DOI: 10.1111/fire.12368

The Financial Review WILEY

ORIGINAL ARTICLE

Good risk measures, bad statistical assumptions, ugly risk forecasts

Michael Michaelides[1] | **Niraj Poudyal**[2]

[1]Cyprus University of Technology, Limassol, Cyprus

[2]Kathmandu University, Lalitpur, Nepal

Correspondence
Michael Michaelides, Department of Finance, Accounting and Management Science, Cyprus University of Technology, 30 Arch. Kyprianos Str., Limassol 3036, Cyprus.
Email: michael.michaelidis@cut.ac.cy

Abstract

This paper proposes the time-heterogeneous Student's *t* autoregressive model as an alternative to the various volatility forecast models documented in the literature. The empirical results indicate that: (i) the proposed model has better forecasting performance than other commonly used models, and (ii) the problem of reliable risk measurement arises primarily from the model risk associated with risk forecast models rather than the particular risk measure for computing risk. Based on the results, the paper makes recommendations to regulators and practitioners.

KEYWORDS
Basel III, financial risk forecasting, market risk, time-heterogeneous Student's *t* AR model, Value-at-Risk

JEL CLASSIFICATION
C18, C51, C52, C53, C58, G17, G32

1 | INTRODUCTION

A risk measure is a mathematical method for computing risk. Risk measures are widely used by various institutions (even individuals) to quantify potential downside risk within a firm, portfolio, or position over a specified period of time. For example, depository institutions use risk measures as the basis for determining their regulatory capital requirements. What is often unappreciated in practice is that behind every risk measure, there is a statistical volatility model, which is used to calculate risk forecasts. If the statistical assumptions underlying a volatility model are *bad*, the produced risk forecasts are likely to be *ugly*, regardless of whether the risk measure used is *good* or not. Inaccurate risk forecasting often leads to either overestimating or underestimating risk, which can be detrimental for the decision-making process in risk management.

This paper is primarily motivated by the recent changes in the regulatory framework of the Basel Committee on Banking Supervision (henceforth "BCBS" or "the Committee"), particularly the implementation of a new risk measure for determining market risk capital requirements. As argued in subsequent sections, the choice of the risk measure is not the only source of risk involved in the process of risk measurement. This paper makes recommendations and puts forward suggestions for further strengthening the market risk regulatory capital framework. Furthermore, given that risk measures are commonly used by nonregulatory institutions for a variety of applications, the paper also makes recommendations to the average practitioner.

1.1 | Historical background and regulatory framework of BCBS

At the end of 1974, the central bank governors of the G10 countries established a Committee on Banking Regulations and Supervisory Practices in response to a series of adverse events and disruptions in the international financial markets.[1] Later renamed BCBS, the Committee was designed as a forum for regular cooperation between its member countries on banking supervisory matters. In 1988, the Committee made the first attempt to assess capital. The Basel Capital Accord (Basel Committee on Banking Supervision, 1988), or Basel I as it has been known more recently, introduced a set of minimum capital requirements, mainly in relation to credit risk. After several revisions, the Committee refined the original framework to address risks other than credit risk. In 1996, the Committee issued the Market Risk Amendment to the Capital Accord (Basel Committee on Banking Supervision, 1996), designed to incorporate a capital requirement for the market risk arising from banks' exposures to foreign exchange, traded debt securities, equities, commodities, and options. An important aspect of the Market Risk Amendment was that banks with well-established risk management functions were, for the first time, allowed to use internal models as a basis for measuring their market risk capital requirements.

Over the years, the reforms in the successive accord of Basel II (Basel Committee on Banking Supervision, 2004) improved the effectiveness of supervision and strengthened the regulatory capital framework. Likewise, banks themselves refined the practice of risk management by the development of new techniques respecting the internal risk measurement. Nevertheless, the financial crisis of 2007–2009 revealed structural weaknesses in the financial system, regulation, and supervision. In response to these weaknesses, the Committee proposed a stronger regulatory framework, known as Basel III, concerning better quality of capital, increased coverage of risk for capital market activities, and better liquidity standards, among other benefits. As part of the review of its regulatory approach, the Committee has proposed a shift from Value-at-Risk (VaR) with a confidence level of 99% to Expected Shortfall (ES) with a confidence level of 97.5% for the internal models-based approach (Basel Committee on Banking Supervision, 2016, 2019b). The decision of the Committee to replace the most commonly used market risk measure of VaR stemmed from a number of weaknesses associated with VaR, including its inability to capture tail risk and lack of coherence (Basel Committee on Banking Supervision, 2011, 2012, 2013, 2016).

1.2 | Related literature and contributions of the paper

This paper proposes the time-heterogeneous Student's *t* autoregressive model (henceforth "t-StAR") as an alternative to the various volatility forecast models documented in the literature. To the best of our knowledge, this model has not been used before as a risk forecast model (we provide R code for the model in Appendix B). The forecasting

[1] Namely, the breakdown of Bretton Woods system of managed exchange rates, the closure and liquidation of Bankhaus Herstatt, and the collapse of Franklin National Bank of New York.

performance of this model is compared to that of the most commonly used risk models. The empirical results indicate that: (i) the proposed model has better forecasting performance than other commonly used models, and (ii) the problem of reliable risk measurement arises primarily from the model risk associated with risk forecast models rather than the particular risk measure for computing risk.

Our paper contributes to the existing literature in two main dimensions. First, the paper adds to the broad risk modeling/forecasting literature through the introduction of the t-StAR model. Previous studies have shown that an inappropriate choice of risk models ("model risk") produces inaccurate risk forecasts, primarily because the statistical assumptions underlying those models are often invalid (Beder, 1995; Hendricks, 1996; Alexander & Sarabia, 2012; Boucher et al., 2014; Danielsson et al., 2016). Within the same context, a number of studies have shown that risk models are more likely to perform poorly in times of financial crisis (Danielsson, 2002; O'Brien & Szerszen, 2017; Danielsson et al., 2016).

We argue that, even though it is impossible to identify a single risk model, which can always forecast risk perfectly, the choice of a *data-coherent* model (Hendry & Richard, 1983) can considerably improve the accuracy of risk forecasts: a rich probabilistic structure that can describe all the relevant features of financial data that could be accounted for over a period of time. The results in this paper show that the proposed t-StAR model performs well during periods of both financial crisis and stability because its probabilistic structure captures common features exhibited by financial data. Its underlying statistical assumptions of Student's *t*, heteroskedasticity, and temporal dependence contribute to the modeling of the fat-tailed, non-Gaussian, and nonlinearly dependent financial returns. Further, and perhaps more important, the model incorporates orthogonal trend polynomials with the aim to model the time-varying volatility exhibited by financial returns. This is a distinctive characteristic of the proposed model because its conditional variance is a function of both the lagged conditioning variables (heteroskedastic) and time (time-heterogeneous).

Second, the paper relates to the ongoing debate around the use of VaR versus ES. This debate is nothing new (see, for example, Yamai & Yoshiba, 2005), yet it grew in popularity after the decision of BCBS to move from VaR to ES for regulatory purposes. Overall, there is little doubt among the literature that the transition from VaR to ES will lead to higher level of capital requirements since ES has the ability to capture the fat-tailed behavior of risk, which leads to more capital. This is important, of course, because more capital makes banks more resilient and reduces the probability of financial distress (Repullo & Suarez, 2013). Despite the obvious advantage of ES to capture tail risk, however, the decision of BCBS has been heavily criticized by the recent literature. A number of studies have focused on the theoretical failure of ES to satisfy the mathematical property of elicitability (Gneiting, 2011). These studies suggest that elicitability should be taken into consideration when choosing a risk measure because nonelicitable measures are difficult to backtest (Emmer et al., 2015; Ziegel, 2016). In addition, important work has been done to analyze how VaR and ES react to different sources of model risk (Kellner & Rösch, 2016), as well as to evaluate the forecasting accuracy of the two risk measures (Danielsson & Zhou, 2015). The findings show that both VaR and ES are vulnerable to model risk and yield inaccurate risk forecasts.

Our point of view is that the decision of BCBS is solely motivated by the theoretical strengths of ES over VaR. From a theoretical perspective, there is no question that ES is superior to VaR. Yet, theoretical superiority alone does not automatically lead to accurate risk forecasts because the process of forecasting volatility involves statistical modeling. Supporting the argument of Danielsson and Zhou (2015) that VaR and ES are related by a small constant, we argue that, in the presence of model risk, both VaR and slightly greater ES risk forecasts are likely to be inaccurate. The results in this paper suggest that it is of primary importance to minimize the model risk associated with the risk forecast models used by banks. As the empirical results show, the choice of an appropriate model can considerably improve the accuracy of risk forecasts, even during periods of financial crisis. Our regulation recommendation is that the BCBS should restrict the scope of internal modeling to risk models, which are likely to pass some kind of statistical adequacy test.

Even though much of the discussion in this paper is centered around the market risk regulatory framework of BCBS, our results apply equally to financial institutions not regulated by the Basel Accords, including pension funds, insurance companies, mutual funds, and hedge funds, among others. Applications may include, but are not limited to, risk budgeting, economic capital, survival analysis, long-term risk analysis for pension plans, risk management on the trading floor, and risk reporting (Danielsson, 2011). Hence, we also provide recommendations to the average practitioner, who may or may not have access to sophisticated expertise, such as model development and validation.

The remaining of the paper is organized as follows. Section 2 defines the concepts of VaR and ES, by paying particular attention to the delineation of the underlying differences between the perspectives of theory and probability. Section 3 presents the risk models used to forecast volatility, while Section 4 lays out the data and methodology employed in the analysis. Section 5 presents and discusses the empirical results. Section 6 concludes the study, and makes recommendations to regulators and practitioners.

2 | VALUE-AT-RISK AND EXPECTED SHORTFALL

The VaR analysis was formally introduced as a market risk management tool in the technical document Riskmetrics, published by J.P. Morgan. However, it was the endorsement of the use of VaR for regulatory purposes by BCBS in 1996 that established VaR as the most commonly used risk measure.

VaR. Given some confidence level $\alpha \in (0,1)$, the VaR of a portfolio with loss L at the confidence level α is given by the smallest number l, such that the probability of loss L exceeding l is no larger than $(1-\alpha)$:

$$\begin{aligned} VaR_{\alpha} &= \inf\left\{l \in \mathbb{R} : \Pr(L > l) \leq (1-\alpha)\right\}, \\ &= \inf\left\{l \in \mathbb{R} : F_L(l) \geq \alpha\right\}, \end{aligned} \tag{1}$$

where $F_L(l) = \Pr(L \leq l)$ is the loss distribution function.

