PORTFOLIO
MANAGEMENT
RESEARCH
By With Intelligence

the journal of PORTFOLIO management

volume 49 number 3

FEBRUARY 2023

jpm.pm-research.com

Forecasting Stock Market Volatility

Michael Stamos

Allianz Global Investors is a leading active asset manager, managing EUR 521 billion in assets for individuals, families and institutions worldwide. By being active and investing for the long term, our goal is to elevate the investment experience for our clients and generate value every step of the way.

Active is: Allianz Global Investors

Data as of 30 September 2022

Michael Stamos

Dr. Michael Stamos, CFA, is leading the AllianzGI Multi Asset team's development of systematic investment strategies focusing on risk management, asset allocation, risk premia investing, portfolio construction. Michael specialized in the management of Dynamic Allocation, Managed Futures, Risk Premia, and Risk Parity Funds. Michael has 20 years of R&D experience and 15 years of portfolio management experience. Michael received a PhD (summa cum laude) in Finance and holds the Chartered Financial Analyst designation. Prior to joining the firm in 2007, Michael was a Research Assistant at Frankfurt University focusing on optimizing asset allocation and retirement solutions. He has published his work in top-tier journals such Review of Financial Studies, Journal of Banking and Finance, Journal of Portfolio Management, Journal of Risk and Insurance, Journal of Economic Dynamics and Control, Insurance Mathematics and Economics, and Journal of Pension Finance and Economics, and presented his work at many conferences.

Forecasting Stock Market Volatility

Michael Stamos

Michael Stamos
is the head of Global R&D Multi Asset at Allianz Global Investors in Frankfurt, Germany.
michael.stamos@allianzgi.com

KEY FINDINGS

- The task of volatility forecasting can be simplified when avoiding modeling overhead and by treating it identically to other forecasting problems.
- More complex models like autoregressive conditional heteroskedasticity (ARCH) and generalized ARCH (GARCH) do not improve forecasting accuracy for the US stock market when benchmarked against simple estimators. The use of option implied volatility data only marginally improves forecasts.
- Negative return-based and intraday data-based models provide significant added value in forecasting accuracy compared to benchmark models.

ABSTRACT

Volatility as a measure of investment risk is widely accepted by academic researchers and industry professionals and has become ubiquitous in investment analysis. Furthermore, it is among the few financial variables that exhibit predictable time variation. Hence, there is an extensive amount of literature describing volatility models and assessing their forecasting power. This article provides a discussion of the prominent models and compares them in a unified notation framework. The empirical analysis shows that it is hard to outperform even simple trailing variance-type models. Autoregressive conditional heteroskedasticity (ARCH), generalized ARCH (GARCH), implied volatility, asymmetric, and seasonal models hardly improve forecasts despite added complexity. In this study, only momentum-based and intraday data-based models improved predictive accuracy significantly.

Volatility may be—behind holding-period return—the most used quantitative measure in financial analysis. The measurement and prediction of volatility is done countless times daily within financial and insurance industries. Therefore, many academic and financial industry researchers come across numerous methods and models that determine and predict volatility.

It is fascinating to see the different approaches of modeling volatility; some use prespecified estimators, for instance, realized trailing variance or exponentially weighted variance,[1] and then there are more rigorous models such as autoregressive conditional heteroskedasticity (ARCH) as introduced by Engle (1982) and generalized ARCH (GARCH) as proposed by Bollerslev (1986), with all existing variants. Furthermore, there are models rooting from the financial engineering fields, such as stochastic volatility models. When reading different volatility studies, there seems to be hardly common ground with respect to modeling approaches and notation. In fact, most well-known volatility models use vastly different notation and are applied

[1] A good reference is Zangari (1994) describing the methodology used in RiskMetrics™.

in different model worlds, making it difficult to comprehend what exactly the theoretical differences and comparative model benefits are, for instance, of simple trailing volatility, a GARCH model, an implied volatility model, and an exponential variance model. Further comparisons are often done only pairwise and for different types of data and time frame.

