UNIVERSITÀ DELLA SVIZZERA ITALIANA

DOCTORAL THESIS

Derivative Securities in Risk Management and Asset Pricing

Author:
Chiara LEGNAZZI

Supervisor:
Giovanni BARONE-ADESI

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in

Finance
Università della Svizzera Italiana
Swiss Finance Institute

October 8, 2018

To the memory of my grandfather Augusto Legnazzi (1929-2007),
one of the most genial and curious people I have ever met.

Abstract

The high informational content and the ease of accessibility are among the most attractive features which make derivative securities particularly useful in financial applications. With a special focus on risk management and asset pricing, I present several methodologies which involve the use of option and futures data in the estimation process. This doctoral thesis consists of three chapters. The first one, "Forward-looking VaR and CVaR: an application to the Natural gas Market", presents, backtests and compares point risk forecasts for the natural gas market using a novel methodology which introduces derivative securities into a classical calibration setting. The second chapter, "A Bayesian Estimate of the Pricing Kernel", is a joint work with G. Barone-Adesi and A. Mira. The article sets the pricing kernel estimation into a Bayesian framework, which enables to combine the use of derivative and historical data in the physical density calibration. Thanks to their higher accuracy and flexibility, the resulting pricing kernel estimates display a monotonic decreasing shape over a large range of returns, consistently with the classical theory. The third chapter, "S&P 500 Index, an Option-Implied Risk Analysis", is a joint work with G. Barone-Adesi and C.Sala. Tested on the US equity market, the article presents a detailed analysis on the performance of the option-implied risk metrics both in absolute terms and relative to the existing historical-based risk metrics.

Keywords: Derivative Securities, Hybrid and Option-implied VaR and CVaR, Risk Neutral Measure, Pricing Kernel, Poisson-Dirichlet process.

Acknowledgments

I would like to express my sincere gratitude to my advisor Prof. Giovanni Barone Adesi for the continuous support during my Ph.D study and related research, for his help, motivation, and profound knowledge. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Antonio Mele, Prof. Antonietta Mira and Prof. Hélyette Geman, for their insightful comments and encouragement, but also for the challenging questions which incented me to widen my research from various perspectives.

I gratefully acknowledge the Swiss National Science Foundation for having provided the funding sources that made my Ph.D. work possible.

I would like to thank my family for their unconditional love and encouragement. For my parents who raised me with a love of science and curiosity about the world. For the presence of my sister Livia, who always kept me in highest regard, making me acting and feeling as a better person. For my boyfriend Lorenz, who always looked at me with lover's eyes and supported me in any choice.

Contents

Part I

Forward-looking VaR and CVaR: an application to the Natural gas Market

1 Introduction

The article presents a methodology finalized at the derivation of forward-looking point risk forecasts. The proposed approach is denoted as "hybrid" because it exploits the advantages and, at the same time overcomes the limits, of the classical and option-implied approaches already existing in the risk management literature. To test the validity of the hybrid methodology, the article presents a direct application of the proposed approach to the US natural gas market.

Over the last decades energy commodities have attracted considerable interest as alternative financial investments. The intrinsic features of these assets, such as high-volatility, high correlation with inflation, counter-cyclicality within the economy and low correlation with bonds and equities, have driven the development of a liquid derivative market. Being the closest substitute to oil, natural gas has gained increasing importance both in real and financial markets. From 2007 onwards in the US real market, the hydraulic fracturing and the horizontal drilling have allowed for the extraction of shale gas, thus prompting the sharp increase of the US gas production and the subsequent drop of the price (see figure 3). In financial markets, starting from the deregulation of wellhead prices in 1989, natural gas has experienced the development of an active derivative market. The growth of US natural gas futures market volumes[1] (see figure 1) has been accompanied by an increase in the share of non-commercial investors[2] actively trading in this commodity (see figure 2). Besides supply/demand-linked risks[3], the increasing trading activity of non-commercial agents has strongly amplified the magnitude of unanticipated changes into the natural gas dynamics (see Geman (2009)[50]), thus making profit and loss forecasting in this market a challenging task to perform.

This article relates to two streams of research. First, from a methodological point of view,

[1]In a quarterly review in 2007, the Bank of International Settlements (BIS) documented an overall increase of open interest in the futures and options markets in the period from 1994 to 2004.
[2]According to the US Commodity Futures Trading Commission (CFTC) definition, an investors is "non-commercial" if he/she does not trade primarily for hedging purposes.
[3]For example, risk associated to inventories, to the availability and unexpected revisions of the proven reserves.

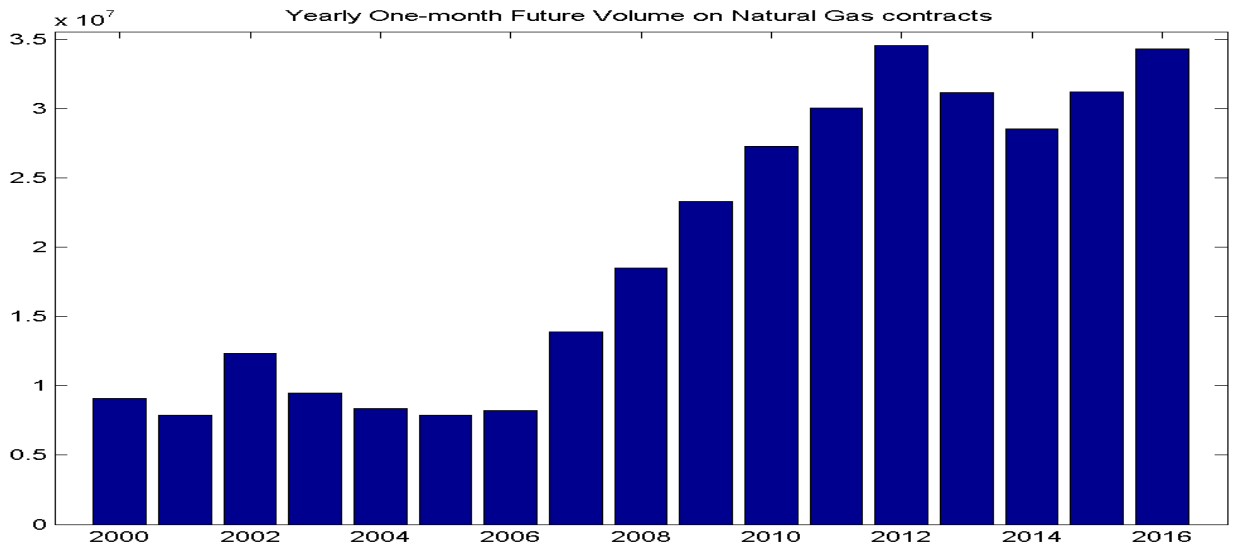

Figure 1: Yearly volumes of the Henry Hub natural gas futures contracts with maturity one-month from January 2000 to December 2016.

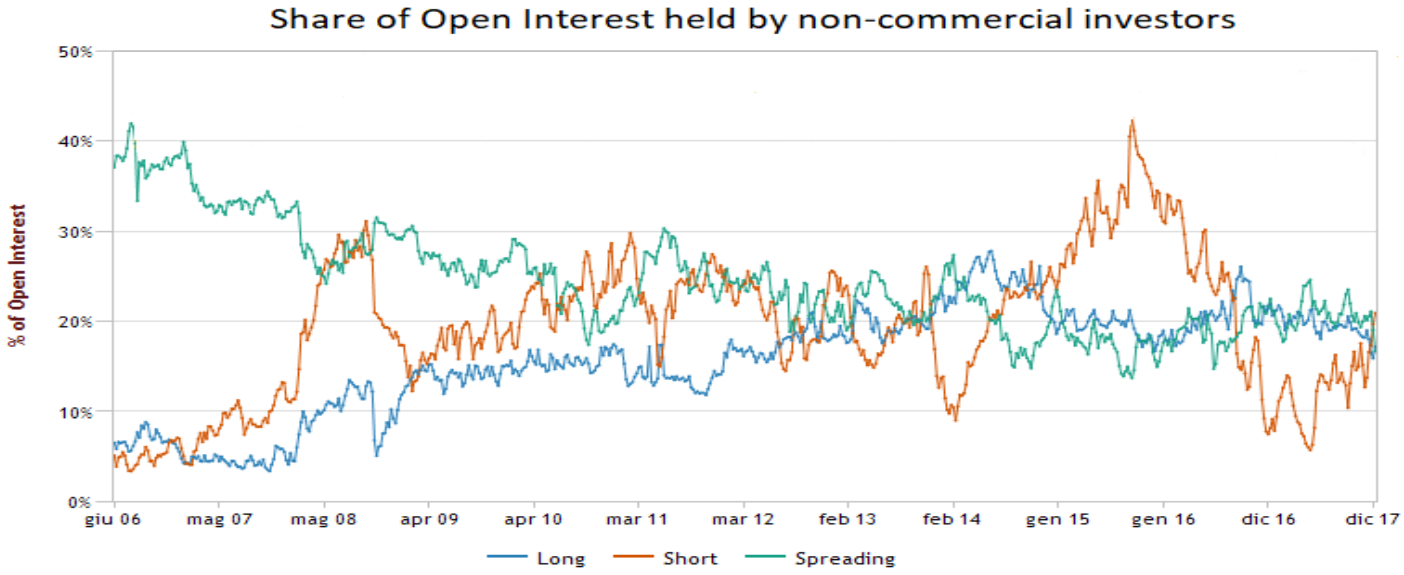

Figure 2: Evolution of the share of open interest in options and futures contracts on the Henry Hub natural gas hold by non commercial investors from January 2006 to December 2017.

the option-based VaR and CVaR estimates are derived using a novel methodology, substantially different from the majority of risk models existing in the literature. Second, the empirical application of the proposed approach to the natural gas market and the comparison with alternative risk

models enriches the existing risk management literature in energy markets. In the financial risk management literature the majority of risk models bases their risk predictions on historical data only (these models are henceforth denoted as "classical"). Since statistical analysis on past data may be not informative about the future dynamics[4] of the underlying asset, classical estimates often turn out to be biased. Thanks to their higher informational content (see Merton (1973)[82] for a complete review), financial instruments traded in the derivative market can naturally provide a solution. To date, the use of derivative data in risk management is limited to a small number of researches (see for instance Christoffersen et al. (2016)[33], Barone Adesi (2016)[7] and Bellini and Di Bernardino (2017)[17]). Although the option-implied approach of Barone Adesi delivers accurate risk estimates for several asset classes[5], its applicability is restricted only to those underlying assets with a liquid derivative market and to a limited choice of investment horizons. To date, in the energy commodities literature the majority of the academic articles have analysed and measured the impact on natural gas of risks strictly connected to real markets[6]. Indeed, for a long time the derivation of financial point risk forecasts in energy markets has been widely overlooked. Few examples of researches in this field are those by Pedraz et al. (2014)[57], Youssef et al. (2015)[97] and Pantoja et al. (2017)[87].

The contribution of the article to the existing literature is fourfold. First, the hybrid methodology enables one to derive option-implied risk forecasts at any frequency and investment horizon, even in periods of low trading activity in the option market. Because of these advantages, the range of applicability of the hybrid approach is very wide and includes financial instruments (for example individual companies) which do not normally have a high trading activity in the derivative market. Second, the hybrid methodology is quick and easy to implement. Futures and option data are publicly available and easy to collect. The model calibration is not time consuming, thus making possible to perform any risk forecasting exercise on a timely basis. Third, the article presents an extensive and detailed analysis on the validation and the comparative performance of the hybrid, option-implied and classical risk metrics. Fourth, the obtained results can be valuable for a wide set of economic agents. Existing and potential investors can assess the risk associated to a specific underlying in a simple and effective way. Individual firms can gain insights about their percep-

[4]See for instance the discussion about the "Peso problem" in Krasker (1980)[71].
[5]See Barone-Adesi et al. (2016)[13] and Barone-Adesi et al. (2016)[12] for an empirical application on the oil and the US equity markets, respectively.
[6]For example, supply/demand-linked risks (Geman and Ohana (2009)[52]), convenience yield risk and counterparty risk .

tion among the investors and derive alternative risk estimates to those delivered by their internal models.

The remainder of the paper is as follows: Section 2 presents the methodology - section 2.1 and 2.2 briefly describe the characteristics and limits of the classical and the existing option-implied risk measures. Section 2.3 presents the hybrid methodology. Section 3 describes the dataset and its relevant characteristics. Section 5 presents the monthly and 10-days hybrid VaR and CVaR risk forecasts and section 5 presents alternative risk estimates, namely the monthly Black & Scholes option-implied VaR and CVaR and the monthly and 10-days VaR estimates under the classical FHS approach. Section 6 validates all the proposed VaR and CVaR estimates on the basis of the backtesting results. Section 7 ranks and compares the VaR forecasts across the proposed methodologies. Section 8 presents conclusions.

2 The Methodology

To date the majority of risk management models derives point risk forecasts based on historical data under the physical measure. Recently, Barone-Adesi (2016)[7] has extracted the VaR and the CVaR from derivative securities under the risk neutral measure. This methodology is henceforth denoted as "option-implied" and is used as a benchmark to evaluate the hybrid approach. The proposed hybrid methodology shares some similarities both with the classical and the option-implied approaches. Sections 2.1 and 2.2 outline the definitions of VaR and CVaR under the physical and risk neutral measures, respectively. Section 2.3 presents the methodology to derive the hybrid VaR and CVaR.

2.1 The classical VaR and CVaR

The VaR is a measure of downside risk based on the current value of the underlying assets. By definition, the Value-at-Risk, $\text{VaR}_{\alpha,t,\tau}$, is the possible maximal loss associated with a portfolio, S, occurring with probability $(1-\alpha)$ at time t and over a given time horizon $\tau = T - t$ (see Jorion (1997)[67]). The estimated loss, L, is defined with respect to the initial investment, S_0, and equals $L(S) = S_0 - S$. In this framework, the VaR equals the negative α-quantile of the P&L distribution

$$\text{VaR}| \qquad \int_{-\infty}^{-VaR} f(L(S))dS = \alpha \tag{1}$$

where $f(L(S))$ is the probability density distribution of the future P&L based on the value of the portfolio today, S_0, and α is the desired risk level. The high sensitivity of S to the statistician's modelling choices makes the probability density function of the future losses the most challenging element to estimate. Taking into account the magnitude of the losses beyond a specified risk level, the CVaR is a weighted tail expectation, defined as

$$CVaR = \frac{1}{\alpha}\int_{-\infty}^{\alpha} L(S)f(S)dS \tag{2}$$

where $f(S)$ is the probability distribution function of the future portfolio values. Despite the advantage of being a coherent risk measure, the lack of elicitability prevents from making reliable inference and meaningful comparisons with other risk metrics based on the backtesting results.

2.2 The option-implied VaR and CVaR

Following Barone Adesi (2016)[7], this section briefly describes how to derive point risk forecasts directly from option prices. The option-implied VaR is the difference between the initial portfolio value, S_t, and the strike of the European put option[7] identified by the desired risk level α. For continuous distributions,

$$VaR_{\alpha,t,\tau} = S_t - K_{\alpha,t,\tau} \qquad \text{such that} \qquad \int_{-\infty}^{K} f(S)dS = F(K) = \alpha \tag{3}$$

where the subscript $\{\alpha, t, \tau\}$ corresponds to quantities estimated at time t with time to maturity/investment horizon τ and identified by the risk level α. In economic terms, K is the strike price of that put option which investors think will expire in the money with $\alpha\%$ probability. By shifting the lower integration limit to zero[8], the option-implied representation of the CVaR follows

$$CVaR_{\alpha,t,\tau} \quad = \quad VaR_{\alpha,t,\tau} + e^{r\tau}\frac{p_{\alpha,t,\tau}}{\alpha} \tag{4}$$

From a modelling perspective, the computation of α can be performed either nonparametrically or parametrically. Following Barone-Adesi and Elliot (2007)[8], α relative to the middle option, $\alpha_{[2]}$,

[7]The same analysis can be performed using call options contracts.
[8]It is equivalent to assume a portfolio with limited liability.

can be computed analytically as

$$\alpha_{[2]} = e^{r\tau}\frac{\partial p}{\partial K} = e^{r\tau}\cdot\frac{1}{2}\left[\frac{p_2 - p_1}{K_2 - K_1} + \frac{p_3 - p_2}{K_3 - K_2}\right]; \tag{5}$$

where $p_1 < p_2 < p_3$ are the prices of three options at time t and with time to maturity τ with contiguous strikes ($K_1 < K_2 < K_3$, respectively). Needing data on three options trading at consecutive strike prices to derive just one estimate of alpha, the applicability of this approach (henceforth denoted as "Model-Free (MF) methodology") is limited to underlying assets with a liquid option market only. Alternatively, under the assumption that the observed option prices are generated by the Black and Scholes (B&S) model (1973)[20], α equals the probability of a European put to expire in the money

$$\alpha = \mathrm{N}(-d_2) = 1 - \mathrm{N}\left(\frac{\ln\left(\frac{F}{K}\right) - \frac{\sigma^2}{2}\tau}{\sigma\sqrt{\tau}}\right) \tag{6}$$

Even though the resulting α estimates rely on the specification of an option pricing model, this approach (henceforth denoted as "B&S methodology") allows one to derive an estimate of α for each observed option price. The B&S methodology is outperformed by the MF method in terms of accuracy and precision, but provides a larger amount of VaR and CVaR estimates when the market liquidity is lower.

2.3 The hybrid methodology

The hybrid methodology shares some common features with both the option-implied and the classical approaches. Similarly to the former ones, it uses, though requiring a much lower amount, option prices as inputs and is derived under the risk neutral measure. At the same time, as the classical approaches, it specifies a stochastic process to model the P&L distribution. Since many asset classes do not have a highly liquid derivative market, the hybrid methodology can be very attractive. Indeed, even when the derivative market of the underlying asset is not well developed, this approach enables one to derive forward-looking risk forecasts, squeezing all the informative content of option data. Moreover, the hybrid approach delivers point risk forecasts at any frequency[9] and at any investment horizon. Indeed, while the option-implied MF and B&S approaches enable one to derive VaR and CVaR risk estimates only at the investment horizons equal to the time to maturity

[9]As long as at any observation date few options are traded.

of the underlying option contracts[10]; the hybrid methodology can deliver point risk forecasts at any investment horizon by simulating the underlying asset dynamics up to the desired date. However, the specification of a mathematical model prevents the hybrid approach to reach the same flexibility of the option-implied MF and B&S methodologies. The choice of an appropriate mathematical structure to describe the natural gas futures price dynamics is the most delicate and challenging task. To gain a better understanding of the characteristics of futures data over the sample period under consideration, the relevant empirical moments are computed and the existence of correlation among residuals is tested. Table 1 reports the first four annualized empirical moments of the daily log-returns on one-month natural gas futures contracts over the years 2012-2016. The high storage

	Mean	Standard deviation	Skewness	Kurtosis
2012	0.0765	0.5040	0.3063	3.7025
2013	0.2677	0.3103	-0.0861	3.4043
2014	-0.4010	0.4854	0.0348	5.1702
2015	-0.2497	0.4610	0.1918	4.5270
2016	0.4654	0.4796	0.3016	3.6845

Table 1: First four annualized empirical moments of daily log-returns on natural gas Henry Hub futures with fixed maturity one month over the years 2012-2016.

costs and the strict relationship with the electricity market (the most volatile energy commodity) are the major drivers of the large volatility of the natural gas market. The skewness is positive and the kurtosis is well above 3, thus indicating that the distribution is asymmetric and leptokurtic [11]. As confirmed by the autocorrelation and partial autocorrelation functions[12] and by the results of the Ljung-Box (Ljung and Box (1978)[77]) (LB) test, residuals are autocorrelated. The Lagrange Multiplier (Engle (1984)[47]) test and the LB test applied to squared residuals are implemented to test for conditional heteroskedasticity of the residuals. Results confirm that the null hypothesis of no conditional heteroskedasticity is always rejected.

In light of these results, a GARCH(1,1)[13] with filtered historical simulated (FHS)[10] innovations is selected to model the futures price dynamics and is fitted to the cross section of OTM call and put options. At each date t, the derivation of the risk neutral density requires a three-steps procedure: (1) the set of physical GARCH parameters, θ_t^P, is derived by fitting a GARCH(1,1) model to the

[10]In this specific case, the option-implied estimates have a monthly frequency because they are based on futures and option contracts with fixed maturity one month.
[11]It has "fatter" tails than a Gaussian distribution
[12]Figures are available upon request.
[13]See Bollerslev (1986)[23].

historical log-returns of the futures price, (2) using θ_t^P as starting values, the risk-neutral GARCH parameters, θ_t^Q, are found by minimizing the mean squared error of the difference between the GARCH option prices and the actual prices of the traded options, (3) based on θ_t^Q, the futures price can be simulated and the risk neutral P&L distribution can be obtained via nonparametric kernel estimation. At each observation date t, under the physical measure the futures prices dynamics are

$$\begin{aligned} log\frac{F_t}{F_{t-1}} &= \mu + \epsilon_t \qquad \epsilon_t = \sigma_t z_t \\ \sigma_t^2 &= \omega + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 \end{aligned} \tag{7}$$

where F_t is the futures price at time t, μ is the drift term and z_t is the standardized historical innovation at time t. The simple GARCH(1,1) model captures the conditional heteroskedasticity of the residuals and ensures a faster convergence to a local maximum compared to its more tailored specifications[14]. The FHS distribution enables one to model the relevant high-order moments features of the data and to use the obtained residuals as a set of empirical innovations for the underlying asset. Each innovation is sampled from its empirical density function, which is obtained by dividing the estimated return innovation, $\hat{\epsilon}_t$, by the estimated conditional volatility, $\hat{\sigma}_t$. The empirical density function generated by this set of estimated scaled innovations embeds excess skewness, kurtosis, and other extreme behaviours which would have been neglected under a normal density estimation. By running a Pseudo Maximum Likelihood estimation, the physical GARCH parameters are obtained and used as starting values to derive the risk neutral parameters. At each observation date and for each set of risk neutral parameters, N=50.000 return paths are simulated up to time $t+\tau$

$$\frac{F_{t+\tau}}{F_t} = \exp\left(\mu^*\tau + \sum_{i=1}^{\tau}\sigma_{t+i}z_{[i]}\right) = \exp\left(\sum_{i=1}^{\tau}\sigma_{t+i}z_{[i]}\right) \tag{8}$$

where τ is the desired investment horizon, $z_{[i]}$ is the estimated past innovation and μ^* is the drift. Following the intuition underneath the "risk neutrality argument" (Cox and Ross (1976)[34] and Harrison and Kreps (1979)[62]), the drift μ^* is equal to zero as a consequence of the martingale property of the futures price. More precisely, the absence of an initial cash outlay to enter the futures contract from either the long or short sides implies that traders are not entitled to receive any compensation for their investment. By the same token, the price of any given fixed term futures is

[14]FOr example, fractionally integrated, nonparametric and multiplicative component GARCH models

not affected by seasonalities and does not require the modelling of mean-reverting behaviours. Similarly, the convenience yield accrues to the holders of the physical commodity, not to the investors holding a derivative contract written on it. As a consequence, the drift term μ^*, which accounts for the convenience yield, seasonalities and mean-reversion, does not influence the dynamics of the futures prices as the premium for holding the risky futures position is already embedded into the risk neutral measure[15]. The implementation of the Empirical Martingale Simulation method of Duan and Simonato (1998)[40] ensures that the martingale condition is fulfilled. Then, the corresponding put prices are equal to the present value of the average simulated payoff, $\sum_{n=1}^{N} \max(K - F_{n,t+\tau})$, where K is the strike price and n identifies the n-th simulation[16]. The risk neutral GARCH parameters minimize the mean squared pricing error, $\frac{1}{N}\left[\sum_{i=1}^{D_t} \psi_t(K_i, \tau)\right]$, where D_t is the number of options on date t and $\psi_t(K_i, \tau)$ is the difference between the simulated GARCH option prices and the observed market prices of the OTM options across strikes K_i. For each set of risk neutral GARCH parameters θ_t^Q, 50.000 futures prices at a daily frequency from t to τ are simulated as in eq. (8). By subtracting to each simulated price at time $(t + \tau)$ the value of the natural gas futures at time t, the conditional profit and loss distribution is obtained via nonparametric kernel density estimation. Finally, the VaR and the CVaR are derived as in equations (1) and (2), respectively. Differently from the classical point risk forecasts, the hybrid VaR and CVaR are derived under the risk neutral measure. Since the discounted expectations under the physical and the risk neutral measures converge to the same state price density when the time to maturity is short, the option data provide reliable approximations of the VaR and CVaR under the real world measure (Barone-Adesi (2016)[7]). Supportive evidence is provided in the work of Martin (2017)[80]. Under the physical measure the probability of observing a crash, $\alpha_p \propto \left[\frac{\partial p}{\partial K} - \frac{p(K)}{\alpha F_t}\right]$, is always lower than the corresponding probability under the risk neutral measure, $\alpha_q \propto \frac{\partial p}{\partial K}$. As a result, at a given risk level the identified option under the physical measure has a higher strike compared to the one identified under the risk neutral one, thus making the VaR and CVaR estimates under the physical measure less conservative.

[15]This no-arbitrage assumption is the necessary and sufficient condition ensuring that futures prices are unbiased predictors of future spot prices (see chp. 5.3 in Geman (2009)[50]).
[16]A similar procedure is applied to derive call option prices.

3 The Dataset

The methodology presented in section 2 is tested on the natural gas futures contracts and the European put options written on them. Even thought the transformation of natural gas in its liquid form has substantially reduced the transportation costs and made it possible to reach many areas/countries without the construction of expensive infrastructures, natural gas still has a very fragmented market[17]. Since there does not exist a global reference price for this commodity, this research considers futures and option contracts traded on the New York Mercantile Exchange (NYMEX) and written on the Henry Hub natural gas index (for brevity, the Henry Hub specification in reference to futures and option contracts is henceforth omitted). The choice of the US market is motivated by the fact that North America was the first natural gas market being liberalized and its price is determined by the competitive interaction between supply and demand[18]. The size of a standard futures contract is 10.000 Million British Thermal Units (MMBTUs) and delivery/sale occurs at Henry Hub in Louisiana. An option contract is written on one natural gas futures contract and on any trading day it has a maximal set of 20 strike prices in increments of 0.05$ above and below the at-the-money (ATM) strike[19]. Daily one-month fixed-maturity futures prices are from Bloomberg and monthly option data are from Datastream. The sample period ranges from January 2012 to December 2016 with 1274 total observations for the daily natural gas futures price and 60 observation dates for the monthly option contracts. The choice of this sample period is motivated by the increasingly important role gained by natural gas over the last years and by the lack of available option data prior to January 2011. To avoid issues linked to the use of overlapping data, the monthly VaR and CVaR have investment horizons equal to one month[20], with a total number of 60 monthly observations for the hybrid, B&S and classical methodologies over the sample period from January 2012 to December 2016. Similarly, the frequency of the 10-days hybrid and classical FHS VaR is 10 days and the total number of estimates is 123 over the sample period from January 2012 to December 2016. Due to the lack of a sufficiently liquid market for OTM options, the option-implied model-free approach cannot be implemented over the sample period under study.

