Forecasting the VIX in the midst of COVID-19

Rutger Lit[(a,b)*]

[(a)]Time Series Lab

[(b)]Vrije Universiteit Amsterdam, The Netherlands

June 11, 2020

Abstract

We study the behavior of the Volatility Index (VIX) time series in the period leading up to the COVID-19 outbreak. Time-varying location/scale models are used to extract several time-varying components from the VIX time series. The time-varying components are driven by the score of the predictive density. These so called score-driven models have proven to be successful in extracting time-varying components like autoregressive processes and seasonal components. A range of model specifications is used to forecast the VIX in the COVID-19 period that spans the first quarter of 2020. Explanatory variables are used to improve in-sample model fit and out-of-sample forecast accuracy. All model computations are carried out with the *Time Series Lab* software package.

Key words: VIX, COVID-19, Time Series Lab, Time-varying components, Time Series, Forecasting, Score-driven models

1 Introduction

The VIX, also called fear index or fear gauge, is a measure of the overall market sentiment and reflects the investors' risk appetite. The VIX is calculated by the Chicago Board Options Exchange (CBOE) and estimates implied volatility by aggregating the weighted prices of S&P 500 put and call options over a wide range of strike prices. A detailed description of the calculation of the VIX is given by CBOE in their white paper, see CBOE (2019). The VIX is quoted in percentage points and represents the expected annualized change in the S&P 500 index over the next 30-day time period within a one standard deviation confidence interval. VIX today more often than not overstates the level of actual volatility experienced in the next 30 days, see Edwards and Preston (2017a) and Edwards and Preston (2017b). Two explanations have been given, the

*The author is grateful for comments from Paolo Gorgi, Siem Jan Koopman, and Binte Groen. All errors are his own.

first one stems from a behavioral finance perspective. Risk averse investors are willing to pay a premium to hedge against risk. The second one comes from an option pricing perspective. Implied volatility takes into account rare events that occur in the (often non-Gaussian) tails of the distribution, while realized volatility will only include rare events if they have occurred in the period for which the realized volatility is calculated. Since on average rare events will only occur with a small frequency, realized volatility tends to underestimate the potential for large losses most of the time.

Blair et al. (2001) present comparisons of volatility forecasts and show that VIX provides more accurate forecasts than either low- or high-frequency index returns, regardless of the definition of realised volatility and the horizon of the forecasts. Fernandes et al. (2014) study the statistical behavior of the VIX time series use a HAR model with additional explanatory variables so as to account for the relationships between the VIX index and financial and macroeconomic variables. Koopman et al. (2005) explore the forecasting value of implied volatility and use the VIX as input for volatility forecasting models. Psaradellis and Sermpinis (2016) use a neural network approach to forecast the VIX.

In this paper we present our study into the behavior of the Volatility Index (VIX) time series and forecast daily closing levels of the VIX for the first quarter of 2020. This time period was chosen deliberately because in March 2020 the VIX fluctuated as never seen before due to the COVID-19 outbreak. We focused on time-varying location/scale models with the aim of identifying the model with the most accurate forecast of the VIX in turbulent times. With accurate forecasts of implied volatility, market participants could potentially form profitable option trading strategies, something which can have implications about the efficiency of the option markets as well. The VIX data and additional explanatory variables of this research have a daily frequency, which means that we calculate forecasts for day $t+1$ with information up to and including time t. The author of this paper is fully aware that in today's high-frequency markets and 24h economy, forecasting one day ahead for a liquid product is a long forecast horizon and by the time the forecast period is reached, new information has already become available that could potentially improve forecasts. However, if statistically significant conclusions can be drawn from forecasts one day ahead, it can assist us in decision making processes at the current day.

We modelled the VIX time series with a wide range of probability distributions, dynamic components, and explanatory variables and compared these models on in-sample model fit and out-of-sample forecast accuracy. All model calculations in this paper were carried out with the software package Time Series Lab - Score Edition which is available from `https://timeserieslab.com` for free and can therefore be easily replicated. Screen captures of the software are presented in the online supplement to highlight the chosen settings in each step of the modelling process.

The remainder of this paper is organized as follows. Section 2 discusses the score-driven modelling framework. Section 3 presents a description of the data, and Section 4 discusses modelling results. Section 5 concludes.

2 The model

The VIX time series is assumed to have observations coming from a probability distribution $p(y_t|\mu_t, \sigma_t)$ with time-varying location μ_t and scale σ_t for $t = 1, \dots, T$ where T is the length of the time series. The mechanism to update the location and scale over time is the scaled score of the likelihood function. These so called score-driven models, or generalized autoregressive score models, were proposed in their full generality by Creal et al. (2013) and for time-varying location/scale volatility models by Harvey (2013). The score-driven approach provides a unified and consistent framework for introducing time-varying parameters in a wide class of nonlinear models. Score-driven models encompass several well-known models like the GARCH model of Engle (1982) and the ACD model of Engle and Russell (1998).

