Has the VIX index been manipulated?

Atanu Saha¹ · Burton G. Malkiel² · Alex Rinaudo¹

Revised: 10 December 2018 / Published online: 27 December 2018
© The Author(s) 2018

Abstract Recently, an influential academic study and many lawsuits have claimed that the VIX index has been manipulated since 2008. In this paper, we construct a regression model with explanatory variables that are exogenous to the index and examine the model prediction errors. We find that the movements in the daily levels of the VIX index are explained by market fundamentals and not by manipulation. We also specifically examine the VIX futures expiration days and demonstrate that the VIX closing values and VIX futures settlements prices on those days are consistent normal market forces and are not artificial.

Keywords VIX · VIX futures · Manipulation · VIX options

Introduction

The Chicago Board Options Exchange Volatility Index (VIX) is the most popular measure of the market's expectation of volatility over the near-term future. Introduced in 1993, the VIX is considered to be the premier gauge of investor sentiment. Since the index tends to rise

during falling markets, it is often called 'the fear index.' The VIX hit a high of over 80 during the 2008 market meltdown. During the slowly rising market of 2017, the VIX averaged around 11. There has been a growing acceptance of VIX and VIX-linked products (such as VIX futures and options) for use as risk management tools, and trading of these instruments has expanded dramatically over time. Because of its excellent liquidity and its negative correlation with broad stock market movements, VIX-linked products are particularly useful hedging instruments. Portfolio managers can mitigate downward movements in the general level of stock prices by buying volatility, i.e., by buying VIX futures and options.

The market for VIX-related financial instruments, such as futures, options and exchange-traded funds, has risen steadily over the years. While estimates of VIX exposure vary, one analyst suggests it could be as high as \$60 billion.¹ Given the size and reach of numerous financial products linked to the VIX, any artificial inflation or deflation of the index can have widespread ramifications, including substantial damages suffered by various parties.

The claims that the VIX has been manipulated

The VIX index gauges the implied 30-day volatility of the market calculated from options on the Standard and Poor's 500 stock index (S&P 500). Futures and options on the VIX itself have a relatively large volume of trading. The value of the VIX is calculated from a wide range of out-of-the-money options, and some of the far out-of-the-money options are thinly traded. In a recent study, Griffin and Shams (2018) have argued that a market participant could

✉ Atanu Saha
asaha@econone.com

Burton G. Malkiel
bmalkiel@princeton.edu

Alex Rinaudo
arinaudo@econone.com

¹ EconOne, New York, USA

² Chemical Bank Chairman's Professor of Economics, Emeritus, Princeton University, Princeton, USA

¹ Christopher R. Cole, "Volatility and the Alchemy of Risk", Artemis Capital Management, October 2017.

manipulate the VIX futures settlement prices by trading in the thinly traded, far out-of-the-money less liquid options used to calculate the VIX. They found that during settlement periods, volume spikes have occurred in these thinly traded options referencing the VIX. They concluded that such trading patterns are consistent with market manipulations during the period 2008 to the present.

The Griffin and Shams (2018) study has been quite influential and has been cited in many recent lawsuits alleging manipulation of the VIX. The plaintiffs in these suits claim, for example, that ‘a select group of financial institutions and trading firms with sophisticated, expensive technology’² are engaging in ‘rampant manipulations of the VIX index.’³ The news media outlets have also paid considerable attention to this issue. For example, *The New York Times* cites markets experts who believe that traders ‘who persistently short the VIX have distorted the market,’⁴ and *Barron’s* reported in April 2018 that investors ‘suspected that someone was trying to manipulate the VIX, which had spiked suddenly just a few weeks before, roiling financial markets.’⁵

The goal of this paper is to take a direct approach to examine the hypothesis that the VIX has been manipulated from 2008 to the present. Surely, the proof of the pudding is whether VIX levels themselves have displayed a different pattern during the period from 2008 to the present than it did in earlier periods. We also examine whether the VIX levels were artificially inflated or deflated on the VIX futures and options expirations days. It is important to note that our analysis uses daily closing levels of the VIX. As a result, in this paper we do not examine whether the effect, if any, of the alleged manipulation of the VIX lasted only for a brief period of time within the futures expirations days. Such an analysis, which requires intraday data, is beyond the scope of this paper.

This paper’s analyses have two components. First, we examine the daily closing levels of the VIX over the past two decades. We fit a model that regresses the VIX on a set of regressors using data from the 10-year period 1998–2007 (the ‘in sample’); we then use the same set of regressors to predict the VIX over the period at issue: 2008 through April 2018 (the ‘out of sample’). If the VIX were manipulated since 2008, then the out-of-sample prediction errors (i.e., the regression residuals) would be expected to be higher than the ‘in-sample’ errors. This is because, if

there was manipulation, then its effects would become evident in the errors, which reflect the portion of the VIX’s movement unexplained by the model’s regressors, which are free from any manipulation claims. Various tests show that the ‘out-of-sample’ prediction errors are not higher than the ‘in-sample’ ones. These results imply that in both periods—the period at issue and the 10 years preceding it—market forces explain the VIX and there is no indication of artificiality in the level of the VIX.⁶ The only time period within 2008–2018 where the prediction errors are unusually large is the fourth quarter of 2008. However, as is widely known, these 3 months were characterized by extreme market turbulence and the VIX spiked to record levels. Thus, large prediction errors during this 3-month period are to be expected because a simple linear regression model cannot capture the full impact of the unprecedented market disruptions and the resulting jumps in the VIX index.

The second component of this paper’s analyses focuses on the expiration days of VIX futures. Recall, Griffin and Shams found unusual trading patterns on these expiration days, and thus, we inspect whether there is evidence of manipulation in these specific days. In particular, we examine how many of the 234 VIX futures and options expiration days in the 2008 to April 2018 period have statistically significant residuals. We find that the number of significant days is no different than what one would expect based on random chance, and thus, our results do not support a claim of manipulation on VIX futures expiration days. We also examine whether the settlement prices of the VIX futures contracts were artificial. Again, we find that empirical evidence does not support the claim of artificiality of the settlement prices. In sum, we have examined the claim of manipulation for: (a) the period 2008 to the present, (b) the VIX closing values on settlement days and (c) the VIX futures settlement prices. All three analyses suggest that the VIX is explained by market fundamentals⁷ and strongly reject the hypothesis of manipulation.

