International Journal of Forecasting 40 (2024) 1275–1301

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting day-ahead expected shortfall on the EUR/USD exchange rate: The (I)relevance of implied volatility[☆]

Štefan Lyócsa[a,b,c], Tomáš Plíhal[a,*], Tomáš Výrost[a,d,e]

[a] Institute of Financial Complex Systems, Masaryk University, Lipova 41a, 602 00 Brno, Czech Republic
[b] Faculty of Management, University of Presov, Konstantinova 16, 080 01 Presov, Slovakia
[c] Slovak Academy of Sciences, Sancova 56, 811 05 Bratislava, Slovakia
[d] Faculty of Commerce, University of Economics, Donozemska cesta 1, 852 35 Bratislava, Slovakia
[e] Faculty of Economics, Technical university of Košice, Němcovej 32, 040 01 Košice, Slovakia

ARTICLE INFO

Keywords:
Finance
Expected shortfall
Implied volatility
Volatility models
Forecasting

ABSTRACT

The existing literature provides mixed results on the usefulness of implied volatility for managing risky assets, while evidence for expected shortfall predictions is almost nonexistent. Given its forward-looking nature, implied volatility might be more valuable than backward-looking measures of realized price fluctuations. Conversely, the volatility risk premium embedded in implied volatility leads to overestimating the observed price variation. This paper explores the benefits of augmenting econometric models used in forecasting the expected shortfall, a risk measured endorsed in the Basel III Accord, with information on implied volatility obtained from EUR/USD option contracts. The day-ahead forecasts are obtained from several classes of econometric models: historical simulation, EGARCH, quantile regression-based HAR, joint VaR and ES model, and combination forecasts. We verify whether the resulting expected shortfall forecasts are well-specified and test the models' accuracy. Our results provide evidence that the information provided by forward-looking implied volatility is more valuable than that in backward-looking realized measures. These results hold across multiple model specifications, are stable over time, hold under alternative loss functions, and are more pronounced during periods of higher market uncertainty when risk modeling matters most.

1. Introduction

Trade in assets offers owners the possibility of obtaining uncertain future payoffs—in the case of currencies, in the form of capital gains or interest payments. This uncertainty manifests in specific future return distributions of such assets, where the properties of such distributions are a subject of interest in market risk modeling. In the context of the Basel II Accord, one of the most popular measures of the return distribution has become value at risk (VaR) (Basel Committee on Banking Supervision, 1996), complemented by the expected shortfall (ES), which addresses the limitations of VaR and was endorsed in the subsequent Basel III Accord (Basel Committee on Banking Supervision, 2010, 2019). Today, VaR and ES are widely employed to measure financial market risk for regulatory and internal risk management purposes.

In this study, we are primarily interested in predicting and backtesting the expected shortfall of the next-day EUR/USD exchange rate. In a recent study, Hoga (2021) derived asymptotically valid extreme value theory-enhanced risk forecasts for APARCH-X models. And in

☆ This research was supported by the Czech Science Foundation (GACR), nr. 22-27075S.

* Corresponding author.
E-mail addresses: stefan.lyocsa@econ.muni.cz (Š. Lyócsa), tomas.plihal@econ.muni.cz (T. Plíhal), tomas.vyrost@econ.muni.cz (T. Výrost).

https://doi.org/10.1016/j.ijforecast.2023.11.003
0169-2070/

an empirical setting, several risk measures, including ES, were predicted with implied and realized measures (median realized variance) using data from the US equity market. The results showed higher accuracy for implied volatility indices. The novelty of our research lies not only in our exploration of the role of implied volatility for forecasting expected shortfall on the foreign exchange market but also in providing evidence across multiple classes of forecasting models and using directly traded implied volatilities with different maturities. Note that the (derivative) trading on the foreign exchange market is quite different from equity markets with a decentralized market structure, 24-hour trading, and high liquidity. The inherent relationship between VaR and ES, as well as ES properties, leads to a joint estimation and evaluation of forecasting accuracy of VaR and ES.

Implied volatility represents the volatility of the underlying security that is implicit in an option's market price according to a particular model. In other words, the implied volatility is the level of volatility in the Black–Scholes formula (for its modification for currencies, see e.g. Garman & Kohlhagen, 1983) that leads to a calculated option price equal to the current market price. In contrast to historical volatility, implied volatility reflects the market's expectations regarding future price fluctuations. This assumption holds under a lognormal distribution model with risk-neutral market representative agents. However, with risk-averse representative agents and non-Gaussian returns, the link between implied volatility (option price) and expected volatility is less straightforward. The major drivers of implied volatility are investor preferences and demand, whose impact on prices could be time varying. For example, Andersen, Fusari, and Todorov (2020) found that the risk premium associated with adverse tail events displays persistent shifts unrelated to volatility. This risk premium is contained in implied volatility, potentially creating an upward bias compared to realized volatility. Moreover, implied volatility is a risk-neutral measure. Using it to forecast a physical measure could create time-varying estimation errors that tend to be higher in periods of increased risk aversion.

However, despite these theoretical drawbacks, the importance of time-varying estimation errors is unclear and does not indicate that implied volatility is useless. On the contrary, there has been a growing trend toward using implied volatility as a critical variable in financial investment decisions, risk management, derivative pricing, market making, market timing, and portfolio selection. Many studies evaluate the usefulness of the information provided by implied volatility (IV) compared to historical realized volatility (RV), usually in the context of forecasting future realized volatility. The majority of studies have focused on stock markets, finding that IV either subsumes the information content of past volatility (Christensen & Prabhala, 1998; Poon, Blair, & Taylor, 2001) or at least provides additional information on future levels of volatility beyond what is captured in RV (Blair, Poon, & Taylor, 2001; Busch, Christensen, & Nielsen, 2011; Christensen & Prabhala, 1998; Chun, Cho, & Ryu, 2019; Day & Lewis, 1992; Han & Park, 2013; Kambouroudis, McMillan, & Tsakou, 2016; Kourtis, Markellos, & Symeonidis, 2016; Liang, Wei, & Zhang, 2020; Pati, Barai, & Rajib, 2018; Taylor, Yadav, & Zhang, 2010; Wang & Wang, 2016; Wayne, Lui, & Wang, 2010). Only a few studies have failed to achieve comparable results and have emphasized RV's superiority over IV (Becker & Clements, 2008; Becker, Clements, & White, 2006, 2007; Bentes, 2015; Canina & Figlewski, 1993).

Regarding foreign exchange markets (FX), the literature suggests that IV provides additional information or even dominates RV in terms of volatility forecasting (Busch et al., 2011; Charoenwong, Jenwittayaroje, & Low, 2009; Covrig & Low, 2003; Jorion, 1995; Plíhal & Lyócsa, 2021; Xu & Taylor, 1995). A partially contradictory finding is provided by Pong, Shackleton, Taylor, and Xu (2004), who finds evidence that RV can beat IV forecasts using high-frequency data for short forecasting horizons of one day and one week. As the literature indicates, IV tends to be useful for future volatility forecasting. As the tails of a return distribution are related to volatility, it seems reasonable to assume that IV is beneficial for estimating tail returns as well, even though forecasting volatility and forecasting quantiles of the return distribution are different objectives.

The literature is not entirely clear on the role of IV in VaR prediction. IV might perform poorly because of the mentioned theoretical drawbacks with this risk-neutral measure, such as nonlinear and regime-changing dynamics of the volatility risk premium (Bams, Blanchard, & Lehnert, 2017). More specifically, in an effort to hedge portfolios, the demand for protection against large price movements leads to an overestimation of extreme price movements, and hence to overestimated implied volatility. The size of this volatility risk premium differs across types of markets. Some authors argue that while IV is forward looking, the embedded volatility risk premium makes IV less useful for predictive purposes.

When VaR is estimated using only implied volatility, it is often dominated by simple estimators based on realized volatility (see Chong (2004), Christoffersen and Mazzotta (2005) for currency markets, and Bams et al. (2017) for the US equity market). A possible alternative is to combine both measures, RV and IV. . In several studies (for various assets) this approach leads to the outperformance of predictions based solely on RV or IV (Giot, 2005; Jeon & Taylor, 2013; Kambouroudis et al., 2016; Kim & Ryu, 2015; Leiss & Nax, 2018; Siu, 2018). Slim, Dahmene, and Boughrara (2020) conclude that the volatility risk premium and resulting benefits for VaR predictions differ across equity markets.

Surprisingly, little is known about the role of IV in predicting ES. While the VaR and ES are related, they are not the same, and evidence for the role of IV in predicting ES is missing. Our study fills this gap in the literature, and we follow the dominant approach in the literature in that we combine RV and IV in our models. Specifically, we are interested in implied volatility's role in predicting the day-ahead expected shortfall of the EUR/USD currency pair, which we select because it is the most traded and researched currency pair with unique implied volatility-based trading of options that does not require us to formulate an explicit option pricing modeling assumption. This trading arrangement is unique to the FX market. Although in Section 5.5 we also offer guidance for five-, 10-, and 22-day-ahead forecasts, our in-depth analysis is based on day-ahead forecasting periods for several reasons. First, as noted in the MAR99 "Guidance on use of the internal models approach" (Basel Committee on Banking Supervision, 2023), major trading institutions commonly experience significant changes in portfolio composition relative to initial positions over different multi-day periods, while backtesting frameworks are often calibrated to a one-day holding period. Second, according to Hull and Basu (2018), when considering market risks, analysts frequently initiate calculations of VaR or ES specifically for a one-day time horizon, making it relatively easy to approximate forecasts for multi-day periods. Third, one-day-ahead forecasts are prevalent in the majority of the academic literature (e.g. Bams et al., 2017; Chong, 2004; Giot, 2005; Jeon & Taylor, 2013; Kim & Ryu, 2015; Patton, Ziegel, & Chen, 2019; Slim et al., 2020; Taylor, 2019, 2020).

With this work, we contribute to the literature along multiple lines. First, instead of volatility indices, we use implied volatility measures from options with daily, weekly, and monthly maturities. The motivation is based on the heterogeneous market hypothesis, where it is recognized that market participants have different preferences with respect to investment horizons. It follows that volatility aggregated across different time horizons might be different as well (Dacorogna, Mller, Olsen, & Pictet, 2001; Dacorogna, Müller, Pictet, & Olsen, 1997; Müller et al., 1993). Moreover, our volatility measures are not explicitly derived under an assumption of an option pricing model. Instead, we use implied volatility directly as traded by currency options traders on a well-known trading platform. Second, we propose several quantile-based models to predict the expected shortfall; these models are easy to estimate, and we show that they tend to perform better than standard time-varying volatility models used in the literature. Third, we provide strong evidence that when implied volatility is utilized, we achieve more accurate predictions of value at risk and the expected shortfall. This result is stronger when the euro weakens, and holds across time and alternative evaluation frameworks.

The remainder of the paper is structured as follows. The next sections describe our data and methodology. We present quantile and expected shortfall prediction models and two evaluation strategies. Section 4 shows key empirical results, while Section 5 explores the stability of our results across different conditions. Section 6 summarizes our key findings.

2. Data

2.1. EUR/USD exchange rates

Our dataset spans 17 years, from June 4, 2003, to October 10, 2020, which is the longest period available from our data sources. We focus on the foreign exchange market and choose the most traded currency pair, EUR/USD.[1] The reason for our choice of the EUR/USD FX market is twofold. First, contrary to other traded assets (e.g. equities), FX options are traded on the over-the-counter (OTC) market, and prices are directly quoted in implied volatilities (see Section 2.2, below). Second, we exploit the advantage of using such a highly liquid currency pair in that it reduces the possibility of the data featuring large spreads and microstructure noise that could influence the results. We work with historical high-frequency OTC quoted spot prices of EUR/USD provided by Dukascopy.[2] The foreign exchange market is influenced by intraday and intraweek seasonality, which causes dramatic changes in trade volume. These effects are connected with the opening hours in the relevant markets (Aloud, Fasli, Tsang, Dupuis, & Olsen, 2013). Therefore, we remove weekends and set the end of the trading day to 4:00 p.m. UTC, which is in accordance with implied volatility calculations provided by Bloomberg and described in the next section. We consider the whole 24-hour trading day, starting at 4:01 p.m. and ending at 4:00 p.m. the following day, to capture all relevant price movements.

2.2. Raw implied volatility data

Implied volatility data for the EUR/USD currency pair are collected from Bloomberg, which uses highly liquid OTC market currency derivative contracts to calibrate at-the-money (ATM) implied volatilities at given maturities. The trading of options on the OTC foreign exchange market is usually realized by quoting the implied volatility directly. Bloomberg aggregates these data on a daily basis from the brokers and dealers of large banks and insurance companies.[3] This procedure reduces the possible idiosyncratic effect specific to individual market participants who provide quotes, and it ensures the high quality and accuracy of implied volatility data. Our analysis uses implied volatility quotes from Bloomberg for three different maturities: one day (overnight volatility), one week, and one month. All quotes are obtained for ATM options with the same delta.[4] However, liquidity differs considerably for other strike prices, and it is especially low for out-of-money options with shorter maturity; thus, objective data constraints limit us from pursuing this course of action. Note that apart from using the level of the implied volatility as in this study, one might also utilize at-the-money implied volatility to calculate slopes and the curvature across maturities. In a recent study, Clements, Liao, and Tang (2022) utilized the level, slope, and curvature of implied volatilities with longer maturities (from 30 to 365 days), but the volatility forecasting improvement of the S&P 500 ETF was somewhat marginal against a standard HAR model.