[17]See Geman and Liu (2016)[51] for a detailed analysis of the integration on the natural gas market.
[18]Also in the United Kingdom natural gas has been liberalized for many years, however, compared to the US case, its price is much more dependent on the crude oil price. So far in Europe, Russia satisfies about 30% of the natural gas needs. Moreover, six EU countries (Bulgaria, Estonia, Finland, Lithuania, Latvia and Slovakia) import gas exclusively from Russia, thus being substantially exposed to supplier bargaining and disruptions.
[19]The ATM strike is the closest strike to the previous day close of the underlying futures contract.
[20]Being fixed at τ=1 month, in sections 5 and 5 the subscript τ is omitted.

4 The hybrid VaR and CVaR estimates

Based on the methodology outlined in section 2.3, this section presents the monthly and 10-days hybrid VaR and CVaR estimates at multiple risk levels, namely $\alpha = 1\%, 5\%, 10\%$ and 15%, and over the sample period from January 2012 to December 2016. The hybrid methodology imposes a structure for the underlying asset dynamics and is not bounded to the availability of a liquid option market, hence there are no missing observations over the whole sample period. The VaR and the CVaR are decreasing functions of the risk level and the CVaR is always larger than the VaR. As a consequence of the high volatility of the natural gas futures price, the VaR and CVaR vary over a wide support across all risk levels. Indeed, as depicted in figure 3, in the first months

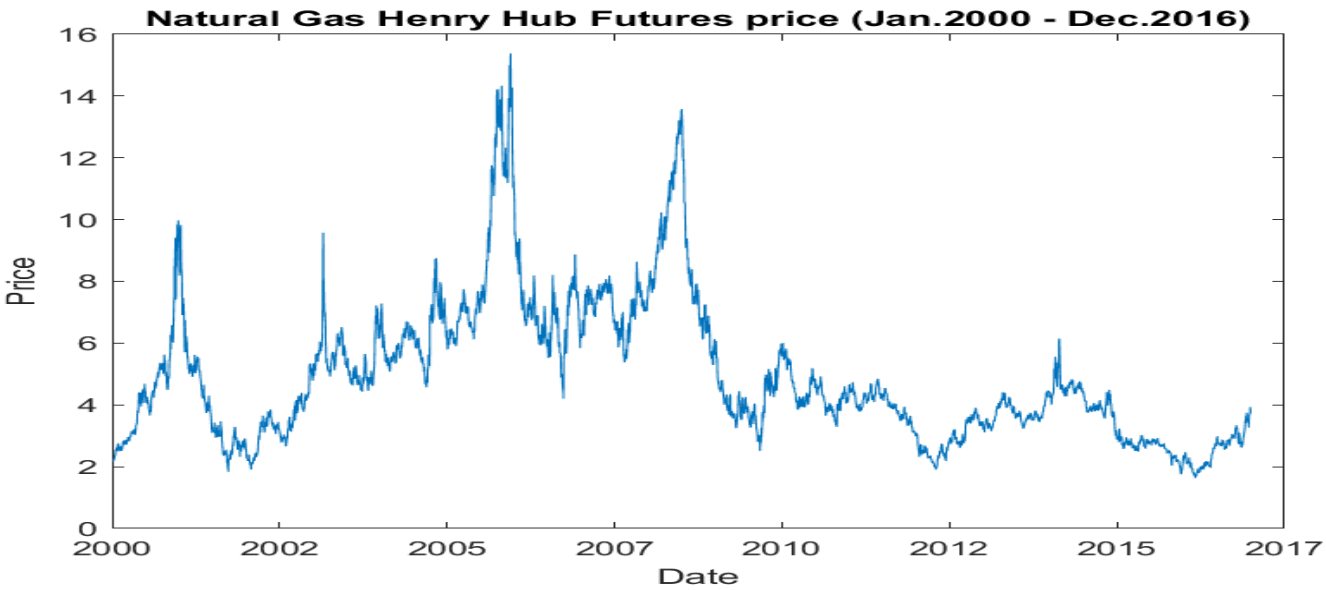

Figure 3: Price of the natural gas Henry Hub futures contracts traded on the NYMEX with fixed maturity one-month from January 2000 to December 2016.

of 2012 the natural gas futures price dropped below \$2, then it more than doubled its value in two years, with a peak hitting the price of almost \$6. In correspondence to this point of maximum in February 2014, the VaR and CVaR estimates registered a clear upward jump at all risk levels. In order to deliver point risk forecasts at a higher frequency, the 10-days hybrid VaR is derived over the sample period from January 2012 to December 2016. The 10-days estimates exhibit a series of small fluctuations which reflect the volatile behaviour of the natural gas market.

5 Alternative VaR and CVaR estimates

This section presents the monthly option-implied risk estimates using the B&S approach and the monthly and 10-days classical FHS VaR estimates at multiple risk levels, namely $\alpha = 1\%, 5\%, 10\%$ and 15%. The monthly B&S estimates are based on the left tail of the distribution, i.e. using European put option contracts with fixed maturity one month.

By assuming an option pricing model generating the observed prices, the B&S approach suffers in a limited way from the lower liquidity of the option market over some time intervals and can, in general, be implemented over the whole sample period. As shown in table 2, across all risk

	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 15\%$
# Missing observations (%)	3.33	6.66	8.33	11.66

Table 2: Proportion of missing observations in the left tail at $\alpha = 1\%, 5\%, 10\%$ and 15% under the B&S methodology when $\tau = 1$ month and over the sample period from January 2012 to December 2016 (total obs.=60).

specifications the proportion of missing observations is increasing in the risk level. The positive relation between α and the number of missing observations suggests that the market of deeply OTM options is more active than that of OTM options with strikes closer to the at-the-money threshold. The VaR and the CVaR are decreasing functions of the risk level and the CVaR is always larger than the VaR. The gap between the two risk forecasts, Δ_α, equals the compounded value of the put price discounted for the probability of that scenario, i.e. alpha. From an economic viewpoint, Δ represents the additional capital at risk perceived by the market if a specific negative event occurs. Table 3 reports summary statistics on Δ under the B&S methodology at $\alpha = 1\%, 5\%, 10\%$ and 15%. At each risk level, the average distance between the CVaR and the VaR is not biased by the

	B&S Δ_α - τ=1 month (2012-2016)			
	Min.	Max.	Mean	Median
$\alpha = 1\%$	0.0557	0.3195	0.0999	0.0912
$\alpha = 5\%$	0.0506	0.4218	0.1069	0.0929
$\alpha = 10\%$	0.0457	0.7754	0.1459	0.1099
$\alpha = 15\%$	0.0773	0.7754	0.1772	0.1348

Table 3: Left tail - Summary statistics on the monthly Δ_α= CVaR_α - VaR_α at $\alpha = 1\%, 5\%, 10\%$ and 15% . The monthly VaR and CVaR estimates are obtained under the assumption that the observed option prices are generated by a Black and Scholes model.

presence of outliers since the median almost coincides with the mean and, on average, Δ increases

with the risk level. Keeping alpha fixed, the gap between the CVaR and the VaR is driven by the put option price identified by the risk level, p_α. When alpha is small, the identified price is low as it corresponds to a deep OTM option. Nevertheless, in periods of uncertainty or turmoil, investors may become more risk averse, thus willing to pay more to buy hedging positions[21].

Since the big majority of risk management models use historical data to produce risk estimates under the physical measure, the classical VaR estimates are obtained by calibrating a GARCH(1,1) with FHS innovations (see eq.(7)) on the historical log-returns of the natural gas futures price. In this way, the underlying asset dynamics can be simulated and the $\text{VaR}_{\alpha,\tau}$ is derived as the α-quantile of the simulated P&L distribution up to the investment horizon τ. The resulting VaR estimates are denoted as "classical FHS" in the rest of the analysis. Compared to other classical approaches, the FHS methodology models the volatility of the underlying asset in a more accurate way, without relying on the square-root rule to extend the VaR forecasting horizon. Similarly to the hybrid approach, the FHS methodology can deliver point risk metrics at any frequency and over any investment horizon. Besides to the monthly estimates, to increase the number of observation dates and hence the power of the backtests, the classical FHS VaR at 10-days investment horizon is derived, backtested and ranked with respect to the corresponding hybrid estimates.

6 Backtesting

The backtesting is a form of model validation procedure, aiming at verifying that the forecasted losses are in line with the actual losses. Based on the backtesting results, the performance of any risk metric can be evaluated.

6.1 VaR backtesting

This section presents the five validation tests chosen for the VaR backtesting. First, validation tests of the VaR are usually based on exceedences count (see Jorion (1997)[67]). An exceedence occurs whenever the actual losses are above the forecasted ones or, equivalently, when the put option identified by α expires in the money. The total number of exceedences at a specified α level

[21] A long put option may be interpreted as buying insurance.

is defined as

$$x = \sum_{t=1}^{N} \mathrm{I}_{\{VaR_{\alpha,t,\tau} < L_{t,\tau}\}} \tag{9}$$

where $L_{t,\tau}$ is the realized loss in the time interval $[t, t+\tau]$. Since the power of exceedences-based tests is low, decreasing the VaR confidence level helps in reducing the probability of type II error[22]. To improve the power of the VaR backtesting, the exceedences-count test is performed both at high (99% and 95%), medium (90%) and low (85%) confidence levels.[23].

Second, following McNeil et al. (2015)[81], the VaR is backtested by looking at the process of realized exceedences, $(\mathrm{I}_t)_{t\in\mathbb{N}}$. At each risk level α, the process $\mathrm{I}_{\{VaR_{\alpha,t,\tau} < L_{t,\tau}\}}$ is such that its conditional expectation up to the information at time (t-1), $\mathbb{E}(\mathrm{I}_t|\mathcal{F}_{t-1})$, is equal to the corresponding risk level α. This condition can be restated in terms of the calibration function

$$\mathbb{E}(\mathrm{I}_{\{VaR_{\alpha,t,\tau} < L_t\}} - \alpha|\mathcal{F}_{t-1}) = 0 \tag{10}$$

Once the realized amount of exceedences is computed, $\hat{\mathrm{I}}_t$, the martingale difference property of equation (10) implies that the number of exceedences follows a Binomial distribution - $Bin(N, \alpha)$ - where N is the total number of observations. Following Christoffersen (1998b)[30], the coverage test (CT) on the binomial behaviour can be formulated as a standard likelihood ratio test

$$LR_{CT} = -2\log\left[\frac{(1-\alpha)^{N-x}\alpha^x}{(1-\frac{x}{N})^{N-x}}\frac{x}{N^x}\right] \sim \chi(s-1) = \chi(1) \tag{11}$$

where x is the total number of exceedences as defined in eq. (9) and s is the number of possible outcomes (i.e. 2). Being a two-sided test, rejection of the null implies that the VaR either underestimates or overestimates risk. However, as this test considers the total amount of exceedences only, it has no power in detecting the existence of periods during which the exceedences cluster in a time-dependent fashion. Third, following Christoffersen (1998b)[30], a likelihood ratio test is implemented to test for the time independence of the exceedences. The resulting test statistic is

$$LR_{IT} = -2\log\left[\frac{(1-\pi)^{(n_{00}+n_{10})}\pi^{(n_{01}+n_{11})}}{(1-\pi_0)^{n_{00}}\pi_0^{n_{01}}(1-\pi_1)^{n_{10}}\pi_1^{n_{11}}}\right] \sim \chi((s-1)^2) = \chi(1) \tag{12}$$

[22]The type II error refers to the probability of accepting an incorrect model

[23]These confidence levels correspond to $\alpha = 1\%, 5\%, 10\%$ and 15%, respectively.

where n_{00} (n_{01}) is the number of periods with no exceedences followed by a period with no (one) exceedence, n_{10} (n_{11}) is the number of periods with an exceedence followed by a period with no (one) exceedence, π_0 (π_1) is the conditional probability of observing an exceedence at date t given that no (one) exceedence occurred at date (t-1) and π is the probability of observing an exceedence at date t. Being independent on the true coverage, the independence test (IT) provides useful insights into the interval forecasts dynamics without considering the true error distribution. Fourth, to gain a complete view, the coverage and the independence tests can be implemented jointly. In the resulting conditional coverage test (CCT), the null of unconditional coverage is tested against the alternative of the independence test. The test statistic is directly connected to the CT (eq. (11)) and the IT (eq. (12)), as

$$LR_{CCT} = LR_{CT} + LR_{IT} \sim \chi(s) = \chi(2) \tag{13}$$

Fifth, to go beyond the backtesting of the specific α-quantile which a VaR measure purports to calculate, a distribution test is performed in order to assess the accuracy of the predicted loss distribution as a whole. Since losses are likely to be dependent over time, loss data are transformed into quantile data by means of the random variable $u_t = \Phi_{L_t}(L_t)$, which identifies the quantile at which each realized loss occurs[24]. Conditional on the independence of $u_1, u_2, ..., u_T$, under the null hypothesis the quantile loss data are uniformly distributed in the interval (0,1). Following Berkowitz (2001)[18], an inverse normal transformation is applied to u_t, i.e. $n_t = \Phi^{-1}(u_t)$, in order to obtain a sample following a Normal distribution under the null hypothesis. Normality is tested by performing the Anderson-Darling (1954)[4] test at 95% confidence level. Rejection of the null suggests that the VaR loss distribution is not correctly specified.

6.2 CVaR backtesting

To date, there does not exist an agreed way to backtest the CVaR. The majority of the validation methods focus on the size of the losses beyond the VaR, conditional on the occurrence of a VaR exceedence. In this research three conditional tests on the CVaR and a joint VaR-CVaR backtest are implemented. First, following Acerbi and Szekely (2014)[1], the CVaR is directly tested using

[24]In the transormation Φ is the cumulative distribution function of the losses according to the chosen risk model.

the Z-statistic[25], $Z(\overrightarrow{X})$, defined as

$$Z(\overrightarrow{X}) = \sum_{t=1}^{N} \frac{X_t \mathrm{I}_t}{N\alpha CVaR_{\alpha,t,\tau}} + 1 \tag{14}$$

where X_t is the realized loss at time t. This test is sensitive to both the frequency and the magnitude of the exceedences and is easy and fast to implement, as it does not involve the use of Monte Carlo simulations. Under the null hypothesis of a well calibrated model, the expected value of Z is zero. Conversely, if the alternative is true the expected value of Z is negative and the model underestimates risk. Testing the unilateral alternative only, this test does not discard models overestimating risk, which involve costs due to the inefficient allocation of the resources. The statistics Z has a high stability of the significant thresholds across a wide choice of tail distributions (see Acerbi and Szekely (2015)[2]), hence the cut-off value $-z_\alpha$ at 99.9% confidence level can be considered fixed at -1.8[26]. Second, the CVaR is backtested by computing the average excess loss beyond the VaR conditional on a VaR exceedance, $\frac{1}{N}\sum_{t=1}^{N}[\mathrm{CVaR}_{\alpha,t,\tau} - \mathrm{VaR}_{\alpha,t,\tau}]I_t$, The choice to perform a conditional CVaR test has the objective of removing the effect of the variation and the estimation error in the VaR through time when computing the expected exceedence. The average conditional expected and realized excess losses are compared by conducting a nonparametric Mann-Whitney U test[27] on the difference between the two means. Being a two-sided test, the Mann-Whitney U test enables one to disentangle both a risk under/overestimation of the CVaR. If the null hypothesis of a zero difference between the two means cannot be rejected, there is no evidence to reject the null of the CVaR as a correct predictor of future losses. Third, following McNeil et al. (2015)[81], the CVaR can be tested by defining the calibration function

$$K_t = h^{(1)}(VaR_{\alpha,t,\tau}, CVaR_{\alpha,t,\tau}, L_{t,\tau}) = \left(\frac{L_{t,t+\tau} - CVaR_{\alpha,t,\tau}}{CVaR_{\alpha,t,\tau}}\right) \cdot I_t \tag{15}$$

[25]See eq.(5) pp.3 of Acerbi and Szekely (2014)[1]

[26]Since the null states the expected value of Z should be zero, the value corresponding to the location zero (2^{nd} column) is considered.

[27]See Mann and Whitney (1947)[78].

Since $(K_t)_{t\in\mathbb{N}}$ is a martingale difference, it is possible to test the null of a CVaR correctly predicting risk against the alternative of a risk underestimation

$$\begin{cases} \mathrm{H}_0 : \mathbb{E}(\hat{K}_t|\mathcal{F}_{t-1}) = 0 \\ \mathrm{H}_1 : \mathbb{E}(\hat{K}_t|\mathcal{F}_{t-1}) > 0 \end{cases} \tag{16}$$

where $\hat{K}$ is the calibration function obtained from the estimated VaR and CVaR. Following Efron and Tibisharani (1994)[43], a bootstrap hypothesis test is performed to test the mean-zero hypothesis for $\hat{K}_t$. Fourth, since the VaR and CVaR are jointly elicitable (Lambert et al.(2008)[74]), McNeil et al. (2015)[81] define the following calibration function

$$S_t = h_2^{(\alpha)}(VaR_{\alpha,t,\tau}, CVaR_{\alpha,t,\tau}, L_{t,t+\tau}) = \frac{\mathrm{I}_t}{CVaR_{\alpha,t,\tau}} - \alpha \tag{17}$$

to test whether the VaR and the CVaR jointly predict risk in a correct way. The realizations of the process, $(\hat{S}_t)_{t\in\mathbb{N}}$, are expected to behave as a series of martingale difference sequences. By formulating a test as in eq. (16), the alternative of a joint VaR-CVaR risk underestimation can be tested.

6.3 Backtesting Results

This section presents the backtesting results for all the VaR and CVaR estimates derived in sections 5 and 5. Following the procedure presented in sections 6.1 and 6.2, tables 4 and 5 report summary statistics on the backtesting results of the hybrid risk forecasts at a monthly and 10-days horizon, respectively. On a monthly horizon, at each risk level the proportion of exceedences is close to its

	Hybrid monthly VaR and CVaR backtesting results - $\tau = 1$ month (2012-2016)							
	# I_t (%)	ρ_{CT}	ρ_{IT}	ρ_{CCT}	Acerbi Z-stat.	ρ_{MW}	ρ_K	ρ_S
$\alpha = 1\%$	1.67	0.6161	0.8501	0.8663	0.5336	1.0000	0.7420	1.0000
$\alpha = 5\%$	5.00	1.0000	1.0000	1.0000	1.1639	0.1000	0.9191	1.0000
$\alpha = 10\%$	11.67	0.6160	0.1525	0.3169	0.7266	0.2345	0.9901	1.0000
$\alpha = 15\%$	16.67	0.6498	0.2657	0.4856	0.5396	0.5543	0.9977	1.0000

Table 4: Monthly hybrid CVaR and VaR backtesting summary statistics at $\alpha = 1\%, 5\%, 10\%$ and 15% when $\tau = 1$ month and over the sample period from January 2012 to December 2016.

theoretical value. This conclusion is confirmed by the p-values of the coverage, independence and

conditional coverage test , which, being always larger than 5%, allow to reject the null hypothesis at 95% confidence level. When backtesting the entire loss distribution, the p-value of the Anderson-Darling test equals 0.0078, thus meaning that the null of a correctly specified loss distribution under the hybrid approach is rejected at 99% confidence level. As concerns the CVaR, at each risk level the value of the Z-statistic is positive, thus indicating that the hybrid CVaR correctly forecasts losses at 99.9% confidence level. The p-values of the Mann-Whitney U test indicate that the predicted and realized excess losses are not significantly different from zero in all cases at 95% confidence level, thus meaning that the hybrid CVaR does not involve neither costs linked to risk underestimation nor to risk overestimation. Testing the underestimation alternative only, the p-values of the bootstrap hypothesis tests on $\hat{K}_t$ and $\hat{S}_t$ are always close to 1, thus indicating the null hypothesis can never be rejected at 95% confidence level. As a consequence, the CVaR and the VaR and CVaR joint forecasts do not show evidence of being affected by negative biases. The forecasting power of the hybrid VaR is satisfactory also when the time horizon of the estimates is shortened. At 10 days investment horizon the proportion of exceedences is close to its theoretical

	Hybrid VaR - τ=10 days (2012-2016)			
	# I_t (%)	ρ_{CT}	ρ_{IT}	ρ_{CCT}
$\alpha = 1\%$	0.80	0.8972	0.7745	0.8233
$\alpha = 5\%$	5.69	0.9972	0.8513	0.8875
$\alpha = 10\%$	7.31	0.7642	0.7865	0.7774
$\alpha = 15\%$	8.13	0.0201	0.6546	0.4533

Table 5: 10-days VaR backtesting summary statistics under the hybrid methodology at $\alpha = 1\%, 5\%, 10\%$ and 15% when $\tau = 10$ days and over the sample period from January 2012 to December 2016.

value, except when $\alpha = 15\%$. Nevertheless, being always larger than 5%, the p-values of the coverage, independence and conditional coverage tests suggest that the hybrid VaR delivers accurate forecasts of the potential losses across all risk levels.

The same backtesting analysis is implemented on the monthly B&S VaR and CVaR over the sample period from January 2012 to December 2016. The results are reported in table 6. As confirmed by the results of the coverage test, the proportion of VaR exceedences is never significantly different from the corresponding theoretical values. However, the p-values of the independence test are close to zero, thus meaning that the VaR exceedences cluster over time in a time-dependent fashion. As a consequence, the null hypothesis of joint independence and correct coverage of the exceedences is always rejected at 95% confidence level, but at $\alpha = 15\%$. When backtesting the

	B&S monthly VaR and CVaR backtesting results - $\tau = 1$ month (2012-2016)							
	# I_t (%)	ρ_{CT}	ρ_{IT}	ρ_{CCT}	Acerbi Z-stat.	ρ_{MW}	ρ_K	ρ_S
$\alpha = 1\%$	3.33	0.1523	0.0042	0.0060	-0.7690	0.6667	0.8705	0.1322
$\alpha = 5\%$	6.67	0.5721	0.0122	0.0368	-1.2542	0.6857	0.5205	0.2961
$\alpha = 10\%$	8.33	0.6586	0.0145	0.0457	-0.2714	0.1508	0.0972	0.4595
$\alpha = 15\%$	13.33	0.7134	0.0457	0.1270	-0.2580	0.1304	0.0817	0.3618

Table 6: Left tail - Monthly B&S CVaR and VaR backtesting summary statistics at $\alpha = 1\%, 5\%, 10\%$ and 15% when $\tau = 1$ month and over the sample period from January 2012 to December 2016.

entire loss distribution, the p-value of the Anderson-Darling test equals 0.0064; hence the null of a correct VaR loss distribution is rejected at 95% confidence levels. This result can be a consequence of the restrictive assumptions underneath the B&S approach (constant volatility and interest rates), which reduce the flexibility of the model and worsen the accuracy of the forecasts. At each risk level, the value of the Z-statistic is above -1.8, hence the CVaR forecasts cannot be rejected at 99.9% confidence level. This result is robust to the nonparametric two-sided Mann-Whitney U test: the p-values are always larger than 5%, hence the null of a correctly calibrated CVaR can never be rejected. Both tests proposed by McNeil et al. (2015)[81] suggest that the $\hat{K}_t$ and $\hat{S}_t$ behave as martingale differences; hence the CVaR alone and the VaR and CVaR jointly do not show evidence of risk underestimation.

Table 7 reports the backtesting results on the monthly and 10-days FHS classical VaR. On

	FHS method - $\tau = 1$ month (2012-2016)				FHS method - $\tau = 10$ days (2012-2016)			
	# I_t (%)	ρ_{CT}	ρ_{IT}	ρ_{CCT}	# I_t (%)	ρ_{CT}	ρ_{IT}	ρ_{CCT}
$\alpha = 1\%$	1.67	0.6357	0.8527	0.8786	6.50	0.0020	0.0001	0.0001
$\alpha = 5\%$	5.00	1.0000	0.5707	0.8515	13.82	0.0000	0.0102	0.0003
$\alpha = 10\%$	15.00	0.2256	0.6975	0.4450	17.88	0.0014	0.0625	0.0010
$\alpha = 15\%$	21.67	0.1700	0.4993	0.3104	21.13	0.0031	0.0678	0.0022

Table 7: Monthly classical VaR backtesting summary statistics under the classical GARCH(1,1)-FHS methodology at $\alpha = 1\%, 5\%, 10\%$ and 15% when $\tau = 1$ month (1^{st}-4^{th} columns) and 10 days (5^{th}-8^{th} columns) and over the sample period from January 2012 to December 2016.

a monthly horizon the discrepancy between the realized proportion of the exceedences and the theoretical value is small at $\alpha = 1\%$ and 5%, but substantially widens as the risk level increases. Nevertheless, at 95% confidence level the coverage, independence and conditional coverage tests do not detect any bias in the classical VaR estimation, as the p-values are always larger than 5% across all risk levels. Conversely, as the investment horizon shortens, the classical FHS VaR

estimates suffer from severe biases. At ten days horizon the proportion of exceedences is always well above the theoretical risk level , thus suggesting that the FHS VaR tends to underestimate risk. The p-values of the coverage and the independence tests indicate that the null of correct coverage and of time independence of the exceedences is always rejected. The unsatisfactory performance of the FHS VaR in terms of coverage and independence is reflected in the conditional coverage test, whose null is always rejected at 95% confidence level.

7 Model Comparison

The ranking of several methodologies involves a problem of model selection. Following Bayer (2017)[16], the performance of each methodology is ranked based on the quantile regression (QR) approach, which assigns the best ranking to the model minimizing the tick loss function[28]. The VaR tick loss function is defined as

$$\rho_\alpha(u) = (\alpha - I_{\{u \leq 0\}})u \qquad \text{where} \qquad u_{t+\tau} = r_{t+\tau} - \beta_{0,t}(\alpha) - \boldsymbol{VaR_{t+\tau|t}}(\alpha)\boldsymbol{\beta_t(\alpha)}$$

where $r_{t+\tau}$ are the realized losses over the horizon τ, $\beta_{0,t}(\alpha)$ is an intercept term introduced to perform a bias correction[29] and $\boldsymbol{VaR_{t+\tau|t}}$ are the set of VaR estimates of the models to be compared. The consistency of the loss function ensures that the optimal forecast combination weights minimize the expected loss of the forecast error,

$$(\hat{\beta}_{0,t}(\alpha), \hat{\boldsymbol{\beta}}_{\boldsymbol{t}}(\boldsymbol{\alpha})) = \underset{\beta_{0,t}(\alpha), \boldsymbol{\beta_t(\alpha)}}{argmin} E[\rho_\alpha(r_{t+\tau} - \beta_{0,t}(\alpha) - \boldsymbol{VaR_{t+\tau|t}(\alpha)\beta_t(\alpha)})|\mathcal{F}_t]$$

The relative ranking among the methodologies is based on the optimal weights $\boldsymbol{\beta_t(\alpha)}$. The largest is the absolute value of the optimal weight, the highest is the ranking of a model. The QR approach is implemented on both the monthly and the 10-days risk forecasts. Results show that the B&S and the hybrid methodologies perform similarly and both outperform the classical FHS VaR in more than 95% of the cases. This conclusion holds also when conducting a pairwise QR ranking between classical FHS VaR and any option-implied (hybrid and B&S) VaR. To quantify the costs

[28]Notice that, since the CVaR is not elicitable, a consistent loss function can not be defined for this risk metric.