This paper is organized as follows. In the next section, we provide a brief background for the VIX and review the relevant literature. In the third section, we describe the data and methodology used for our analyses; we also set out the regression model and the general framework for the analysis. In the fourth section, we discuss the regression results, focusing on the various parametric and nonparametric tests of the difference between ‘in-sample’ and ‘out-of-sample’

² Jeffery Tomasulo v. CBOE et al., Illinois Northern District Court, 2018.

³ Siegel v. CBOE et al Complaint, Illinois Northern District Court, 2018.

⁴ Stephen Voss, “Was the VIX Fixed?” *New York Times*, February 14, 2018.

⁵ Crystal Kim, “No Tricks with the VIX,” *Barron’s*, April 24, 2018.

⁶ This sentence has been crafted for the ease of exposition. A more formal statistical phrasing of our finding would be: The null hypothesis that market forces explain the movement of VIX is not rejected.

⁷ Again, more formally, the null hypothesis that market fundamentals explain the movement of VIX is not rejected.

regression residuals. In the fifth section, we describe the data on VIX futures expiration days and then examine whether the residuals on those days are statistically significant. In the final section of the paper, we briefly discuss why our results are not necessarily incongruent with the findings of Griffin and Shams.

Relevant background and the literature

The Chicago Board Options Exchange (CBOE 2018) provides an excellent description of the history and mechanics of the VIX. At its inception in 1993, the VIX was based on the implied volatility of at-the-money S&P 100 option prices. The intent was to provide a reliable estimate of short-term stock market volatility and to ‘offer a market volatility ‘standard’ upon which derivative contracts may be written.’⁸ The method of calculation of the index was revised in 2003. The new method, currently employed, estimates expected volatility by using the weighted average prices of a wide range of strikes of puts and calls of S&P 500 options expiring in approximately 30 days. The methodology was further refined in 2014 to include S&P 500 weekly options.

CBOE (2018) documents that an inverse relationship between the market and the VIX tends to hold roughly 80% of the time. Despite the fact that the VIX is often viewed as a hedge against market downturns or a proxy for investor sentiment, it is important to note that the VIX is simply a formulaic representation of derived 30-day forward volatility expectations based upon S&P 500 option prices. As co-creator Devish Shah explains, the VIX is akin to measuring ‘the temperature outside. If it’s the winter it’s going to be really [cold], and if it’s summer it’s going to be really hot. It’s not the cold index or the heat index, it’s just the temperature.’⁹

The prior literature

The prior literature relevant to our study falls into two broad areas: The first has examined the market forces that drive the VIX, and the second relates to the methods for detecting manipulation.

Three recent studies have focused on factors explaining the movement of the VIX. Hait (2016) examined the relationship between the VIX and S&P 500 returns and found that 98.8% of the daily variation in the VIX can be explained by current S&P returns and lagged VIX values.

Vodenska and Chambers (2013) compared the movement in the VIX with 22-day realized S&P volatility and found that over 80% of the variation in the VIX is explained by this one variable alone. Ozair (2014) focused on the impact of market shocks on the index. He found that the impacts of a shock persisted in the VIX and these shocks account for nearly 70% of the variation in the VIX.

Other studies have looked at the asymmetric movement of the VIX. For example, Zakamulin (2016) found that the durations for the periods of rising and falling VIX are unequal: The timespan of falling periods exceeds that of rising VIX by a factor of 1.4. This finding suggests that the VIX often experiences sudden leaps, which take a fairly long time to subside. Chow et al. (2018) found that the VIX estimation errors—between realized volatility and the VIX—are considerably larger during volatile markets.

Turning next to the literature on manipulation, we note that investors, academics and regulators generally fail to agree on the definition of manipulation. Furthermore, they disagree on how manipulation can be discerned from transactions, if at all. This is, in large part, because in any market, each order or transaction, particularly larger ones, can affect market prices. As a result, distinguishing manipulative transactions from legitimate ones using price effects can be challenging.

One branch of the literature on market manipulation posits that the intent behind the trading activity is the key to determining whether manipulation occurred. For example, Perdue (1987–1988) focused on whether the conduct of the relevant party involved was reasonable. If the conduct was uneconomical or irrational, then such conduct could indicate manipulative intent.

Rather than focusing on unobservable ‘intent’ of a trade, Pirrong (1996) proposed an economic model of manipulation based on observable variables and discussed the conditions that facilitate manipulation. In a subsequent article, Pirrong (2004) set out a number of econometric tests to detect price and quantity patterns symptomatic of manipulation.

Abrantes-Metz et al. (2013) used data on prices, bids, quotes, spreads, market shares, and especially volumes to identify patterns that are anomalous or highly improbable, and concluded that such patterns could indicate manipulation. The Griffin and Shams (2018) study has a similar theme. They focused on the ‘highly unusual’ trading activity in the underlying options used to determine settlement prices on VIX futures expiration dates. The final settlement price for expiring VIX futures is determined during a 30-min period known as the Special Opening Quotation (SOQ) on futures expiration days. Griffin and Shams (2018) observed volume spikes in the S&P 500 option book during the SOQ. They then examined various alternative explanations of these volume spikes and found

⁸ CBOE (2018).

⁹ Max Abelson and Joe Weisenthal, “An Inventor of the VIX: ‘I Don’t Know Why These Products Exist,’” Bloomberg, February 6, 2018.

that none was supported by data. As a result, they concluded that these spikes were consistent with manipulation of the VIX.

Most prior empirical studies on manipulation have examined the price movements of the financial instruments or commodity at issue and not the trades of the alleged manipulator. This is because data on an alleged manipulator's trades are proprietary and are virtually impossible to acquire. However, empirical studies related to the Amaranth matter were exceptions. Amaranth LLC was a hedge fund that closed down in 2006 and faced allegations of manipulation of natural gas futures. The Senate investigation yielded data on Amaranth's natural gas futures trades, and a number of studies have relied on this dataset. For example, Marthinsen and Gai (2010a) used a Granger causality model to analyze whether Amaranth's trades affected the prices of natural gas futures in 2006. In a follow-up article, the authors examined whether Amaranth's spread trading affected prices of calendar spreads, particularly the winter–summer spreads (Marthinsen and Gai, 2010b). Saha and Petersen (2012) also used the Amaranth dataset. They proposed a method to examine both whether prices were artificial and whether alleged manipulator's trades caused the price artificiality. Their methodology involved creating a model to explain futures prices using market fundamentals and then examining the correlation between the 'errors' (that is, the difference between actual and model-predicted prices) and the alleged manipulator's trades.