[1] The EUR/USD currency pair represents approximately 24% of total foreign exchange market turnover. See the Bank of International Settlements report https://www.bis.org/statistics/rpfx19_fx.pdf.

[2] The data are publicly available at https://www.dukascopy.com/swiss/english/marketwatch/historical/.

[3] The number of institutions varies but is usually approximately 80, including e.g. GFI Group, HSBC, Raiffeisen, Societe Generale, and InvestAZ.

[4] It might be more straightforward to use estimates from a range of strike prices and consequently estimate the expected shortfall. For example, one might consider using out-of-money options to predict extreme price movements, which would better match the purpose of our empirical strategy with the underlying instrument.

3. Methodology

3.1. Returns

In this study, our key observable variables of interest are daily close-to-close returns.[5] As noted above, the closing price P_t is recorded at 4:00 p.m. UTC, where $t = 1, 2, \ldots, T$ denotes the usual time index that corresponds to a given day. The daily return is defined as:

$$R_t = 100 \times \frac{P_t - P_{t-1}}{P_{t-1}} \tag{1}$$

3.2. Market risk measures: VaR and ES

Value at risk ($VaR_{R_t}(\tau), \tau \in \langle 0, 1\rangle$) is the minimum loss (in terms of returns $R_t, t = 1, 2, \ldots$) incurred in the $\tau \times 100\%$ worst returns of a portfolio (Acerbi & Tasche, 2002). $VaR_{R_t}(\tau)$ has been a popular measure of market risk since it was introduced in the 1980s (Nadarajah, Zhang, & Chan, 2014). Private firms use $VaR_{R_t}(\tau)$ as an internal risk management tool, monitoring the exposure to risky positions and portfolios. The usage of $VaR_{R_t}(\tau)$ was further endorsed by regulators through the Basel Accords (particularly Basel II).

Although $VaR_{R_t}(\tau)$ is conceptually simple and easy to estimate and evaluate (it is elicitable), it suffers from deficiencies that led it to be replaced by other measures. Most notably, $VaR_{R_t}(\tau)$ does not sufficiently account for tail risk (Du & Escanciano, 2017), as it does not consider the size of the losses realized below the chosen return threshold (τ-quantile). As a consequence, several portfolios with the same $VaR_{R_t}(\tau)$ may experience dramatically different losses below the VaR threshold (Acerbi & Tasche, 2002). $VaR_{R_t}(\tau)$ does not seem to work well during times of crisis, and it is neither subadditive (Artzner, Delbaen, Eber, & Heath, 1999) nor convex (Basak & Shapiro, 2001). This lack of subadditivity implies that diversification, an essential strategy for portfolio optimization and in finance theory in general, does not necessarily reduce the $VaR_{R_t}(\tau)$.

To address these shortcomings, the expected shortfall ($ES_{R_t}(\tau)$, sometimes referred to as conditional VaR) has been introduced as an alternative risk measure. It may be defined as the expected loss incurred in the $\tau \times 100\%$ worst cases of a portfolio (Acerbi & Tasche, 2002). Given a specific quantile $\tau \ll 0.5$, we can express ES via *VaR* as:[6]

$$\begin{aligned} VaR_{R_t}(\tau) &= Q_{R_t}(\tau | I_{t-1}) \\ ES_{R_t}(\tau) &= E\left[R_t | R_t \leq VaR_{R_t}(\tau), I_{t-1}\right] \end{aligned} \tag{2}$$

where VaR_{R_t} denotes the estimated value at risk using some function $Q_{R_t}(.)$ and all information (I_{t-1}) known at time $t-1$. The expected shortfall represents the expected return below the chosen τ-quantile. The equation above shows that $VaR_{R_t}(\tau)$ can be perceived as an upper bound on $ES_{R_t}(\tau)$; thus, $ES_{R_t}(\tau)$ can be understood as a more restrictive risk measure than $VaR_{R_t}(\tau)$. ES_{R_t} can be expressed as:

$$ES_{R_t}(\tau) = \frac{1}{\tau}\int_0^\tau VaR_{R_t}(u)du \tag{3}$$

As $ES_{R_t}(\tau)$ is explicitly endorsed by Basel III, current research focuses more on $ES_{R_t}(\tau)$ than on $VaR_{R_t}(\tau)$. Practical applications using measures of market risk often replace $VaR_{R_t}(\tau)$ at the probability level of $\tau = 0.01$ with $ES_{R_t}(\tau)$ at the level of $\tau = 0.025$. A more thorough comparison of VaR and ES, along with their advantages and disadvantages, is presented, for example, in Embrechts, Liu, and Wang (2018), Embrechts, Puccetti, Rüschendorf, Wang, and Beleraj (2014). While $VaR_{R_t}(\tau)$ is elicitable with a piecewise-linear loss function, expected shortfall, while being subadditive (and a coherent measure of market risk), is not elicitable, which means that there is no loss function that can be optimized to estimate it (Patton et al., 2019). The main consequence for $ES_{R_t}(\tau)$ is that backtesting is not straightforward, a fact that has given rise to a rich area of research on possible backtesting approaches in recent years. Given the forward-looking nature of derivative contracts, we explore whether implied volatility can improve predictions of day-ahead ES for the left tail $\tau = 0.01, 0.025, 0.05$ and right tail $(1 - \tau) = 0.95, 0.975, 0.99$ of the EUR/USD FX returns. As the usual notation follows risk applications in primarily focusing on the left tail of the return distribution (with $\tau \ll 0.5$), the analysis of the right tail $(1 - \tau)$ may be performed by taking the negative of the ES from the left tail (τ) of the distribution of the opposite returns (additive inverse, i.e. the left tail of the distribution of original returns multiplied by -1).

3.3. Overview of the empirical strategy

To assess the possible benefits of utilizing the information on implied volatility, our empirical strategy involves comparing the forecasting accuracy of models with and without implied volatilities. To summarize, we have two groups of models that predict $ES_{R_t}(\tau)$:

- The first group does not utilize information from implied volatilities:
  - Historical models assuming flexible but static return distributions.
  - The exponential generalized autoregressive conditional heteroskedasticity model of Nelson (1991), the EGARCH model.
  - Quantile regression models (QR models, henceforth) with approximate ES estimation and specifications, including realized price variation measures.
  - The joint VaR and ES model of Dimitriadis and Bayer (2019) (the DB model, henceforth) with specifications involving realized price measures.

[5] We prefer using gross returns instead of the often-used log returns in our empirical setting for two reasons. First, although the differences between log and gross returns are very small in most applications, the differences between the two will increase for extreme returns. As extreme returns are of interest in our study, we incline towards gross returns. Second, in real-world scenarios, one receives gross (realized) returns instead of log returns.

[6] The right tail ES is defined analogously.

- The second group augments models with information on implied volatility obtained from options on the EUR/USD currency market with daily, weekly, and monthly maturities:
  - The EGARCH model with implied volatility in the variance equation.
  - QR models with specifications also including implied volatilities.
  - The DB model with specifications also including implied volatilities.

To evaluate these multiple prediction models, we follow three recent advances in evaluating ES predictions. In the first statistical approach, we use the strict ES regression-based backtesting of Bayer and Dimitriadis (2022), where the hypothesis of interest is a well-specified forecast of the expected shortfall, i.e. a forecast that does not under- or overestimate ES.[7] In the second statistical approach, we follow two recent studies: Taylor (2019) and Patton et al. (2019). They both utilize the fact that given a certain class of loss functions, VaR and ES are jointly elicitable (using the results of Fissler & Ziegel, 2016). It follows that VaR and ES can be jointly evaluated.[8] Our main results are based on the FZ^0 loss function of Patton et al. (2019). To evaluate the accuracy of the 38 prediction models, including simple combination forecasts, we employ Hansen, Lunde, and Nason (2011)'s model confidence test to find a set of superior models (for a similar approach, see Gerlach, Naimoli, & Storti, 2020). Third, we perform the pairwise weak dominance test of Ziegel, Krüger, Jordan, and Fasciati (2020), where we test whether it is more likely that models that utilize IV dominate models that do not.

3.4. Historical and GARCH class models

Our first class of models comprises historical simulation models based on two return distributions. Second, we use forecasts based on conditional time-varying parameters with autoregressive GARCH models that are common in the literature (e.g. Du & Escanciano, 2017). The former approach can be expressed as a special case of the latter; thus, we can write the GARCH model as:

$$
\begin{aligned}
R_t &= \mu_0 + \phi R_{t-1} + \epsilon_t, \\
\epsilon_t &= \sigma_t \eta_t, \\
\eta_t &\sim iid\,(0,1) \\
\ln \sigma_t^2 &= \omega + \alpha z_{t-1} + \gamma\left(|\eta_{t-1}| - E|\eta_{t-1}|\right) \\
&\quad + \beta \ln \sigma_{t-1}^2 + \sum_{k=1} \nu_k X_{k,t-1} \qquad (4)
\end{aligned}
$$

where ϕ is the autoregressive coefficient, and the variance equation that we employ is the exponential GARCH of Nelson (1991), with η_t representing standardized innovations, and coefficients α and γ capture the sign and size effects. Other choices of variance model are possible, e.g. Couperier and Leymarie (2020), Du and Escanciano (2017), Francq et al. (2019), Löser, Wied, and Ziggel (2018), who use a simpler GARCH models with different error structures.[9]

Historical simulation-based models assume that in Eq. (4), $\phi = \alpha = \gamma = \beta = 0 \wedge \forall_i : \nu_i = 0$ and that η_t follows a specific distribution with given parameters estimated with data from a given estimation window. Using the estimated parameters, we predict the quantiles of interest and the expected shortfall directly via numerical integration. We consider two distributions: (i) the skewed Student's t-distribution (the *SSK* model, henceforth) (Ferreira & Steel, 2006), and (ii) Johnson's SU distribution (the *JSU* model, henceforth) (Johnson, 1949a, 1949b). These are flexible enough to incorporate asymmetries and heavy tails. Johnson's distribution has recently been employed in several studies modeling market returns (e.g. Horváth, Lyócsa, & Baumöhl, 2018; Lyócsa, Vŷrost, & Baumöhl, 2019), while Choi and Nam (2008), Corlu and Corlu (2015), Gurrola (2007) specifically model FX returns and find that flexible distributions lead to a better fit.

Restricting $\forall_i : \nu_i = 0$ in Eq. (4) leads to the EGARCH model of Nelson (1991). Depending on the underlying distribution, we have either EGARCH-SSK or EGARCH-JSU models.

In the variance equation of Eq. (4), the exogenous variables $X_{k,t-1}$ are either (i) historical past levels of average daily (RV^D), weekly (RV^W), and monthly (RV^M) price variation, which leads to the EGARCH-RV-SSK and EGARCH-RV-JSU models; or with (ii) implied volatilities with one-day (IV^D), one-week (IV^W), and one-month (IV^M) maturities. These lead to the EGARCH-IV-SSK and EGARCH-IV-JSU models, respectively.

More specifically, let $j = 1, 2, \ldots, N$ correspond to intraday five-minute sampling observations, so that $P_{t,j}$ is the price at the jth five-minute interval on day t.[10] The realized volatility is given by:

$$
\begin{aligned}
RV_t(K) &= (K+1)^{-1} \sum_{k=0}^{K} \sum_{j=2}^{N} \left(lnP_{t-k,j} - lnP_{t-k,j-1} \right)^2 \\
&= (K+1)^{-1} \sum_{k=0}^{K} \sum_{j=2}^{N} R_{t-k,j}^2 \qquad (5)
\end{aligned}
$$

[7] Another alternative ES-specific backtest is proposed, e.g., by Barendse, Kole, and van Dijk (2023).

[8] It also follows that one can create corresponding VaR and ES forecasting models. Taylor (2019, 2020) and Meng and Taylor (2020) proposed several such models, while Patton et al. (2019) also provided theoretical results supporting the validity of the Taylor (2019) approach. Several other contributions along these lines are the studies of Candila, Gallo, and Petrella (2020) and Le (2020), who enrich such models by allowing mixed data sampling frequencies.

[9] In a previous version of the manuscript we used the Glosten, Jagannathan, and Runkle (1993) model, with qualitatively similar results.

[10] This sampling scheme is often preferred in the literature, as it is considered to lead to a good balance between precision and microstructure noise. In an influential study, Liu, Patton, and Sheppard (2015) provide evidence that, in general, other frequencies rarely lead to more accurate estimates. The given statement is related to estimation accuracy. In forecasting settings, the five- and 15-minute truncated realized variance seems more appropriate, as it disentangles jumps from the integrated variance.