In what follows, this article provides an overview of prominent volatility models using a common notation to make models directly comparable. Using a common notational framework allows identifying exactly where models contrast and where they overlap and intersect. This is done by embedding those volatility models into the notation of standard time-series models, marrying econometrical and financial notation. Then, an empirical analysis compares the forecast accuracy of the well-known and additional less conventional models for the common dataset that is US equities while controlling for complexity by restricting the number of tunable parameters. The study's time frame is January 2000 to December 2020 using daily data. The additional use of intraday data to predict daily volatility does not permit going further back in history. On the other hand, the study covers the three major stock market crises, the dot-com bubble, the Global Financial Crisis, and the 2020 COVID crash, which all represent very different types of market corrections.

THE VOLATILITY FORECASTING PROBLEM

To allow formulaic comparisons between different volatility models, the common notational ground must be defined first. This section provides a model-free specification of the volatility forecasting problem. Let us start with the definition of time-varying volatility of asset returns r_t available for the periods $t = 1, \ldots, T$

$$\sigma(t) = \sqrt{VAR[r_t]}$$

with VAR denoting variance, and variance defined as

$$VAR[r_t] = E[r_t^2] - E[r_t]^2$$

To simplify notation, set $E[r_t]$ = c, and $c = 0$ without loss of generality. For all practicality of forecasting financial time series, $E[r_t]^2$ can often be assumed to be negligible because $E[r_t]$ is close to zero and the square of it even more so. Now, we define the variable y_t, which is the squared return of an asset $y_t = r_t^2$. Hence, we get

$$VAR[r_t] = E[y_t]$$

This takes us back to standard methodology to derive estimated forecasts for y_t that get the label $\hat{E}[y_t]$ and to use the standard ways to assess forecasting accuracy. Forecasting variance is the same as forecasting squared returns. The advantage of looking at the forecasting problem in this way is the parsimonious nature and that volatility forecasting is treated as any other forecasting problem. Forecasted volatility is then defined by $\hat{\sigma}(t) = \sqrt{\hat{E}[y_t]}$.

For calibration, the standard estimation framework can be used. For instance, using ordinary least squares, the preceding equations can be fitted by minimizing the mean squared error in

$$MSQE = 1 / T \sum_{t=1}^{T} (y_t - \hat{E}[y_t])^2$$

TIME-SERIES MODELS

Having defined the model-free volatility forecasting problem as a straightforward time-series forecasting problem allows for plugging in standard time-series models to create the forecasts and to evaluate their predictive accuracy. Standard time-series models are defined as follows:

1. White noise: $y_t = a + \varepsilon_t$
2. Auto-regressive AR(p) model: $y_t = a + \Sigma_{i=1}^{p} b_i y_{t-i} + \varepsilon_t$
3. Moving-average MA(q) model: $y_t = a + \Sigma_{i=1}^{q} c_i \varepsilon_{t-i} + \varepsilon_t$
4. Auto-regressive moving average ARMA(p,q) model: $y_t = a + \Sigma_{i=1}^{p} b_i y_{t-i} + \Sigma_{i=1}^{q} c_i \varepsilon_{t-i} + \varepsilon_t$
5. ARMAX(p,q,m) model to include *m* exogenous predictor variables $x_{i,t}$:

$$y_t = a + \sum_{i=1}^{p} b_i\, y_{t-i} + \sum_{i=1}^{q} c_i\, \varepsilon_{t-i} + \sum_{i=1}^{m} d_i\, x_{i,t-1} + \varepsilon_t$$

6. Any other function: $y_t = f(x, y) + \varepsilon_t$
   With errors $\varepsilon_t = y_t - E[y_t]$, $E[\varepsilon_t] = 0$

Modeling-wise that is sufficient to empirically forecast volatility in a parsimonious fashion avoiding large purely financial modeling overhead. The quality of the forecast will mostly depend on the predictive ability of the predictor variables rather than the complexity of the model.