This article contributes to the existing literature by developing a framework for the examination of VIX manipulation claims using the relationship between the VIX and market fundamentals. Like Saha and Petersen (2012), we examine the pattern of the estimation model's 'errors' to determine whether manipulation of the index occurred. As noted earlier, some prior studies have included lagged value of the VIX as an explanatory variable in the regression model for the VIX. We do not include this variable in our model since our goal is to examine whether or not manipulation occurred, and therefore, inclusion of a lagged value of a variable which itself may contain the effects of alleged manipulation would contaminate our analysis.

The data and the general methodology

The CBOE Web site provides historical data on the VIX using the current methodology backfilled to 1990. As noted earlier, CBOE revised its methodology of calculating the VIX in 2003 and in 2014. However, CBOE provides data on the VIX going back to its inception, with the daily closing values of the index calculated using the most

current (i.e., 2014) methodology. Throughout this paper, we have used these backfilled data on VIX, which precludes the possibility of the VIX values being affected by changes in the methodology of calculating the index.

Our analysis uses the daily closing values of the index for the roughly 20-year period, 1998 through April 2018. These 2 decades allow us to use a 'clean' period (i.e., 1998–2007) with a similar length of time to the period in question (i.e., January 2008–April 2018). We also gathered data on the daily closing values of the S&P 500 index and computed its daily log-returns, denoted by 'Spr' in the table below. This variable was then used to compute the two key regressors: the 20-day rolling volatility and the 5-day rolling mean of the S&P daily returns. These two variables are denoted by 'Spv' and 'Spm.' We also created two indicator variables for a day with a positive (sp₊) and a negative (sp₋) return for the S&P index. The summary statistics for these variables are reported in Table 1.¹⁰

Before undertaking the regression analysis, we examined whether the time series on the VIX index is stationary. In particular, we implemented the augmented Dickey–Fuller (DF) test to determine whether the VIX follows a unit root process.¹¹ The test results strongly rejected the null hypothesis that the VIX follows a non-stationary process, that is, it has a unit root.¹² Based largely on these results of the test for stationarity, and as in many prior studies, we have chosen to use a levels model of the VIX. The explained variable in the regression model is the level of VIX rather than the daily changes in the index. However, we observed that the mean level of the VIX in the 2008–2018 period was slightly lower than the preceding 10 years' level. We thus 'de-mean-ed' the explained variable, that is, for each of the two sub-periods we subtracted the respective means from the daily VIX values.

The explanatory variables chosen for the regression are generally consistent with the prior studies discussed earlier. In particular, we posit that the VIX is explained by two key variables: the 20-day realized volatility and the realized

¹⁰ See 'Summary statistics for two sub-periods' of "Appendix" for the summary statistics for the two separate periods: 1997–2007 and 2008–2018.

¹¹ The augmented *DF* test was implemented by fitting the model $\Delta Vix_t = \beta Vix_{t-1} + \sum_{j=1}^k \alpha_j \Delta Vix_{t-j} + \varepsilon_t$ (Dickey and Fuller 1979). Testing $\beta = 0$ is equivalent to testing that the process has a unit root, i.e., that it is non-stationary. For the VIX, the null hypothesis $H_0 : \beta = 0$ was tested for up to 20 lags (i.e., $k = 20$) and was overwhelmingly rejected in each case. See 'Dickey–Fuller test for unit root' of "Appendix" for further details.

¹² We ran these tests through the entire 20-year time period, as well as the pre-2008 and post-2008 periods, separately. Finally, we also ran these tests by excluding the fourth quarter of 2008. In each case, the results of the analyses were the same: The non-stationarity of VIX was rejected. For details, see 'Dickey–Fuller test for unit root' of "Appendix."

Table 1 Summary statistics (Jan 1998–April 2018)

Variable	Variable name	Mean	SD	Min	Max
VIX	Vix	20.3414	8.5626	9.14	80.86
SP daily return	Spr	0.0002	0.0121	− 0.0947	0.1096
SP 20-day rolling volatility	Spv	0.0104	0.0064	0.0021	0.0537
SP 5-day rolling mean	Spm	0.0002	0.0050	− 0.0405	0.0350
Indicator variable for +SP day	sp ₊	0.5338	0.4989	0	1
Indicator variable for −SP day	sp _−	0.4662	0.4989	0	1
	# of Obs	5114			

average 5-day returns of the S&P 500 index; each of these two variables is lagged by a day and interacted with indicator variables for a positive or a negative return day. The regression equation is shown in (1):

$$VIX_t = \beta_0 + \beta_1 \cdot Spv_{t-1} \cdot sp_+ + \beta_2 \cdot Spv_{t-1} \cdot sp_- + \beta_3 \cdot Spm_{t-1} \cdot sp_+ + \beta_4 \cdot Spm_{t-1} \cdot sp_- + \varepsilon_t \quad (1)$$

In choosing the regressors in (1), we were careful not to include variables that could have been affected by the alleged manipulation. For example, as indicated earlier, this was our rationale for not including a lagged value of the VIX.¹³ Similarly, we chose not to use contemporaneous values of the regressors and used lagged values instead.¹⁴ One might argue that any given day's level of the VIX can potentially affect that day's realized volatility, thereby creating a simultaneity problem in the regression analysis. The inclusion of the lagged values of the regressor avoids this problem because any given day's VIX level cannot affect the preceding day values of realized volatility.