Despite its simple appearance, VaR has been subject of sustained criticism, with Artzner et al. (1997, 1999) being the most noticeable. In their seminal papers, they identified two major weaknesses of VaR. The first weakness is associated with the inability of VaR to capture "tail risk." More specifically, VaR does not provide any information beyond what is the most optimistic of the worst case scenarios. The second weakness of VaR is that it is not a coherent risk measure (Artzner et al., 1999) because it does not always satisfy the axiom of subadditivity. The violation of this axiom is of particular concern since it contradicts with the principle of diversification. For instance, if the regulator uses a non-subadditive risk measure in determining the regulatory capital for banks, then banks have an incentive to legally break up into various subsidiaries in order to reduce their regulatory capital requirements (McNeil et al., 2005).

Undoubtedly, the two weaknesses of VaR influenced the decision of BCBS to propose a shift from VaR to ES. This is clearly noted in Basel Committee on Banking Supervision (2011): "ES avoids the major flaws of VaR" (Basel Committee on Banking Supervision, 2011, p. 25) because it "does account for the severity of losses beyond the confidence threshold" and "is always subadditive and coherent" (Basel Committee on Banking Supervision, 2011, p. 20). According to Basel Committee on Banking Supervision (2013, 2016), the use of ES will "strengthen model standards" (Basel Committee on Banking Supervision, 2013, p. 5) and "help to ensure … capital adequacy during periods of significant financial market stress" (Basel Committee on Banking Supervision, 2016, p. 1).

The risk measure of ES has been proposed by Artzner et al. (1997) to overcome the tail risk and lack of coherence weaknesses inherent in VaR.

ES. Given some confidence level $\alpha \in (0,1)$, the ES of a portfolio with loss L at the confidence level α is defined as

$$ES_{\alpha} = E(L \mid L \geq VaR_{\alpha}), \tag{2}$$

where VaR_{α} is the VaR of a portfolio as defined in Equation (1).

whereas VaR does not account for the tail of the distribution, ES is the expected loss conditional on VaR being violated; thus, ES is larger or equal to VaR. Moreover, ES is a coherent risk measure since it satisfies the axiom of subadditivity (Artzner et al., 1999). Nevertheless, ES has been proven not to be elicitable (Gneiting, 2011), which makes

ES difficult to backtest, as opposed to VaR. For this reason, BCBS kept VaR as the risk measure for backtesting purposes (Basel Committee on Banking Supervision, 2019a).[2]

2.1 | Theory versus probability

In practice, the calculation of VaR[3]

$$VaR_{\alpha,t} = -\hat{\sigma}_t \times F^{-1}_{1-\alpha} \times W_{t-1}, \tag{3}$$

and ES[4]

$$ES_{\alpha,t} = -\hat{\sigma}_t \times \frac{f(F^{-1}_{1-\alpha})}{1-\alpha} g(.) \times W_{t-1}, \tag{4}$$

involve the estimated volatility forecast, $\hat{\sigma}_t$.[5]

As it is clear from Equations (3) and (4), the calculations of both VaR and ES are severely affected by the estimated volatility forecast, $\hat{\sigma}_t$. This estimation depends crucially on the validity of the statistical assumptions underlying the model used to forecast risk. Hence, model risk is a source of risk that arises solely from the use of inappropriate risk forecast models.

On the other hand, the theoretical weaknesses of VaR and ES are stemming from the very nature of their definitions. In fact, there is very little to be done from the perspective of probability that could address such theoretical weaknesses. First, the tail risk weakness of VaR and lack of elicitability weakness of ES will remain for any underlying distribution. Second, although the axiom of subadditivity is only violated for fat-tailed distributions (Artzner et al., 1999; Danielsson et al., 2013), employing the normal or another thin-tailed distribution in an attempt to satisfy subadditivity will falsely assume less risk than actually is present.[6]

Therefore, even though ES is superior to VaR from a theoretical perspective, from a purely probabilistic perspective both risk measures can yield equally inaccurate risk forecasts. On the whole, model risk, which is purely probabilistic in nature, is a major source of risk in the process of risk measurement, and can be far more crucial than any of the theoretical weaknesses associated with the risk measures of VaR and ES.

[2] For additional information on the concepts and definitions, see (McNeil et al., 2005, Chapter 2).

[3] The definition of VaR can be expressed as

$$\Pr(r_t W_{t-1} \leq -VaR_{\alpha,t}) = 1-\alpha,$$

where $r_t = \frac{W_t - W_{t-1}}{W_{t-1}}$ is the change in the value of the portfolio. Thus,

$$\begin{aligned} 1-\alpha &= \Pr\left(\frac{r_t}{\sigma_t} \leq -\frac{VaR_{\alpha,t}}{\sigma_t W_{t-1}}\right), \\ F^{-1}_{1-\alpha} &= -\frac{VaR_{\alpha,t}}{\sigma_t W_{t-1}}, \\ VaR_{\alpha,t} &= -\sigma_t \times F^{-1}_{1-\alpha} \times W_{t-1}. \end{aligned}$$

[4] The ES for a given distribution is calculated by direct integration. It involves the general term, $\frac{f(F^{-1}_{1-\alpha})}{1-\alpha} g(.)$, where $f(.)$ and $F(.)$ denote the probability density function and cumulative distribution function of the underlying distribution, respectively, and $g(.)$ is a non-negative function that depends on the underlying distribution. For the normal standard distribution, $g(.) = 1$. For other standard distributions, $g(.) > 1$; see McNeil et al. (2005) for details.

[5] In the two equations, $F(.)$ and $f(.)$ denote the cumulative distribution function and probability density function of the underlying distribution, respectively, $g(.)$ is a non-negative function that depends on the underlying distribution, and W_{t-1} is the value of the portfolio at time $t-1$. Thus, unlike $\hat{\sigma}_t$, which must be estimated, the rest of the components in the equations are known.

[6] The trade-off between coherence and underestimation of risk occurs because the tails of financial return distributions are much thicker than for the normal; see, for example, Mandelbrot (1963).

3 | RISK FORECAST MODELS

3.1 | Commonly used models

In this paper, we first consider the most commonly used risk forecast models (Danielsson et al., 2016): historical simulation (HS), moving average (MA), exponentially weighted moving average (EWMA)[7], normal generalized autoregressive conditional heteroskedasticity (NGARCH), and Student's *t* GARCH (StGARCH). These models are commonly used by various financial institutions, including banks regulated by the Basel Accords. The conditional variances of the most commonly used risk forecast models (except HS, which is nonparametric) are stated below[8]; for a detailed description of these models, see Danielsson (2011).

HS. The VaR at confidence level α is defined as the negative $(T \times (1-\alpha))$ th value in the sorted return vector, multiplied be the monetary value of the portfolio.

MA. The conditional variance of MA is given by

$$\hat{\sigma}_t^2 = \frac{1}{W_E}\sum_{i=1}^{W_E} r_{t-i}^2, \tag{5}$$

where r_t is the return for day t, and W_E denotes the length of the estimation window.

EWMA. The conditional variance of EWMA is given by

$$\hat{\sigma}_t^2 = (1-\lambda)r_{t-1}^2 + \lambda\hat{\sigma}_{t-1}^2, \tag{6}$$

where $0 < \lambda < 1$ denotes a decay factor.

NGARCH. The conditional variance of NGARCH(p, q) is given by

$$\hat{\sigma}_t^2 = \omega + \textstyle\sum_{i=1}^{p}\alpha_i r_{t-i}^2 + \sum_{j=1}^{q}\beta_j\hat{\sigma}_{t-j}^2, \tag{7}$$

where ω, $\{\alpha_i, i = 1, 2, \dots, p\}$, and $\{\beta_j, j = 1, 2, \dots, q\}$ are estimable parameters. Restrictions imposed on parameters are necessary to ensure positive volatility forecasts.

StGARCH. The residuals of StGARCH$(p, q; \nu)$ are Student's *t* distributed with ν degrees of freedom. The degrees of freedom, ν, is estimated as an extra parameter along with the NGARCH model parameters. Similar to NGARCH, restrictions imposed on parameters are necessary to ensure positive volatility forecasts.

3.2 | The proposed model

In addition to the most commonly used models, we also consider the proposed t-StAR model. This model is a traditional autoregressive model that assumes the Student's *t* distribution and incorporates orthogonal trend polynomials.

The conditional variance of t-StAR$(p; \nu)$ is given by

$$\hat{\sigma}_t^2 = \left(\frac{\nu\omega^2}{\nu+p-2}\right)\left[1 + \frac{1}{\nu}\left[\textstyle\sum_{i=1}^{p}\left(r_{t-i} - \mu_{t-i}(t)\right)\Sigma^{-1}\left(r_{t-i} - \mu_{t-i}(t)\right)\right]\right], \tag{8}$$

[7] EWMA is not permitted under the Basel Accords for the purpose of calculating VaR because its exponential weights decline to zero very quickly (Danielsson, 2011). However, this model is an option for nonregulatory financial institutions.

[8] For all parametric models, the volatility forecast for day t (see Equation 3) is calculated as the square root of the conditional variance forecast for day t, that is, $\hat{\sigma}_t = \sqrt{\hat{\sigma}_t^2}$.

where ω^2 is a scaling variance constant, Σ is the variance–covariance matrix, and $\mu_{t-i}(t)$ is the time-varying unconditional mean of r_{t-i}, for $i = 1, 2, \ldots, p$. The degrees of freedom of the Student's t distribution, ν, is an estimable parameter. The degrees of freedom for the conditional Student's t distribution is equal to ν plus the number of lagged conditioning variables, p.[9]

The proposed model has several advantages. First, it does not assume that returns are normally distributed. Going back at least to Mandelbrot (1963), many studies argue that the peak of the return distribution is much higher and the tails are much thicker than for the normal distribution. This non-normal behavior of returns suggests replacing the normal distribution with another distribution from the elliptically symmetric family (Fang et al., 1990). Distributions within the elliptical family allow for a heteroskedastic conditional variance while retaining the bell-shaped symmetry and linearity of the conditional mean (Kelker, 1970). Among the members of the elliptical family, the Student's t is the most commonly used non-normal distribution for modeling financial returns (Bollerslev, 1987) and for forecasting risk (Danielsson, 2011).

Second, the conditional variance of the model ensures positive definiteness of the variance–covariance matrix estimate without requiring complicated and unverifiable parameter restrictions (Heracleous, 2006), which often cause computational problems such as singularities. Third, the model incorporates orthogonal trend polynomials[10] to capture the time-varying volatility exhibited by returns (Michaelides & Spanos, 2020). Therefore, the conditional variance of the model is not only a function of the conditioning variables (heteroskedastic), but also a function of time (time-heterogeneous) via the unconditional mean $\mu(t)$.

