Surprisingly, though, volatility was traditionally assumed to be constant by both financial economists and market participants. However, the 1970's oil crises and the economic policies adopted to mitigate their effects showed that the assumption of constant volatility in financial markets was untenable. The "black Monday" of October 1987, when most stock markets around the world fell in unison, convinced even the most recalcitrant believers. Nowadays neither academics nor financial market participants question the time varying nature of volatility, which manifests itself through three different channels:

a. volatility clusters, with fairly calmed – low uncertainty - phases of reasonable but random length in which price changes tend to be small followed by more turbulent – high uncertainty - periods with large price changes,

b. fat tailed unconditional distributions for returns with both too many big movements and too many modest ones relative to the "normal", and

c. substantial serial correlation in the magnitudes (but not the levels) of returns.

1.2 Volatility measures

Volatility is a crucial ingredient in many different areas of finance: asset allocation, option pricing, risk management or systemic risk measurement. Unfortunately, in real life the true distribution of the returns on an asset between $t-1$ and t, given the information available to investors at time $t-1$, is unknown. We only observe ex post what has happened, a single number, not all that might happen ex ante. We would need a machine that could travel across the multiverse of "parallel universes" to observe all possible return realisations and their relative frequencies. Although some theoretical physicists dream of this machine, it remains science fiction. Therefore, volatility is one of those elegant concepts favoured by economists which cannot be unequivocally measured, a fact that sometimes raises eyebrows among scientists from other disciplines.

Still, economists have developed ingenious ways of filtering volatility out from data. In fact, there are at least four common ways of measuring volatility in financial markets:

1. Sample standard deviations over rolling windows The simplest vol estimate is the square root of the so-called historical variance

$$his_{t,S} = \frac{1}{S}\sum_{s=1}^{S} r_{t-s}^2, \tag{2}$$

which is the average of the daily square excess returns computed over a rolling window of the previous S trading days.[7] This is a fairly natural way of estimating σ_t^2 because the use of the most recent observations helps keep track of changes in volatility. However, it has at least two shortcomings. First, the length of the rolling window must be chosen somehow. If S is too large, $his_{t,S}$ will be essentially constant. If, on the other hand, S is very small, say 2 or 3, estimated volatility will be very noisy. Therefore, in practice a compromise has to be found, usually by trial and error. But even if one chooses a sensible value for S, there is a second, more fundamental problem. The expression above gives equal weight to all observations in the current rolling window, but zero weights to all prior observations. However, when volatility changes slowly over time, it seems more natural to give proportionally more weight to the most recent observations and less weight to the most distant ones. The popular commercial risk management software

[7] There are several popular slight variations of this expression. Some subtract the mean excess return before squaring, as in expression (1), while others use geometric returns instead of arithmetic ones.

RiskMetrics does precisely that. The default estimate of volatility at the core of its code is based on the following exponentially weighted moving average (EWMA)

$$RM_t = (1-\lambda)\sum_{s=1}^{\infty}\lambda^{s-1}r_{t-s}^2 = (1-\lambda)r_{t-1}^2 + \lambda RM_{t-1}, \tag{3}$$

where λ is a decay parameter between 0 and 1, which, according to the original RiskMetrics (1996) manual, should be set to .94 for daily observations. In effect, this decay parameter plays the role that S played in the historical vol estimate (2), the half-life of the weights being 11.2(=ln(2)/ ln(.94)) days (i.e. the weight on day 11 is approximately half of the weight on day 1).

2. Econometric models The acronym ARCH (Autoregressive conditional heteroskedasticity) refers a class of parametric time series models for volatility proposed by Robert Engle (1982). His original baseline specification, the ARCH(1) model, is not empirically realistic, but it led to the development of an entire subdiscipline: Financial Econometrics. In fact, his proposal became so influential that the Royal Swedish Academy of Sciences awarded him the Nobel Memorial Prize in Economic Sciences in 2003 "for methods of analyzing economic time series with time-varying volatility". The most popular version is the GARCH(1,1) specification put forward by Tim Bollerslev (1986), one of Engle's PhD students, which is such that

$$\sigma_t^2 = \frac{\theta}{1-\alpha-\beta} + \alpha\sum_{s=1}^{\infty}\beta^{s-1}r_{t-s}^2 = \theta + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2,$$

where θ, α and β are parameters to be estimated from the data. In this sense, expression (3) can be regarded as a special case of a GARCH(1,1) model in which $\theta = 0$, $\alpha = 1-\beta$ and $\beta = .94$.[8] There are many other variations of the ARCH model, including the GQARCH(1,1) process I proposed in Sentana (1995):

$$\sigma_t^2 = \theta + \alpha r_{t-1}^2 + \gamma r_{t-1} + \beta\sigma_{t-1}^2,$$

which allows volatility to increase more after negative returns than after positive ones, thereby capturing an empirical phenomenon known as the "leverage effect", but without sacrificing analytical tractability.[9]

[8] A GARCH(1,1) model in which $\alpha + \beta = 1$ is said to be Integrated. Nelson (1990) proved that an Integrated (IGARCH) process with $\theta = 0$ would converge to 0 with probability 1, so the RiskMetrics model (3) can only be understood as a volatility filter, as opposed to the true data generating process.

[9] This model was used by BARRA, another major commercial provider of risk management solutions for the financial industry, in its US equity model to capture the volatility of the thousands of stocks traded in the New York Stock Exchange [see BARRA (1998)].

The parameters of this and many other variants can be readily estimated by maximum likelihood[10] using the numerical optimisation routines provided by the dominant commercially available econometric packages, such as EVIEWS and STATA [see IHS Global Inc (2015) and StataCorp LP (2015)]. A minimum requirement for a GARCH model is that it should not generate negative volatilities. In the standard GQARCH(1,1) case, this is guaranteed when $4\alpha\theta - \gamma^2 \geq 0$, $\alpha \geq 0$ and $\beta \geq 0$, but in most higher-order models the required positivity conditions remain largely unknown.[11]

The main conceptual problem with the ARCH family of models is that they all implicitly assume that the volatility of an asset is a function of its own past returns only, thereby ignoring all other possible sources of publicly available information. The stochastic volatility literature attempts to break this link by assuming that volatility follows a process of its own, with shocks that cannot be fully accounted for by past returns [see Andersen and Shephard (2009)]. From the practitioners' point of view, though, the estimation of the parameters characterising stochastic volatility models require fairly sophisticated simulation techniques [see e.g. Kim, Shephard and Chib (1998) or Calzolari, Fiorentini and Sentana (2004)], which have not yet been coded by the most popular econometric packages.

3. Realised volatility using intraday data In the 1990s, the ultra high frequency intraday observations on foreign exchange quotes and prices made available to researchers by Charles Goodhart at the LSE and Richard Olsen in Zurich allowed them to compute daily measures of "realised" volatility analogous to the rolling window ones, but based on geometric returns over 15-minutes, 5-minutes, 1-minute or even higher frequencies.[12]

Following an explosion of work formally linking these model-free measures to quadratic variation, which is a fundamental concept in the theory of stochastic processes [see e.g. Barndorff-Nielsen and Shephard (2007)], these realised volatility measures have become very popular among researchers and financial market participants, not only because they can be estimated in real time during a trading session, but also because they provide better forecasts of future realised volatility than the measures based on daily

[10] This is a classical estimation procedure which effectively chooses the parameter values that maximise the ex-ante probability of observing the data we see in the sample.

[11] But see Nelson and Cao (1992) and Demos and Sentana (1992) for some special cases.

[12] French, Schwert and Stambaugh (1987) carried out an early application of this approach in which they used daily data to compute realised volatility at the monthly frequency.

observations that I have previously discussed [see e.g. Andersen, Bollerslev and Diebold (2009)].

Although many academic studies have proposed clever modifications of these simple realised volatility measures that deal with market micro-structure effects arising from the fact that in practice a financial asset has not a single price but an entire pricing curve which depends on the number of units an investor might want to buy or sell [see Barndorff-Nielsen et al (2006) and Zhang, Mykland and Aït-Sahalia (2005)] subsampling realised variance estimates based on the simple average of the variances of the 5 possible series of 5-minute returns one can construct on any given day (covering minutes 1-5, 2-6, 3-7, 4-8 and 5-9, etc.) remain the most popular [see Liu, Patton and Sheppard (2015)]. Nevertheless, further work is needed in two important practical areas: the treatment of overnight returns, and infrequently traded assets [see Hansen and Lunde (2005)].

4. Implied vols obtained from financial derivatives Formally, a European call option with strike price K and expiration T written on an asset whose current price is P_t is a contract that gives the holder the right but not the obligation to buy the asset at the expiration date T at the exercise (or strike) price K.[13] Therefore, the holder of this option will presumably exercise it if and only if the price of the asset at expiration, P_T, exceeds the strike price. The seller of the option, in contrast, has the obligation to sell the asset for K to any buyer who wants to exercise her option, even if she could buy the asset in the spot market at a lower price. Black and Scholes (1973) and Merton (1973) derived a closed-form expression for the price of a European call option on a non-dividend paying asset under the assumption that its geometric return between t and T follows a normal distribution with constant variance σ^2 for all T.[14] Their formula, known as the Black - Scholes formula, makes the price of the call option a function of four observable quantities: the current price, P_t, the strike price, K, the time to maturity, $\tau = T - t$, and

[13] There are many other varieties of call options, including American options, which give the right to buy the asset at any point before the expiration date, Bermudan options, in which the early exercise can only happen at some pre-specified dates, and more exotic versions, such as Canary and Verde options, whose names reflect the fact that their namesake archipelagos lie in the middle of the Atlantic ocean, which separates the European and American continents. In addition, there are put options, whose European variety gives the holder the right but not the obligation to sell the underlying asset at the strike price on the expiration date.

[14] The idea is to create a self-financing portfolio of the underlying and safe assets that replicates the final payoff of the call, so the value of the option at any time is equal to the value of this portfolio. The investment strategy is dynamic in the sense that the weights of the two assets are continuously adjusted according to the evolution of the price of the underlying asset.

the constant (continuously compounded) safe interest rate $\ln R_s$; and one unobservable one: the volatility of the underlying asset, σ.[15] Specifically,

$$Call^{BS}(P_t, K, \tau, r, \sigma) = P_t\Phi(d) - KR_s^{-\tau}\Phi(d - \sigma\sqrt{\tau}),$$

where $\Phi(.)$ is the cumulative distribution function of a standard normal variable (see Figure 1), which is computed as the integral of its density function $\phi(.)$ such that $\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(u)du$, and

$$d = \frac{1}{\sigma\sqrt{\tau}}\left[\ln\left(\frac{P_t R_s^\tau}{K}\right) + \frac{1}{2}\sigma^2\tau\right].$$

Importantly, $\ln(P_t R_s^\tau/K)$, which is the (log) ratio of the current price of a forward contract written on the asset to the strike price, is a measure of the "moneyness" of the option. Thus, "in the money" options that at expiration would always be exercised will tend to have a positive value for this "moneyness" indicator, which will become negative for "out of the money" options that at expiration would never be exercised. Finally, at the money options are those whose strike price (approximately) coincides with the forward price.

By combining the actual market price of a call option with P_t, K, τ and R_s, one can invert the Black - Scholes formula to obtain the implied value σ. For that reason, this estimate of σ is known as the Black - Scholes implied vol. Given that the call price is monotonically increasing in σ, the existence and uniqueness of the solution is guaranteed. In addition, given that the payments at expiration of a call option are equal to those of a portfolio with one put option, one unit of the risky asset, and $-K$ units of the safe asset, the law of one price implies the model-free put-call parity condition:

$$Call_t(K, \tau) = Put_t(K, \tau) + P_t - KR_s^{-\tau},$$

where $Put_t(K, \tau)$ and $Call_t(K, \tau)$ are the market prices at time t of European put and call options written on P_t with strike K and time-to-maturity τ, which allows one to obtain implied vols from put option prices too.

In practice, however, the Black-Scholes-Merton assumptions are empirically unrealistic, and instead of a single vol number, one ends up with volatility smiles or smirks,

[15] Robert Merton and Myron Scholes also received the Nobel Memorial Price in Economics in 1997 "for a new method to determine the value of derivatives". Fischer Black had passed away two years before.

which depict an entire curve of values of Black-Scholes implied vols obtained from call and put options with different strike prices but the same expiration date. In fact, there is a term structure of implied volatility smiles and smirks which varies with the time to maturity τ, usually represented by means of an implied vol surface.

Initially, Black-Scholes implied volatility measures were computed from at the money options because they are often the most actively traded. However, this procedure ignored all other option prices. Although there are many possible ways of summarising the information in those volatility smiles in a single number, nowadays the industry standard is, in effect, a nonparametric procedure based on an alternative over the counter derivative asset known as a variance swap.