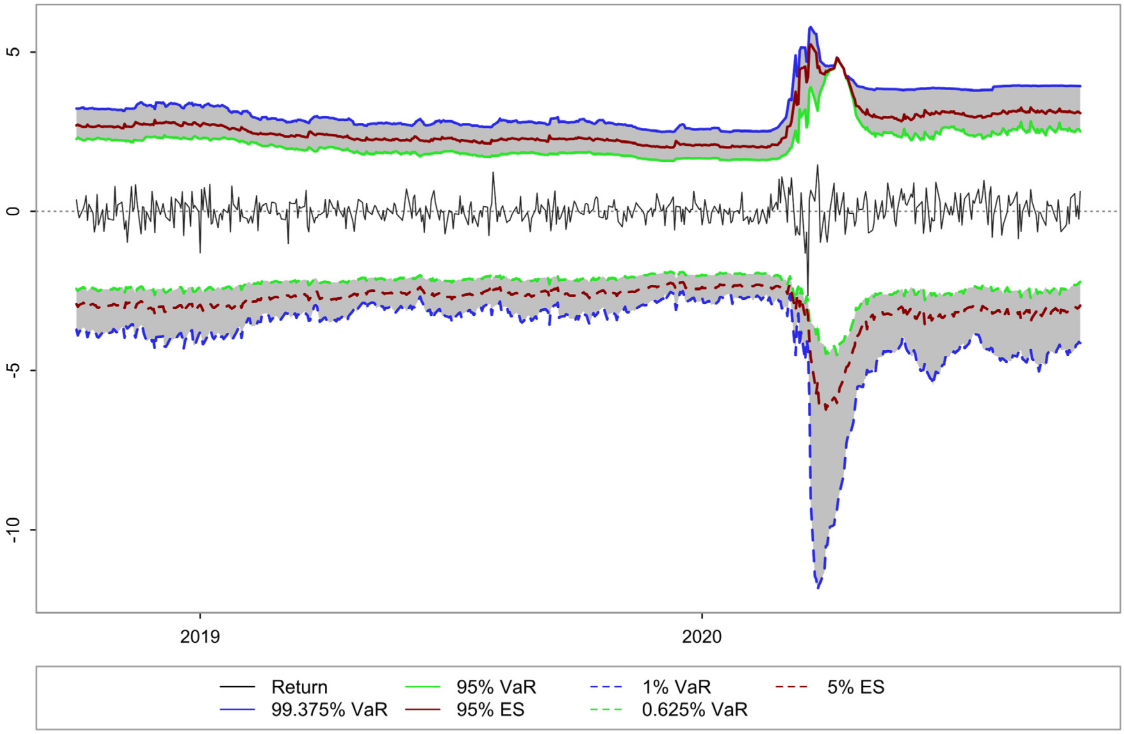

Fig. 1. Example of the approximation of the expected shortfall.

Here, $k = 1, 2, \ldots, T$ for $t \geq K$ corresponds to given days, so the daily (RV_t^D), weekly (RV_t^W), and monthly (RV_t^M) realized volatility is given with $K = 0, 4,$ and 21.

3.5. Approximate expected shortfall

As noted by Couperier and Leymarie (2020), ES can be approximated via the sum:

$$ES_{R_t}(\tau) \approx \frac{1}{p} \sum_{j=1}^{p} \widehat{VaR}_{R_t}(u_j) \tag{6}$$

In Eq. (6), u_j corresponds to $j = 1, 2, \ldots, p$ quantile levels, and $\widehat{VaR}_{R_t}(u_j)$ to the predicted VaR. This approximation suggests that we can use a two-step ES estimation strategy. First, we estimate p VaR quantiles. Second, we find $ES_{R_t}(\tau)$ using Eq. (6). Theoretically, the larger the number of quantiles p, the more accurate the approximation. However, in a finite sample, we cannot accurately estimate the extreme quantiles needed for larger p. We set $p = 8$. Following Couperier and Leymarie (2020), for $\tau < 0.5$, we can find which quantiles are of interest via $u_j = 1 - \left[(1-\tau) + (j-1)(1-(1-\tau))p^{-1}\right]$. Fig. 1 demonstrates the connection between VaR and ES. We plot the 5% VaR (the upper bound on ES) and the 0.625% VaR, the lowest needed to approximate the ES with $p = 8$, with the gray area in between. The estimated ES at 5% is denoted by the red line.

3.6. QR models

Following the ideas proposed in the previous section, a prediction of the future ES can be approximated via predicted quantiles. Motivated by earlier works of Haugom, Ray, Ullrich, Veka, and Westgaard (2016) and Lyócsa, Todorova, and Vỳrost (2021), we use QRs, where R_t is a linear function of a set of K variables, $\boldsymbol{x}_{t-1} = (x_{1,t-1}, \ldots, x_{K,t-1})'$, $\boldsymbol{z}_{t-1} = (1, \boldsymbol{x}'_{t-1})$. The u_j-th conditional quantile of EUR/USD returns is $\boldsymbol{z}'_{t-1}\boldsymbol{\beta}(\boldsymbol{u_j})$, $P(R_t \leq \boldsymbol{z}'_{t-1}\boldsymbol{\beta}(u_j)|\boldsymbol{x}_{t-1}) = u_j$. Apart from the simplicity of the QR models, an attractive feature of estimating $ES_t(\tau)$ via different quantiles is that the vector of coefficients $\beta(u_j)$ is likely to vary across quantiles u_j. We can therefore exploit the potentially nonlinear relationship between the explanatory variables and FX market returns. In the case of unsatisfactory ES forecasts, we can not only identify which quantiles are more difficult to estimate but also identify factors that are likely to improve quantile forecasts (and thus likely expected shortfall). On the other hand, the overall number of parameters tends to be much higher than in GARCH models, and we might encounter quantile-crossing issues. The estimation is carried out using the standard asymmetric loss function:

$$\rho(u_j, v) = v(u_j - I(v < 0)) = 2^{-1}\left[|v| + (2u_j - 1)v\right] \tag{7}$$

By minimizing the loss function, we obtain the estimates:

$$\hat{\boldsymbol{\beta}}(u_j) = arg \min_{\boldsymbol{\beta}(u_j)} \sum_{t=1}^{T} \rho(R_t - \boldsymbol{z}'_{t-1}\boldsymbol{\beta}(u_j)) \tag{8}$$

To estimate QR models, we use the Barrodale and Roberts (1978) simplex algorithm for l1-regressions, as described in Koenker and d'Orey (1994) and implemented in Koenker (2020). The next sections describe estimated specifications used to predict all quantiles of interest.

3.6.1. Realized variance models: QR-RV

Our benchmark model is motivated by the previous studies of Haugom et al. (2016) and Lyócsa, Todorova, and Vŷrost (2021). The model is denoted as QR-RV, and the specification can be expressed as:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j} RV^D_{t-1} + \beta_{2,u_j} RV^W_{t-1} + \beta_{3,u_j} RV^M_{t-1} \quad (9)$$

The right-hand-side specification given in Eq. (9) is the same as that of the standard heterogeneous autoregressive (HAR, henceforth) model of Corsi (2009) for predicting volatility. During periods of higher volatility, extreme returns are more likely to occur. Therefore, variables known to predict next-day volatility well should also be useful in predicting next-day extreme returns.

Many variations of the HAR model exist that can also be employed in the QR context. We use four popular alternative specifications. The next specification, denoted QR-RV-CJ, is given by:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j} CC_{t-1} + \beta_{2,u_j} JC_{t-1} + \beta_{3,u_j} RV^W_{t-1} + \beta_{4,u_j} RV^M_{t-1} \quad (10)$$

The terms CC_{t-1} and JC_{t-1} are the continuous and jump components, respectively. As the continuous component represents the part of the price variation not driven by discontinuous price movements, it should be more persistent and thus more indicative of future price variation. The jump component tends to be much less persistent. However, if price discontinuities are clustered over time, their inclusion in our specification might be useful for predicting extreme price movements. The estimation of these components follows the procedure in Andersen, Dobrev, and Schaumburg (2012). Specifically, the continuous component is given as:

$$CC_t = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{N}{N-2}\right) \sum_{j=2}^{N-1} med\{|R_{t,j-1}|, |R_{t,j}|, |R_{t,j+1}|\}^2 \quad (11)$$

The jump component (JC_t) is given as $RV_t - CC_t$ if positive, or the jump component is equal to 0 if the difference is negative.[11]

The specification denoted as QR-RV-CQ combines the continuous components with volatility measurement uncertainty:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j} CC_{t-1} + \beta_{2,u_j} CQ_{t-1} + \beta_{3,u_j} RV^W_{t-1} + \beta_{4,u_j} RV^M_{t-1} \quad (12)$$

with:

$$CQ_t = CC_t \times \sqrt{MedRQ_t} \quad (13)$$

where $MedRQ_t$ is the median realized quarticity, given by Andersen et al. (2012) as:

$$MedRQ_t = \frac{3\pi N}{9\pi + 72 - 52\sqrt{3}} \left(\frac{N}{N-2}\right) \times \sum_{j=2}^{N-1} \left[med|r_{t,j-1}|, |r_{t,j}|, |r_{t,j+1}|\right]^4 \quad (14)$$

The original observation of Bollerslev, Patton, and Quaedvlieg (2016) was that during periods of high volatility, uncertainty regarding the volatility estimate itself increases. Therefore, volatility persistence should decrease with the increase in uncertainty surrounding the volatility estimate ($MedRQ_t$). We utilize the same idea in our quantile regressions that predict extreme returns by assuming that the explanatory power of volatility, $MedRV_t$, should be lower in times of higher uncertainty, $MedRQ_t$.

Next, two specifications are motivated by the decomposition of the realized variance into positive and negative semivariances (Barndorff-Nielsen, Kinnebrock, & Shephard, 2008). In this regard, Patton and Sheppard (2015) find that negative price variation is more persistent and a better predictor of the next-day price variation on US equity markets. Specifically, the negative (NV) and positive (PV) semivariances are defined as:

$$NV_t = \sum_{j=2}^{N} r^2_{t,j} \times I\left(r_{t,j} < 0\right)$$
$$PV_t = \sum_{j=2}^{N} r^2_{t,j} \times I\left(r_{t,j} \geq 0\right) \quad (15)$$

where $I(.)$ is an indicator function returning 1 if the intraday return is either negative (for NV_t) or not negative (for PV_t), and 0 otherwise. We expect that the negative semivariance is more indicative of the next-day extreme left tail return and the positive semivariance more indicative of the next-day extreme right tail return; thus, we include both in our QR-RV-SV model:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j} PV_{t-1} + \beta_{2,u_j} NV_{t-1} + \beta_{3,u_j} RV^W_{t-1} + \beta_{4,u_j} RV^M_{t-1} \quad (16)$$

We follow the previous ideas of Barndorff-Nielsen et al. (2008), Andersen et al. (2012) and Patton and Sheppard (2015) by using the continuous and signed jump components, with the latter being the difference between the positive and negative semivariances:

$$SJ_t = \left(\sum_{j=2}^{N} r^2_{t,j} \times I\left(r_{t,j} \geq 0\right) - \sum_{j=2}^{N} r^2_{t,j} \times I\left(r_{t,j} < 0\right)\right) \quad (17)$$

Barndorff-Nielsen et al. (2008) show that the difference between the realized variance components leads to signed jump variation; that is, as opposed to JC_t defined above, SJ_t is positive when positive jumps dominate the process, while SJ_t is negative when negative price variation is more prevalent. As our model includes the signed jump, we combine it with the continuous component, which leads

11 Alternatively, nonzero jump components might be designated only for statistically significant discontinuities. We use the asymptotic results in Eq. (6) of Andersen et al. (2012) to test for the significance of the jumps, where we use median realized quarticity as an estimate of the integrated quarticity. However, we found that the resulting time series behave similarly to the original series, since the correlation between the resulting series with that we use in our study is 0.920 with Pearson's coefficient and 0.988 with Spearman's.

to the QR-RV-SJ model:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j}CC_{t-1} + \beta_{2,u_j}SJ_{t-1} + \beta_{3,u_j}RV^W_{t-1} + \beta_{4,u_j}RV^M_{t-1} \quad (18)$$

3.6.2. Implied volatility models: QR-IV

To test whether implied volatility matters for predicting the ES, we augment the QR-RV models with implied volatilities. As noted in Section 2, we use data on implied volatility for the EUR/USD currency pair of at-the-money options with the same delta. However, we incorporate implied volatility with different maturities: one day, denoted as IV^D_t; weekly, denoted as IV^W_t; and monthly, as denoted as IV^M_t. The first specification, the QR-IV model, uses only the implied volatilities:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j}IV^D_{t-1} + \beta_{2,u_j}IV^W_{t-1} + \beta_{3,u_j}IV^M_{t-1} \quad (19)$$

Adding the three implied volatility components into the different realized volatility models defined in the previous section leads to our remaining specifications. Specifically, the QR-IV-RV model:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j}RV^D_{t-1} + \beta_{2,u_j}RV^W_{t-1} + \beta_{3,u_j}RV^M_{t-1} + \beta_{4,u_j}IV^D_{t-1} + \beta_{5,u_j}IV^W_{t-1} + \beta_{6,u_j}IV^M_{t-1} \quad (20)$$

the QR-IV-CJ model:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j}CC_{t-1} + \beta_{2,u_j}JC_{t-1} + \beta_{3,u_j}RV^W_{t-1} + \beta_{4,u_j}RV^M_{t-1} + \beta_{5,u_j}IV^D_{t-1} + \beta_{6,u_j}IV^W_{t-1} + \beta_{7,u_j}IV^M_{t-1} \quad (21)$$

the QR-IV-CQ model:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j}CC_{t-1} + \beta_{2,u_j}CQ_{t-1} + \beta_{3,u_j}RV^W_{t-1} + \beta_{4,u_j}RV^M_{t-1} + \beta_{5,u_j}IV^D_{t-1} + \beta_{6,u_j}IV^W_{t-1} + \beta_{7,u_j}IV^M_{t-1} \quad (22)$$

the QR-IV-SV model:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j}PV_{t-1} + \beta_{2,u_j}NV_{t-1} + \beta_{3,u_j}RV^W_{t-1} + \beta_{4,u_j}RV^M_{t-1} + \beta_{5,u_j}IV^D_{t-1} + \beta_{6,u_j}IV^W_{t-1} + \beta_{7,u_j}IV^M_{t-1} \quad (23)$$

and the QR-IV-SJ model:

$$VaR_{R_t}(u_j) = \beta_{0,u_j} + \beta_{1,u_j}CC_{t-1} + \beta_{2,u_j}SJ_{t-1} + \beta_{3,u_j}RV^W_{t-1} + \beta_{4,u_j}RV^M_{t-1} + \beta_{5,u_j}IV^D_{t-1} + \beta_{6,u_j}IV^W_{t-1} + \beta_{7,u_j}IV^M_{t-1} \quad (24)$$

3.7. A joint quantile and expected shortfall regression framework

Following the recent development based on the elicitability results of Fissler and Ziegel (2016), we complement our analysis using a recent contribution to the forecasting of VaR and ES market risk measures via the joint VaR and ES regression model of Dimitriadis and Bayer (2019). Technically, the approach of Dimitriadis and Bayer (2019) is very similar to the work of Patton et al. (2019), who provide an M-estimator for joint regression models for VaR and ES. We utilize the Dimitriadis and Bayer (2019) framework for two main reasons. First, while the Patton et al. (2019) approach is based on a choice of a single strictly consistent loss function (albeit a very sensible one), the Dimitriadis and Bayer (2019) model has a more general approach allowing for a larger class of loss functions. Second, the Patton et al. (2019) approach has more general assumptions about the data generation processes, intending to allow for autoregressive modeling in their forecasting framework. On the other hand, Dimitriadis and Bayer (2019) provide the asymptotic theory for ES forecasting models that allow for the inclusion of exogenous variables, which makes their approach more suitable in our endeavor to verify the benefits of including the information on implied volatility in the ES modeling framework.

