RedFame

Applied Economics and Finance
Vol. 6, No. 5; September 2019
ISSN 2332-7294 E-ISSN 2332-7308
Published by Redfame Publishing
URL: http://aef.redfame.com

Distribution of Historic Market Data – Implied and Realized Volatility

M. Dashti Moghaddam[1], Zhiyuan Liu[1] & R. A. Serota[1]

[1] Department of Physics, University of Cincinnati, Cincinnati, Ohio, United States

Correspondence: R. A. Serota, Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221-0011, United States.

Received: July 1, 2019 Accepted: July 24, 2019 Available online: July 30, 2019

doi:10.11114/aef.v6i5.4416 URL: https://doi.org/10.11114/aef.v6i5.4416

Abstract

We undertake a systematic comparison between implied volatility, as represented by VIX (new methodology) and VXO (old methodology) and realized volatility. We do not find substantial difference in accuracy between VIX and VXO. We compare visually and statistically the distributions of realized and implied variance (volatility squared) and study the distribution of their ratio. The ratio distributions are studied both for the known realized variance (for the current month) and for the predicted realized variance (for the following month). We show that the ratio of the two is best fitted by a Beta Prime distribution, whose shape parameters depend strongly on which of the two months is used.

Keywords: volatility, implied, realized, VIX, fat tails

1. Introduction

The implied volatility index VIX was created in order to estimate, looking forward, the expected realized volatility (RV). CBOE introduced the original VIX (now VXO) in 1986. It was based on an inverted Black-Scholes formula, where S&P 100 near-term, at-the-money options were used to calculate a weighted average of volatilities. However, the Black-Scholes formula assumes that the volatility in the stock returns equation is either a constant, or at least does not have a stochastic component, while in reality it was already understood that volatility itself is stochastic in nature. A number of well-studied models of stochastic volatility have emerged, such as Heston (HM) (Dragulescu & Yakovenko, 2002; Heston, 1993) and multiplicative (MM) (Ma & Serota, 2014; Nelson, 1990). Consequently, a need arose for an implied volatility index, which would not only be based on stochastic volatility but would also be agnostic to a particular model of the latter (Bollerslev, Mathew & Zhou, 2004; Zhou & Chesnes, 2003).

CBOE introduced its current VIX methodology on September 22, 2003 (*The CBOE volatility index - VIX*, 2003) to fulfill the above requirements and was based on (Demeterfi, Derman, Kamal, & Zou, 1999b, 1999a), where a closed-form formula for the expected value of RV (Barndorff-Nielsen & Shephard, 2002) was derived using call and put prices. Notably, it utilized the S&P 500 index, which is far more representative of the total market, both near-term and next-term options and a broader range of strike prices. CBOE publishes historic data using both methodologies, VIX (new) and VXO (old) dating back to 1990 (*VIX Options and Futures Historical Data*, n.d.), (historic stock prices used in calculation of RV can be found at (*S&P500*, n.d.)). Here we call 1990 through September 19, 2003 VIX Archive and VXO Archive and from September 22, 2003 through December 30, 2016 VIX Current and VXO Current.

Naturally, the question arises of whether VIX, designed to be a superior methodology, has a better track record than VXO (or even be trusted, given a recent surmise that VIX can be manipulated (*Are Traders Manipulating the VIX?*, 2017; Griffin & Shams, 2017) and the fact that Nasdaq is working on its own volatility index (*Watch Out VIX: Nasdaq Amps Up Volatility Game*, 2017) -- we will not address this issue here). The short answer from our study is that it is unclear. All-in-all, VIX/VXO are still too young to have accumulated sufficient amount of data and only time will tell how reliable they are in predicting realized volatility. Still, one of our notable observations discussed below is that the ratio of realized to implied variance (squared RV, RV^2) is best fitted with a fat-tailed (power-law) Beta Prime distribution, which clearly signals occasional large discrepancies between prediction and realization. This is not surprising, given that we are trying to predict the future (by pricing options) based on what we know today and thus are unaware of unexpected future events that can spike the volatility.

As a reminder to the reader, the probability density function (PDF) of BP distribution is given by

$$\mathrm{BP}(p, q, \beta, x) = \frac{(1+\frac{x}{\beta})^{-p-q}(\frac{x}{\beta})^{p-1}}{\beta\, B[p,q]} \tag{1}$$

where β is a scale parameter, p and q are shape parameters and $B[p,q]$ is the Beta function; $BP \propto x^{p-1}$ for $x \ll \beta$ and $BP \propto x^{-q-1}$ for $x \gg \beta$. To wit, q is relatively small for the ratio of RV^2 of the predicted month to VIX^2 and VXO^2, which underscores the above point of unpredicted future spikes of volatility.

On the other hand, we also find that the distributions of the ratio of the current RV^2, that is the one calculated for the *preceding* month, to VIX^2 and VXO^2, while also fitted best with BP, has large values of q and $p \approx q$. As explained below, the inverse of the BP is also BP and, in this case, the inverse ratios, of VIX^2 and VXO^2 to RV^2, have approximately the same distributions as direct ratios. The interpretation of the fact that the distributions are still fat-tailed but with greatly suppressed tails might be that VIX and VXO, which are both aware of past spikes and are in possession of current information, are trying to predict future spikes. It should be pointed out that currently, there exists a robust body of research comparing implied and realized volatility (Andersen & Benzoni, 2014; Chrstensen & Prabhala, 1998; Han & Park, 2013; Kownatzki, 2016; Russon & Vakil, 2017; Vodenska & Chambers, 2013), including using high-frequency trading data (Andersen & Benzoni, 2014; Han & Park, 2013) and within-sample and out-of-sample predicted RV (Han & Park, 2013). The bulk of this research concentrated on regression analysis and modeling, the latter allowing for forecasting (Andersen & Benzoni, 2014; Han & Park, 2013). It can be summarized by stating that VIX and VXO a generally better predictors than past RV of future RV (at least long-term; near-term past RV may do better (Han & Park, 2013).)

While we also conduct regression analysis, this work is different in that it is based on the comparison of the probability density distributions of implied and realized variance: visually, statistically and via study of the ratio distributions. We believe that such approach provides a more in-depth picture of correlations between realized and implied volatilities. It should be also emphasized that we work only with a full monthly term and that this work is completely model-independent.

This paper is the second in a series devoted to analysis of historic market data, the other two discussing, respectively, stock returns (Liu, Dashti Moghaddam & Serota, 2019) and relaxation and correlations (Dashti Moghaddam, Liu & Serota, 2019b). It is organized as follows. In Section 2 we give a detailed visual and statistical comparison between RV and implied volatility represented by VIX and VXO based on 1990-2016 market data. In particular, we discuss the definitions of these quantities and proceed to compare distributions of realized variance RV^2 with VIX^2 and VXO^2. Towards this end, we underscore the necessity of rescaling, given that RV^2 is calculated based on 252 trading days a year (21 a month), while VIX^2 and VXO^2 are calculated for 365 (30). We proceed to briefly discuss properties of RV^2. We use KS statistic to gauge the proximity of VIX^2 and VXO^2 distributions to RV^2 distribution. We also analyze KS statistic of fits of RV^2/VIX^2 and RV^2/VXO^2 by various distributions, from normal to fat-tailed. Finally, we provide a qualitative summary of our results and elementary regression analysis. In Appendix we split the data into 1990-2003/2003-2016 and further into 2003-2010/2010-2016 units.

2. Comparing Distributions of RV^2 and VIX^2

2.1 Definitions, Rescaling, Distribution of RV^2

Realized variance (index) is defined as follows

$$\mathrm{RV}^2 = 100^2 \times \frac{252}{n} \sum_{i=1}^{n} r_i^2 \tag{2}$$

Where

$$r_i = \ln \frac{S_i}{S_{i-1}} \tag{3}$$

are daily returns and S_i is the reference (closing) price on day i. Time-averaged realized variance can be calculated from stochastic volatility σ_t (Barndorff-Nielsen & Shephard, 2002; Liu et al., 2019) as

$$\frac{1}{\tau}\int_0^{\tau} \sigma_t^2\, dt \tag{4}$$

Evaluation of the implied volatility is based on the evaluation of the expectation value of (4) (Demeterfi et al., 1999b, 1999a). VIX uses options prices to estimate this expectation value via the generalized formula (*The CBOE volatility index - VIX*, 2003)

$$\mathrm{VIX}^2 = 100^2 \times \frac{2}{T} \sum_i \frac{\Delta K_i}{{K_i}^2} e^{RT} Q(K_i) - \frac{1}{T}\left[\frac{F}{K_0} - 1\right]^2 \tag{5}$$

where T is the time to expiration; F is the forward index level derived from index option price; K_0 is the first strike below the forward index level, F; K_i is the strike price of ith out-of-money option: a call if $K_i > K_0$, a put if $K_i < K_0$ and both a put and a call if $K_i = K_0$; ΔK_i is the interval between strike prices, that is half the difference between the strike on either side of K_i, $\Delta K_i = (K_{i+1} - K_{i-1})/2$; R is the risk-free interest rate to expiration and $Q(K_i)$ is the midpoint of bid-ask spread for each option with strike K_i. This formula is then used for near- and next-term options (*The CBOE volatility index - VIX*, 2003) and the final expression for VIX is effectively an average between the two so the latter and the sum in (5) are intended to approximate the time average in (4).

VIX and VXO were designed to measure a 30-day expected volatility. However, in their final form VIX^2 and VXO^2. are annualized by the ratio of $365/30 \approx 12$ (*The CBOE volatility index - VIX*, 2003). As is clear from (2), RV^2 is also annualized and for comparison with VIX/VXO we should take $n = 21$, so that $252/21 = 12$; unlike VIX/VXO, RV is calculated based on the number of trading days. Accordingly, to compare the distributions of VIX^2 and VXO^2 with RV^2 we must rescale one of them with the ratio of their mean values. Table 1 lists ratios of the mean of VIX^2 and VXO^2 over the mean of RV^2. In what follows, the distributions of RV^2 are rescaled with the respective ratios from Table 1. We also analyze data for VIX Current and VXO Current both in aggregate form and split nearly evenly for a period covering the financial crisis and after (see Appendix).

An intriguing question arises as to an actual form of the RV^2 distribution (as well as those of VIX^2 and VXO^2). Were daily returns uncorrelated, for $n = 21$ in (2) RV^2 would be expected to approach a normal or a stable distribution by central and generalized central limit theorem respectively, depending on whether the variance of the daily PDF exists or not. Single-day returns seem to be better described by power-law-tailed distributions (Fuentes, Gerig & Vicente, 2009; Gerig, Vicente & Fuentes, 2009; Ma & Serota, 2014) with existing variance, while intra-day data seem to point to very long tails with a diverging variance (Behfar, 2016) (with a usual caveat that the tail behavior is hard to pinpoint, especially with smaller data sets; for multi-day returns, see (Dashti Moghaddam & Serota, 2018; Liu et al., 2019)). Our own work (Dashti Moghaddam, Liu & Serota, 2019a) indicates that correlations fall off quickly, as a power law, over a period of about five days and then persist to slowly decay exponentially. Fig. 1 indicates a tailed distribution for RV^2 which saturates to its final shape over about five days as well. As per our current results (Dashti Moghaddam et al., 2019a), it is best fitted -- and with high precision -- by Generalized Beta Prime distribution – a generalization of BP – and Beta Prime distribution. Conversely, while VIX^2 and VXO^2 are best fitted by these two distributions as well, the precision is considerably worse, which may be another indicator of their deficiencies.

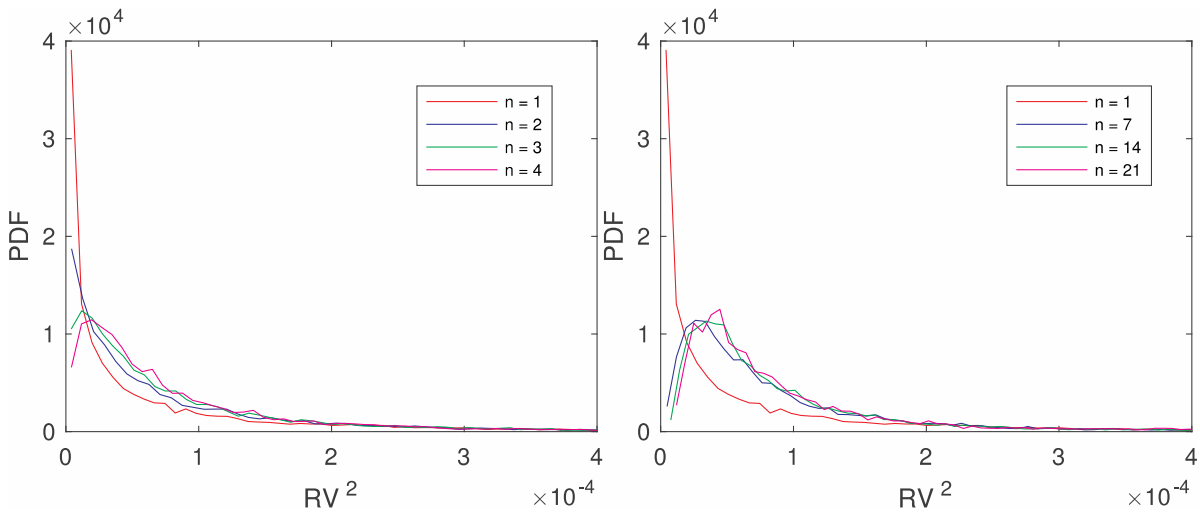


Figure 1. PDFs of $\frac{1}{n}\sum_i^n r_i^2$ for $n = 1,2,3,4$ (left) and $n = 1,7,14,21$ (right)

2.2 Qualitative Assessment

As previously mentioned, realized variance RV^2 is scaled by entries in Table 1 and raw time series plotted in Fig 2 - 3. Here we give a qualitative assessment of the tables and figures shown in the rest of this Section, Sections 2.3 and 2.4.

Comparing RV^2 distribution with those VIX^2 and VXO^2 using KS statistic (Table 2) we observe that the distributions of implied variance approximate the distribution of realized variance rather poorly, given large KS numbers.

By visually comparing PDF of RV^2 with VIX^2 and VXO^2 (Figs. 4 - 5) we observe the following features (in agreement with (Russon & Vakil, 2017) see also (Andersen & Benzoni, 2014; Chrstensen & Prabhala, 1998; Han & Park, 2013;

Kownatzki, 2016; Vodenska & Chambers, 2013). First, VIX^2 and VXO^2 have lower high-volatility probabilities relative to RV^2, including shorter fat tails, indicating that volatility indices do not predict accurately large values of RV, including the largest volatility spikes. In other word, volatility indices underestimate future large RV. Second, VIX^2 and VXO^2 have higher mid-volatility probabilities relative to RV^2, indicating that volatility indices overestimate future mid-level RV. Third, VIX^2 and VXO^2 have lower low-volatility probabilities relative to RV^2, indicating that volatility indices underestimate future low RV.

For the distributions of the ratios RV^2/VIX^2, RV^2/VXO^2, it is important to notice that, since realized and implied volatilities are correlated, we cannot construct them simply as the quotient distributions of two independent variable. For the *month predicted* by the volatility indices we observe the following attributes (Figs. 6-9). First, the distributions have fat tails, indicating again that VIX and VXO underestimate future values of RV, in particular volatility spikes. Second, very small ratios are suppressed, as manifested by a very large power exponent, indicating that it is rare that RV is considerably smaller than the one predicted by the volatility indices. Third, the tail exponents of the ratio distributions are larger than that of either RV^2 or VIX^2 and VXO^2, pointing to that for the RV^2 values taken from the tails, the values of VIX^2, VXO^2 are also more likely to come from the tails.