EMBEDDING PROMINENT VOLATILITY MODELS INTO STANDARD TIME-SERIES MODELS

Interestingly, prominent volatility models can be easily reconciled with the unifying notation shown earlier, achieving a better formulaic understanding of their interdependencies. Using $\varepsilon_t = y_t - E[y_t]$ and $\sigma_t^2 = E[y_t]$ gives the following relationships:

1. *Constant variance* $\sigma_t^2 = \sigma^2$ is a white-noise model with $\sigma^2 = a$.
2. *Realized variance* $\sigma_t^2 = 1/N \times \Sigma_{i=1}^{n}\ r_{t-i}^2$, with N being the number of trailing observations, is a restricted form of an AR(p) model with $a = 0,\ b_i = \dfrac{1}{p}, i = 1, \ldots, p$.
3. *Exponential variance* $\sigma_t^2 = \lambda\ \sigma_{t-1}^2 + (1-\lambda) r_{t-1}^2$ with weighting factor $0 \le \lambda \le 1$ is an ARMA(p,q) case with $a = 0, p = 1, b = 1, q = 1,\ c = -\lambda$.
4. *ARCH(m)* model $\sigma_t^2 = \omega + \Sigma_{i=1}^{m} \alpha_i\, r_{t-i}^2$ is an AR(p) model with $m = p,\ \omega = a,\ \alpha_i = b_i$.
5. *GARCH(m,n)* $\sigma_t^2 = \omega + \Sigma_{i=1}^{m} \alpha_i\ r_{t-i}^2 + \Sigma_{i=1}^{n} \beta_i\, \sigma_{t-i}^2$ is an ARMA(p,q) model with $m = p,\ n = q, \omega = a, \alpha_i = b_i, \beta_i = c_i$.

Volatility forecast models that were hardly comparable before can now be better contrasted by aligning their notation using standard time-series models framing. All of the presented volatility models are in fact versions of a standard time-series model. The author believes that the connection between volatility models and standard time-series models has not been clarified before succinctly and hopes this article can facilitate the understanding of many students, academics, and industry professionals. The intent is that the simplified volatility forecasting framework allows us to seamlessly extend the parsimonious forecasting models aiming to improve forecasts, as demonstrated in the following analysis.

EMPIRICAL ANALYSIS TO FORECAST THE VOLATILITY OF US EQUITIES

Using the preceding definitions, we run a case study as a proof of concept. We use daily returns of the S&P 500 Index realized during the years 2000 to 2020. The dependent variable of the forecast problem is defined as the daily squared return as described in "The Volatility Forecasting Problem."[2] We run the estimations for the models described in "Embedding Prominent Volatility Models into Standard Time-Series Models" with the idea to serve as benchmarks:

1. White noise/Constant variance: $y_t = a + \varepsilon_t$
2. Restricted AR(p)/Realized variance: $y_t = a + \frac{1}{p}\Sigma_{i=1}^{p} b_i\ y_{t-i} + \varepsilon_t$
3. Restricted ARMA(1,1)/Exponential variance: $y_t = y_{t-1} + c\ \varepsilon_{t-1} + \varepsilon_t$
4. AR(p)/ARCH: $y_t = a + \Sigma_{i=1}^{p} b_i\ y_{t-i} + \varepsilon_t$
5. ARMA(p,q)/GARCH: $y_t = a + \Sigma_{i=1}^{p} b_i\ \ y_{t-i} + \Sigma_{i=1}^{q} c_i\ \ \varepsilon_{t-i} + \varepsilon_t$
   Furthermore, we add five models to have parsimonious cases of the general model class $y_t = f(x, y) + \varepsilon_t$, with x being exogenous predictive variables
6. *Asymmetric exponential variance*: $y_t = \begin{cases} y_{t-1} + c^u \varepsilon_{t-1} + \varepsilon_t,\ if\ r_t \geq 0 \\ y_{t-1} + c^d\ \varepsilon_{t-1} + \varepsilon_t,\ if\ r_t < 0 \end{cases}$
7. *Seasonal*: $y_t = a + d_1 \times week_t + d_2 \times week_t^2 + \varepsilon_t$, with $week_t$ equaling the current week number of year minus 26.
8. *Implied volatility*: $y_t = a + d \times IV_{t-1}^2 + \varepsilon_t$, with $IV_t^2 = \lambda^{RV}\ IV_{t-1}^2 + (1 - \lambda^{IV}) \times iv_t^2$, with iv_t being the at-the-money implied volatility of S&P 500 options at close of day t and IV_t^2 the exponentially weighted average of single-day implied variances. The daily time series of implied volatilities is sourced from Bloomberg. Using option implied volatilities for forecasting has been studied before with mixed results. For instance, Jorion (1995) tested currency option volatilities without finding a marginal added value while Blair, Poon, and Taylor (2001) found that option implied data provides the best forecast accuracy among their set of tested alternatives. Christensen and Prabhala (1998) showed for monthly equity market data that option implied information improves volatility forecasts relative to pure historical ones. Poon and Granger's (2005) meta-analysis found that most empirical studies indicate that option implied data models dominate pure time-series models.
9. *Negative momentum:* $y_t = a + d \times mom_{t-1}^2 + \varepsilon_t$, with $mom_t = \lambda^{mom}\ mom_{t-1} + (1 - \lambda^{mom}) r_t \times I_{\{r_t < 0\}}$ denoting the exponentially weighted past average negative market return.
10. *Intraday:* $y_t = a + d \times RV_{t-1} + \varepsilon_t$, with $RV_t = \lambda^{RV}\ RV_{t-1} + (1 - \lambda^{RV}) rv_t$, with rv_t being the realized variance of 10-minute returns on day t and RV_t the exponentially weighted average of single-day variances. The 10-minute-wise S&P 500 Index time series is provided by Refinitiv. Previously, Blair, Poon, and Taylor (2001) tested the value of intraday returns for forecasting S&P 100 volatility and found it to be insignificant.