In a variance swap, the buyer will pay an amount that depends on the realised variance of the daily geometric (log) returns of the underlying asset between its inception and expiration in exchange for a fixed amount (the strike), which is set at the time the contract is signed. Under certain assumptions, a variance swap may be replicated using a portfolio of European call and put options with weights inversely proportional to the square of their strikes. Specifically, the "variance swap rate" can be computed from observations across all strikes as follows:

$$\tau\sigma_\tau^2(t) = 2\int_0^{P_t} \frac{\{1+\ln[P_t/K]\}}{K^2} Put_t(K,\tau)dK + 2\int_{P_t}^{\infty} \frac{\{1-\ln[K/P_t]\}}{K^2} Call_t(K,\tau)dK \quad (4)$$

[see Bakshi, Kapadia and Madan (2003) for further details, and Martin (2011) for a critical review of the variance swap and a more robust alternative derivative asset].[16]

The Chicago Board Options Exchange (CBOE) volatility index, widely known by its ticker symbol VIX, is the best known application of this methodology, and has effectively become the standard measure of volatility risk for investors in the US stock market, at least judged by the amount of prime time devoted to it in the main business channels. Although it was originally introduced in 1993 to track the Black-Scholes implied volatilities of options on the S&P100 with near-the-money strikes, its rise to stardom came in 2003 when the CBOE redefined it as a model-free measure and released a time series of daily closing prices starting in January 1990. Nowadays, VIX is computed in real time using as inputs the mid bid-ask market prices for most calls and puts on the S&P500

[16] To compute the integral that yields the desired value of the variance swap, $\sigma_\tau^2(t)$, one would need a continuum of option prices. In practice, one can compute a discretised approximation to those integrals based on the available options. To do so, one can linearly interpolate their prices inside the observable range, and extrapolate them by keeping the implied volatilities constant at the extremes.

index for the front month and the second month expirations [see CBOE (2009a) for details].

Formally, it is the square root of the risk neutral expectation of the integrated variance of the S&P500 over the next 30 calendar days, reported on an annualised basis. Despite this rather technical definition, both financial market participants and the media pay a lot of attention to its movements. To some extent, its popularity is due to the fact that VIX changes are negatively correlated to changes in stock prices, the most plausible explanation being that investors trade options on the S&P500 to buy protection in periods of market turmoil, which increases the value of the VIX.

The advantage of these implied vol measures is that they are forward looking. In that regard, Blair, Poon and Taylor (2001) present evidence showing that there is little additional information in high frequency data for the purposes of predicting future vol once lagged implied vols are taken into account. But as Andersen and Bodarenko (2007) and many others show, implied vol often contains a positive risk premium, in the sense that it almost uniformly exceeds realised volatility because investors are on average willing to pay a sizeable premium to acquire a positive exposure to future equity-index volatility. For that reason, some commentators refer to the VIX as the market's fear gauge, even though a high value does not necessarily imply negative future returns. In fact, there is a complex dynamic relationship between volatility and returns: over the long run, higher volatility is usually associated with higher average returns, but over the short run, sudden increases in volatility often coincide with market falls, which in turn lead to further increases in volatility.

The VIX has been so successful that the CBOE currently applies the same methodology to an ever increasing set of financial assets, including 3-month options on the S&P500, as well as 1-month options on the most important US stock market indices: DJIA, S&P100, Nasdaq 100 and Russell 2000. They also construct analogous short term volatility indices for several actively traded individual stocks, including Amazon, Apple, Goldman Sachs, Alphabet (formerly known as Google) and IBM, as well as international stock indices for developed markets, emerging markets, China and Brasil. In addition, there are volatility indices for 10-year Treasury notes, interest rate swaps, crude oil, gold and the US \$/euro, US \$/yen and US \$/sterling exchange rates.

1.3 Stylised facts

Figure 2a displays the temporal evolution of the VIX over its entire history. Visually, one can easily see two salient characteristics, which are also shared by the RiskMetrics EWMA estimates, volatilities generated by GARCH models and realised volatilities computed from ultra high frequency data:

a. Sudden spikes, in which volatility jumps from (relatively) low levels to (relatively) high ones, often in the course of a single day,

b. Slow, exponential declines, which bring down volatility to more reasonable levels following a spike.

Recent examples of the first phenomenon are the surge in volatility in August 2015 motivated by concerns about the Chinese economy, or the recent increase on February 5 this year following some indicators of inflation pressures building up in the US at the time of a changing of the guard at the Federal Reserve and an unusually expansionary fiscal policy on the part of the Trump administration. The sudden spikes make predicting volatility movements a very difficult task. Nevertheless, mean reversion implies that even in the worst days of the 2008 global financial crisis, investors expected volatility to return to more reasonable values, eventually [see Schwert (2011)]. This is confirmed in Figure 2b by the fact that during the autumn of 2008, the VIX3M, which has a constant three-month horizon, was clearly below the VIX, which has a one-month horizon. In contrast, in the fourth quarter of 2017, the opposite was true.

However, there is a third characteristic which can only be observed over such a long time span: volatility contains a slowly moving trend which is itself very slowly mean reverting to the long run mean, with high volatility phases alternating with low volatility ones that can last for several quarters or even years.[17] For example, volatility was remarkably low between February 2006 and July 2007, with values well below the long run historical average of around 20. In fact, the lowest value over this period was 9.89 on January 24, 2007, in what some called "the calm before the storm".[18] Over the following year, VIX increased to values between 20 and 35. Finally, in the autumn of 2008 it reached unprecedented levels, with the largest historical closing price (80.86) taking

[17] This phenomenon has been related to the long memory properties of the Nile flooding cycle, popularly known for the biblical reference to seven years of plenty and seven years of famine.

[18] Even lower values were observed in the last few months of 2017 and again in January 2018, with an all-time lowest intraday value of 8.56 on Friday 24 November 2017.

place on November 20, 2008, although on October 24 the VIX reached an intraday value of 89.53. After this peak, VIX followed a decreasing trend over the following months until the beginning of April, 2010, when the Greek debt crisis unfolded.

1.4 Volatility derivatives

Nowadays it is possible to directly invest in volatility by means of VIX derivatives [see Rhoads (2011)]. Specifically, in March 2004, trading in futures on the VIX began on the CBOE Futures Exchange (CFE). Further, in February 2006, European-style options on the VIX index were also launched on the CBOE. Like VIX futures, they are cash settled according to the difference between the value of the VIX at expiration and their strike price. More recently, several volatility-related Exchange Traded Notes (ETNs) have provided investors with equity-like long and short exposure to constant maturity futures on the VIX, and even dynamic combinations of long-short exposures to different maturities. Although the poor performance of investors with long positions on some of these derivative assets during steadily decreasing volatility periods, as well as those holding short positions during volatility spikes, have raised some serious concerns about their risks and their suitability for retail investors, especially those with a heart condition, trading in VIX-related assets recently reached unprecedented levels, representing a non-neglibigle fraction of total volume trade on U.S. exchanges.

One of the main reasons for the high interest in these products is that VIX derivative positions can be used to provide protection against the risks inherent in the S&P500 index, especially in downturns. At the same time, VIX derivatives allow investors to achieve exposure to S&P500 volatility more cheaply than by using traditional derivatives on this broad stock market index.

Although these new assets certainly offer additional investment and hedging opportunities, their correct use requires reliable valuation models that adequately capture the features of the VIX, which is an index and not a tradeable asset.

In Mencía and Sentana (2013, 2016), we developed two complementary approaches to price VIX derivatives. Although they differ in many important details, both of them exploit the empirically relevant features of the VIX mentioned in the previous section. The markedly different volatility periods observed over recent years provide a very useful testing ground to assess the empirical performance of the different pricing models that we consider: a continuous time model for the log of the VIX and a discrete time process for its level, which explicitly takes into account the positivity of this index.

As usual, we analyse the discrepancies between actual and theoretical derivatives prices. But we also go beyond pricing errors, and analyse the implications of our models for the term structures of VIX futures and options, which are of considerable independent interest. Since we combine futures and options data, we can also assess which features of our models are more relevant for pricing futures, and which ones are more important for options. Moreover, by combining data on VIX derivatives with historical data on the VIX itself in the estimation of the model parameters, we can not only cover a much longer time span with different volatility phases, but also look at risk premia.

2 Diversification

2.1 Portfolio choice and diversification gains

So far I have considered a single risky financial asset, but in practice there are plenty to choose from. For that reason, it is of the utmost practical importance to consider multiple assets simultaneously. Let $\mathbf{R} = (R_1, R_2, \ldots, R_N)'$ the vector of gross returns on a finite set of N risky assets. The discussion in section 1.1 allows us to think about the N marginal distributions of those risky assets on an asset by asset basis, which in turn allows us to figure out the N associated expected returns $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_N)'$ and standard deviations, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$. However, those objects tell us nothing whatsoever about the joint probability distribution of returns.

The relevant object is the joint density, such as those depicted in Figure 3 for the case of two risky assets whose distribution is either bivariate normal, a symmetric Student t with 8 degrees of freedom or an asymmetric Student t with 8 degrees of freedom but skewness parameters (-2,-2) [see again Mencía and Sentana (2012) for details]. Once again, we can interpret the volume under those functions over a square with opposite vertices (a_1, a_2) and (b_1, b_2) as providing the probability that the first return takes values in the range (a_1, b_1) **and** the second one takes values in the range (a_2, b_2) **simultaneously**. If most investors suffer from information overload when looking at a single marginal distribution, one can only wonder how their brain scans will look like when they try to visualise these joint densities in their minds.

Financial engineering has a bad press reputation but at the end of the day its simple objective is to combine existing assets to create new ones. Under my maintained assumption of constant returns to scale in the investment technology, the payoffs to a portfolio of the riskless asset and the N risky ones with fixed weights given by w_s and the vector $\mathbf{w} = (w_1, w_2, \ldots, w_N)'$, respectively, will be given by $p = w_s R_s + w_1 R_1 + \ldots + w_N R_N$. Therefore, investors could safely ignore joint distributions if they could figure out the probability density function of p for every conceivable combination of weights $\mathbf{w}$. Unfortunately, except in some special cases I will discuss in section 3.1, this is far from trivial. For that reason, investors often characterise dependence by means of the pairwise covariance between any two risky financial returns.

A straightforward generalisation of the variance concept, covariance is defined as the sum of the cross products of the differences of each return to its expected value, weighted by the probability that such a pair of returns simultaneously realises. More formally,

$$\sigma_{ij} = cov(R_i, R_j) = \int_{R_{i\min}}^{R_{i\max}} \int_{R_{j\min}}^{R_{j\max}} (u_i - \nu_i)(u_j - \nu_j) dF(u_i, u_j),$$

where (u_i, u_j) are a possible combination of values for the two returns, while $dF(u_i, u_j)$ is the probability that R_i and R_j take values arbitrarily close to u_i and u_j, respectively. Trivially, $\sigma_{ij} = \sigma_{ji}$, so there is effectively a single covariance between any two assets. In addition, $V(R_i)$ can be understood as the covariance of an asset return with itself.

If combinations of two asset returns either above their means or below their means are more likely to occur than alternative combinations in which one asset return is above while the other one below their respective means, the covariance will typically be positive. In contrast, if the opposite happens then the covariance will typically be negative. Finally, when the returns of the two assets are (stochastically) independent, so that the probability of observing a given pair is exactly equal to the product of the probabilities of observing each of its components, and the “sum” above is well defined, their covariance will be 0. However, zero covariance is far weaker than independence. For example, one can easily construct counterexamples such as $R_2 = R_1^2$ in which there is extreme dependence but zero covariance if the distribution of R_1 is symmetric around its mean. As in the case of a single asset, therefore, mean, variances and covariances only provide a partial characterisation of a joint distribution. This point is clearly illustrated by the markedly different shapes of the densities in Figure 3, even though all three underlying bivariate random vectors have zero means, unit standard deviations and zero covariance by construction.

Interestingly, the covariance between the excess returns of assets i and j, $r_i = R_i - R_s$ and $r_j = R_j - R_s$, coincides with the covariance between their gross returns, the reason being once again that the safe asset is riskless.

Although as I explain below covariances play a fundamental role in the classical theory of optimal portfolio allocation, they are not ideal as descriptive measures because they are influenced by the way in which one measures returns (% or pure numbers, annualised, leveraged, etc.). A far more common measure of linear dependence is Pearson correlation coefficient

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \cdot \sigma_j},$$

which is effectively the covariance between two excess returns leveraged or deleveraged so that they both have unit volatility. This unitless measure always lies between -1 and +1, which facilitates comparisons across asset pairs. A correlation of 1 implies that there must be an upward sloping deterministic, linear relationship between the two variables while a value of -1 implies a downward sloping one.