For the *month preceding* the volatility indices calculation we observe the following characteristics (Figs. 10-13). First tails of the distributions are much shorter than those for the predicted month, reflecting the fact that volatility indices account for past RV. Second, the tail exponents of the distributions are almost identical to those of their inverse, VIX^2/RV^2 and VXO^2/RV^2 distributions, indicating, as above, strong correlations.

For the ratio distribution of nRV^2 of the predicted (next) month to pRV^2 of the preceding month, we observe the following (Figs. 14-15). First, the exponent of the fat tail is smaller than those of the RV^2/VIX^2 and RV^2/VXO^2 distributions, that is the tails are longer. Second, the power-law exponent at very small ratios is much smaller for this distribution than for RV^2/VIX^2 and RV^2/VXO^2 that is those ratios are far less suppressed.

By both measures, VIX and VXO are better predictors of the future RV than the past RV.

2.3 Visual Comparison of Realized volatility and VIX/VXO

Figs. 2 and 3 show the time series of scaled RV and scaled RV^2 vis-a-vis their volatility indices counterparts. While we concentrate on realized and implied variance, we present RV and VIX/VXO as well since they are prominently figured in everyday use and are more familiar to the reader.

Figs. 4-5 show normalized (data PDF) histograms and their contour plots of RV^2 vis-a-vis VIX^2 and VXO^2. Qualitative assessment of these plots is conducted in Sec. 2.2.

Table 2 shows KS statistic for comparing distributions of VIX^2 and VXO^2 with the scaled RV^2. Lower KS numbers correspond to a better fit. The absolute KS values in Table 2 indicate a poor correspondence between VIX/VXO and RV. Compare those, for instance with an order-of-magnitude lower numbers for BP fits in Secs. 2.4.1-2.4.3.

Table 1. Ratio of mean (365/252 = 1.4484, 30/21 = 1.4286), left VIX^2, right VXO^2

Date	Ratio
1990-2016	1.4911
1990-2003	1.6691
2003-2016	1.3446
2003-2010	1.2861
2010-2016	1.4104

Date	Ratio
1990-2016	1.5257
1990-2003	1.8372
2003-2016	1.2985
2003-2010	1.2850
2010-2016	1.3097

Table 2. KS test results, right VIX^2, left VXO^2

Date	KS Statistic
1990-2016	0.1723
1990-2003	0.1478
2003-2016	0.2394
2003-2010	0.2215
2010-2016	0.2734

Date	KS Statistic
1990-2016	0.1589
1990-2003	0.1632
2003-2016	0.2157
2003-2010	0.2034
2010-2016	0.2376

2.4 Ratio Distribution

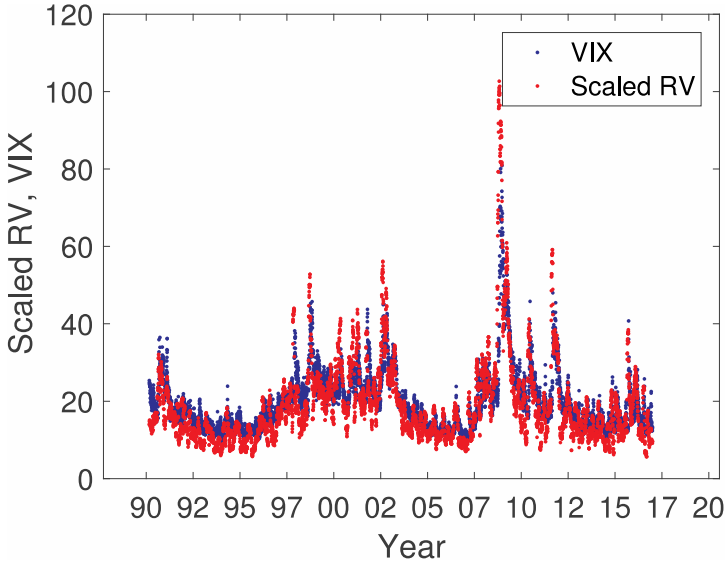
To further compare the volatilities, we examined the ratios RV^2/VIX^2 and RV^2/VIX^2. In plots below we show their time series and distribution functions (PDF). The latter are fitted using maximum likelihood estimation (MLE) and the parameters of the fits and KS statistic are collected in the tables. Seven functions -- Normal (N), Lognormal (LN), Inverse Gamma (IGa), Beta Prime (BP), Gamma (Ga), Weibull (Wbl) and Inverse Gaussian (IG) were used but only four best fits are shown with the data PDFs. Clearly, the fat-tailed BP was the best fit. The hypothesis is that fat tails signify sudden, unforeseen spikes in RV.

The inverse distributions VIX^2/RV^2 and VXO^2/RV^2 are also given to glean into whether there were unexpected surges in VIX and to illustrate the consistency of MLE. In particular, we observe that under transformation of the variable to its inverse, $x \to 1/x$, distributions transform as $\mathrm{IGa} \to \mathrm{Ga}$, $\mathrm{LN} \to \mathrm{LN}$ and $\mathrm{BP} \to \mathrm{BP}$. For BP, specifically, $\beta \to 1/\beta$, $\mathrm{p} \to \mathrm{q}$ and $\mathrm{q} \to \mathrm{p}$. Large values of p indicate suppressed low values, mimicking IGa-like behavior; large values of q indicate suppressed tails, mimicking Ga-like behavior.

In 2.4.1, we analyze the ratio distribution of the *predicted* month for which VIX/VXO is calculated. For instance, if on March 31 VIX/VXO predict RV for April, we compare those to RV for April.

In 2.4.2, we analyze the ratio distribution of the *preceding* month for which VIX/VXO is calculated. For instance, if on March 31 VIX/VXO predict RV for April, we compare those to RV for March. This is to test how much known RV affects volatility indices.

In 2.4.3, we analyze the ratio distribution of the *predicted* month (labeled "n" for next) to the *preceding* month (labeled "p"). This serves as one of the tests of whether VIX/VXO are a better predictor of the future RV than the past RV, of which they are already aware.



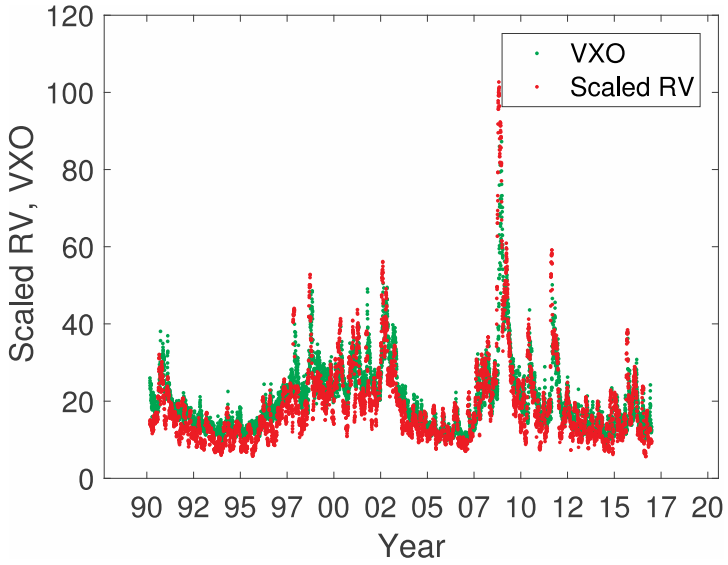
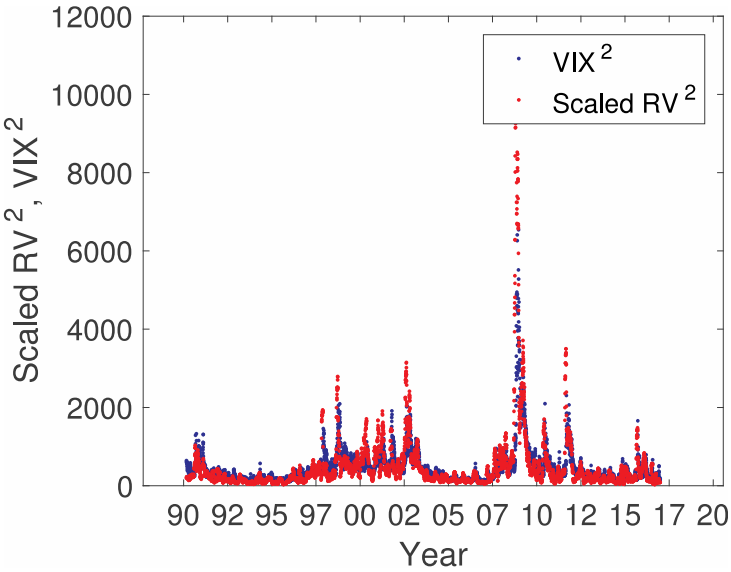


Figure 2. VIX (top) and VXO (bottom) with scaled RV, from Jan 2nd, 1990 to Dec 30th, 2016



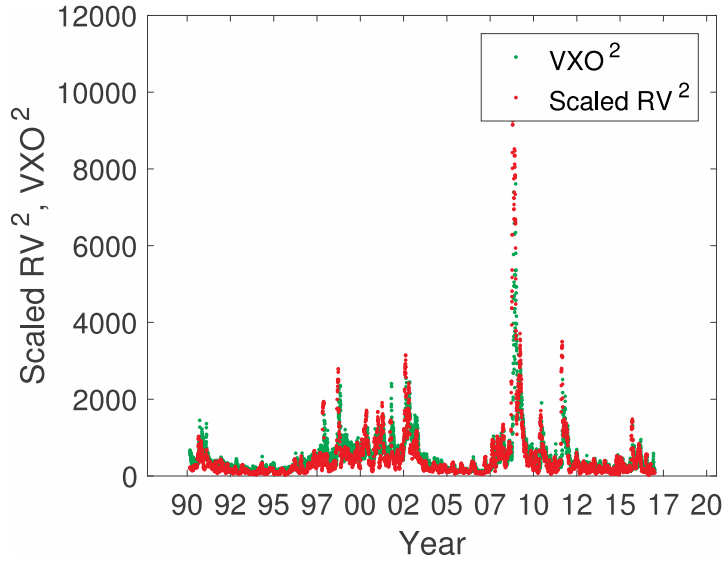


Figure 3. VIX^2 (top) and VXO^2 (bottom) with Scaled RV^2, from Jan 2nd, 1990 to Dec 30th, 2016