Let the time-varying parameter vector α_t be updated over time using the following updating function

$$\alpha_{t+1} = \omega + \sum_{i=1}^{q} A_i s_{t-i+1} + \sum_{j=1}^{p} B_j \alpha_{t-j+1},$$

where ω is a vector of constants, A and B are fixed coefficient matrices and s_t is the scaled score function which is the driving force behind the updating equation. The unknown coefficients ω, A and B depend on the static parameter vector ψ which is estimated by the method of Maximum Likelihood. The definition of s_t is

$$s_t = S_t \cdot \nabla_t, \qquad \nabla_t = \frac{\partial \, \log \, p(y_t|\alpha_t, \mathcal{F}_{t-1}; \psi)}{\partial \alpha_t}, \tag{1}$$

for $t = 1, \dots, T$, and where ∇_t is the score vector of the density $p(y_t|\alpha_t, \mathcal{F}_{t-1}; \psi)$. In many situations, it is natural to consider a form of scaling that depends on the variance of the score and therefore the inverse Fisher information is often taken as score scaling. The information set $\mathcal{F}_{t-1}$ usually consists of lagged variables of α_t and y_t but can contain exogenous variables as well. The connection between the distribution parameters μ_t and σ_t is made by the link function $f(\cdot)$ in the following way

$$\mu_t = f(Z_t^{\mu} \alpha_t), \qquad \sigma_t = f(Z_t^{\sigma} \alpha_t),$$

where at time t, the selection matrices Z_t^{μ} and Z_t^{σ} select components from the time-varying parameter vector α_t. A typical example of a link function is the exponential link function to ensure positivity. Another convenient feature of the score-driven framework is the ease in which missing values can be taken into account. If a missing values is encountered at time t, the corresponding scaled score is set to zero and the filter can continue naturally without score updating at time t.

Several time-varying components can be included in α_t. Among them are non-stationary components like Trend and Seasonal but also stationary components like Autoregressive processes of order p. Furthermore, the score-driven framework easily takes explanatory variables into account as well, something which we will use for the modelling of the VIX time series. Finally, several probability distribution $p(y_t|\mu_t, \sigma_t)$ can be applied to the data and the resulting new models can be compared for their model fit and forecast accuracy. Applying a different distribution is in general a routine task as only the log likelihood, which is available in closed-form, and the scaled score need to be re-calculated.

3 Data description

The daily VIX time series was downloaded from Yahoo finance (`https://finance.yahoo.com`) and descriptive statistics are presented in Table 1. Note that the downloaded raw data has weekends and banking holidays removed so that no missing values are present in the data. To be able to model a possible weekly, or biweekly pattern we added missing values for the banking holidays so that each 5 consecutive data points constitutes exactly 1 week. The VIX time series is presented in Figure 1 for the full sample and the first quarter of 2020. Our approach is to estimate model parameters from period 1 (05-01-2004 to 31-12-2019) and use these to forecast period 2 (01-01-2020 to 31-03-2020). This poses several challenges since both periods differ strongly when it comes to their statistical characteristics. For example, mean and standard deviation are much higher in period 2. Also, the data is skewed which is somewhat alleviated after taking logs from the series but not completely. Furthermore, kurtosis switches from excess kurtosis (> 3) in period 1 to leptokurtic (< 3) in period 2. Most noteworthy is the strong rejection of a unit root in period 1 and the strong acceptance of a unit root in period 2.

The following lag-1 explanatory variables were used in an attempt to increase forecast accuracy: the κ-day S&P 500 continuously compounded positive returns and negative returns for $\kappa = 1, 5, 10,$ 20. The κ-day continuously compounded return of WTI Crude oil price. The S&P 500 volume first difference of logs. The first difference of the logarithm of the foreign exchange value of the US dollar against the Euro, British pound, Japanese yen, and Bitcoin. The spread between 10-Year and 3-month Treasury Constant Maturities (credit spread). The difference between the three-month Treasury bill and the three-month LIBOR based in US dollars (TED spread) and dummy variables for the beginning and end of the trading month.

Explanatory variables were downloaded from Yahoo finance and the Federal Reserve Economic Data base (`https://fred.stlouisfed.org`).