4 | DATA AND METHODOLOGY

4.1 | Data

The data used in this paper are a sample from the major assets classes, and includes equities, fixed income, commodities, and currencies. We present results for nine individual assets (see Table 1)[11], which are treated as univariate portfolios. In order to determine the relative change in the price of the assets over trading days, asset prices are converted to log returns using adjusted daily closing prices.[12] While we only present results for nine assets treated as univariate portfolios, we have also examined other assets, as well as bivariate and multivariate portfolios containing assets from various asset classes. Reasonable variations in the definitions of portfolios do not seem to particularly affect the overall picture of the empirical results, and thus do not alter the main conclusions of this paper.

4.2 | Sample period

The sample period consists of the estimation window, $\{W_{E_i},\ i = 1, 2, \ldots, m\}$, and the testing window, $\{W_{T_i},\ i = 1, 2, \ldots, n\}$. Therefore, the entire sample period has a length of $m + n$ trading days, with m being the length of the estimation

[9] For additional information about the (conditional) Student's t distribution, see Fang et al. (1990). For additional information regarding the conditional variance of Student's t regression and autoregression models, see Spanos (1994) and Heracleous (2006).

[10] In the analysis, we use the Gram–Schmidt orthonormal trend polynomials. These polynomials offer the flexibility of extending them to higher orders without giving rise to collinearity problems; see Michaelides and Spanos (2020) for details.

[11] Six of the nine assets are the same as the ones used by (Danielsson, 2002, tab. I).

[12] On April 20, 2020, the front-month West Texas Intermediate (WTI) crude oil contract dropped 306% for the session, to settle at negative $37.63 (−$37.63) a barrel on the New York Mercantile Exchange. To determine the relative change in the price of crude oil on April 20 and April 21, 2020, simple returns were used instead of log returns.

TABLE 1 Dataset information.

Asset	Asset class	Description	Ticker
S&P 500	Equity	U.S. equity index	SPX
Nikkei	Equity	Japan equity index	NKY
Hang Seng	Equity	Hong Kong equity index	HSI
Microsoft	Equity	Microsoft Corporation stock price	MSFT US
U.S. bond	Fixed Income	S&P U.S. treasury bond total return index	SPBDUSBT
Oil	Commodity	Crude oil, West Texas Intermediate, price	CL1
Gold	Commodity	Gold, 100 oz., price	GC1
GBP/USD	Currency	British pound/U.S. dollar cross rate	GBPUSD
MYR/GBP	Currency	Malaysian ringgit/British pound cross rate	MYRGBP

Note: The table lists the assets used in the analysis of this paper. These assets include equities, fixed income, commodities, and currencies. Asset prices are obtained from Bloomberg. The table provides descriptions of the assets and their Bloomberg ticker symbols. In the analysis, these assets are treated as univariate portfolios.

window and *n* being the length of the testing window. The forecasting of daily volatility across the testing window is carried out over different estimation windows; specifically, five different estimation windows of 100, 250, 500, 1000, and 2000 days. The testing window starts on the first trading day after the detected structural break of the Great Recession, and runs through the last trading day of May 2021. As discussed below, the structural break point date of the Great Recession varies from one portfolio to another; thus the length of the testing window is not the same for all portfolios.

4.3 | Structural breaks and testing periods

For each of the portfolios considered, we divide the testing window into four testing periods: the entire roughly 14-year period starting at the beginning of the Great Recession and ending on the last trading day of May 2021, the two volatile subperiods of the Great Recession and COVID-19 recessions, and the relatively stable subperiod from the end of the Great Recession to the start of the COVID-19 recession. To detect the volatility structural breaks associated with the Great Recession and COVID-19 recessions, we use the sup MZ test[13] of Ahmed et al. (2017). The rationale for using this test is that it has been shown to have better empirical performance in detecting structural changes in the presence of heteroskedasticity (typical of financial time series) than other widely used tests, such as the sup *F* test of Quandt (1960); see Ahmed et al. (2017).

Figure 1 plots the daily log returns of the S&P 500 (blue line). The red line illustrates the sup MZ statistics produced by the sup MZ test. The positions of the maxima of the sup MZ function are used to locate the structural breaks in volatility associated with the Great Recession and COVID-19 recessions. As shown in Figure 1, the structural break point dates for the S&P 500 are determined as July 19, 2007, June 22, 2009, and February 21, 2020. For the purpose of our analysis, these dates represent the baseline of the start of the Great Recession, the end of the Great Recession (or, equivalently, the start of the post-Great Recession/pre-COVID-19 subperiod), and the start of COVID-19 recession, respectively. Comparing these dates with the National Bureau of Economic Research (NBER) official recession

[13] The sup MZ test is a natural extension of the MZ test of Maasoumi et al. (2010). While the latter tests for structural changes at a fixed and known breakpoint, the sup MZ test extends it to the case of the unknown breakpoint.

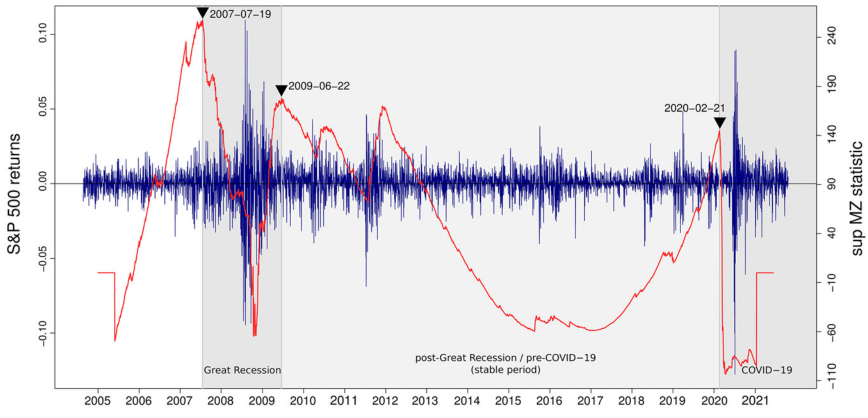

FIGURE 1 Structural breaks for the S&P 500 returns. The figure is used as an illustrative example to show how the sup MZ test of Ahmed et al. (2017) is employed to identify the structural break point dates for the S&P 500. The blue line represents the S&P 500 returns. The red line represents the sup MZ statistics produced by the sup MZ test. The structural break point dates for the S&P 500 are determined as July 19, 2007, June 22, 2009, and February 21, 2020. These three dates represent the baseline of the start of the Great Recession, the end of the Great Recession (or, equivalently, the start of the post-Great Recession/pre-COVID-19 subperiod), and the start of the COVID-19 recession, respectively.

dates,[14] it seems that the sup MZ test identifies volatility structural breaks rather accurately.[15] The sup MZ test does not detect a structural change in volatility to determine the location of the structural break date associated with the end of the COVID-19 recession. This implies that the impact of the COVID-19 pandemic on the volatility of financial markets was ongoing at the time we performed our analysis. The last day in our sample is the last trading day of May 2021.

As can be seen in Figure 1, the sup MZ test identifies additional structural changes in volatility within the post-Great Recession/pre-COVID-19 subperiod. For instance, the test identified two structural breaks in mid-2010 and late 2011. These short periods of market turbulence are, for simplicity, ignored. Therefore, we assume that the subperiod from the end of the Great Recession to the start of COVID-19 recession is a stable period.

Similar figures to Figure 1 are not presented for the rest of the portfolios. The structural break dates identified by the sup MZ test, and thus the four testing periods, do not vary significantly from one portfolio to another. Panel A of Table 2 reports the structural break point dates for each portfolio, as identified by the sup MZ test. These dates represent the baseline start dates of the three testing subperiods. The date ranges of these testing subperiods are reported in Panel B of Table 2. Each subperiod begins on the first trading day after the corresponding structural break point date, and ends on the next identified structural break point date. The end date of the COVID-19 subperiod is the

[14] The NBER's Business Cycle Dating Committee (BCDC) maintains a chronology of U.S. business cycles. The chronology identifies the months of peaks and troughs that frame economic expansions and recessions based on a range of monthly measures of aggregate real economic activity, including real personal income less transfers, nonfarm payroll employment, employment as measured by the household survey, real personal consumption expenditures, wholesale-retail sales adjusted for price changes, and industrial production. According to the NBER's BCDC, the official beginning and ending dates of the Great Recession and COVID-19 recessions in the United States are December 2007 to June 2009 (18 months) and February 2020 to April 2020 (2 months), respectively.

[15] Though financial markets and the real economy do not move together at the same time, it seems that the structural break point dates detected by the sup MZ test are similar to the NBER official recession dates. The baseline end date of the Great Recession (June 22, 2009) and baseline start date of the COVID-19 (February 21, 2020) recessions detected by the sup MZ test are about identical to the NBER official dates (June 2009 and February 2020). On the other hand, the baseline start date of the Great Recession detected by the sup MZ test (July 19, 2007) is about 6 months prior to the NBER official date (December 2007). This baseline date does not seem unreasonable. Halfway through 2007, United States publicly traded companies filed for bankruptcy at a record base. The latter was almost immediately reflected in the United States and world financial markets. Moreover, this baseline date is in support of the belief that financial markets often move in anticipation of the real economy (see, for example, Stock & Watson, 2003).

TABLE 2 Structural breaks and testing periods.