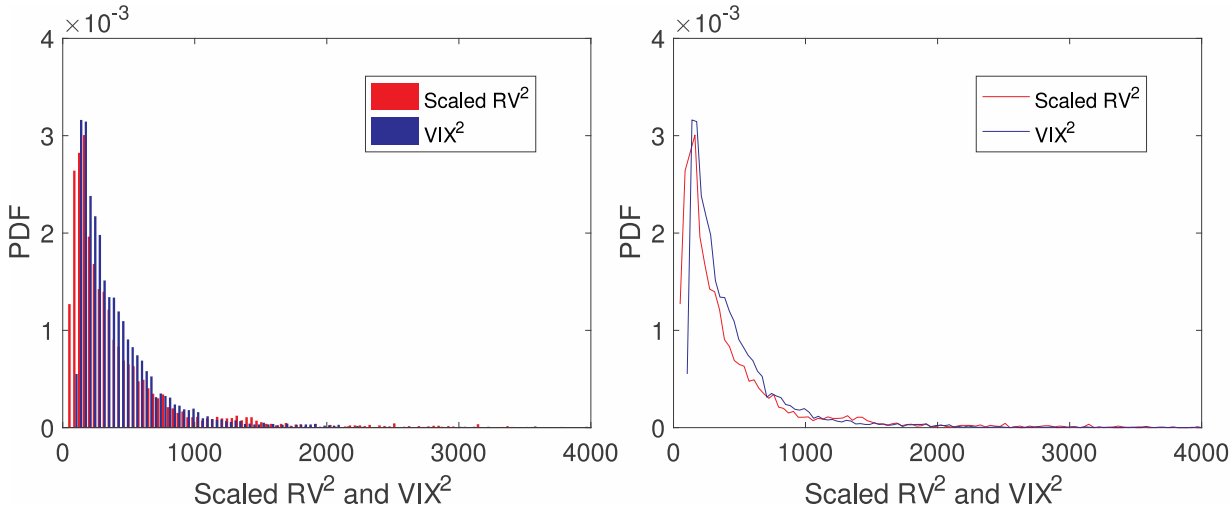


Figure 4. PDFs of scaled RV^2 and VIX^2 from Jan 2nd, 1990 to Dec 30th, 2016

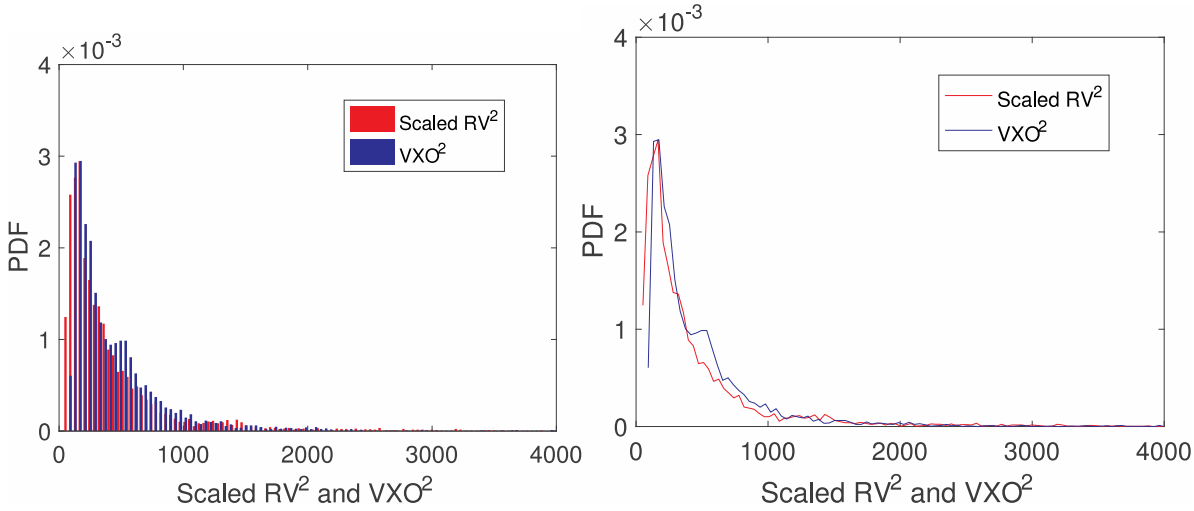


Figure 5. PDFs of scaled RV^2 and VXO^2 from Jan 2nd, 1990 to Dec 30th, 2016

2.4.1 Predicted Month

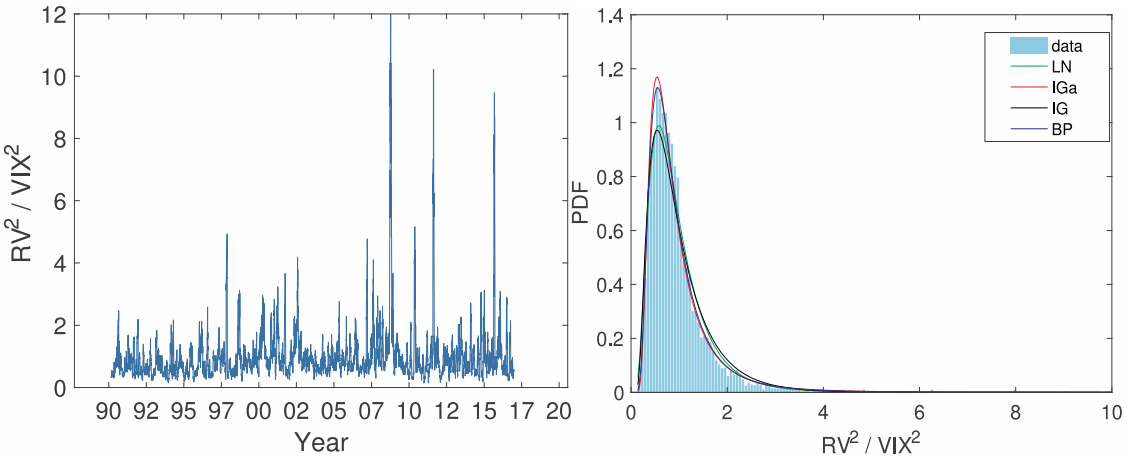


Figure 6. RV^2/VIX^2, from Jan 2nd, 1990 to Dec 30th, 2016

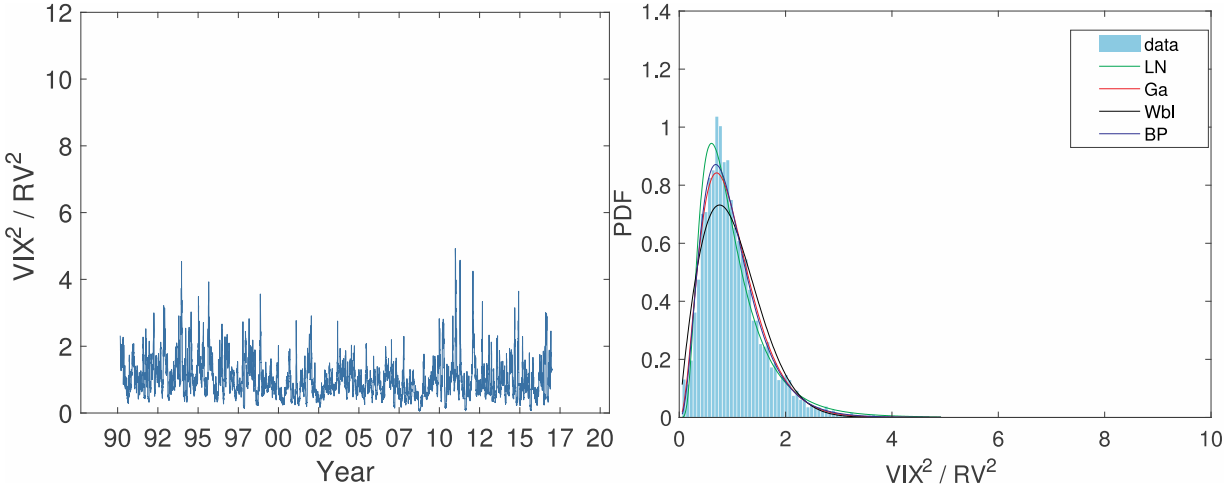


Figure 7. VIX^2/RV^2, from Jan 2nd, 1990 to Dec 30th, 2016

Table 3. MLE results for RV^2/VIX^2 and VIX^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000, 0.907)	0.194
LogNormal	LN(-0.203, 0.587)	0.045
IGa	IGa(3.359, 2.347)	0.024
Gamma	Ga(2.622, 0.381)	0.098
Weibull	Wbl(1.112, 1.400)	0.122
IG	IG(1.000, 2.317)	0.061
BP	BP(27.228, 3.805, 0.101)	0.020

type	parameters	KS Statistic
Normal	N(1.000, 0.562)	0.097
LogNormal	LN(-0.156, 0.587)	0.045
IGa	IGa(2.622, 1.831)	0.098
Gamma	Ga(3.359, 0.298)	0.025
Weibull	Wbl(1.131, 1.888)	0.050
IG	IG(1.000, 2.317)	0.073
BP	BP(3.805, 27.228, 6.891)	0.020

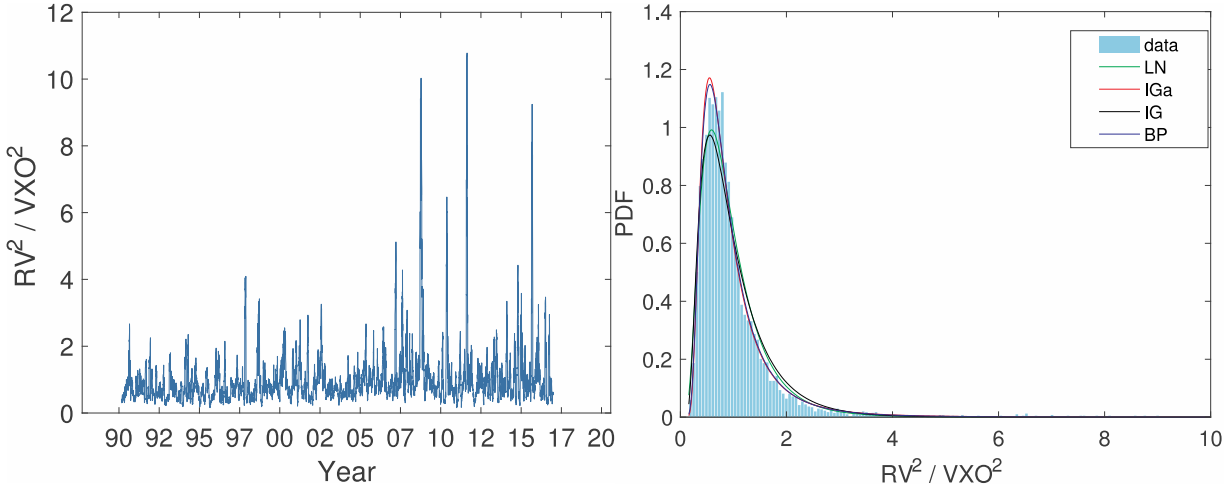


Figure 8. RV^2/VXO^2, from Jan 2nd, 1990 to Dec 30th, 2016

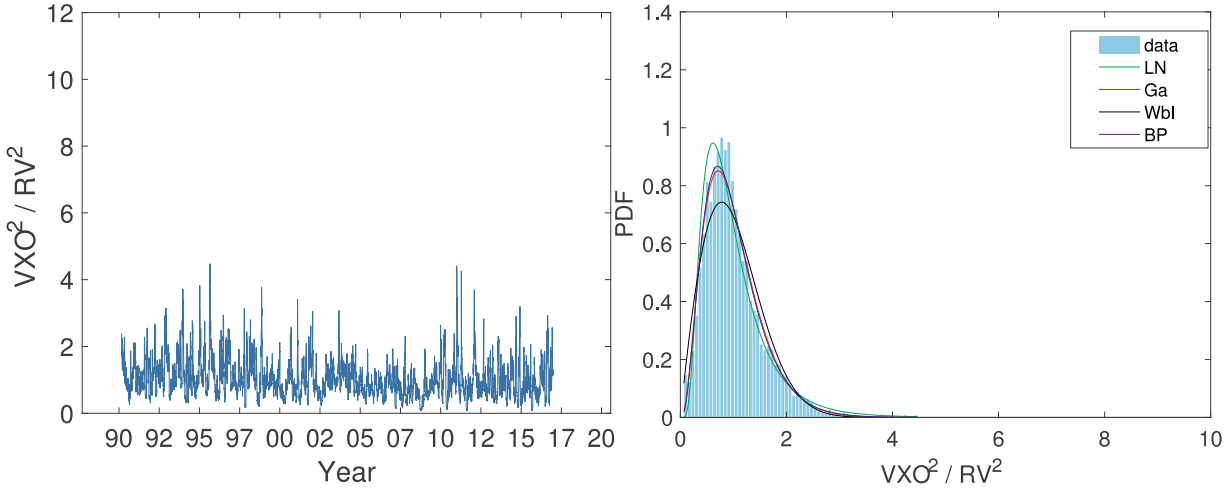


Figure 9. VXO^2/RV^2, from Jan 2nd, 1990 to Dec 30th, 2016

Table 4. MLE results for RV^2/VXO^2 and VXO^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000, 0.875)	0.191
LogNormal	LN(-0.197, 0.579)	0.045
IGa	IGa(3.463, 2.444)	0.022
Gamma	Ga(2.690, 0.372)	0.097
Weibull	Wbl(1.115, 1.426)	0.123
IG	IG(1.000, 2.398)	0.061
BP	BP(47.600, 3.716, 0.056)	0.018

type	parameters	KS Statistic
Normal	N(1.000, 0.547)	0.092
LogNormal	LN(-0.151, 0.579)	0.045
IGa	IGa(2.690, 1.898)	0.097
Gamma	Ga(3.463, 0.289)	0.022
Weibull	Wbl(1.131, 1.938)	0.050
IG	IG(1.000, 2.398)	0.073
BP	BP(3.716, 47.600, 12.541)	0.018

2.4.2 Preceding Month

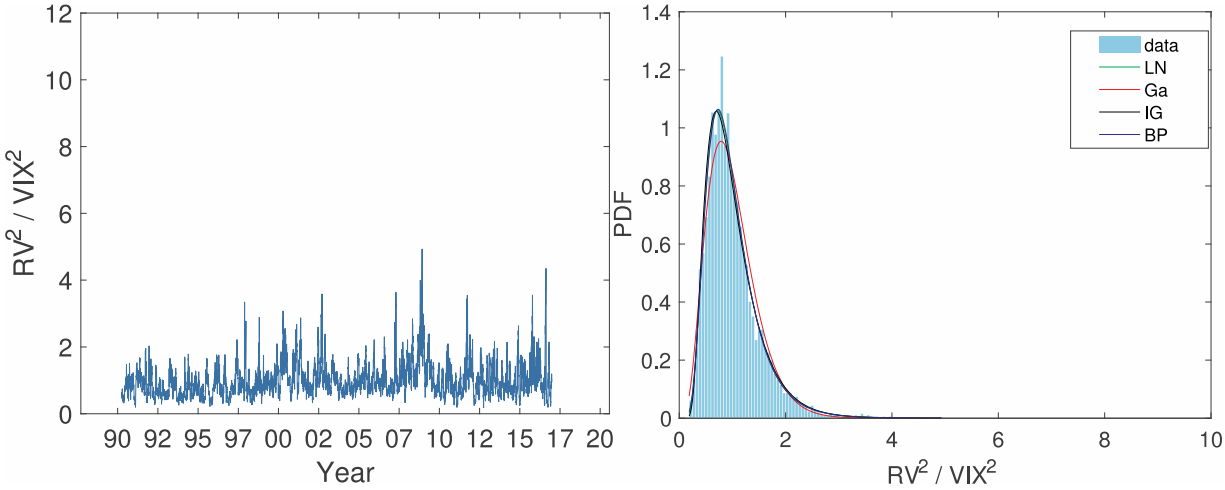


Figure 10. RV^2/VIX^2, from Jan 2nd, 1990 to Dec 30th, 2016

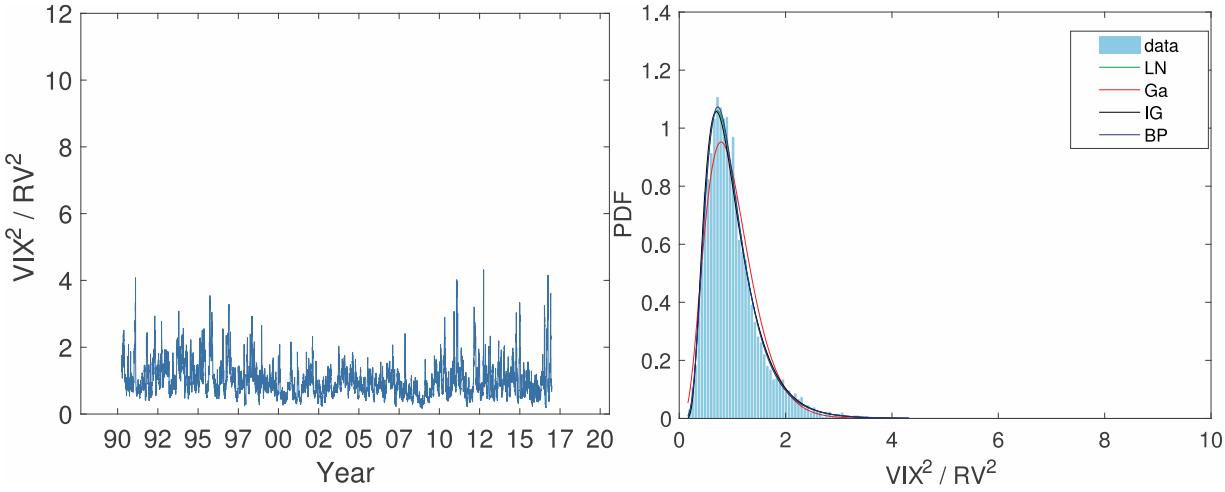


Figure 11. VIX^2/RV^2, from Jan 2nd, 1990 to Dec 30th, 2016

Table 5. MLE results for RV^2/VIX^2 and VIX^2/RV^2

type	parameters	KS Statistic
Normal	N (1.000, 0.497)	0.099
LogNormal	LN(-0.110, 0.469)	0.015
IGa	IGa(4.689, 3.762)	0.043
Gamma	Ga(4.711, 0.212)	0.038
Weibull	Wbl(1.132, 2.125)	0.067
IG	IG(1.000, 4.058)	0.021
BP	BP(9.223, 9.985, 0.974)	0.012

type	parameters	KS Statistic
Normal	N(1.000, 0.499)	0.106
LogNormal	LN(-0.110, 0.469)	0.015
IGa	IGa(4.711, 3.779)	0.038
Gamma	Ga(4.689, 0.213)	0.043
Weibull	Wbl(1.132, 2.119)	0.075
IG	IG(1.000, 4.058)	0.016
BP	BP(9.985, 9.233, 0.824)	0.012