Let $R \in \mathbb{R}$ be the returns and $\mathbf{X}_q \in \mathbb{R}^{k_1}$, $\mathbf{X}_e \in \mathbb{R}^{k_2}$, $k_1, k_2 \in \mathbb{N}$ the vectors of explanatory variables. The regression framework jointly modeling the VaR and ES of R, given the covariates $\mathbf{X}_q$ and $\mathbf{X}_e$ for some fixed-level $\tau \in (0, 1)$, is given by:

$$R = \mathbf{X}_q'\boldsymbol{\theta}_0^q + u^q, \qquad \text{and} \qquad R = \mathbf{X}_e'\boldsymbol{\theta}_0^e + u^e \quad (25)$$

where the requirements on the error terms are $VaR_{u^q}(\tau|\mathbf{X}_q, \mathbf{X}_e) = 0$ and $ES_{u^e}(\tau|\mathbf{X}_q, \mathbf{X}_e) = 0$. We follow Dimitriadis and Bayer (2019), who propose an M-estimation and Z-estimation procedure for the regression parameter vectors $\boldsymbol{\theta}_0^q$ and $\boldsymbol{\theta}_0^e$ and derive their asymptotic properties. As Dimitriadis and Bayer (2019) found, the Z-estimator is unstable, and the M-estimator is used. With respect to the specific loss function used in the estimation, we opt for $G_1(z) = 0$ and $\mathcal{G}_2 = -\log(-z)$ for $z < 0$ (for details, see Fissler and Ziegel (2016) and particularly Section 4.2 in Dimitriadis and Bayer (2019)).

3.7.1. Realized variance models: DB-RV

We use the Dimitriadis and Bayer (2019) joint VaR and ES regression framework to estimate the VaR and ES directly. The specifications that we use follow those defined for the QR framework. Specifically, we estimate five specifications with realized volatility estimators—DB-RV, DB-RV-CJ, DB-RV-SV, DB-RV-CQ, and DB-RV-SJ—which are analogous to the QR models in that they adopt a HAR model structure, continuous and jump components, semivariances, a measurement error component, and signed jumps, respectively.

3.7.2. Implied volatility models: DB-IV

With respect to the variations that employ implied volatilities, we use five specifications: DB-IV, DB-IV-CJ, DB-IV-CQ, DB-IV-SV, and DB-IV-SJ. The DB-IV uses only daily, weekly, and monthly implied volatilities. The remaining models are augmented with continuous and jump components, a measurement error component, semivariances, or signed jumps.

3.8. Combination forecasts

An accurate prediction of extreme price movements is constrained by the limited number of observations around the extreme price return events. An additional source of prediction error stems from unknown data-generating processes (DGPs), as the true DGP is unknown or might change over time. We therefore combine multiple individual forecasts (of ES and VaR) into a single forecast by taking a simple average across forecasts from different models. As argued by Timmermann (2006), this allows us to address model choice uncertainty and reduce our estimation error, potentially leading to more accurate quantile and ES predictions. Despite the importance of VaR and ES, the literature on combining quantiles is limited. Recent applications for equity indices include Lyócsa and Stašek (2021), who perform multiple quantile combinations to improve the prediction of the volatility distribution. In the context of risk measures (VaR and ES), Taylor (2020) advocates for quantile combinations and notes that the nonexistence of a loss function for ES only makes (weighted) combinations of ES problematic. In our study, we combine estimates of VaR and resulting ES using a simple arithmetic average.

We create nine combination forecasts:

1. Two EGARCH models that use realized variance (EGARCH-RV-SSK and EGARCH-RV-JSU).
2. Two EGARCH models that use implied volatility (EGARCH-IV-SSK and EGARCH-IV-JSU).
3. Five QR models that use realized variance components (QR-RV, QR-RV-CJ, QR-RV-CQ, QR-RV-SV, and QR-RV-SJ).
4. Five QR models that also use implied volatility (QR-IV, QR-IV-CJ, QR-IV-CQ, QR-IV-SV, and QR-IV-SJ).
5. Five DB models that use realized variance components (DB-RV, DB-RV-CJ, DB-RV-CQ, DB-RV-SV, and DB-RV-SJ).
6. Five DB models that also use implied volatility (DB-IV, DB-IV-CJ, DB-IV-CQ, DB-IV-SV, and DB-IV-SJ).
7. Fourteen models that use realized variance components including historical simulation-based models.
8. Twelve models that use only realized variance components.
9. Twelve models that also use implied volatility.

By comparing individual forecasts and combination forecasts, we can draw broader conclusions with respect to the information used (implied volatility), the different estimation methods employed (EGARCH, QR, or the Dimitriadis and Bayer (2019) model), and the role of model choice uncertainty (individual vs. combination forecasts).

3.9. Forecasting procedure

We use a rolling window estimation approach with a window size of 1500, which corresponds to 6 years of data. Parameters of all models are re-estimated after each observation. It follows that our first ES forecast is for the 1501st day (March 4, 2009) and ends on the 4522nd observation (October 2, 2020), leading to 3022 out-of-sample forecasts.

Within each estimation window, the realized measures used in forecasting models are subject to automatic screening for outlier observations. Specifically, all values above the 0.995 quantile (i.e. 1500 observations from the estimation window $\times(1-0.995) \rightarrow$ 7 or 8 highest numbers) are winsorized to the nearest value below the 0.995 quantile. For signed jumps, the only ones that can attain negative realization, values below the 0.005 quantile are also winsorized. Implied volatility measures, as our key variable of interest, were not subject to winsorization.[12]

With respect to quantile forecasts, three procedures are used to rule out implausible predictions. First, values six times the standard deviation above (below) the last observed return are substituted with the previous-period prediction. Second, if predicted left-tail (right-tail) quantiles are above (below) the historical 0.35 (0.65) percentile, as estimated within the estimation window, the given values are substituted with the last observed prediction. Third, we have to ensure quantile non-crossing, which is not guaranteed under QR models. If two predicted consecutive quantiles cross, the higher one is substituted with the lower one plus 0.0001.[13]

3.10. Expected shortfall forecast evaluation

Next, we present three approaches that we use to evaluate the ES predictions. The first is based on the ideas of Mincer and Zarnowitz (1969) and proposed by Bayer and Dimitriadis (2022). It tests that a series of ES forecasts, the risk measure of interest, is correctly specified with respect to the observed returns, i.e. whether the forecasts are systematically over- or under-estimated. The second approach relies on a loss function that jointly evaluates the VaR and ES predictions. The lower the loss, the more accurate the joint prediction of the VaR and ES prediction. The third approach performs a pair-wise weak dominance test of relevant models (with and without implied volatilities) across a class of loss functions as proposed by Ziegel et al. (2020).

[12] Ideally, the preferred approach is to identify reasons for outliers carefully. That is often impossible; the reasons for their presence are unknown, or extreme observations, although rare, are part of the underlying distribution. We used winsorization, a simple and practical approach to deal with extreme observations in day-to-day operations. However, to assess the impact of the winsorization on the resulting time series, we calculated the correlation between the winsorized and original time series. Among all daily, weekly, and monthly realized price measures (e.g., realized variance, positive and negative semivariance, median realized variance, jump component, signed jump, and integrated quarticity), the average Pearson's correlation is 0.9732 with a median value of 0.9970. In our data, winsorization resulted in only minor changes in the time series. We have also decided to test how the winsorization of implied volatilities would affect forecasting model parameter estimates. We have compared the implied volatility parameter estimates from the Dimitriadis and Bayer (2019) DB-IV and DB-RV-IV models (see Tables 2 and 3) for VaR and ES across all quantiles. We find that out of 72 coefficients (Panels B in Tables 2 and 3), 68 had the same sign. Among the 48 significant coefficients (as reported in our main results), 47 had the same sign. We thus conclude that within the Dimitriadis and Bayer (2019) model, the results are generally unaffected by winsorization.

[13] For example, if the predicted 0.01 quantile is equal to -2.5 and the next consecutive quantile of interest is 0.025, and the corresponding prediction is actually lower at -2.6, the latter value is substituted with $-2.5+0.0001$, which ensures non-crossing.

3.10.1. The strict ES regression-based backtest

We employ the strict ES backtest of Bayer and Dimitriadis (2022) to verify whether the ES measures are correctly specified. While ES backtesting necessarily has to address the problems of non-elicitability and non-identifiability (see e.g. Fissler & Ziegel, 2016; Fissler, Ziegel, & Gneiting, 2015; Gneiting, 2011), the Bayer and Dimitriadis (2022) strict ES backtest has the advantage of using only the history of realized returns and ES forecasts, without requiring the VaR forecasts for testing purposes.

Formally, the backtests are based on regressing the returns R_t on the ES forecasts $\hat{e}_t$. The general approaches to ES modeling (Dimitriadis & Bayer, 2019; Patton et al., 2019) are based on the joint regressions:

$$R_t = \mathbf{V}_t'\boldsymbol{\beta} + u_t^q, \qquad \text{and} \qquad R_t = \mathbf{W}_t'\boldsymbol{\gamma} + u_t^e \tag{26}$$

where $\mathbf{V}_t$ and $\mathbf{W}_t$ are some k-dimensional vectors, and the error terms should satisfy $VaR_{u_t^q}(\tau) = 0$ and $ES_{u_t^e}(\tau) = 0$ almost surely. In the case of the strict ES backtest, the ES forecasts $\hat{e}_t$ are used as covariates in both equations:

$$R_t = \beta_0 + \beta_1\hat{e}_t + u_t^q, \qquad \text{and} \qquad R_t = \gamma_0 + \gamma_1\hat{e}_t + u_t^e \tag{27}$$

Using this specification, Bayer and Dimitriadis (2022) test the hypothesis:

$$\mathbb{H}_0 : (\gamma_0, \gamma_1) = (0, 1), \qquad \text{against} \qquad \mathbb{H}_1(\gamma_0, \gamma_1) \neq (0, 1) \tag{28}$$

Under $\mathbb{H}_0$, $\hat{e}_t = ES_{R_t}(\tau)$ holds for the forecasts almost surely. The hypothesis test itself is conducted using a Wald-type test statistic given in Bayer and Dimitriadis (2022), who also provide the asymptotic theory for the joint VaR and ES regression under potential model misspecification.

3.10.2. Statistical evaluation of expected shortfall predictions

Note that while many models might be well-specified, their accuracy might differ. At the same time, some ill-specified models might show even higher accuracy than well-specified models. Backtesting as described above might not be enough to discriminate between competing models. We therefore complement our analysis by calculating the accuracy of forecasts via a loss function, as indicated by Fissler and Ziegel (2016) and Patton et al. (2019). The loss function evaluates the ability of a given model to jointly predict VaR and ES. The following loss function assumes that $ES_t(\tau) \leq VaR_{R_t}(\tau) < 0$, and that $L_t = I(R_t < VaR_{R_t}(\tau))$ is an indicator function. Then, the loss function given by Patton et al. (2019) is:

$$FZ_t^{(0)} = \frac{L_t(R_t - VaR_{R_t}(\tau))}{\tau ES_t(\tau)} + \frac{VaR_{R_t}(\tau)}{ES_t(\tau)} + \log(-ES_t(\tau)) - 1 \tag{29}$$

The most accurate models have the lowest average value of $FZ_t^{(0)}$. We follow the work of Gerlach et al. (2020), Lyócsa, Todorova, and Vŷrost (2021) and employ the model confidence set approach of Hansen et al. (2011), which allows us to identify a set of superior models within our initial set of 38 VaR and ES prediction models. We use the 90% and 75% confidence levels.

3.10.3. Weak dominance test

Many different scoring functions may be used to compare ES forecasts, and the literature does not provide a universally accepted and preferred choice of a single scoring function. We therefore evaluate a pair of ES forecasts using the weak dominance test of Ziegel et al. (2020), examining whether one forecast method dominates another under a whole relevant class of scoring functions.