Panel A: Structural break point dates			
	Great Recession (GR)	Post-GR/pre-COVID-19 (stable period)	COVID-19
S&P 500	2007-07-19	2009-06-22	2020-02-21
Nikkei	2007-08-16	2009-05-19	2020-02-28
Hang Seng	2007-07-26	2009-06-10	2020-03-06
Microsoft	2007-07-31	2009-07-28	2020-02-19
U.S. bond	2007-06-12	2009-07-31	2020-02-20
Oil	2007-08-15	2009-07-08	2020-03-05
Gold	2007-08-02	2009-04-06	2020-02-27
GBP/USD	2007-07-24	2009-06-02	2020-03-13
MYR/GBP	2007-07-26	2009-06-02	2020-03-02

Panel B: Testing subperiods			
	Great Recession (GR)	Post-GR/pre-COVID-19 (stable period)	COVID-19
S&P 500	2007-07-20 – 2009-06-22 (485 days)	2009-06-23 – 2020-02-21 (2685 days)	2020-02-24 – 2021-05-28 (320 days)
Nikkei	2007-08-17 – 2009-05-19 (426 days)	2009-05-20 – 2020-02-28 (2639 days)	2020-03-02 – 2021-05-31 (305 days)
Hang Seng	2007-07-27 – 2009-06-10 (459 days)	2009-06-11 – 2020-03-06 (2650 days)	2020-03-09 – 2021-05-31 (303 days)
Microsoft	2007-08-01 – 2009-07-28 (502 days)	2009-07-29 – 2020-02-19 (2658 days)	2020-02-20 – 2021-05-28 (322 days)
U.S. bond	2007-06-13 – 2009-07-31 (535 days)	2009-08-03 – 2020-02-20 (2638 days)	2020-02-21 – 2021-05-28 (319 days)
Oil	2007-08-16 – 2009-07-08 (477 days)	2009-07-09 – 2020-03-05 (2686 days)	2020-03-06 – 2021-05-28 (312 days)
Gold	2007-08-03 – 2009-04-06 (422 days)	2009-04-07 – 2020-02-27 (2745 days)	2020-02-28 – 2021-05-28 (317 days)
GBP/USD	2007-07-25 – 2009-06-02 (485 days)	2009-06-03 – 2020-03-13 (2813 days)	2020-03-16 – 2021-05-31 (316 days)
MYR/GBP	2007-07-27 – 2009-06-02 (483 days)	2009-06-03 – 2020-03-02 (2804 days)	2020-03-03 – 2021-05-31 (324 days)

Note: Panel A of the table reports the structural break point dates. These dates are identified by the sup MZ test of Ahmed et al. (2017), and represent the baseline start dates of the three testing subperiods. Panel B of the table reports the date range of the three subperiods within the entire testing period. Each subperiod begins on the first trading day after the corresponding structural break point date, and ends on the next identified structural break point date. The end date of the COVID-19 subperiod is the last trading day of May 2021. The number of trading days contained in each testing subperiod is reported in parentheses. The entire testing period (testing window) is determined by the first date of the Great Recession subperiod and the last date of the COVID-19 subperiod. Thus, the number of trading days contained in the entire testing period is the summed number of days of the three testing subperiods.

last trading day of May 2021. The entire testing period (testing window) is determined by the first date of the Great Recession subperiod and the last date of the COVID-19 subperiod.

4.4 | Descriptive statistics

Table 3 provides basic descriptive statistics for each portfolio over the four testing periods. The reported descriptive statistics are not surprising or out of the ordinary. For instance, certain asset classes, such as equities and

TABLE 3 Descriptive statistics.

	Great Recession (GR)							Post-GR/pre-COVID-19 (stable period)						
	Obs	Mean	Std	Skew	Kurt	Min	Max	Obs	Mean	Std	Skew	Kurt	Min	Max
S&P 500	485	−0.11	2.25	−0.05	3.85	−9.47	10.96	2685	0.05	0.94	−0.48	4.29	−6.90	4.84
Nikkei	426	−0.13	2.55	−0.25	4.17	−12.11	13.23	2639	0.03	1.31	−0.55	5.11	−11.15	7.43
Hang Seng	459	−0.05	2.87	0.18	2.91	−13.58	13.41	2650	0.01	1.17	−0.29	2.01	−6.02	5.52
Microsoft	502	−0.04	2.77	0.36	4.71	−12.46	17.06	2658	0.08	1.43	−0.10	6.24	−12.10	9.94
U.S. bond	535	0.03	0.35	−0.01	1.90	−1.67	1.74	2638	0.01	0.20	−0.09	1.18	−1.02	0.91
Oil	477	−0.04	3.72	0.20	2.45	−13.07	16.41	2686	−0.01	2.07	0.05	3.20	−10.79	13.69
Gold	422	0.06	1.79	0.28	2.09	−6.05	8.62	2745	0.02	1.00	−0.71	6.56	−9.82	4.62
GBP/USD	485	−0.05	0.85	−0.49	1.97	−3.47	2.93	2813	−0.01	0.56	−1.19	18.79	−8.40	3.00
MYR/GBP	483	0.04	0.85	0.67	3.93	−3.29	5.44	2804	0.00	0.65	0.44	4.75	−2.95	6.48
	COVID-19							Entire period						
	Obs	Mean	Std	Skew	Kurt	Min	Max	Obs	Mean	Std	Skew	Kurt	Min	Max
S&P 500	320	0.07	2.00	−0.92	10.04	−12.77	8.97	3490	0.03	1.32	−0.55	12.67	−12.77	10.96
Nikkei	305	0.10	1.55	0.12	4.13	−6.27	7.73	3370	0.02	1.54	−0.46	7.76	−12.11	13.23
Hang Seng	303	0.04	1.44	−0.46	1.83	−5.72	4.92	3412	0.01	1.53	−0.02	8.41	−13.58	13.41
Microsoft	322	0.09	2.55	−0.38	7.62	−15.95	13.29	3482	0.06	1.81	0.01	9.80	−15.95	17.06
U.S. bond	319	0.01	0.31	0.48	9.49	−1.69	1.79	3492	0.01	0.24	0.09	5.21	−1.69	1.79
Oil	312	−0.26	19.53	−11.4	195.20	−305.97	126.60	3475	−0.04	6.27	−30.97	1672.65	−305.97	126.60
Gold	317	0.05	1.31	−0.21	3.63	−5.11	5.78	3484	0.03	1.15	−0.25	5.86	−9.82	8.62
GBP/USD	316	0.05	0.61	−0.49	5.75	−3.78	2.70	3614	−0.01	0.61	−0.95	12.27	−8.40	3.00
MYR/GBP	324	−0.03	0.60	0.50	2.90	−2.19	3.25	3611	0.01	0.68	0.52	4.80	−3.29	6.48

Note: The table reports descriptive statistics of the data of daily log returns of the adjusted closing price. Descriptive statistics are provided for the two volatile testing subperiods of the Great Recession and COVID-19, the stable post-Great Recession/pre-COVID-19 testing subperiod, and the entire testing period. Obs is the number of return observations in each period. Mean, Std, Skew, Kurt, Min, and Max are the period mean, standard deviation, skewness, excess kurtosis, minimum, and maximum, respectively.

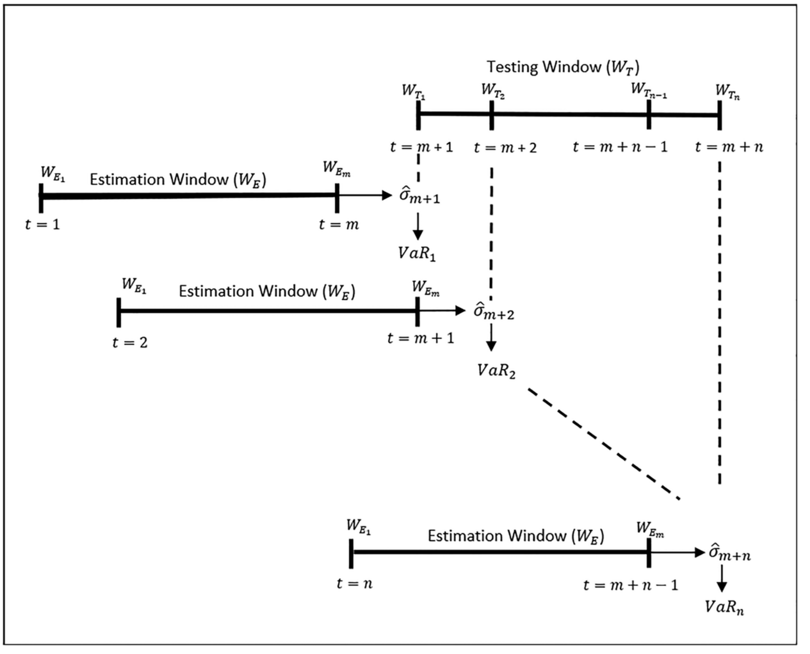

FIGURE 2 VaR calculation process. The figure graphically summarizes the VaR calculation process. W_E denotes a moving estimation window of length *m*. W_T denotes a testing window of length *n*. The entire sample period has a length of $m + n$ trading days. $(\hat{\sigma}_{m+1}, \hat{\sigma}_{m+2}, \ldots, \hat{\sigma}_{m+n})$ denote *n* 1-day-ahead volatilities forecasted by moving W_E throughout W_T. The *n* 1-day-ahead volatilities are subsequently used to calculate *n* daily VaR forecasts, denoted as $(VaR_1, VaR_2, \ldots, VaR_n)$.

commodities, are on average more volatile than others, such as fixed income and currencies.[16] These more volatile asset classes exhibit larger price swings in either direction over a period of time. Besides, it is apparent that the volatility of returns increases during periods of financial crisis. Moreover, the descriptive statistics indicate clear deviations from the normal distribution. More specifically, the positive excess kurtoses indicate strong evidence of leptokurticity of returns, while the skewnesses suggest possible deviations from symmetry of returns.

4.5 | VaR calculation

The logical starting point when calculating daily VaR forecasts is to use an estimation window (W_E) of length *m* to forecast the one-day-ahead volatility ($\hat{\sigma}_{m+1}$). This one-day-ahead volatility is the forecasted volatility of the first trading day in the testing window (W_T). The process is repeated by moving the estimation window throughout the testing window. Assuming the length of the testing window is equal to *n*, then *n* number of one-day-ahead volatilities, denoted as $(\hat{\sigma}_{m+1}, \hat{\sigma}_{m+2}, \ldots, \hat{\sigma}_{m+n})$, are forecasted; that is, one volatility for each trading day within the testing window. Using Equation (3), the *n* forecasted volatilities are subsequently used to calculate *n* number of daily VaR forecasts, denoted as $(VaR_1, VaR_2, \ldots, VaR_n)$. Figure 2 graphically summarizes the VaR calculation process.

In our analysis, we forecast the one-day-ahead volatilities using estimation windows of 100, 250, 500, 1000, and 2000 days, while the daily VaR forecasts are calculated for the regulatory confidence level of 99%. The latter implies

[16] The two currency pairs used in the analysis are not particularly volatile. Other, less stable currency pairs, are likely to exhibit higher volatility.

that there is a 1% chance of having a realized percentage loss that exceeds the VaR forecast on a particular trading day. The VaR calculation process is performed for each portfolio using different risk forecast models (see Section 3).