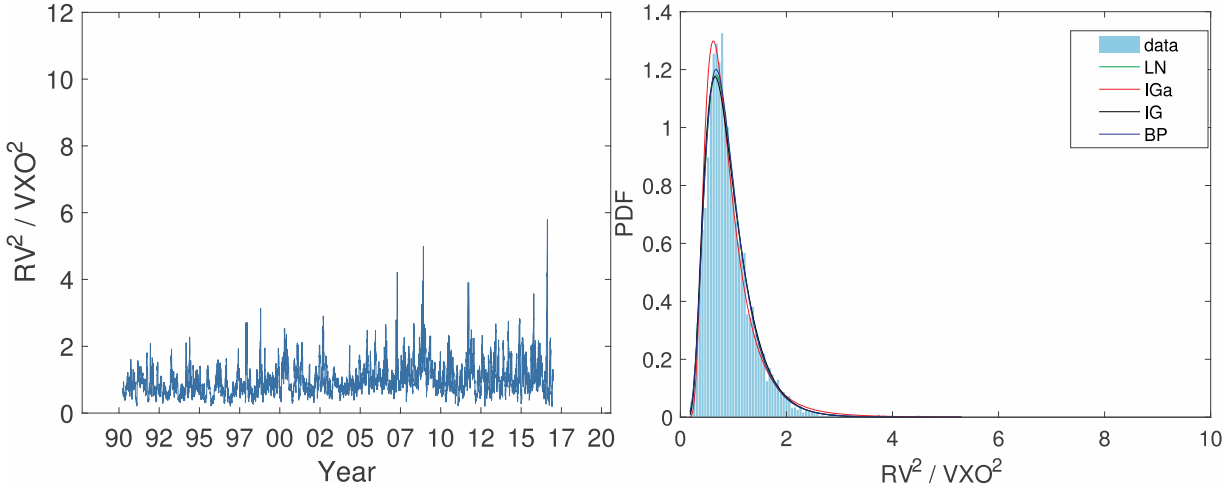


Figure 12. RV^2/VXO^2, from Jan 2nd, 1990 to Dec 30th, 2016

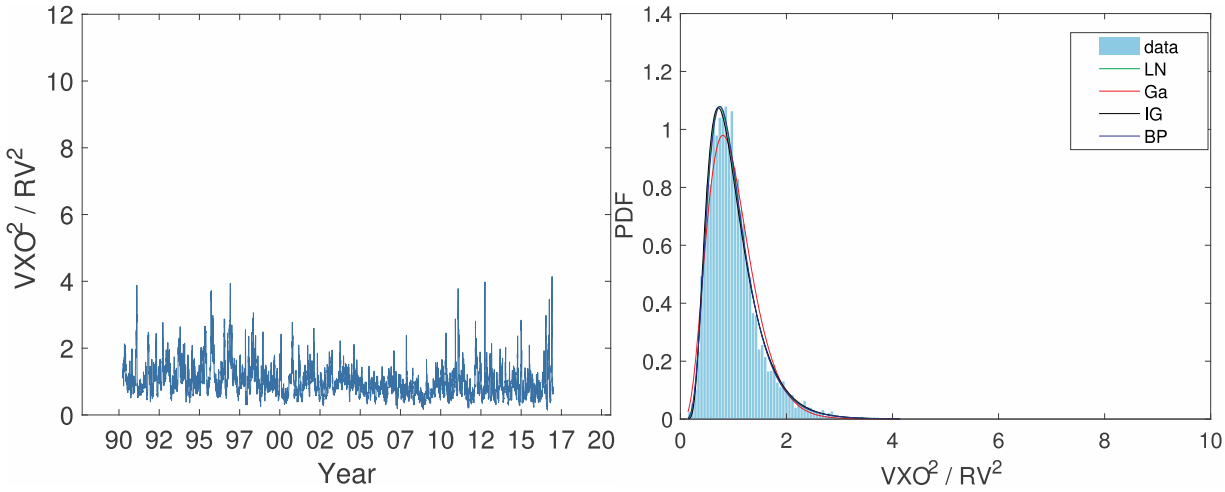


Figure 13. VXO^2/RV^2, from Jan 2nd, 1990 to Dec 30th, 2016

Table 6. MLE results for RV^2/VXO^2 and VXO^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000,0.491)	0.106
LogNormal	LN(-0.104, 0.454)	0.015
IGa	IGa(5.035, 4.095)	0.033
Gamma	Ga(4.962, 0.201)	0.045
Weibull	Wbl(1.132, 2.138)	0.073
IG	IG(1.000, 4.355)	0.020
BP	BP(11.169, 9.403, 0.752)	0.013

type	parameters	KS Statistic
Normal	N(1.000, 0.477)	0.093
LogNormal	LN(-0.103, 0.454)	0.015
IGa	IGa(4.962, 4.035)	0.045
Gamma	Ga(5.035, 0.199)	0.033
Weibull	Wbl(1.132, 2.210)	0.069
IG	IG(1.000, 4.355)	0.021
BP	BP(9.403, 11.169, 1.081)	0.013

2.4.3 Ratio of Realized Variances of Two Adjacent Months

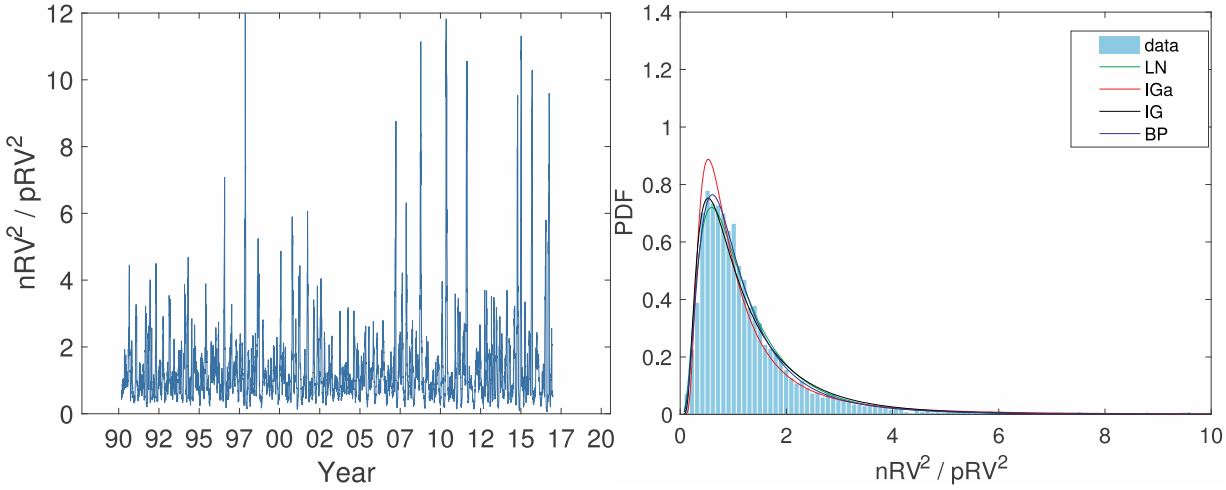


Figure 14. Ratio of next-month realized variance to that of the preceding month, from Jan 2nd, 1990 to Dec 30th, 2016

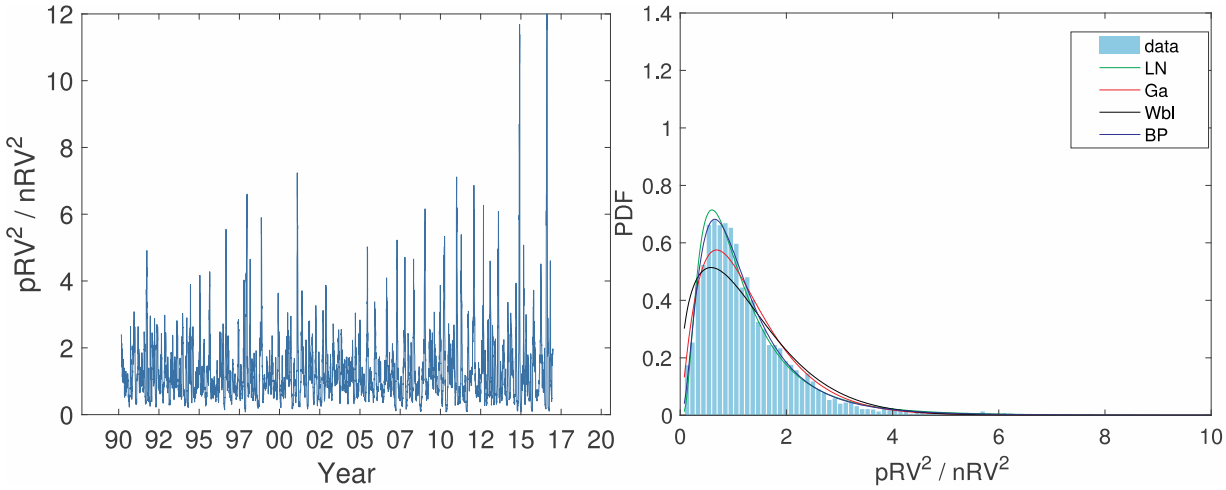


Figure 15. Ratio of preceding-month realized variance to that of the following month, from Jan 2nd, 1990 to Dec 30th, 2016

Table 7. MLE results for nRV^2/pRV^2 and pRV^2/nRV^2

type	parameters	KS Statistic	type	parameters	KS Statistic
Normal	N(1.317, 1.258)	0.181	Normal	N(1.292, 1.078)	0.142
LogNormal	LN(-0.004, 0.721)	0.024	LogNormal	LN(0.004, 0.721)	0.024
IGa	IGa(2.129, 1.647)	0.047	IGa	IGa(1.939, 1.472)	0.080
Gamma	Ga(1.939, 0.679)	0.080	Gamma	Ga(2.129, 0.607)	0.047
Weibull	Wbl(1.440, 1.287)	0.092	Weibull	Wbl(1.430, 1.395)	0.061
IG	IG(1.317, 1.874)	0.034	IG	IG(1.292, 1.838)	0.051
BP	BP(5.877, 3.489, 0.555)	0.012	BP	BP(3.489, 5.877, 1.799)	0.012

2.5 Correlation Analysis

Tables 8 and 9 list Pearson correlation coefficients (PCC). Here "n" labels the "next" month that is the month for which VIX and VXO were predicting the implied RV; "p" labels the preceding month; "r" labels a "random" month. All RV^2 are scaled, as explained in Sec. 2.1. Tables 10 and 11 list Kolmogorov-Smirnov (KS) statistic for comparison of the two plots. Of note in Tables 8 and 9 are very close PCC for $pRV - nRV$ and $VIX/VXO - nRV$ and much higher PCC for $pRV - VIX/VXO$.

Table 8. PCC VIX

	pRV^2	nRV^2	VIX^2	rRV^2
pRV^2	1	0.70	0.88	0.005
nRV^2	0.70	1	0.71	0.002
VIX^2	0.88	0.71	1	0.003
rRV^2	0.005	0.002	0.003	1

Table 9. PCC VXO

	pRV^2	nRV^2	VIX^2	rRV^2
pRV^2	1	0.70	0.87	0.001
nRV^2	0.70	1	0.72	0.004
VIX^2	0.87	0.72	1	0.002
rRV^2	0.001	0.004	0.002	1

Table 10. KS VIX

	$\frac{pRV^2}{VIX^2}$	$\frac{nRV^2}{VIX^2}$	$\frac{pRV^2}{nRV^2}$	$\frac{rRV^2}{rVIX^2}$	$\frac{rRV^2}{rRV^2}$	$\frac{nRV^2}{pRV^2}$
$\frac{pRV^2}{VIX^2}$	0	0.06	0.13	0.20	0.26	-
$\frac{nRV^2}{VIX^2}$	0.06	0	-	0.18	0.23	0.13
$\frac{pRV^2}{nRV^2}$	0.13	-	0	0.17	0.15	-

$\frac{rRV^2}{rVIX^2}$	0.20	0.18	0.17	0	0.06	0.17
$\frac{rRV^2}{rRV^2}$	0.26	0.23	0.15	0.06	0	0.16
$\frac{nRV^2}{pRV^2}$	-	0.13	-	0.17	0.16	0

Table 11. KS VXO

	$\frac{pRV^2}{VXO^2}$	$\frac{nRV^2}{VXO^2}$	$\frac{pRV^2}{nRV^2}$	$\frac{rRV^2}{rVXO^2}$	$\frac{rRV^2}{rRV^2}$	$\frac{nRV^2}{pRV^2}$
$\frac{pRV^2}{VXO^2}$	0	0.06	0.13	0.22	0.26	-
$\frac{nRV^2}{VXO^2}$	0.06	0	-	0.19	0.23	0.16
$\frac{pRV^2}{nRV^2}$	0.13	-	0	0.17	0.16	-
$\frac{rRV^2}{rVXO^2}$	0.22	0.19	0.17	0	0.06	0.18
$\frac{rRV^2}{rRV^2}$	0.26	0.23	0.16	0.06	0	0.17
$\frac{nRV^2}{pRV^2}$	-	0.16	-	0.18	0.17	0

3. Conclusions

We presented the results of comparison between implied variance, as represented by squared volatility indices VIX and VXO, and realized variance (squared realized volatility). The main idea was to study the probability density functions (distributions) of these quantities and of their ratios in order to glean a better insight into correlations between them. Towards that end we compared implied variance for the predicted month -- predicted realized variance -- with the actual realized variance for that month. Additionally, we compared implied variance with the realized variance of the preceding month and the realized variances of the adjacent months. We used 1990-2016 market data, with the 1990-2003/2003-2016 and 2003-2010/2010-2016 slices investigated separately to reflect on the introduction of VIX and financial crisis respectively.

For proper comparison, we rescaled the data to bring implied and realized variance into alignment, given that the former is calculated daily and the latter only for trading days. We employed various statistical measures and techniques. We compared the distributions visually and using Kolmogorov-Smirnov (KS) statistic. We fitted ratio distributions with seven distributions which ranged from short-tailed to fat-tailed. We also conducted a simple regression analysis and compared various ratio distributions using KS statistic. All our analyses are completely model-independent. In addition to quantitative assessment listed in Sec. 2.2, our main findings are as follows.

First, we have not found substantial difference between VIX and VXO. For both of them, KS statistic for comparison between implied and realized variance distributions produces rather poor results.

Second, implied volatility is a better predictor of the realized volatility than the past realized volatility. This is born out both by the study of distributions and regression analysis. This is not surprising, given that the estimation of future volatility incorporates knowledge of the past volatility through option pricing. On the other hand, implied variance is only marginally better than the past realized variance in prediction quality.

Third, all ratio distributions are best fitted by the fat-tailed Beta Prime distribution, given by (1). For the ratio of

predicted month variance to its implied variance, the tail exponent q is rather small, that is the tail is really "fat," indicating presumably that the implied variance misses future volatility spikes. For the ratio of next month variance to the preceding month, the tail is even "fatter." On the other hand, the large p for the ratio involving implied variance, as opposed to small one involving past realized variance indicates that the former does better for low volatilities, which is probably why the implied variance works slightly better than the past realized variance. At the same time, the ratio of implied variance to realized variance of the preceding month and its inverse are roughly the same, with $p \approx q$ being rather large, which points to why it is only marginally better. This ratio distribution analysis is also fully consistent with the regression analysis.

Presently, we are studying the actual distributions of realized and implied variances. As per our current results (Dashti Moghaddam et al., 2019a), realized variance is best fitted -- and with high precision -- by a Generalized Beta prime distribution (and almost as well by BP). Conversely, while squared implied volatility indices are best fitted by it as well, the precision is considerably worse, which may be another indicator of their deficiencies as predictors of future volatility. In a future work, we will directly investigate the possible relationship between volatility spikes and fat tails of the ratio distributions.

Appendix A. VIX/VXO Archive Current: 1990-2003 -- 2003-2016 Split

A.1 Visual Comparison

Here we split the 1990-2016 data in two, before and after introduction of the current VIX.