4 Results

The ACF of the VIX time series shows a slowly decaying pattern while the PACF shows 2 significant lags. We therefore started our analysis with a benchmark model consisting of a time-varying location that consists of an AR2 process and a constant scale parameter. The most extensive model has a location parameter that is decomposed into an AR2 process, an AR1 process, a weekday seasonal specification, explanatory variables, and a scale parameter consisting of an AR1 process. All components, except the explanatory variables component, are driven by the score as described in Section (2). In-sample model fit and out-of-sample forecast accuracy is reported and compared among other models including simple benchmark models in Table 2. The losses were evaluated by the Diebold and Mariano (1995) statistic to test for equal predictive accuracy. A higher (in-sample) log likelihood value results in almost all cases in lower in-sample loss, however, this is not true for out-of-sample forecast accuracy. The more extensive and better fitting in-sample models are often not better at forecasting as indicated by the higher out-of-sample losses. This can be explained by overfitting of the time series, a phenomenon where simple time series models outperform more extensive models when it comes to forecasting. We discuss estimation results per category.

Distribution

Looking at point forecasts (RMSE, MAE, MAPE) in Table 2, the Gaussian distribution outperforms the Student t and Generalized Error Distribution (GED) in several cases and is never significantly worse.

This strong performance on point forecasting is in contrast to the performance when the full distribution is taken into account, e.g. better tail fitting behavior. The in-sample log likelihood value is much higher for the Student t distribution, a clear sign of fatter tails compared to the Gaussian distribution. Parameter estimates of the best in-sample performing model are presented in Table 3 from which we see that the degrees of freedom parameter is estimated at 7.40 indicating fatter tails than the Gaussian distribution. The out-of-sample log loss function, which is defined as minus the log of the predictive density, is also lower for the Student t models.

The better performance of the Gaussian distribution on point forecasting might be surprising at first but are actually in line with theory, since the Gaussian distribution minimizes the squared errors. It is for the practitioner to decide which is valued higher, point forecasts or density forecasts.

Autoregressive components

We included several time-varying components in the model. The extracted components for location and scale are presented in Figure 2 and 3. We note that the signal for location is decomposed into a persistent AR2 processes ($\phi_1 + \phi_2 = 0.996$) and an AR1 process that takes the short(er) run shocks into account albeit relatively persistent as well with $\phi = 0.863$. The sum of two autoregressive components can be seen as a long memory process, see (Harvey, 2013, p91) for long memory in location/scale models and Fernandes et al. (2014) for long memory in VIX modelling.

Seasonal

A small but significant weekday seasonal is extracted from the VIX time series. Table 2 shows the parameters of the seasonal components. First, the seasonal is not time-varying since the updating parameter (seasonal κ) was estimated close to zero and subsequently fixed at zero during re-estimation. Second, there is a weekday effect in which the VIX is on average slightly higher on Monday and Tuesday and lower on Friday. The parameters for Wednesday and Thursday are not significantly different from zero. Note that the seasonal components must sum to zero for identification and therefore the Friday parameter can be deduced from the parameters belonging to the rest of the week. We refer to Figure 2 for a plot of the weekday pattern.

Explanatory variables

After optimizing the model with the full set of explanatory variables, corresponding standard errors were calculated. The least significant explanatory variable was removed and the process of estimating and calculating standard errors was repeated until a significant set of explanatory variables remained. The final model consists of the explanatory variables S&P negative return lag-1, WTI Close return lag-1 (oil price), and a dummy for the beginning and end of the trading month. We refer to Table 2 for the parameter estimates and standard errors of the explanatory variables. The negative sign in front of the coefficients corresponding to S&P500 negative returns lag1 and WTI closing price lag1 show an inverse relationship with the VIX.

Model extensions

The inclusion of higher lags of the autoregressive processes for both location $(p > 2)$ and scale $(p > 1)$ did not lead to improvements. Furthermore, a seasonal pattern could not be extracted from the scale parameter. Finally, multiple lags of the score were included in the model as well without improvement on model fit or forecast-accuracy.

The effect of a constant scale parameter

The poor out-of-sample forecasting results of the Student t model with **constant** scale is striking. We investigated this further and provide Figure 4 and 5 to illustrate matters. We see from Figure 4 that the mean of the Student t distribution does not adapt quickly enough to keep up with the sharp increase in the VIX level due to the COVID-19 outbreak. The score function of the Gaussian and Student t distribution provide more insight. With the definition of the scaled score given in (1) and the inverse Fisher information taken as the score scaling we have for Gaussian $p(y_t|\mu_t, \sigma_t)$ with time-varying location $\mu_t = \exp(\alpha_t)$ and constant scale σ

$$s_t = \frac{y_t - \mu_t}{\mu_t}, \tag{2}$$

and for Student t

$$s_t = \frac{\nu + 3}{\nu \mu_t} \cdot \frac{y_t - \mu_t}{1 + \frac{(y_t - \mu_t)^2}{\nu \sigma^2}}. \tag{3}$$

The constant scale σ enters in the score of the Student t distribution but not in the score of the Gaussian distribution. If the scale is allowed to be time-varying, we see that σ_t will increase in turbulent times and hence the Student t score with time-varying scale will be larger due to the way σ enters in (3). The forecast results for the Student t model with time-varying scale is given in Figure 5. We see that it tracks the VIX level much better, something which is confirmed by the much higher likelihood as well (Table 3).