4.6 | Backtesting and violation ratios

To assess and compare the forecasting performance of the various risk models, we use the procedure of backtesting; see Danielsson (2011) for details. This procedure is similar to the one employed by banks when calculating backtesting multipliers, which are used when determining market risk capital requirements. Backtesting involves the comparison of the daily VaR forecasts, $(VaR_1, VaR_2, \dots, VaR_n)$, to the realized returns, $(r_1, r_2, \dots, r_n)$. When the realized return exceeds the VaR forecast on a particular trading day t, a violation, denoted as $\eta_t = 1$, is recorded. The repeated procedure of comparing ex ante VaR forecasts to ex post returns generates a sequence of violations:

$$\begin{aligned} &\eta := (\eta_1, \eta_2, \dots, \eta_n), \\ &\text{where } \eta_t = \begin{cases} 1 \text{ if } r_t \leq -VaR_t \\ 0 \text{ if } r_t > -VaR_t \end{cases}, \ \forall t = 1, 2, \dots, n. \end{aligned} \tag{9}$$

The sequence of violations, which is simply a sequence of 0's and 1's, is subsequently used to calculate violation ratios. A violation ratio is calculated as the total number of violations divided by the expected number of violations within a testing period. For instance, the violation ratio for the entire testing period is calculated as

$$VR = \frac{\sum \eta}{(1-\alpha) \times n}, \tag{10}$$

where $\sum \eta$ is the total number of violations within the entire testing period, α denotes the confidence level (99% in our analysis), n denotes the length of the entire testing period, and $((1-\alpha) \times n)$ is the expected number of violations within the entire testing period.

In our analysis, we calculate violation ratios for the four testing periods considered; these are, the three testing subperiods of the Great Recession, post-Great Recession/pre-COVID-19, and COVID-19, and the entire testing period. In general, violation ratios close to 1 are considered to be the greatest, reflecting adequate forecasting performance of a risk model, whereas ratios below and above 1 are indicative of over- and under-forecasting performance of a risk model, respectively. Of course, it is important to take into consideration that violation ratios are quite sensitive to the length of the testing period, which makes the ratios not interpretable in an absolute sense but only relative to other ratios. Table 4 presents the best possible violation ratios achievable for each portfolio over the four testing periods. As can be seen, the best possible violation ratios are close to but not exactly equal to 1. Hence, the violation ratios reported in the analysis are used for comparative purposes in order to identify the best performing model among the risk forecast models considered.

5 | EMPIRICAL RESULTS

Table 5 presents the 99% VaR violation ratios for the risk forecast models considered in this paper. Violation ratios are reported for the two volatile testing subperiods of the Great Recession and COVID-19, the stable post-Great Recession/pre-COVID-19 testing subperiod, and the entire testing period. These violation ratios are calculated using Equation (10). Violation ratios in bold indicate the best performing risk forecast model. These violation ratios are the ratios closest to 1. Violation ratios below and above 1 are indicative of over- and under-forecasting performance of a risk model, respectively. An asterisk (*) denotes whether a violation ratio is the best possible violation ratio. Numbers in parentheses refer to the number of singularities encountered during the VaR calculation process. When singularities are encountered, they are removed from the calculation of the violation ratios.

TABLE 4 Best possible violation ratios.

	Great Recession (GR)			Post-GR/pre-COVID-19 (stable period)			COVID-19			Entire period		
	Obs	Under	Over	Obs	Under	Over	Obs	Under	Over	Obs	Under	Over
S&P 500	485	0.82	1.03	2685	0.97	1.01	320	0.94	1.25	3490	0.97	1.00
Nikkei	426	0.94	1.17	2639	0.99	1.02	305	0.98	1.31	3370	0.98	1.01
Hang Seng	459	0.87	1.09	2650	0.98	1.02	303	0.99	1.32	3412	1.00	1.03
Microsoft	502	1.00	1.20	2658	0.98	1.02	322	0.93	1.24	3482	0.98	1.01
U.S. bond	535	0.93	1.12	2638	0.99	1.02	319	0.94	1.25	3492	0.97	1.00
Oil	477	0.84	1.05	2686	0.97	1.01	312	0.96	1.28	3475	0.98	1.01
Gold	422	0.95	1.18	2745	0.98	1.02	317	0.95	1.26	3484	0.98	1.00
GBP/USD	485	0.82	1.03	2813	1.00	1.03	316	0.95	1.27	3614	1.00	1.02
MYR/GBP	483	0.83	1.04	2804	1.00	1.03	324	0.93	1.23	3611	1.00	1.02

Note: The table reports the best possible violation ratios achievable for each portfolio. The best possible violation ratios are provided for the two volatile testing subperiods of the Great Recession and COVID-19, the stable post-Great Recession/pre-COVID-19 testing subperiod, and the entire testing period. Obs is the number of return observations in each period. Under and Over represent the best possible violation ratios below and above 1, respectively. For instance, the best possible violation ratios for S&P 500 for the testing subperiod of the Great Recession are calculated as 4/(0.01×485) and 5/(0.01×485). The violation ratios reported in this table are calculated under the assumption that no singularities are encountered during the VaR calculation process.

5.1 | HS, MA, and EWMA

HS, MA, and EWMA[17] are the simplest among the risk forecast models considered, in the sense that they are computationally easy to implement. It is worth mentioning that HS is the most common risk forecast model preferred in the industry (O'Brien & Szerszen, 2017; Danielsson et al., 2016); it is commonly used by banks regulated by the Basel Accords,[18] while it is perhaps almost exclusively used by nonregulatory financial institutions.

It is clear from the violation ratios that the forecasting performance of the three models is not satisfactory. The violation ratios for portfolios that exhibit relatively high volatility (e.g., S&P 500) are consistently way above 1. This indicates that these simple models underestimate risk for volatile portfolios in times of both financial crisis and stability. The only exception is the violation ratios of HS for the stable post-Great Recession/pre-COVID- 19 subperiod. These ratios are closer to 1, yet they appear to be quite sensitive to the length of the estimation window. The three models seem to have a slightly better forecasting ability for the more stable portfolios (e.g., U.S. Bond) whose returns are not particularly volatile. However, those violation ratios are not consistent across different testing periods and estimation windows.

5.2 | NGARCH and StGARCH

The violation ratios reported in Table 5 are for the most common specifications of the NGARCH(p, q) and StGARCH($p, q; \nu$) models that employ only one lag (see Danielsson, 2011), resulting in the NGARCH(1,1) and StGARCH(1,1;ν) with ν degrees of freedom.

[17] In the analysis, the decay factor, λ, in EWMA is set at 0.94, as suggested by Riskmetrics.

[18] For instance, many large banks, such as Bank of America and J.P. Morgan, calculate trading risk via HS for their annual reports (Danielsson et al., 2016).

TABLE 5 99% VaR violation ratios.

Testing period	W_E	HS	MA	EWMA	NGARCH	StGARCH	t-StAR
Panel A: S&P500							
Great Recession (GR)	100	1.65	4.33	3.30	3.51	0.41 (1)	**1.03***
	250	2.68	7.63	3.30	3.92	0.41	**0.82***
	500	5.36	8.66	3.30	4.33	0.21	**1.03***
	1000	7.63	11.55	3.30	4.54	1.65	**1.24**
	2000	6.39	8.66	3.30	3.71	**1.03***	**0.82***
Post-GR/pre-COVID-19	100	**1.04**	2.46	2.38	2.50	1.08 (11)	0.89
(stable period)	250	0.56	2.42	2.38	2.46	0.67	**1.01***
	500	0.78	2.20	2.38	2.12	0.60	**1.01***
	1000	0.74	1.38	2.38	2.09	0.56	**0.97***
	2000	0.48	**0.97***	2.38	2.12	0.74	0.93
COVID-19	100	1.56	4.06	4.06	2.81	1.56	**1.25***
	250	2.50	4.06	4.06	3.12	**1.25***	**1.25***
	500	3.12	4.06	4.06	3.44	1.56	**0.94***
	1000	3.75	5.62	4.06	3.44	**1.25***	**0.94***
	2000	5.31	6.88	4.06	3.75	1.56	**0.94***
Entire period	100	1.17	2.87	2.66	2.66	1.04 (12)	**0.95**
	250	1.03	3.30	2.66	2.72	0.69	**0.97***
	500	1.63	3.27	2.66	2.55	0.63	**1.00***
	1000	1.98	3.18	2.66	2.55	0.77	**1.03**
	2000	1.75	2.58	2.66	2.49	0.86	**0.92**
Panel B: Nikkei							
Great Recession (GR)	100	1.41	3.52	1.88	2.58	**1.17***	**1.17***
	250	1.64	4.93	1.88	1.64	**0.94***	**0.94***
	500	3.52	6.34	1.88	1.64	**1.17***	**0.94***
	1000	5.16	7.51	1.88	1.88	**1.41**	**1.41**
	2000	5.40	7.75	1.88	1.41	1.41	**0.94***
Post-GR/pre-COVID-19	100	1.10	2.12	2.31	2.39	0.84 (7)	**1.06**
(stable period)	250	0.91	1.93	2.31	2.12	0.61	**0.99***
	500	**1.02***	1.52	2.31	2.01	0.68	0.95
	1000	0.87	1.29	2.31	2.08	0.64	**0.99***
	2000	0.49	**0.99***	2.31	2.08	0.80	0.95
COVID-19	100	1.64	2.62	1.97	2.30	**0.98***	**0.98***
	250	**0.98***	2.62	1.97	3.28	0.66	**0.98***
	500	1.31	2.95	1.97	2.62	0.00	**0.98***
	1000	1.97	3.61	1.97	2.62	0.00	**0.98***
	2000	1.97	2.62	1.97	2.62	0.00	**0.98***

(Continues)

TABLE 5 (Continued)