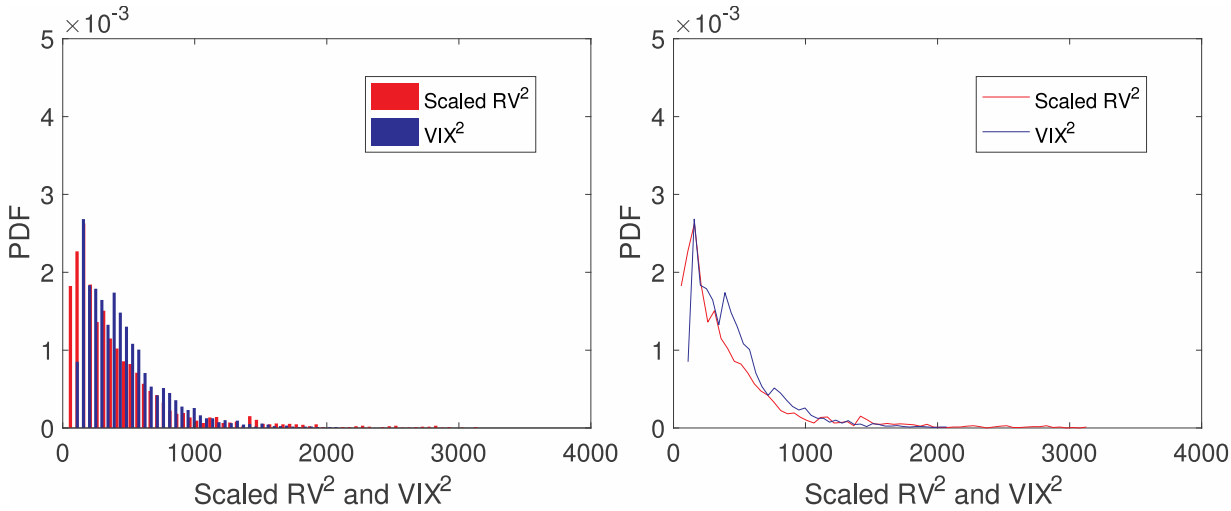


Figure A.16. PDFs of scaled RV^2 and VIX^2 from Jan 2nd, 1990 to Sep 19th, 2003

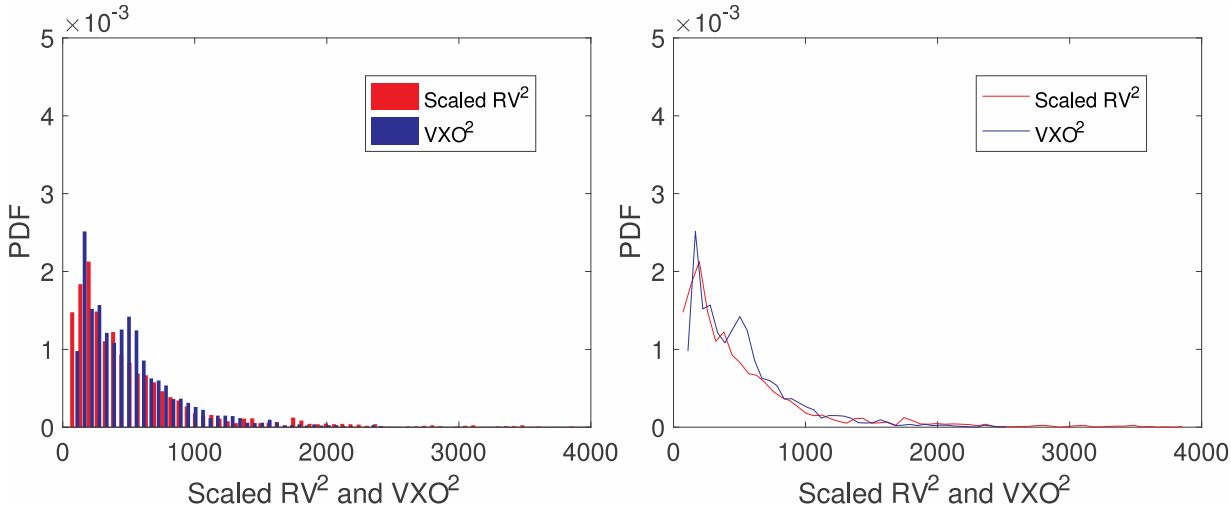


Figure A.17. PDFs of scaled RV^2 and VXO^2 from Jan 2nd, 1990 to Sep 19th, 2003

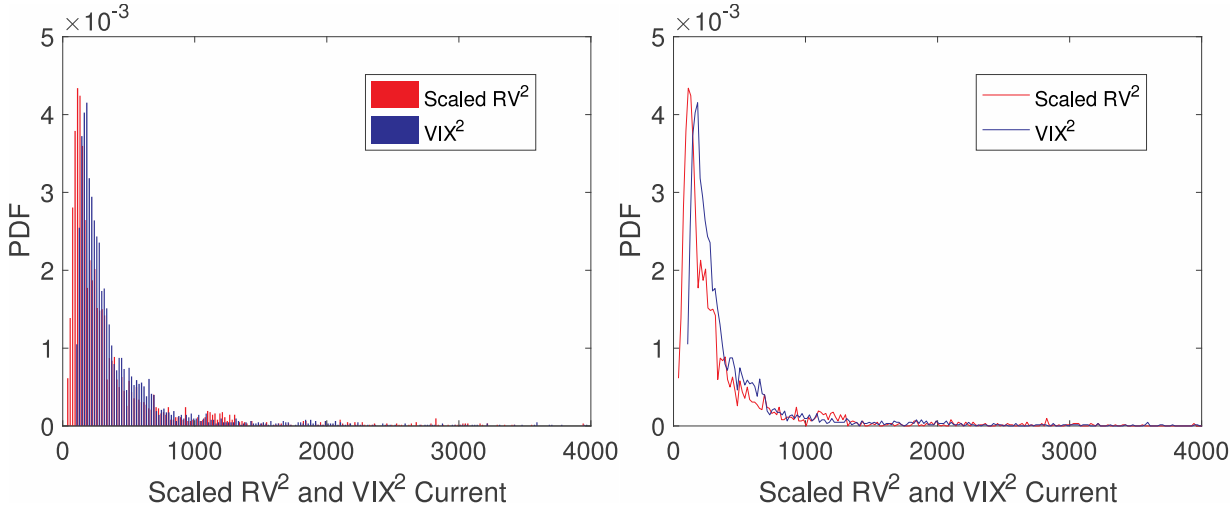


Figure A.18. PDFs of scaled RV^2 and VIX^2 from Sep 22nd, 2003 to Dec 30th, 2016

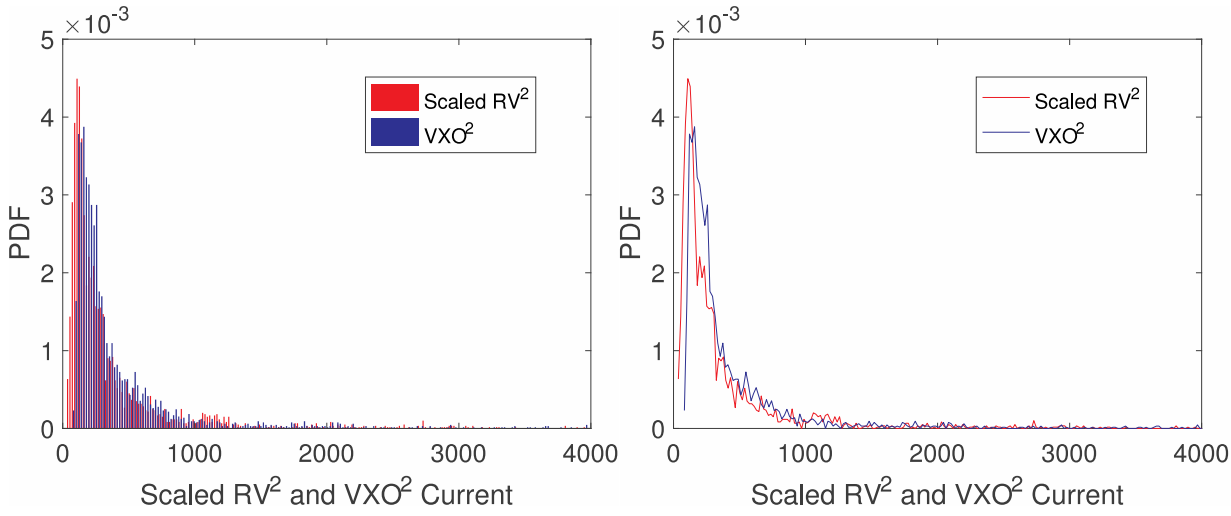


Figure A.19. PDFs of scaled RV^2 and VXO^2 from Sep 22nd, 2003 to Dec 30th, 2016

A.2 Ratio Distribution

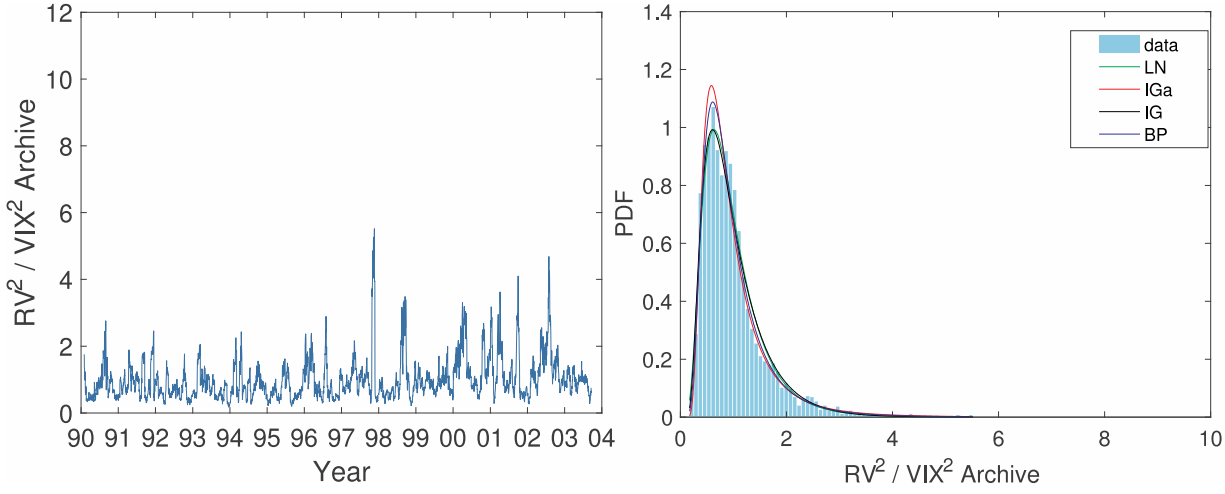


Figure A.20. RV^2/VIX^2, from Jan 2nd, 1990 to Sep 19th, 2003

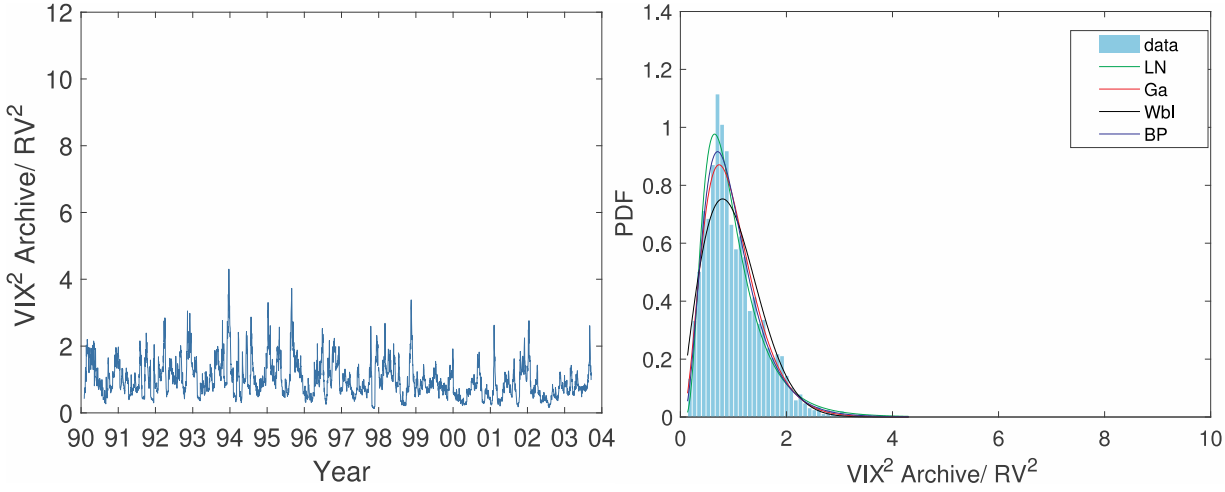


Figure A.21. VIX^2/RV^2, from Jan 2nd, 1990 to Sep 19th, 2003

Table A.12. MLE results for RV^2/VIX^2 and VIX^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000, 0.638)	0.142
LogNormal	LN(-0.157, 0.546)	0.028
IGa	IGa(3.696, 2.743)	0.036
Gamma	Ga(3.342, 0.299)	0.064
Weibull	Wbl(1.131, 1.726)	0.091
IG	IG(1.000, 2.877)	0.034
BP	BP(18.433, 4.582, 0.195)	0.022

type	parameters	KS Statistic
Normal	N(1.000, 0.537)	0.101
LogNormal	LN(-0.141, 0.546)	0.028
IGa	IGa(3.342, 2.480)	0.064
Gamma	Ga(3.696, 0.270)	0.036
Weibull	Wbl(1.133, 1.982)	0.058
IG	IG(1.000, 2.877)	0.042
BP	BP(4.582, 18.433, 3.805)	0.022

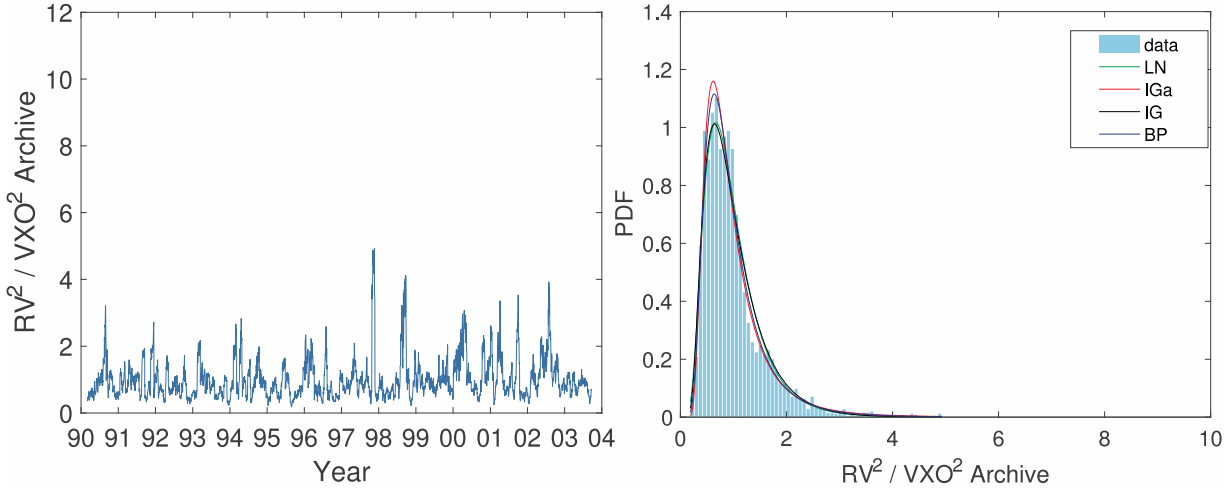


Figure A.22. RV^2/VXO^2, from Jan 2nd, 1990 to Sep 19th, 2003

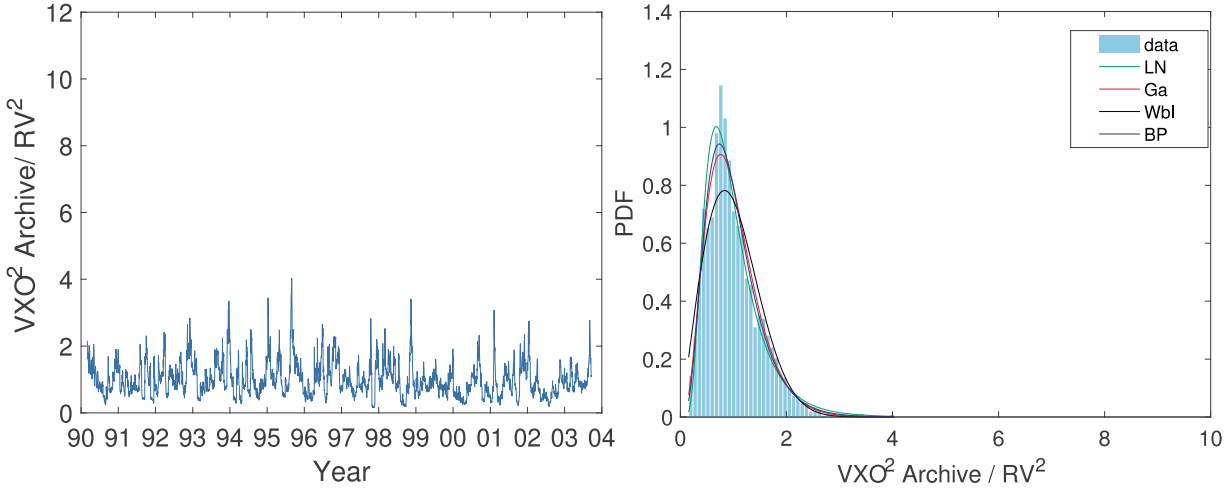


Figure A.23. VXO^2/RV^2, from Jan 2nd, 1990 to Sep 19th, 2003

Table A.13. MLE results for RV^2/VXO^2 and VXO^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000, 0.601)	0.145
LogNormal	LN(-0.197, 0.515)	0.035
IGa	IGa(4.129, 3.163)	0.027
Gamma	Ga(3.718, 0.269)	0.072
Weibull	Wbl(1.133, 1.813)	0.098
IG	IG(1.000, 3.276)	0.041
BP	BP(24.157, 4.943, 4.685)	0.019