5 Conclusion

We used the time series package Time Series Lab - Score Edition to extract signal from the VIX time series. We modelled the time-varying location and scale of the model for several probability distributions and combinations of time-varying components. Especially the time-varying location parameter was a rich source of information since two autoregressive processes and a weekday seasonal pattern was extracted.

The Gaussian distribution outperformed the Student t and Generalized Error Distribution (GED) when it comes to point forecasts (RMSE, MAE, MAPE). This strong performance on point forecasting is in contrast to the performance when the full distribution is taken into account. These results are in line with theory since the Gaussian distribution minimizes the squared errors. We found long memory in the VIX time series due to the 2 persistent AR process that were extracted. A small but significant weekday seasonal is extracted from the VIX time series as well. There is a weekday effect in which the VIX is on average slightly higher on Monday and Tuesday and lower on Friday. The final model consists of the explanatory variables S&P negative return lag-1, WTI Close return lag-1, and a dummy for the beginning and end of the trading month. The negative sign in front of the coefficients corresponding to S&P500 negative returns lag1 and WTI closing price lag1 show an inverse relationship with the VIX.

References

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

CBOE (2019). Vix white paper cboe volatility index®. http://www.cboe.com/micro/vix/vixwhite.pdf.

Creal, D. D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28(5)*, 777–795.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*, 253–265.

Edwards, T. and H. Preston (2017a). A practitioner's guide to reading vix®.

Edwards, T. and H. Preston (2017b). Reading vix®: Does vix predict future volatility?

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 987–1007.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 1127–1162.

Fernandes, M., M. C. Medeiros, and M. Scharth (2014). Modeling and predicting the cboe market volatility index. *Journal of Banking & Finance 40*, 1–10.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*, Volume 52. Cambridge: Cambridge University Press.

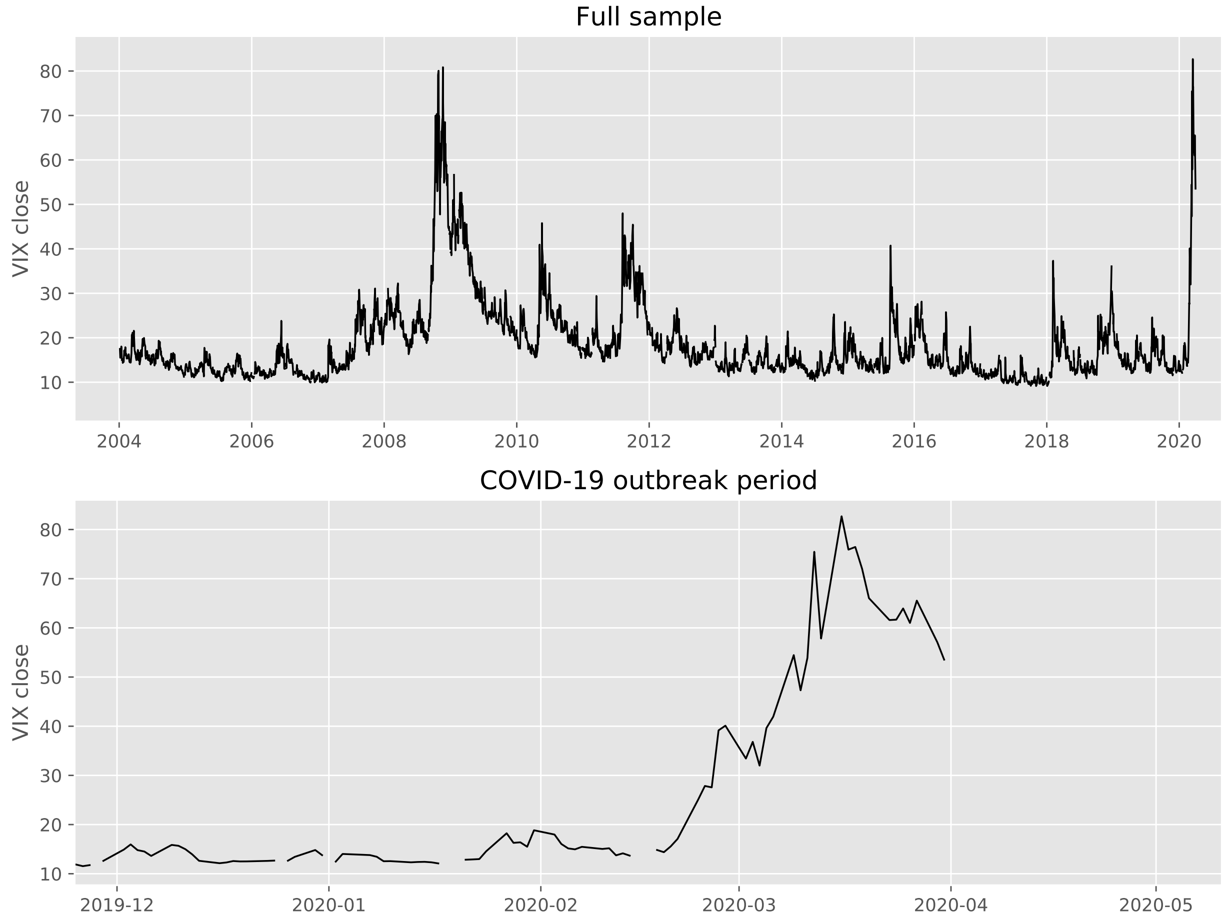
Koopman, S. J., B. Jungbacker, and E. Hol (2005). Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance 12*(3), 445–475.