The model in (1) was estimated using daily data from the 'clean' period or the in-sample period (i.e., January 1998–December 2007). We then used the estimated coefficients to predict the daily levels of the VIX during the out-of-sample period, that is, the period at issue (i.e., January 2008–April 2018). We then compared the regression errors (i.e., the residuals) between the actual and the predicted levels of the VIX in the two periods to determine whether they are statistically significantly different. If manipulation of the index was evident, then one would expect the residuals (absolute or squared value) to be larger in the post-2008 period.¹⁵

¹³ Had we included the VIX_{t-1} in the right-hand side of (1), the explanatory power of the model would have been significantly higher (the adjusted R^2 becomes 0.96 versus 0.74 without including VIX_{t-1}).

¹⁴ Had we utilized same day (i.e., un-lagged) values of the regressors in the right-hand side of (1), the explanatory power of the model improves somewhat and all the main findings of the paper remain unchanged.

¹⁵ Note, because the VIX has been de-measured, the mean value of the index in the two periods is exactly zero. Thus, the difference in the absolute value or squared value of the residuals from the two periods cannot be explained by the differences in the two period's average level of the VIX, the explained variable in the model.

Table 2 Regression results

Dependent variable: daily closing value of VIX (1998–2007)		
Explanatory variable	Coefficient	T-stat
20-day Vol*(+SP day)	1162.33	71.55
20-day Vol*(−SP day)	1309.95	79.31
SP 5-day rolling mean*(+SP day)	− 259.09	− 12.92
SP 5-day rolling mean*(−SP day)	− 382.91	− 17.90
Intercept	− 12.78	− 73.88
# of Obs		2514
Adjusted R^2		0.7392

As noted earlier, we addressed the issue of whether VIX levels were different across the two periods by de-meaning VIX in both periods. We have also tested whether the volatility of VIX was different in the post-2008 period relative to the pre-period. For example, if the volatility of VIX (the dependent variable in the regression) was lower in the 2008–2018 period, it could lead to smaller residuals for that period; however, in that case, the smaller residuals would not indicate a better fit of the model, but rather would be an artifact of lower volatility of the dependent variable. However, our tests showed that the volatility of VIX was actually higher in the 2008–2018 period. The details of these tests and the results are contained in the 'Bartlett's test for equal variance of VIX' of 'Appendix' to this paper.

The results of the analyses

The regression results are shown in Table 2. The estimation model's predictions are the focus of our analysis and not the statistical significance of the estimated coefficients.

Footnote 15 continued

Furthermore, since the residuals during the period at issue are out-of-sample residuals, a potential manipulator would be able to move the VIX level to the model-predicted value (thereby decrease the size of the residuals) only if the manipulator could know with certainty the in-sample modeled relationship between the regressors and VIX, and be certain that the same relationship would continue to hold during the period at issue.

Fig. 1 Actual and predicted VIX index: 1998–April 2018

However, as the t statistics in Table 2 indicate, all the coefficients are statistically significant.¹⁶

In Fig. 1, we display the actual and predicted levels of the VIX, for both the in-sample and out-of-sample periods. As is evident from this figure, the model performs well in predicting the daily level of the VIX during both periods.¹⁷ The plot of the regression residuals is shown in the ‘Regression residuals’ of ‘Appendix.’

In order to test the predictive accuracy of the model, we examined several measures. One such widely used measure is the Theil’s U statistic.¹⁸ A lower value of the statistic indicates higher predictive accuracy. We found that the value of Theil’s U was considerably lower in the 2008–2018 period than in the 10 years preceding it. This result shows that the regression model has a *higher*

predictive accuracy during the period at issue than in the pre-2008 period.¹⁹ This finding was also corroborated by other measures of the predictive accuracy [such as mean absolute error (MAE) and root mean squared error (RMSE)] discussed later in the paper.

Testing for structural break

Before undertaking a comparative analysis of the in-sample and out-of-sample residuals, we tested whether the relationship between the market fundamentals and the VIX shows evidence of structural break in the post-2008 period. In particular, we estimated our regression Eq. (1) using two different periods’ data: 1998–2007 and 2008–2018. We then tested whether the estimated coefficients of the regressions using the two periods’ data were significantly different.²⁰ The results of the tests for equality of the estimated coefficients are shown in Table 3. In this table, we report the p values of the test under the null hypothesis

¹⁶ We also estimated the model using the Newey–West correction for heteroskedasticity and autocorrelation. All coefficients remain statistically significant.

¹⁷ For the purposes of depiction of the actual and predicted values of the VIX in Fig. 1, we have used a model that has the actual and not de-meaned value of the index. However, with the exception of this chart, we have consistently used the de-meaned value of VIX for all empirical analyses in this paper.

¹⁸ Greene (1997); p. 373.

¹⁹ The details of the Theil’s U computation are discussed in ‘Theil’s U statistics’ of ‘Appendix.’

²⁰ Further details of the tests for structural break are contained in the ‘Testing for structural break’ of ‘Appendix.’

Table 3 Test for structural break

Explanatory variable	The p values for test of equality of estimated coefficients	
	[A] '98-'07 = '08-'18	[B] Excluding 4Q 2008 '98-'07 = '08-'18
20-day Vol*(+SP day)	0.000***	0.401
20-day Vol*(-SP day)	0.000***	0.588
SP 5-day rolling mean*(+SP day)	0.122	0.129
SP 5-day rolling mean*(-SP day)	0.595	0.163
Intercept	0.000***	0.384

***Denotes significant at 1% level

that the two periods' estimated coefficients are equal to each other, which would imply an absence of structural break.

The p values in column [A] of this table are indicative of a structural break: Three of the five coefficients are significantly different from each other. However, further investigation of this issue reveals that the results are driven by data from the fourth quarter of 2008, when the VIX rose to its highest level ever recorded. In column [B], we undertake the same tests of equality of the coefficients, but this time, the data for the first regression remain the same, while the second regression uses data for 2008–2018 period excluding the fourth quarter of 2008 (henceforth 2008 Q4). The p values in column [B] show that none of the coefficients are significantly different from each other, indicating the absence of a structural break. In other words, the relationship between the explanatory variables and the VIX is essentially unchanged in the two periods, when the second period excludes the 2008 Q4. The importance of these unusually volatile 3 months will also be apparent when we undertake the comparative analysis of the prediction errors, next.