[29]This term is important when all methodologies taken singularly systematically underestimate or overestimate risk.

associated to each methodology, the unpredicted losses, $L^*_{j,\alpha,\tau}$, are computed as

$$L^*_{j,t,\alpha,\tau} = [L_{R,t,\tau} - \text{VaR}_{j,\alpha,\tau}] \cdot I_t$$

where j identifies the model and $L_{R,t,\tau}$ is the realized loss over the investment horizon τ. Tables 8 and 9 report the minimum, maximum, mean and median of the unpredicted \$ losses associated to one futures contract at one month and 10 days horizons across all the proposed methodologies, respectively. On a monthly basis, the ranking between the B&S and the hybrid VaR often reverts, thus meaning that the two estimates perform similarly.

	Option-implied B&S L^*				Hybrid L^*				Classical FHS L^*			
	Min	Max	Mean	Median	Min	Max	Mean	Median	Min	Max	Mean	Median
$\alpha = 1\%$	0.06	0.18	0.07	0.07	0.11	0.18	0.14	0.13	1.62	1.62	1.62	1.62
$\alpha = 5\%$	0.06	0.08	0.07	0.07	0.07	0.47	0.18	0.26	1.22	1.36	1.30	1.32
$\alpha = 10\%$	0.04	0.31	0.16	0.16	0.01	0.76	0.24	0.06	0.55	1.46	1.02	0.97
$\alpha = 15\%$	0.03	0.46	0.20	0.19	0.04	0.78	0.23	0.13	0.49	1.35	0.85	0.85

Table 8: Summary statistics on the monthly unpredicted losses, L^*, defined as $L^* = L_{R,\tau} - VaR_{j,\tau}$, where $L_{R,\tau}$ is the realized loss over the time horizon τ (i.e. one month) and $VaR_{j,\tau}$ is the forecasted one-month VaR of model j (i.e. B&S, hybrid and classical FHS) over the sample period from January 2012 to December 2016.

Hence, in general, the hybrid approach achieves the best trade-off between availability and accuracy. The hybrid methodology always outperforms the classical FHS VaR. Under the classical FHS methodology the average unpredicted \$ losses are from almost 50% up to 300% higher than those under the hybrid approach. At 10 days horizon the gap between the hybrid and the classical FHS VaR narrows; nevertheless, the average unpredicted losses under the latter approach still are from 5% up to 25% higher compared to the hybrid methodology.

	Hybrid L^*				Classical FHS L^*			
	Min	Max	Mean	Median	Min	Max	Mean	Median
$\alpha = 1\%$	0.16	0.16	0.16	0.16	0.03	0.32	0.18	0.19
$\alpha = 5\%$	0.02	0.42	0.20	0.18	0.01	0.47	0.21	0.22
$\alpha = 10\%$	0.01	0.48	0.19	0.19	0.02	0.59	0.25	0.20
$\alpha = 15\%$	0.01	0.51	0.20	0.17	0.02	0.66	0.27	0.20

Table 9: Summary statistics on the 10-days unpredicted losses, L^*, defined as $L^* = L_{R,\tau} - VaR_{j,\tau}$, where $L_{R,\tau}$ is the realized loss over the time horizon $\tau = 10$ days and $VaR_{j,\tau}$ is the forecasted 10-days VaR of model j (i.e. hybrid and classical) over the sample period from January 2012 to December 2016.

Due to the high unforeseen costs associated to the classical FHS VaR, this approach is inferior to the hybrid methodology and, more generally, fails to provide an adequate capital protection.

8 Conclusions

The high volatility and the pronounced non-normal features are just few aspects which pose several modelling challenges when performing a risk forecasting exercise in energy markets. The article presents a new methodology to derive point risk metrics for the natural gas market. Exploiting the informative content of derivatives (i.e. futures and options), the hybrid VaR and CVaR estimates are not bounded to the availability of a liquid option market and are reliable predictors of the future losses, even during prolonged periods of turmoil. Indeed, the proposed risk metrics can be implemented on a regular basis, at several time horizons/frequencies and to any underlying asset with a small number of traded options. Compared to the classical FHS VaR, the hybrid risk forecasts always obtain a superior ranking and involve unforeseen costs about three times lower in average terms.

Part II

A Bayesian Estimate of the Pricing Kernel

1 Introduction

The Pricing Kernel is an asset-independent random quantity which summarizes preferences of an investor about risk and allows to price any security traded in the market. Under regularity conditions[30], a unique PK exists and equals the discounted value of the ratio between the risk neutral and the physical densities. According to the economic theory, the PK should be a monotonic decreasing function of the asset returns. However, empirical evidence shows that the PK is often locally increasing around the 0% return interval (see Jackwerth (2000)[65]). Documented for the first time by Bates (1996)[15], this phenomenon, known as "Pricing Kernel puzzle", implies that the return of a contingent claim written on the market portfolio stochastically dominates the return of the market portfolio itself[31].

To date, the majority of academics agrees that market incompleteness, deviations from risk aversion in investors behaviour, misestimation of the investors and/or model misspecification errors are the three major drivers of the PK puzzle. The presence of at least one of these conditions is sufficient to break down the monotonicity requirement. First, market incompleteness stems from the presence of background risks, market frictions and/or differences in traded option volumes due to asymmetric information[32]. When markets are incomplete, agents may choose different portfolio allocations, which correspond to distinct pricing kernels. In this context the representative agent does not exist anymore and the PK is no longer unique, thus being potentially increasing in the complete subspace. Second, even though investors are risk averse in general, on some return intervals they show the tendency to gamble (see Kahneman and Tversky (1979)[69]). As a consequence, the utility function becomes convex, thus allowing in principle for an increasing PK[33]. Third, investors

[30] Absence of arbitrage and market completeness.

[31] There exists a portfolio of Arrow securities delivering the same return of the market portfolio at a lower price. This empirical fact contradicts the basic principle in economics and finance according to which payoffs in good states should be higher than payoffs in bad states of the world.

[32] While there exists a large number of liquid at-the-money options, there are much fewer options for extreme strikes. Moreover, the liquidity of out-the-money (OTM) options is asymmetric (see Dennis (2000)[38]), thus increasing the steepness of the option volatility smile.

[33] Nevertheless, many studies argue that local convexity does not influence the utility maximization problem (see Levy and Levy (2002)[76]). As an alternative, state-dependent utility functions can be used (see for instance Kreps and Porteus (1978)[72] and Gul and Pesendorfer (2006)[60]).

hold incorrect beliefs when the subjective probability weights are different from those estimated under the objective probability measure. The mismatching is usually originated by misestimation and/or distortion (i.e. heterogeneous beliefs, framing or other behavioural biases). Misestimation arises when either the agent and/or the statistician wrongly assign probabilities to the events (see Chabi-Yo et al. (2007)[28]). Nevertheless, biased or distorted beliefs of the agents in isolation are in general insufficient to justify an increasing PK (see Hens and Reichlin (2013)[64]). Conversely, model misspecification can be a major driver of the PK puzzle. Recently Leisen (2014)[75] has shown how the PK puzzle can be originated by small sample biases in the physical and risk neutral density estimation. Even thought the consequences of estimation errors of the agent and the econometrician are similar, the latter ones can not be easily detected. Kan and Zhou (1999)[70] show indeed that one of the major weaknesses of the stochastic discount factor methodology consists in its low power in identifying misspecified models.

In the context of model misspecification, this article focuses on the discrepancy between the information set used in practice to derive the physical density, $\mathcal{H}_t$, and the information set theoretically required, Ψ_t. By definition, both the physical and the risk neutral densities should be conditional to all the available information at a specified time t, Ψ_t. In practice, a complete information set is difficult, if not impossible, to achieve. The attempt to reduce the information bias and use information sets as complete as possible in the estimation process of both densities is the starting point of this research. In general, while the risk neutral density is well extracted from option data, the estimation of the conditional physical density is often rather imprecise[34]. To date, the majority of models rely on past returns/prices, which do not include a significant portion of the (required) information based on which the physical density should be extracted . Compared to historical data, derivative securities are more informative about the underlying asset dynamics (Bollerslev and Todorov (2011)[23]), as they naturally embed investors' expectations about future market scenarios. As a consequence, the combined use of derivative and historical data as inputs in the physical density estimation enables one to obtain an information set closer to the one theoretically required. To date, few authors have tried to solve the information bias arising in the PK estimation. Among them, Ross (2015)[91] with the Recovery Theorem has proposed a methodology to directly extract the physical density from the risk neutral one. However, this approach requires the transition-independence of the PK[35], a condition rarely fulfilled in real world modelling. Recently, Sala et

[34]Indeed, the use of historical data only makes this density almost unconditional.

[35]Transition independence implies that the PK depends on the marginal rate of substitution between future and current consumption only.

al.(2017)[92] have estimated the physical density by using a Bayesian Dirichlet process to combine the information coming from the option market and the historical data. Besides the inability of the Dirichlet process to capture important characteristics of financial data, the authors impose several assumptions on the shape of the posterior distribution which undermine the flexibility and the accuracy of the methodology.

Using a similar approach as starting point, the proposed methodology derives the physical density using jointly option and historical data as inputs. The weights associated to each information source are modelled with a Bayesian nonparametric Poisson Dirichlet process. The article is interesting both at a theoretical and practical level. From a modeling perspective, the proposed methodology sheds light on the advantages of setting the PK estimation problem in a Bayesian framework. From a practical point of view, given the key role of the PK in finance theory and its properties, estimates of the PK are valuable for asset pricing, as well as for risk and portfolio management purposes. The contribution of the article to the existing literature is fourfold. First, it proposes a flexible and accurate methodology to estimate the conditional physical density, modeled as the posterior distribution of a Poisson-Dirichlet process. Since the Dirichlet process is a special case of the Poisson Dirichlet one, the proposed approach is a generalization of the methodology used in Sala et al. (2017)[92]. The PD process allows one to capture a wide range of tail behaviours, whereas the Dirichlet one allows for tails decaying exponentially only, a pattern not generally followed by financial data. As a consequence, the resulting estimates behave in a more regular way in the tails and respect the monotonicity requirement also for extreme positive/negative returns. This improvement is crucial especially during high volatility periods and at short time horizons. Second, this is the first article which proposes a fully Bayesian approach at each step of the PK estimation process. In the first phase the simulated asset prices under the physical measure are partitioned into clusters before deriving the Bayesian physical density without imposing any assumption neither on the PD posterior distribution, nor on the structure of the data. In the second stage, at each observation date and time to maturity the PD parameters are estimated using a Bayesian Markov Chain Monte Carlo (MCMC) technique. Third, the derivation of time-varying estimates of the PD process parameters allows one to formulate predictions about the relationships among the model parameters, the time to maturity and the market volatility. Fourth, the existence of a monotonic relation between the PK and the market returns is formally tested by using the Monotonic Relation (MR) test developed by Patton and Timmermann (2010)[88]. The results validate the predictions: in almost all cases the PK is monotonic decreasing in the returns, even at investment horizons

shorter than one month.

The remainder of the article is as follows: section 2 presents the methodology - section 2.1 describes the estimation procedure of the classical physical and the risk neutral densities and section 2.2 introduces the Bayesian approach and motivates the choice of a Poisson-Dirichlet process for modelling the physical density. Section 3 describes the dataset and its relevant characteristics. Section 4 presents and compares the estimates of the Poisson-Dirichlet parameters using the Maximum Likelihood and the Bayesian approaches. Section 5 presents the PK estimates on a day-specific basis and over the whole sample and the results of the MR test on all the obtained estimates. As a robustness check, section 6 repeats the same analysis based on a more recent sample. Section 7 concludes.

2 The Methodology

If the market is arbitrage-free and complete, the PK equals the present value of the ratio of two conditional densities. Starting from this definition, the proposed Bayesian PK, $\mathbb{M}^{\dagger}$, equals

$$\mathbb{M}^{\dagger}_{t,T} = e^{-r_t(T-t)} \left[\frac{q_{t,T}(S_T|S_t, \Psi_t)}{p^*_{t,T}(S_T|S_t, \Psi_t)} \right] \tag{18}$$

where r_t is the risk-free rate at time t, S_t is the value of the market portfolio at time t, $q_{t,T}(S_T|S_t, \Psi_t)$ and $p^*_{t,T}(S_T|S_t, \Psi_t)$ are the risk neutral and the Bayesian physical densities at time t over the time horizon $(T-t)$ and conditional on Ψ_t, the available information at time t. To simplify the notation, henceforth the conditional risk neutral and physical densities are denoted with q and p^*, respectively. By definition the set Ψ_t should contain *all* the available information at time t. As anticipated in section 1, at a practical level investors are generally unable to obtain the complete information set Ψ_t with the data at their disposal. As a consequence, the set of "achievable" information, $\mathcal{H}_t$, set is strictly included in Ψ_t, i.e. there exists an information bias. This article presents a methodology to reduce the entity and the severity of the information bias. Exploiting already the superior informative content of derivative securities, the risk neutral density is already conditional on an information set which can be hardly improved in terms of completeness. Conversely, using only historical data as inputs, the physical density estimation strongly suffers from the identified

information bias[36]. Being more informative about the underlying asset dynamics than historical records, option data can be helpful in reducing the gap between $\mathcal{H}_t$ and Ψ_t. Mathematically, the combination of option and past data information is performed by modelling the physical density as the posterior distribution of a PD process with q^*[37] as base distribution and p (the physical density extracted from historical data) as empirical density. The derivation of $\mathbb{M}^\dagger$ follows a two-steps procedure. First, the risk neutral, q, and the physical, p, densities are derived from the cross-section of OTM options and from the time series of the historical returns, respectively. In the rest of the paper, p and q are labelled as "classical densities". Second, the modified physical density, p^*, is modelled as the posterior of a PD process. Section 2.1 describes the estimation methodology for the classical densities, p and q, and section 2.2 motivates the choice of the Poisson-Dirichlet process for modelling p^*.

2.1 The classical densities estimation

The physical and the risk neutral densities are estimated following the methodology proposed in Barone et al.(2008)[14]. At each observation date t, the physical density p is estimated by fitting a Glosten Jagannathan and Runkle (GJR)-GARCH (see Glosten et al. (1993)[54]) model to the historical daily log-returns of the underlying asset. Calibrated on a pre-specified rolling window, at each observation date t the model under the physical measure is:

$$log\frac{S_t}{S_{t-1}} = r_t = \mu + \epsilon_t \tag{19}$$

$$\epsilon_t = \sqrt{\sigma_t^2} z_t \tag{20}$$

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \mathbb{I}_{\epsilon_{t-1}<0}\gamma\epsilon_{t-1}^2 \tag{21}$$

where μ is the drift term, ϵ_t is the return innovation, z_t is the scaled return innovation and $\mathbb{I}_{\epsilon_{t-1}<0}$ is the indicator function controlling for the leverage effect. The use of a GJR-GARCH allows one to capture the asymmetric impact of the negative shocks: when the innovation in the previous time period is negative (i.e. $\epsilon_{t-1} < 0$) the conditional volatility increases more than when a positive innovation of the same absolute value occurs. The estimates of the GARCH parameters

[36]Examples of articles using this estimation approach for p and q are those by Jackwerth (2000)[65], Aït Sahalia and Lo (2000)[3], Rosenberg and Engle (2002)[46], Christoffersen et al. (2013)[31].
[37]Compared to q, q^* differs only in the drift term, which is rescaled for the equity premium (details follow in section 2.1).

$\hat{\theta}_t = (\hat{\omega}_t, \hat{\beta}_t, \hat{\alpha}_t, \hat{\gamma}_t)$ are obtained by maximizing the Gaussian pseudo maximum likelihood (PML) under the assumption of normal innovations. Importantly, even though the true distribution of the innovations is not normal, the results still remain consistent (see Gourieroux et al. (1984)[58]). Finally, the physical density p is estimated by smoothing the empirical kernel density obtained from the simulated prices.

As concerns q, at each observation date t the GJR-GARCH model in equations (19-21) is fitted to the cross-section of the OTM European call and put options written on the underlying asset. Based on the risk neutral GARCH parameters, $\theta_t^* = (\omega_t^*, \beta_t^*, \alpha_t^*, \gamma_t^*)$, $N = 10.000$ return paths are simulated for each holding period,

$$\frac{S_T}{S_t} = \exp\left((T-t)\mu^* + \sum_{i=1}^{T-t} \sigma_{t+i} z_i\right) \tag{22}$$

where z_i is the past innovation updating the volatility in the next period, σ_{t+i}. In order to exclude arbitrage violations, the simulated paths are corrected by means of the Empirical Martingale Simulation (EMS) method by Duan and Simonato (1998)[41], which ensures that μ^* equals the risk free rate. In this way, for a given time to maturity τ and strike K, the GARCH call option price equals the present value[38] of the average simulated payoffs, i.e. $e^{-r_f(T-t)} \sum_{n=1}^{N} \frac{\max(S_T^n - K, 0)}{N}$. The derivation of put option prices follows a similar procedure. The estimated GARCH parameters $\theta_Q^* = (\omega^*, \beta^*, \alpha^*, \gamma^*)$ minimize the mean squared pricing error

$$\frac{1}{N} \sum_{i=1}^{D_t} \psi_t(K_i, \tau_i) \tag{23}$$

where D_t is the number of options on date t, $\psi_t(K_i, \tau_i)$ is the difference between the simulated GARCH option prices and the observed market prices of the OTM options across strikes K_i and times to maturity τ_i. At each observation date t, the derivation of q is performed by running $N = 50.000$ Monte Carlo simulations and then implementing a nonparametric kernel density estimation.

[38] All the payoffs are discounted at the risk-free rate.

2.2 The Bayesian approach

In this analysis, the Bayesian structure is attractive because of its high degree of flexibility, ease of application[39] and suitability in representing the economic intuition behind the proposed model. The Poisson-Dirichlet process (see Pitman and Yor (1997)[90]) is a stochastic process whose sample path is a infinite discrete probability distribution, where all the atoms are drawn from the base distribution G and the weights follow a Poisson-Dirichlet distribution. It includes a wide range of distributions and enjoys useful properties which make it computationally easy to handle.

First, the process is fully characterized by three parameters only: the base distribution, G, the discount parameter, d, and the strength (or concentration) parameter, α. The base distribution is directly linked to the first moment and represents the mean of the process. The discount parameter controls the tail behaviour and its support ranges in the interval $[0, 1)$. When d is equal to 0 the PD process collapses to the Dirichlet process, according to which tails can only decay exponentially. Increasing values of d in the interval $(0, 1)$ imply a higher probability of observing values in new clusters (i.e. values less close to the mean of the process). In the limit case of d converging to 1, the model produces a high number of small clusters consistently with a power law distribution. The strength parameter controls the weight associated to the base distribution and its support ranges in the interval $(-d, \infty)$. Large values of α imply distributional shapes concentrated around the mean of the process. Second, being nonparametric, the process does not impose any restriction on the functional form of the PK, thus ensuring a satisfactory degree of flexibility. Third, the process is conjugate. This feature implies that given the random distribution H following a PD process with parameters (G, α, d) and the sequence of independent draws $\theta_1, \theta_2, ..., \theta_n$ from H, the posterior distribution given the observed parameters still follows a Poisson-Dirichlet process with updated parameter (G, α^*, d), where $\alpha^* = \alpha + Nd$. The conjugacy property makes the process very easy to handle analytically, as the posterior distribution can be derived in closed form. Mathematically, the modified physical density, p^*, equals the posterior distribution of a Poisson-Dirichlet process

$$p^* = \frac{\alpha + Kd}{\alpha + N} G + \sum_{i=1}^{N} \frac{(1-d)}{\alpha + N} \delta_{X_i^*} \tag{24}$$

where α and d are the concentration and the discount parameters, G is the base distribution, N is the number of observations, K is the number of clusters and $\delta_{X_i^*}$ is the dirac mass function applied

[39]Tractability issues linked to the shape of the posterior distribution can be largely overcome by means of Markov Chain Monte Carlo (MCMC) algorithms.

to each unique simulated price X_i^*. The base distribution is the risk neutral density rescaled for the equity premium, q^*, which differs from q in the drift term only

$$d_t^{q^*} = d_t^q + \pi_t; \qquad \text{where} \qquad d_t^q = r_{f,t} - div_t$$

where d_t^q is the drift under the risk neutral measure, $r_{f,t}$ is the risk free rate, div_t is the continuously compounded dividend and π_t is the equity premium. This drift correction is required in order to rule out any arbitrage opportunity.

3 The Dataset

The methodology presented in section 2 is tested on the cross-section of the OTM European call and put options (SPX) traded on the S&P 500 with time to maturity ranging between 10 days and one year. Options data and S&P 500 prices are from OptionMetrics. Being among ones of the most traded financial assets, mispricings are likely to occur rarely and the degree of liquidity is always satisfactory. Figure 4 reports the volumes of the OTM European call and put options written on the S&P 500 index from 1998 to 2014. Option volumes have followed an increasing trend over the recent years and the amount of OTM call options has always been lower compared to the put options one. Starting from January 2002 to December 2004, the sample period consists of 157

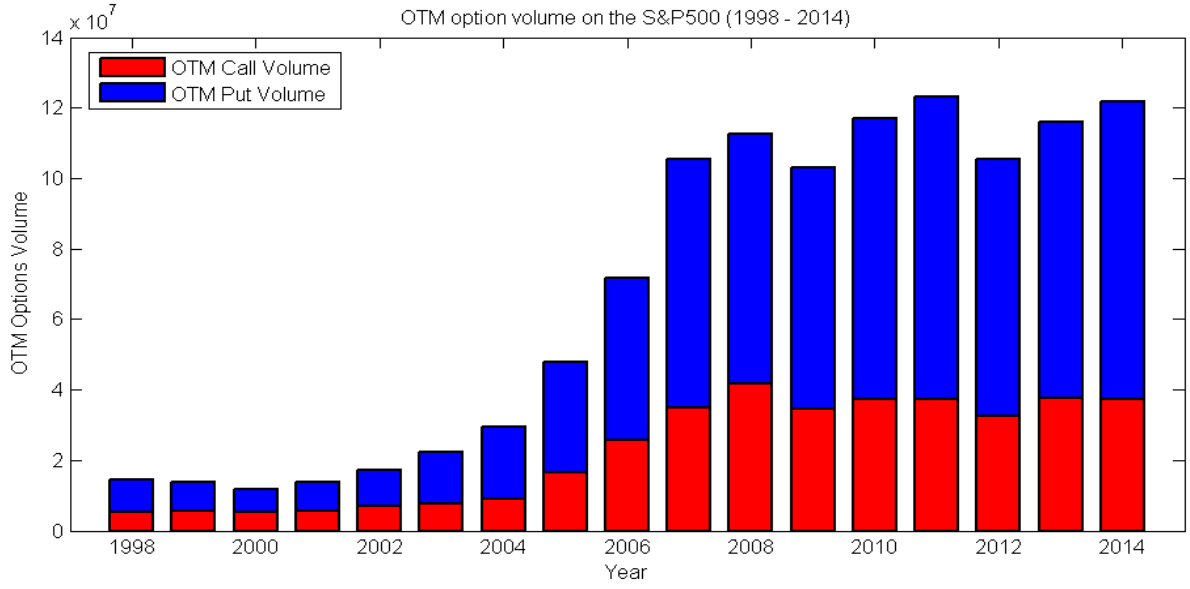

Figure 4: Evolution of OTM options volume on the S&P500 from 1998 to 2014.

weekly observations. The choice of this sample period stems from the willingness to make a direct comparison between a fully specified nonparametric Bayesian model and the hybrid[40] Bayesian approach proposed in Sala et al. (2017)[92]. The frequency is weekly in order to limit the high computational time intensity of the optimization problem, which involves a large cross-section of option contracts at each observation date. The use of Wednesday data aims at reducing both the probability of losing observation dates because of holidays and the incidence of day-of-the-week effects.To estimate the classical physical density, a GJR-GARCH is fitted on a rolling window of 14 years[41] of historical log-returns. The GJR-GARCH model provides a good fit when residuals are auto-correlated, and thus when volatility clusters, in time-series regressions. The model and the length of the log-returns time series used for the calibration are validated from the results of the Lagrange Multiplier ARCH test (Engle (1982)[48]). By interpolation, daily risk-free interest rates are obtained from the time series of US treasury zero coupon yield curve downloaded from OptionMetrics. For each time-to-maturity (τ), the previous (τ^-) and next (τ^+) time period zero-coupon whose values straddle the time τ are linearly interpolated. The time series of discrete S&P 500 dividends from January 2002 to December 2004 are from CRSP and used to compute the continuous daily dividend yield.

4 Parameters Estimation

This section presents the initial assumptions and the methodology to derive the base distribution, the concentration and the discount parameters. Differing from q in the drift term only, the estimation of the rescaled risk neutral density follows the methodology outlined in section 2.1 and also requires the estimation of the equity premium. The equity premium, π, is defined as the historical average differential return of the market portfolio (i.e. the S&P 500 index) over risk-free debt[42]. Following Merton (1980)[82], across its most common specifications in terms of risk-free security (T-bills or long term government bonds) and type of average differential return (geometric or arithmetic), the equity premium on the S&P 500 ranges between 4% and 8% and is assumed to

[40]Data are not clustered when deriving the Dirichlet posterior distribution and the strength parameter of the process is estimated empirically from option data.
[41]The choice of this time window is motivated by the intention to exclude the period of turmoil during the oil crisis in 1987.
[42]Given that the investors consider the same time interval and market index, the historical equity premium should be unique.

vary as a deterministic function of the unconditional annual GARCH long run volatility[43] under the risk neutral measure. Table 10 reports the equity premia and the starting values of the PD parameters associated to the three identified volatility regimes, describing a low ($\sigma_t \leq 17\%$), normal ($17\% < \sigma_t \leq 22\%$) and high ($\sigma_t > 22\%$) volatility environment, and used to initialize the maximum likelihood algorithm. Since investors require a higher remuneration when the uncertainty increases, in the high volatility scenario π equals its highest value.

	Low Vol. ($\sigma_{i,j} \leq 17\%$)			Med. Vol. ($17\% < \sigma_{i,j} \leq 22\%$)			High Vol. ($\sigma_{i,j} > 22\%$)		
	$\tau = 24$	$\tau = 73$	$\tau = 185$	$\tau = 24$	$\tau = 73$	$\tau = 185$	$\tau = 24$	$\tau = 73$	$\tau = 185$
π	4%	4%	4%	6%	6%	6%	8%	8%	8%
	Poisson-Dirichlet parameters								
α_0	0.75	0.75	0.75	1.5	1.5	1.5	2.25	2.25	2.25
d_0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Table 10: Starting values for the Maximum Likelihood algorithm for each volatility regime (defined as a deterministic function of the unconditional GARCH volatility $\sigma_{i,j}$ at time i using innovations $j = FHS, Gaussian$) and time to maturity (τ)).