Lit, R., S. J. Koopman, and A. C. Harvey (2020). Time Series Lab - Score Edition. https://timeserieslab.com.

Psaradellis, I. and G. Sermpinis (2016). Modelling and trading the us implied volatility indices. evidence from the vix, vxn and vxd indices. *International Journal of Forecasting 32*(4), 1268–1283.

Tables and Figures

Figure 1
VIX closing levels



The figure shows closing levels of the VIX. Top panel: full sample. Bottom panel: COVID-19 outbreak period. The top panel clearly shows the sharp rise of the VIX in a very short time period at the end of the sample. Equal levels were reached during the financial crisis however the period leading up to the financial crisis was more turbulent than the period leading up to the COVID-19 jump. The bottom panel shows that the VIX calmly started the beginning of 2020 and than abruptly rose.

Table 1

Descriptive statistics of the VIX time series

The table reports descriptive statistics of the VIX time series for the closing price and log closing price. Period 1 ranges from 05-01-2004 to 31-12-2019. Period 2 spans the first quarter of 2020 in which the COVID-19 outbreak took place (01-01-2020 to 31-03-2020). The p-values of the Jarque-Bera test for normality and the Augmented Dickey-Fuller (ADF) test for unit root are reported. The number of lags in the ADF test are selected using the Bayesian information criterion.

	Closing price			Log closing price		
Characteristic	Period 1	Period 2	Full sample	Period 1	Period 2	Full sample
Mean	18.22	31.22	18.42	2.83	3.21	2.83
Standard dev.	8.58	22.05	9.08	0.36	0.67	0.37
Median	15.60	16.73	15.61	2.75	2.82	2.75
Minimum	9.14	12.10	9.14	2.21	2.49	2.21
Maximum	80.86	82.69	82.69	4.39	4.42	4.42
Kurtosis	13.41	2.27	13.74	4.57	1.59	4.73
Skewness	2.75	0.88	2.84	1.17	0.50	1.23
T	4172	66	4238	4172	66	4238
Missings	146	4	150	146	4	150
Jarque-Bera	0.0000	0.0068	0.0000	0.0000	0.0164	0.0000
ADF	0.0002	0.9502	0.0019	0.0000	0.9516	0.0000

Table 2

In-sample and out-of-sample results for a range of model specifications

The table reports in-sample model fit and out-of-sample forecast results for a wide range of model specifications. All models are estimated with Time Series Lab - Score Edition. All dynamic components are driven by the score. The lowest loss per category is given a blue shade. The Diebold and Mariano (1995) statistic is used to assess forecast accuracy among competing models. The second Student t model with Location: exp(AR2) and Scale: exp(AR1) is set as benchmark model. All out-of-sample losses with a dagger sign † next to them perform significantly worse compared to the benchmark model.