type	parameters	KS Statistic
Normal	N(1.000, 0.506)	0.091
LogNormal	LN(-0.126, 0.515)	0.035
IGa	IGa(3.718, 2.849)	0.072
Gamma	Ga(4.129, 0.242)	0.027
Weibull	Wbl(1.133, 2.098)	0.054
IG	IG(1.000, 3.276)	0.048
BP	BP(4.943, 24.157, 0.163)	0.019

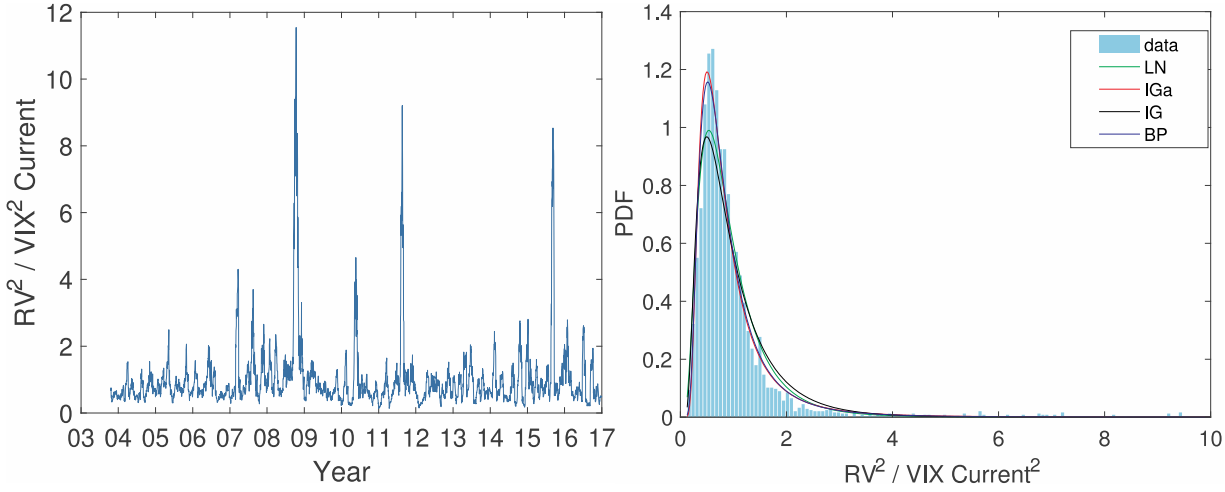


Figure A.24. RV^2/VIX^2, from Sep 22nd, 2003 to Dec 30th, 2016

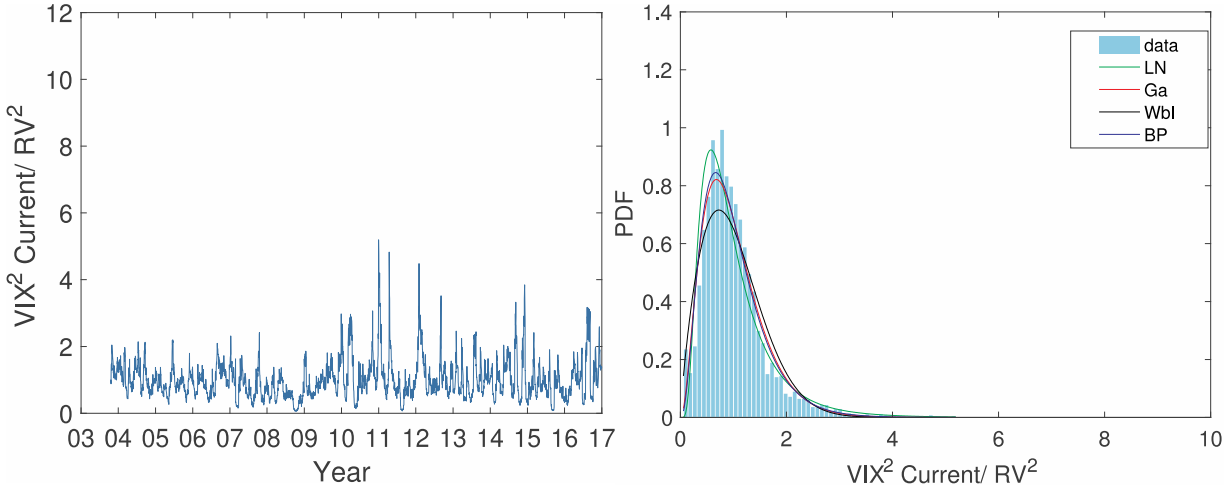


Figure A.25. VIX^2/RV^2, from Sep 22nd, 2003 to Dec 30th, 2016

Table A.14. MLE results for RV^2/VIX^2 and VIX^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000, 1.034)	0.227
LogNormal	LN(-0.239, 0.620)	0.056
IGa	IGa(3.111, 2.068)	0.036
Gamma	Ga(2.244, 0.446)	0.118
Weibull	Wbl(1.098, 1.292)	0.134
IG	IG(1.000, 1.983)	0.081
BP	BP(31.307, 3.433, 0.075)	0.032

type	parameters	KS Statistic
Normal	N(1.000, 0.586)	0.096
LogNormal	LN(-0.169, 0.620)	0.056
IGa	IGa(2.244, 1.491)	0.118
Gamma	Ga(3.111, 0.321)	0.036
Weibull	Wbl(1.128, 1.811)	0.056
IG	IG(1.000, 1.983)	0.095
BP	BP(3.433, 31.307, 8.827)	0.032

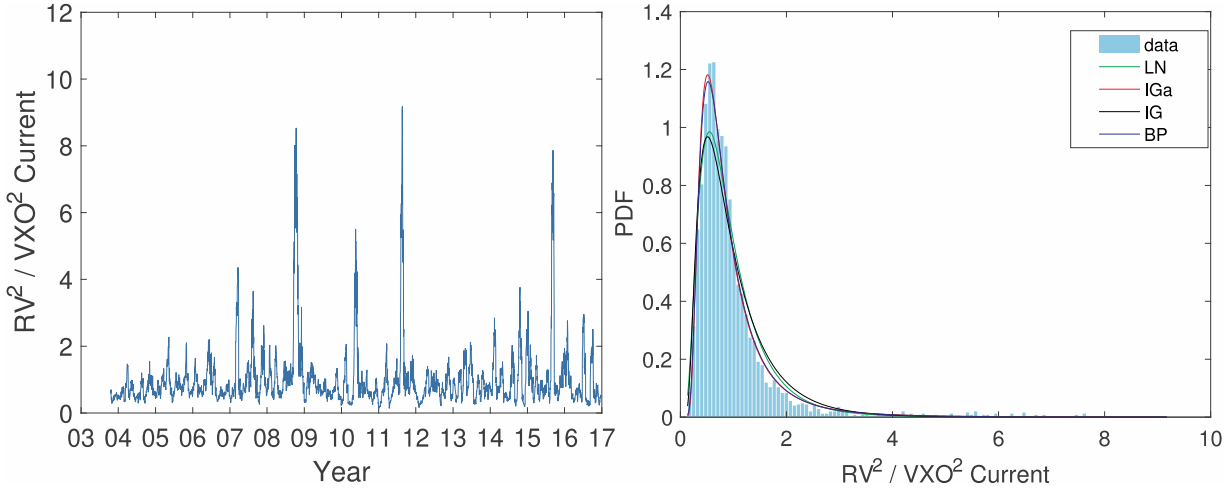


Figure A.26. RV^2/VXO^2, from Sep 22nd, 2003 to Dec 30th, 2016

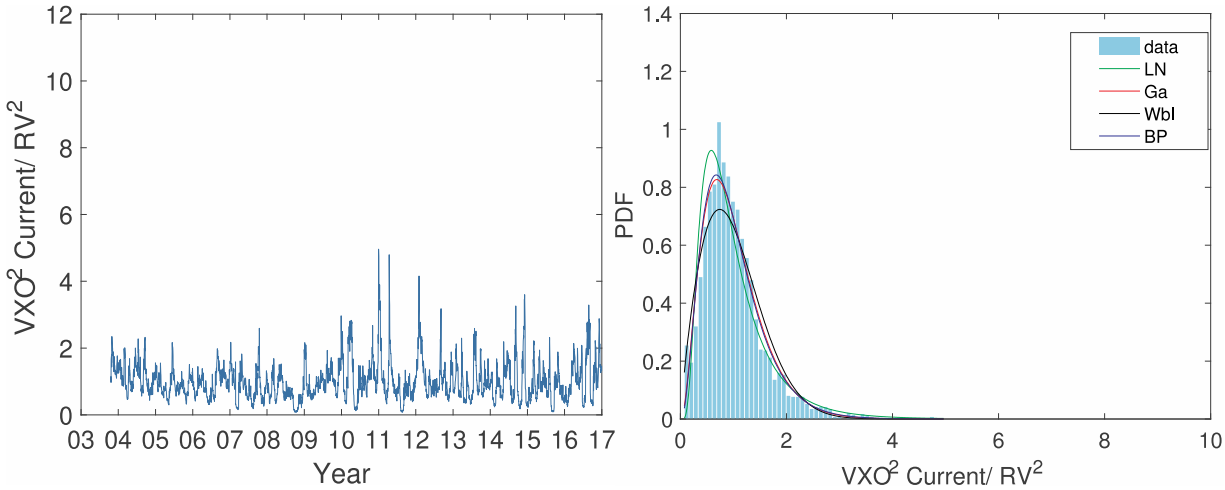


Figure A.27. VXO^2/RV^2, from Sep 22nd, 2003 to Dec 30th, 2016

Table A.15. MLE results for RV^2/VXO^2 and VXO^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000, 0.491)	0.106
LogNormal	LN(-0.104, 0.454)	0.015
IGa	IGa(5.035, 4.095)	0.033
Gamma	Ga(4.962, 0.201)	0.045
Weibull	Wbl(1.132, 2.138)	0.073
IG	IG(1.000, 4.355)	0.020
BP	BP(11.169, 9.403, 0.752)	0.013

type	parameters	KS Statistic
Normal	N(1.000, 0.477)	0.093
LogNormal	LN(-0.103, 0.454)	0.015
IGa	IGa(4.962, 4.035)	0.045
Gamma	Ga(5.035, 0.198)	0.033
Weibull	Wbl(1.132, 2.209)	0.069
IG	IG(1.000, 4.355)	0.021
BP	BP(9.403, 11.169, 1.081)	0.013

Appendix B. VIX/VXO Current: 2003-2010 -- 2010-2016 Split

Here we split 2003-2016 data in two roughly equal time periods, before and after the financial crisis.