Panel B: Nikkei							
Entire period	100	1.19	2.34	2.23	2.40	0.89 (7)	**1.07**
	250	**1.01***	2.37	2.23	2.17	0.65	1.04
	500	1.36	2.26	2.23	2.02	0.68	**1.01***
	1000	1.51	2.28	2.23	2.11	0.68	**0.98***
	2000	1.25	1.99	2.23	2.05	0.80	**0.98***
Panel C: Hang Seng							
Great Recession (GR)	100	2.18	3.70	2.18	2.40	**0.87***	**0.87***
	250	2.40	5.23	2.18	2.61	**0.87***	**1.09***
	500	4.58	7.19	2.18	2.40	**0.87***	**0.87***
	1000	6.97	11.33	2.18	2.40	0.44	**1.31**
	2000	6.54	10.68	2.18	2.40	0.65	**0.87***
Post-GR/pre-COVID-19	100	1.09	2.15	2.26	2.45	0.64 (9)	**0.98***
(stable period)	250	0.64	1.92	2.26	2.19	0.57 (13)	**1.06**
	500	**0.98***	1.92	2.26	2.04	0.64	**1.02***
	1000	0.45	1.17	2.26	1.92	0.49	**0.94**
	2000	0.34	0.87	2.26	1.85	0.68	**0.91**
COVID-19	100	2.31	3.36	2.31	1.98	0.66	**0.99***
	250	1.98	2.64	2.31	2.64	0.66	**1.32***
	500	2.31	2.97	2.31	2.64	**0.99***	**1.32***
	1000	2.64	3.30	2.31	2.97	0.66	**0.99***
	2000	2.97	3.30	2.31	2.64	0.66	**0.99***
Entire period	100	1.35	2.49	2.26	2.40	0.68 (9)	**1.03***
	250	**1.00***	2.43	2.26	2.29	0.62 (13)	**1.03***
	500	1.58	2.73	2.26	2.14	0.70	**1.00***
	1000	1.52	2.73	2.26	2.08	0.50	**1.00***
	2000	1.41	2.40	2.26	1.99	0.67	**0.97**
Panel D: Microsoft							
Great Recession (GR)	100	**1.20***	2.39	1.79	2.39	**1.20***	**1.00***
	250	1.59	3.78	1.79	2.79	**1.00***	**1.00***
	500	3.59	5.58	1.79	2.79	**1.00***	**1.00***
	1000	5.38	6.97	1.79	3.19	0.60	**1.00***
	2000	2.59	4.58	1.79	2.19	0.80	**1.20***
Post-GR/pre-COVID -19	100	1.39	1.69	1.73	1.69	0.60	**0.98***
(stable period)	250	0.79	1.47	1.73	1.54	0.45	**1.02***
	500	0.83	1.58	1.73	1.43	0.26	**0.98***
	1000	0.75	1.17	1.73	1.32	0.23	**0.94**
	2000	0.45	0.75	1.73	1.35	0.19	**0.90**

(Continues)

TABLE 5 (Continued)

Panel D: Microsoft							
COVID-19	100	2.17	3.42	2.48	1.86	0.62	**1.24***
	250	1.86	2.80	2.48	2.17	0.62	**1.24***
	500	2.17	3.11	2.48	2.48	**0.93***	**1.24***
	1000	3.11	4.97	2.48	2.80	0.31	**0.93***
	2000	4.35	4.97	2.48	3.11	0.31	**1.24***
Entire period	100	1.44	1.95	1.81	1.81	0.69	**1.01***
	250	**1.01***	1.92	1.81	1.78	0.55	**1.01***
	500	1.35	2.30	1.81	1.72	0.43	**0.98***
	1000	1.64	2.35	1.81	1.72	0.29	**0.95**
	2000	1.12	1.69	1.81	1.64	0.29	**1.01***
Panel E: U.S. Bond							
Great Recession (GR)	100	**1.12***	**1.12***	**1.12***	0.75	0.29 (196)	0.75
	250	1.87	2.06	**1.12***	**1.12***	0.37 (266)	**0.93***
	500	3.36	4.30	**1.12***	1.50	0.72 (397)	**1.12***
	1000	4.67	5.79	**1.12***	**0.93***	0.00 (535)	**1.12***
	2000	2.06	3.74	**1.12***	0.56	0.00 (535)	**0.93***
Post-GR/pre-COVID-19	100	**1.10**	1.44	1.25	1.33 (4)	0.26 (716)	0.76
(stable period)	250	**0.91**	1.25	1.25	1.22 (5)	0.33 (1126)	0.72
	500	**0.91**	1.29	1.25	1.06 (1)	0.00 (1528)	0.68
	1000	0.80	1.14	1.25	**1.02***	0.00 (2521)	0.61
	2000	0.53	0.76	1.25	**1.02***	0.00 (2638)	0.57
COVID-19	100	**1.25***	2.51	2.51	2.19	1.00 (19)	**0.94***
	250	**0.94***	**0.94***	2.51	1.88	0.97 (9)	**0.94***
	500	1.57	1.57	2.51	1.88	1.03 (27)	**1.25***
	1000	2.19	1.57	2.51	2.19	0.00 (285)	**0.94***
	2000	3.45	3.45	2.51	2.51	0.00 (319)	**0.94***
Entire period	100	**1.12**	1.49	1.35	1.32 (4)	0.35 (931)	0.66
	250	**1.06**	1.35	1.35	1.26 (5)	0.43 (1401)	0.80
	500	1.35	1.78	1.35	1.20 (1)	0.26 (1952)	**0.83**
	1000	1.52	1.89	1.35	**1.12**	0.00 (3341)	0.83
	2000	**1.03**	1.46	1.35	1.09	0.00 (3492)	0.60
Panel F: Oil							
Great Recession (GR)	100	**1.05***	1.26	0.63	**0.84***	0.42	**1.05***
	250	1.68	3.56	0.63	1.26	0.42	**1.05***
	500	3.98	5.87	0.63	**0.84***	0.00	**0.84***
	1000	6.29	6.92	0.63	**1.05***	0.21	**0.84***
	2000	5.45	6.08	0.63	**1.05***	0.21	**0.84***

(Continues)

TABLE 5 (Continued)

Panel F: Oil							
Post-GR/pre-COVID-19	100	0.86	2.23	2.23	2.08	0.67	**1.04**
(stable period)	250	0.86	2.08	2.23	1.94	0.78	**1.01***
	500	1.27	2.35	2.23	1.71	0.86	**1.01***
	1000	0.86	1.79	2.23	1.71	0.86	**0.97***
	2000	0.48	1.30	2.23	1.82	0.78	**0.93**
COVID -19	100	**1.28***	2.24	1.92	2.88	**1.28***	**1.28***
	250	**1.28***	2.24	1.92	1.28	**0.96***	**1.28***
	500	2.56	4.17	1.92	1.60	**0.96***	**1.28***
	1000	4.49	4.81	1.92	1.92	**1.28***	**0.96***
	2000	5.13	4.81	1.92	12.90 (281)	1.92	**1.28***
Entire period	100	0.92	2.10	1.99	1.99	0.69	**1.01***
	250	**1.01***	2.30	1.99	1.78	0.75	0.95
	500	1.76	2.99	1.99	1.58	0.75	**1.01***
	1000	1.93	2.76	1.99	1.64	0.81	**0.98***
	2000	1.58	2.27	1.99	1.82 (281)	0.81	**1.01***
Panel G: Gold							
Great Recession (GR)	100	1.42	2.37	1.90	2.37	**0.95***	**1.18***
	250	1.66	3.79	1.90	1.90	**1.18***	**1.18***
	500	2.13	3.55	1.90	1.66	**0.95***	**1.18***
	1000	2.61	4.74	1.90	1.90	**0.95***	**1.18***
	2000	3.08	6.64	1.90	1.90	0.71	**1.18***
Post-GR/pre-COVID-19	100	0.91	1.97	2.15	2.22	0.59 (16)	**1.02***
(stable period)	250	0.73	1.93	2.15	1.93	0.59 (23)	**1.06**
	500	0.84	1.60	2.15	1.86 (1)	0.51 (17)	**1.02***
	1000	0.51	0.84	2.15	1.79	0.47	**0.95**
	2000	0.44	0.87	2.15	1.46	0.51	**0.91**
COVID-19	100	**1.26***	2.52	2.21	1.89	1.58	**1.26***
	250	**1.26***	2.21	2.21	2.21	**0.95***	**0.95***
	500	2.84	3.79	2.21	3.15	**0.95***	**1.26***
	1000	3.79	5.36	2.21	3.47	**1.26***	**0.95***
	2000	2.84	4.10	2.21	2.52	**1.26***	**1.26***
Entire period	100	**1.00***	2.07	2.12	2.21	0.72 (16)	0.95
	250	0.89	2.18	2.12	1.95	0.69 (23)	**0.98***
	500	1.18	2.04	2.12	1.95 (1)	0.61 (17)	**1.00***
	1000	1.06	1.72	2.12	1.95	0.60	**1.03**
	2000	**0.98***	1.87	2.12	1.61	0.60	**1.00***

(Continues)

TABLE 5 (Continued)

Panel H: GBP/USD							
Great Recession (GR)	100	2.27	4.12	2.68	2.68	0.66 (31)	**0.82***
	250	2.06	5.57	2.68	2.47	0.67 (35)	**0.82***
	500	3.71	7.42	2.68	2.47	0.73 (76)	**0.82***
	1000	4.95	6.60	2.68	2.47	1.25 (5)	**1.03***
	2000	5.77	6.60	2.68	2.47	1.24	**1.03***
Post-GR/pre-COVID-19 (stable period)	100	**1.03***	1.64	1.81	1.56	0.50 (203)	**1.03***
	250	0.78	1.49	1.81	1.49	0.48 (119)	**1.00***
	500	**0.96**	1.24	1.81	1.56	0.52 (111)	0.92
	1000	0.92	1.39	1.81	1.64	0.38 (184)	**0.96**
	2000	0.60	**1.07**	1.81	1.64	0.52 (112)	0.92
COVID-19	100	**0.95***	**1.27***	**1.27***	**1.27***	0.32 (5)	**0.95***
	250	0.32	0.63	**1.27***	0.63	0.00 (3)	0.63
	500	0.63	**0.95***	**1.27***	**0.95***	0.32	**0.95***
	1000	**0.95***	**1.27***	**1.27***	**1.27***	0.32	**0.95***
	2000	**0.95***	**1.27***	**1.27***	0.63	0.32	**0.95***
Entire period	100	1.19	1.94	1.88	1.69	0.50 (239)	**0.94**
	250	0.91	1.96	1.88	1.55	0.46 (157)	**1.02***
	500	1.30	2.05	1.88	1.63	0.53 (187)	**1.02***
	1000	1.47	2.08	1.88	1.72	0.50 (189)	**1.05**
	2000	1.33	1.83	1.88	1.66	0.60 (112)	**1.00***
Panel I: MYR/GBP							
Great Recession (GR)	100	1.45	**1.04***	**0.83***	**0.83***	0.24 (58)	**0.83***
	250	2.07	1.66	**0.83***	**0.83***	0.49 (74)	**0.83***
	500	3.52	3.93	**0.83***	**1.04***	0.21 (2)	**1.04***
	1000	3.52	4.14	**0.83***	**0.83***	0.41	**1.04***
	2000	4.14	4.76	**0.83***	**0.83***	0.41	**0.83***
Post-GR/pre-COVID-19 (stable period)	100	**1.00***	1.32	1.50	1.53	0.55 (52)	**1.00***
	250	**1.03***	1.60	1.50	1.28	0.36 (14)	**1.03***
	500	**0.96**	1.32	1.50	1.07	0.36	0.89
	1000	**1.00***	1.50	1.50	1.39	0.39	0.96
	2000	**0.96**	1.36	1.50	1.50	0.46	0.93
COVID-19	100	**1.23***	0.62	**0.93***	1.85	0.64 (10)	0.62
	250	0.62	**1.23***	**0.93***	**1.23***	0.00	**0.93***
	500	**0.93***	**1.23***	**0.93***	**1.23***	0.00	**0.93***
	1000	0.31	**0.93***	**0.93***	**0.93***	0.00	0.62
	2000	0.31	**0.93***	**0.93***	0.62	0.00	0.62