Ziegel et al. (2020) first define a set of scoring functions they call consistent and show that the scoring functions in the widely used form introduced by Fissler and Ziegel (2016) are consistent (consequently, FZ^0, FZG, and AL (Taylor, 2019) are all consistent scoring functions). Next, they define the so-called elementary scores, given $\tau \in (0, 1)$, parameter $\eta \in \mathbb{R}$, return $R \in \mathbb{R}$, and $(v, e) \in \{(v, e) \in \mathbb{R}^2 : v \geq e\}$ as:

$$S_\eta(v, e, y) = \mathbb{1}\{\eta \leq e\}\left(\frac{1}{\tau}\mathbb{1}\{y \leq v\}(v - y) - (v - \eta)\right) + \mathbb{1}\{\eta \leq y\}(y - \eta) \tag{30}$$

The Ziegel et al. (2020) procedure provides a statistical test to evaluate the methods for forecasting expected shortfall. Forecasting method A, leading to forecasts V^A of VaR and E^A of ES, is said to be weakly dominant over forecasting method B, leading to forecasts V^B and E^B, if and only if:

$$\mathbb{E}\left(S_\eta(V^A, E^A, R)\right) \leq \mathbb{E}\left(S_\eta(V^B, E^B, R)\right), \qquad \text{for all } \eta \in \mathbb{R}. \tag{31}$$

4. Results

In the following sections, we present our empirical results. First, we describe some stylized facts observed in our data. Next, we provide in-sample evidence from selected models about the different roles of implied and realized volatility used in the tail prediction of the return distribution (VaRs). The third subsection presents our key results based on a horse race among a battery of competing models, searching for well-specified models that also belong to the set of superior models according to a chosen loss function.

4.1. Stylized facts on realized and implied volatility

In Figs. 2 and 3, we plot the EUR/USD exchange rate and its returns, respectively. The blue vertical lines correspond to selected events related to the US, while the red lines correspond to events related to the eurozone's economy and institutions. We observe a weakening of the US dollar (against the euro) that followed the onset of the financial crisis and, similarly, a weakening of the euro (against the US dollar) during the economic and debt crisis in Greece. Moreover, a sharp weakening of the US dollar is visible during the first wave of the Covid-19 pandemic at the end of our sample. The return series (Fig. 3) shows several clusters of high-volatility periods; notably, the financial crisis appears to represent the most volatile period for the EUR/USD exchange rate. However, Fig. 3 also shows that extreme price movements are spread

Table 1
Data characteristics.

Name	Notation	Mean	SD	2.5%	50.0%	97.50%	$\rho(1)$	$\rho(5)$	$\rho(22)$	EL	Skew.	Kurt.
Panel A: EUR/USD daily return												
Daily return	R_t	0.002	0.592	−1.229	0.009	1.172	−0.010	0.020	0.000	0.700	0.070	5.070
Panel B: Realized measures												
Realized var.	RV_t	88.37	85.59	12.97	65.33	322.65	0.67	0.61	0.42	0.00	3.44	20.04
Positive realized semivar.	PV_t	44.37	44.58	6.12	31.85	169.76	0.60	0.55	0.36	0.00	3.43	19.70
Negative realized semivar.	NV_t	43.65	43.56	6.35	31.11	167.15	0.65	0.59	0.41	0.00	3.27	17.49
Continuous comp.	CC_t	76.98	76.20	11.70	56.15	294.34	0.73	0.65	0.46	0.00	3.64	22.31
Jump comp.	JC_t	11.74	18.43	0.00	6.11	59.96	0.09	0.14	0.06	0.00	4.08	25.31
Signed jump	SJ_t	1.25	31.79	−56.15	0.04	56.34	0.04	0.01	−0.01	0.03	5.51	92.50
Measurement error component	CQ_t	1272.18	3972.92	13.22	325.72	8993.39	0.45	0.34	0.21	0.00	8.08	82.99
Panel C: Implied volatility measures												
Implied volatility with daily mat.	IV_t^D	120.56	131.00	15.60	88.69	453.08	0.68	0.74	0.52	0.00	4.70	40.98
Implied volatility with weekly mat.	IV_t^W	99.02	91.01	19.56	76.58	337.36	0.96	0.84	0.73	0.00	3.99	28.06
Implied volatility with monthly mat.	IV_t^M	96.60	79.09	22.09	78.10	314.00	0.99	0.94	0.84	0.00	3.42	20.24

Notes: SD denotes standard deviation; 2.5%, 50%, and 97.5% are percentiles; $\rho(.)$ is the autocorrelation of the given order; EL represents the p-value of the Escanciano and Lobato (2009) automatic portmanteau test of serial correlation; and Skew. and Kurt. are skewness and kurtosis, respectively. Statistics are computed over the whole sample of 4522 observations, starting from June 4, 2003, and ending on October 10, 2020. The out-of-sample analysis covers the last 3022 observations.

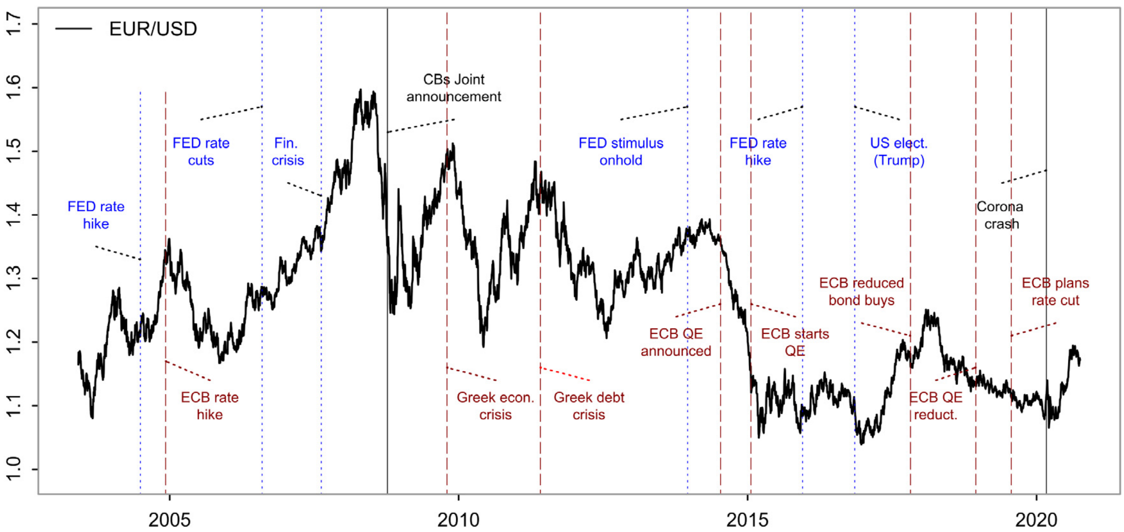

Fig. 2. EUR/USD FX rate with selected major events.

over the whole sample period and are not limited to just a few high-volatility periods. While daily returns are symmetric (see Table 1), they are also subject to extremes in both tails of the return distribution, as suggested by the kurtosis of 5.07.

As seen in Table 1, the data exhibit several properties well known in the volatility modeling literature. Most notably, the realized variance shows right skewness (3.44), excess kurtosis (20.04), and long-memory properties, as even at the 22nd lag, autocorrelation is still 0.42. As expected, the price variation due to the discontinuous price movements shows little persistence (e.g. Andersen, Bollerslev, & Diebold, 2007; Lyócsa, Plíhal, & Vỳrost, 2021), regardless of whether it is estimated via JC_t as in Andersen et al. (2012) or SJ_t as in Patton and Sheppard (2015). After we remove the variation due to price discontinuities from the overall price variation, the resulting continuous component, CC_t, shows longer memory than the realized variance. This finding also corresponds to known stylized facts about the continuous volatility component. Our models also use the positive and negative price variation. In the case of the EUR/USD exchange rate, the price variation due to dollar weakening corresponds to the positive realized semivariance, PV_t. This component is slightly less persistent than the price variation corresponding to the weakening of the euro, the negative realized semivariance, NV_t.

Finally, Panel C in Table 1 shows the characteristics of daily, weekly, and monthly implied volatilities (IVs). As opposed to realized measures, IVs are forward looking and incorporate the volatility risk premium; thus, it is not surprising that the daily IV overestimates the observed realized variance. Moreover, the daily IV is much more volatile (SD at 131 as opposed to 85.59 for realized variance), persistent (autocorrelation of 0.52 at the 22nd lag), skewed and heavy tailed. The weekly and monthly

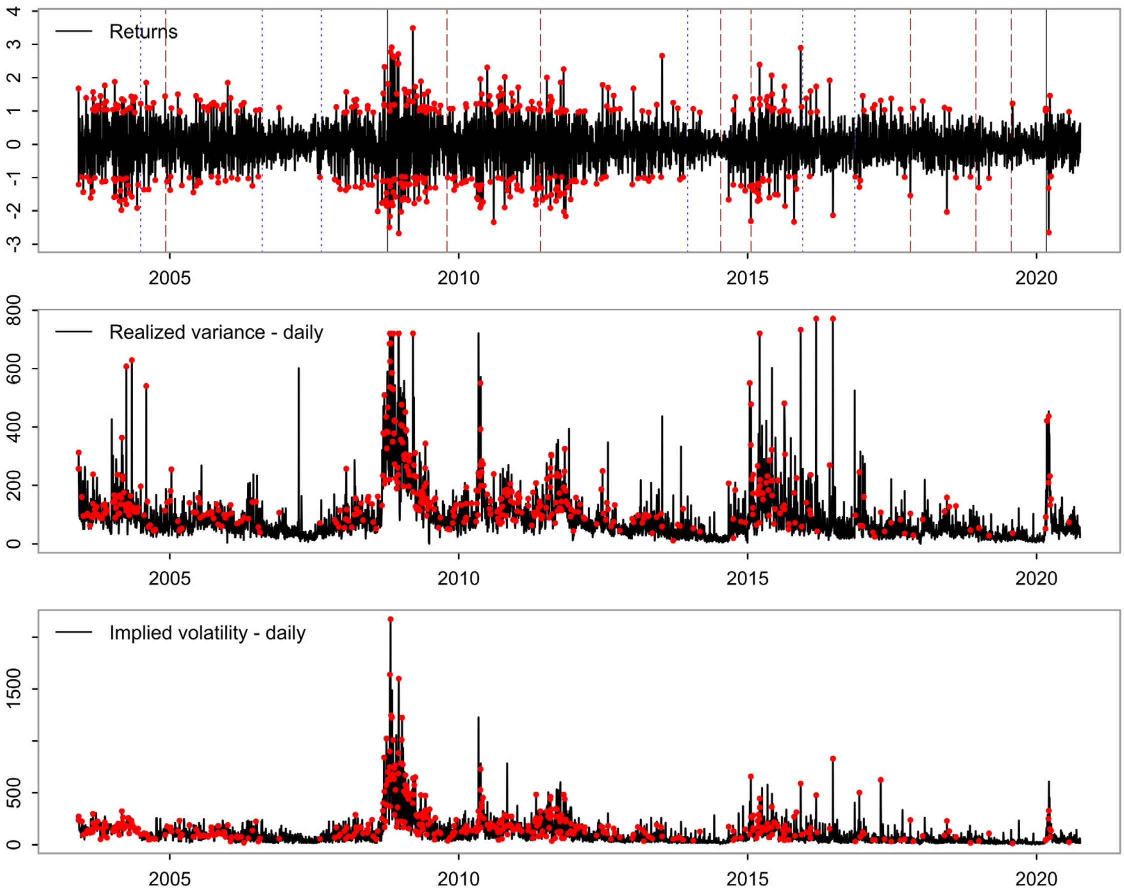

Fig. 3. EUR/USD returns, realized and implied volatilities.

implied volatilities appear to smooth out the uncertainty over future price variation and are much more persistent. It therefore appears that implied volatility overstates observed price variation, consistent with the literature on the volatility risk premium, as a consequence of excess demand for options protecting against large price swings. The volatility risk premium is correspondingly pronounced for the EUR/USD exchange rate, as investors holding either EUR or USD demand protection against both tails of the return distribution.

Fig. 3 also showcases the days when returns are below (above) the 5th (95th) historical quantile (red points). Interestingly, many "hits" are captured by both the realized and implied volatility measures, suggesting that extreme price variation is related to extreme daily price movements as well. The correlation between RV_t and IV_t^D is 0.79, which suggests an overlap of information content between these two market uncertainty measures. The following section investigates whether the unique part of the two measures leads to different expected shortfall predictions.

4.2. Relevance of implied volatility: In-sample evidence

Tables 2 and 3 provide parameter estimates from three specifications of the joint VaR and ES regression model of Dimitriadis and Bayer (2019). The DB-RV model uses only the historical realized variance, the DB-IV model uses only the implied volatility components, and the DB-RV-IV model combines both. The models reported in Tables 2 and 3 differ only with respect to the quantile of interest. By comparing the coefficients and optimized loss values, we might observe whether the implied volatility components supplement or complement the measures of realized price variation.

Table 2 presents left-tail daily return predictions, i.e. extreme weakening of the euro. Regarding the DR-RV model that uses only a historical price variation measure, we see that VaR and ES are both mainly driven by the weekly and monthly components that have negative coefficients; thus, in periods of higher volatility, extreme price declines are larger. The weekly component seems to be stronger for more extreme price declines. As both weekly and monthly volatility are highly persistent, this suggests that extreme price declines are also likely clustered in high-volatility periods.

Under the DB-IV model, the optimized loss achieved lower values than under the DB-RV model. This suggests that implied volatility is more informative about future price declines. The losses declined from 1.627 ($\tau = 0.01$), 1.581 ($\tau = 0.025$) and 1.543 ($\tau = 0.05$) to 1.611, 1.572 and 1.536. This observation is further strengthened by observing that all implied volatilities are loaded negatively onto extreme returns, as well as VaR and ES. The distinctive behavior of the past and forward-looking price variation measures is visible in that, contrary to the realized measures in the DB-RV model, short-term implied volatilities are also relevant in the DB-IV model.