The strength and the discount parameters are derived using both the Maximum Likelihood (ML) and the Bayesian MCMC techniques. Both methodologies start from the PD posterior distribution, defined as in eq. (24). The specification of a feasible set of starting values (ML approach) and an informative prior distribution (Bayesian method) for α and d relies on the economic interpretation of the PD parameters. Using the estimation results in Sala et al. (2017)[92] as starting point, for each time to maturity τ an initial value for the strength parameter, α_0 [44], is readily available. Following the intuition that in periods of turmoil historical data are not helpful in modeling the underlying asset future dynamics, α is further assumed to be positively correlated with the volatility level. The relationship is formalized in the model by setting increasing starting values of α from the lowest to the highest identified volatility regimes. The parameter d controls the number of clusters generated by the process and its support ranges between 0 and 1. Reflecting the choice of

[43]The long run GARCH volatility at time t equals

$$\sigma_t = \sqrt{\frac{\omega_t}{1 - \alpha_t - \beta_t - \frac{\gamma_t}{2}}}$$

Notice that σ_t and the corresponding regime identification at time t are deterministic once that the GARCH parameters have been estimated.

[44]The estimates of α are derived empirically under the assumption of a Dirichlet process driving the data distribution. Once that the time to maturity is fixed, α is assumed to be constant across all the sample period.

an uninformative prior, the starting value of d is always set equal to 0.5[45].

4.1 The Maximum Likelihood estimators

Following Carlton (1999)[26] and generalizing the two-parameter Ewens Sampling Formula to a sample of size N, the likelihood function is

$$\mathbb{L}_N(\alpha, d; x_1, ..., x_N) = \frac{N(a)}{(\alpha+1)_{N-1}} \prod_{l=1}^{k-1}(\alpha + ld) \prod_{j=2}^{n}[(1-d)_{j-1}]^{a_j} I\left(\sum_{j=1}^{N} j a_j = N\right) \tag{25}$$

where $A = (a_1, a_2, ..., a_n)$ is the counting vector[46], a_j is the number of clusters with exactly j elements inside, $N(a) = \frac{N!}{\prod_{j=1}^{N}(j!^{a_j} a_j!)}$ is the number of the sample orderings prescribed by a, $k = \sum_{j=1}^{N} a_j$ and $(x)_n$ is the forward product $\frac{\Gamma(x+N)}{\Gamma(x)}$. Taking the logarithm of eq.(25), the log-likelihood is:

$$\begin{aligned} l(\alpha, d) &= \log \mathbb{L}_N(\alpha, d; a) \\ &= \log N(a) - \log(\alpha+1)_{N-1} + \sum_{l=1}^{k-1} \log(\alpha + ld) + \sum_{j=2}^{N} a_j \log(1-d)_{j-1} \\ &= \log N(a) - \sum_{m=1}^{n-1} \log(\alpha + m) + \sum_{l=1}^{k-1} \log(\alpha + ld) + \sum_{j=2}^{N} a_j \sum_{i=1}^{j-1} \log(i - d) \end{aligned} \tag{26}$$

And the first partial derivatives with respect to d and α respectively are:

$$l_d(\alpha, d) = \sum_{l=1}^{k-1} \frac{l}{\alpha + ld} - \sum_{j=2}^{N} a_j \sum_{i=1}^{j-1} \frac{1}{i - d}$$

$$l_\alpha(\alpha, d) = -\sum_{m=1}^{n-1} \frac{1}{\alpha + m} + \sum_{l=1}^{k-1} \frac{1}{\alpha + ld}$$

[45]This value is the mean of the (uninformative) Uniform prior distribution over the support (0,1).

[46]The counting vector summarizes all the information of the sample and it is subject to the following restrictions: (1) $(a_1, a_2, ..., a_n)$ are non-negative and integer-valued and (2) $\sum_{j=1}^{N} a_j = N$.

Beyond the first order conditions, the joint derivation of the ML estimators [47], $(\hat{\alpha}, \hat{d})$, requires the determinant of the Hessian matrix, $|H|$, to be positive. The Hessian matrix, H, equals:

$$H = \begin{bmatrix} l_{dd} & l_{\alpha d} \\ l_{d\alpha} & l_{\alpha\alpha} \end{bmatrix}$$

where

$$l_{dd} = -\sum_{l=1}^{k-1} \frac{l^2}{(\alpha + ld)^2} - \sum_{j=2}^{n} a_j \sum_{i=1}^{j-1} \frac{1}{(i-d)^2} \qquad l_{\alpha\alpha} = \sum_{m=1}^{n-1} \frac{1}{(\alpha+m)^2} - \sum_{l=1}^{k-1} \frac{1}{(\alpha+ld)^2}$$

$$\text{and} \qquad l_{d\alpha} = l_{\alpha d} = -\sum_{l=1}^{k-1} \frac{l}{(\alpha+ld)^2}$$

By definition the determinant of H equals $|H| = l_{dd} l_{\alpha\alpha} - (l_{\alpha d} l_{d\alpha})$. Hence, to ensure the positivity of $|H|$ it is sufficient to verify that:

$$l_{\alpha\alpha} = -\frac{l_{d\alpha}^2}{|l_{dd}|}$$

Table 11 reports the average values[48] of the Maximum Likelihood estimators for each volatility regime. While the ML estimates of α do not deviate much from the starting values, the estimated discount parameters are significantly below the corresponding values of d_0. Low estimates of d imply that the process generates a small number of clusters, a result in contrast with the power law nature of the data.

	Maximum Likelihood Poisson-Dirichlet parameters $(\hat{\alpha}, \hat{d})$.								
	Low Vol.			Med. Vol.			High Vol.		
	$\tau = 24$	$\tau = 73$	$\tau = 185$	$\tau = 24$	$\tau = 73$	$\tau = 185$	$\tau = 24$	$\tau = 73$	$\tau = 185$
$(\hat{\alpha}_G, \hat{d}_G)$	$(2.8, 0.7)$	$(1.19, 0.6)$	$(1.48, 0.7)$	$(2.31, 0.7)$	$(1.19, 0.6)$	$(0.99, 0.7)$	$(1.81, 0.7)$	$(0.81, 0.6)$	$(0.55, 0.7)$
$(\hat{\alpha}_F, \hat{d}_F)$	$(2.85, 0.6)$	$(2.8, 0.7)$	$(2.8, 0.7)$	$(2.2, 0.65)$	$(2.7, 0.75)$	$(2.7, 0.7)$	$(1.99, 0.6)$	$(2.8, 0.7)$	$(2.8, 0.7)$

Table 11: Average Maximum Likelihood estimates $(\hat{\alpha}, \hat{d})$ under Gaussian (G) and FHS (F) innovations for $\tau = 24, 73$ and 185 days over the sample period Jan.2002 - Dec.2004.

[47] Estimates when one of the two parameters is assumed to be known are available upon request.
[48] We compute the mean conditional on the convergence of the Maximum Likelihood algorithm (30% of the observation dates).

4.2 The Bayesian approach

The Bayesian approach integrates the information contained in the likelihood function with the subjective beliefs about the parameters distribution. The posterior distribution, $\pi(\alpha, d|x)$, has the following structure:

$$\pi(\alpha, d|x) \propto \mathbb{L}(\alpha, d|x) \cdot p(\alpha) \cdot p(d);$$

$$\alpha \sim NTrunc(\alpha_0, \sigma) \qquad d \sim U(a, b)$$

where $\theta = (\alpha, d)$, $\mathbb{L}(\alpha, d|x)$ is the likelihood function as in eq.(25) and $p(\cdot)$ is the prior distribution. Given the data characteristics, α_i is assumed to follow a Truncated Normal distribution with mean $\alpha_{0,i}$ and standard deviation $\sigma = 2^{49}$ over the support $[-d_{0,i}, +\infty)$. Choosing for an uninformative prior, the discount parameter is uniformly distributed on the support $(0, 1)$. Complexity and precision can be further enhanced by assuming a prior distribution also on the PD model hyperparameters.Using the Metropolis-Hastings algorithm (see Hastings (1970)[63]), samples from the posterior distribution are generated and, to guarantee the convergence of the chain to the true value of the parameter, the degree of mixing of the simulated chains is checked by looking at the autocorrelation plots[50]. Finally, under the assumption of a squared-error loss function, the Bayesian estimators are computed as the mean of the simulated values.

Bayesian estimates of the PD parameters $(\alpha^\dagger, d^\dagger)$

For each volatility regime and time to maturity, the first rows of table 12 present the average strength and discount parameters, $(\alpha^\dagger, d^\dagger)$.

	Bayesian Poisson-Dirichlet parameters $(\alpha^\dagger, d^\dagger)$								
	Low Vol.			Med. Vol.			High Vol.		
2002-2004	$\tau = 24$	$\tau = 73$	$\tau = 185$	$\tau = 24$	$\tau = 73$	$\tau = 185$	$\tau = 24$	$\tau = 73$	$\tau = 185$
$(\alpha_G^\dagger, d_G^\dagger)$	$(5.28, 0.83)$	$(5.31, 0.83)$	$(5.44, 0.84)$	$(5.45, 0.87)$	$(5.45, 0.85)$	$(5.45, 0.95)$	$(5.47, 0.92)$	$(5.46, 0.95)$	$(5.54, 0.96)$
$(\alpha_F^\dagger, d_F^\dagger)$	$(5.42, 0.85)$	$(5.44, 0.87)$	$(5.46, 0.86)$	$(5.43, 0.89)$	$(5.48, 0.88)$	$(5.49, 0.89)$	$(5.44, 0.94)$	$(5.49, 0.95)$	$(5.54, 0.97)$

Table 12: Average Poisson-Dirichlet parameters $(\alpha^\dagger, d^\dagger)$ when $\tau = 24, 73$ and 185 days using Gaussian (G) and FHS (F) innovations over the sample period Jan.2002 - Dec.2004.

The estimates are validated by the results of the convergence analysis. At each observation date as the number of lags increases, the autocorrelation converges to zero after a reasonable number

[49]This is a common choice for an informative prior.

[50]In general, the simulated draws of α and d induce an autocorrelation converging to zero after a reasonable number of draws.

of lags, thus indicating that the chains have a satisfactory degree of mixing. The concentration parameter $\alpha^{\dagger}$ is always above its proposed starting values, thus suggesting that, in general, the weight assigned to the prior information should be higher than what predicted under the Dirichlet Process in Sala et al. (2017)[92]. For each time to maturity, $\alpha^{\dagger}$ is a positive function of the volatility. Indeed, during periods of turmoil historical data have a limited informative content and the benefits of exploiting the forward-looking contained in option data become greater. The average discount parameter is always close to the upper bound of its support, thus showing that the process generates a large number of clusters, independently on the volatility environment. Even thought there is not a clear trend within each volatility regime, across all times to maturity the discount parameter is a positive function of the GARCH long run unconditional volatility and always fluctuates in the upper bound of its support.

5 Results

Based on the methodology explained in section 2, the following sections present the PK estimates on a day-specific basis using the Bayesian estimators, $\alpha^{\dagger}$ and $d^{\dagger}$, as inputs in the derivation of p^{*}[51]. At each date and for selected times to maturity, the PK estimates are derived both using Gaussian and FHS innovations in the GJR-GARCH model calibration. Section 5.1 presents and compares the PK estimates under the Dirichlet and the Poisson-Dirichlet process on a day-specific basis. At fixed times to maturity, sections 5.2 and 5.3 present the PK estimates across the whole sample and the results of the corresponding monotonic relation test.

5.1 Day-specific PK Estimates

For each volatility regime two sample dates[52] are randomly selected and the corresponding PK estimates under the PD and Dirichlet[53] are presented. As anticipated in section 2.2, the PD process collapses to the Dirichlet one when the discount parameter is equal to zero. The Dirichlet process generates a low number of clusters and the posterior distribution is concentrated around the a priori mean, i.e. the base distribution. Conversely, increasing values of d generate an increasing number

[51]The PK estimates using the ML estimators are available upon request.
[52]Under the risk neutral measure the unconditional GARCH volatility changes with the innovations type, therefore, for each volatility regime one (different) sample day must be selected.
[53]Estimates of the concentration parameter are as in Sala et al. (2017)[92]

of small clusters and values less close to the mean of the process. While the Dirichlet and the PD processes share many characteristics and properties[54], the modeling of the tails of the posterior distribution can be very different. More precisely, the PD process captures a wide variety of patterns in the tail behaviour, whereas the Dirichlet process enables one to model distributions which stay very concentrated around their a priori mean only. The consequences are immediate: even though the Dirichlet process can be a good choice in periods of financial stability, it is unsuitable during high volatility periods, when extreme fluctuations become more likely and a correct PK modeling is needed the most. For each volatility regime, the day-specific results are reported at multiple times to maturity, namely $\tau = 24, 73$ and 185 days. When the volatility is low[55], the first row of figure 5 presents the estimates under the Dirichlet and the PD processes. Even though in a low volatility environment the gap between the Dirichlet and the PD should not be remarkable[56], for medium/long times to maturity ($\tau = 185$ days and to some extent $\tau = 73$ days) the PK estimates exhibit some differences. The estimation exercise at long time horizons naturally embeds more uncertainty[57], as the dynamics of the underlying asset must be forecasted up to a date further in the future. As a consequence, even though the initial volatility is low, the underlying price can deviate in a substantial way from its mean, thus making the Dirichlet process inadequate at long investment horizons. When $\tau = 185$ days, the Dirichlet PK estimates are locally increasing in the left tail of the distribution, thus being unable to solve the PK puzzle. On the contrary, the PD process performs similarly to the Dirichlet one at short time horizons and, at the same time, delivers consistent PK estimates also at longer time horizons. The gap between the two processes is driven by the discount parameter, which impacts on the expected number of clusters and regulates the tail behaviour of the distribution. Under normal market conditions[58], the second row of figure 5 plots the PK estimates at $\tau = 24, 73$ and 185 days. As in the low volatility case, the PK estimates under the PD process ensure better results in terms of monotonicity in the left tail of the distribution. In the case of FHS innovations the gap with the Dirichlet process is remarkable across all times to maturity. While the estimates under the two processes are very close around the 0% return interval [59], the Dirichlet PK estimates are increasing when returns are negative, thus

[54]Both the processes are conjugate and include a wide range of distributions.
[55]On the 155-th Wednesday of the sample, the unconditional risk neutral GARCH volatilities under FHS innovations, $\sigma^{Q}_{155,FHS}$, is equal to $= 12.97\%$.
[56]As the underlying should not largely deviate from its mean.
[57]Compared to estimations performed at short τs.
[58]The selected observation date is the 122^{th} Wednesday, on which $\sigma^{Q}_{122,FHS} = 19.65\%$.
[59]At each given date, the spot price of the S&P500 is indicated with a black star (*) on the x-axis of each graph.

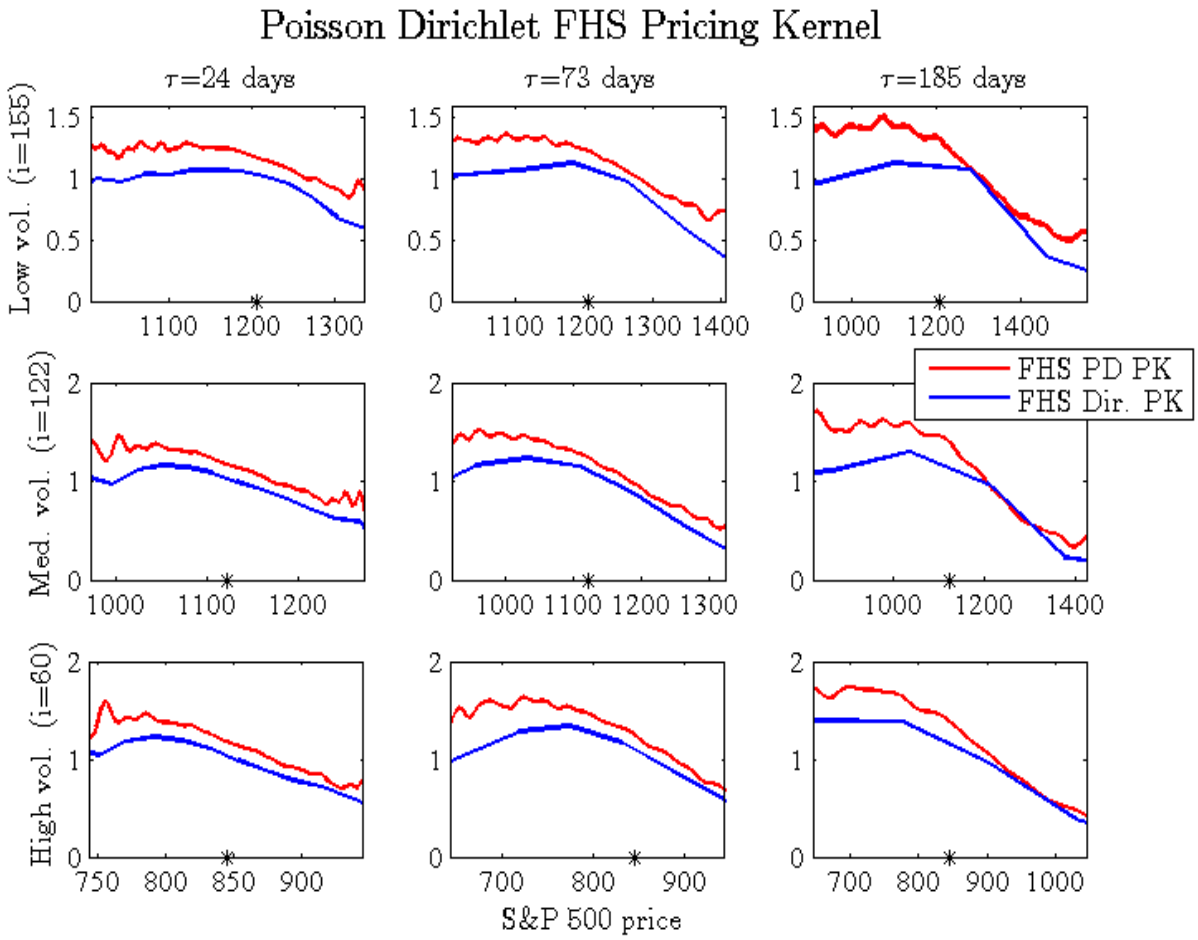

Figure 5: Day-specific PK estimates under the Dirichlet and Poisson-Dirichlet processes in the case of FHS innovations during the sample period Jan.2002 - Dec.2004).

being inconsistent with the classical theory. The ability of the PD process to model the PK as a decreasing function across almost all return intervals sheds light on the possibility to solve, or at least to substantially reduce, the PK puzzle by means of an accurate model choice and calibration. The last row of figure 5 presents the PK estimates in the high volatility regime[60]. During periods of turmoil the probability of observing extreme events increases and a correct modeling of the tails becomes crucial. In this context the PD process outperforms the Dirichlet one thanks to its ability to capture and model in a more flexible way specific patterns of the data in the tails of the distribution. When the innovations are filtered historical simulated the estimates under the Dirichlet process exhibit a puzzling upward slope in the left tail, thus failing to meet the

[60]The selected date is the 60^{th} Wednesday of the sample, date on which the volatility under FHS innovations is 25.29%.

monotonicity requirement. Conversely, even though some irregularities are still present, the PD process provides a set of estimates closer to the predictions of the classical theory.

5.2 The PK Estimates across the whole sample

To gain a complete overview of the estimation results, this section presents the three dimensional plots of the PK estimates across the whole sample period when the innovations are filtered historical simulated[61]. The analysis considers the same times to maturity used in section 5.1. At each observation date, figure 6 plots the PK estimates under the PD process under FHS innovations when $\tau = 24$ days. Across the whole sample period the slope of the PK is decreasing but almost

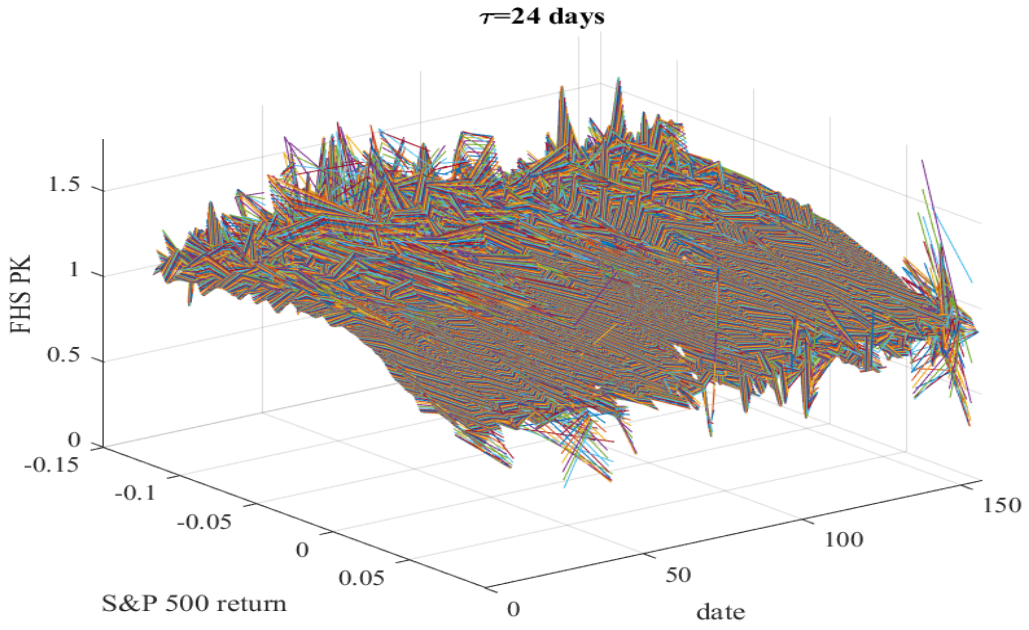

Figure 6: PK estimates across all sample in the case of FHS innovations ($\tau = 24$ days).

flat and the PK fluctuates around 1, thus meaning that the risk neutral and the modified physical densities are almost equal. In the left tail of the distribution, the frequency of the irregularities is still quite high (especially in the case of FHS innovations), nevertheless deviations from monotonicity are always bounded. As concerns the right tail, the PK distribution presents some spikes, which are partially driven by the lack of a dense amount of OTM traded call options which makes statistical inference on some sample dates unreliable.

[61]Results under Gaussian innovations are available upon request.

Consistently with the theoretical result on the divergence between risk neutral and physical densities, as the time horizon gets longer $\mathbb{M}_t^\dagger$ increases under both types of innovations. Figure 7 presents the estimation results in the case of FHS innovations when $\tau = 73$ days. Bounded spikes are still

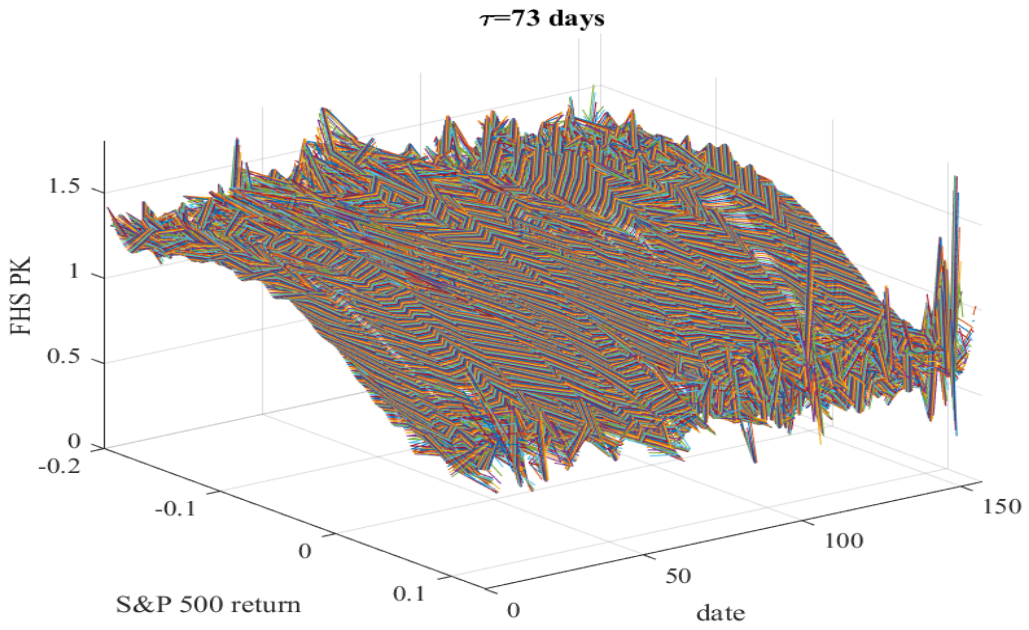

Figure 7: PK estimates across all sample in the case of FHS innovations ($\tau = 73$ days).

present in the tails of the distributions; nevertheless they only locally affect monotonicity for small and extreme return intervals. As shown in figure 8, when $\tau = 185$ days $\mathbb{M}_t^\dagger$ exhibits a monotonic decreasing shape across the whole sample period.

This result confirms the findings in the literature, according to which for long time horizons the monotonicity assumption of the PK cannot typically be rejected. Few isolated irregularities still remain in the right tail of the distribution due to the lack of a liquid market for OTM call options. Importantly, the isolated deviations from monotonicity are always bounded, hence the model is helpful in reconciling the PK estimates with the observed option prices.