(Continues)

TABLE 5 (Continued)

Panel I: MYR/GBP							
Entire period	100	1.08	1.22	1.36	1.47	0.52 (120)	**0.97**
	250	1.14	1.58	1.36	1.22	0.34 (88)	**1.00***
	500	1.30	1.66	1.36	1.08	0.30 (2)	**1.00***
	1000	1.27	1.80	1.36	1.27	0.36	**0.97**
	2000	1.33	1.77	1.36	1.33	0.42	**1.00***

Note: The table presents the 99% VaR violation ratios for the risk forecast models considered in this paper. Panels A–I of the table report the violation ratios for each of the nine portfolios employed in this paper. In each panel, violation ratios are independently reported for the two volatile testing subperiods of the Great Recession and COVID-19, the stable post-Great Recession/pre-COVID-19 testing subperiod, and the entire testing period. Furthermore, violation ratios are individually reported for estimation windows (W_E) of 100, 250, 500, 1000, and 2000 days. Each column of each panel presents violation ratios for each of the risk forecast models considered. These models are the Historical Simulation (HS), Moving Average (MA), Exponentially Weighted Moving Average (EWMA), Normal Generalized Autoregressive Heteroskedastic (NGARCH), Student's *t* GARCH (StGARCH), and the proposed time-heterogeneous Student's *t* Autoregressive (t-StAR). The reported 99% VaR violation ratios are calculated using Equation (10). Violation ratios in bold indicate the best performing risk forecast model. These violation ratios are the ratios closest to 1. Violation ratios below and above 1 are indicative of over- and under-forecasting performance of a risk model, respectively. An asterisk ($*$) denotes whether a violation ratio is the best possible violation ratio. Numbers in parentheses refer to the number of singularities encountered during the VaR calculation process. When singularities are encountered, they are removed from the calculation of the violation ratios.

The violation ratios of the two GARCH models are fairly different. First, the ratios of NGARCH are comparable to the ones of HS, MA, and EWMA. Specifically, the NGARCH model consistently underestimates risk for volatile portfolios, whereas its forecasting performance appears to be slightly better for stable portfolios. In contrast, the violation ratios of the StGARCH model are mainly below 1. This indicates that StGARCH consistently overestimates risk. The only exception is the relatively good performance of the model for extremely volatile situations, that is, when risk is forecasted for volatile portfolios across volatile periods.

One concern regarding these GARCH models, especially StGARCH, is the occurrence of singularities due to the parameter restrictions imposed to ensure positive volatility. Tables A1 and A2 in Appendix A present the violation ratios for estimations of the NGARCH(p, q) and StGARCH($p, q; \nu$) models containing different lag polynomial degrees. As can be seen from this table, the violation ratios do not seem to improve when the models contain additional lags, while it is particularly noteworthy that the number of singularities encountered during the estimation process increases with the number of lags.

5.3 | t-StAR

For consistency purposes, the violation ratios reported in Table 5 are for the t-StAR($p; \nu$) model that employs only one lag, resulting in the t-StAR($1; \nu$) with ν degrees of freedom. The violation ratios of t-StAR indicate clearly that this model performs well for all portfolios and testing periods, whether volatile or not. Moreover, the ratios are not sensitive to the length of the estimation window, while no singularities were encountered during the estimation process.

Table A3 in Appendix A presents the violation ratios for estimations of the t-StAR($p; \nu$) model containing different lag polynomial degrees while keeping the degrees of freedom parameter constant. Similar to the NGARCH(p, q) and StGARCH($p, q; \nu$) models, the violation ratios are not particularly sensitive to the number of lags. The interesting observation here is the direct relationship between degrees of freedom and violation ratios. In times of financial crisis (i.e., times of high volatility), the best violation ratios are obtained for the lower degrees of freedom in the model. Conversely, in times of financial stability (i.e., times of low volatility), the best violation ratios are obtained for the

TABLE 6 Model comparison.

	Cases	HS	MA	EWMA	NGARCH	StGARCH	t-StAR
Volatile portfolios	90	10 (9)	2 (2)	0	4 (4)	28 (27)	86 (72)
Stable portfolios	45	18 (12)	12 (11)	20 (20)	15 (15)	0	30 (29)
Volatile periods	90	15 (15)	11 (11)	20 (20)	17 (17)	28 (27)	85 (82)
Stable periods	45	13 (6)	3 (2)	0	2 (2)	0	31 (19)
Volatile × Volatile	60	7 (7)	0	0	4 (4)	28 (27)	60 (57)
Volatile × Stable	30	3 (2)	2 (2)	0	0	0	26 (15)
Stable × Volatile	30	8 (8)	11 (11)	20 (20)	13 (13)	0	25 (25)
Stable × Stable	15	10 (4)	1 (0)	0	2 (2)	0	5 (4)
Volatile and stable	135	28 (21)	14 (13)	20 (20)	19 (19)	28 (27)	116 (101)
Entire period	45	9 (6)	0	0	1 (0)	0	38 (26)
All	180	37 (27)	14 (13)	20 (20)	20 (19)	28 (27)	154 (127)

Note: The table reports the number of times each risk forecast model is selected as the best performing model. Numbers in parentheses refer to the number of times the best possible violation ratios are achieved. Cases is the total number of model evaluations (portfolios × testing periods × estimation windows). The groups are as follows. Volatile portfolios: S&P 500, Nikkei, Hang Seng, Microsoft, Oil, Gold; Stable portfolios: U.S. Bond, GBP/USD, MYR/GBP; Volatile periods: Great Recession, COVID-19; Stable periods: post-Great Recession/pre-COVID-19. The interaction groups are groups of interacting cases within volatile/stable portfolios and volatile/stable periods. The entire testing period is not included in the grouping; it is reported separately and included in the "all" cases combined group. The sum of the reported numbers may exceed the number of cases because more than one models may simultaneously be selected as the best performing model.

higher degrees of freedom in the model. The latter is not as surprising because the area in the tails of the Student's t distribution gets smaller with increasing degrees of freedom.

5.4 | Model comparison

Table 6 reports the number of times each risk forecast model is selected as the best performing model. Numbers in parentheses refer to the number of times the best possible violation ratios are achieved. The numbers are independently reported for volatile and stable portfolios, volatile and stable periods, interactions within volatile/stable portfolios and volatile/stable periods, the entire testing period, and all cases combined.

The model evaluation results indicate clearly that the proposed t-StAR model has the best forecasting performance among the models considered. The model seems to provide accurate risk forecasts across different portfolios, testing periods, and estimation windows. Specifically, the model provided more accurate risk forecasts than the rest of the models in more than 85% of the times, while of those only in less than 20% of the times it did not achieve the best possible violation ratios.

The model evaluation numbers in Table 6 also show clearly that models with simple probabilistic structure and models that assume the normal distribution cannot be considered reliable. The forecasting performance of those models is in general poor, as well as inconsistent and unstable across different portfolios, testing periods, and estimation windows. Such models seem to provide comparatively better accurate risk forecasts for stable portfolios that exhibit low volatility of returns, yet even for those portfolios their performance is not good enough. The forecasting performance of t-StAR is at least four times as good as that of models with simple probabilistic structure and models that assume the normal distribution. Hence, such models should be avoided in practice as they tend to consistently underestimate risk.

An interesting observation is that HS has the best performance (better than t-StAR) for extremely stable situations, that is, when risk is forecasted for stable portfolios across stable periods. This result indicates that the main problem with HS is its slow adjustment to volatility movements (O'Brien & Szerszen, 2017). Hence, even though HS does not seem to be a good model for forecasting risk of portfolios that exhibit high volatility, it appears to be a very good option for forecasting risk of stable portfolios.

The StGARCH model is the only one among the commonly used models considered that appears to forecast risk relatively well for portfolios that exhibit high volatility of returns. However, the model consistently overestimates risk for stable portfolios, as well as for volatile portfolios evaluated across stable periods. Actually, StGARCH has the worst forecasting performance among all models for such portfolios. Overall, the forecasting performance of StGARCH is at least five times worse than that of t-StAR.

By comparing the results of StGARCH to the ones of the proposed t-StAR model, it is apparent that the latter has at least two crucial advantages over the former. First, as argued by Heracleous (2007), the StGARCH model provides biased and inconsistent estimates of the degrees of freedom. As observed during the estimation process, the estimated degrees of freedom parameter in StGARCH remains relatively constant across portfolios and periods. This may explain why the performance of the model differs significantly between portfolios and periods. When adopting the Student's *t* distribution, it is important to choose the most appropriate degrees of freedom. In this regard, the proposed t-StAR model seems to reach a higher standard than StGARCH. Second, the conditional variance of t-StAR does not rely on any parameter restrictions, thus naturally avoids singularity problems. This is a clear advantage over StGARCH, which commonly encounters singularities.

6 | CONCLUSIONS AND RECOMMENDATIONS

6.1 | Concluding remarks

From a theoretical perspective, ES has been generally viewed as being superior to VaR. This theoretical superiority has substantially affected the decision of BCBS to replace VaR with ES. However, the ability of ES to capture tail risk and its coherence neither guarantees accuracy of risk forecasts nor adequacy of risk capital requirements. This is because the process of risk forecasting relies heavily on the validity of the statistical assumptions underlying the internal risk forecast models used by banks.

In general, a risk model is expected to have good forecasting performance when its actual error probabilities approximate closely the nominal ones. One may think of the violation ratio in Equation (10) as a post-comparison of actual and nominal error probabilities, where the former is directly related to the actual number of violations and the latter to the expected number of violations. In order for the actual and expected number of violations to approximate closely each other, the actual error probabilities must approximate closely the nominal ones at the outset. This approximation is solely related to the statistical assumptions underlying the models used to forecast risk, and has very little to do with the choice of the risk measure for computing risk.

The results in this paper suggest that an appropriate choice of risk models can considerably improve risk forecasts, even in times of financial crisis. The model proposed in this paper seems to have sufficient forecasting performance, primarily because its rich probabilistic structure can describe relevant features of financial data that are not accounted for by other commonly used models. This indicates that it is of crucial importance to formally assess the validity of the statistical assumptions underlying the risk forecast models used by banks and other financial institutions on a regular basis.