	In-sample				Out-of-sample			
Model description	Log L	RMSE	MAE	MAPE	RMSE	MAE	MAPE	Log loss
Distribution: Gaussian								
Location: exp(AR2) Scale: constant	-7895.50	1.72	1.02	5.18	5.91	3.37	8.64	7.36†
Location: exp(AR2) Scale: exp(AR1)	-6687.61	1.74	1.01	5.13	6.05	3.40	8.63	2.91
Location: exp(AR2 + AR1) Scale: exp(AR1)	-6649.01	1.74	1.00	5.04	6.49	3.44	8.55	2.94
Location: exp(AR2 + AR1 + Seas(5)) Scale: exp(AR1)	-6589.67	1.74	0.99	4.97	6.47	3.53	8.80	2.91
Location: exp(AR2+AR1+Seas(5)+$X\beta$) Scale: exp(AR1)	-6536.45	1.69	0.98	4.93	6.27	3.68	9.09	2.90
Distribution: Student t								
Location: exp(AR2) Scale: constant	-7293.05	2.05	1.11	5.49	18.14†	11.09†	21.89†	7.15†
Location: exp(AR2) Scale: exp(AR1)	-6306.96	1.80	1.04	5.19	5.96	3.33	8.86	2.61
Location: exp(AR2 + AR1) Scale: exp(AR1)	-6270.03	1.83	1.03	5.13	6.70†	3.64	9.34	2.65
Location: exp(AR2 + AR1 + Seas(5)) Scale: exp(AR1)	-6209.58	1.83	1.03	5.06	6.64†	3.68	9.49	2.66
Location: exp(AR2+AR1+Seas(5)+$X\beta$) Scale: exp(AR1)	-6162.03	1.80	1.01	5.00	7.16†	4.09†	10.08†	2.69
Distribution: GED								
Location: exp(AR2) Scale: constant	-7485.44	1.82	1.08	5.50	7.11†	3.99†	10.20†	6.33†
Location: exp(AR2) Scale: exp(AR1)	-6464.19	1.81	1.04	5.24	6.00	3.26	8.60	2.71
Location: exp(AR2 + AR1) Scale: exp(AR1)	-6428.11	1.85	1.04	5.19	6.57	3.43	8.83	2.74
Location: exp(AR2 + AR1 + Seas(5)) Scale: exp(AR1)	-6367.55	1.84	1.03	5.11	6.54	3.42	8.83	2.74
Location: exp(AR2+AR1+Seas(5)+$X\beta$) Scale: exp(AR1)	-6317.81	1.78	1.01	5.04	7.11†	3.95†	9.64	2.78

Table 3
Parameter estimates for the Student t model

The table reports parameter estimates obtained from maximizing the likelihood function of the most extensive Student t model with specification Location: exp(AR2+AR1+Seas(5)+$X\beta$) and Scale: exp(AR1). P-values should be taken with caution. They are directly interpretable for regression coefficients but, due to boundary issues, can give misleading results for parameters corresponding to dynamic components or variances.

Parameter type	Value	Sig.Lvl	Asymp.SE	p-value	-1.96 SE	+1.96 SE
Log location						
AR2 ω	0.0091	***	0.0015	0.0000	0.0062	0.0120
AR2 κ	0.3279	***	0.0422	0.0000	0.2451	0.4107
AR2 ϕ_1	1.0072	***	0.0003	0.0000	1.0066	1.0078
AR2 ϕ_2	-0.0113		0.0692	0.8703	-0.1470	0.1244
init	2.8509	***	0.0525	0.0000	2.7480	2.9538
AR1 (2nd) κ	0.4898	***	0.0434	0.0000	0.4048	0.5748
AR1 (2nd) ϕ	0.8628	***	0.0316	0.0000	0.8009	0.9247
seasonal κ	0.0000					
init seasonal 1	0.0075	***	0.0010	0.0000	0.0056	0.0094
init seasonal 2	0.0050	***	0.0011	0.0000	0.0029	0.0071
init seasonal 3	0.0000					
init seasonal 4	0.0000					
init seasonal 5	-0.0125					
β dummy em	0.0184	***	0.0034	0.0000	0.0116	0.0251
β dummy bm	0.0084	*	0.0035	0.0158	0.0016	0.0153
β S&P neg return la..	-0.3428	*	0.1571	0.0292	-0.6508	-0.0348
β WTI Close lag1	-0.2426	***	0.0314	0.0000	-0.3041	-0.1812
Log scale						
AR1 ω	0.0052		0.0036	0.1483	-0.0019	0.0123
AR1 κ	0.1536	***	0.0107	0.0000	0.1327	0.1746
AR1 ϕ	0.9259	***	0.0091	0.0000	0.9081	0.9437
Additional						
Degrees of freedom	7.4019	***	0.7310	0.0000	5.9691	8.8347