One of the most important advantages of creating new assets by means of portfolios of risky assets is that they can have lower risk than any of the two original ones. In particular, in the case of two risky assets with gross returns R_i and R_j, the variance of a convex combination of them with weights w and $1-w$ will be:

$$\sigma^2(w) = w^2\sigma_i^2 + (1-w)^2\sigma_j^2 + 2w(1-w)\sigma_{ij}.$$

This expression is minimised for

$$w_{MV} = \frac{\sigma_j^2 - \sigma_{ij}}{\sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}},$$

which yields

$$\sigma^2(w_{MV}) = \frac{\sigma_i^2\sigma_j^2 - \sigma_{ij}^2}{\sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}},$$

so that

$$\sigma_i^2 - \sigma^2(w_{MV}) = \frac{(\sigma_i^2 - \sigma_{ij})^2}{\sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}} \geq 0$$

because the denominator is the variance of a portfolio that takes a long position on one of the assets and a short position on the other one. This (weak) inequality confirms that we can generally find a portfolio of unit cost whose variance is lower than the variance of each of its constituents. The only exception arises when the two assets are in fact identical. This simple mathematical argument is behind the idea of diversification gains, which is often expressed by the adage "don't put all your eggs in one basket".

Nevertheless, several observations are in order. First, in deriving w_{MV}, I have not imposed any leverage or short-selling restrictions on w, which could be either negative or bigger than 1. In the presence of short sale constraints, diversification gains diminish but they do not disappear. Second, and more important, minimising risk by minimising variance usually involves an opportunity cost in terms of reducing expected returns too. In fact, there is a trade off between risk and return. Modern portfolio choice theory, also known as mean-variance portfolio analysis, is effectively built around this trade-off.

Despite its simplicity, and the fact that over six and a half decades have elapsed since Markowitz published his seminal work on the theory of portfolio allocation under uncertainty [Markowitz (1952)], mean-variance analysis remains the most widely used asset allocation method. There are several reasons for its popularity. First, it provides a very intuitive assessment of the relative merits of alternative portfolios, as their risk and expected return characteristics can be compared in a two-dimensional graph. Second, it can be shown that return mean-variance frontiers are spanned by only two funds, a property that simplifies their calculation and interpretation. Finally, mean-variance analysis is fully compatible with expected utility maximisation if we assume Gaussian or elliptical distributions[19] for asset returns [see e.g. Chamberlain (1983), Owen and Rabinovitch (1983) and Berk (1997)], or if the mutual fund separation conditions hold [see Ross (1978)]. Not surprisingly, Harry Markowitz received the 1990 Nobel Memorial Prize in Economics, together with Merton Miller and William Sharpe, "for their pioneering work in the theory of financial economics".

Let $\mathbf{\Sigma}$ denote the $N \times N$ covariance matrix of returns, a square array whose diagonal elements contain the individual variances of the N assets and whose off-diagonal elements contain the covariances of all possible pairs. It is easy to see that the mean and variance of a portfolio with weights $\mathbf{w}$ written on these assets will be given by

$$\nu(\mathbf{w}) = \sum\nolimits_{i=1}^{N} w_i \nu_i = \mathbf{w}'\boldsymbol{\nu}$$

[19]Spherically symmetric distributions are a generalisation of the multivariate normal with zero mean vector and identity covariance matrix, in the sense that the density contours are concentric circles, but with flexibility is the relative assignment of probability between the centre and the tails. Examples include the multivariate normal and Student t depicted in Figure 3, as well as the multivariate Laplace and all scale mixtures of normals [see Amengual and Sentana (2010) for further details]. Elliptical distributions are simple affine transformations of spherical ones.

and

$$\sigma^2(\mathbf{w}) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij} = \sum_{i=1}^{N} w_i^2\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} w_i w_j \sigma_{ij} = \mathbf{w}'\mathbf{\Sigma}\mathbf{w}.$$

When a safe asset is available, the unrestricted optimal solution in the mean - variance sense is to take a unit position on the safe asset, and N positions on the zero cost portfolios underlying each of the excess returns with weights proportional to $\mathbf{\Sigma}^{-1}\boldsymbol{\mu}$.[20]

The lower the absolute value of the weights, the lower the degree of leverage of the position and the lower the expected returns one can obtain. However, the risk of the optimal portfolio decreases commensurately, the extreme case being given by a unit position on R_s but zero positions on all the risky assets, which eliminates all risks. More generally, if $\mathbf{P}$ denotes the correlation matrix and $\mathbf{s} = (\mu_1/\sigma_1, \ldots, \mu_N/\sigma_N)$ the vector of Sharpe ratios for the underlying assets, the aggregate risk-return trade off in the presence of a safe asset will be given by $\sqrt{\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}} = \sqrt{\mathbf{s}'\mathbf{P}^{-1}\mathbf{s}}$, which is the highest Sharpe ratio any investor can achieve. This quantity corresponds to the slope of the mean - variance frontier in mean - standard deviation space.

Nevertheless, a successful practical implementation of mean variance analysis is far from trivial, as it requires assigning reasonably accurate values to the vector of risk premia $\boldsymbol{\mu}$ and the covariance matrix $\mathbf{\Sigma}$, whose elements can change over time.[21] Given the focus of this paper, I will concentrate on $\mathbf{\Sigma}$ henceforth.

2.2 Correlation measures

As I have just explained, the covariance matrix plays a fundamental role in the theory of optimal portfolio allocation. However, the volatility measures discussed in section 1.2

[20] In the case of two asset returns, the optimal weights are proportional to:

$$\{E(r_1) - [cov(r_1,r_2)/V(r_2)]E(r_2)\}/[V(r_1) - cov^2(r_1,r_2)/V(r_2)]$$
$$\{E(r_2) - [cov(r_1,r_2)/V(r_1)]E(r_1)\}/[V(r_2) - cov^2(r_1,r_2)/V(r_1)]$$

The numerators of those expressions are the intercepts in the linear least squares projection of one excess return on a constant and the other one. In turn, the denominators are the mean square errors of those projections. $\{E(r_1) - [cov(r_1,r_2)/V(r_2)]E(r_2)\}/\sqrt{V(r_1) - cov^2(r_1,r_2)/V(r_2)}$ is called the "information ratio" of the first asset relative to the second one, which coincides with its Sharpe ratio if and only if $cov(r_1,r_2) = 0$.

[21] For that reason, simpler investment rules such as (i) equally weighted portfolios, which is mean - variance optimal when all risk premia in $\boldsymbol{\mu}$ are identical and $\mathbf{\Sigma}$ has the equicorrelated structure I discuss in the next section, or (ii) risk parity portfolios, which in turn are mean - variance optimal when all Sharpe ratios μ_i/σ_i are identical and $\mathbf{\Sigma}$ is diagonal but not scalar, have become very popular among both retail and institutional investors.

only contain information about its diagonal. Therefore, it is necessary to obtain its off-diagonal elements too. Next, I briefly discuss the generalisation of the four methods in section 1.2 to the case of multiple assets.

1. Sample covariance matrices over rolling windows The simplest estimate is the so-called historical covariance matrix

$$\mathbf{his}_{t,S} = \frac{1}{S}\sum_{s=1}^{S} \mathbf{r}_{t-s}\mathbf{r}'_{t-s}, \tag{5}$$

which is the average of the squares and cross-products of the daily returns computed over a rolling window of the previous S trading days.[22] The criticisms I made in the univariate case apply here too.

Similarly, RiskMetrics EWMA version uses

$$\mathbf{RM}_t = (1-\lambda)\sum_{s=1}^{\infty} \lambda^s \mathbf{r}_{t-s}\mathbf{r}'_{t-s} = (1-\lambda)\mathbf{r}_{t-1}\mathbf{r}'_{t-1} + \lambda \mathbf{RM}_{t-1}, \tag{6}$$

where λ is the common decay parameter.

Both these measures have two important practical advantages: they contain the corresponding (square) volatility estimate for each asset along their diagonal, and they give rise to non-negative variances for any portfolio one could think of.

2. Econometric models Multivariate versions of ARCH models are conceptually straightforward. In the GARCH(1,1) case, we would have

$$vech(\mathbf{\Sigma}_t) = vech(\mathbf{\Theta}) + \mathbf{A}vech(\mathbf{r}_{t-1}\mathbf{r}'_{t-1}) + \mathbf{B}vech(\mathbf{\Sigma}_{t-1}),$$

where $\mathbf{\Theta}$ is an $N \times N$ matrix of intercepts, $\mathbf{A}$ and $\mathbf{B}$ are $N(N+1)/2 \times N(N+1)/2$ matrices of coefficients, and $vech()$ is an operator which re-arranges the $N(N+1)/2$ different elements of the conditional covariance matrix $\mathbf{\Sigma}_t$ by stacking its columns on top of each other in vector form and getting rid of the duplications [see Bauwens, Laurent and Rombouts (2006) for a survey].

However, these models have been far less successful than their univariate counterparts because of two main reasons. First, the unrestricted version involves a large number of parameters in $\mathbf{A}$ and $\mathbf{B}$, which increases with the fourth power of N, an extreme version

[22] Again, some slight variants of this expression subtract the mean excess returns before computing squares and cross-products, while others use geometric returns instead of arithmetic ones.

of the so-called curse of dimensionality. Second, except in some special cases, it is incredibly difficult to derive conditions on the coefficients that guarantee that they will give rise to non-negative variance for every conceivable portfolio.[23]

Some restricted versions have been popular, though. The simplest multivariate GARCH(1,1) model is the so-called scalar version [see Ding and Engle (2001)], in which

$$\mathbf{\Sigma}_t = \mathbf{\Theta} + \alpha \mathbf{r}_{t-1}\mathbf{r}'_{t-1} + \beta \mathbf{\Sigma}_{t-1}.$$

In fact, the RiskMetrics model (6) can be regarded as a special case of this structure in which $\beta = 1 - \alpha$ and $\mathbf{\Theta} = \mathbf{0}$.

Other versions combine several univariate GARCH models with a simplified correlation structure. The simplest possible example is the conditional correlation model of Bollerslev (1987), which allows for time-varying volatilities but forces correlations to be constant. Given its lack of realism, this model has been largely superseded by the dynamic conditional correlation (DCC) models of Engle (2002) and Tse and Tsuy (2002), which impose an approximately scalar GARCH(1,1) structure on the conditionally standardised returns r_{it}/σ_{it} $(i = 1, \ldots, N)$ [see Aielli (2013) for details].

Another popular version is the conditionally heteroskedastic analogue to the common factor model that has been used to represent large covariance structures for stock returns for decades, and which inspired Ross' (1976) Arbitrage Pricing Theory [see Connor, Goldberg and Korajczik (2010)]. In the single factor case put forward by Diebold and Nerlove (1989),

$$\mathbf{r}_t = \boldsymbol{\mu} + \mathbf{b} f_t + \mathbf{u}_t, \tag{7}$$

where f_t is a common risk factor which affects all asset returns simultaneously, $\mathbf{u}_t = (u_{1t}, \ldots, u_{Nt})$ are N conditionally uncorrelated idiosyncratic factors which affect one asset only and $\mathbf{b}$ are the coefficients which measure the sensitivity of the assets to the common factor. Assuming that f_t and $\mathbf{u}_t$ have conditional variances given by λ_t and $\mathbf{\Gamma}_t = diag(\gamma_{1t}, \ldots, \gamma_{Nt})$, respectively, the conditional covariance matrix will be given by

$$\mathbf{\Sigma}_t = \mathbf{b}\mathbf{b}'\lambda_t + \mathbf{\Gamma}_t.$$

The diagonal elements of this matrix, which represent the conditional variances of

[23] In He, Sentana and Teräsvirta (2008), we made some progress by writting multivariate ARCH models as vector autoregressions (VARs) with random coefficients.

each of the returns, are given by

$$\sigma_{it}^2 = b_i^2 \lambda_t + \gamma_{it},$$

while the conditional covariances in the off-diagonal elements will be

$$\sigma_{ijt} = b_i b_j \lambda_t.$$

As a result, the conditional correlation between the returns of assets i and j will be

$$\rho_{ijt} = \frac{b_i b_j \lambda_t}{\sqrt{b_i^2 \lambda_t + \gamma_{it}}\sqrt{b_j^2 \lambda_t + \gamma_{jt}}}.$$

Assuming that the factor loadings b_i and b_j are both positive, King, Sentana and Wadhwani (1994) proved that $\partial \rho_{ijt}/\partial \lambda_t > 0$ and $\partial \rho_{ijt}/\partial \gamma_{it} < 0$, which implies that in this model correlations will increase (decrease) when, ceteris paribus, the volatility of the common (specific) risk component increases. I will revisit this issue in section 2.3.