5.3 Monotonicity test

In this section the monotonic relation between PK and market returns is formally tested using the Monotonic Relation (MR) test by Patton and Timmermann (2010)[88]. The MR test presents several attractive features. First, being a nonparametric test, it does not impose any assumption

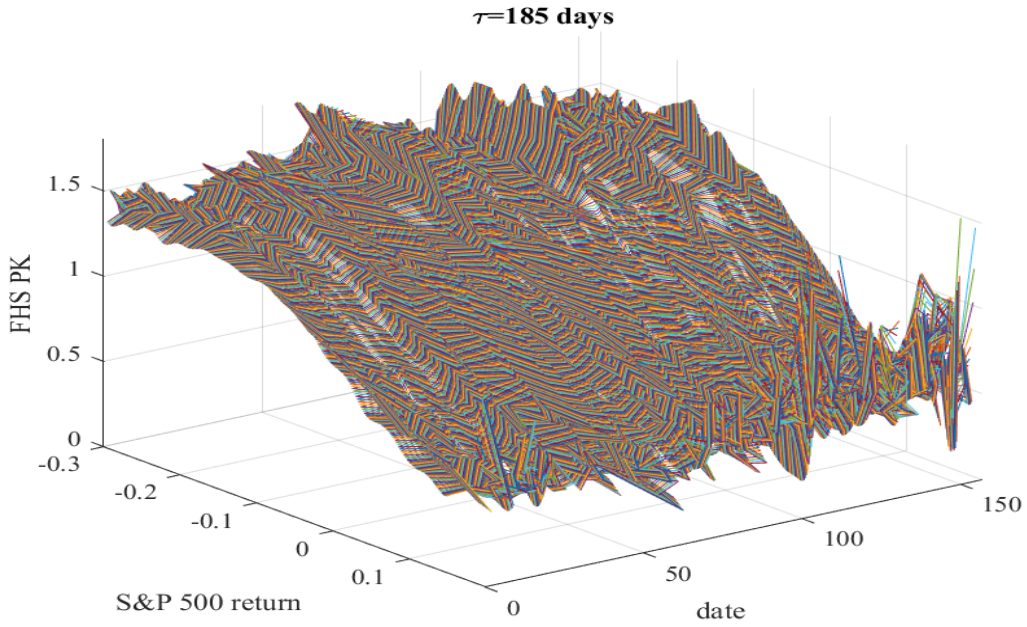

Figure 8: PK estimates across all sample in the case of FHS innovations ($\tau = 185$ days).

neither on the returns distribution nor on the functional form of the relation between the returns and the PK. Second, the design of the test substantially reduces the probability of the type II error. Indeed, under the null hypothesis the MR test states that the PK is identical or weakly increasing in the returns, while under the alternative it is monotonically decreasing. In this way the theoretical prediction of a monotonic decreasing relation is confirmed only if the data contain a sufficiently strong evidence to support it. Differently from other widely used tests which entertain monotonicity under the null hypothesis (see Wolak (1989)[94] and Gourieroux et al. (1982)[58]), a rejection of the null of the MR test represents a strong empirical endorsement of the theory[62].

For each investment horizon and innovations type, the PK estimates are tested over the whole sample period. The first two rows of table 13 report the number of times the null hypothesis is rejected at 95% confidence level when the Bayesian PK is derived using FHS and Gaussian innovations, respectively. In the big majority of the cases (at least 93% of the observation dates), the null is rejected, thus implying that the data support the decreasing monotonic relation between the Bayesian PK and the market returns[63]. Differently from recent findings in the literature[64], the

[62]The same conclusion is not valid for aforementioned tests, whose results can be biased by limited power issues. Indeed, short time series of observations or noisy data may increase the probability of a type II error.
[63]The relationship is tested over a sufficiently wide range of market returns, namely ±20%, ±30% and ±40% when $\tau = 24, 73$ and 185 days.
[64]See Cuesdeanu and Jackwerth (2016)[35].

% cases H_0 is rejected			
	τ=24	τ=73	τ=185
2002-2004 (FHS)	95%	93%	94%
2002-2004 (GAUSS)	94%	94%	96%
2012-2014 (FHS)	68%	89%	97%

Table 13: Proportion of observation days on which the null hypothesis of the MR test is rejected over the sample periods Jan.2002 - Dec. 2004 and from Jan. 2012 to Dec. 2014.

monotonic relation is still valid on investment horizons shorter than one month, thus corroborating the consistency of the obtained PK estimates with the classical theory.

6 A more recent sample

This section presents an out-of-sample analysis in order to assess the validity of the proposed methodology in a different market situation. Following the same procedure, the PK estimates are derived on the same underlying but on a more recent sample period.

6.1 The Dataset

Starting from January 2012 to December 2014, the sample period consists of 154 weekly observations. To derive the risk neutral density, European call and put options written on the S&P500 with time to maturity ranging between 10 days and one year are downloaded from OptionMetrics. The time series of the S&P500 is from the same database. Given the atypical market situation characterized by negative interest rates, the risk free rate is extracted from at the money option prices using the put-call parity relation. Discrete dividends are downloaded from CRSP and used to derive the continuous dividend yield for each observation date. The choice of a GJR-GARCH model is validated by the Lagrange Multiplier ARCH test (Engle (1982)[48]). In this part of the analysis innovations are assumed to be filtered historical simulated only.

6.2 Parameters calibration

The model parameters are derived following the Bayesian technique presented in section 4[65]. As before, conditioning onto the unconditional long run GARCH volatility, three regimes and their

[65]Given the impossibility to derive a consistent set of estimates under the Frequentist approach, Maximum Likelihood estimates are not considered in this part of the analysis.

corresponding starting values of the equity premium π are identified. Volatility thresholds and starting values for the PD parameters and the equity premium are defined as in the first sample and are presented in table 10. Similarly, the prior distributions of α and d are assumed to be the same of section 4.

Table 14 reports the Bayesian estimates, $(\alpha^\dagger, d^\dagger)$, at multiple time horizons.

	Bayesian Poisson-Dirichlet parameters $(\alpha^\dagger, d^\dagger)$								
	Low Vol.			Med. Vol.			High Vol.		
2012-2014	$\tau = 24$	$\tau = 73$	$\tau = 185$	$\tau = 24$	$\tau = 73$	$\tau = 185$	$\tau = 24$	$\tau = 73$	$\tau = 185$
$(\alpha_F^\dagger, d_F^\dagger)$	$(0.14, 0.84)$	$(0.18, 0.88)$	$(0.20, 0.92)$	$(0.17, 0.84)$	$(0.19, 0.88)$	$(0.22, 0.93)$	$(0.19, 0.89)$	$(0.25, 0.90)$	$(0.27, 0.94)$

Table 14: Average Poisson-Dirichlet parameters $(\alpha^\dagger, d^\dagger)$ when $\tau = 24, 73$ and 185 days using FHS (F) innovations over the sample period Jan.2012 - Dec.2014.

The discount parameter always fluctuates in the upper bound of its support and it is positive function both of the volatility and of the time to maturity. Consistently with the theoretical result on the convergence between physical and risk neutral densities as τ approaches 0[66], the concentration parameter is negatively correlated with the time to maturity. Hence, the weight assigned to q^* decreases with τ as the option market contains less additional information to be included in the physical density estimation and the asymmetry between the information sets under the two measures shrinks. Compared to the results based on the first sample, the estimates of $\alpha^\dagger$ are lower. Being so close to zero, the concentration parameter assigns little weight to the prior distribution, i.e. the option-based information. This difference stems from the different market outlook in the two samples. From January 2002 to December 2004 the S&P 500 index dropped from above \$1150 to \$776 in October 2002 and then rose again to above \$1200. Conversely, from January 2012 to December 2014 the US market maintained a positive outlook and the S&P 500 index constantly rose from \$1277 up to above \$2000. During periods of turmoil the discrepancy between the information contained in the historical prices and that embedded in the option data naturally increases, thus becoming necessary the integration of option data in the information set under the physical measure. Conversely, in periods of stability or with a continuing positive trend the option-based information plays a minor role, the estimated values of $\alpha^\dagger$ decrease and historical data can be a sufficient and satisfactory input for the modeling of the modified physical density. These differences in the general market outlook are naturally reflected in the Poisson-Dirichlet parameters estimates, and are confirmed by the obtained $(\alpha^\dagger, d^\dagger)$ reported in table 12. While the general period of uncertainty during the first sample implies estimates of $\alpha^\dagger$ fluctuating around 5,

[66]This conclusion follows from the fact that $\mathbb{M}_t$ converges to 1 as the time to maturity shortens.

the stability of the US market from 2002 to 2004 induces a sharp decrease of the concentration parameter estimates, which converge to 0 across all sample dates.

6.3 Results

As in section 5, for the same times to maturity ($\tau = 24, 73$ and 185 days) the PK estimates are derived on a day-specific basis and over the whole sample. For each volatility regime, one sample date is randomly chosen and the corresponding PKs are plotted at multiple τs. The monotonicity requirement around the 0% interval is met across all specifications. For $\tau = 24$ days, especially when the volatility is in the low and medium regime, some irregularities are still present in the tails of the distribution and are motivated by the lack of a dense amount of deep OTM options. As τ increases, the OTM option market becomes more dense, irregularities vanish and the PK estimates are consistent with the classical paradigm also for extreme positive and negative returns. Compared to the day-specific results of the first sample, the estimates behave in a more regular way in the tails thanks to the increasing amount of traded OTM call and put options (see figure 4) over the recent years and to the different market conditions characterizing the two periods. More precisely, during the first sample period the S&P 500 experienced a larger upward/downward movements which made the estimation exercise more challenging. Similar conclusions can be inferred by observing the estimates based on the whole samplewhen $\tau = 24, 73$ and 185 days. Due to the lack of a dense amount of OTM options trading at short times to maturity, the estimates for $\tau = 24$ days present many irregularities in the tails and do not generally respect the monotonicity requirement. As options expiration date gets further in the future ($\tau = 73$ and 185 days), the option market becomes more liquid and extreme movements in the tails are ruled out. As shown in figures 9 and 10, for medium and long times to maturity the estimated PK is monotonically decreasing in the returns around the 0% interval with few and bounded spikes in the distribution occurring at extreme positive/negative returns.

The null hypothesis of identical or weakly increasing PK against the alternative of a monotonic decreasing relationship is tested by using MR test. As reported in the last row of table 13), when $\tau = 24$ days the MR test has the lowest acceptance rate, as the PK estimates are consistent with the classical theory in almost 70% of the cases only. Conversely, at longer investment horizons the PK respects the monotonic decreasing relation with the market returns for more than 89% and 97% of the sample observations dates when τ equals 73 and 185 days, respectively.

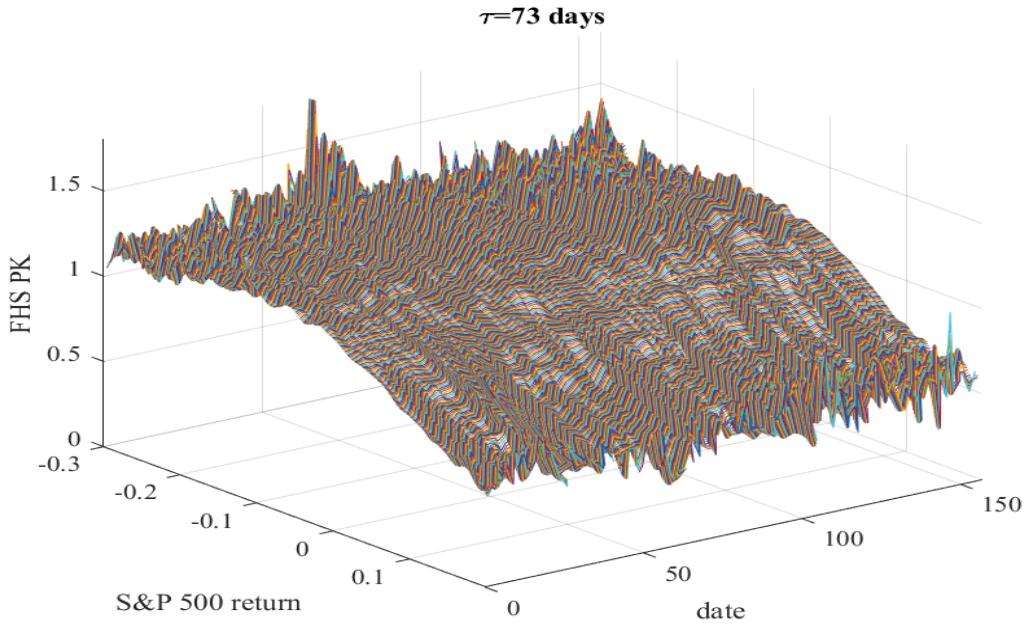

Figure 9: PK estimates across all sample ($\tau = 73$ days).

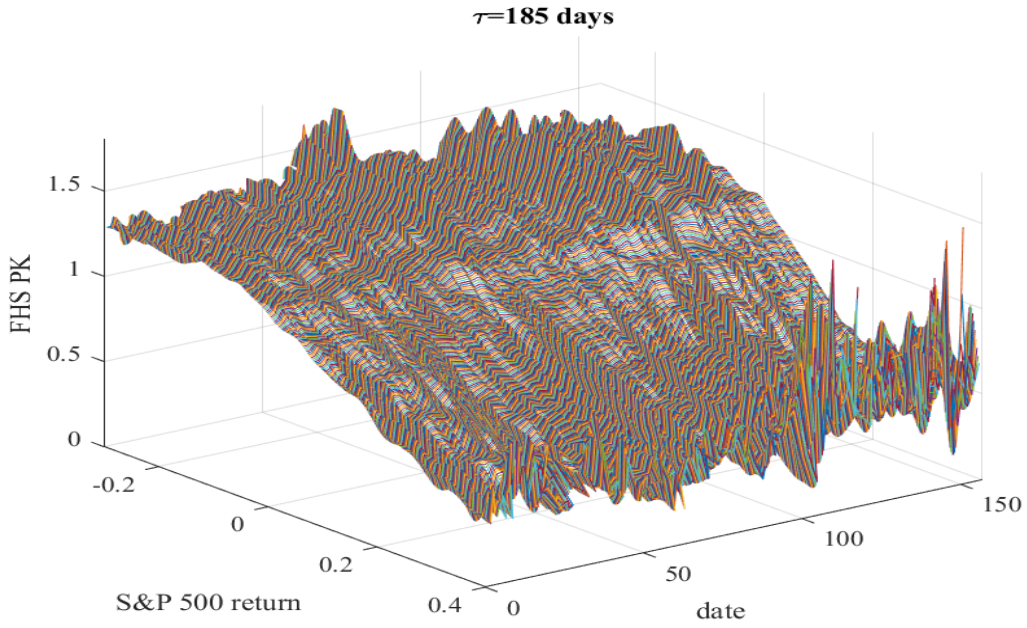

Figure 10: PK estimates across all sample ($\tau = 185$ days).

7 Conclusions

The implementation of a nonparametric Bayesian approach in the physical density estimation and the parameters calibration steps allows one to enhance the precision and the accuracy of the PK estimates. The Bayesian approach perfectly adapts to the setting under consideration without imposing any additional assumption on the functional shape of the PK. The Poisson-Dirichlet parameters capture specific characteristics of the financial data under study and enable one to model in a flexible way the PK dynamics both during high and low volatility periods. In agreement with the economic theory, the concentration parameter, α, is negatively correlated with the time to maturity and implies the convergence of the physical and risk neutral densities when the investment horizon gets shorter. The estimates of α are positively correlated with the volatility level; hence, as expected, the contribution of option-based information becomes more relevant during periods of turmoil. The estimated discount parameter is always close to the upper bound of its support, consistently with the (unconditionally) power law nature of the data under study. The PK estimates gain precision either in the tails of the distribution, thus delivering results consistent with the observed option prices, either in the interval around the 0% return, thus reducing, at least locally, the entity and the severity of the PK puzzle. In the tails of the PK distribution some deviations from monotonicity are still present and can be attributable to the lack of a dense OTM call market. The proposed fully Bayesian methodology to estimate the physical density allows one to eliminate the identified key driver of the PK puzzle, i.e. the information bias. At each observation date, the proposed methodology delivers estimates of the PD parameters and of the PK consistent with findings in the literature and with the economic intuition underneath the model. The estimation procedure is reliable also when applied to a recent sample period. Compared to previous results in the literature, the PK estimates are always monotonic decreasing around the 0% interval and in the tails in general[67], thanks to the ability of the PD discount parameter to correctly model the tails of the data distribution. Hence, results are consistent over a larger support (i.e. a wider range of returns) and the PK estimates are reconciled with the observed price dynamics. To conclude, the results provide supportive evidence to the classical theory postulating the PK monotonicity, thus showing that the PK puzzle is, at least partially, an expression of the information bias.

[67] Some exceptions still occur for short times to maturity for extreme positive/negative returns.

Part III

S&P 500 Index, an Option-Implied Risk Analysis

1 Introduction

Thanks to their ease of application and to several empirical and theoretical convenient features[68], the Value at Risk (VaR) and the Conditional Value at Risk (CVaR) have emerged and still are the most widely used risk measures in the finance community. Summarizing the market risk in a single number, both risk measures assess the maximal loss at a specified probability level over a fixed time horizon. The risk management literature proposes a wide variety of approaches to estimate the VaR and CVaR. Nevertheless, to date, most of proposed models infer their estimates from past returns (henceforth: "statistically-based" models), a backward-looking type of input, which often poorly predicts the dynamics of financial markets, especially during crashes. Following the last financial crisis, the recent efforts of regulators and governments to provide more solid and comprehensive risk models best demonstrates how the statistically-based risk measures are still far from providing a reliable capital protection.

Being a direct expression of investors' future beliefs, option prices are by construction forward-looking and market-based financial assets. Using option market data as inputs and adopting a nonparametric approach, this paper tests if the option-based (henceforth: "option-implied") VaR and CVaR deliver more accurate risk forecasts. Performing well, both at short and long time horizons and at different risk levels, our results show that a comprehensive capital protection scheme should include forward-looking information in the analysis. Or, alternatively, that "option-implied" risk measures can be considered as valid alternatives to the classical "statistically-based" risk metrics.

While the role of both the VaR and CVaR is theoretically well understood, the existing statistically-based models deliver estimates suffering from several economic and econometric biases. First, even though at different extents, all statistically-based risk models involve errors in the modelling of the underlying asset volatility, thus being exposed to the "risk that the risk will change" (Engle (2009)[44] and (2011)[45]). Second, the statistically-based risk models often use historical data as inputs, thus making the resulting estimates almost unconditional. More formally, at each trading

[68]Refer to Dowd (1999)[39], Duffie and Pan (1997)[42] and Penza and Bansal (2000)[89] and Jorion (2007)[67] for a complete overview on the topic.

day only one price/return is observable. Being just a single scalar, it is thus not possible to make proper inference out of it. To tackle this problem, a large amount of past data is often used in the estimation process. Unfortunately, when piling up many historical returns to estimate the future profit and loss (P&L) distribution, the last values become almost uninformative. As a consequence, also in presence of efficient and informative markets, the obtained statistically-based risk measures turn out to be almost unconditional with respect to the actual market situation and strongly dependent on the time window used in the estimation process. Third, it is well-known[69] that the statistical properties of market prices depend on the general market outlook. In stable periods, the heterogeneity among investors' behaviour increases, whereas during high volatility periods agents' actions become more similar (e.g.: bank runs and flight to safety). As a consequence, the statistical analysis conducted in stable periods can rarely be helpful during highly volatile ones, and viceversa. Fourth, short-time horizons risk measures[70], although more robust and easier to extract than those at longer horizons, might not be helpful if one has to liquidate a large (and possibly illiquid) financial position. Risk estimates over longer term time horizons are then needed but, as shown by Brownlees et al. (2011)[25], while the performances of most of risk measures are more or less good at short time horizons, they become highly unsatisfactory in the long term.

Option market data are helpful in order to overcome the aforementioned problems. First, compared to historical records, option market data have a superior informative power, especially as concerns the modelling of the underlying asset volatility[71]. Second, at each trading day the implied moments of the daily option surface reflect investors' expectations, thus being a naturally forward-looking indicator of investors beliefs. Third, at each observation date, option data have a matrix structure from which is more informative in the derivation of the P&L distribution. By pricing several states of the world in terms of strikes and times to maturity, the initial dataset is naturally richer than the one based on historical data and directly delivers estimates at longer time horizons [72]. Surprisingly, despite the increasing attention of academics and practitioners on the predictive content of derivative securities, the use of option data in the risk management context is still widely unexplored. To the best of our knowledge, only Äit-Sahalia and Lo(2000)[3], Bali

[69] See e.g.: Danielsson (2002)[37] and references therein.
[70] One day ahead estimates.
[71] From Merton (1973)[83] this is a large and still very dynamic and open stream of literature.
[72] Directly observing the market prices of call/put options enables one to extract option-implied risk forecasts with time horizon equal to the time to maturity of the underlying option contract.

et al. (2011)[6], Samit (2012)[93] and Mitra (2015)[85] relate to the topic in a similar manner[73]. After presenting the option-implied methodology developed by Barone Adesi (2016)[7], this paper proposes an extensive empirical analysis on how the option-implied risk measures perform compared to the statistically-based ones. Several backtesting results show that the option-implied VaR and CVaR estimates are accurate in the short (weekly) and long (monthly) term, independently on the tail of the distribution used in the estimation process. On a relative level, the option-implied methodology outperforms the statistically-based one using an AR(1) - GJR GARCH (1,1) model with filtered historical simulated (FHS)[11] innovations. The choice of this benchmark model stems from Brownlees et al. (2011)[25], who shows that, for the S&P 500, the asymmetric GJR GARCH (1,1) model is the best performer among a big family of different risk models[74]. The option-implied risk forecasts are fully market driven and nonparametric; hence, the lack of liquidity or the existence of mispricings may potentially bias the final results. Results could be of interest for regulators, central banks and single companies. Regulators and central banks can derive risk estimates of large companies[75] without knowing the precise composition of their portfolio. Large companies can compare the option-implied estimates with those delivered by their internal models.

The remainder of the article is as follows: section 2 reviews the VaR and CVaR risk measures. Section 3 derives the option-implied VaR and CVaR. Section 4 describes the dataset and its relevant characteristics. Section 5 presents the option-implied nonparametric (model-free) and parametric (Black and Scholes) VaR and CVaR. Section 7 presents the statistically-based VaR and CVaR estimates using a GJR-GARCH model with FHS innovations. Section 8 analyses and compares the backtesting results for all the aforementioned risk estimates. Section 9 concludes.

2 VaR and CVaR

Defined as the quantile of the projected profit and loss distribution, the VaR (α,t,T) is the loss at time t, associated with a generic portfolio value ($S_{t,T}$), at a given time horizon (T$< \infty$) and at $(1-\alpha)$ confidence, where $\alpha \in [0,1]$ defines the chosen risk level. In this framework the $VaR^{\alpha}_{t,T}$ is

[73]Less directly related but still worth mentioning, Bali and Murray (2013)[6], Xing, Zhang, and Zhao (2010)[95], and Yan (2011)[96] show how option-implied skewness or jump risk measures have significant explanatory power in pricing the cross-section of asset returns.

[74]We also performed the same analysis on many other models like the variance-covariance model, the historical model, the expected weighted moving average model and many different ARCH specifications. Results confirm the findings of Brownlees et al. (2011). For ease of space we only propose the AR(1)-GJR GARCH (1,1) estimates but all the additional results are available upon request.

[75]Those companies with a sufficiently liquid derivative market to implement the option-implied methodology.

defined as the α-quantile of the return distribution:

$$\text{VaR}_{t,T}^{\alpha} = \int_{-\infty}^{K} f(S_{t,T}) dS_{t,T} = \alpha \tag{27}$$

where $f(S_{t,T})$ represents the probability distribution function of the *future* portfolio values based on its value today, S_t, and projected at the future horizon $T \geq t$. Because of its forward-looking nature, the pricing probability distribution of the *future* portfolio values, $f(S_{t,T})$, is the most challenging element to estimate. Although empirically robust and easy to implement, the lack of convexity and the quantile nature of the estimation makes the VaR an unsatisfactory and unreliable risk measure in controlling for tail risks. The lack of sub-additivity[76] implies that the VaR is not weakly coherent[77] and may discourage diversification.

Also known in literature as expected shortfall, expected tail loss or average value at risk, the CVaR has been proposed by Artzner et al. (1999)[5] and strongly supported by the Basel Committee as an alternative to overcome the VaR limitations. The CVaR is defined as the conditional expected loss determined in eq. (27):

$$\text{CVaR}_{t,T}^{\alpha} = \frac{1}{\alpha} \int_{-\infty}^{K} L(S_{t,T}) f(S_{t,T}) dS_{t,T} \qquad L(S_{t,T}) = S_t - S_{t,T} \tag{28}$$

where $L(S_{t,T})$ defines the future loss function. Unfortunately, the lack of elicitability (Gneiting (2011)[56] and Ziegel (2014)[98]) of the CVaR makes it inferior with respect to the VaR when performing a backtesting analysis[78].

3 Linking VaR and CVaR to the option market

This section shows how to derive the option-implied VaR and the CVaR. As a main advantage, extracting the cross-section of risk measures from the option surface allows one to derive economic-grounded risk measures.

As defined in equation 27, the VaR is a quantile, a single numeric value determined at a specific

[76] It is well known that the VaR is sub-additive only if all marginals of the joint distribution are elliptical. This condition coincides with the assumption underneath the 1952[79] Markowitz's variance-minimizing portfolio, which makes the VaR calculations not even required once that the variance is known.

[77] A risk measure is defined as weakly coherent if it lacks of the positive homogeneity (for details see Carr et al. (2001)[27]).

[78] According to Acerbi and Skezely (2015)[**?**] the CVaR is only backtestable if the profit and loss distribution is known.

threshold over the cumulative distribution of the profit and losses. Under the Arrow-Debreu representation and following Breeden and Litzenberger (1978)[24], the first derivative of a put price, $p = e^{-rT}\int_0^K (K-S)f(S)dS$, over its strike price, K, is:

$$\frac{dp_{t,T}}{dK} = \frac{d[e^{-r_{t,T}T}\int_0^K (K-S_T)f(S_{t,T})dS_{t,T}]}{dK} \tag{29}$$

$$= e^{-r_{t,T}T}\int_0^K f(S_T)dS_{t,T} \tag{30}$$

$$= e^{-r_{t,T}T}F(K) \tag{31}$$

$$= e^{-r_{t,T}T}\alpha \tag{32}$$

where $r_{t,T}$ is the risk-free rate over the time interval (T-t). With no loss of generality, the lower bound of the integral has been changed from $-\infty$ to 0, reflecting the assumption of holding a portfolio with limited liability. As defined in section 2, α represents the risk level (i.e. 1 - confidence level). It follows that at each observation date t, the option-implied $\text{VaR}^{\alpha}_{t,T}$ is the difference between the initial portfolio value and the strike price of a European put option at level α:

$$\text{VaR}^{\alpha}_{t,T} = S_t - K^{\alpha}_{t,T} \tag{33}$$

where $K^{\alpha}_{t,T}$ is the strike price of the option contract traded at time t with time to maturity T and identified by the risk level α. Being alpha proportional to the probability that the portfolio value is below $K^{\alpha}_{t,T}$, the obtained VaR is naturally *forward-looking* and directly linked to the perceived future market's beliefs. The use of put options is related to the analysis of the left tail of the distribution. By the same token, the use of call options leads to similar results, but linked to the right tail of the distribution. It follows that the VaR of a short position is:

$$\text{VaR}^{\alpha}_{t,T} = K^{\alpha}_{t,T} - S_t \tag{34}$$

where $K^{\alpha}_{t,T}$ is now the strike price identified by α of an European call option[79].