[2] Regarding the length of the forecast horizon. In a sense, the model calibration selects the right forecast horizon itself. When optimizing the different models, the parameters will define how fast the forecasts are moving. Fast-moving forecasts imply high short-term predictability, whereas slow-moving forecasts imply low predictability. Slow-moving forecasts also imply that the effective forecast horizon is longer. A hypothetical case is to forecast the daily market return itself. Due to very low predictability, models will automatically produce very slow-moving forecasts, hence effectively very long-term forecasts.

EXHIBIT 1
Calibration Results

Model	MSQE × 10,000,000	Average Daily Change of Volatility per Annum (p.a.) Forecast	Best Model Specification
Constant Variance	3.22	0.00	Constant $a = 0.0001551$
Realized Variance	2.43	1.05	Length of trailing window $N = 11$
Exponential Variance	2.42	1.39	$c = -0.84$
Asymmetric Exponential Variance	2.41	1.27	$c^u = -0.89$, $c^d = -0.82$
ARCH(1)	2.88	3.28	$a = 0.0001046$, $b_1 = 0.3255$
ARCH(2)	2.46	3.36	$a = 0.0000645$, $b_1 = 0.2006$, $b_2 = 0.3837$
GARCH(1,1)	2.39	1.27	$a = 0.0000053$, $b_1 = 0.9655$, $c_1 = -0.8115$
Seasonal	3.22	0.01	$a = 0.0001506$, $d_1 = 24.4/100000000$, $d_2 = 1.91/100000000$
Implied Volatility	2.39	1.20	$a = -0.000095$, $d = 0.0057$, $\lambda^{IV} = 0$
Negative Momentum	2.18	1.60	$a = 0.000043$, $d = 0.0004$, $\lambda^{MOM} = 0.867$
Intraday	2.12	2.62	$a = -0.0000044$, $d = 0.70$, $\lambda^{RV} = 0.358$

NOTES: We report calibration results of various optimized models. Models are fitted to predict volatility of the S&P 500 over time. We use daily returns data spanning January 2000 to December 2020.

Notably, the number of tuning parameters is three or lower for all models.[3] The calibration of these models is done by numerically minimizing the mean of the squared errors (MSQE). The results of the calibration are reported in Exhibit 1.

Exhibit 2 presents the development of volatility forecasts of optimal models over time. From optical inspection, the optimal calibrated models behave in a similar way throughout the period, except—of course—the constant variance model and the purely seasonal one. Notable differences can be observed in the noisiness of the forecasts over time, with the ARCH model forecasts on the high side because only very limited trailing data are used as well as for the intraday model.