A special case of this model is one in which $\mathbf{b}$ is (proportional to) a vector of N ones and $\mathbf{\Gamma}_t$ is scalar. Then, all pairwise correlations will be given by

$$\frac{\lambda_t}{\lambda_t + \gamma_t}.$$

This model is called an equicorrelated model, and has been extensively used in many other contexts [see e.g. Vasicek's (1987) loan portfolio model]. Engle and Kelly (2012) proposed a generalisation in which the equicorrelated structure is effectively applied to conditionally standardised returns.

Multivariate versions of stochastic volatility models have evolved along similar lines [see Chib, Nardari and Shephard (2006)], but their use remains limited.[24]

3. Realised covariance matrices using intraday data Again, it is possible to use ultra high frequency intraday observations to compute daily measures of covariances and correlations. Guaranteeing positivity of the portfolio variances for every portfolio is also tricky but several methods have been proposed that achieve this [see e.g. Barndorff-Nielsen et al (2011)]. However, in addition to the issues highlighted in section 1.2, the

[24] In fact, if the conditional variances of the common and idiosyncratic latent factors in (7) follow univariate Garch processes, then this model effectively becomes a stochastic volatility model, in the sense that there are more shocks than observed returns. Fiorentini, Sentana and Shephard (2004) and Sentana, Calzolari and Fiorentini (2008) propose simulation methods for the estimation of the model parameters in that case, which they apply to moderately large cross-sections of stock returns.

main practical difficulty here is dealing with return pairs in which one of the assets trades far less frequently than the other. In addition, in some cases it is simply impossible to compute a realised covariance, the obvious example being the US and Japanese stock market indices, whose trading hours do not overlap.

4. Implied correlations obtained from financial derivatives The triangular equality of exchange rates implies that, absent arbitrage opportunities, the (log) capital gain obtained by trading the US dollar - Japanese Yen cross rate has to be the difference between the (log) capital gain obtained by trading the US dollar - euro and Japanese Yen - euro rates. As a result, in a Black-Scholes world, the implied square vol of the dollar - Yen rate should be equal to the implied square vol of the dollar - euro rate plus the implied square vol of the dollar - Yen rate minus twice the implied covariance between the last two. Therefore, if there are actively traded options for all three currencies, one can easily obtain a forward-looking, market-based measure of correlation, known as the implied correlation [see e.g. Campa and Chang (1998)].

More generally, one could analogously obtain the implied correlation between two assets by combining the implied vols of each of them with the implied vol of any portfolio based on them, regardless of the portfolio weights. Unfortunately, there are very few "basket" options, that is, options written on portfolios, and they are frequently over the counter products bespoke to satisfy the specific needs of some large investors. The main exception are options written on stock market indices, which are incredibly popular. The complication there is that they usually involve many assets, not just one pair, so without further assumptions it is impossible to obtain the $N(N-1)/2$ correlations involved in an index that combines N risky assets. In the case of the S&P 500 index, the number of potentially different correlations is 124,750!

There are at least two practical approaches to deal with this problem. The first one is to assume an equicorrelated structure, as in the previous section. This is the methodology of the CBOE S&P 500 Implied Correlation Index [see CBOE (2009b)]. Specifically, the τ-maturity implied correlation index can be obtained as

$$\rho_{Index,\tau}(t) = \frac{\sigma^2_{Index,\tau}(t) - \sum_{i=1}^{N} w_i^2 \sigma^2_{i,\tau}(t)}{2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} w_i w_j \sqrt{\sigma^2_{i,\tau}(t)\sigma^2_{j,\tau}(t)}}, \tag{8}$$

where w_i $(i = 1, \ldots, N)$ are the index weights.

In recent years, though, the numerator of (8), known as the overdispersion, which measures the difference between the actual variance of the index and its hypothetical value if the returns to all its constituents were mutually uncorrelated, has gained increased attention among both practitioners and researchers [see e.g. Jones and Vischer (2015)].[25] Unlike $\rho_{Index,\tau}(t)$, this measure makes no restrictive assumptions on the correlation structure of the individual constituents of an index.

2.3 Stylised facts

Two stylised facts related to time-varying correlations are that (i) there are certain periods when markets seem to move persistently in unison and others when the correlation between them appears to be systematically low; and (ii) periods when markets are increasingly correlated are also times when markets are volatile [see King and Wadhwani (1990) and Roll (1989) for some early evidence]. Indeed, King and Wadhwani (1990) argued that this might be because a rise in volatility might lead agents to pay greater attention to other markets. As I mentioned in the previous section, conditionally heteroskedastic factor models in general, and equicorrelated structures in particular, account for these observations. For example, King, Sentana and Wadhwani (1994) confirmed both the time-varying nature of correlations and the fact that periods of high volatility in the common global risk factors are also periods of increased correlation across national stock markets, which in turn substantially reduces the gains from international diversification.

Figure 4a compares the CBOE implied vols of three major US stock indices: S&P500, Nasdaq 100 and Russell 2000. The S&P 500 index contains the 500 largest companies in terms of capitalisation whose shares trade on U.S. exchanges. In contrast, the Russell 2000 index contains the 2000 smallest companies in the Russell 3000 index, which in turn contains the 3000 largest companies whose stocks trade on U.S. exchanges. Finally, the Nasdaq 100 is made up of the largest non-financial stocks on the Nasdaq, which is heavily specialised in technology stocks. Therefore, while there should be no overlap between S&P500 and Russell 2000, there is an increasing degree of overlap between S&P500 and Nasdaq 100 because of the current prominence of Apple, Amazon, Facebook, Microsoft

[25] Although in theory the overdispersion should always be non-negative and the implied correlation below 1, the indirect way in which one computes these quantities might occasionally lead to violations of those bounds. For example, the implied correlation for the S&P500 officially reported by the CBOE exceeds 1 occassionally.

and Alphabet. Not surprisingly, the correlations between the geometric returns on those indices are very high:

return correlations	Nasdaq 100	Russell 2000
S&P 500	.9216	.9177
Nasdaq 100		.8698

However, the correlations between their implied vols are even higher:

vol correlations	Nasdaq 100	Russell 2000
S&P 500	.9789	.9805
Nasdaq 100		.9418

In turn, Figure 4b compares the VIX with the V2X, which is the analogue volatility index for the Euro Stoxx 50, a market-based capitalisation index of the largest and most liquid publicly traded companies in the eurozone. Once again, this picture confirms the strong comovements in international stock markets. In this case, however, the difference between the correlation of those two volatilities (.942) and the correlation of the underlying (geometric) returns (.615) is even more striking.

More generally, Bollerslev et al (2018) show that the high degree of correlation across volatilities that I have just documented happens not only within an asset class but also across asset classes.

Turning now to the correlations between returns, Amengual and Sentana (2017a) present a simple decomposition of the implied vol of the IBEX 35 index into two components: the implied vol that this index would have if all the pairwise correlations between its constituents were 0, which we denote by "constituents", and the rest, denoted by "dispersion", which is exclusively due to the off-diagonal elements of the conditional covariance matrix. Their figure 8a, which I reproduce here as Figure 5a, confirms that both the volatility of the individual constituents and their dependence rises in crisis periods. Nevertheless, the importance of both components is roughly similar, except in the last part of the sample, when a large fraction in the reduction of the implied vol of the index seems to be due to the reduction of the implied vols of its constituents. One possible explanation for this phenomenon is that the largest Spanish companies are internationally diversified, with approximately two third of their revenues from foreign markets, so that their income sources are more affected by sectoral factors than country risks.

Amengual and Sentana (2017a) also compute the implied correlation of the IBEX 35, which I reproduce in Figure 5b. The median value of this implied correlation is 67%, and although it positively correlated with the implied vol of the IBEX, the relationship is not too strong. Interestingly, the persistence of the IBEX implied correlation series, as measured by its correlogram, is lower than the persistence of its implied vol series, which might be due to the cross-sectional averaging involved in the implied correlation calculation.

Correlations across asset classes, though, are far more difficult to capture. The paradigmatic example is the correlation between US stocks and bonds, which changes sign from time to time [see e.g. Campbell, Pflueger and Viceira (2015)].

2.4 Value at risk

The 1996 Amendment to the first Basel Capital Adequacy Accord forced banks and other financial institutions to develop models to quantify their market risks accurately [see Basel Committee on Banking Supervision (1997)]. In practice, most institutions chose the so-called Value at Risk (VaR) framework popularised by JP Morgan in 1994 in order to determine the capital necessary to cover their exposure to those risks. Nowadays, retail investors routinely receive information on this statistic when they buy a mutual fund, although sometimes its meaning is not properly explained. As I will show in the rest of this section, VaR is another concept intimately related to volatility and correlation.

To understand it better, consider again the portfolio selection problem of an investor with an initial wealth W that can be invested in N risky assets with gross returns $\mathbf{R}$ with means $\boldsymbol{\nu}$ and covariance matrix $\boldsymbol{\Sigma}$, and a safe asset with gross return R_s. Let $\mathbf{w}$ denote the vector that represents the proportion of his wealth invested in the risky assets, so that $w_0 = 1 - \mathbf{w}'\boldsymbol{\ell}_N$ is the fraction invested in the riskless asset, with $\boldsymbol{\ell}_N$ denoting a vector of N ones.

The random wealth of the investor at date 1 will be $W(R_s + \mathbf{w}'\mathbf{r})$, so she will bear losses when $W(R_s + \mathbf{w}'\mathbf{r}) < W$, i.e., when the gross return of her portfolio, $R_s + \mathbf{w}'\mathbf{r}$, is less than 1. In this context, for a given a confidence level $1 - \alpha$ (say 99%), with $0 < \alpha < 1/2$, the *Value at Risk* (*VaR*) associated to portfolio $\mathbf{w}$ is the $(1-\alpha)^{th}$ quantile of the probability distribution of losses, i.e., the critical value $VaR(\mathbf{w})$ such that

$$P\left[1 - (R_s + \mathbf{w}'\mathbf{r}) > VaR(\mathbf{w})/W\right] = \alpha.$$

For convenience, though, the portfolio VaR is often reported in fractional form as

$-VaR(\mathbf{w})/W$. Hence,

$$VaR(\mathbf{w})/W = -F^{-1}(\alpha;\mathbf{w})\sigma(\mathbf{w}) - R_s + 1 - \mu(\mathbf{w}),$$

where $\mu(\mathbf{w}) = \mathbf{w}'\boldsymbol{\mu}$, $\sigma^2(\mathbf{w}) = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$, $F(\cdot;\mathbf{w})$ denotes the cdf of the portfolio excess returns $\mathbf{w}'\mathbf{r}$ after having been standardised and $F^{-1}(\alpha;\mathbf{w})$ is its inverse (i.e. the quantile function).[26] It is important to emphasise that $VaR(\mathbf{w})/W$ is a quantile (lower bound) on the losses that may materialise. If the model used to compute it is correct, we know that, on average, $100\alpha\%$ of the time the losses will be at least as large. For that reason, expected shortfall, which estimates the expected losses beyond the VaR threshold, has become increasingly popular [see Artzner et al (1999)].

There are many ways of computing the VaR of a portfolio. A rather popular procedure employed by many financial institutions all over the world is the so-called historical method, which relies on the empirical quantiles of the returns to the current portfolio over the last S observations. As a result, it suffers from the same problem as historical volatilities, although it does not make any assumption about the distribution of returns.