Following the same intuition, it is possible to derive the option-implied CVaR. Starting again

[79]A natural knowledge of the upside risk is of extreme importance in the commodities market, where sellers (buyers) are mostly affected by the downside (upside) risk.

from the definition of a put option:

$$p_{t,T} = e^{-r_{t,T}T} \int_0^{K_{t,T}} (K_{t,T} - S_T) f(S_{t,T}) dS_{t,T} \tag{35}$$

and setting the today stock value of the loss function as $S_t = K_{t,T} + S_t - K_{t,T}$, from equations (27), (28) and (35) it follows that the option-implied CVaR based on the left tail (downside risk)[80] is

$$\text{CVaR}^{\alpha}_{t,T} = \text{VaR}^{\alpha}_{t,T} + e^{r_{t,T}T} \frac{p^{\alpha}_{t,T}}{\alpha_{t,T}} \tag{37}$$

where p^{α} is the price of the put option contract identified by the risk level α. Just relying on the option market data, the option-implied CVaR equals the sum of the corresponding VaR and an additional term, which equals the market put price compounded at the risk-free rate and divided by the probability of that negative scenario, i.e. alpha.

It is worth noting that all the quantities in eq. (33), (34), (36) and (37) are easily accessible and readily available in the market. The only parameter to be estimated is the risk level $\alpha_{t,T}$, which can be computed either nonparametrically, e.g.: using more or less sophisticated numerical differentiation approaches, or parametrically, e.g.: by specifying an option pricing model. In this paper $\alpha_{t,T}$ is derived (i) nonparametrically using the interpolation method of Barone-Adesi and Elliot (2007)[8] and (ii) parametrically under a Black and Scholes economy. While the parametric method has the advantage of being less data demanding, the nonparametric one provides more reliable and accurate results. For the nonparametric approach the time series of alpha at different risk levels are extracted from the daily cross-section of put/call options. Free of any distributional assumption, the estimates based on this methodology are denoted as "model-free (MF)" option-implied risk measures. From equations (29-32), the MF alpha is defined as the average of the first order condition of three contiguous option prices with maturities equal to the desired investment

[80] By the same token, the upside risk (right tail) is defined as:

$$\text{CVaR}^{\alpha}_{t,T} = \text{VaR}^{\alpha}_{t,T} + e^{r_{t,T}T} \frac{c^{\alpha}_{t,T}}{\alpha_{t,T}} \tag{36}$$

horizon T:

$$\alpha_{t,T} = e^{r_{t,T}T}\left[\frac{1}{2}\left(\frac{dp_{t,T}^{\mathrm{UP}}}{dK^{\mathrm{UP}}} + \frac{dp_{t,T}^{\mathrm{DOWN}}}{dK^{\mathrm{DOWN}}}\right)\right] \tag{38}$$

$$= e^{r_{t,T}T}\left[\frac{1}{2}\left(\frac{p_3 - p_2}{K_3 - K_2} + \frac{p_2 - p_1}{K_2 - K_1}\right)\right] \tag{39}$$

where $p_3 > p_2 > p_1$ and $K_3 > K_2 > K_1$. In presence of non equidistant option prices, eq. (39) can be easily generalized by performing a weighted average. Finally, as defined in eq. (32), $\alpha_{t,T}$ is corrected for the daily market risk-free rate, $r_{t,T}$. Two major advantages make the MF approach numerically efficient and accurate. First, it eliminates the first-order error in the Taylor expansion of the derivative. Second, it eliminates the first-order error due to the implied volatility changing across strike prices. The dataset used in our empirical exercise as already sufficiently rich to exclude the necessity of any interpolation/smoothing of the available data[81]. Differently from the MF approach, the option-implied parametric methodology implicitly assumes that the observed option prices are generated from a Black and Scholes model (henceforth denoted as "B&S"). Following Black and Scholes (1976)[21], $\alpha_{t,T}$ is the probability that the today option contract will expire in the money at expiration T, hence:

$$\begin{aligned}
\alpha_{t,T} &= Pr(S_T < K_{t,T}^{\alpha}) \\
&= N(-d_2) \\
&= 1 - N\left[\frac{\ln(\frac{S_t}{K_{t,T}}) + (r_{t,T} - \frac{\sigma_{t,T}^2}{2})(T-t)}{\sigma_{t,T}\sqrt{(T-t)}}\right]
\end{aligned} \tag{40}$$

The values of $\alpha_{t,T}$ obtained numerically in eq. (39) are compared with the ones derived in eq. (40) at the same confidence level. As presented in the empirical section, the nonparametric nature of the MF approach delivers more reliable estimates, while the B&S is a valid alternative in periods of low liquidity of the option market. Econometrically, the proposed option-implied MF and B&S risk estimates are obtained from a *finite* number of option prices, a characteristic which makes the resulting estimates more accurate with respect to those ones based on the VIX and SKEW indexes, which require the existence of an infinite amount of put and call options. In real world modelling such a dense support of prices is rarely available. As a result, the aforementioned measures are affected by large truncation errors which become larger in the tails estimation, where the liquidity

[81] Nevertheless,these techniques are easy to implement and might be a solution in case of liquidity problems.

dries up quicker.

4 Dataset

Supported by the evidence[82] that a large class of investors holds the market as a form of investment, the underlying asset of the analysis is the S&P 500 index. The methodology presented in section 3 is tested on the European options (SPX) traded on the S&P 500 with 7 and 30 days time to maturity. The 7-days and 30-days maturity[83] options are used to estimate the weekly and monthly risk metrics, respectively. Sharing similar patterns over time, the proposed risk forecasts show that weekly and monthly options are integrated markets. To propose a broader analysis, the risk levels used for the weekly time horizons are 1%, 2.5% and 5%, while for the monthly time horizons are 5%, 10% and 15%[84]. To lighten the notation in sections 5 and 6, the subscript T of the VaR and CVaR estimates is omitted and assumed to be equal either to 7 or 30 days[85] for the weekly and monthly estimates, respectively.

Starting from January 1, 2012 to August 31, 2015 the sample period of the weekly estimations consists of 159 weekly observations. While introduced by the CBOE in October 2005, it is only from the beginning of 2012 that the liquidity of the S&P 500 weeklies is sufficiently high to implement the option-implied methodology. Although it took a while for the market to become mature, nowadays, some weekly options (i.e. Fridays, Wednesdays and Mondays expirations) account for more then 40% of the total options daily volume traded on the market[86]. The monthly analysis allows us to extend the sample period (2005-2015) and to verify the performance of the estimates during the 2007-2009 financial crisis. Given the monthly expiration convention, the presented risk measures refer to the third Thursday of each single month. Going from very short, to very long (LEAPS)

[82] As reported by different surveys, a big fraction of U.S. investors relies on indexing policies for investment. Analyzing 60 years of market data, Bogle (2005)[22] shows how, since 2000, index funds account for roughly one-third of equity fund cash inflows and represent about one-seventh of the total amount of equity fund assets.

[83] The VIX, probably the most famous options-based risk measures, also has a 30 days time horizon. Its portfolio composition and the auction system at the base of the index determination make this index not totally reliable. Among the others, possible truncation, discretization, expansion, and maturity interpolation errors (Jing and Tian (2007)[66]) and manipulations (Griffin and Shams (2018)[59]) are under deep investigation.

[84] While omitted for space reason we have performed the analysis over the entire set of risk levels - from 1% to 15% - and for both time horizons. Results are available upon request to the authors

[85] It will be made clear if we will refer to a short (7 days) or a long (30 days) time horizon estimation.

[86] The great success of the S&P 500 weekly options convinced the CBOE to propose other similar products, e.g.: new even shorter expirations (2/3 days) and new underlying assets (weekly VIX). For more details, see the CBOE website.

time horizons, option market data allows for a very natural and broad risk analysis at different time horizons[87]. All contracts have strike price intervals of 5$ and 25$ for at-the-money and out-the-money (ATM and OTM) options and deeply OTM options, respectively. To discard possible mispricings, prices violating classical no-arbitrage lower bounds are excluded from the dataset:

$$p_{t,T} \geq \max[K_{t,T}\,\mathrm{e}^{-r_{t,T}T} - S_t + d_t, 0] \tag{41}$$

$$c_{t,T} \geq \max[S_t - K_{t,T}\,\mathrm{e}^{-r_{t,T}T} - d_t, 0] \tag{42}$$

where d_t is the continuously compounded dividend at time t.

From the time series of the daily zero coupons, the curve of the daily risk-free interest rates is obtained by numerical interpolation. For each time-to-maturity $(T-t)$, the previous $(T-t)_-$ and next $(T-t)_+$ period zero-coupon whose values straddle the time $(T-t)$ are linearly interpolated. Finally, the time series of S&P 500 index dividend yield from January 1, 2012 to August 31, 2015 (weekly) and from January 1, 2005 to August 31, 2015 (monthly) are used to compute the ex-dividend spot index level.

To validate the option-implied VaR and CVaR, the same risk measures are compared with the statistically-based VaR and CVaR derived using the Filtered Historical Simulation approach. For the GARCH estimations used in the FHS approach, the 2005-2015 time series of closing prices (top part of figure 11) is extended in order to include the previous 15 years of observations and the corresponding log returns are derived (bottom part of figure 11). All data are from OptionMetrics.

5 Model-free (MF) option-implied CVaR and VaR

Following the methodology presented in section 3, the weekly and monthly option-implied MF CVaR and VaR are derived based on the left and right tails of the distribution. For the entire analysis the risk levels are set at 1%, 2.5% and 5% (5%, 10% and 15%) for the weekly (monthly) estimates. In this section the time series of $\alpha_{t,T}$ is estimated nonparametrically as in eq. (39).

The weekly estimates do not substantially vary across the whole sample period, except for a high spike on August 2015, in correspondence of the so-called China's "Black Monday" which carried a correction of about 10% on the S&P 500 value. As concerns the monthly estimates, during the

[87] While omitted for space reason, our analysis has been performed also at very short (2 days) or longer than 30 days time horizons. Results are available upon request to the authors.

financial crisis period (from June 2007 to December 2009, according to the NBER) the MF CVaR and VaR exhibit several pronounced upward spikes across all risk levels, thus reflecting the negative outlook of the US market in that period. The CVaR is always larger than the VaR and the distance between the two measures, defined as $\Delta_{\alpha,t} = \text{CVaR}_{\alpha,t} - \text{VaR}_{\alpha,t}$, represents the future market beliefs with respect to the negative scenario occurring with $\alpha\%$ probability. The decomposition of eq. (37) shows that $\Delta_{\alpha,t}$ represents the compounded put option price divided by the corresponding α level. This quantity is highly sensitive to the future market expectations. In periods of market turmoil, for example, investors are more risk averse and the prices of put options are likely to increase, thus implying a larger difference between the CVaR and VaR estimates. The upper part of table 17 and 18 report basic statistic information about the weekly and monthly $\Delta_{\alpha,t}$, respectively. As the option-implied methodology strongly relies on the liquidity of the option market, the upper part of table 15 reports the proportion of missing values for each risk level.

Model-Free approach	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$
% missing observations (Left)	1.26	0.00	0.00
% missing observations (Right)	1.88	1.25	18.23
Black and Scholes approach	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$
% missing observations (Left)	0.00	0.00	0.00
% missing observations (Right)	0.00	0.00	2.52

Table 15: Proportion of missing observations at $\alpha = 1\%, 2.5\%, 5\%$ under the model free (MF) methodology (upper part) and the Black and Scholes approach (lower part) for both the left and right tails of the distribution.

Over the sample period under consideration the liquidity of weekly put option contracts (1^{st} row) is always satisfactory as the proportion of missing estimates is above 1% only for the lowest risk level. The same analysis is then repeated using call options, thus assuming a portfolio of

Model-Free approach	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 15\%$
% missing observations (Left)	6.25	9.38	13.28
% missing observations (Right)	9.38	7.81	12.50
Black and Scholes approach	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 15\%$
% missing observations (Left)	0.78	0.00	3.13
% missing observations (Right)	1.56	1.56	4.69

Table 16: Proportion of missing observations for $\alpha = 5\%, 10\%, 15\%$ extracted with the model free (MF) methodology (upper part) and the Black and Scholes approach (lower part) for both the left and right tails of the distribution.

short positions. The derivation of the risk measures based on right tail of the distribution has

Left Tail	Model-Free				Black and Scholes			
CVaR-VaR	Min	Max	Mean	Median	Min	Max	Mean	Median
$\alpha = 1\%$	2.00	122.51	38.13	36.66	14.95	56.61	26.30	25.23
$\alpha = 2.5\%$	3.00	140.00	41.44	38.00	13.57	86.95	22.92	21.44
$\alpha = 5\%$	14.37	123.00	32.49	30.00	12.40	76.48	20.26	18.67
Right Tail	Model-Free				Black and Scholes			
CVaR-VaR	Min	Max	Mean	Median	Min	Max	Mean	Median
$\alpha = 1\%$	7.50	610.02	25.48	20.00	5.03	30.01	8.44	7.97
$\alpha = 2.5\%$	6.66	53.75	14.61	13.75	11.56	69.16	28.53	27.16
$\alpha = 5\%$	6.66	54.44	12.46	11.70	21.26	84.27	38.57	36.54

Table 17: Summary statistics of Δ_α=CVaR$_\alpha$-VaR$_\alpha$ at $\alpha = 1\%, 2.5\%, 5\%$ under the Model-Free and the Black and Scholes methods for the left (upper portion) and right (lower portion) tails of the distribution.

two interesting properties. First, it provides a benchmark for the estimates obtained using the left one. Second, it allows to make a density free inference on the degree of tail asymmetry in the underlying return distribution. Based on the right tail of the distribution, both the CVaR and the VaR are decreasing in the risk level and display similar patterns to the corresponding metrics based on the left tail, both on a weekly and monthly basis. However, keeping α fixed, both the

Left Tail	Model-Free				Black and Scholes			
CVaR-VaR	Min	Max	Mean	Median	Min	Max	Mean	Median
$\alpha = 5\%$	2.36	150	55.34	49.42	22.44	82.07	40.35	40.24
$\alpha = 10\%$	27.35	142.35	56.08	54.55	22.18	84.99	41.65	41.17
$\alpha = 15\%$	28.51	129.45	59.48	49.43	23.22	88.03	43.14	42.66
Right Tail	Model-Free				Black and Scholes			
CVaR-VaR	Min	Max	Mean	Median	Min	Max	Mean	Median
$\alpha = 5\%$	11.94	114.24	24.78	22.51	10.07	75.80	20.84	19.05
$\alpha = 10\%$	11.53	117.68	23.67	21.36	11.79	83.53	24.01	21.58
$\alpha = 15\%$	12.31	59.94	24.23	23.03	13.22	90.35	26.94	24.48

Table 18: Summary statistics on Δ_α=CVaR$_\alpha$-VaR$_\alpha$ for $\alpha = 5\%, 10\%, 15\%$ under the Model-Free and the Black and Scholes methods for the left (upper portion) and right (lower portion) tails of the distribution.

weekly and monthly MF VaR and CVaR based on the right tail are almost always below those obtained from the left one. This result can be motivated by the shape of the pricing distribution. As for the left tail, the second rows of table 15 and 16 and the left lower portion of table 17 and

18 report the proportion of missing data and the summary statistics on $\Delta_{\alpha,t}$ for the weekly and monthly estimates, respectively. In almost all cases the number of missing observations is larger compared to the one in the left tail, thus confirming the higher degree of liquidity of OTM put options compared to OTM call contracts. It is also worth noting that, for each α level, the average $\Delta_{\alpha,t}$s are much lower than the ones based on the left tail. For α fixed, this discrepancy is driven by the price of the identified option, thus confirming that OTM call options are, on average, cheaper than the corresponding OTM put ones.

6 Black and Scholes (BS) option-implied CVaR and VaR

In this section we repeat the same analysis, but under the assumption that the observed option prices are generated by a Black and Scholes pricing model. As a consequence, the parametric α is derived as in eq.(40). First, for both put and call options, the option-implied volatilities are backed-out by numerical inversion from the daily S&P 500 option market prices:

$$\sigma_t^{IV} = f(S_t, K_t, r_t, q_t, \tau, p_t/c_t) \tag{43}$$

where S_t represents the time t price of the S&P 500 Index, K_t is the time t strike price, r_t the daily risk-free rate, q_t the time t dividend, T is the time-to-maturity (which is always fixed to 7 or 30 days) and p_t/c_t the market prices for European put (p_t) and call (c_t) options, respectively.
Then, the probability of the underlying asset of being below a predetermined threshold K_t is computed as in equation (40). The same holds for the call options, with $Pr(S_T > K_{t,T}^{\alpha}) = N(d_2)$. Once that the time series of α_t is derived, the weekly and monthly B&S VaR and CVaR are computed as in equations (33) and (37). The biggest advantage of this approach, as for most of the Black and Scholes results, is the existence of a closed-formula solution for the value of the threshold α. Not requiring any contiguous option price for the interpolation, each alpha value can be derived as soon as there is at least one market price. The lower part of table 16 reports the proportion of missing values for the entire sample. As expected, the proportion of missing data across all α specifications is almost zero and always much smaller than that obtained under the MF methodology (upper part of table 16). The higher stability of the B&S values conflicts with the accuracy of the final results. Indeed, the B&S paradigm implies larger estimation errors when compared with the MF

method. A reduced smile effect is remarkable into the deepest area of the tails, where the B&S model provides approximative estimates. Among the others, these effects can be summarized by the difference of alpha under the two methodologies, $\Gamma_t = \alpha_t^{\text{MF}} - \alpha_t^{\text{BS}}$. Table 19 shows summary statistics of Γ_t for both the weekly (upper part) and monthly (lower part) observations. For both

Γ_t	Mean	Median	Min	Max
Left tail (W)	-0.0156	-0.0072	-0.2026	0.2225
Right tail (W)	0.0201	0.0113	-0.1790	0.2513
Left tail (M)	-0.0035	-0.0096	-0.3390	0.2430
Right tail (M)	0.0284	0.0229	-0.2027	0.2481

Table 19: Summary statistics on $\Gamma_t = \alpha_t^{\text{MF}} - \alpha_t^{\text{BS}}$ when using options with weekly(W) and monthly(M) time to maturity.

time horizons and tails, the behaviour of the estimates is overall comparable both in terms of mean and median values. As expectable, the min-max range is larger, thus confirming the non-normality of the dataset. Interestingly, also the highest and lowest statistical values show a similar behaviour across the sample for both put and call options and for both short and long time horizons.

Graphically, the evolution over time and risk level of the weekly VaR and CVaR estimates confirm most of the MF results, but at different magnitudes. At $\alpha = 1\%$, the B&S VaR and CVaR present a few large spikes, which are motivated by the high (possibly irrational) put prices on that dates. No similar spikes are detected at other risk levels, but for the very last one (end of 2015), which is present in all graphs (although at different magnitudes). The B&S VaR and CVaR are accurate risk predictors over the financial crisis. Indeed, the NBER recession period from June 2007 to December 2009 coincides with the highest VaR and CVaR values. It is worth noting how, also with a parametric α, option prices are able to forecast the quiet pre-crisis period and anticipate the subsequent crisis. As a confirmation, the biggest spike within the recession area represents the pre and post Lehman Brother failure (September 15, 2008). The subsequent jumps correspond to the downward S&P 500 movements of June 2010 (S&P 500 down of \$200 at \$1022.58 from \$1217.28 of two months before) and the period of turmoil from July to November 2011 during which the S&P 500 index dropped from above \$1300 losing more than \$200. At the rightmost part of the figure, also the crisis of Summer 2015 is well identified. Similar results hold for the estimates based on the right tail of the distribution.

Compared with the MF results, while the mean and median values of the weekly Δs are comparable, the min-max range substantially differs. Keeping alpha and the compounding effect fixed,

the two methodologies work differently into the tails of the distribution, with the B&S approach that deals with higher values. As expectable, this results confirms that the MF approach better estimates the tails of the distribution, while on average also the parametric B&S approach might work well.

Both the weekly and monthly estimates, once compared with the nonparametric approach (presented in the right part of tables 18 and 17) show, as expectable, a more "conservative" behaviour. At most levels in fact the extreme values are less dispersed and the central moments are lower than non-parametric results. At all levels, values that differ from the log-normal distribution implied by the Black and Scholes approach are evidence in favour of a non-parametric and non-normal distribution of returns, thus justifying the use of a data driven MF approach. For the monthly estimates, as α_t grows, the B&S VaR and CVaR estimates decrease and the distance between the two measures, Δ, increases.

7 Statistically-based Risk Estimations

Economically, the main difference between the option-implied and the statistically-based historical risk measures stems from the risk premium. It is well known that the risk premium is unobservable and that its estimation is still a very open debate in the finance community (Damodaran (2016)[36]). Indeed, investors are generally not risk neutral and the ex-ante quantification of their risk preferences is not a trivial task to perform, as their subjective probabilities are unobservable quantities. It follows that, under the physical measure, the absence of a reliable methodology to estimate the risk premium is the main drawback faced by any risk metric. While the difference between the risk neutral and the physical measures is theoretically well known, their distance is indeed small and almost negligible at short time horizons. In fact, by definition, the pricing kernel converges to unity as time to expiration shrinks:

$$S = e^{rT} q \cdot 1 = e^{RT} p \cdot 1 \tag{44}$$

where S is the risk-neutral expected payoff both of a contingent claim, q, and of the relative physical measure, p. If the no arbitrage assumption applies, the risk-neutral value grows at the risk-free rate, r, whereas the physical value grows at the risk-adjusted risk-free rate, R. As the time to maturity T approaches zero, both quantities converge to the same state price density, S;

conversely, when T increases their divergence increases at the rate $(R-r)T$. The impossibility to directly observe the physical measure and the risk premium makes the estimation of the pricing kernel rather problematic. If follows that, as long as T is not too big, adopting a model-free risk-neutral approach, which avoids *any* numerical implementation and parametrization to model future dynamics of the underlying asset, may be a convenient choice. Supportive evidence is provided in the work of Martin (2017)[80]. Under the physical measure the probability of observing a crash, $\alpha_p \propto \left[\frac{\partial p}{\partial K} - \frac{p(K)}{\alpha F_t}\right]$, is always lower than the corresponding probability under the risk neutral measure, $\alpha_q \propto \frac{\partial p}{\partial K}$. As a result, at a given risk level the identified option under the physical measure has a higher strike compared to the one identified under the risk neutral one, thus making the VaR and CVaR estimates under the physical measure less conservative. We test the above assertion by comparing the option-implied VaR with the statistically-based VaR calibrated on an asymmetric Glosten, Jagannathan and Runkle (1993)[55] AR(1) - GARCH model using FHS innovations.

To capture the clustering and the heteroskedastic feature shared by the market returns, an asymmetric autoregressive model with filtered historical simulation (FHS) innovations is fitted to the historical log-returns of the S&P 500 index. Starting from January 20, 2005 the estimation window goes back up to 10 years of daily observations (January 2, 1996 for a total of 2279 daily observations, see figure 11)[88]. The GJR-GARCH model is defined as:

$$\log \frac{S_t}{S_{t-1}} = r_t = a + \rho r_{t-1} + \epsilon_t \tag{45}$$

$$\epsilon_t = \sqrt{\sigma_t^2} z_t \tag{46}$$

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \epsilon_{t-1}^2 + \gamma \mathbb{I}_{\epsilon_{t-1}<0} \epsilon_{t-1}^2$$

where $\mathbb{I}_{\epsilon_t - 1<0}$ describes the indicator function accounting for the leverage effect (Black (1976)[19]). Table 20 shows summary statistics on the estimated physical parameters $\theta_t = (\omega_t, \alpha_t, \beta_t, \gamma_t)$ obtained by performing a Gaussian Pseudo Maximum Likelihood (PML) estimation. Among the wide family of autoregressive specifications, the choice of using an AR(1) GJR GARCH (1,1) model is motivated by its capacity of fitting the S&P 500 data well (see Ghysels et al.(1996)[53] and Christoffersen and Jacobs (2004)[32]) and by the results of the Lagrange Multiplier ARCH test by Engle (1984)[48][89]. With no loss of generality, the same approach can be also performed under other

[88]The Portmanteau test of Ljung and Box and the Lagrange Multiplier (MA) ARCH test present evidence that the proposed model and the number of observations used are robust.
[89]Results are omitted but available upon request.

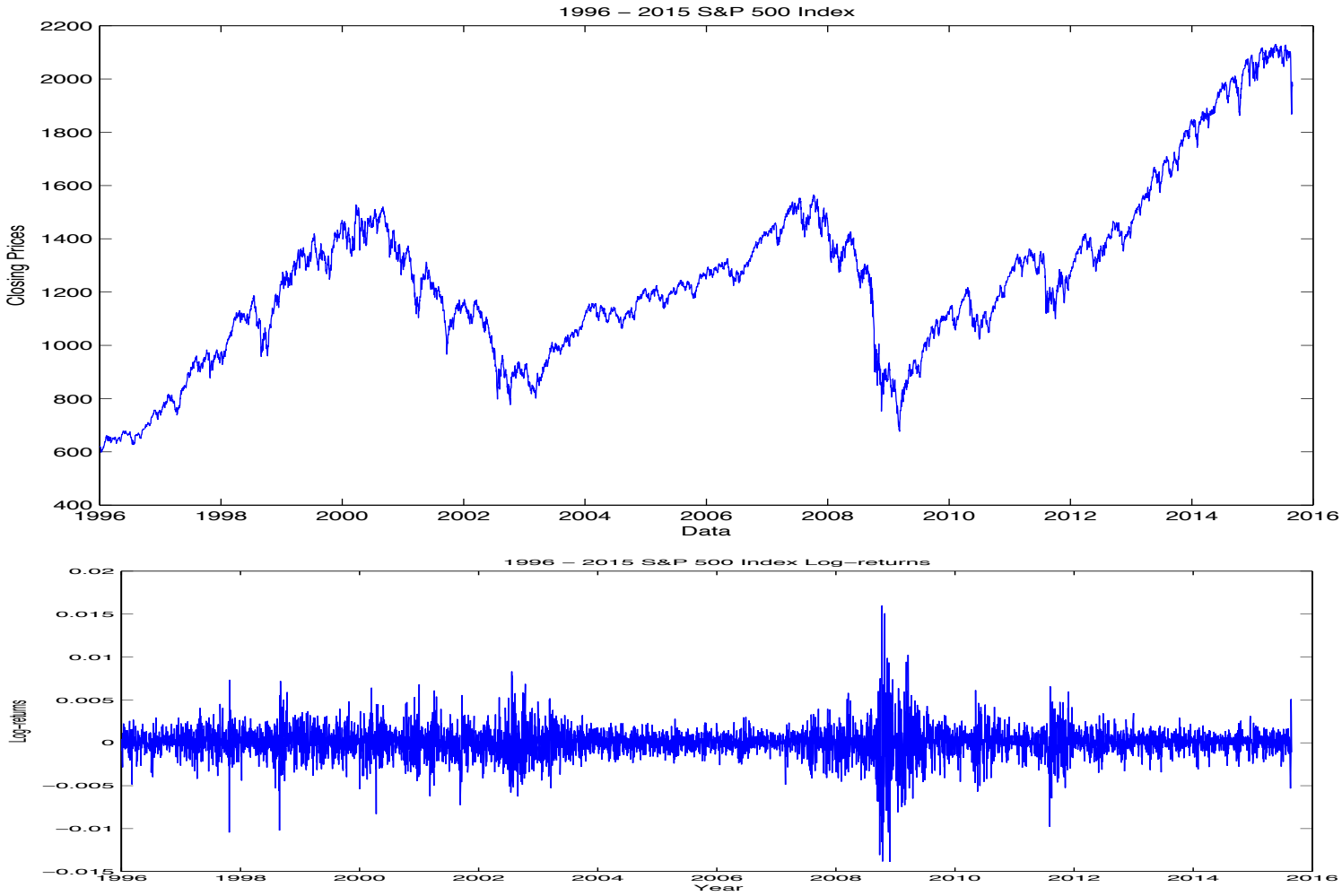

Figure 11: Daily closing prices and log returns of S&P 500 Index over the period 1996-2015.