6.2 | Recommendations to regulators

Recent reports of the BCBS state that banks must "ensure that their internal models have been adequately validated" when "initially developed and when any significant changes are made to the model[s]." In addition, the reports state that "models must be periodically revalidated, particularly when there have been significant structural changes in the market or changes to the composition of the portfolio which might lead to the model no longer being adequate."

It is emphasized that model validation must include "tests to demonstrate that any assumptions made within the internal model are appropriate and do not underestimate risk." These reports, however, do not provide any specific information regarding the assumptions to be tested and the testing approaches to be utilized. The only explicit recommendation the Committee has made is to suggest that model validation "may include the assumption of the normal distribution" ((Basel Committee on Banking Supervision, 2016, p. 55); see also (Basel Committee on Banking Supervision, 2019a, p. 69)).

Our regulation recommendation is that BCBS should address model risk separately from the choice of risk measures because the two are distinct sources of risk. Therefore, we recommend the Committee consider providing clearer guidelines for model validation procedures. These guidelines may include the following:

- potential alternatives to the normal distribution;
- key nondistributional statistical assumptions to be examined;
- formal and graphical test procedures for distributional and nondistributional assumptions;
- diagnostic tests for estimating and determining appropriate parameters; and
- tests for structural breaks.[19]

Beyond providing clear guidelines for model validation, we recommend that BCBS consider restricting the scope of internal modeling to risk forecast models, which are likely to pass some kind of statistical adequacy test. The introduction of a formal statistical adequacy test to be monitored by regulatory agencies can serve as an important tool in determining the most appropriate risk forecast model for any given situation.

It is important to emphasize that the proposed recommendations to regulators are neither intended nor expected to universally "penalize" banks by increasing their capital requirements; they are rather intended to help better determine the *appropriate* level of capital for each bank. During times of financial stability, the proposed recommendations are expected to significantly increase capital requirements only for banks holding riskier positions than are natural for them. An increase in their capital requirements can provide many benefits for those banks, including enhancing their protection against risk, increasing their probability of survival, and improving their efficiency, profitability, and performance (Berger & Bouwman, 2013; Bitar et al., 2016). Moreover, the proposed recommendations are not only expected to benefit individual banks themselves, but also the entire financial system by helping to reduce the overall systemic risk (Laeven et al., 2016), which often contributes to the probability of failure of banks and poses a threat to the entire financial system.

Furthermore, during periods of financial crisis, the proposed recommendations are expected to help banks determine their appropriate levels of capital earlier than is currently possible. This is because most banks currently use the HS method, which is very slow to adjust to the crisis period conditions (O'Brien & Szerszen, 2017). This leads to the underestimation of risk in the early stages of a financial crisis, consequently increasing overall systemic risk. Hence, during periods of financial crisis, the proposed recommendations can help to mitigate systemic risk, stabilize the financial system earlier in order to avoid deep recessions, and offer better prospects for bank efficiency, profitability, and performance improvements.

6.3 | Recommendations to practitioners

The results in the paper apply equally to financial institutions not regulated by the Basel Accords. Therefore, the aforementioned recommendations to regulators can be extended to the average practitioner who may or may not have access to sophisticated expertise, such as model development and validation. First, financial institutions with access to

[19] Because we do not wish to single out any particular assumption or test, we do not provide specific references herein. We refer the interested reader to the voluminous econometrics and statistics literatures on misspecification testing.

sophisticated expertise should pay more attention to model risk, including enhancing practices around model review and validation. Second, financial institutions with moderate to limited access to sophisticated expertise should develop internal rules of thumb to help them identifying the most appropriate risk forecast model to be used for a particular situation. For example, one rule of thumb might be using the HS method during periods of financial stability, but during periods of financial crisis replacing HS with a richer risk forecast model. One such model is the one proposed in this paper (R code is provided in Appendix B). Third, financial institutions with limited to no access to sophisticated expertise may use universally the HS method. However, when considered necessary in terms of risk and during periods of financial crisis, it would be wise to multiply potential risk by a factor greater than 1 before using risk forecasts for decision-making in order to account for the fact that risk is being underestimated. These factors should be chosen on the basis of specific risks and market conditions. All in all, these recommendations are expected to improve the decision-making process in risk management.

ACKNOWLEDGMENTS

We thank the editors and anonymous referees for their valuable comments and suggestions to improve our paper.

REFERENCES

Ahmed, M., Haider, G., & Zaman, A. (2017). Detecting structural change with heteroskedasticity. *Communications in Statistics - Theory and Methods*, *46*(21), 10446–10455.

Alexander, C., & Sarabia, J. M. (2012). Quantile uncertainty and Value-at-Risk model risk. *Risk Analysis*, *32*(8), 1293–1308.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1997). Thinking coherently. *Risk*, *10*(11), 68–71.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228.

Basel Committee on Banking Supervision. (1988). *International convergence of capital measurement and capital standards*. Bank for International Settlements. Retrieved from https://www.bis.org/publ/bcbs04a.pdf

Basel Committee on Banking Supervision. (1996). *Amendment to the capital accord to incorporate market risks*. Bank for International Settlements. Retrieved from https://www.bis.org/publ/bcbs24.pdf

Basel Committee on Banking Supervision. (2004). *International convergence of capital measurement and capital standards: A revised framework*. Bank for International Settlements. Retrieved from https://bis.org/publ/bcbs107.pdf

Basel Committee on Banking Supervision. (2011). *Messages from the academic literature on risk management for the trading book* [Working paper no. 19]. Bank for International Settlements. Retrieved from http://bis.org/publ/bcbs_wp19.pdf

Basel Committee on Banking Supervision. (2012). *Consulative document: Fundamental review of the trading book*. Bank for International Settlements. Retrieved from http://bis.org/publ/bcbs219.pdf

Basel Committee on Banking Supervision. (2013). *Consulative document: Fundamental review of the trading book: A revised market risk framework 5*. Bank for International Settlements. Retrieved from http://bis.org/publ/bcbs265.pdf

Basel Committee on Banking Supervision. (2016). *Minimum capital requirements for market risk*. Bank for International Settlements. Retrieved from https://www.bis.org/bcbs/publ/d352.pdf

Basel Committee on Banking Supervision. (2019a). *Minimum capital requirements for market risk*. Bank for International Settlements. Retrieved from https://www.bis.org/bcbs/publ/d457.pdf

Basel Committee on Banking Supervision. (2019b). *The market risk framework: In brief*. Bank for International Settlements. Retrieved from https://www.bis.org/bcbs/publ/d457_inbrief.pdf

Beder, T. S. (1995). VAR: Seductive but dangerous. *Financial Analysts Journal*, *51*(5), 12–24.

Berger, A. N., & Bouwman, C. H. (2013). How does capital affect bank performance during financial crises? *Journal of Financial Economics*, *109*(1), 146–176.

Bitar, M., Saad, W., & Benlemlih, M. (2016). Bank risk and performance in the MENA region: The importance of capital requirements. *Economic Systems*, *40*(3), 398–421.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, *69*(3), 542–547.

Boucher, C. M., Danielsson, J., Kouontchou, P. S., & Maillet, B. B. (2014). Risk models-at-risk. *Journal of Banking and Finance*, *44*, 72–92.

Danielsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking and Finance*, *26*(7), 1273–1296.

Danielsson, J. (2011). *Financial risk forecasting: The theory and practice of forecasting market risk with implementation in R and Matlab*. John Wiley and Sons Ltd.

Danielsson, J., James, K. R., Valenzuela, M., & Zer, I. (2016). Model risk of risk models. *Journal of Financial Stability*, *23*, 79–91.

Danielsson, J., Jorgensen, B. N., Samorodnitsky, G., Sarma, M., & de Vries, C. G. (2013). Fat tails, VaR and subadditivity. *Journal of Econometrics*, *172*(2), 283–291.

Danielsson, J., & Zhou, C. (2015). *Why risk is so hard to measure* [Discussion paper no 36]. LSE Systemic Risk Centre.

Emmer, S., Kratz, M., & Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, *18*(2), 31–60.
Fang, K.-T., Kotz, S., & Ng, K.-W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall.
Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762.
Hendricks, D. (1996). Evaluation of Value-at-Risk models using historical data. *Economic Policy Review*, *2*(1), 39–70.
Hendry, D. F., & Richard, J. F. (1983). The econometric analysis of economic time series. *International Statistical Review*, *51*(2), 111–148.
Heracleous, M. (2007). *Sample kurtosis, GARCH-t and the degrees of freedom issue* [Working paper ECO 2007/60]. European University Institute.
Heracleous, M., & Spanos, A. (2006). The Student's *t* dynamic linear regression: Re-examining volatility modeling. *Advances in Econometrics*, *20*, 289–319.
Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhya: The Indian Journal of Statistics, Series A*, *32*(4), 419–430.
Kellner, R., & Rösch, D. (2016). Quantifying market risk with Value-at-Risk or Expected Shortfall? – Consequences for capital requirements and model risk. *Journal of Economic Dynamics and Control*, *68*, 45–63.
Laeven, L., Ratnovski, L., & Tong, H. (2016). Bank size, capital, and systemic risk: Some international evidence. *Journal of Banking and Finance*, *69*, S25–S34.
Maasoumi, E., Zaman, A., & Ahmed, M. (2010). Tests for structural change, aggregation, and homogeneity. *Economic Modelling*, *27*(6), 1382–1391.
Mandelbrot, B. (1963). The variation of certain speculative returns. *Journal of Business*, *36*(4), 394–419.
McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative risk management*. Princeton University Press.
Michaelides, M., & Spanos, A. (2020). On modeling heterogeneity in linear models using trend polynomials. *Economic Modelling*, *85*, 74–86.
O'Brien, J. M., & Szerszen, P. J. (2017). An evaluation of bank VaR measures for market risk during and before the financial crisis. *Journal of Banking and Finance*, *80*, 215–234.
Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, *55*(290), 324–330.
Repullo, R., & Suarez, J. (2013). The procyclical effects of bank capital regulation. *Review of Financial Studies*, *26*(2), 452–490.
Spanos, A. (1994). On modeling heteroskedasticity: The Student's *t* and elliptical linear regression models. *Econometric Theory*, *10*(2), 286–315.
Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, *41*(3), 788–829.
Yamai, Y., & Yoshiba, T. (2005). Value-at-Risk versus expected shortfall: A practical perspective. *Journal of Banking and Finance*, *29*(4), 997–1015.
Ziegel, J. F. (2016). Coherence and elicitability. *Mathematical Finance*, *26*(4), 901–918.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Michaelides, M., & Poudyal, N. (2023). Good risk measures, bad statistical assumptions, ugly risk forecasts. *Financial Review*, 1–25. https://doi.org/10.1111/fire.12368