On the other hand, the original RiskMetrics methodology ignores the mean component $\mu(\mathbf{w})$ on the grounds that expected overnight returns are usually tiny relative to their volatility [but see Meddahi and Yamashita (2017)], and uses the covariance matrix in (6) to compute the variance of the current portfolio as $\mathbf{w}'\mathbf{RM}_t\mathbf{w}$. A second generation version of RiskMetrics offered the choice of a Gaussian distribution or a Student t distribution with 5 degrees of freedom to compute the quantile [see Zumbach (2007)], while the latest one allows for normal-gamma mixtures. In that regard, it is often argued that the fact that the distribution of returns has fatter tails than the normal invalidates the Gaussian VaR. While this is strictly speaking true, the effects of misspecifying the distribution, which depend on a complex manner on the interaction between confidence level and true distribution, are not necessarily important [see Amengual, Fiorentini and Sentana (2013) for further discussion]. For example, when $\alpha = .025$, it is well known that the Gaussian critical value is $\Phi^{-1}(\alpha) = 1.9599$. What is far less known, however, is

[26] In Sentana (2003), I proved that the combinations of portfolio means and standard deviations which give rise to the same (proportional) VaR lie on the positively sloped straight line in (μ, σ) space, $\mu = (1 - R_s - VaR/W) - F^{-1}(\alpha)\sigma$, which I call an "IsoVaR", provided that $F(\alpha;\mathbf{w})$ does not in fact depend on the portfolio weights $\mathbf{w}$. This result does not rely on normality, and it will hold as long as the distribution of $\mathbf{w}'\mathbf{r}$ depends on $\mathbf{w}$ only through its mean and variance. In particular, it will hold under for elliptical distributions too. The superposition of these *IsoVaRs* on the mean - variance frontier immediately determines the solution to a mean - variance portfolio problem with a VaR constraint [see Sentana (2003) for further details].

that 1.9599 is also the 97.5% quantile of a standardised Student t distribution with 3.222 degrees of freedom despite the fact that this distribution has unbounded fourth moments. Intuitively, the reason is the fact that the area under the curve of any valid density is necessarily one because that is the probability of observing some return. Fatter tails and a higher peak with more mass near the mean return implies that the Student t density must be leaner in between (see again Figure 1). As a result, the cumulative distribution functions of standardised version of the normal and Student t random variables cross not only at 0, but also near each of the extremes.

3 Contagion

3.1 Alternative dependence measures

The analysis in section 2 might give the impression that diversification is a panacea. Unfortunately, financial markets that show little correlation in normal periods show a tendency to fall together during crisis periods. For instance, when Russia defaulted in August 1998, most stock markets around the world fell at unison. This had really terrible consequences for those investors who thought that they could eliminate most of the risk of their portfolios by holding assets that looked seemingly uncorrelated.[27]

The blame falls partly on Pearson correlation coefficients, which only provide a complete description of dependence under multivariate normality, in which case the conditional mean of one variable given others is a linear function of the latter and the corresponding conditional variance constant. Attractive alternatives are provided by Spearman rank correlation coefficients, which are formally defined as Pearson correlation coefficients of the marginal probability integral transforms (PITs) of each of the returns. These coefficient are computed as follows. First, one instantaneously transforms each of the elements of $\mathbf{R}$ into N uniform random variables $u_i = G_{1i}(R_i)$ $(i = 1, \ldots, N)$, where $G_{1i}(.)$ is the marginal cumulative distribution function of R_i. Although this procedure may sound a little daunting, in any given sample this initial step simply replaces the original returns on each asset by their observed ranks divided by the number of observations S $(+1)$, so that the smallest sample value becomes $1/(S+1)$ and the largest $S/(S+1)$. Then, one computes the usual correlation coefficient between u_i and u_j. Thus,

[27] The 1998 Russian default also created havoc for seemingly riskless portfolios that exploited almost perfect arbitrage opportunities (convergence trades) along the yield curve, leading to the collapse of the hedge fund Long Term Capital Management, in which Merton and Scholes were heavily involved.

the resulting coefficient will also be between -1 and 1 by construction. Interestingly, a Spearman rank correlation of 1 implies that there must be a monotonically increasing, deterministic relationship between the two variables, while a value of -1 implies a monotonically decreasing one. Given that Pearson correlation coefficients only attain those extreme values when the deterministic relationship is in fact linear, these rank correlation coefficients can capture non-linear relationships between returns that might go undetected with the usual procedures.

However, the correlation between ranks is difficult to visualise in a two dimensional scatter plot. In that regard, a third possibility is the so-called Gaussian rank correlation, which is Pearson correlation coefficient of the Gaussian ranks, defined as the Gaussian quantiles corresponding to the uniform ranks u_i. More formally, the Gaussian rank of the s^{th} observation on the gross return on the i^{th} asset is defined as $\Phi^{-1}[G_{1i}(R_{is})]$, where $\Phi^{-1}(.)$ denotes the quantile function of a standard normal variable, which is defined as the inverse of its cdf $\Phi(.)$ so that $\Phi[\Phi^{-1}(u_i)] = u_i$. These ranks have standard normal marginal distributions by construction, which makes the interpretation of their scatter plots easier [see Amengual and Sentana (2018)].

Nevertheless, both Spearman and Gaussian rank correlations continue to look at the entire distribution, while investors might worry more about certain regions. For that reason, Longin and Solnik (2000) proposed a more targeted measure of "tail dependence", which they called exceedance correlation. This is simply a sequence of Pearson correlations defined over nested sections of the third quadrant, in which both returns are negative, and the first quadrant, when the returns are simultaneously positive. More formally,

$$\rho_{ij}(\kappa) = \begin{cases} cor(r_i, r_j | r_i > \kappa, r_j > \kappa) & \text{if } \kappa > 0 \\ cor(r_i, r_j | r_i < \kappa, r_j < \kappa) & \text{if } \kappa < 0 \end{cases}$$

Figure 6 presents the pattern of exceedance correlation for the three bivariate distributions in Figure 3, whose marginal elements have zero mean and unit variance and whose Pearson correlation coefficients are 0 by construction. As expected, in the Gaussian distribution $\rho_{ij}(\kappa) = 0$ for all κ because lack of correlation implies independence. On the other hand, there is tail dependence in the symmetric Student t distribution, but it is the same in the first and third quadrants because of its spherical nature. In contrast, the asymmetric Student t distribution shows much higher exceedance correlation in the third quadrant than in the first one.

Therefore, negative tail dependence is a phenomenon that the multivariate normal distribution cannot account for, and elliptical distributions struggle with. For that reason, it is convenient to consider more flexible multivariate distributions. The two main problems with non-elliptical distributions is that the number of parameters may increase very rapidly with the number of assets, and that it is not always easy to find the marginal distributions that they imply for the returns of portfolios constructed from the original risky assets. In this second regard, the generalised hyperbolic family, finite mixtures of multivariate Gaussian distributions or Hermite expansions are examples of joint distributions whose marginal components belong to the same class [see Mencía and Sentana (2009, 2012), as well as Amengual and Sentana (2015)].

An alternative way of modelling dependence is through copulas. Formally, copulas are joint distribution functions with uniform marginals defined over the unit hypercube in $\mathbb{R}^N$. Intuitively, the idea is very simple. After transforming again the returns on each asset into their ranks by means of the corresponding marginal probability integral transform, one models the dependence of the random vector of uniform ranks $\mathbf{u} = (u_1, \ldots, u_K)'$ through a joint density function, known as the copula density.

Nowadays copulas are extensively used in many economic and finance applications. Although there are many well known examples of bivariate copulas, some of them are popular simply because they are mathematically convenient, as opposed to being motivated by empirical observations on real life phenomena. More importantly, they are difficult to generalize to multiple dimensions. On the other hand, the Gaussian copula is a popular choice both in bivariate and multivariate contexts since it is easily scalable. Unfortunately, it rules out non-linear dependence, particularly in the lower tail. For that reason, the validity of this copula in finance has been the subject of considerable public debate, to the extent that the media declared it "the formula that felled Wall Street" [see the provocative article by Salmon (2009), the more nuanced analysis by MacKenzie and Spears (2012), and the academic response by Donnelly and Embrechts (2010)]. In Amengual and Sentana (2017b), we develop specification tests for a Gaussian copula against the copulas associated to some of the flexible generalisations of the multivariate normal distribution I mentioned above.

3.2 Financial contagion

Financial crises are not a modern phenomenon, but their incidence, and the extent to which they have gone from being a domestic problem to a regional one, and more recently, a truly global concern, is definitely a sign of our times. This largely undesired globalisation of financial crises is intimately related to the concept of contagion. By analogy to the epidemiological concept of communicable disease, financial contagion literally means that the economic and financial difficulties in a country get quickly transmitted to other more or less closely linked economies. The 1982 Latin America debt crisis, the 1994 Mexican peso "tequila" crisis, the 1997 Asian financial crisis and the following year's Russian one, the 2007-2008 global financial crisis and the 2010-2012 European sovereign debt crisis are obvious examples of this phenomenon. Although several authors have developed detailed narratives for each of them [see e.g. Baldwin et al (2015) for the eurozone], all those crises share an important characteristic: they typically start in a single country but they rapidly spread to others, even though the direct economic and financial links between the original country and some of the affected ones are fairly weak. For that reason, financial contagion at such a grand scale has been a fundamental driving force behind financial regulation, both at the domestic level and at the international one.

Nevertheless, the academic definition of contagion is not necessarily the same as the informal definition used by the media and the national and international financial regulators. Two confounding facts that I mentioned before are that (i) there are certain periods when markets seem to move in unison and others when the correlation between them appears to be low; and (ii) periods when markets are increasingly correlated are also times when markets are volatile. But these correlation increases cannot be really called contagion. For that reason, most previous approaches to identify real contagion episodes have started from a factor model for returns of the type that I discussed in section 2.2 [see Dungey et al (2005) for a survey, with special emphasis on methodological aspects]. According to Forbes and Rigobon (2002), contagion would then arise if and only if there is a "significant increase in cross-market linkages after a shock to an individual country or group of countries".

The empirical finance literature on contagion has focus almost exclusively on stock returns. Nevertheless, stock returns on their own are not necessarily informative enough about tail dependence between financial markets, the reason being that by definition such simultaneous events only occur occasionally. As recently illustrated by Andersen, Fusari and Todorov (2016, 2017), financial derivatives in general, and out-of-the money put options, in particular, contain a lot of information about the left tail. For that reason, in Amengual and Sentana (2017a) we combine data on several national stock market indices with options written on them. In that regard, our approach is related to the literature on risky debt, both corporate and sovereign, which has successfully combined data on bond yields with data on credit default swaps (CDSs) to increase our knowledge of default risk and loss given default, and the common movements in credit ratings observed in practice [see e.g. Pan and Singleton (2008) or Duffie et al (2009)].

3.3 Volatility, correlation and dependence during the recent financial crises

The euro, formally launched on January 1st, 1999, brought about a period of low interest rates that coincided with a reduction in inflation levels and an increase in growth rates across the world. However, important imbalances began to build up very soon. The so-called peripheral countries, or GIIPS (Greece, Ireland, Italy, Portugal and Spain) accumulated important current account deficits, which were effectively financed by the core countries (Austria, Belgium, Finland, France, Germany, Luxembourg and the Netherlands). More importantly, they suffered a progressive deterioration of competitiveness due to their inflation differentials.

Still, Spain had a modest ratio of public debt to GDP of 36% at the outset of the global financial crisis. In addition, the initial effects of the European sovereign debt crisis on the country were arguably triggered by events in Greece and elsewhere. As it is usually the case in episodes of financial contagion, though, Spain had some serious economic problems of its own. Its seemingly good fiscal position was largely due to the huge construction-related tax revenues generated during the housing bubble, which allowed for large increases in local, regional and federal government spending in public consumption and infrastructure (not necessarily guided by economic efficiency criteria) to remain blurred in the aggregate figures. While NINJA (no income, no jobs, no assets) mortgages did not exist as such, and regular residential mortgages were a relative minor problem for banks thanks to their recoursing character and the safety net provided by families, competition for market share earlier on, implemented through generous revolving credit facilities to property developers, quickly increased the volume of “zombie” loans on the books of some financial institutions, which were betting on resurrection hoping for “green shoots”. The socialist government first, and the conservative one that was elected at the end of 2011, implemented three tepid and arguably misguided financial reforms, but to no avail. On June 1st, 2012, just three weeks after nationalising Bankia, a bank resulting from the merger two years earlier of seven regional savings banks in difficulties, the Spanish 10-year bond reached 548 basis points above its German counterpart on intraday trading. The Spanish government eventually agreed to a very large European financial rescue package for its banks, but in the second half of July the spread on German bunds reached 6%, and the irreversibility of EMU was in doubt. The famous “whatever it takes” speech by the president of the ECB backed by the German chancellor, and a far more aggressive reform of the Spanish financial system as a result of the Memorandum of Understanding signed with the EU, which included the creation of a bad bank, reduced those spreads by 140 basis points in only five days. By mid October, interest rate differentials were a mere 2.4%, and the Spanish Treasury was able to issue short term debt at less than 1%.