ARCH specifications. From the set of estimated physical parameters, i[90] log-returns are simulated:

$$\hat{S}_{i,t} = S_t \exp\left(\text{Drift} - \frac{\hat{\sigma}^2_{t,\text{sim}}}{2}\right) dt + \hat{\sigma}^2_{t,\text{sim}} dW_t \tag{47}$$

where S_t represents the S&P 500 price at day t, the drift input accounts for the risk premium, $\hat{\sigma}^2_{t,\text{sim}}$ is the simulated stochastic variance, dt is fixed at one day and dW_t represents the canonical Brownian motion. As anticipated above, the physical measure strongly depends on the risk premium estimation, which directly impacts the drift. Following Merton (1980)[84], the drift is set at daily fixed percentages, namely 4%, 6% or 8%. From the estimated log-returns the simulated probability density functions (pdf) and cumulative density functions (cdf) are extracted by means

[90]Where i represents the number of simulations. The choice of this parameter is subjective and constrained by the available computing power.

Year	ω	α	β	γ	Persistence
2005	2.43e-06	2.20e-13	0.91	0.15	0.98
2006	1.56e-06	1.16e-13	0.92	0.15	0.99
2007	1.16e-06	3.35e-10	0.93	0.12	0.99
2008	1.24e-06	1.24e-13	0.93	0.12	0.99
2009	1.21e-06	1.37e-13	0.93	0.13	0.99
2010	1.15e-06	1.24e-13	0.93	0.12	0.99
2011	1.25e-06	1.10e-13	0.93	0.12	0.99
2012	1.68e-06	2.45e-10	0.91	0.14	0.98
2013	1.85e-06	9.54e-14	0.91	0.15	0.98
2014	2.08e-06	1.43e-13	0.90	0.17	0.98
2015	2.37e-06	2.45e-13	0.90	0.15	0.98

Table 20: Yearly average values of the physical GARCH parameters $\theta_t = f(\omega, \alpha, \beta, \gamma)$ calibrated each third Thursday of the month over the sample period from January 2005 to December 2015.

of kernel regressions[91]. Over the sample period the distance between CVaR and VaR is more stable than under the option-implied approaches, thus denoting a lower sensibility for changing market scenarios. Indeed, being linked to a stream of past returns, the empirical innovations and the volatility of the asymmetric GARCH model slowly react to market changes.

8 Backtesting

To test the validity of our estimates we perform different types of backtests. Backtesting refers to the validation of an estimated risk measure based on realized losses. The chosen backtesting measures for the VaR are: the number of times that the options expire in the money, the Binomial test (Bin), the Traffic Light (TL) test, the Proportion Of Failure (POF) test, the Time Until First Failure (TUFF) test, the Conditional Coverage Independence (CCI) test, the Conditional Coverage (CC) test, the Time Between Failure Independence (TBFI) test and the Time Between Failure (TBF) test.

At a specified α level, $\sum_{t=1}^{N} I_{[S_T<K_{\alpha,t}]}$ and $\sum_{t=1}^{N} I_{[S_T>K_{\alpha,t}]}$ deliver the number of option contracts expiring ITM based on the left and right tail of the distribution, respectively. where S_T is the price of the underlying asset at the maturity of the option contract (i.e. $T = 7, 30$ days for weekly and monthly estimates, respectively), $K_{\alpha,t}$ is the strike price of the option contract identified by the risk level α set at date t, $I_{[.]}$ is the indicator function and N=128 is the total number of observation

[91]Extra details concerning the estimation in Barone Adesi et al. (2008)[9] and Sala et al. (2016)[92].

dates. For all tests: R means rejection and A acceptance. The only exception is the output of the TL test which has three possible outcomes (details follow). The Binomial test of Jorion (2011)[68] compares the observed with the expected number of exceedances through the following test statistic:

$$\text{Bin} = \frac{I - Np}{\sqrt{Np(1-p)}} \tag{48}$$

where I represents the number of exceedances observed, p is a positive value between 0 and 1 and is defined as $p = 1 - \alpha$, and N is the total number of observations.

Proposed by the Basel Commitee (1996)[86], the Traffic Light test compares the number of observed and expected exceedances. Given a number of exceedances, I, this test computes and ranks the probability of observing up to I exceedances. While the Binomial test - as well as all other tests - only gives two mutually exclusive results, this test classifies risk into three regions: red, yellow and green (R, Y and G respectively[92]).

The POF and TUFF refer to the Proportion Of Failure and the Time Until First Failure tests of Kupiec (1995)[73]. Starting from the Binomial test and adding a likelihood ratio test, the POF test suggests that the VaR estimates can be rejected if the likelihood ratio exceeds a given critical value. Asymptotically distributed as a chi square with one degree of freedom, the POF test statistic is defined as:

$$POF = -2\text{Log}\left(\frac{(1-p)^{N-I}p^I}{\left(1-\frac{I}{N}\right)^{N-I}\left(\frac{I}{N}\right)^I}\right) \tag{49}$$

Still based on the likelihood ratio - but with a geometric distribution - the TUFF test checks when the first exceedance has occurred through the following statistic:

$$TUFF = -2\text{Log}\left(\frac{p(1-p)^{d-1}}{\left(\frac{1}{d}\right)\left(1-\frac{1}{d}\right)^{d-1}}\right) \tag{50}$$

where d accounts for the number of days until the first rejection. Also this statistic is asymptotically distributed as a Chi squared distribution with one degree of freedom.

The CCI and CC refer to the Conditional Coverage Independence and the Conditional Coverage of Christofferesen (1998)[29]. Both tests check if the observed failures are among them dependent between consecutive days. Asymptotically distributed as a χ^2 with one degree of freedom, the CCI

[92]The red region is linked to a number of exceedances whose probability goes over 99.9%, the probability of the yellow are goes from 99.9% up to 95% while the green area refers to all values below the yellow area.

test statistic is defined as:

$$CCI = -2\text{Log}\left(\frac{(1-\pi_2)^{d_{nn}+d_{yn}}\,\pi_2^{d_{ny}+d_{yy}}}{(1-\pi_0)^{d_{nn}}\pi_0^{d_{ny}}(1-\pi_1)^{d_{yn}}\pi_1^{d_{yy}}}\right) \tag{51}$$

where all powers account for the different combinations of d days with no failures, n, or with failures, y, thus producing four possible combinations. The probabilities are so defined:

- π_0: probability of having a time t failure given no $t-1$ failure: $\pi_0 = \frac{d_{ny}}{(d_{nn}+d_{ny})}$
- π_1: probability of having a time t failure given a $t-1$ failure: $\pi_1 = \frac{d_{yy}}{(d_{yn}+d_{yy})}$
- π_2: probability of having a time t failure: $\pi_2 = \frac{(d_{ny}+d_{yy})}{(d_{nn}+d_{ny}+d_{yn}+d_{yy})}$

The CC test is defined as:

$$CC = POF + CCI \tag{52}$$

and is asymptotically distributed as a chi square with two degrees of freedom.
The TBFI and TBF refer to the Time Between Failure Independence and the Time Between Failure of Haas (2001)[61]. While the TUFF checks only for the first exceedance, thus possibly leaving important information out of the analysis, the TBFI of Haas (2001)[61] considers all exceedances. Asymptotically distributed as a chi square with I degrees of freedom, the TBFI test is defined as:

$$TBFI = -2\sum_{i=1}^{I}\text{Log}\left(\frac{p(1-p)^{d_i-1}}{\left(\frac{1}{d_i}\right)\left(1-\frac{1}{d_i}\right)^{d_i-1}}\right) \tag{53}$$

As a crucial modification with respect to the TUFF test, d is enriched with the subscript i, which accounts for the number of exceedances between i and $i-1$. As for the CC test, the POF can be used jointly with the TBFI test to define the TBF statistic:

$$TBF = POF + TBFI \tag{54}$$

which is asymptotically distributed as a chi square with $I+1$ degrees of freedom.

As concerns the CVaR, three backtests are performed. First, the average predicted excess loss

beyond the VaR based on the left (eq. (55)) and right (eq. (56)) tail of the distribution:

$$L_{exp} = \frac{1}{N}\sum_{t=1}^{N}\left[\text{CVaR}_{\alpha,t} - \text{VaR}_{\alpha,t}\right] I_{[S_{t+T}<K_{\alpha,t}]} \qquad \text{Left tail} \tag{55}$$

$$L_{exp} = \frac{1}{N}\sum_{t=1}^{N}\left[\text{CVaR}_{\alpha,t} - \text{VaR}_{\alpha,t}\right] I_{[K_{\alpha,t}>S_{t+T}]} \qquad \text{Right tail} \tag{56}$$

is compared with the average realized loss conditional on the observation of a VaR exceedance, $L_{real} = \frac{1}{N}\sum_{t=1}^{N}\left[-(S_T - S_t) - \text{VaR}_{\alpha,t}\right] I_{[S_{t+T}<K_{\alpha,t}]}$. When computing the expected excess loss, the choice of performing a conditional test has the objective to remove the effect of variation and estimation error in the VaR through time. The mean conditional expected excess loss is compared with the mean conditional realized loss exceeding the VaR by conducting a nonparametric MannWhitney U[93] test (Mann and Whitney (1947)[78]) on the difference between the two means. If the null hypothesis of a zero difference between the two means cannot be rejected, the option-implied CVaR forecasts future losses in a consistent way.

Second, following McNeil et al. (2015)[81], the null of a correctly calibrated CVaR is tested against the alternative of a risk underestimation through the following calibration function:

$$K_t = h_1^{(\alpha)}(VaR_{\alpha,t}, CVaR_{\alpha,t}, L_t) = \left(\frac{L_t - CVaR_{\alpha,t}}{CVaR_{\alpha,t}}\right) \cdot I_t \tag{57}$$

where $\text{VaR}_{\alpha,t}$ and $\text{CVaR}_{\alpha,t}$ are the time t VaR and CVaR at the risk level α, L_t is the realized loss and I_t is the indicator function of a VaR exceedance. Under the null, the expected value of K_t conditional to the information up to time $(t-1)$, $\mathbb{E}(K_t|\mathcal{F}_{t-1})$, is equal to zero. Under the alternative, $\mathbb{E}(K_t|\mathcal{F}_{t-1})$ is strictly positive, thus suggesting that the CVaR has a negative bias.

Third, since the VaR and the CVaR are jointly elicitable (see Lambert (2008)[74] and Fissler et al. (2015)[49]), the joint calibration function:

$$S_t = h_2^{(\alpha)}(VaR_{\alpha,t}, CVaR_{\alpha,t}, L_t) = \frac{I_t}{CVaR_{\alpha,t}} - \alpha \tag{58}$$

enables to test the null hypothesis of a correct joint calibration of the VaR and CVaR. Rejection of the null implies that the VaR and/or the CVaR underestimate risks. As for K_t, S_t is expected to behave as a martingale difference (i.e. $\mathbb{E}(K_t|\mathcal{F}_{t-1}) = 0$) under the null and to be strictly positive

[93] In risk management, forecasted losses smaller/larger than the realized ones are both undesirable. While the former ones imply a risk underestimation and hence an inadequate capital protection, the latter ones induce an inefficient allocation of the resources.

under the alternative. Table 21 presents basic results for the nine aforementioned VaR tests (Bin, TL, POF, TUFF, CCI, CC, TBFI and TBF). Based on both tails of the distribution, the upper and lower parts of the table report the acceptance/rejection results for the weekly and monthly VaR, respectively. As shown in the upper part of table 21, the weekly MF VaR based on the

Weekly VaR estimates									
Model	α	Bin	TL	POF	TUFF	CCI	CC	TBFI	TBF
MF	1%	A/R	G/R	A/R	A/A	A/A	A/R	A/R	A/A
MF	2.5%	A/R	G/Y	A/R	A/A	A/A	A/R	A/A	A/A
MF	5%	A/A	G/G	A/A	A/A	A/A	A/A	A/A	A/A
Monthly VaR estimates									
Model	α	Bin	TL	POF	TUFF	CCI	CC	TBFI	TBF
MF	5%	A/A	G/G	A/A	A/A	A/A	A/A	A/A	A/A
MF	10%	R/A	G/G	R/A	A/A	A/A	A/A	A/A	A/A
MF	15%	A/A	G/G	A/A	A/A	A/A	A/A	A/A	A/A

Table 21: Backtesting results of the Model Free (MF) VaR computed for the left (right) tail using put (call) options.The upper (lower) part of the table refers to the weekly (monthly) estimates. On both horizons, the first/second values refer to the test results on the MF VaR based on the left/right tail of the distribution.

left tail of the distribution perfectly performs across all the proposed tests, thus showing that the MF approach delivers accurate results over short investment horizons and at multiple risk levels, namely 1%, 2.5% and 5%. To further investigate over these results, tables 22, 23 and 24 present detailed statistics on all the proposed tests. In all cases the corresponding p-values are almost

MF VaR	Traffic Light	Probability	Type I	Increase
$\alpha = 5\%$	Green	0.0981	0.95	0.00
$\alpha = 2.5\%$	Green	0.2473	0.91	0.00
$\alpha = 1\%$	Green	0.2112	1.00	0.00

Table 22: Detailed backtesting results of the Traffic Light test for the MF VaR based on the left tail of the distribution at short (weekly) time horizon.

always substantially larger that the threshold level 5%, thus corroborating the soundness of the MF estimates. When α=5% two exceedances occur at close dates at the end of 2014, thus potentially undermining the time independence among the VaR exceedances. Nevertheless, the detailed results of the independence (CCI) test (see first three rows of table 24) show that the p-values (4^{th} column) are always larger than 5%. As a consequence, the null of time-independent exceedances cannot be rejected at 95% confidence level. The upper left part of table 25 presents the backtesting results on

MF VaR	Bin	Z Score	P Value	Test Level
$\alpha = 5\%$	Accept	-1.4236	0.077281	0.95
$\alpha = 2.5\%$	Accept	-0.9936	0.1602	0.95
$\alpha = 1\%$	Accept	-1.2553	0.1047	0.95
MF VaR	POF	Likelihood Ratio	P Value	Test Level
$\alpha = 5\%$	Accept	2.4559	0.11708	0.95
$\alpha = 2.5\%$	Accept	1.2023	0.27286	0.95
$\alpha = 1\%$	Accept	3.1357	0.0766	0.95
MF VaR	TUFF	Likelihood Ratio	P Value	Test Level
$\alpha = 5\%$	Accept	3.5780	0.0585	0.95
$\alpha = 2.5\%$	Accept	1.8368	0.1753	0.95
$\alpha = 1\%$	Accept	–	–	0.95
MF VaR	CC	Likelihood Ratio	P Value	Test Level
$\alpha = 5\%$	Accept	2.6651	0.2638	0.95
$\alpha = 2.5\%$	Accept	1.2539	0.5342	0.95
$\alpha = 1\%$	Accept	3.1357	0.2085	0.95
MF VaR	POF	Likelihood Ratio	P Value	Test Level
$\alpha = 5\%$	Accept	2.4559	0.1171	0.95
$\alpha = 2.5\%$	Accept	1.2023	0.2728	0.95
$\alpha = 1\%$	Accept	3.1357	0.0766	0.95

Table 23: Detailed backtesting results of the Bin, TUFF, CC an POF tests for the MF VaR based on the left tail of the distribution at short (weekly) time horizon.

MF VaR	CCI	Likelihood Ratio	P Value	N00	N10	N01	N11	Test Level
$\alpha = 5\%$	Accept	0.2091	0.6474	149	4	4	0	0.95
$\alpha = 2.5\%$	Accept	0.0516	0.8203	158	2	2	0	0.95
$\alpha = 1\%$	Accept	0	1	155	0	0	0	0.95
MF VaR	TBFI	Likelihood Ratio	P Value	Min/Max	Q1	Q2	Q3	Test Level
$\alpha = 5\%$	Accept	6.7574	0.1493	3/83	18	35	60	0.95
$\alpha = 2.5\%$	Accept	1.8478	0.39697	36/120	36	78	120	0.95
$\alpha = 1\%$	Accept	–	–	–/–	–	–	–	0.95

Table 24: Detailed backtesting results of the CCI and TBFI tests for the MF VaR based on the left tail of the distribution at short (weekly) time horizon.

the weekly MF CVaR based on the left tail of the distribution. Being conditional on the occurrence of a VaR exceedance, the p-values of the Mann-Whitney test (ρ_{MW}) and of the two tests proposed by McNeil (ρ_K and ρ_S) are not available at $\alpha = 1\%$, as no VaR exceedances occur at the lowest risk level. At $\alpha = 2.5\%$ and 5%, the results of all tests suggest that the MF CVaR estimates provide good forecasts of future risks.

Left tail	Weekly Model-Free					Weekly Black and Scholes				
	% I_t	L_{exp}	L_{real}	ρ_K	ρ_S	% I_t	L_{exp}	L_{real}	ρ_K	ρ_S
α=1%	0.00	–	–	–	–	0.00	–	–	–	–
α=2.5%	1.26	26.77	10.89	0.88	1.00	0.00	–	–	–	–
α=5%	2.52	29.61	21.16	0.85	1.00	0.63	19.46	44.11	0.10	1.00
Right tail	Weekly Model-Free					Weekly Black and Scholes				
	% I_t	L_{exp}	L_{real}	ρ_K	ρ_S	% I_t	L_{exp}	L_{real}	ρ_K	ρ_S
α=1%	6.29	23.37	16.54	0.66	0.01	1.26	6.12	29.01	0.13	0.23
α=2.5%	6.29	12.51	24.04	0.08	0.03	5.03	26.33	17.09	0.85	0.12
α=5%	5.66	10.67	21.64	0.03	0.18	6.92	39.91	19.79	0.98	0.41

Left Tail	Monthly Model-Free					Monthly Black and Scholes				
	% I_t	L_{exp}	L_{real}	ρ_K	ρ_S	% I_t	L_{exp}	L_{real}	ρ_K	ρ_S
$\alpha = 5\%$	1.56	43.22	46.26	0.35	1.00	0.78	38.84	46.81	0.11	1.00
$\alpha = 10\%$	4.68	50.90	46.77	0.43	1.00	2.34	44.32	52.38	0.14	1.00
$\alpha = 15\%$	6.25	55.99	30.42	0.92	1.00	7.03	43.75	41.34	0.41	1.00
Right Tail	Monthly Model-Free					Monthly Black and Scholes				
	% I_t	L_{exp}	L_{real}	ρ_K	ρ_S	% I_t	L_{exp}	L_{real}	ρ_K	ρ_S
$\alpha = 5\%$	2.34	30.07	55.96	0.87	1.00	3.91	22.30	26.11	0.12	0.72
$\alpha = 10\%$	4.68	26.92	41.73	0.96	1.00	5.46	24.73	57.14	0.15	1.00
$\alpha = 15\%$	6.25	23.09	37.79	0.88	1.00	7.81	28.93	55.86	0.09	0.96

Table 25: The upper portion of the table reports the backtesting results of the weekly option-implied VaR and CVaR under the MF and B&S methodology at $\alpha = 1\%, 2.5\%, 5\%$ based on the left and right tail of the distribution. The lower portion of the table reports the backtesting results of the monthly option-implied VaR and CVaR under the MF and B&S methodology at $\alpha = 5\%, 10\%, 15\%$ based on the left and right tail of the distribution.

On a monthly basis, figure 12 indicates with black stars the dates on which a monthly VaR exceedance occurs in the left tail under the model-free methodology. Tables 26, 27 and 28 present detailed statistics on the tests proposed in table 21. At a monthly horizon, even taking into account the number of missing data and for both tails, the proportion of exceedances is always below the corresponding α level (see the first three rows in the 1^{st} column of table 25). For $\alpha = 5\%, 10\%, 15\%$, the

MF VaR	Traffic Light	Probability	Type I	Increase
$\alpha = 5\%$	Green	0.0666	0.9815	0
$\alpha = 10\%$	Green	0.0195	0.9929	0
$\alpha = 15\%$	Green	0.0603	0.9686	0

Table 26: Detailed backtesting results for the Traffic Light test for the MF VaR based on the left tail of the distribution at long (monthly) time horizon.

proportion of exceedances equals $1.56\%, 4.68\%$ and 6.25% for the left tail and $2.34\%, 4.58\%, 6.25\%$

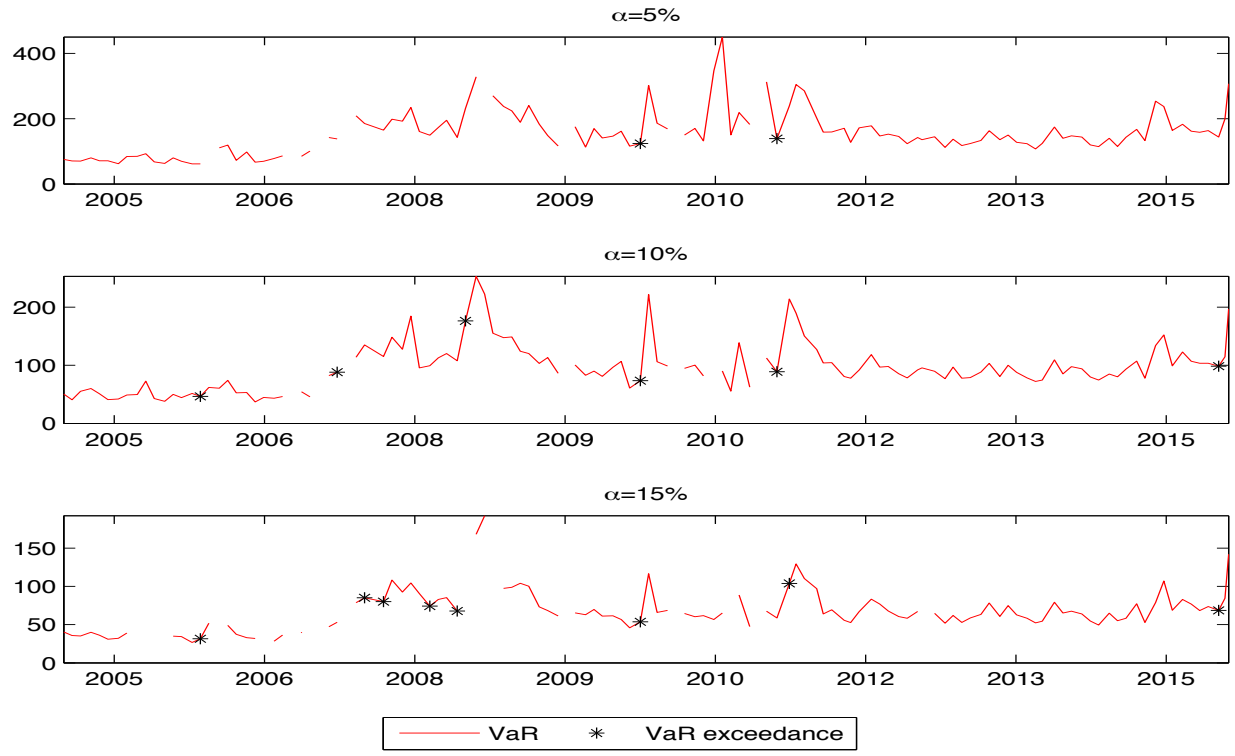

Figure 12: VaR exceedances on the left tail monthly MF VaR at α=5%, 10% and 15%.

for the right one. The proposed VaR risk measures are then conservative forecasts of the potential losses. This result is reflected in the p-values of the Binomial and POF tests (see the first six rows in the 4^{th} column of table 27). The discrepancy between the proportion of exceedances and their corresponding theoretical values drives down the p-values of the Binomial and POF tests, whose null hypothesis are rejected when the risk level is 10%. Conversely, at $\alpha = 5\%$ and 15% the difference between the proportion of realized and theoretical exceedances is not statistically significant and the null of correct coverage cannot be rejected at 95% confidence level. According to all the other tests and across each risk level, the monthly MF VaR correctly predicts risk. The lower left part of table 25 presents the backtesting results on the monthly MF CVaR based on the left tail of the distribution. When α equals 5% and 10%, the null hypothesis cannot be rejected at 99% confidence level. The estimated CVaR thus provide an adequate capital protection. Conversely, at $\alpha = 15\%$ the null is rejected. Being a two-tailed test, the MannWhitney U test does not provide information whether the CVaR is under/over-estimating the actual losses. At $\alpha = 15\%$, the difference between the sample means of the expected and the realized losses is largely positive (i.e. 25.57\$). This result implies that the option-implied CVaR is a conservative risk metric and is overestimating future losses at 95% confidence level. The right upper part of table 25 reports the

MF VaR	Bin	Z Score	P Value	Test Level
$\alpha = 5\%$	Accept	-1.6189	0.05274	0.95
$\alpha = 10\%$	Reject	-2.0647	0.01947	0.95
$\alpha = 15\%$	Accept	-1.638	0.05071	0.95
MF VaR	POF	Likelihood Ratio	P Value	Test Level
$\alpha = 5\%$	Accept	3.4707	0.0624	0.95
$\alpha = 10\%$	Reject	5.3160	0.0211	0.95
$\alpha = 15\%$	Accept	3.0313	0.0816	0.95
MF VaR	TUFF	Likelihood Ratio	P Value	Test Level
$\alpha = 5\%$	Accept	1.6679	0.1965	0.95
$\alpha = 10\%$	Accept	0.2846	0.59367	0.95
$\alpha = 15\%$	Accept	0.6252	0.4291	0.95
MF VaR	CC	Likelihood Ratio	P Value	Test Level
$\alpha = 5\%$	Accept	3.5415	0.1702	0.95
$\alpha = 10\%$	Accept	5.7666	0.0559	0.95
$\alpha = 15\%$	Accept	4.1933	0.1228	0.95
MF VaR	POF	Likelihood Ratio	P Value	Test Level
$\alpha = 5\%$	Accept	3.4707	0.0625	0.95
$\alpha = 10\%$	Accept	5.3160	0.0211	0.95
$\alpha = 15\%$	Accept	3.0313	0.0816	0.95

Table 27: Detailed backtesting results of the Bin, TUFF, CC an POF tests for the MF VaR based on the left tail of the distribution at long (monthly) time horizon.