Analysis of the prediction errors

The estimated coefficients from the base 1998–2007 period shown in Table 2 are then used to predict the VIX and to compute the in-sample and out-of-sample residuals. We then run various tests to see whether the two sets of residuals are statistically different from each other. However, before undertaking these tests, it is important to note that tests strongly rejected the hypothesis that the residuals are normally distributed.²¹ Consequently, traditional tests of equality of means (t test, Z test, etc.) might be unreliable, since they are based on the distributional assumption of normality. However, in the interest of completeness, we have undertaken both parametric tests (assuming

normality) and nonparametric (i.e., distribution-free) tests. The results are shown in Table 4.

Because the out-of-sample period includes 2008 Q4, a period of 3 months marked by large and unprecedented spikes in the VIX, we compared the means (and medians) of the in-sample residuals with two sets of out-of-sample residuals: one that includes 2008 Q4 and the other that does not. As is evident from the results in Table 4, the residuals for 2008 Q4 are generally much larger and their exclusion makes a large difference in the test of means.²² However, the medians are far less affected by inclusion of 2008 Q4, since the median is less sensitive to outliers than the mean. We compared the means and medians of both measures of the residuals: their absolute and their squared values.

As shown in Table 4, when the absolute value of the residuals is compared, the out-of-sample residuals are in fact *smaller* than the in-sample ones, and this holds true whether or not 2008 Q4 is included in the out-of-sample period. We also report the p value of the one-sided test of means.²³ In this test, the null hypothesis is that the means of the two periods are equal; it is tested against the alternative that the residuals in the out-of-sample period are larger.

For the squared residuals, we report the widely used measure of forecast accuracy: the root mean squared error (RMSE), which is the square root of the mean of the squared residuals.²⁴ The in-sample RMSE is significantly larger than the out of sample; the p value of 0.02 rejects the null that the two means are equal. However, this difference in means is driven by the large residuals in 2008 Q4. When that quarter is excluded from the out-of-sample period, the RMSE actually becomes lower than the in-sample period's, although the difference is not statistically significant (the

²¹ Details of the tests for normality are contained in the 'Testing for normality of residuals' of "Appendix."

²² This is consistent with Chow et al. (2018), who found that the estimation errors were considerably larger during volatile markets.

²³ Each of these tests of means was undertaken using a t test, as the test statistics for the difference of means test is distributed as t distribution. Mood: Graybill et al. (1974), p. 435.

²⁴ Greene (1997), p. 372.

Table 4 Analysis of regression residuals

Measures of forecast accuracy	In sample	Out of sample		Test	Test w/o 2008 Q4
	1998–2007	2008–2018	2008–2018–ex2008Q4	<i>p</i> value	<i>p</i> value
<i>Panel A: parametric tests</i>					
Mean of absolute errors (MAE)	2.8089	2.7243	2.6013	0.8991	0.9995
Median of absolute errors	2.4238	2.0695	2.0193		
Root mean of squared errors (RSME)	3.5386	3.7482	3.4968	0.0194	0.6786
Median of squared residuals	5.8750	4.2827	4.0775		
			2008–2018% of draws	2008–2018–ex2008Q4	
<i>Panel B: nonparametric test (Monte Carlo) of squared residuals</i>					
Mean out of sample $\geq$ mean in sample			73.60%	43.49%	
Median out of sample $\geq$ median in sample			7.22%	4.49%	
Wilcoxon signed-rank test					
Null hypothesis: in and out of samples have the same distribution			<i>p</i> value: 0.2173		
	In sample	Out of sample			
	1998–2007	2008 Q4		2008–2018–ex2008Q4	
<i>Panel C: statistically significant observations (%)</i>					
Nonparametric	4.57%	43.75%		4.56%	
Assuming normality	5.69%	45.31%		5.17%	

p value is 0.68).²⁵ Importantly, when one compares the median of the squared residuals, the out-of-sample median is lower, regardless of whether 2008 Q4 is included in the out-of-sample period. These tests show that one cannot find in the data any evidence of unusual values of the VIX during the period of alleged manipulation.

Panel B of the table contains the results of the nonparametric test. This test is undertaken through a Monte Carlo (akin to bootstrapping) exercise: Two samples of squared residuals, each with 253 observations, are randomly drawn (with replacement) from the in-sample and the out-of-sample periods. We chose a random sample size of 253 data points because it constitutes approximately 10% of the observations in each period.²⁶ For each random draw, we compare the means and medians of the two samples. This process is repeated 10,000 times. Then, we calculate the percentage of the draws where the average out-of-sample squared residuals were greater than the average in-sample ones. If this percentage was greater than

50%, this would indicate that more than half of the time the average out-of-sample residuals were larger than the average in-sample ones, that is, on average, the model's out-of-sample forecasts were worse than the in-sample ones. By contrast, a percentage lower than 50% would imply superior forecasting performance of the estimation model in the 2008–2018 period, relative to the estimation period of 1998–2007.

The fact that large residuals are clustered in 2008 Q4 is particularly clear in the Monte Carlo exercise: The percentage of random draws where the mean of the out-of-sample squared residuals is larger drops from 73.6 to 43.5% when 2008 Q4 is excluded from the out-of-sample period. By contrast, when the medians of the randomly drawn samples are compared, less than 8% of the time the median of the out-of-sample residuals is larger, and this is true regardless of whether one includes 2008 Q4 in the out-of-sample period. Therefore, the nonparametric results are consistent with the traditional test results.

In panel B of the table, we also report the results of the nonparametric Wilcoxon signed-rank test, which is used to determine whether two samples were selected from populations having the same distribution. The *p* value of the test, 0.22, also cannot reject the hypotheses that the residuals in the out-of-sample period come from the same distribution as those of the in-sample period.

Panel C of the table provides further evidence on the clustering of large residuals in 2008 Q4. As expected,

²⁵ We are aware that throwing out the most extreme quarter of the test sample without throwing out the most extreme quarter from the estimation sample could introduce a bias. However, the bias, if any, would work in favor of finding that the residuals in the test sample are larger than in the estimation sample, because throwing out the most extreme quarter from the estimation sample can only reduce the size of its residuals.

²⁶ Our results remain virtually unchanged if somewhat smaller or larger random samples were drawn.

roughly 5% of the in-sample residuals are statistically significant.²⁷ But approximately 44% of the residuals are statistically significant in 2008 Q4, and approximately 5% of the residuals are statistically significant in the remaining out-of-sample period excluding 2008 Q4.