The results in terms of forecast accuracy are reported in Exhibit 3. The exhibit shows the reduction of the MSQE relative to the one of the constant variance model. Overall, most models provide an improvement of forecasting accuracy of roughly 25% compared with the best constant variance estimate. Hence, from an average predictive accuracy level point of view, we cannot confirm the forecasting powers of around 50% as documented in Poon and Granger (2005) anymore; the reason might be their studied time frame ended in July 2003 so that volatility levels seen during the Great Financial Crisis of 2008 and the coronavirus crash of 2020 were simply not in the data yet. There is very limited fluctuation of more traditional volatility forecasts around the 25%. For example, realized variance, exponential variance, asymmetric exponential variance, ARCH(2), GARCH(1,1), and implied volatility all deliver almost the same predictive power. Outliers to the downside are ARCH(1) and the seasonal volatility model, which have rather low explanatory power. Outliers to the upside are the negative momentum and the intraday forecast models, which deliver improvements of 32% and 34%, respectively, which is a significant step up from 25%.

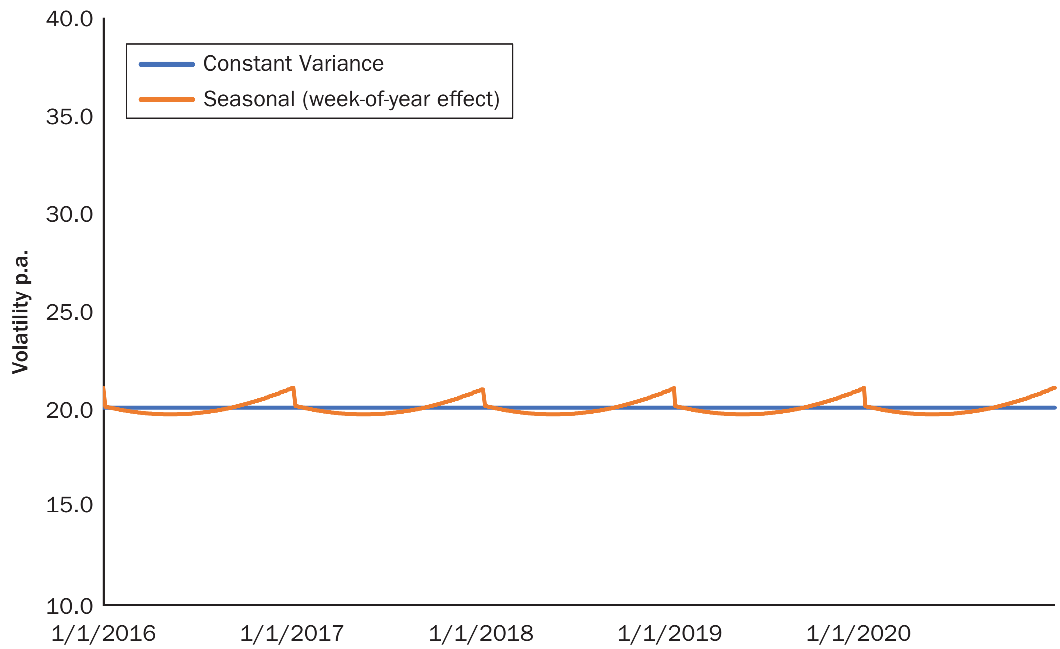
Regarding the models presented, it should be noted which models were chosen not to be presented although they have been tested. For instance, regarding seasonalities, we checked for other daily and monthly seasonal patterns, but this did not change the analysis's result that the tested seasonalities do not improve forecasts.