Figures 5a and 5b clearly indicate that the events in Greece, Ireland and Portugal substantially affected the implied volatility of the Spanish stock market. To assess the extent to which the channel was the negative feedback loop between banks and weak sovereigns, in Amengual and Sentana (2017a) we compare the implied vols of the two largest Spanish banks (Santander and BBVA) with the implied vols of its two largest utility companies (Iberdrola and Gas Natural). Figures 9a and 9b in that paper, which I reproduce here as Figures 7a and b, show that banks had larger implied vols on average than utilities, which to some extent reflects the procyclical nature of their revenue base but also the leveraged nature of their capital structure. More importantly, those figures show that the banks reacted more strongly to the European sovereign debt crisis than the implied vols of the utilities, which might be interpreted as evidence in favour of the diabolic loop. Somewhat surprisingly, while both Santander and BBVA were affected by the turmoil in financial markets after the Lehmann Brothers collapse, the effects of the European sovereign debt crisis on them was relatively mild in the spring of 2010, when the first Greek aid package was been discussed, and especially in the second half of 2011, a period in which the details of the private debt re-structuring deal for Greece were being discussed. Nevertheless, the dramatic increase of all four series in the spring of 2012, which is precisely when the yields on Spanish public debt rocketed, reflects the extent to which investors were concerned about the prospects of the Spanish economy before the government decided to negotiate with the European institutions an assistance package for the financial sector.

4 Conclusion

In this paper I describe in detail the concepts of volatility, diversification and contagion, three basic keys to understand the seemingly whimsical behaviour of financial markets. The presentation is deliberately not very technical and largely self-contained, with most required concepts defined along the way. Nevertheless, the analysis is mostly empirically oriented, with an emphasis on the methods that have been proposed to measure those concepts and a discussion of the stylised facts that the resulting measures imply. I also use those measures to study the effects of the financial crisis of 2007-2008 and the euro sovereign debt crisis of 2010-2012 on Spain.

The three concepts that I discuss illustrate the interest of financial markets participants in having at their disposal accurate and timely measures of risk on which to base their decisions. At the same time, they also raise many research possibilities for academics. Life is indeed uncertain, but the interest in volatility, diversification and contagion is bound to increase over time.

References

Aielli, G.P. (2013): "Dynamic conditional correlation: on properties and estimation", *Journal of Business and Economic Statistics* 31, pp. 282-299.

Amengual, D., Fiorentini, G. and Sentana, E. (2013): "Sequential estimators of shape parameters in multivariate dynamic models", *Journal of Econometrics* 177, pp. 233-249.

Amengual, D. and Sentana, E. (2010): "A comparison of mean-variance efficiency tests", *Journal of Econometrics* 154, pp. 16-34.

Amengual, D. and Sentana, E. (2015): "Is a normal copula the right copula?", CEMFI Working Paper 1504.

Amengual, D. and Sentana, E. (2017a): "Financial contagion in the eurozone", mimeo, CEMFI.

Amengual, D. and Sentana, E. (2017b): "Is a normal copula the right copula?", mimeo, CEMFI.

Amengual, D. and Sentana, E. (2018): "Gaussian rank correlation: a coefficient with attractive statistical properties", mimeo, CEMFI.

Andersen, T. G., Bollerslev, T., Diebold, F. X. (2009): "Parametric and nonparametric volatility measurement", in: Aït-Sahalia, Y. and Hansen, L.P. (eds.), *Handbook of Financial Econometrics* vol 1, pp. 67-138, North-Holland Press.

Andersen, T.G. and Bondarenko, O. (2007): "Construction and interpretation of model-free implied volatility", in I. Nelken (ed.), *Volatility as an asset class*, pp. 141-181, Risk Publications.

Andersen, T.G., Fusari, N. and Todorov, V. (2016): "Pricing short-term market risk: evidence from weekly options", *Journal of Finance*, forthcoming.

Andersen, T.G., Fusari, N. and Todorov, V. (2017): "The pricing of tail risk and the equity premium: evidence from international option markets", mimeo, Kellogg School of Business, Northwestern University.

Andersen, T.G. and Shephard, N. (2009): "Stochastic volatility: origins and overview", in T.G. Andersen, R.A. Davis, J.P. Kreiss and T. Mikosch (eds.), *Handbook of Financial Time Series*, pp. 233-254, Springer.

Artzner, P., Delbaen, F., Eber, J. M. and Heath, D. (1999): "Coherent measures of risk", *Mathematical Finance* 9, pp. 203-228.

Baldwin, R., Beck, T., Bénassy-Quéré, A., Blanchard, O., Corsetti, G-C., de Grauwe, P., den Haan, W., Giavazzi, F., Gros, D., Kalemli-Ozcan, S., Micossi, S., Papaioannou, E., Pesenti, P., Pissarides, C., Tabellini, G. and Weder di Mauro, B. (2015): "Rebooting the eurozone: step 1 – agreeing a crisis narrative", http://voxeu.org/article/ez-crisis-consensus-narrative.

Bakshi, G. S., Kapadia, N. and Madan, D.B. (2003): "Stock returns characteristics, skew laws, and the differential pricing of individual equity options", *Review of Financial Studies* 16, pp. 101–143.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008): "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise", *Econometrica* 76, pp. 1481-1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2011): "Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading", *Journal of Econometrics* 162, pp. 149-169.

Barndorff-Nielsen, O.E. and Shephard, N. (2007): "Variation, jumps and high frequency data in financial econometrics", in R. Blundell, T. Persson and W.K. Newey (eds.) *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, pp. 328-372, Cambridge University Press.

BARRA (1998): "United States equity risk model handbook", version 3, BARRA Inc.

Basel Committee on Banking Supervision (1997): "Amendment to the Capital Accord to incorporate market risk", mimeo, Bank of International Settlements.

Bauwens, L, Laurent, S. and Rombouts, J. (2006): "Multivariate GARCH models: a survey", *Journal of Applied Econometrics* 21, pp. 79-109.

Blair, B.J., Poon, S.-H. and Taylor, S.J. (2001): "Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns", *Journal of Econometrics* 105, pp. 5-26.

Berk, J. (1997): "Necessary conditions for the CAPM", *Journal of Economic Theory* 73, pp. 245-257.

Bollerslev, T. (1986): "Generalized ARCH", *Journal of Econometrics* 31, pp. 307-327.

Bollerslev, T. (1990): "Modelling the coherence in short-run nominal exchange rates: a multivariate Generalized ARCH model", *Review of Economics and Statistics* 72, pp.

498-505.

Bollerslev, T., Hood, B., Huss, J. and Pedersen, L.H. (2018): "Risk everywhere: modeling and managing volatility", CEPR Discussion Paper 12687.

Calzolari, G., Fiorentini, G. and Sentana , E. (2004): "Constrained indirect estimation", *Review of Economic Studies* 71, pp. 945-973.

Campa, J.M. and Chang, P.H.K. (1998): "The forecasting ability of correlations implied in foreign exchange options", *Journal of International Money and Finance* 17, pp. 855-880.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997): *The econometrics of financial markets,* Princeton University Press.

Campbell, J.Y., Pflueger, C. and Viceira, L.M. (2015): "Monetary policy drivers of bond and equity risks", mimeo, Harvard Business School.

Chib, S., Nardari, F. and Shephard, N. (2006): "Analysis of high dimensional multivariate stochastic volatility models", *Journal of Econometrics* 134, pp. 341-371.

Chicago Board Options Exchange (2009a): "The CBOE volatility index - VIX. White Paper" (available at www.cboe.com/micro/vix).

Chicago Board Options Exchange (2009b): "S&P 500 implied correlation index. White Paper" (available at www.cboe.com/micro/vix).

Chamberlain, G. (1983): "A characterization of the distributions that imply mean-variance utility functions", *Journal of Economic Theory* 29, pp. 185-201.

Connor, G., Goldberg, L.R. and Korajczik, R.A. (2010): *Portfolio risk analysis*, Princeton University Press.

Demos, A. and Sentana, E. (1991): "Inequality constraints on GQARCH models", mimeo, London School of Economics.

Diebold, F.X. and Nerlove, M. (1989): "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model", *Journal of Applied Econometrics* 4, pp. 1-21.

Ding, Z. and Engle, R.F. (2001): "Large scale conditional covariance matrix modeling: estimation and testing", Academia Economic Papers V29N2.

Donnelly, C. and Embrechts, P. (2010): "The devil is in the tails: actuarial mathematics and the subprime mortgage crisis", *ASTIN Bulletin* 40, pp. 1–33.

Duffie, D., Eckner A., Horel, G. and Saita, L. (2009): "Frailty correlated default", *Journal of Finance*, 64(5), pp. 2089–2123.

Dungey, M., Fry, R., Gonzalez-Hermosillo, B. and Martin, V.L. (2005): "Empirical modelling of contagion: a review of methodologies", *Quantitative Finance,* 5, pp. 9–24.

Engle, R.F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica* 50, pp. 987-1007.

Engle, R.F. (2002): "Dynamic conditional correlation - a simple class of multivariate GARCH models", *Journal of Business and Economic Statistics* 20, pp. 339-350.

Engle, R.F. and Kelly, B. (2012): "Dynamic equicorrelation", *Journal of Business and Economic Statistics* 30, pp. 212-228.

Financial Conduit Authority (2018): "Statement on communications in relation to PRIIPs", https://www.fca.org.uk/node/47296

Fiorentini, G., Sentana, E. and Shephard, N. (2004): "Likelihood estimation of latent generalised ARCH structures", *Econometrica* 72, pp. 1481-1517.

Forbes, K.J. and Rigobon, R. (2002): "No contagion, only interdependence: measuring stock market comovements", *Journal of Finance* 57, pp. 2223–2261.

French, K. R., Schwert, G. W., Stambaugh, R. F. (1987): "Expected stock returns and volatility", *Journal of Financial Economics* 19, pp. 3-30.

Hansen, P. R. and Lunde, A.: (2005): "A realized variance for the whole day based on intermittent high-frequency data", *Journal of Financial Econometrics* 3, pp. 525-554.

He, C., Sentana, E. and Teräsvirta, T. (2008): "The connection between vector ARCH and vector random coefficient autoregressive models with applications", mimeo, CEMFI.

IHS Global Inc (2015): *Eviews 9: Command and Programming Reference.*

Jones, R. and Vischer, A. (2015): "A primer on Euro Stoxx implied correlation", available at http://www.eurexchange.com/

Kim, S., Shephard, N. and Chib, S. (1998): "Stochastic volatility: likelihood inference and comparison with ARCH models", *Review of Economic Studies* 65, pp. 361–393.

King, M.A., Sentana, E. and Wadhwani, S.B. (1994): "Volatility and links between national stock markets", *Econometrica* 62, pp. 901-933.

King, M.A. and Wadhwani, S.B. (1990): "Transmission of volatility between stock markets", *Review of Financial Studies* 3, pp. 5-33.

Liu, L.Y., Patton, A. J. and Sheppard, K. (2015): "Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes", *Journal of Econometrics* 187, pp. 293-311.

Lo, A.W. (2002): "The statistics of Sharpe ratios", *Financial Analysts Journal* 58, pp. 36–52.

Longin, F. and Solnik, B. (2001): "Extreme correlation of international equity markets", *Journal of Finance* 56, pp. 649-676.

MacKenzie, D. and T. Spears (2012): "The formula that killed Wall Street? The Gaussian copula and the material cultures of modelling", mimeo, School of Social & Political Science, University of Edinburgh.

Markowitz, H. (1952). "Portfolio selection", *Journal of Finance* 8, pp. 77-91.

Martin, I. (2011): "A simple variance swap", NBER Working Paper 16884.

Meddahi, N. and Yamashita, M. (2017): "Return predictability and risk management", mimeo, Toulouse School of Economics.

Mencía, J. and Sentana, E. (2009): "Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation", *Journal of Econometrics* 153, pp. 105-121.

Mencía, J. and Sentana, E. (2012): "Distributional tests in multivariate dynamic models with normal and Student t innovations", *Review of Economics and Statistics* 94, pp. 133-152.

Mencía, J. and Sentana, E. (2013): "Valuation of VIX derivatives", *Journal of Financial Economics* 108, pp. 367-391.

Mencía, J. and Sentana, E. (2016): "Volatility exchange traded assets: an empirical investigation", forthcoming in the *Journal of Business and Economic Statistics.*

Merton, R.C. (1973): "Theory of rational option pricing", *Bell Journal of Economics and Management Science* 4, pp. 141-183.

Nelson, D.B. (1990): "Stationarity and persistence in the GARCH(1,1) model", *Econometric Theory* 6, pp. 318-334.

Nelson, D.B. and Cao, C.Q. (1992): "Inequality constraints in the univariate GARCH model", *Journal of Business and Economic Statistics* 10, pp. 229-235.

Owen, J. and R. Rabinovitch (1983): "On the class of elliptical distributions and their applications to the theory of portfolio choice", *Journal of Finance* 38, pp. 745-752.

Pan, J. and Singleton, K.J. (2008): "Default and recovery implicit in the term structure of sovereign CDS spreads", *Journal of Finance*, 63, pp. 2345–2384.

Rhoads, R. (2011): *Trading VIX derivatives: trading and hedging strategies using*

VIX futures, options and exchange traded notes, John Wiley and Sons.