These results make intuitive sense: The linear regression model cannot fully capture the large spikes in the VIX, like those in 2008 Q4. Thus, one would expect that the residuals—which reflect the portion of the VIX’s movement unexplained by the model—to be large in 2008 Q4 when markets were in massive disarray. Furthermore, the regressors in the estimation model are lagged by a day [day $t - 1$]; as a result, they do not capture the changes in the market happening on a given day, which, of course, is reflected in that day’s VIX [i.e., on day t], the model’s explained variable.

In sum, the comparative analyses of the in-sample and out-of-sample model errors undertaken in this section of paper do not support the hypothesis of manipulation of VIX. If one excludes just 3 months of the 2008 Q4, the errors in the period at issue are actually *smaller* than in 1998–2007. And the larger errors in 2008 Q4 are explained by the unprecedented jump in the VIX precipitated by the financial crisis. Thus, the movements of the VIX throughout the 20-year period analyzed appear to be consistent with normal market forces and do not support the conjecture of artificiality or manipulation of the index.

The analysis of the VIX futures and options expiration days

In the preceding section, while we examined a 10-year period for signs of manipulation, one might argue that the effect of the manipulation was confined to specific days when the VIX futures and options expired. This line of argument is consistent with the Griffin and Shams (2018) study, which found unusual trading activities during VIX futures expiration days and therefore concludes that settlement prices were manipulated.

Testing the statistical significance of the model errors on futures expiration days

There are currently both monthly and weekly VIX futures contracts; monthly futures expire on the same day (typically 3rd Wed of a month). Weekly futures expire on

Wednesday. Monthly futures started trading in May 2004; weekly futures were introduced in August 2015.

In any month, 1 weekly option expiration day coincides with the monthly option and futures expiration day. There are 274 unique expiration days (monthly and weekly combined) in our data set. Of these, 40 are in the pre-2008 period; so, there are 234 expiration days in the period at issue.

To test the significance of the regression errors on the expiration days, we used the same regression model as in (1) estimated using data for the time period 1998–2018 but excluding the 234 expiration days from the estimation sample. We then generated regression residuals for these 234 days (thus, they are out-of-sample residuals for the 234 days) and then tested their statistical significance. Our approach is akin to widely-used event study analysis, where the residual value on a given event day is tested for statistical significance. Details of these tests are contained in the ‘Testing for statistically significant residuals on settlement days’ of ‘Appendix.’

The results are shown in Table 5.

Panel A in Table 5 shows that 11 days or 4.7% of the 234 expiration days at issue are statistically significant.²⁸ By random chance alone, one would expect 5% of the expiration days to be statistically significant, and that is confirmed by the Monte Carlo results shown in Panel B of the table. Consistent with the results in the preceding section of the paper, we also observe, if one excludes 2008 Q4, the proportion of statistically significant expiration days falls under 4%. We have tested the significance of the futures expiration days using both a distribution-free non-parametric approach and also using the standard t test; the results under both approaches are identical.

Testing of evidence of manipulation in settlement prices

So far in this paper we have focused on the daily closing value of the VIX. As noted earlier, the analysis by Griffin and Shams (2018) was focused specifically on the VIX futures’ settlement prices and trading during the auction period during which those prices are determined. In this subsection, we will examine the settlement prices to see if there appears to be any artificiality in VIX futures contracts.

The settlement price for an expiring futures contract is determined using an auction process called the Special Opening Quotation (SOQ). The SOQ takes place on the morning of each expiration day. The CBOE provides the

²⁷ In the nonparametric approach, the statistical significance is based on the empirical distribution of the absolute value of the residuals. We determine a residual data point to be statistically significant if its value is equal to or greater than the 95th percentile value of all residuals.

²⁸ The results are virtually unchanged if the prediction errors are computed using the opening level of the VIX on expiration days or the closing level of the VIX on the day preceding the expiration days.

Table 5 Statistically significant expiration days

	Significant days using VIX closing values		
	Expiration days	Using VIX closing values	
<i>Panel A</i>			
2008–2018	234	11	4.70%
2008 Q4	3	2	
2008–2018–ex 2008Q4	231	9	3.90%
% of days statistically significant			
<i>Panel B</i>			
10,000 random draws of 234 non-expiration days			5.02%

Table 6 Statistically significant expiration days

	Significant days using settlement price		
	Expiration days	# of days	
2008–2018	234	10	4.27%
2008 Q4	3	2	
2008–2018–ex 2008Q4	231	8	3.46%

settlement price for each futures contract back to 2013 on its website. We were able to obtain settlement prices for futures contracts prior to 2013 from an alternate data provider and confirmed the accuracy of data using the periods that overlap, i.e., since 2013.

Here we undertake a variant of the foregoing analysis by substituting the VIX futures settlement prices for the VIX closing values on expiration days, i.e., the settlement days. Thus, under this approach, a regression error on a futures expiration day is the difference between the settlement price (as opposed to the VIX closing value) and the model-predicted VIX value. The results of our analysis are shown in Table 6.

Table 6 shows that when settlement prices are used, the proportion of statistically significant days drops to 4.3% (which is lower than the percentage (approximately 5%) one would expect by chance alone) even when we include 4Q 2008. This result provides compelling evidence that the settlement prices on the VIX futures on the expiration days were not artificial.

In sum, the results of our analysis of the VIX futures expiration days do not support the hypothesis of manipulation of the VIX, even on the specific dates of VIX futures expirations. During the period at issue, 2008 through the present, the number of statistically significant expiration dates is consistent with random chance, regardless of whether we use VIX closing values or settlement prices of VIX futures. Thus, our findings imply that the level of the VIX on futures expiration days is explained by normal market forces.²⁹

Concluding comments

In this paper, we examined the daily level of the VIX index for signs of manipulation, as has been alleged during the period January 2008 to the present. We constructed a model using explanatory variables that are exogenous to the index and found that the results strongly support the movement in the VIX being explained by market fundamentals: the results overwhelmingly do not support a claim of manipulation. We also specifically examined the VIX futures expiration days and found that the VIX closing values as well as the VIX futures settlements prices on these expiration days are consistent normal market forces and do not show evidence of manipulation.