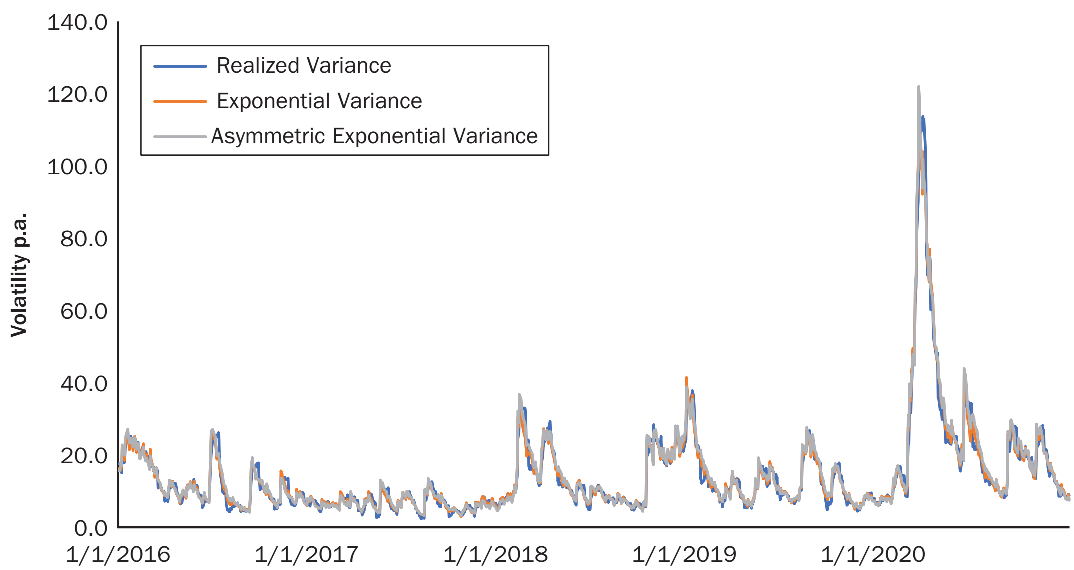
[3] The author is a strong believer in John von Neumann's quote, "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."

Obviously, another researcher might find a seasonality that is predictive. The article shows an effective way to model and benchmark them. Also, with respect to the ARCH and GARCH models, larger lags up to four were tested with no significant improvement in forecasting accuracy despite a much higher number of degrees of freedom.