If the two measures of price variation complement each other, combining both into a single model might improve the overall fit. However, as the results from the DB-RV-IV model show, this does not appear to be the case when the US dollar weakens. The fit improves only negligibly, while systematically (across VaR and ES), only the long-memory (monthly) realized measures tend to matter. It appears that realized and implied volatilities do not complement one another. Instead, with regard to next-day extreme returns, the implied volatility likely encompasses information from the realized measures, particularly for $\tau = 0.05$.

In Table 3, we observe the role of the realized and implied measures for the weakening of the US dollar. In the DB-RV models, VaR is driven by the weekly and monthly realized measures with expected positive signs. The results for ES are quite different, with negative signs for the daily ($1 - \tau = 0.95, 0.975$ and 0.99) and monthly realized volatilities ($1 - \tau = 0.95$ and 0.975). These results are interesting, as they show that past levels of volatility have different effects on the VaR event and the loss incurred.

The role of implied volatility seems to be weaker when the US dollar weakens (DB-IV model in Table 3), as we observe only a minor decline in the optimal loss for the DB-IV model as opposed to the DB-RV model. As with the realized measures, we again observe that VaR and ES are impacted differently with IVs. We observe, as expected, a mostly positive effect of IV on VaR, but the results are mixed with respect to ES. As DB-RV-IV shows minor improvement to the fit of the model, we find little evidence that implied volatility encompasses or complements realized variance.

What these results underscore is the distinctive role of historical and forward-looking price variation measures in understanding tail risks. Implied volatility is more relevant for the dollar strengthening than for the dollar weakening. Moreover, for dollar weakening, realized (DB-RV model) or implied (DB-IV model) variables show different effects on the VaR and ES, implying that the same observed market uncertainty can have slightly different implications for the next day's VaR and ES; i.e. a different response is expected when managing risky assets.

4.3. Relevance of implied volatility: Out-of-sample evidence

The previous section suggests that IV should be more useful for US dollar strengthening for the left tail of the EUR/USD return distribution. However, it is unclear whether in-sample evidence translates into the out-of-sample framework. In Fig. 4, we present ES forecasts from one of the best-performing prediction models within their category.[14] The blue line corresponds to the models that do not employ information from implied volatility, while the red line corresponds to predictions from the models that do utilize this information. There are periods when the two lines do not entirely overlap, suggesting different levels of risk assessment. The implied volatility model forecasts also appear to have more frequent spikes. What is interesting to observe is how the predictions behave when volatility regimes change. For example, for the period before 2015, there is a sudden increase in volatility leading to multiple hits, i.e. tail events. The non-implied volatility model's predictions are more conservative, having a tighter band prior to this event.

In Table 4, we report p-values from the joint hypothesis on the intercept and the slope of the strict ES backtesting procedure of Bayer and Dimitriadis (2022). The non-rejection of the hypothesis that the intercept and slope are equal to 0 and 1, respectively, suggests that we can expect that the forecasting model generates unbiased forecasts. Table 4 shows the results for each out-of-sample model. For example, in Panel A, we report results from historical simulation distribution-based models that ignore time variation in return volatility. Across multiple quantiles (except for those predicting ES at $1 - \tau = 0.99$), these models do not appear to behave well.

Accounting for time-varying mean and volatility through an EGARCH model (Panel B) leads to well-specified ES predictions. The joint hypothesis is never rejected, even at the 0.10 significance level. Otherwise, notable differences between the EGARCH-class models (EGARCH-JSU and EGARCH-SSK in Panel B) are not found. Comparing these results with the historically based simulations, it seems that having flexible return distributions is not enough and instead that one needs to allow for time variation in return and volatilities to achieve well-specified predictions of ES. This result also shows why GARCH models are still popular and used as a benchmark in the literature.

After augmenting the EGARCH models with realized (EGARCH-RV in Panel C) or implied (EGARCH-IV in Panel F) volatility measures, we surprisingly observe a different result: rejection of the joint hypothesis. This result holds across almost all quantiles, although we have to add that the rejection tends to happen at between the 0.01 and 0.10 significance levels. As rejection is observed for both measures, while it is not found under all other models that use realized and implied measures of volatility, it might be a model-specific issue. It also remains to be seen whether the not-well-specified EGARCH-RV and EGARCH-IV forecasts of ES translate into worse accuracies under different backtesting criteria.

Approximate ES forecasts from QR models that use only realized variance measures (Panel D) are generally well-specified. With implied volatilities, the results are slightly different (Panel G) showing a well-specified forecast only in the right tail ($1 - \tau = 0.975$ and 0.99).

Under the Dimitriadis and Bayer (2019) model, we expect non-rejection because the forecasts were generated using the same joint VaR and ES estimation framework used in this backtesting procedure. This is indeed the case, as most of the forecasts with realized (Panel E) and

14 According to the FZ^0 loss function.

Table 2
Role of implied volatility measures for predicting 0.05, 0.025, and 0.01 VaR and ES of daily EUR/USD returns: Joint VaR and ES regression framework.

	τ = 0.01						τ = 0.025						τ = 0.05					
	VaR			ES			VaR			ES			VaR			ES		
	DB-RV	DB-IV	DB-RV-IV	DB-RV	DB-IV	DB-RV-IV	DB-RV	DB-IV	DB-RV-IV	DB-RV	DB-IV	DB-RV-IV	DB-RV	DB-IV	DB-RV-IV	DB-RV	DB-IV	DB-RV-IV
Intercept	$\mathbf{-0.65}^{d}$	$\mathbf{-0.58}^{d}$	$\mathbf{-0.58}^{d}$	$\mathbf{-0.96}^{d}$	$\mathbf{-0.76}^{d}$	$\mathbf{-0.80}^{d}$	$\mathbf{-0.57}^{d}$	$\mathbf{-0.55}^{d}$	$\mathbf{-0.51}^{d}$	$\mathbf{-0.79}^{d}$	$\mathbf{-0.72}^{d}$	$\mathbf{-0.60}^{d}$	$\mathbf{-0.48}^{d}$	$\mathbf{-0.44}^{d}$	$\mathbf{-0.41}^{d}$	$\mathbf{-0.69}^{d}$	$\mathbf{-0.64}^{d}$	$\mathbf{-0.61}^{d}$
Panel A: Realized measures ($\times$**1000**)																		
RV_t^D	-1.52		$\mathbf{0.27}^{c}$	-1.72		$\mathbf{0.71}^{c}$	$\mathbf{-0.41}^{b}$		0.64	$\mathbf{-0.76}^{b}$		0.87	$\mathbf{-0.47}^{c}$		-0.10	$\mathbf{-0.43}^{b}$		0.38
RV_t^W	$\mathbf{-4.69}^{d}$		$\mathbf{-1.44}^{c}$	-2.88		$\mathbf{-0.47}^{b}$	$\mathbf{-3.39}^{d}$		-1.63	$\mathbf{-3.82}^{c}$		$\mathbf{-2.17}^{c}$	$\mathbf{-1.41}^{d}$		-0.15	-3.06		-1.56
RV_t^M	$\mathbf{-1.66}^{c}$		$\mathbf{-1.67}^{c}$	$\mathbf{-2.80}^{c}$		$\mathbf{-2.94}^{c}$	$\mathbf{-2.55}^{d}$		-1.85	-2.14		-1.49	$\mathbf{-3.13}^{d}$		-0.37	$\mathbf{-2.32}^{d}$		$\mathbf{-0.63}^{c}$
Panel B: Implied volatility measures ($\times$**1000**)																		
IV_t^D		$\mathbf{-3.23}^{d}$	$\mathbf{-3.22}^{d}$		$\mathbf{-3.98}^{d}$	$\mathbf{-4.46}^{d}$		$\mathbf{-1.80}^{d}$	$\mathbf{-1.67}^{d}$		$\mathbf{-2.66}^{d}$	$\mathbf{-3.07}^{d}$		$\mathbf{-1.59}^{d}$	$\mathbf{-1.59}^{d}$		$\mathbf{-2.19}^{d}$	$\mathbf{-2.34}^{d}$
IV_t^W		$\mathbf{-0.89}^{b}$	-0.64		-1.19	-1.06		$\mathbf{-1.21}^{c}$	$\mathbf{-1.00}^{c}$		$\mathbf{-0.58}^{b}$	0.15		$\mathbf{-1.27}^{c}$	$\mathbf{-0.91}^{c}$		$\mathbf{-1.82}^{c}$	$\mathbf{-0.79}^{c}$
IV_t^M		$\mathbf{-3.21}^{b}$	-0.94		$\mathbf{-2.06}^{b}$	1.26		$\mathbf{-2.05}^{c}$	-0.45		$\mathbf{-2.65}^{c}$	-1.36		-1.80	-1.85		$\mathbf{-1.01}^{c}$	-0.43
Loss	1.627	1.611	1.609				1.581	1.572	1.572				1.543	1.536	1.536			
Loss SD	0.047	0.080	0.024				0.065	0.052	0.018				0.053	0.022	0.057			

Notes: Values correspond to the estimated coefficients from the joint VaR and ES regression framework of Dimitriadis and Bayer (2019). Significance is based on stationary bootstrap of Politis and Romano (1994), with random block length drawn from a geometric distribution with expected value given as in Politis and White (2004), and Patton, Politis, and White (2009) with 1000 bootstrap samples. Coefficients in Panel A and B are multiplied by 10^3. Letters 'a', 'b', 'c', and 'd' denote significance at the 10%, 5%, 1%, and 0.1% levels, respectively. Bolded coefficients were significant at least at the 10% level. The loss of the converged model is denoted as Loss, and Loss SD denotes the standard deviation of the converged models over the bootstrap samples.

Table 3
Role of implied volatility measures for predicting 0.95, 0.975, and 0.99 VaR and ES of daily EUR/USD returns: Joint VaR and ES regression framework.

	$1-\tau$ = 0.95						$1-\tau$ = 0.975						$1-\tau$ = 0.99					
	VaR			ES			VaR			ES			VaR			ES		
	DB-RV	DB-IV	DB-RV-IV	DB-RV	DB-IV	DB-RV-IV	DB-RV	DB-IV	DB-RV-IV	DB-RV	DB-IV	DB-RV-IV	DB-RV	DB-IV	DB-RV-IV	DB-RV	DB-IV	DB-RV-IV
Intercept	$\mathbf{0.51}^{d}$	$\mathbf{0.51}^{d}$	$\mathbf{0.51}^{d}$	$\mathbf{-0.02}^{b}$	$\mathbf{-0.03}^{c}$	-0.01	$\mathbf{0.60}^{d}$	$\mathbf{0.59}^{d}$	$\mathbf{0.60}^{d}$	$\mathbf{0.00}^{b}$	$\mathbf{-0.01}^{a}$	0.00	$\mathbf{0.89}^{d}$	$\mathbf{0.84}^{d}$	$\mathbf{0.80}^{d}$	$\mathbf{0.01}^{c}$	$\mathbf{-0.01}^{b}$	$\mathbf{0.01}^{c}$
Panel A: Realized measures ($\times$**1000**)																		
RV_t^D	$\mathbf{-0.26}^{b}$		-0.67	$\mathbf{-0.44}^{d}$		$\mathbf{-0.59}^{d}$	$\mathbf{-0.26}^{b}$		0.06	$\mathbf{-0.45}^{c}$		$\mathbf{-0.59}^{d}$	0.08		-0.73	$\mathbf{-0.42}^{d}$		$\mathbf{-0.57}^{d}$
RV_t^W	$\mathbf{2.20}^{d}$		$\mathbf{1.41}^{d}$	$\mathbf{0.05}^{b}$		$\mathbf{-0.07}^{b}$	$\mathbf{2.21}^{c}$		0.81	$\mathbf{0.10}^{b}$		$\mathbf{-0.06}^{b}$	4.06		$\mathbf{1.90}^{c}$	$\mathbf{0.04}^{b}$		$\mathbf{0.05}^{b}$
RV_t^M	$\mathbf{2.63}^{d}$		$\mathbf{1.10}^{c}$	$\mathbf{-0.11}^{b}$		$\mathbf{-0.70}^{c}$	$\mathbf{3.82}^{d}$		0.64	$\mathbf{-0.04}^{b}$		$\mathbf{-0.62}^{c}$	$\mathbf{1.47}^{c}$		$\mathbf{-1.66}^{c}$	$\mathbf{0.08}^{b}$		$\mathbf{-0.63}^{c}$
Panel B: Implied volatility measures ($\times$**1000**)																		
IV_t^D		$\mathbf{1.40}^{d}$	$\mathbf{1.49}^{d}$		-0.18	-0.13		$\mathbf{1.45}^{c}$	$\mathbf{1.55}^{c}$		$\mathbf{-0.18}^{b}$	-0.09		$\mathbf{1.82}^{b}$	1.94		$\mathbf{-0.14}^{b}$	-0.07
IV_t^W		0.69	-0.19		-0.23	0.28		$\mathbf{1.26}^{b}$	$\mathbf{-0.09}^{b}$		$\mathbf{-0.13}^{c}$	0.30		$\mathbf{3.49}^{c}$	$\mathbf{3.09}^{b}$		$\mathbf{-0.25}^{b}$	0.18
IV_t^M		1.63	$\mathbf{0.68}^{c}$		0.17	$\mathbf{0.65}^{c}$		$\mathbf{2.60}^{c}$	$\mathbf{2.16}^{c}$		$\mathbf{0.15}^{b}$	$\mathbf{0.60}^{c}$		$\mathbf{-0.25}^{b}$	$\mathbf{0.81}^{c}$		$\mathbf{0.34}^{a}$	$\mathbf{0.69}^{c}$
Loss	1.268	1.268	1.267				1.260	1.260	1.260				1.255	1.254	1.254			
Loss SD	0.040	0.094	0.019				0.038	0.044	0.023				0.092	0.046	0.020			

Notes: Values correspond to the estimated coefficients from the joint VaR and ES regression framework of Dimitriadis and Bayer (2019). Significance is based on stationary bootstrap of Politis and Romano (1994), with random block length drawn from a geometric distribution with expected value given as in Politis and White (2004), and Patton et al. (2009) with 1000 bootstrap samples. Coefficients in Panel A and B are multiplied by 10^3. Letters 'a', 'b', 'c', and 'd' denote significance at the 10%, 5%, 1%, and 0.1% levels. Bolded coefficients were significant at least at the 10% level. The loss of the converged model is denoted as Loss, and Loss SD denotes the standard deviation of the converged models over the bootstrap samples.

implied (Panel H) measures are well-specified. However, it is obvious that non-rejection under implied volatility models is much less certain (lower p-values).