While these findings strongly support the conclusion that the VIX is not manipulated, it is important to note that our findings are not necessarily incongruent with those of Griffin and Shams (2018). As noted earlier, determination of whether any artificiality in the VIX existed for brief periods of time during and after the SOQ would require intraday data and that analysis is beyond the scope of our study. Our findings, however, *do* imply that, notwithstanding Griffin and Shams finding of unusual trading patterns during the SOQ, the effects of these trades do not persist through the close. Our analysis shows that both the closing values of VIX on settlement days and the settlement prices themselves do not show any evidence of

²⁹ More formally, the null hypothesis that the VIX is explained by normal market forces is not rejected.

manipulation. Both are consistent with market fundamentals.

Acknowledgements We are immensely grateful to Heather Roberts for her research assistance for this paper. We would also like to thank Arthur Havenner and Sonya Rauschenbach for comments that greatly improved the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

Summary statistics for two sub-periods

Below we list the mean and standard deviations for the regression variables for the two sub-periods. The two periods have similar statistics, with the exception of the standard deviation of the VIX, which is higher in the out-of-sample period (Table 7).

Table 7 Summary statistics for two sub-periods: 1998–2007 and 2008–2018

Variable	Variable name	Mean		SD	
		In sample	Out of sample	In sample	Out of sample
VIX	Vix	20.691	20.003	6.936	9.874
SP daily return	Spr	0.000	0.000	0.011	0.013
SP 20-day rolling volatility	Spv	0.010	0.010	0.005	0.008
SP 5-day rolling mean	Spm	0.000	0.000	0.005	0.005
Indicator variable for +SP day	sp ₊	0.524	0.543	0.500	0.498
Indicator variable for –SP day	sp _–	0.476	0.457	0.500	0.498
	# of Obs	2514		2600	

Table 8 Dickey–Fuller test results

<i>p</i> value	[A] ‘98–’07	[B] ‘08–’18	[C] ‘98–’18	[D] ‘98–’18 ex 4Q2008
Minimum	1.327E–04	1.811E–04	2.980E–11	2.970E–12
Maximum	0.027	0.029	1.020E–05	4.230E–06

The *p* values are for the test of the null hypothesis shown in footnote 9 of the paper

Dickey–Fuller test for unit root

The results of the Dickey–Fuller tests are reported in Table 8 for multiple time periods. Leybourne et al. (1988) showed that the existence of a structural break can lead to a false rejection of the Dickey–Fuller Test. To confirm that the VIX time series was stationary, we tested the two sub-periods separately, as well as the entire time period (1998–2018) with and without the fourth quarter of 2008 (since once one excludes 4Q 2008 the data do not show structural break). The null hypothesis of a non-stationary process (i.e., existence of a unit root) was rejected in every case, as indicated by the *p* values below.

Bartlett’s test for equal variance of VIX

See Table 9.

Table 9 Test for equality of variance for VIX

Standard deviation of VIX		
In sample	Out of sample	Out-of-sample ex 4Q2008
6.936	9.874	7.672
<i>Bartlett’s test for equal variances</i>		
χ^2 (<i>df</i> = 1)	311.144	25.610
<i>p</i> value	0.000	0.000

Bartlett (1937)

Regression residuals

Theil's U statistics

Theil's U statistics for forecast accuracy was computed as follows:

$$\frac{\sqrt{\frac{1}{T} \sum_{t=1}^T (V_t - \hat{V}_t)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^T (V_t)^2}}$$

where V_t denotes the actual value of the VIX index and $\hat{V}_t$ denotes the predicted value of the index on day t and T denotes the total number of observations. Large values indicate a poor forecasting performance. See Greene (1997), p. 373.

As can be seen below, the out-of-sample forecasting performance of the regression model is better than the in-sample performance (Table 10).

Testing for structural break

For each regressor, x_t^i , in the regression equation given by (1) in the body of the paper, we created two associated regressors by multiplying each by two indicator variables,

Table 10 Theil's U statistics

In sample	Out of sample	
1998–2007	2008–2018	2008–2018–ex2008Q4
0.5103	0.3797	0.4523

D^1 and D^2 , where D^1 takes a value of one if it is in the 1998–2007 period and zero otherwise, and D^2 takes a value of one if it is in the 2008–2018 period and zero otherwise.

Then, we estimated a single equation using the entire dataset (1998–2018):

$$Vix_t = \sum_{i=1}^k \beta^{1i} \cdot D^1 \cdot x_t^i + \sum_{i=1}^k \beta^{2i} \cdot D^2 \cdot x_t^i + \varepsilon_t \quad (2)$$

We then tested the null hypothesis: $H_0 : \beta^{1i} = \beta^{2i}$, $i = 1, \dots, k$, using a t test. Table 3 in the paper contains the p values of these tests. In undertaking

Table 11 Skewness/kurtosis tests for normality

Variable	Obs	p values		
		Skewness	Kurtosis	Joint
Residuals	5114	0.0000	0.0000	0.0000

the tests by excluding the 4Q 2008 (panel B in Table 3), we adopted the same procedure as above, except we dropped the observations from this quarter when estimating Eq. (2).

Estimating the single equation as shown above in (2) and testing the null hypotheses noted above are equivalent to estimating two separate equations, one for each of the two sub-periods (1998–2007 and 2008–2018), and then testing whether the estimated coefficients are the same across the two equations.

Testing for normality of residuals

The Jarque–Bera (Jarque and Bera (1987)) test is to examine whether the skewness and kurtosis of the sample data match that of a normal distribution. The test statistics is defined as follows:

$$JB = \frac{(T - k + 1)}{6} \left(S^2 + \frac{1}{4}(K - 3)^2 \right)$$

where T is the number of observations; k is the number of regressors; S is sample skewness; and K is sample kurtosis. The JB statistic has a *Chi-square* distribution with 2 degrees of freedom.

In Table 11, we report the p values for the individual tests for excess skewness being zero and excess kurtosis (i.e., greater than 3) being zero; the final column labeled ‘Joint’ shows the p value of the JB statistics, noted above. In all three cases, the null hypothesis of normally distributed residuals is rejected.