MF VaR	CCI	Likelihood Ratio	P Value	N00	N10	N01	N11	Test Level
$\alpha = 5\%$	Accept	0.0708	0.7902	111	2	2	0	0.95
$\alpha = 10\%$	Accept	0.4506	0.5020	106	5	5	0	0.95
$\alpha = 15\%$	Accept	1.1620	0.2811	88	8	8	2	0.95
MF VaR	TBFI	Likelihood Ratio	P Value	Min/Max	Q1	Q2	Q3	Test Level
$\alpha = 5\%$	Accept	1.8394	0.3986	13/57	13	35	57	0.95
$\alpha = 10\%$	Accept	0.9089	0.9695	12/18	12	13	16.5	0.95
$\alpha = 15\%$	Accept	12.4840	0.254	1/17	2	4	10	0.95

Table 28: Detailed backtesting results of the CCI and TBFI tests for the MF VaR based on the left tail of the distribution at long (monthly) time horizon.

backtesting results of the weekly B&S CVaR based on the left and right tails. At $\alpha = 1\%$ and 2.5%, no exceedances occur; hence, the conditional CVaR backtests deliver no results. At $\alpha = 5\%$, the proportion of the exceedances (0.63%) is much lower than its theoretical value (5%). The expected excess losses are substantially lower than the realized ones, thus suggesting that the B&S CVaR may underestimate risk. Nevertheless, being this conclusion based on just one VaR exceedance, the

statistical power is not sufficiently high to assess that in general the weekly B&S CVaR forecasts risks in an inadequate way.

The right lower part of table 25 reports the same backtesting statistics for the B&S CVaR and for both tails of the distribution at a monthly and weekly investment horizon. On a monthly horizon, figure 13 graphically shows when the VaR exceedances occur in the left tail under the B&S methodology. For the left tail of the distribution and for $\alpha = 5\%$ and 10%, the proportion of

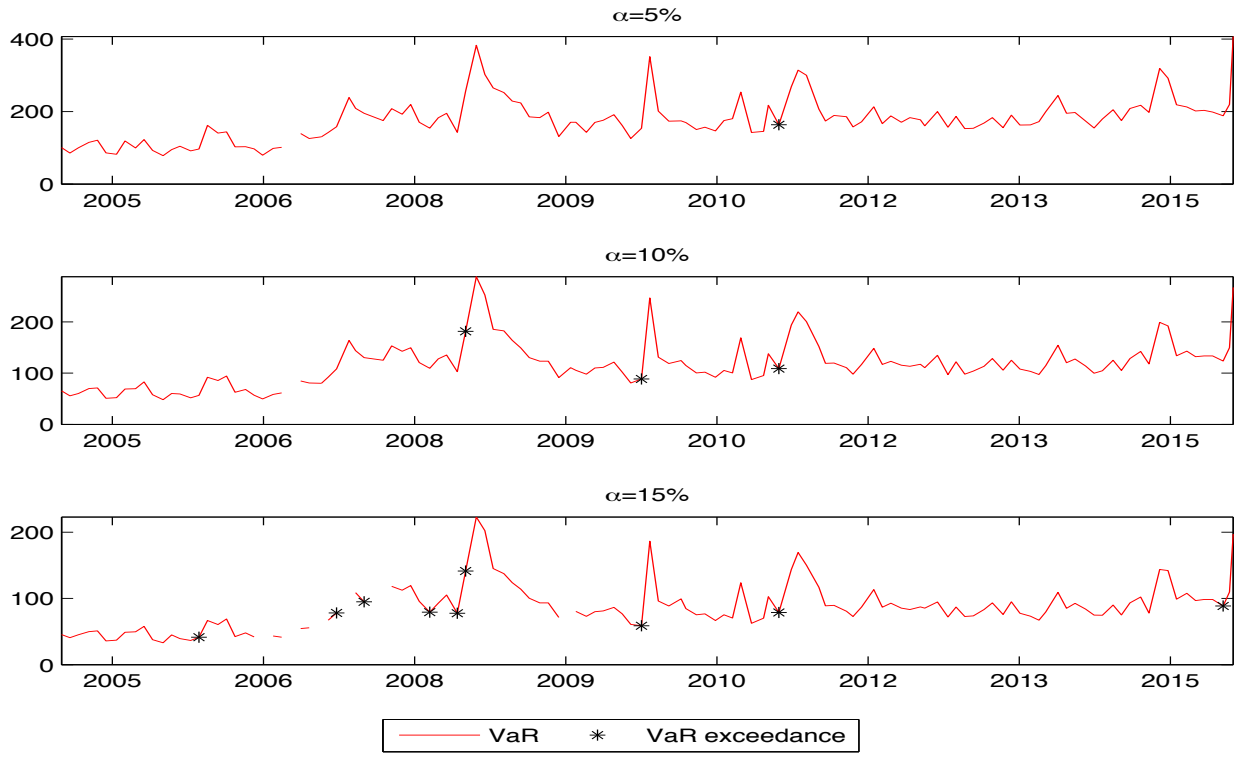

Figure 13: VaR exceedances on the left tail monthly B&S VaR at α=5% , 10% and 15%.

exceedances (upper part of 6^{th} column) is always lower under the B&S assumption than the MF approach. Based on the right tail and at each risk level, the proportion of VaR exceedances (middle part of 6^{th} column) under the B&S approach is always larger. The B&S methodology then delivers less conservative risk forecasts. The MannWhitney U test based on the left tail of the distribution always rejects the null hypothesis at 95% confidence level. Compared to the MF methodology and for $\alpha = 15\%$, the difference between the expected and the realized losses is close enough to zero (i.e. 2.40\$) to fail to reject the null. Based on the right tail of the distribution, the null of a correctly specified risk metric cannot be rejected at any risk level. When performing the tests proposed by McNeil et al. (2015)[81], ρ_K and ρ_S (see 9^{th} and 10^{th} columns) are substantially lower than those obtained under the MF methodology. Nevertheless, for both tails of the distribution at 95%

confidence level, the p-values suggest that the B&S CVaR and the B&S VaR and CVaR jointly do not underestimate risk.

To conclude the analysis we present the backtesting results for the statistically-based measures. While the list of risk estimates that can be classified under the name of statistically-based historical models is huge, as a benchmark model we consider the Filtered Historical Simulation (FHS) approach presented in section 7. Detaching from the parametric approach of the Gaussian model, and trying to combine the main advantages of the Historical Simulation, the EWMA and different Monte Carlo approaches, the FHS has emerged as the leading model to detect the short time risk on the S&P 500 [94]. Table 29 presents backtesting results for the VaR estimated using the FHS approach at weekly (left) and monthly (right) level. Figure 14 graphically plots the VaR exceedances over the sample period.

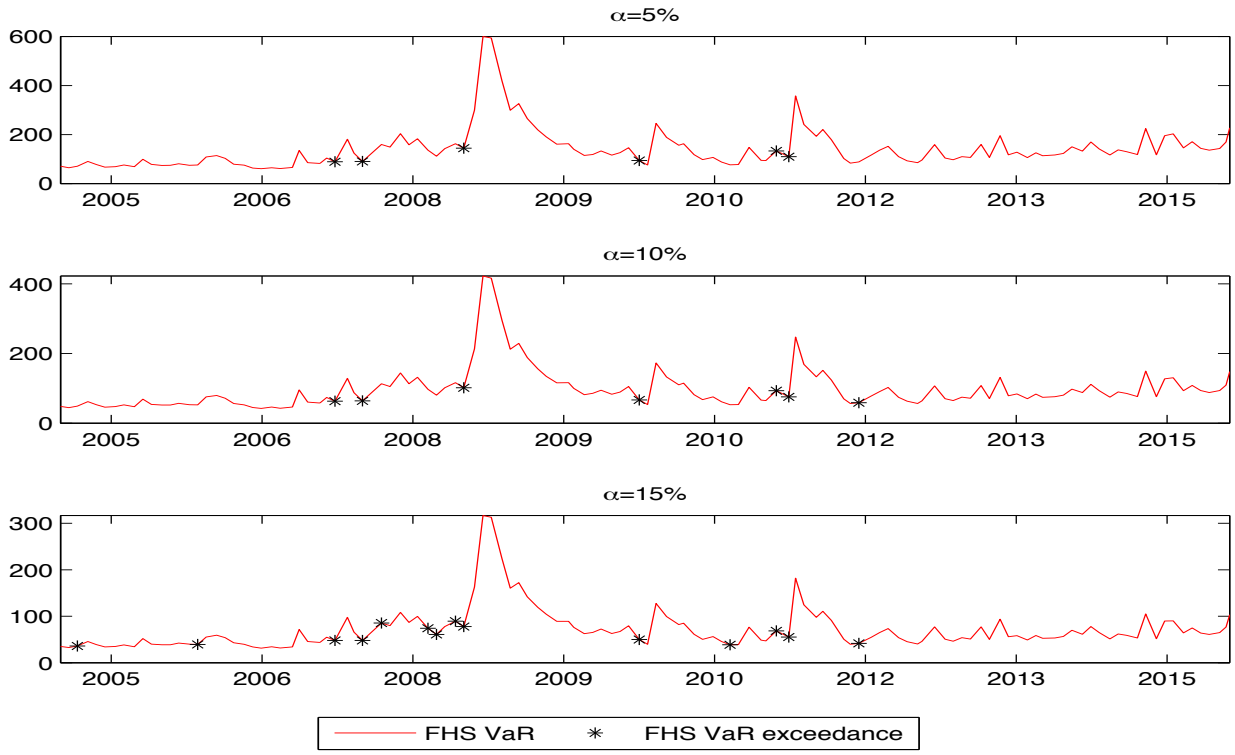

Figure 14: VaR exceedances on the monthly FHS VaR at α=5%, 10% and 15%.

Confirming Brownlees at al. (2011)[25] the risk measures estimated with the FHS approach work well for almost all tests in consideration. There are in fact no rejections at the weekly horizon.

[94] As already mentioned in the introduction, we performed many other test with other statistical-based models. Confirming the literature most of these models fail in detecting the risk at different confidence levels and above all at long time horizons. Results are here omitted for reason of space.

Things change at longer time horizons where the monthly VaR are rejected at 5% for the POF test and at 15% for the TBF. This evidence shows the ability of the option-implied approach for long time horizons forecasting, where the speed of reaction of a GARCH based volatility could be sometimes suboptimal. As it concerns the VaR, at $\alpha = 5\%$ and 15%, the proportion of the exceedances (see the 1^{st} column of table 29) is slightly higher compared to the VaR backtesting results in the left tail under the option-implied methodology, both when using the MF and the B&S approaches (see 1^{st} and 6^{th} columns of table 25). The VaR risk forecasts are rejected in just

Weekly VaR estimates									
Model	α	Bin	TL	POF	TUFF	CCI	CC	TBFI	TBF
FHS	1%	A	G	A	A	A	A	A	A
FHS	2.5%	A	G	A	A	A	A	A	A
FHS	5%	A	G	A	A	A	A	A	A
Monthly VaR estimates									
FHS	5%	A	G	R	A	A	A	A	A
FHS	10%	A	G	A	A	A	A	A	R
FHS	15%	A	G	A	A	A	A	A	A

Table 29: Backtesting results of the weekly (upper part) and monthly (lower part) FHS VaR under the physical measure.

two cases (at $\alpha = 5\%$ the null of the POF test is rejected and at $\alpha = 10\%$ the null of the TBF test is rejected). Besides to the backtesting approaches already presented, the statistically-based CVaR is further backtested by using the Z-statistic proposed by Acerbi and Szekely (2014)[1][95].
At a specified α level, the Z-statistic equals:

$$Z(\overrightarrow{X}) = \sum_{t=1}^{N} \frac{X_t I_t}{N\alpha CVaR_{\alpha,t}} + 1 \tag{59}$$

where X_t is the time t realized loss, I_t is the indicator function of a VaR exceedance, α is the risk level, $\text{CVaR}_{\alpha,t}$ is the time t CVaR simulated from the GJR-GARCH model with FHS innovations and N is the total number of observation dates. The proposed statistic does not test directly the CVaR, but it is subordinated to a preliminary (positive) VaR test[96]. Under the null hypothesis that the forecasting model correctly predicts the risk, the Z-statistic is equal to zero. When the alternative is true the expected value of Z is negative and the CVaR underestimates risk[97].

[95] See eq.(5) pp.3 of Acerbi and Szekely (2014)[1]
[96] To compute a backtesting measure for the CVaR at least one VaR exceedance must occur.
[97] Being a one-sided test, it does not enable to detect a risk overestimation.

Given the high stability of Z across a wide variety of tail distributions[98] (see Acerbi and Szekely (2014)[1]), the cut-off value at 99.9% confidence level can be considered fixed at 1.8. Table 30 reports the backtesting statistics for the weekly (upper part) and monthly (lower part) FHS CVaR at each risk level. At both investment horizons, since the value of the Z-statistics is always above

Weekly FHS CVaR					Monthly FHS CVaR				
α	% I_t	Acerbi Z-stat.	ρ_K	ρ_S	α	% I_t	Acerbi Z-stat.	ρ_K	ρ_S
1%	0.63	1.80	0.82	0.63	5%	4.69	0.13	0.86	0.61
2.5%	1.26	1.91	0.77	1.00	10%	5.47	0.40	0.19	0.97
5%	2.52	1.92	0.81	1.00	15%	10.94	0.30	0.66	0.94

Table 30: Backtesting results of the weekly (left part) and monthly (right part) FHS CVaR under the physical measure.

the threshold -1.8, the null of correct FHS CVaR forecasts cannot be rejected at 99.9% confidence level. There is no clustering of the exceedances in our backtests, supporting the validity of the tested risk models with the possible exception of their VaR results at 15% during the crisis period, with the option-implied measures performing marginally better. The p-values of the tests on K_t and S_t (3^{rd} and 4^{th} columns) are always larger than 5%, thus suggesting that the FHS CVaR alone and the FHS VaR and CVaR jointly do not underestimate risk.

9 Conclusions

This article presents and compares the backtesting results of the short and long time horizons VaR and the CVaR computed on the S&P 500 Index and estimated under two different methodologies, namely the option-implied and the statistically-based approaches. For the long time horizons, the option-implied risk measures perform similarly to those obtained under the physical measure when modelling the dynamics of the underlying asset with an AR(1) - GARCH(1,1) model with FHS innovations. Option-implied risk measures are quick and easy to implement and are naturally forward-looking. As a drawback, they are fully data driven. At a monthly horizon, findings are of particular interest over the 2007-2009 financial crisis period. While almost all statistically-based measures did not always provide an adequate capital protection, the proposed option-implied measures delivered marginally better results. At a weekly horizon, the option-implied risk estimates forecast risk in an accurate way, thus being valid alternatives to statistically-based risk metrics.

[98] Namely, Gaussian and t-Student with 3,5,10 and 100 degrees of freedom.

Being free of any modelling assumption and very simple to implement, the proposed approach can be useful to regulators, single companies and investors. The former ones can quantify the specific risk of a firm without needing to know what is on (and off) its balance sheet. Firms can either gain an overview of their perception among the investors and derive market-based risk metrics to be compared with those obtained by internal models. Finally, investors can assess and evaluate the risk of their existing and potential investments.

Bibliography

[1] C. Acerbi and B. Szekely. Back-testing expected shortfall. *Risk*, page 76, 2014.

[2] C. Acerbi and B. Szekely. Backtestable [slides]. 2015. Retrieved from https://www.cass.city.ac.uk/faculty-and-research/faculties/finance/seminars-and-workshops/financial-engineering-workshops/ACERBI-Carlo-10.03.2015.pdf.

[3] Y. Äit-Sahalia and A.W.Lo. Non parametric risk management and implied risk aversion. *Journal of Econometrics*, 94:9–51, 2000.

[4] T. W. Anderson and D. A. Darling. A test of goodness of fit. *Journal of the American statistical association*, 49(268):765–769, 1954.

[5] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risks. *Mathematical Finance*, 9:203–228, 1999.

[6] T. Bali, N. Cakici, and F. Chabi-Yo. A generalized measure of riskiness. *Management Science*, 57(8):1406–1423, August 2011.

[7] G. Barone-Adesi. Var and cvar implied in option prices. *Journal of Risk and Financial Management*, 9(1), 2016.

[8] G. Barone-Adesi and R. Elliot. Cutting the hedge. *Computational Economics*, 29(2):151–158, 2007.

[9] G. Barone-Adesi, R. Engle, and L. Mancini. A garch option pricing model in incomplete markets. *Review of Financial Studies*, 21(1223-1258), 2008.

[10] G. Barone-Adesi, R. F. Engle, and L. Mancini. A garch option pricing model with filtered historical simulation. *Review of Financial Studies*, 2008.

[11] G. Barone-Adesi, K. Giannopoulos, and Les Vosper. Backtesting derivative portfolios with filtered historical simulation (fhs). *Journal of future market*, 8(1):31–58, March 1999.

[12] G. Barone-Adesi, C. Legnazzi, and C. Sala. S&p 500 index, an option-implied risk analysis. 2016.

[13] G. Barone-Adesi, C. Legnazzi, and C. Sala. Wti crude oil option-implied var and cvar: An empirical application. 2016.

[14] Giovanni Barone-Adesi, Robert F. Engle, and Loriano Mancini. A garch option pricing model with filtered historical simulation. *Review of Financial Studies*, 2008.

[15] D.S. Bates. Testing option pricing models. *NBER Working Paper N.5129*, 1996.

[16] S. Bayer. Combining value-at-risk forecasts using penalized quantile regressions. *Working Paper*, 2017.

[17] F. Bellini and E. Di Bernardino. Risk management with expectiles. *The European Journal of Finance*, 23(6):487–506, 2017.

[18] J. Berkowitz. Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474, 2001.

[19] F. Black. Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association*, pages 171–181, 1976.

[20] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

[21] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, May - Jun 1973.

[22] J.C. Bogle. The mutual fund industry 60 years later: For better or worse? *Financial Analyst Journal*, 61:15–24, 2005.

[23] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[24] D. Breeden and R. Litzenberger. Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–651, October 1978.

[25] C. Brownlees, R. Engle, and B. Kelly. A practical guide to volatility forecasting through calm and storm. *The Journal of Risk*, 14(2):3–22, Winter 2011.

[26] M. A. Carlton. *Applications of the two-parameter Poisson-Dirichlet distribution.* PhD thesis, University of California Los Angeles, 1999.

[27] P. Carr, H. Geman, and D. Madan. Pricing and hedging in incomplete markets. *Journal of Financial Economics*, 62:131–168, 2001.

[28] F. Chabi-Yo, E. Ghysels, and E. Renault. Disentangling the effects of heterogeneous beliefs and preferences on asset prices. *Unpublished Working Paper. Ohio State University and University of North Carolina at Chapel Hill*, 2007.

[29] P. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39:841–862, 1998.

[30] P. Christoffersen. Evaluating interval forecasts. *International economic review*, pages 841–862, 1998b.

[31] P. Christoffersen, St. Heston, and K. Jacobs. Capturing option anomalies with a variance-dependent pricing kernel. *Review of Financial Studies*, 26(8):1963–2006, 2013.

[32] P. Christoffersen and K. Jacobs. Which garch model for option valuation? *Management Science*, 50(9):1024–1221, 2004.

[33] P. Christoffersen, K. Jacobs, and B. Li. Dynamic jump intensities and risk premiums in crude oil futures and options markets. *The Journal of Derivatives*, 24(2):8–30, 2012.

[34] J. C. Cox and S. A. Ross. The valuation of options for alternative stochastic processes. *Journal of financial economics*, 3(1-2):145–166, 1976.

[35] H. Cuesdeanu and J. C. Jackwerth. The pricing kernel puzzle in forward looking data. 2016.

[36] Aswath Damodaran. Equity risk premiums (erp): Determinants, estimation and implications–the 2016 edition. 2016.

[37] J. Danielsson, P Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, H. S. Shin, et al. An academic response to basel ii. *Special Paper-LSE Financial Markets Group*, 2001.

[38] Patrick Dennis and Stewart Mayhew. Implied volatility smiles: Evidence from options on individual equities. *Banking and Finance Department, Terry College of Business, University of Georgia, Working Paper*, 2000.

[39] K. Dowd. *Beyond Value at Risk: The Science of Risk Management.* Wiley, New York, 1999.

[40] J. C. Duan and J. G. Simonato. Empirical martingale simulation for asset prices. *Management Science*, 44(9):1218–1233, 1998.

[41] Jin-Chuan Duan and Jean-Guy Simonato. Empirical martingale simulation for asset prices. *Management Science*, 44(9):1218–1233, 1998.

[42] D. Duffie and J. Pan. An overview of value at risk. *Journal of Derivatives*, 4(3):7–49, 1997.

[43] B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[44] R. Engle. The risk that risk will change. *Journal of Investment Management*, 7:24–28, 2009.

[45] R. Engle. Long-term skewness and systemic risk. *Journal of Financial Econometrics*, 9(3):437–468, 2011.

[46] R. Engle and V. Rosenberg. Empirical pricing kernels. *Review of Financial Studies*, 64(341-372), 2002.

[47] R. F. Engle. Wald, likelihood ratio, and lagrange multiplier tests in econometrics. *Handbook of econometrics*, 2:775–826, 1984.

[48] Robert F Engle. A general approach to lagrange multiplier model diagnostics. *Journal of Econometrics*, 20(1):83–104, 1982.

[49] T. Fissler, J. F. Ziegel, and T. Gneiting. Expected shortfall is jointly elicitable with value at risk-implications for backtesting. *arXiv preprint arXiv:1507.00244*, 2015.

[50] H. Geman. *Commodities and commodity derivatives: modeling and pricing for agriculturals, metals and energy*. John Wiley & Sons, 2009.

[51] H. Geman and B. Liu. Introducing distances between commodity markets: the case of the us and uk natural gas. In *Advanced Modelling in Mathematical Finance*, pages 93–105. Springer, 2016.

[52] H. Geman and S. Ohana. Forward curves, scarcity and price volatility in oil and natural gas markets. *Energy Economics*, 31(4):576–585, 2009.

[53] E. Ghysels, A. C. Harvey, and E. Renault. Stochastic volatility. *Handbook of Statistics*, 14:119–191, 1996.

[54] Lawrence R. Glosten, Ravi Jagannathan, and David E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801, 1993.

[55] L.R. Glosten, R. Jagannathan, and D.E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

[56] T. Gneiting. Quantiles as optimal point forecasts. *International Journal of Forecasting*.

[57] C. González-Pedraz, M. Moreno, and J. I. Peña. Tail risk in energy portfolios. *Energy Economics*, 46:422–434, 2014.

[58] C. Gourieroux, A. Monfort, and A. Trognon. Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700, May 1984.

[59] J. M. Griffin and A. Shams. Manupulation in the vix. *The Review of Financial Studies*, 31(4):1377–1417, 2018.

[60] F. Gul and W. Pesendorfer. Random expected utility. *Econometrica*, 74(1):121–146, 2006.

[61] M. Haas. New methods in backtesting. *Financial Engineering Research Center Caesar, Bonn,*, 2001.

[62] J. M. Harrison and D. M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, 20(3):381–408, 1979.

[63] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[64] T. Hens and C. Reichlin. Three solutions to the pricing kernel puzzle. *Working paper, University of Zurich*, 2011.

[65] J. Jackwerth. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies*, 13:433–451, 2000.

[66] J. Jiang and Y.S. Tian. The model-free implied volatility and its information content. *The Review of Financial Studies*, 18:1305–1342, 2005.

[67] P. Jorion. Value at risk. 1997.

[68] P. Jorion. *Financial Risk Manager Handbook*. Number 6th Edition. Wiley Finance,, 2011.

[69] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, pages 263–291, 1979.

[70] Raymond Kan and Guofu Zhou. A critique of the stochastic discount factor methodology. *The Journal of finance*, 54(4):1221–1248, 1999.

[71] W. S. Krasker. The peso problemin testing the efficiency of forward exchange markets. *Journal of Monetary Economics*, 6(2):269–276, 1980.

[72] D. M. Kreps and E. L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica: journal of the Econometric Society*, pages 185–200, 1978.

[73] P. Kupiec. Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3:73–84, 1995.

[74] N. S. Lambert, D. M. Pennock, and Y. Shoham. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138. ACM, 2008.

[75] D. Leisen. Small sample bias in pricing kernel estimations. *Available at SSRN 2409661*, 2014.

[76] M. Levy and H. Levy. Prospect theory: much ado about nothing? *Management Science*, 48(10):1334–1349, 2002.

[77] G. M. Ljung and G. E. Box. On a measure of lack of fit in time series models. *Biometrika*, pages 297–303, 1978.

[78] M. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

[79] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, March 1952.

[80] I. Martin. What is the expected return on the market? *The Quarterly Journal of Economics*, 132(1):367–433, 2017.

[81] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: Concepts, techniques and tools*. Princeton university press, 2015.

[82] R. C Merton. Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183, 1973.

[83] R. C. Merton. Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183, 1973.

[84] R.C. Merton. On estimating the expected return on the market. *Journal of Financial Economics*, 8:323–361, 1980.

[85] S. Mitra. The relationship between conditional value at risk and option prices with a closed form solution. *The European Journal of Finance*, 21(5):400–425, 2015.

[86] Basel Committee on Banking Supervision. Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements. *http://www.bis.org/publ/bcbs22.htm.*, January 1996.

[87] J. Pantoja, A. Roncoroni, and R. Id Brik. Optimal static hedging of energy price and volume risk: Closed-form results. 2012.

[88] Andrew J Patton and Allan Timmermann. Monotonicity in asset returns: New tests with applications to the term structure, the capm, and portfolio sorts. *Journal of Financial Economics*, 98(3):605–625, 2010.

[89] P. Penza and V.K. Bansal. *Measuring Market Risk with Value at Risk*. Wiley, New York, 2000.

[90] Jim Pitman and Marc Yor. The two-parameter poisson-dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, pages 855–900, 1997.

[91] S. Ross. The recovery theorem. *The Journal of Finance*, 70(2):615–648, April 2015.

[92] C. Sala, G. Barone-Adesi, and A. Mira. Option-implied state prices and misspecified stochastic discount factors. *Swiss Finance Institute Research Paper No. 15-66*, 2017.

[93] A. Samit. Calculating value-at-risk using option implied probability distribution of asset price. *Wilmott Magazine*, 2012(59):56–61, 2012.

[94] Frank A Wolak. Testing inequality constraints in linear econometric models. *Journal of econometrics*, 41(2):205–235, 1989.

[95] Y. Xing, X. Zhang, and R. Zhao. What does the individual option volatility smirk tell us about future equity returns? *Journal of Financial Quantitative Analysis*, 45(3):641–662, 2010.

[96] S. Yan. Jump risk, stock returns, and slope of implied volatility smile. *Journal of Financial Economics*, 99:216–233, 2011.

[97] M. Youssef, L. Belkacem, and K. Mokni. Value-at-risk estimation of energy commodities: A long-memory garch–evt approach. *Energy Economics*, 51:99–110, 2015.

[98] J. F Ziegel. Coherence and elicitability. *Mathematical Finance*, 2014.