RiskMetrics Group (1996): *RiskMetrics Technical Document*, 4^{th}edition, J.P. Morgan/Reuters.

Roll, R. (1989): "Price volatility, international market links, and their implications for regulatory policies", *Journal of Financial Services Research* 3, pp. 211–246.

Ross, S.A. (1976): "The arbitrage theory of capital asset pricing", *Journal of Economic Theory* 13, pp. 341-360.

Ross, S.A. (1978): "Mutual fund separation in financial theory - the separating distributions", *Journal of Economic Theory* 17, pp. 254-286.

Salmon, F. (2009): "Recipe for disaster: the formula that killed Wall Street", *Wired*, 23 February, http://www.wired.com/print/techbiz/it/magazine/17-03/wp_quant

Schwert, G. W. (2011): "Stock volatility during the recent financial crisis", *European Financial Management* 17, pp. 789-805.

Sentana, E. (1995): "Quadratic ARCH models", *Review of Economic Studies* 62, pp. 639-661

Sentana, E. (2003): "Mean-variance portfolio allocation with a value at risk constraint", *Revista de Economía Financiera* 1, pp. 4-14.

Sentana, E., Calzolari, G. and Fiorentini, G. (2008): "Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks", *Journal of Econometrics* 146, pp. 10-25.

Sharpe, W.F. (1966): "Mutual fund performance", *Journal of Business* 39, pp. 119-138.

StataCorp LP (2015): *STATA Time Series Reference Manual Release 14*.

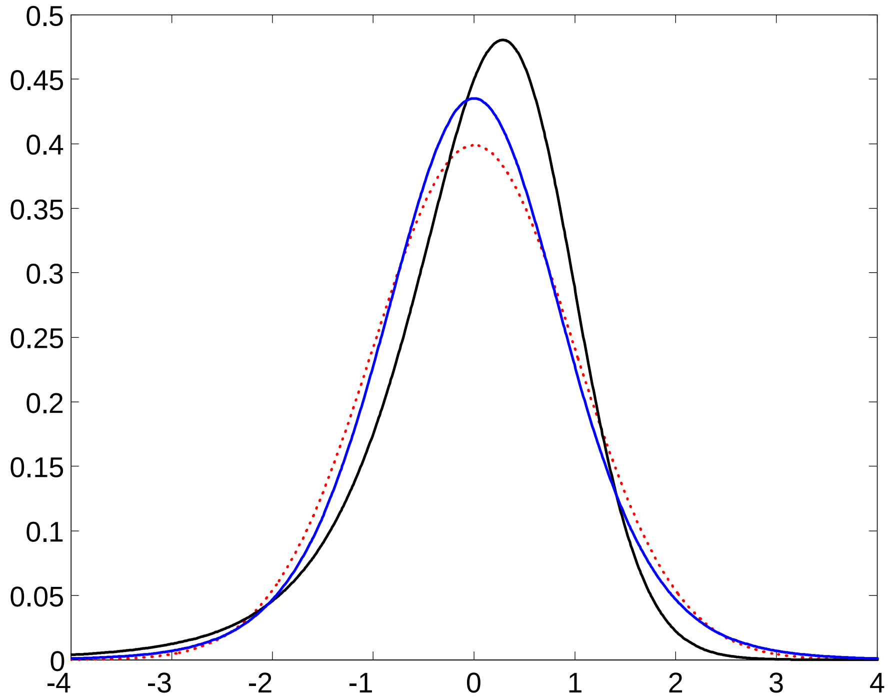
Tse, Y.K. and Tsui, A.K.C. (2002): "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations", *Journal of Business and Economic Statistics* 20, pp. 351-362.

Vasicek, O.A. (1987): "Probability of loss on loan portfolio", KMV Corporation.

Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005): "A tale of two time scales: determining integrated volatility with noisy high-frequency data", *Journal of the American Statistical Association* 100, pp. 1394-1411.

Zumbach, G. (2007): "The RiskMetrics 2006 methodology", mimeo, RiskMetrics Group.

Figure 1: Densities of three standardised random variables: Gaussian, symmetric Student t and asymmetric Student t



Note: The asymmetric Student t has ten degrees of freedom, as its symmetric counterpart, but skewness parameter -1.5. All three distributions, including the Gaussian one (dotted line), have been standardised so that they have zero mean and unit standard deviation.

Figure 2: Temporal evolution of the S&P500 implied volatilities

Figure 2a: 1-month (1990-2018)

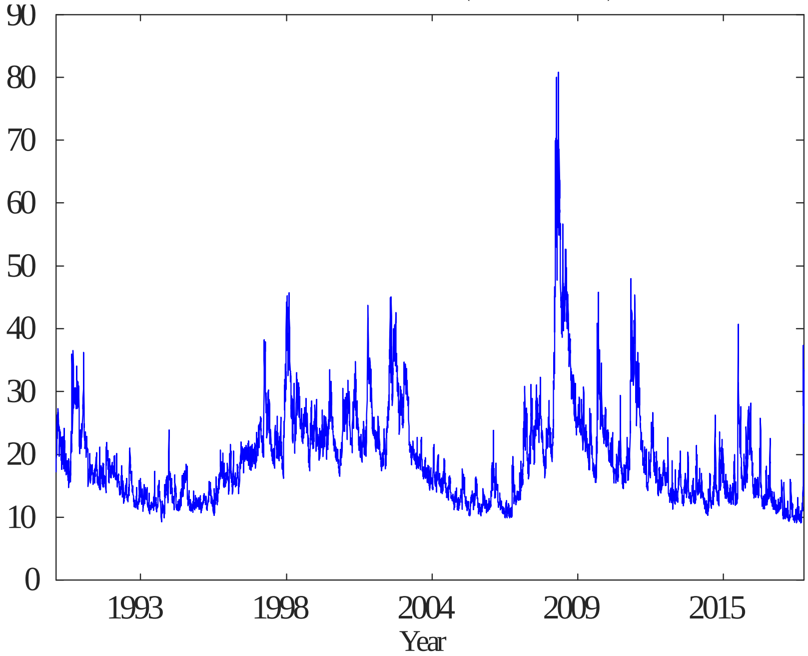
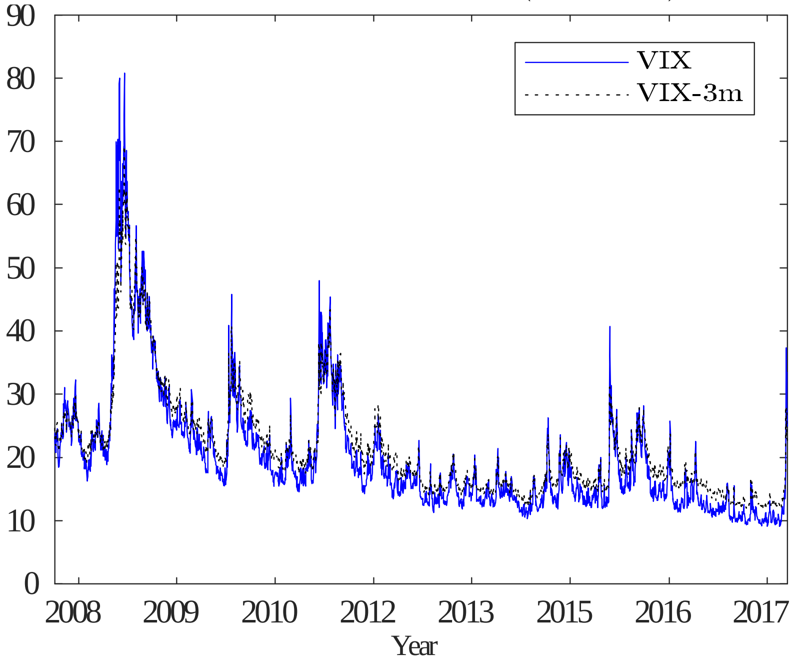


Figure 2b: 1- and 3-months (2007-2018)



Note: Figure 2a displays the temporal evolution of the VIX (one-month horizon) over its entire history, while Figure 2b, which also displays the VIX-3M (three-month horizon), starts at the time of the global financial crisis.

Figure 3a: Standardised bivariate normal density

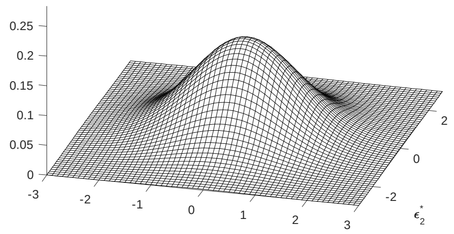


Figure 3b: Contours of a standardised bivariate normal density

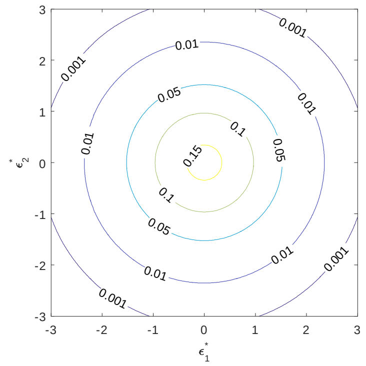


Figure 3c: Standardised bivariate Student t density with 8 degrees of freedom ($\eta = .125$)

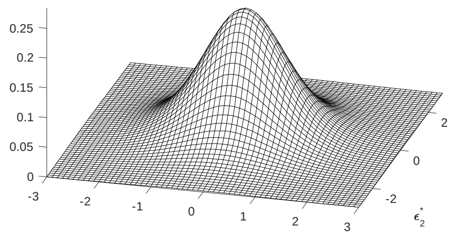


Figure 3d: Contours of a standardised bivariate Student t density with 8 degrees of freedom ($\eta = .125$)

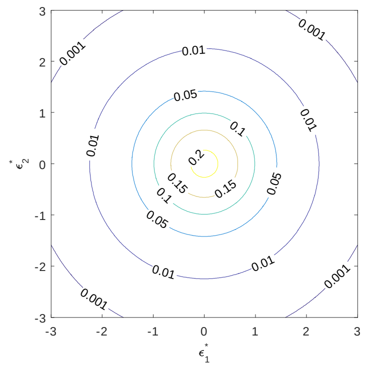


Figure 3e: Standardised bivariate asymmetric t density with 8 degrees of freedom ($\eta = .125$) and $\beta = (-2, -2)$

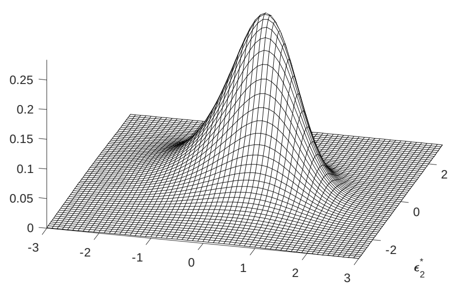
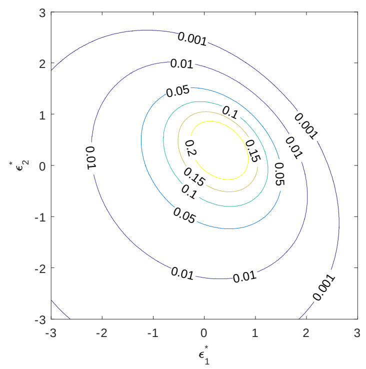


Figure 3f: Contours of a standardised bivariate asymmetric t density with 8 degrees of freedom ($\eta = .125$) and $\beta = (-2, -2)$



Notes: All three bivariate distributions have uncorrelated components and marginals with zero means and unit standard deviations. Panels a-b: Gaussian. Panels c-d: Student t with 8 degrees of freedom. Panels e-f: Asymmetric Student t with 8 degrees of freedom and skewness parameters -2.