Testing for statistically significant residuals on settlement days

We estimated a single equation using the entire dataset (1998–2018):

$$Vix_t = \sum_{i=1}^k \beta^i \cdot x_t^i + \sum_{j=1}^L \alpha^j \cdot S^j + \varepsilon_t \quad (3)$$

where x_t^j are the regressors in the regression equation given by (1) in the body of the paper and S^j denotes an indicator variable that takes a value of one on the j th settlement day, zero otherwise. We then tested the statistical significance of the estimated coefficients, $\hat{\alpha}^j$, $j = 1, \dots, L$, using a t test. This is exactly the approach undertaken in event study analysis.

Note, in Eq. (3), the estimated coefficients, $\hat{\alpha}^j$, $j = 1, \dots, L$, are exactly equal to the residuals we would have obtained if we simply estimated the equation $Vix_t = \sum_{i=1}^k \beta^i \cdot x_t^i + \varepsilon_t$ by excluding the L settlement days and then predicted the residuals for those days.

When using settlement prices instead of VIX closing values (the results reported in Table 6 of the paper), we estimated (3) but substituted the settlement prices for the actual VIX values on the settlement days. The estimated coefficients, $\hat{\alpha}^j$, $j = 1, \dots, L$, in this case are equal to the differences (i.e., residuals) between settlement prices and model-predicted VIX values on the settlement days.

References

- Abelson, M., and J. Weisenthal. 2018. An inventor of the VIX: ‘I Don’t Know Why These Products Exist’. Bloomberg, 6 February. <https://www.bloomberg.com/news/articles/2018-02-06/an-inventor-of-the-vix-i-don-t-know-why-these-products-exist>. Accessed 14 June 2018.
- Abrantes-Metz, R.M., G. Rautenberg, and A. Verstein. 2013. Revolution in Manipulation Law: The New CFTC Rules and the Urgent Need for Economic and Empirical Analyses. *University of Pennsylvania Journal of Business Law* 15(2): 357–418.
- Bartlett, M.S. 1937. Properties of Sufficiency and Statistical Tests. *Proceedings of the Royal Society, Series A* 160: 268–282.
- CBOE Exchange. 2018. White Paper: CBOE Volatility Index. CBOE report. <https://www.cboe.com/micro/vix/vixwhite.pdf>. Accessed 14 June 2018.
- Chow, K.V., W. Jiang, and J. Li. 2018. Does VIX Truly Measure Return Volatility? Working Paper, 24 January 2018. <http://www.ssrn.com/abstract-id=2489345>. Accessed 14 June 2018.
- Dickey, D.A., and W.A. Fuller. 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* 74: 427–431.
- Graybill, F.A., A.M. Mood, and D.C. Boes. (1974). *Introduction to the Theory of Statistics*. McGraw Hill.
- Greene, William H. 1997. *Econometric Analysis*. Upper Saddle River: Prentice Hall.
- Griffin, J., and A. Shams. 2018. Manipulation in the VIX? *Review of Financial Studies* 31(4): 1377–1417.
- Hait, D. 2016 VIX: Fear of What? OptionMetrics Report, 13 October. https://optionmetrics.com/wp-content/uploads/2016/10/VIXFearofWhat_WhitePaper.pdf. Accessed 14 June 2018.
- Jarque, C.M., and A.K. Bera. 1987. A Test for Normality of Observations and Regression Residuals. *International Statistical Review* 2: 163–172.
- Leybourne, S.J., T.C. Mills, and P. Newbold. 1988. Spurious Rejections by Dickey–Fuller Tests in the Presence of a Break Under the Null. *Journal of Econometrics* 87(1): 191–203.
- Marthinsen, J.E., and Y. Gai. 2010a. Did Amaranth advisors LLC engage in interday price manipulation in the natural gas futures market? *Journal of Derivatives and Hedge Funds* 15(4): 261–273.
- Marthinsen, J.E., and Y. Gai. 2010b. Did Amaranth’s absolute, relative and extreme positions affect natural gas futures prices, spreads and volatilities? *Journal of Derivatives and Hedge Funds* 16(1): 9–21.

- Ozair, M. 2014. What Does the VIX Actually Measure? An Analysis of the Causation of SPX and VIX. *Journal of Finance and Risk Perspectives* 3(2): 83–132.
- Perdue, W.C. 1987–1988. Manipulation of Futures Markets: Redefining the Offense. *Fordham Law Review* 56(3): 345–402.
- Pirrong, C. 1996. *The Economics, Law, and Public Policy of Market Power Manipulation*. Norwell, MA: Kluwer Academic Publishers.
- Pirrong, C. 2004. Detecting Manipulation in Futures Markets: The Ferruzzi Soybean Episode. *American Law and Economics Review* 6(1): 28–71.
- Saha, A., and H. Petersen. 2012. Detecting Price Artificiality and Manipulation in Futures Markets: An Application to Amaranth. *Journal of Derivatives and Hedge Funds* 18(3): 254–271.
- Vodenska, I. and W. J. Chambers. 2013. Understanding the Relationship Between VIX and the S&P 500 Index Volatility. In *Proceedings of the 26th Australasian Finance and Banking Conference*; 19 December, Sydney, AU. Sydney: UNSW Business School.
- Whaley, R. 1993. Derivatives on Market Volatility: Hedging Tools Long Overdue. *Journal of Derivatives* 1(1): 71–84.
- Zakamulin, V. 2016. Abnormal Stock Market Returns Around Peaks in VIX: The Evidence of Investor Overreaction? Working Paper, 3 May 2016. <http://www.ssrn.com/abstract-id2773134>. Accessed 14 June 2018.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Atanu Saha is a Managing Director of EconOne, a research and economic consulting firm. He specializes in the application of economics and finance to complex business issues, with a focus on data analysis.

Burton G. Malkiel is the Chemical Bank Chairman's Professor of Economics, Emeritus at Princeton University. His books include *A Random Walk Down Wall Street*. He had served as a member of the Council of Economic Advisers, president of the American Finance Association and dean of the Yale School of Management.

Alexander Rinaudo is a Managing Director of EconOne, a research and economic consulting firm. He has over 15 years of experience bringing together academic research, industry experts and legal counsel to develop economic and financial analysis for complex business issues.