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Figure 2

Decomposition of VIX level into AR2, AR1, and seasonal component

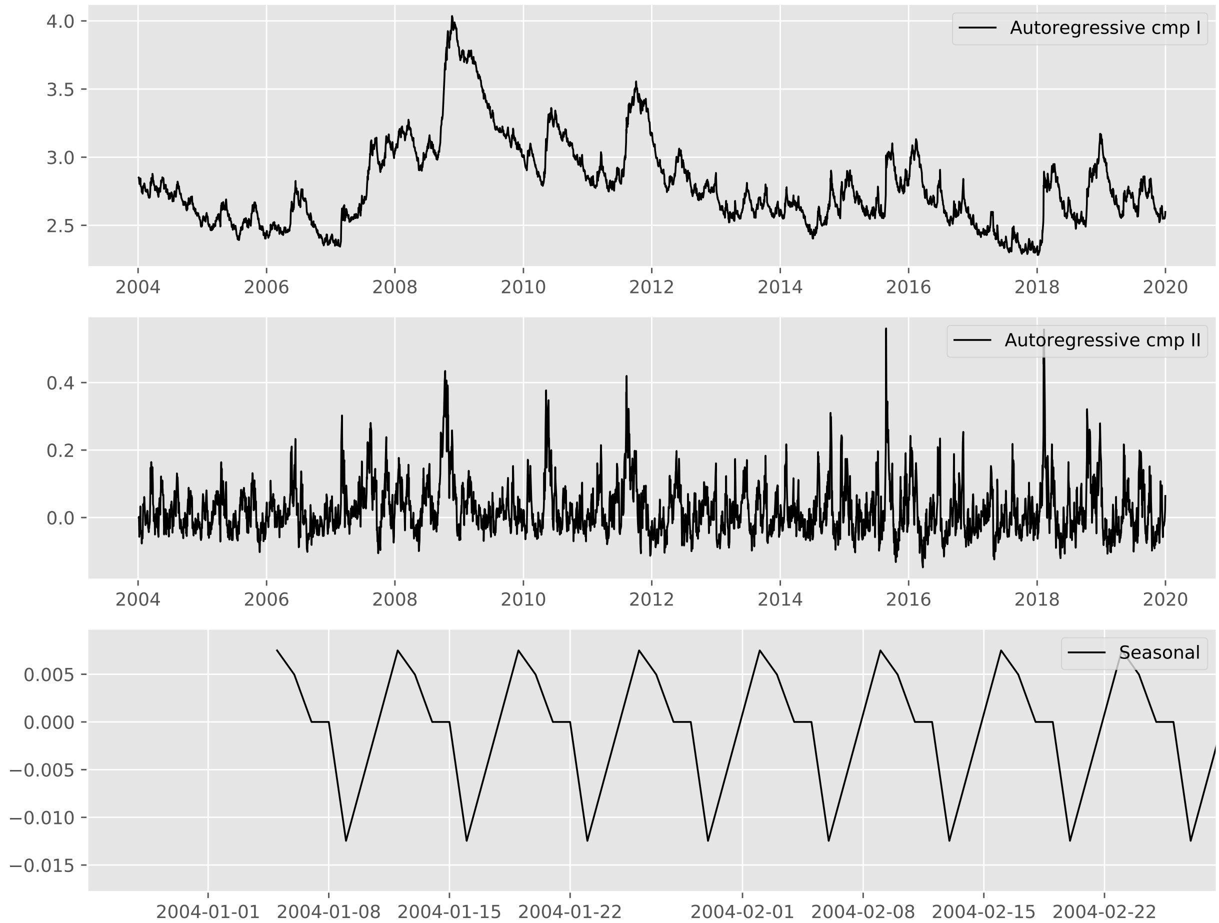


Figure 3

Time-varying standard deviation of VIX time series