To summarize, we find well-specified models under both realized and implied volatility models. However, our results imply that if well-specified ES forecasts are of interest, a forecaster faces considerable model choice uncertainty. We attempt to mitigate model choice uncertainty with combination forecasts. The results in Panel I nicely match previous results, as we see that QR models are more susceptible to rejecting the hypothesis of well-specified forecasts, particularly for models with implied volatilities. This is not the case for the DB class of models. A safe approach is to use a combination across a broad class of models, i.e. ES forecasts from all implied or realized volatility models.

The results reported above show that we have multiple forecasts within each class of models (different models, with or without implied volatilities), that lead to well-specified ES predictions. However, risk managers might prefer models that lead not only to well-specified but, under a specific loss function, also to accurate predictions. To evaluate accuracy, we use the FZ^0 loss function of Patton et al. (2019) to jointly evaluate the VaR and ES predictions. In Table 5, we report the average FZ^0 loss values, together with the † and ‡ symbols that highlight whether the model belongs to the sets of superior models created with a 90% or a 75% (more strict → fewer models) confidence level. We also rank models from the one with the lowest average loss (having the rank of 1) to the one with the highest (having the rank of 38). The results reported in Table 5 allow us to compare not only whether the models belong to the superior set of models but also whether there is an increase in the accuracy of one model relative to that of the others.

Forecasts based on historical simulation from a specific distribution (Panel A in Table 5) are not competitive. Among the 38 models, they are often the worst performing. This is indicated by their higher average loss and lower ranking. Comparing different EGARCH model specifications complements the results reported in Table 4. Across all quantiles, the simpler EGARCH-JSU and EGARCH-SSK (Panel C) perform better than augmented versions with realized (Panel C) or implied (Panel F) volatility measures. In fact, when we examine the weakening of the US dollar ($1-\tau = 0.95$, 0.975, and 0.99), the simple EGARCH models compete with the most accurate models.

The forecasts from the QR and DB models differ in two dimensions: (i) whether implied volatilities are included, and (ii) whether we consider the left or right tail of the

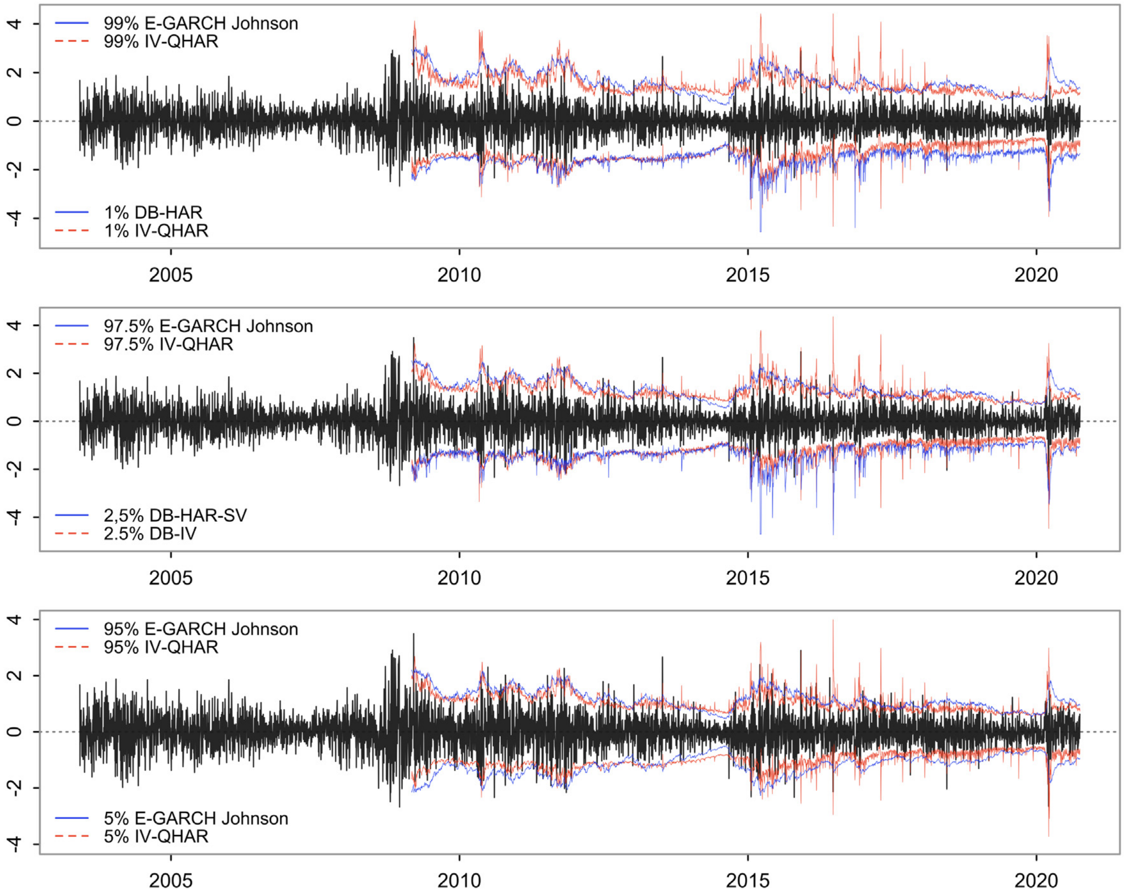

Fig. 4. Comparison of best-performing standard and implied volatility ES models.

EUR/USD distribution. With realized measures only (Panels D and E), both model classes perform even worse than the simpler EGARCH model—an interesting result, given that QR and DB use more accurate high-frequency volatility estimates.[15] The inclusion of implied volatility (Panels G and H) led to more accurate predictions under the FZ^0 loss function. This is one of our key results. We observe that both the QR and DB models led to forecasts that are in the set of superior models and that rank among the best. The result is particularly convincing for the QR models that have a superior ranking even in the right tail, where the DB models perform worse. Generally, we observe that the inclusion of the implied volatility measures is less useful when the US dollar is weakening (right tail). With respect to the FZ^0 loss function, the predictions from the implied volatility models under QR and DB (not EGARCH) are clearly dominant. The best implied volatility model improves the accuracy of the best non-implied volatility model by 28.9% ($\tau = 0.01$), 39.10% ($\tau = 0.025$), 55.5% ($\tau = 0.05$), 11.70% ($1-\tau = 0.95$), 7.30% ($1-\tau = 0.975$), and 11.70% ($1-\tau = 0.99$).[16]

An interesting observation that we also observed with the strict ES backtesting framework is that simpler models with fewer parameters tend to perform better. Note that among non-implied volatility models, EGARCH, QR-RV, and DB-RV perform well, while among implied volatility models, the QR-IV and DB-IV models perform well. This result is in line with the idea that in an out-of-sample context (and for modeling rare events especially), increasing the number of model parameters might eventually lead to a detrimental variance–bias tradeoff.

In practice, knowing which model performs best is not possible. Instead, we recommend relying on predictions based on combination forecasts (Panel I) and comparisons of the average losses across multiple combination forecasts. The implied volatility-based combination forecasts led to improvements of 34.7% ($\tau = 0.01$), 34.3% ($\tau = 0.025$), 60.6% ($\tau = 0.05$), 17.1% ($1-\tau = 0.95$), 7.7%

15 Recently, Lyócsa, Todorova, and Výrost (2021) found that for four energy assets, QR models predict ES with low-frequency measures of price variation approximately as well as QR models with high-frequency measures.

16 In absolute terms, the improvements correspond to 0.111, 0.080, 0.045 for 0.01, 0.025, 0.05 in the left tail, and to 0.013, 0.020, 0.055 for 0.95, 0.975, 0.99 in the right tail of the return distribution.

Table 4
Strict regression-based expected shortfall backtesting.

Model	0.01 ES	0.025 ES	0.05 ES	0.95 ES	0.975 ES	0.99 ES
Panel A: Historical distribution-based expected shortfall forecasting models						
SSK	0.000	0.013	0.004	0.014	0.021	0.130
JSU	0.001	0.015	0.011	0.007	0.037	0.263
Panel B: Dynamic latent volatility expected shortfall forecasting models						
EGARCH-JSU	0.447	0.228	0.248	0.384	0.497	0.175
EGARCH-SSK	0.439	0.242	0.289	0.417	0.503	0.174
Panel C: Dynamic latent volatility expected shortfall forecasting models with realized volatility						
EGARCH-RV-SSK	0.035	0.050	0.045	0.013	0.040	0.054
EGARCH-RV-JSU	0.045	0.033	0.037	0.022	0.044	0.044
Panel D: Quantile regression approximate expected shortfall forecasting models with realized volatility						
QR-RV	0.065	0.138	0.091	0.134	0.292	0.169
QR-RV-CJ	0.014	0.049	0.071	0.090	0.221	0.112
QR-RV-CQ	0.895	0.808	0.416	0.132	0.257	0.082
QR-RV-SV	0.025	0.094	0.035	0.064	0.134	0.075
QR-RV-SJ	0.018	0.189	0.074	0.098	0.155	0.084
Panel E: Dimitriadis and Bayer's (2019) expected shortfall forecasting models with realized volatility						
DB-RV	0.329	0.758	0.755	0.983	0.869	0.249
DB-RV-CJ	0.395	0.699	0.913	0.950	0.513	0.245
DB-RV-SV	0.118	0.739	0.695	0.716	0.389	0.098
DB-RV-CQ	0.356	0.179	0.549	0.760	0.908	0.143
DB-RV-SJ	0.167	0.727	0.855	0.928	0.510	0.148
Panel F: Dynamic latent volatility xpected shortfall forecasting models with implied volatility						
EGARCH-IV-SSK	0.037	0.095	0.050	0.021	0.039	0.043
EGARCH-IV-JSU	0.027	0.050	0.035	0.021	0.063	0.040
Panel G: Quantile regression approximate expected shortfall forecasting models with implied volatility						
QR-IV	0.097	0.026	0.008	0.051	0.133	0.350
QR-IV-RV	0.013	0.013	0.005	0.030	0.098	0.125
QR-IV-CJ	0.005	0.009	0.003	0.018	0.096	0.140
QR-IV-CQ	0.016	0.007	0.003	0.033	0.076	0.105
QR-IV-SV	0.001	0.003	0.003	0.008	0.059	0.098
QR-IV-SJ	0.007	0.013	0.005	0.010	0.071	0.132
Panel H: Dimitriadis and Bayer's (2019) expected shortfall forecasting models with implied volatility						
DB-IV	0.459	0.354	0.164	0.922	0.649	0.580
DB-IV-CJ	0.363	0.706	0.510	0.224	0.244	0.141
DB-IV-CQ	0.036	0.246	0.471	0.427	0.146	0.112
DB-IV-SV	0.130	0.494	0.519	0.213	0.150	0.110
DB-IV-SJ	0.020	0.320	0.410	0.187	0.098	0.120
Panel I: Combination forecasts						
EGARCH realized volatility	0.034	0.034	0.038	0.010	0.043	0.051
EGARCH implied volatility	0.041	0.061	0.035	0.030	0.051	0.051
QR realized volatility	0.120	0.102	0.086	0.110	0.201	0.137
QR implied volatility	0.016	0.008	0.006	0.016	0.096	0.161
DB realized volatility	0.317	0.763	0.890	0.953	0.754	0.235
DB implied volatility	0.907	0.692	0.330	0.540	0.355	0.269
All realized volatility and historical models	0.454	0.587	0.782	0.649	0.551	0.228
All realized volatility	0.957	0.691	0.143	0.838	0.784	0.517
All implied volatility	0.624	0.256	0.091	0.368	0.378	0.444

Notes: Values in the table correspond to the p-values of the strict expected shortfall regression-based backtest hypothesis of Bayer and Dimitriadis (2022), where under the null, the intercept and the slope coefficients are 0 and 1, respectively. A non-rejection of the null-hypothesis implies a well-specified forecast of the expected shortfall. The p-values are based on the stationary bootstrap of Politis and Romano (1994), with random block lengths drawn from a geometric distribution with the expected value given as in Politis and White (2004), and Patton et al. (2009) with 1000 bootstrap samples. In our application, the bootstrapped p-values are in almost 90% of cases lower as the asymptotic p-values. That is, we are more conservative, as it is more difficult not to reject the null of a well-specified model.

$(1-\tau = 0.975)$, and 10.8% $(1-\tau = 0.99)$ over the non-implied volatility-based combination forecasts.[17] Overall, these results provide unambiguous evidence in favor of implied volatility models; it appears that under the FZ^0 loss function, the information encompassed in implied volatility is useful for predicting the next-day EUR/USD VaR and ES.