EXHIBIT 2

Volatility Forecasts over Time

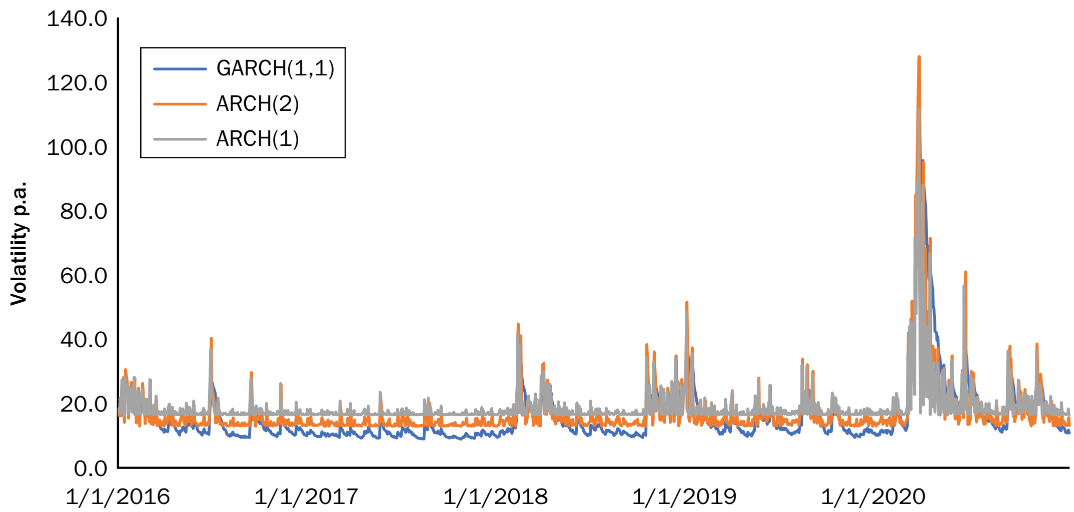


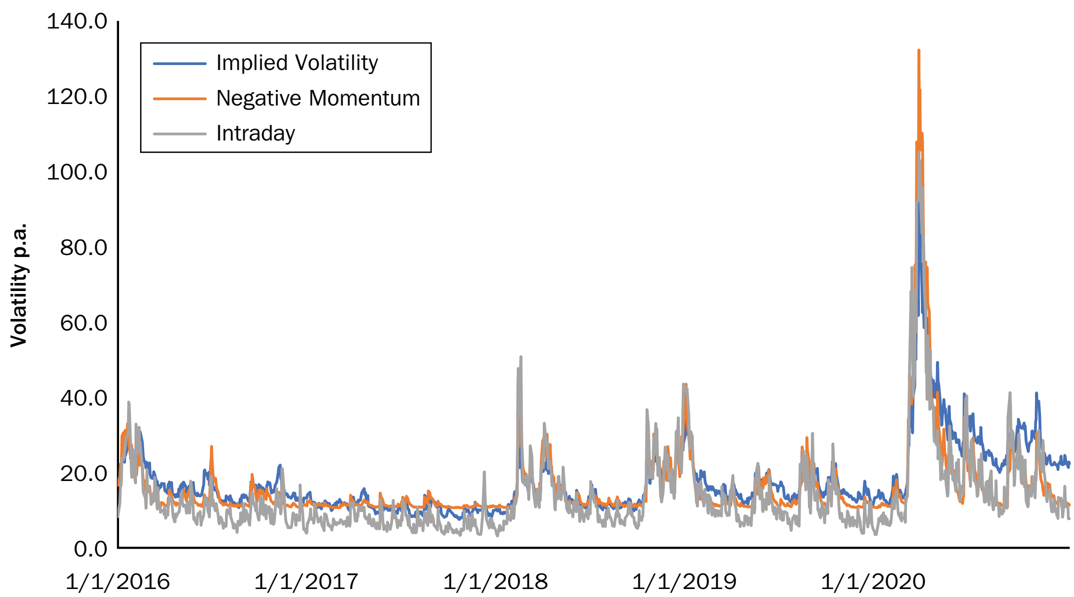


(continued)

EXHIBIT 2 *(continued)*

Volatility Forecasts over Time





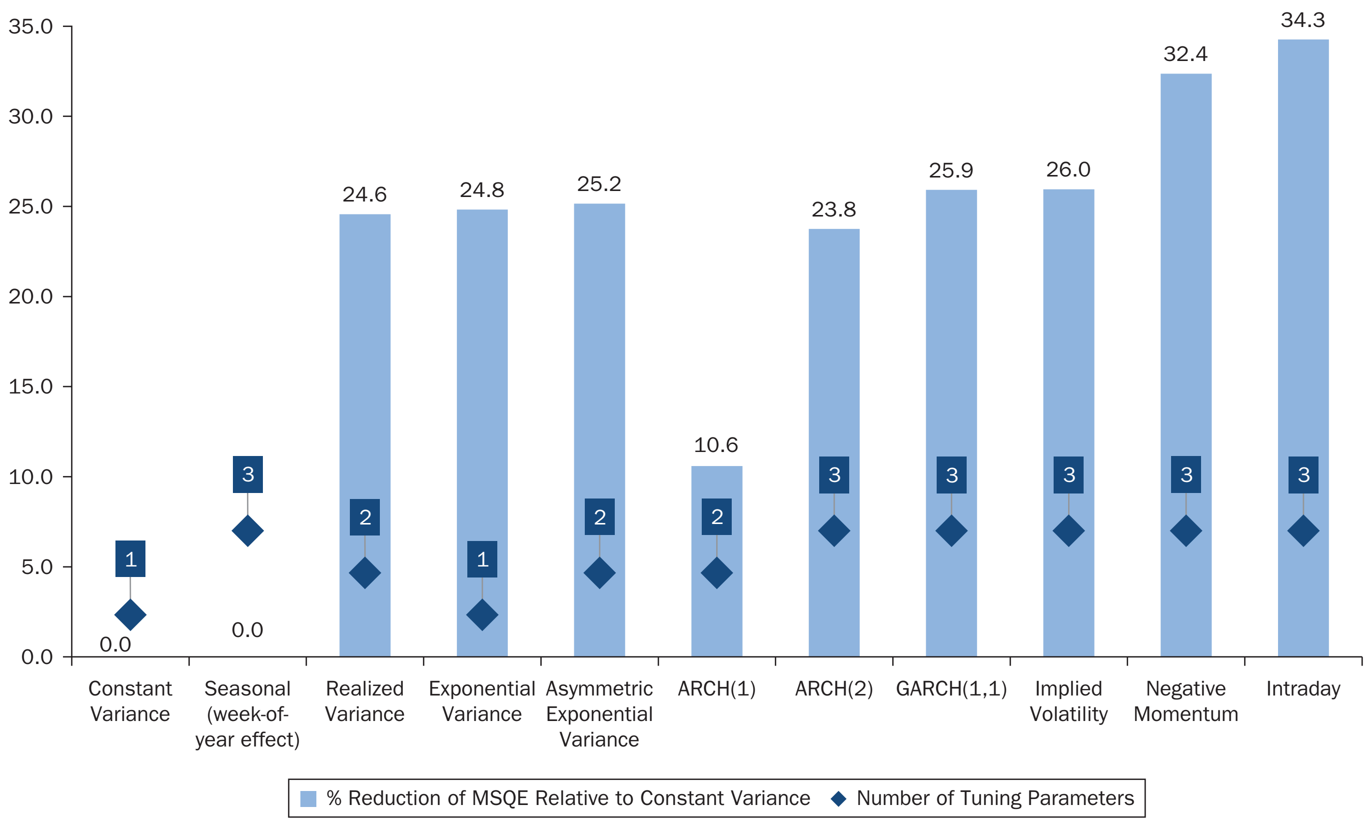
NOTES: We depict S&P 500 volatility forecasts (p.a., %) during the time frame 2016 to 2020 of various optimized models to illustrate the behavior of the estimates; the total time frame used for model estimation is 2000 to 2020. Annualized volatility is computed as $\hat{\sigma}(t) \times \sqrt{260} \times 100$. The *y*-axis is log scaled to allow better visualization of forecast variability when volatility is low and high.