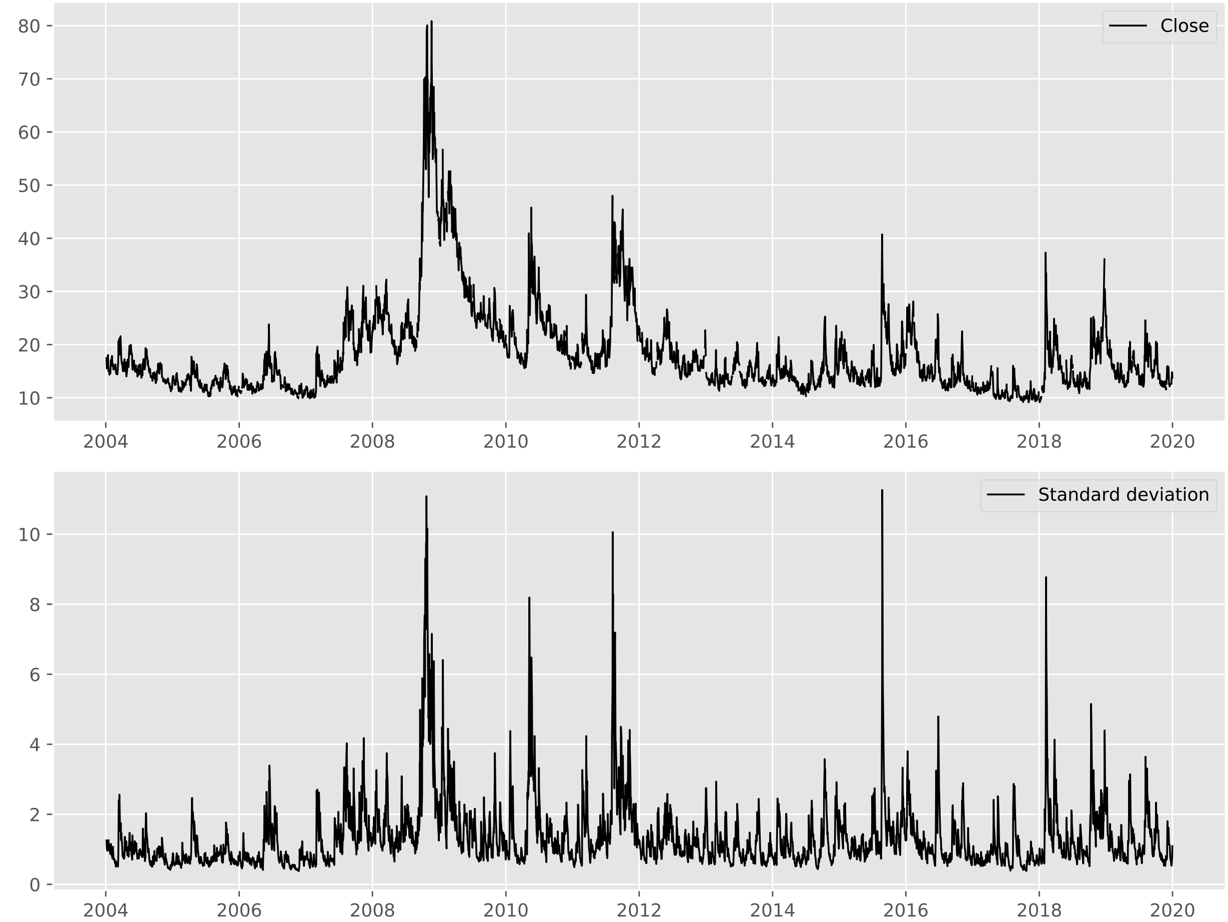


Figure 4

VIX forecast with time-varying mean and constant scale

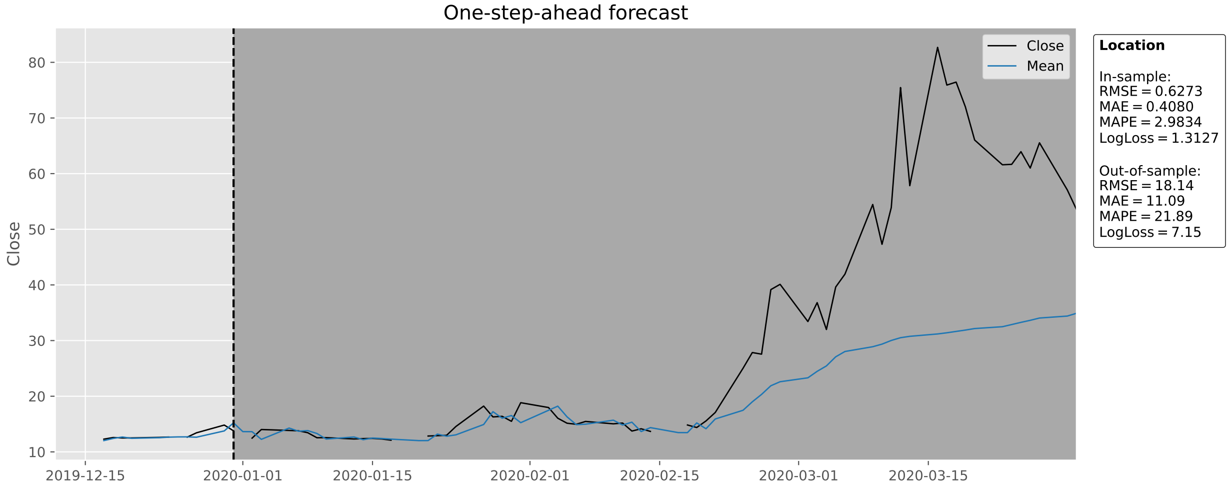


Figure 5

VIX forecast with time-varying mean and time-varying scale