B.1 Visual Comparison

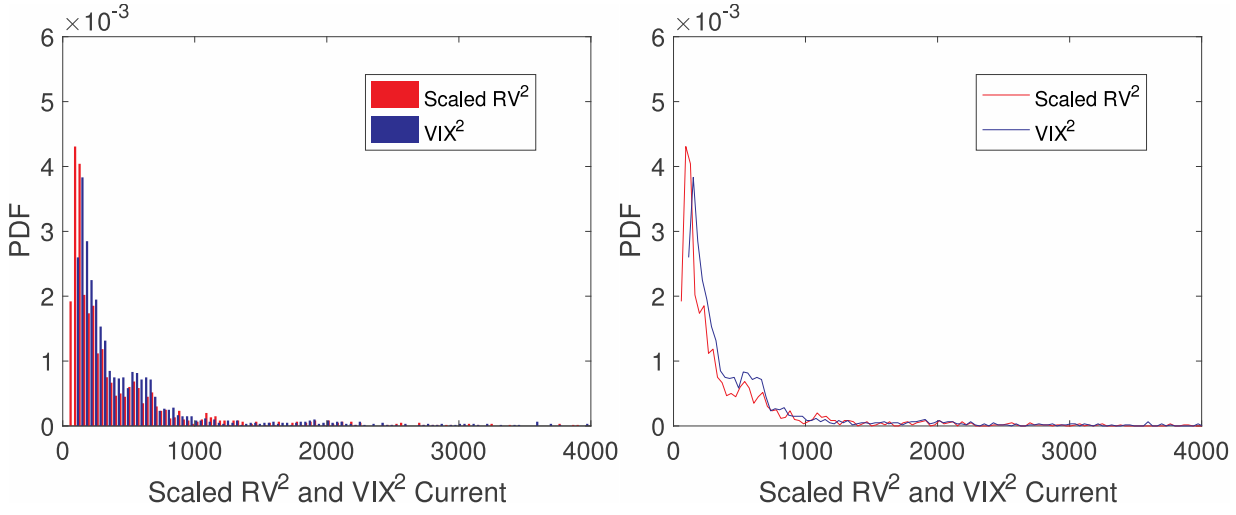


Figure B.28. PDFs of scaled RV^2 and VIX^2 from Sep 22nd, 2003 to Aug 30th, 2010

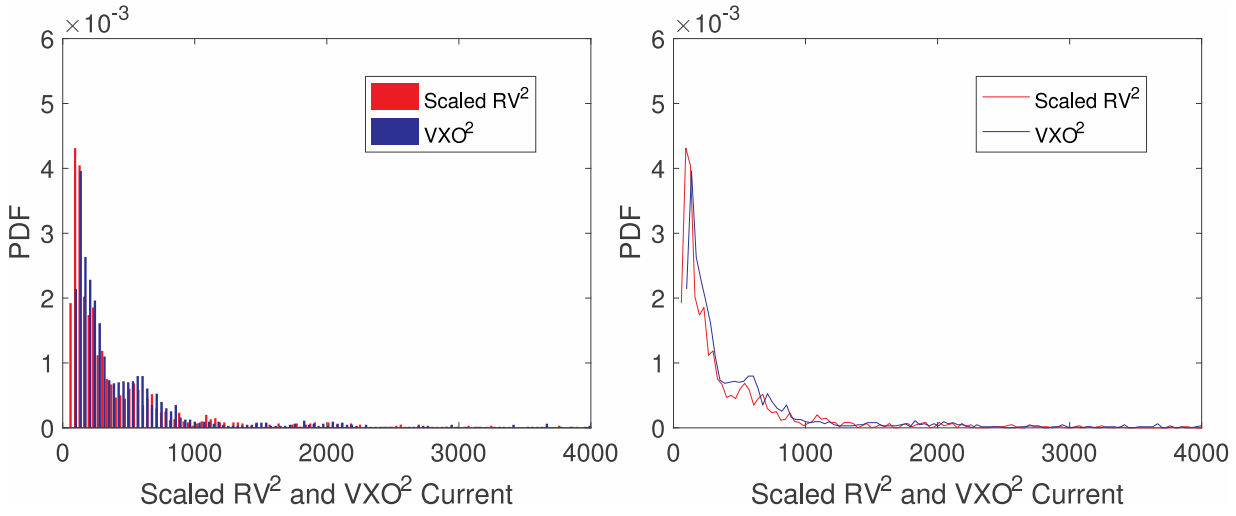


Figure B.29. PDFs of scaled RV^2 and VXO^2 from Sep 22nd, 2003 to Aug 30th, 2010

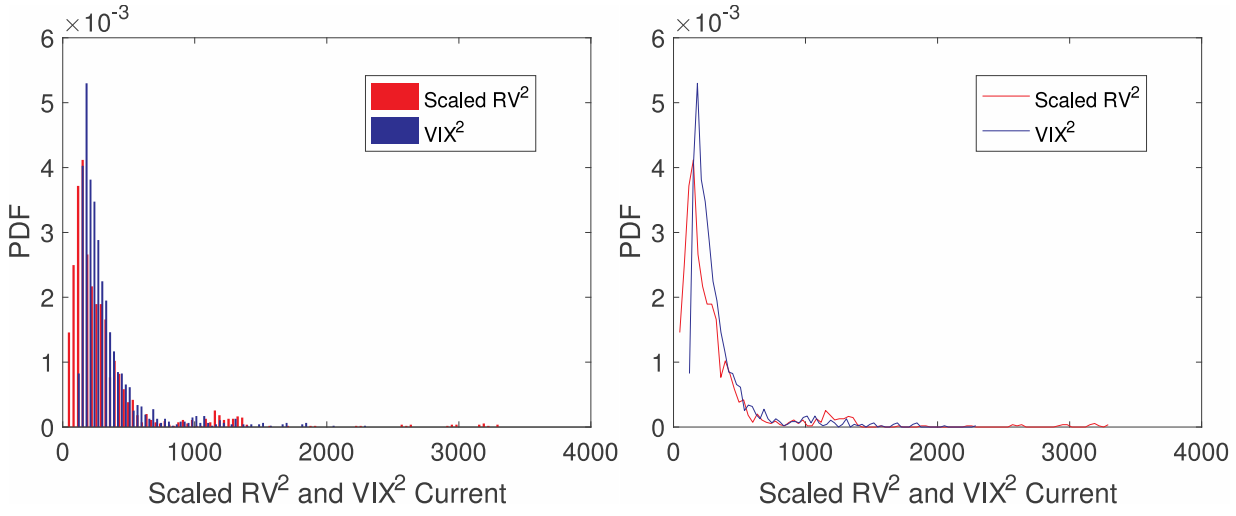


Figure B.30. PDFs of scaled RV^2 and VIX^2 from Aug 31st, 2010 to Dec 30th, 2016

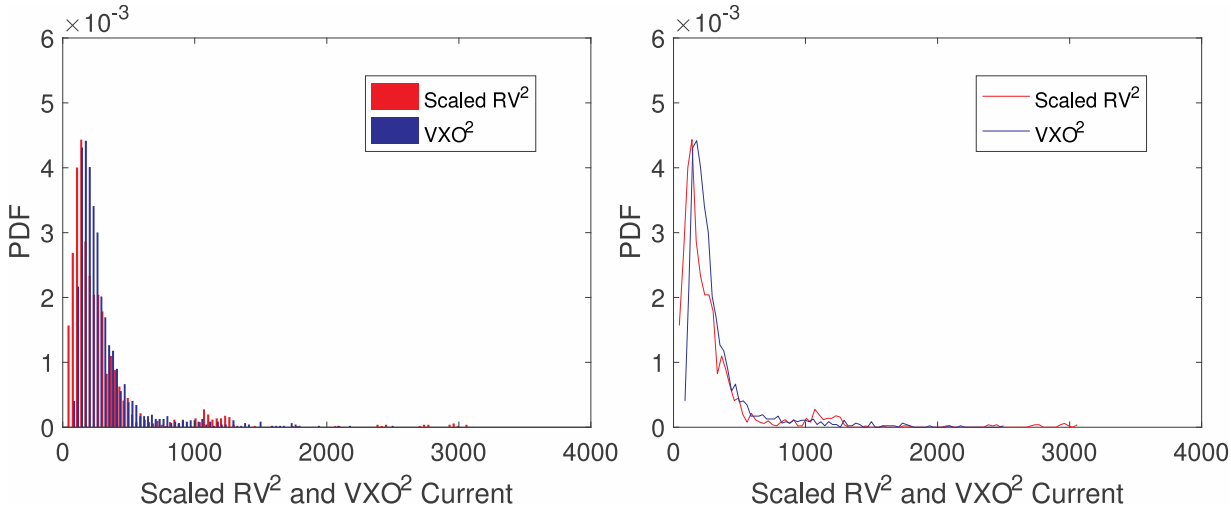


Figure B.31. PDFs of scaled RV^2 and VXO^2 from Aug 31st, 2010 to Dec 30th, 2016

B.2 Ratio Distribution

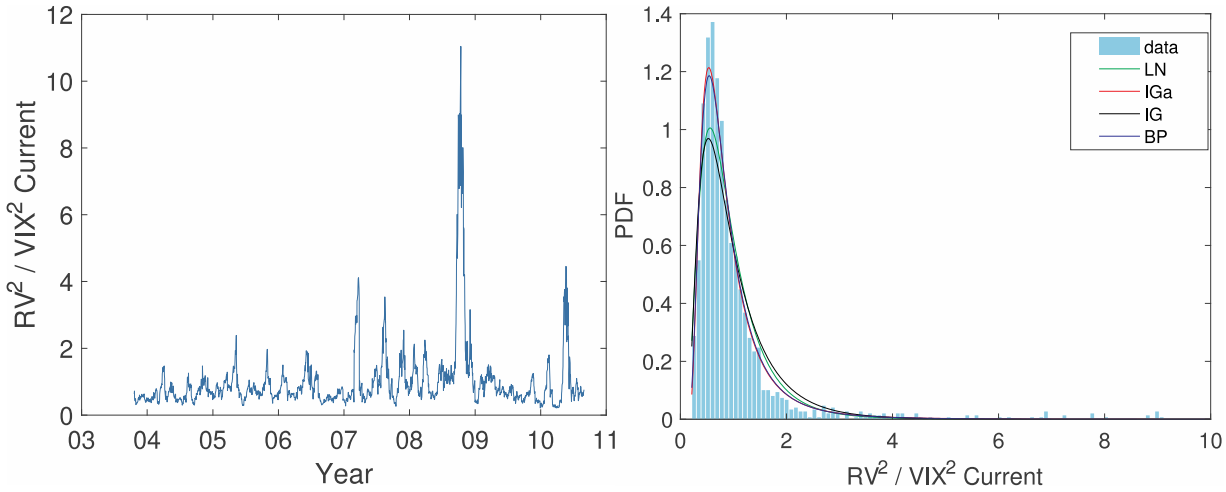


Figure B.32. RV^2/VIX^2, from Sep 22nd, 2003 to Aug 30th, 2010

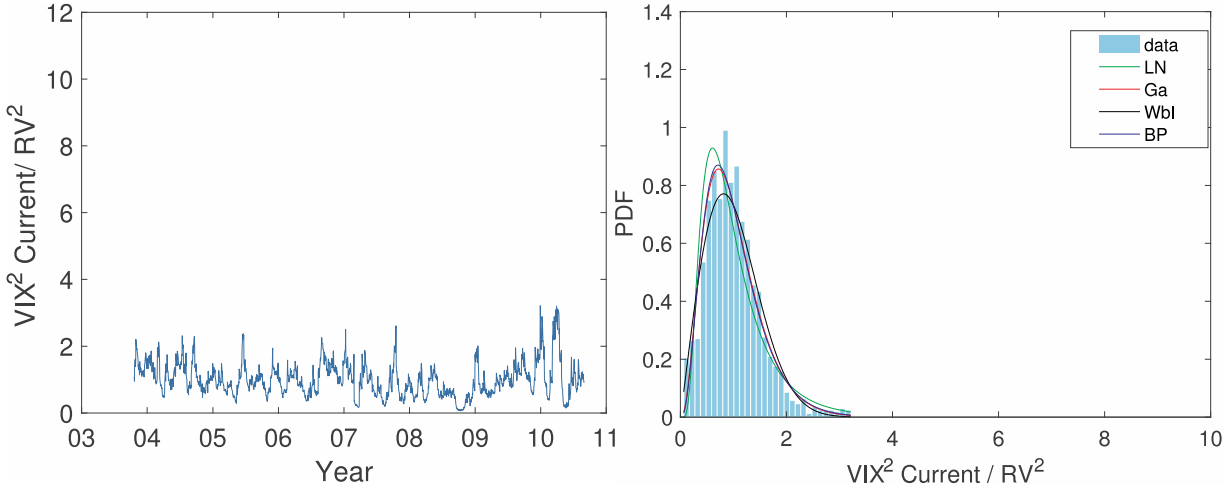


Figure B.33. VIX^2/RV^2, from Sep 22nd, 2003 to Aug 30th, 2010

Table B.16. MLE results for RV^2/VIX^2 and VIX^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000, 1.024)	0.234
LogNormal	LN(-0.228, 0.594)	0.076
IGa	IGa(3.529, 2.422)	0.038
Gamma	Ga(2.346, 0.426)	0.130
Weibull	Wbl(1.101, 1.306)	0.159
IG	IG(1.000, 2.189)	0.097
BP	BP(44.213, 3.782, 0.060)	0.041

type	parameters	KS Statistic
Normal	N(1.000, 0.512)	0.061
LogNormal	LN(-0.148, 0.594)	0.076
IGa	IGa(2.346, 1.610)	0.130
Gamma	Ga(3.529, 0.283)	0.038
Weibull	Wbl(1.130, 2.045)	0.039
IG	IG(1.000, 2.189)	0.113
BP	BP(3.782, 44.213, 12.652)	0.041

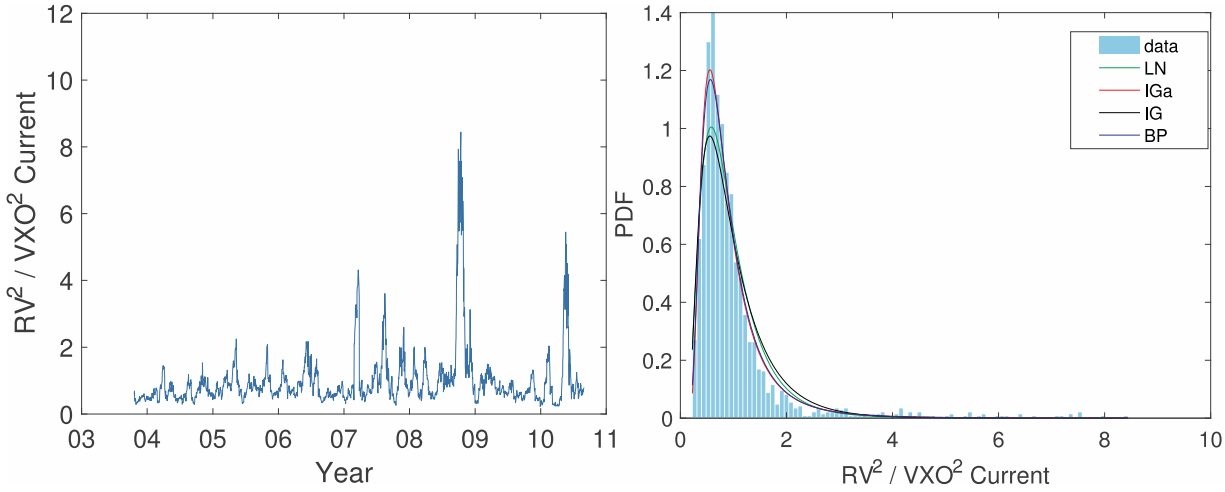


Figure B.34. RV^2/VXO^2, from Sep 22nd, 2003 to Aug 30th, 2010

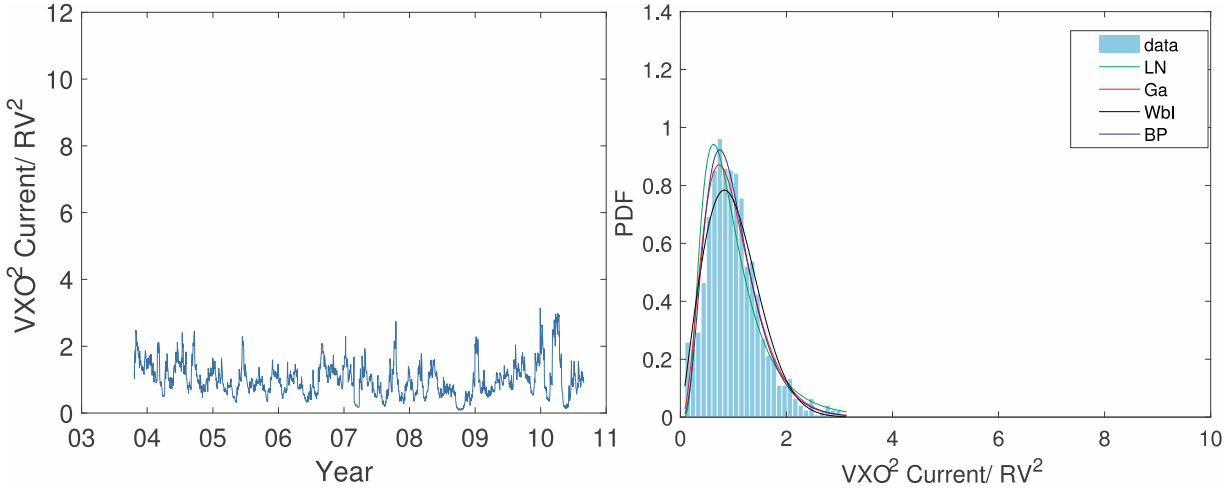


Figure B.35. VXO^2/RV^2, from Sep 22nd, 2003 to Aug 30th, 2010

Table B.17. MLE results for RV^2/VXO^2 and VXO^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000, 0.913)	0.224
LogNormal	LN(-0.207, 0.576)	0.072
IGa	IGa(3.695, 2.608)	0.036
Gamma	Ga(2.571, 0.389)	0.127
Weibull	Wbl(1.112, 1.384)	0.144
IG	IG(1.000, 2.400)	0.092
BP	BP(41.466, 3.512, 0.126)	0.031

type	parameters	KS Statistic
Normal	N(1.000, 0.500)	0.063
LogNormal	LN(-0.141, 0.576)	0.072
IGa	IGa(2.571, 1.815)	0.127
Gamma	Ga(3.695, 0.271)	0.036
Weibull	Wbl(1.130, 2.097)	0.038
IG	IG(1.000, 2.400)	0.105
BP	BP(3.512, 41.466, 8.792)	0.031