SUMMARY

This article tries to improve the understanding of the volatility forecasting problem, which, in the literature, is somehow and unnecessarily treated differently than any other forecasting problem. We unify notation between financial and econometrical time-series models and clarify their links. We show that constant variance models are white-noise models, realized variance models are restricted AR(p) models, exponential variance models are restricted ARMA(p,q) models, ARCH models are AR(p) models, and GARCH models are ARMA(p,q) models.

EXHIBIT 3

Benchmarking Forecasting Accuracy



NOTES: Models are fitted to predict the daily variance of S&P 500 returns over time. We use daily returns data spanning January 2000 to December 2020. The MSQE reduction is defined as –(MSQE (Model)/MSQE (Constant Variance) – 1).

The empirical part of the article compares the forecast accuracy of 10 types of models in a joint dataset and runs a benchmarking analysis. All models are chosen to have three or fewer estimation parameters to control complexity risk and the risks of overfitting. Notably, the study finds that complexity does not automatically improve forecasts. Effectively, it is hard to beat simple trailing variance estimates by a significant degree. For instance, prominent academic models such as ARCH models underperform simple trailing variance models, and GARCH models provide only insignificant improvement despite an increased number of tuning parameters. We also find that other models such as implied volatility–informed or asymmetric and seasonal models hardly demonstrate a comparative improvement relative to simple exponential variance or historical variance models. Only negative momentum–based and intraday return–based models showed significant improvement.

I hope that this article allows for improved understanding of the volatility forecasting problem in financial markets, ultimately leading to simpler but better volatility prediction models that can be used in financial risk management and portfolio management units.

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 (1): 5–26.

Bollerslev, T. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31 (3): 307–327.

Christensen, B. J., and N. R. Prabhala. 1998. "The Relation between Implied and Realized Volatility." *Journal of Financial Economics* 50 (2): 125–150.

Engle, R. F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50 (4): 987–1007.

Jorion, P. 1995. "Predicting Volatility in the Foreign Exchange Market." *The Journal of Finance* 50 (2): 507–528.

Poon, S.-H., and C. Granger. 2005. "Practical Issues in Forecasting Volatility." *Financial Analysts Journal* 61 (1): 45–56.

Zangari, P. 1994. "Estimating Volatilities and Correlations." In *RiskMetrics—Technical Document*, 2nd ed., pp. 43–66. New York: Morgan Guaranty.