Figure 4: Comparison of implied vol measures for stock indices

Figure 4a: S&P500, Nasdaq 100 and Russell 2000

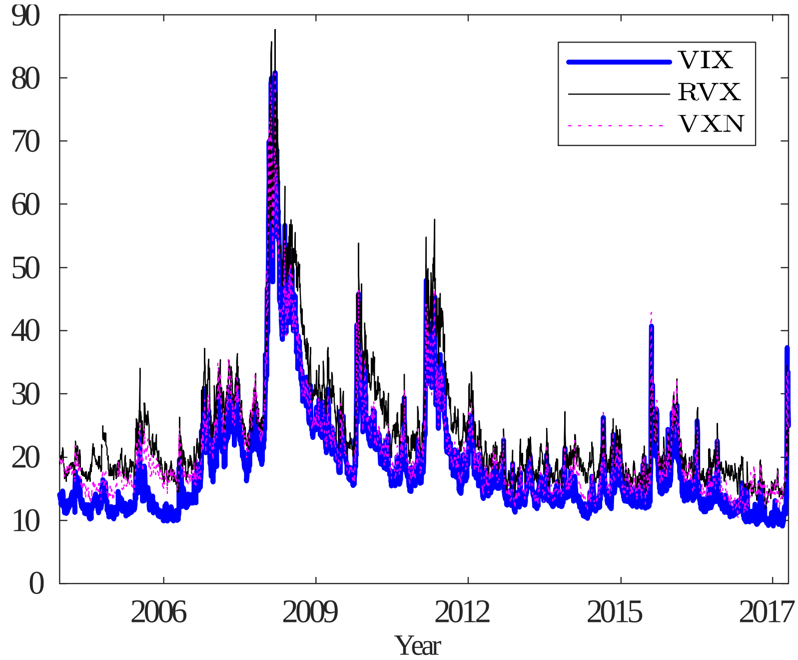
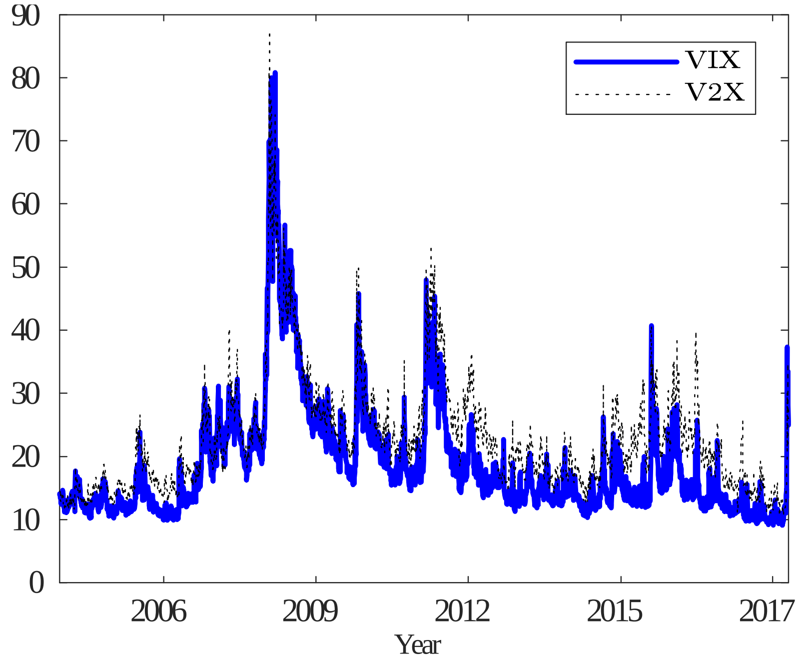


Figure 4b: S&P500 and Euro Stoxx



Note: Figure 4a compares the CBOE volatility indices for the S&P500 (VIX), Nasdaq 100 (VXN) and Russell 2000 (RVX), while Figure 4b compares the VIX with the V2X, which is the analogue volatility index for the Euro Stoxx 50.

Figure 5: Implied vol decomposition and implied correlation for the IBEX 35

Figure 5a: Implied vol decomposition

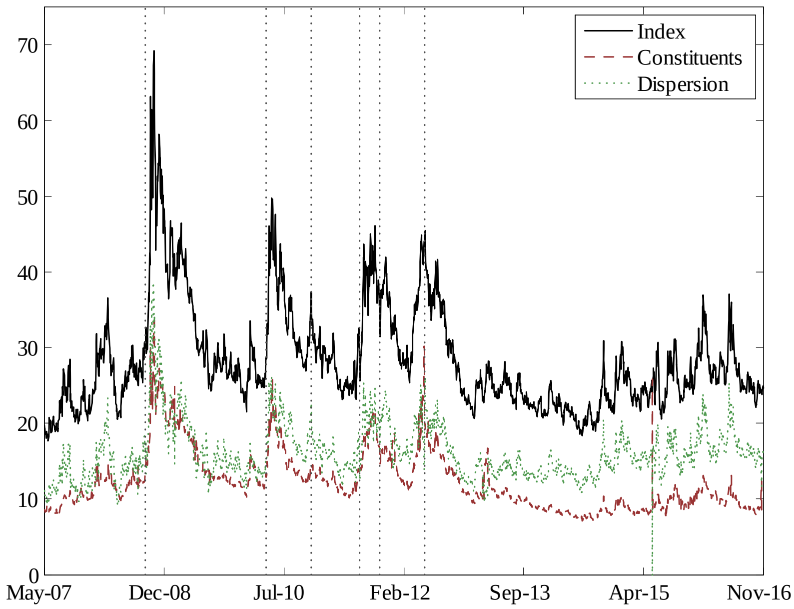
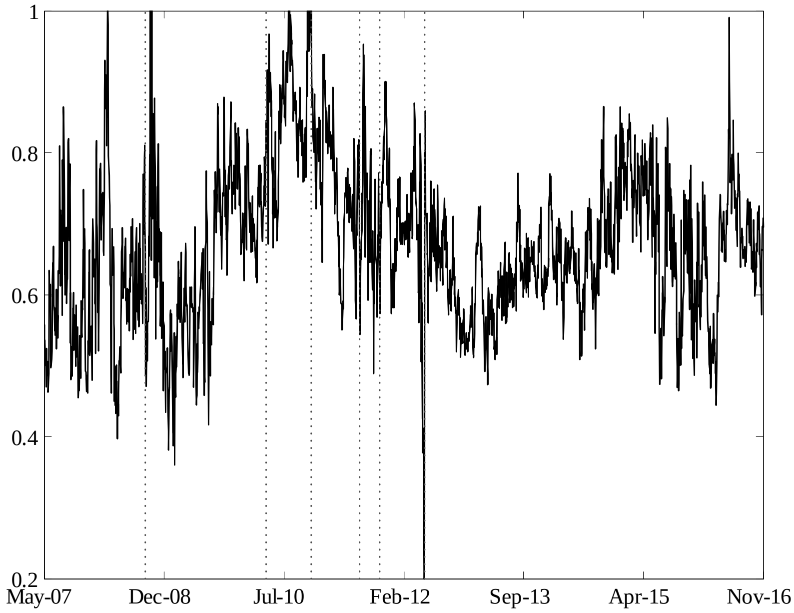
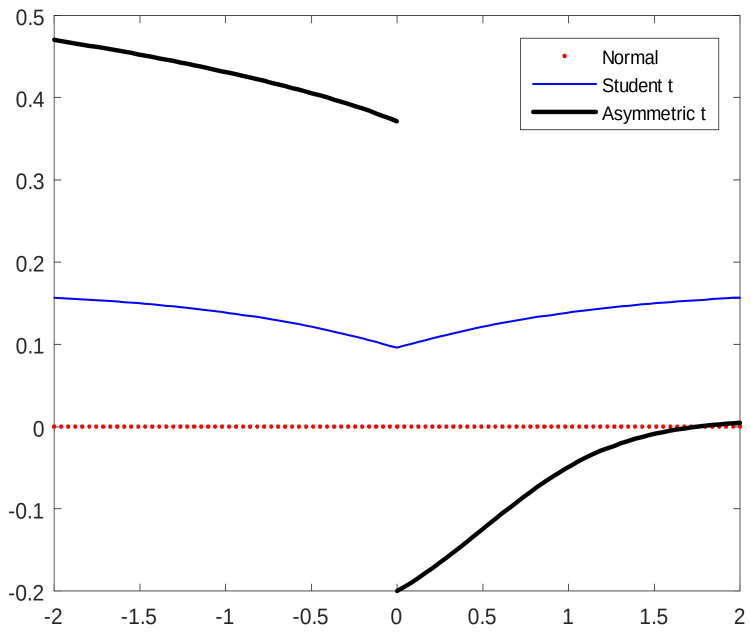


Figure 5b: Implied correlation



Notes: These figures plot the implied vol decomposition of the annualized three-months time-to-maturity variance swap rates for the IBEX 35 (Spain), as well as its implied correlation. The sampling period is from May 14^{th}, 2007 to November 30^{th}, 2016. Variance swap rates for each index/stock are computed using observations across all strikes of out-the-money (OTM) options from the OptionMetrics Ivy DB Europe database, following the methodology proposed by Bakshi, Kapadia and Madan (2003). Vertical lines correspond to 15-Sep-2008, 23-Apr-2010, 29-Nov-2010, 21-Jul-2011, 26-Oct-2011 and 1-Jun-2012.

Figure 6: Excedance correlations for bivariate normal, Student t and asymmetric Student t whose components are uncorrelated



Note: The three bivariate random variables coincide with the ones in Figure 3. Therefore, the asymmetric Student t has eight degrees of freedom, as its symmetric counterpart, but skewness parameters (-2,-2). Furthermore, all three bivariate distributions have been standardised so that their components are uncorrelated and their marginals have zero means and unit standard deviations.

Figure 7: Volatilities from three-month time-to-maturity variance swap rates during the European sovereign debt crisis

Figure 7a: IBEX35 and main Spanish banks

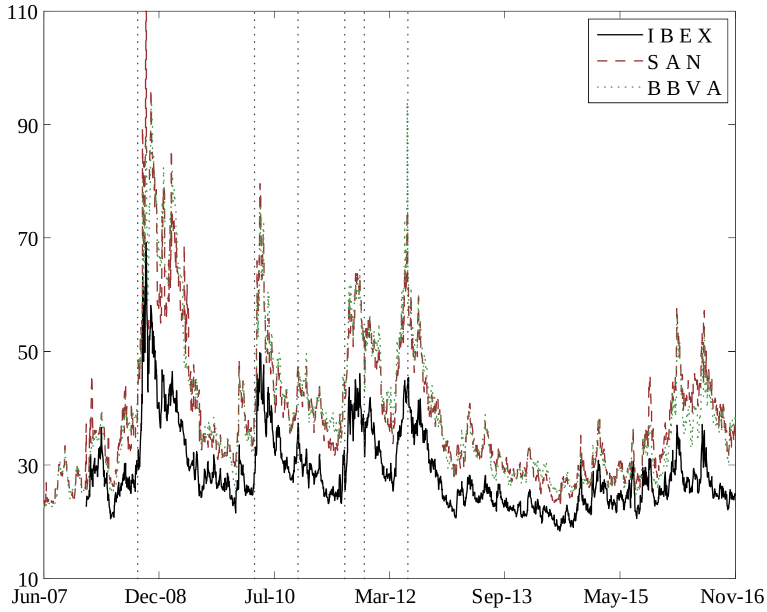
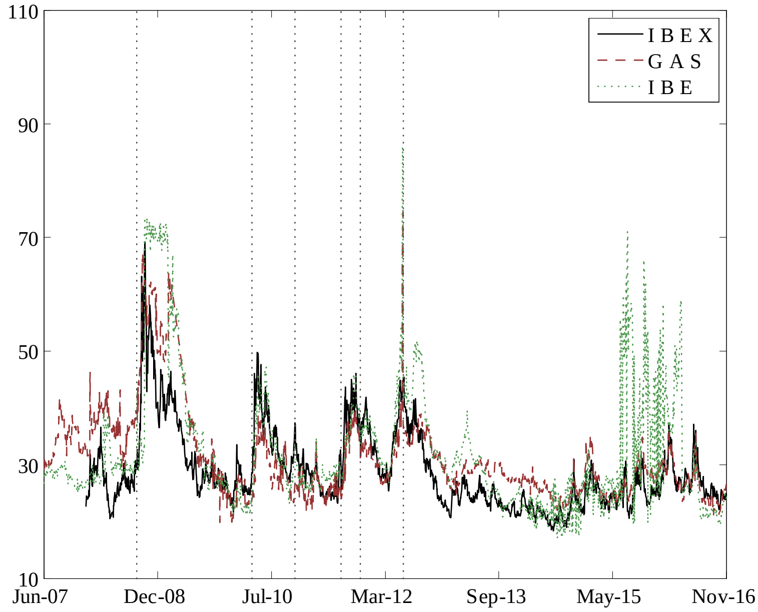


Figure 7b: IBEX35 and main utilities companies



Notes: These figures plot the square root of the annualized variance swap rates for the IBEX35 (largest companies in Spain), Banco Santander (SAN), Banco Bilbao Vizcaya Argentaria (BBVA), Gas Natural SDG (GAS) and Iberdrola (IBE). The sampling period is from June 1^{st}, 2007 to November 30^{th}, 2016. Variance swap rates for each index are computed using observations across all strikes of out-the-money (OTM) options from the OptionMetrics Ivy DB Europe database, following the methodology proposed by Bakshi, Kapadia and Madan (2003). Vertical lines correspond to 15-Sep-2008, 23-Apr-2010, 29-Nov-2010, 21-Jul-2011, 26-Oct-2011 and 1-Jun-2012.