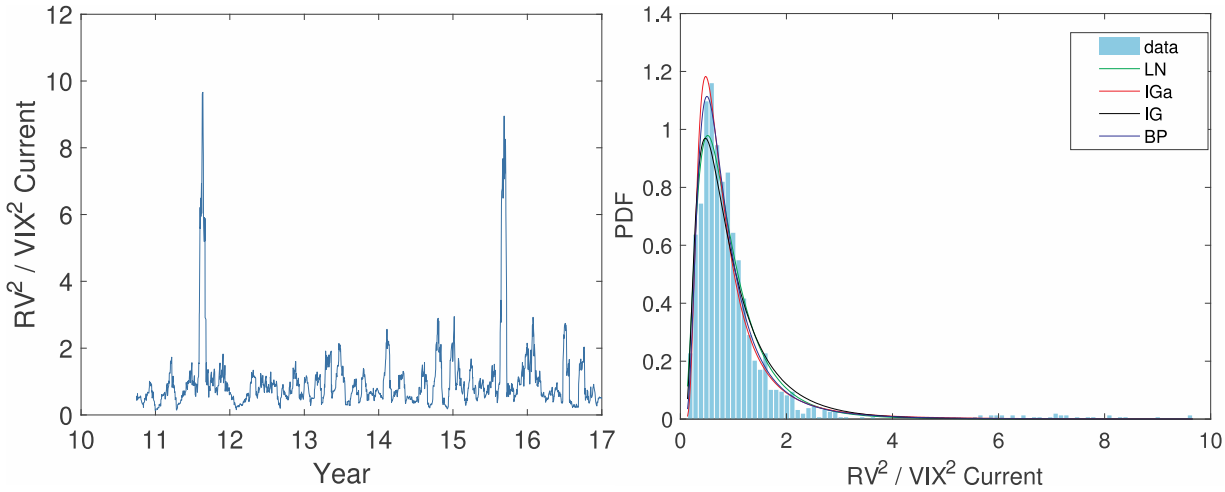


Figure B.36. RV^2/VIX^2, from Aug 31st, 2010 to Dec 30th, 2016

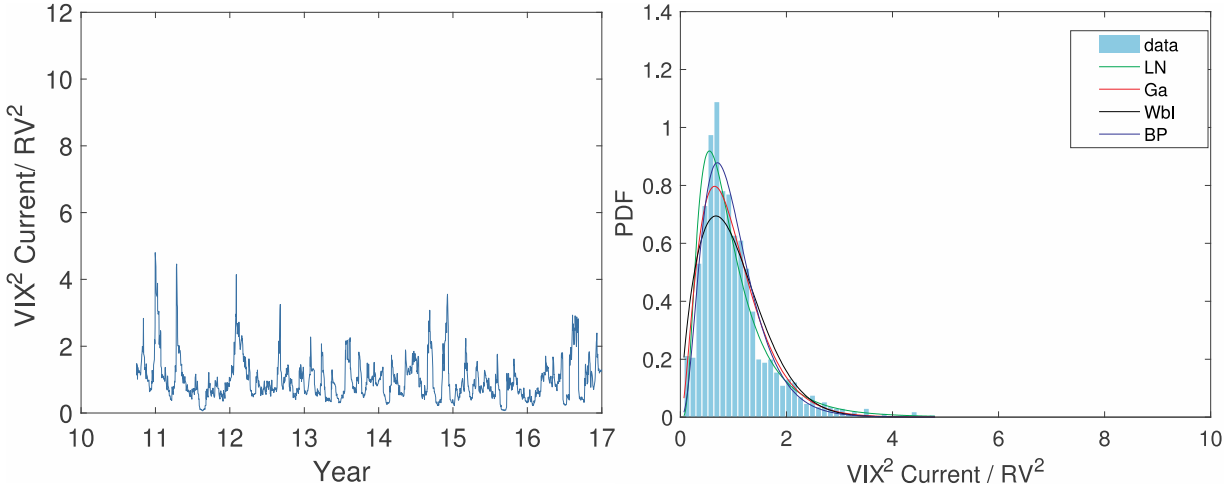


Figure B.37. VIX^2/RV^2, from Aug 31st, 2010 to Dec 30th, 2016

Table B.18. MLE results for RV^2/VIX^2 and VIX^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000, 1.044)	0.225
LogNormal	LN(-0.251, 0.645)	0.051
IGa	IGa(2.819, 1.817)	0.042
Gamma	Ga(2.141, 0.467)	0.108
Weibull	Wbl(1.095, 1.276)	0.125
IG	IG(1.000, 1.814)	0.075
BP	BP(15.139, 3.438, 0.157)	0.033

type	parameters	KS Statistic
Normal	N(1.000, 0.631)	0.111
LogNormal	LN(-0.188, 0.645)	0.051
IGa	IGa(2.141, 1.380)	0.108
Gamma	Ga(2.819, 0.355)	0.042
Weibull	Wbl(1.126, 1.703)	0.065
IG	IG(1.000, 1.814)	0.087
BP	BP(3.438, 15.139, 4.111)	0.033

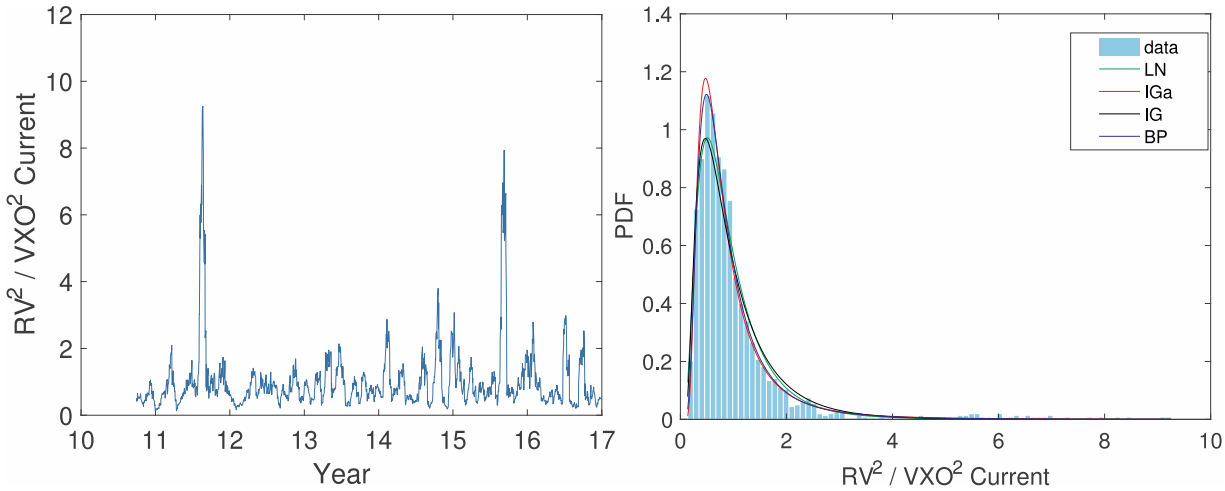


Figure B.38. RV^2/VXO^2, from Aug 31st, 2010 to Dec 30th, 2016

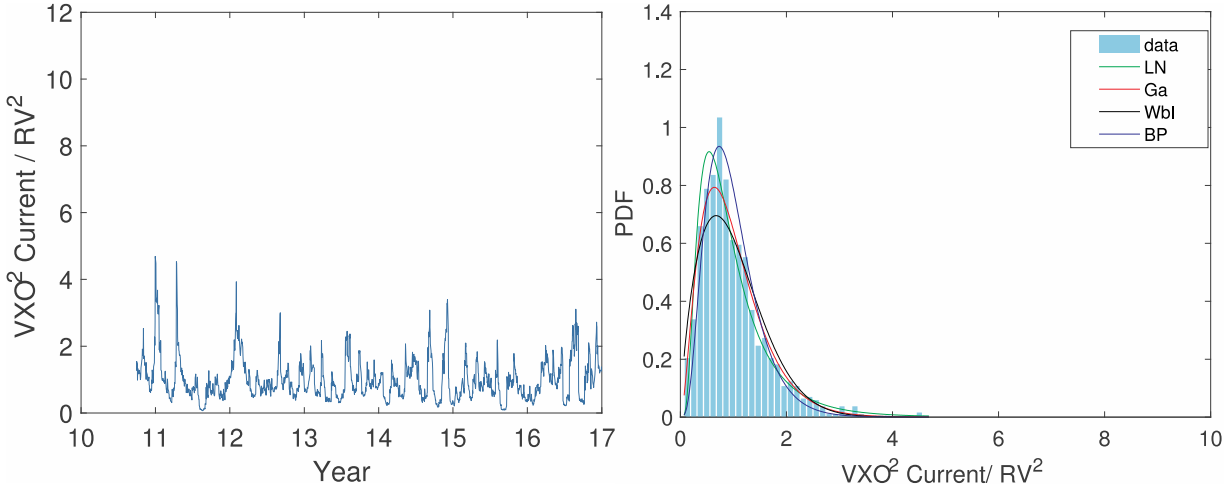


Figure B.39. VXO^2/RV^2, from Aug 31st, 2010 to Dec 30th, 2016

Table B.19. MLE results for RV^2/VXO^2 and VXO^2/RV^2

type	parameters	KS Statistic
Normal	N(1.000, 0.988)	0.206
LogNormal	LN(-0.249, 0.651)	0.050
IGa	IGa(2.777, 1.788)	0.030
Gamma	Ga(2.156, 0.464)	0.109
Weibull	Wbl(1.098, 1.302)	0.115
IG	IG(1.000, 1.808)	0.067
BP	BP(18.957, 3.225, 0.116)	0.024

type	parameters	KS Statistic
Normal	N(1.000, 0.626)	0.106
LogNormal	LN(-0.191, 0.651)	0.050
IGa	IGa(2.156, 1.388)	0.109
Gamma	Ga(2.777, 0.360)	0.030
Weibull	Wbl(1.126, 1.708)	0.052
IG	IG(1.000, 1.808)	0.084
BP	BP(3.225, 18.957, 5.567)	0.024

References

Andersen, T. G., & Benzoni, L. (2014). *Stochastic Volatility*. https://doi.org/10.1007/978-3-642-27737-5_527-3

Are Traders Manipulating the VIX? (2017, May). https://blogs.wsj.com/moneybeat/2017/05/25/are-traders-manipulating-the-vix/.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(2), 253–280. https://doi.org/10.1111/1467-9868.00336

Behfar, S. K. (2016). Long memory behavior of returns after intraday financial jumps. *Physica A: Statistical Mechanics and Its Applications*, *461*, 716–725. https://doi.org/10.1016/j.physa.2016.06.026

Bollerslev, T., Mathew, G., & Zhou, H. (2004). *Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatility*.

Chrstensen, B. J., & Prabhala, N. R. (1998). The Relation Between Implied and Realized Volaility. *Journal of Financial Economics*, *50*, 125–150.

Dashti Moghaddam, M., Liu, J., & Serota, R. A. (2019a). Implied and Realized Volatility: A Study of Distributions and the Distribution of Difference. *ArXiv Preprint ArXiv:1906.02306*. Retrieved from http://arxiv.org/abs/1906.02306

Dashti Moghaddam, M., Liu, Z., & Serota, R. A. (2019b). Distributions of Historic Market Data -- Relaxation and Correlations. *ArXiv Preprint ArXiv:1907.05348*. Retrieved from http://arxiv.org/abs/1907.05348

Dashti Moghaddam, M., & Serota, R. A. (2018). Combined Mutiplicative-Heston Model for Stochastic Volatility. *ArXiv Preprint ArXiv:1807.10793*. Retrieved from https://arxiv.org/pdf/1807.10793.pdf

Demeterfi, K., Derman, E., Kamal, M., & Zou, J. (1999a). A guide to volatility and variance swaps. *The Journal of*

Derivatives, *6*(4), 9–32. https://doi.org/10.3905/jod.1999.319129

Demeterfi, K., Derman, E., Kamal, M., & Zou, J. (1999b). *More than you ever wanted to know* about volatility swaps*.

Dragulescu, A. A., & Yakovenko, V. M. (2002). Probability distribution of returns in the Heston model with stochastic volatility. *Quantitative Finance*, *2*, 445–455. https://doi.org/10.1088/1469-7688/2/6/303

Fuentes, M. A., Gerig, A., & Vicente, J. (2009). Universal behvior of extreme price movements in stock markets. *PLoS ONE*, *4*(12), 1. https://doi.org/10.1371/journal.pone.0008243

Gerig, A., Vicente, J., & Fuentes, M. A. (2009). Model for non-Gaussian intraday stock returns. *Physical Review E*, *80*, 065102(R). https://doi.org/10.1103/PhysRevE.80.065102

Griffin, J. M., & Shams, A. (2017, May). *Manipulation in the VIX?* https://doi.org/10.2139/ssrn.2972979

Han, H., & Park, D. (2013). Comparison of Realized Measure and Implied Volatilityin Forecasting Volatility. *Journal of Forecasting*, *32*, 522–533. https://doi.org/10.1002/for.2253

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, *6*(2), 327–343. https://doi.org/10.1093/rfs/6.2.327

Kownatzki, C. (2016). How good is the VIX as a predictor of market risk? *Journal of Accounting and Finance*, *16*(6), 39–60.

Liu, Z., Dashti Moghaddam, M., & Serota, R. A. (2019). Distributions of historic market data -- stock returns. *European Physics Journal B*, *92*(60), 1–10. https://doi.org/10.1140/epjb/e2019-90218-8

Ma, T., & Serota, R. A. (2014). A model for stock returns and volatility. *Physica A: Statistical Mechanics and Its Applications*, *398*, 89–115. https://doi.org/10.1016/j.physa.2013.11.032

Nelson, D. B. (1990). Arch models as diffusion approximations*. *Journal of Econometrics*, *45*, 7. https://doi.org/10.1016/0304-4076(90)90092-8

Russon, M. D., & Vakil, A. F. (2017). On the non-linear relationship between VIX and realized SP500 volatility. *Investment Management and Financial Innovations*, *14*(2), 200–206. https://doi.org/10.21511/imfi.14(2-1).2017.05

S&P500. (n.d.). https://finance.google.com/finance/historical?cid=626307.

The CBOE volatility index - VIX. (2003). https://www.cboe.com/micro/vix/vixwhite.pdf.

VIX Options and Futures Historical Data. (n.d.). http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data.

Vodenska, I., & Chambers, W. J. (2013). *Understanding the Relationship between VIX and the S&P 500 Index Volatility*. https://doi.org/10.2139/ssrn.2311964

Watch Out VIX: Nasdaq Amps Up Volatility Game. (2017, December). https://www.wsj.com/articles/watch-out-vix-nasdaq-amps-up-volatility-game-1513938600.

Zhou, H., & Chesnes, M. (2003). *VIX index becomes model for and based on S&P 500*.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.