Interestingly, we can also observe evidence of an asymmetric effect, as the implied volatility models appear more useful for predicting the extent of euro weakening, i.e. the left tail. Note the average losses on the tails and the improvements of the implied volatility models, which differ considerably for the left and right tails. Thus, with regard to losses, the benefits from incorporating implied volatility are not distributed equally among investors;

[17] In absolute terms, the improvements correspond to 0.127, 0.075, 0.048 for 0.01, 0.025, 0.05 in the left tail, and to 0.019, 0.022, 0.054 for 0.95, 0.975, 0.99 in the right tail of the return distribution.

Table 5
Average values of the FZ^0 loss function, model confidence set, and average ranks.

Model	0.01 ES			0.025 ES			0.05 ES			LTR	0.95 ES			0.975 ES			0.99 ES			RTR
	FZ^0	M^*	R	FZ^0	M^*	R	FZ^0	M^*	R		FZ^0	M^*	R	FZ^0	M^*	R	FZ^0	M^*	R	
Panel A: Historical distribution-based expected shortfall forecasting models																				
SSK	0.626		36	0.422		36	0.237		37	36.333	0.211		37	0.403		37	0.614		36	36.667
JSU	0.642		38	0.426		37	0.237		37	37.333	0.214		38	0.406		38	0.614		36	37.333
Panel B: Dynamic latent volatility expected shortfall forecasting models																				
EGARCH-JSU	0.508		19	0.290		19	0.126		15	17.667	0.123		9	0.301		8	0.530		11	9.333
EGARCH-SSK	0.508		19	0.290		19	0.126		15	17.667	0.123		9	0.302		9	0.530		11	9.667
Panel C: Dynamic latent volatility expected shortfall forecasting models with realized volatility																				
EGARCH-RV-SSK	0.532		30	0.313		31	0.153		30	30.333	0.143		26	0.328		26	0.558		17	23.000
EGARCH-RV-JSU	0.530		28	0.315		33	0.154		33	31.333	0.145		32	0.330		31	0.558		17	26.667
Panel D: Quantile regression approximate expected shortfall forecasting models with realized volatility																				
QR-RV	0.526		25	0.297		23	0.127		17	21.667	0.136	‡	18	0.316	‡	16	0.567		22	18.667
QR-RV-CJ	0.549		32	0.311		29	0.131		21	27.333	0.141	‡	22	0.321	‡	20	0.589		29	23.667
QR-RV-CQ	0.633		37	0.371		35	0.175		35	35.667	0.139		20	0.327		23	0.597		32	25.000
QR-RV-SV	0.557		33	0.303		26	0.139		23	27.333	0.140		21	0.332		32	0.589		29	27.333
QR-RV-SJ	0.586		34	0.322		34	0.145		25	31.000	0.142		23	0.329		29	0.585		27	26.333
Panel E: Dimitriadis and Bayer's (2019) expected shortfall forecasting models with realized volatility																				
DB-RV	0.493		14	0.291		21	0.135		22	19.000	0.142		23	0.312	‡	13	0.592		31	22.333
DB-RV-CJ	0.517		23	0.301		25	0.145		25	24.333	0.152		34	0.328		26	0.598		33	31.000
DB-RV-SV	0.516		22	0.286		16	0.158		34	24.000	0.153		35	0.340		34	0.606		35	34.667
DB-RV-CQ	0.509		21	0.633		38	0.179		36	31.667	0.146		33	0.320	‡	18	0.629		38	29.667
DB-RV-SJ	0.502		18	0.293		22	0.153		30	23.333	0.154		36	0.324		22	0.602		34	30.667
Panel F: Dynamic latent volatility expected shortfall forecasting models with implied volatility																				
EGARCH-IV-SSK	0.525		24	0.308		27	0.148		27	26.000	0.144		29	0.328		26	0.554		14	23.000
EGARCH-IV-JSU	0.529		27	0.311		29	0.148		27	27.667	0.143		26	0.327		23	0.554		14	21.000
Panel G: Quantile regression approximate expected shortfall forecasting models with implied volatility																				
QR-IV	0.382	‡	3	0.212	‡	3	0.081	‡	2	2.667	0.110	‡	1	0.281	‡	1	0.475	‡	1	1.000
QR-IV-RV	0.436	‡	8	0.240	‡	8	0.092	‡	8	8.000	0.114	‡	3	0.286	‡	3	0.490	‡	2	2.667
QR-IV-CJ	0.486	‡	12	0.264	‡	12	0.104	‡	12	12.000	0.121	‡	7	0.294	‡	6	0.504	‡	6	6.333
QR-IV-CQ	0.470	‡	9	0.245	‡	10	0.102	‡	11	10.000	0.115	‡	4	0.308	‡	11	0.523	‡	9	8.000
QR-IV-SV	0.501	‡	17	0.277		14	0.108	‡	14	15.000	0.126	‡	12	0.304	‡	10	0.518	‡	8	10.000
QR-IV-SJ	0.472	‡	10	0.271	‡	13	0.106	‡	13	12.000	0.121	‡	7	0.298	‡	7	0.513	‡	7	7.000
Panel H: Dimitriadis and Bayer's (2019) expected shortfall forecasting models with implied volatility																				
DB-IV	0.398	‡	5	0.206	‡	1	0.081	‡	2	2.667	0.117	‡	6	0.284	‡	2	0.501	‡	4	4.000
DB-IV-CJ	0.616		35	0.215	‡	5	0.089	‡	5	15.000	0.144		29	0.341		35	0.570		24	29.333
DB-IV-SV	0.396	‡	4	0.249		11	0.099		10	8.333	0.135		16	0.333		33	0.579	‡	26	25.000
DB-IV-CQ	0.433	‡	7	0.218	‡	6	0.091	‡	7	6.667	0.135	‡	16	0.344		36	0.561		20	24.000
DB-IV-SJ	0.473	‡	11	0.242		9	0.096		9	9.667	0.134		15	0.323	‡	21	0.563	‡	21	19.000
Panel I: Combination forecasts																				
EGARCH realized volatility	0.531		29	0.314		32	0.153		30	30.333	0.144		29	0.329		29	0.558		17	25.000
EGARCH implied volatility	0.527		26	0.309		28	0.148		27	27.000	0.142		23	0.327		23	0.554	‡	14	20.000
QR realized volatility	0.532		30	0.298		24	0.128		19	24.333	0.137	‡	19	0.320	‡	18	0.573		25	20.667
QR implied volatility	0.415	‡	6	0.233	‡	7	0.089	‡	5	6.000	0.116	‡	5	0.293	‡	5	0.502	‡	5	5.000
DB realized volatility	0.492		13	0.286		16	0.139		23	17.333	0.143		26	0.318	‡	17	0.587		28	23.667
DB implied volatility	0.365	‡	1	0.211	‡	2	0.081	‡	2	1.667	0.125	‡	11	0.309	‡	12	0.527	‡	10	11.000
All realized volatility and historical models	0.495		15	0.289		18	0.128		19	17.333	0.131	‡	13	0.312	‡	13	0.552	‡	13	13.000
All realized volatility	0.498		16	0.283		15	0.127		17	16.000	0.133	‡	14	0.313	‡	15	0.569		23	17.333
All implied volatility	0.367	‡	2	0.214	‡	4	0.079	‡	1	2.333	0.112	‡	2	0.290	‡	4	0.498	‡	3	3.000

Notes: Values in the table correspond to the average FZ^0 loss of the ES model in a given row. The symbols † and ‡ denote models that are in the superior set of models M^*, as indicated by the model confidence set testing procedure († for 90% confidence, and ‡ for 75% confidence), while their loss is below the average loss across all 38 models. R denotes the rank of the model, and LTR and RTR denote the average rank across left or right tails of the distribution, respectively.

holders of the euro should be more able to improve their risk management with implied volatility than holders of the US dollar.

One explanation for this result might be a stronger overestimation of the true price variation of implied volatility related to options that protect against the weakening of the euro. For currency pairs, using implied volatilities from different options might be useful and shall be explored in future research.

Next, we report results from the weak dominance tests of Ziegel et al. (2020), where a pairwise comparison is implemented (see Section 3.10.3). We always compare two models. One with realized measures only (model A) and another with implied volatilities (model B). The null hypothesis states that model A weakly dominates model B. Assume a rejection of the null hypothesis, which suggests that we do not have evidence that model A dominates model B. Following Ziegel et al. (2020), we can switch the models and test the null that, instead, model B weakly dominates model A. If we are now unable to reject the null hypothesis, together with the previous finding, the evidence suggests that forecasts from model B are preferred over forecasts from model A.

We report the results from the weak dominance tests in Table 6. In each row, we have two models of interest, and the values correspond to bootstrapped p-values. In the first six columns, we observe that across all quantiles of interest, we cannot rule out that the implied volatility model outperforms its counterpart, the model with realized measures only. However, the final six columns show that in the left tail of the return distribution, we have many cases where we can safely reject that the model with realized variances dominates the model with implied volatilities. This is especially true when we compare combination forecasts from QR or DB models or all models. On the individual level, the differences between QR-IV vs. QR-RV and DB-IV vs. DB-RV are significant at the 0.10 threshold level. These results strengthen previously

Table 6
Pair-wise weak dominance tests of Ziegel et al. (2020).

Implied volatility model	Benchmark model	H0: Implied volatility model weakly dominates benchmark						H0: Benchmark weakly dominates implied volatility model					
		0.01 ES	0.025 ES	0.05 ES	0.95 ES	0.975 ES	0.99 ES	0.01 ES	0.025 ES	0.05 ES	0.95 ES	0.975 ES	0.99 ES
Panel A: GARCH class of expected shortfall forecasts													
EGARCH-IV-SSK	EGARCH-SSK	0.403	0.065	0.001	0.967	0.794	0.983	0.310	0.448	0.620	0.830	0.574	0.474
EGARCH-IV-SSK	EGARCH-RV-SSK	0.523	0.570	0.136	0.438	0.362	0.337	0.606	0.459	0.446	0.846	0.732	0.776
EGARCH-IV-JSU	EGARCH-JSU	0.306	0.020	0.004	0.781	0.826	0.961	0.287	0.469	0.714	0.756	0.710	0.507
EGARCH-IV-JSU	EGARCH-RV-JSU	0.505	0.873	0.766	0.298	0.219	0.187	0.715	0.539	0.411	0.616	0.710	0.701
Panel B: Quantile regression approximate expected shortfall forecasts													
QR-IV	QR-RV	0.509	0.567	0.346	1.000	0.993	0.991	0.050	0.021	0.052	0.090	0.116	0.072
QR-IV-CJ	QR-RV-CJ	0.407	0.778	0.918	0.832	0.998	0.990	0.091	0.157	0.106	0.106	0.139	0.128
QR-IV-SV	QR-RV-SV	0.171	0.226	0.162	1.000	0.992	0.992	0.102	0.116	0.029	0.236	0.243	0.108
QR-IV-CQ	QR-RV-CQ	0.942	0.972	0.976	0.999	0.997	0.998	0.101	0.028	0.045	0.448	0.419	0.147
QR-IV-SJ	QR-RV-SJ	0.452	0.699	0.664	1.000	0.997	0.994	0.080	0.048	0.040	0.120	0.152	0.162
Panel C: Dimitriadis and Bayer's (2019) expected shortfall forecasts													
DB-IV	DB-HAR	0.875	0.957	0.995	0.826	0.994	0.997	0.081	0.020	0.072	0.101	0.172	0.103
DB-IV-CJ	DB-RV-CJ	0.302	0.996	0.996	0.999	0.785	0.987	0.032	0.019	0.020	0.427	0.089	0.427
DB-IV-SV	DB-RV-SV	0.739	0.822	0.941	1.000	0.994	0.617	0.060	0.085	0.015	0.437	0.629	0.202
DB-IV-CQ	DB-RV-CQ	0.307	0.948	0.989	1.000	0.915	0.999	0.094	0.018	0.004	0.105	0.687	0.242
DB-IV-SJ	DB-RV-SJ	0.713	0.945	0.979	1.000	0.766	0.373	0.163	0.086	0.043	0.317	0.568	0.312
Panel D: Combination forecasts													
EGARCH implied volatility	EGARCH realized volatility	0.428	0.772	0.151	0.360	0.298	0.241	0.718	0.439	0.382	0.861	0.803	0.842
QR implied volatility	QR realized volatility	0.519	0.716	0.757	1.000	0.989	0.991	0.040	0.029	0.007	0.207	0.121	0.140
DB implied volatility	DB realized volatility	0.839	0.970	0.999	0.999	0.999	0.995	0.043	0.055	0.006	0.189	0.493	0.407
All implied volatility	All realized volatility and hist. models	0.867	0.993	0.987	1.000	0.998	0.985	0.003	0.005	0.014	0.147	0.222	0.258
All implied volatility	All realized volatility	0.995	0.978	0.901	1.000	0.997	0.989	0.016	0.031	0.020	0.208	0.099	0.065

Notes: The table reports bootstrap-based p-values of the weak dominance test of Ziegel et al. (2020). We test both variations under the null hypothesis, that either the implied volatility or the benchmark model outperforms the other. Non-rejection of the hypothesis that implied volatility dominates the benchmark model and rejection of the hypothesis that the benchmark model dominates the implied volatility model are considered evidence that the implied volatility model is preferred against the benchmark model under a broad set of loss functions. The p-values are based on the stationary bootstrap of Politis and Romano (1994), with random block lengths drawn from a geometric distribution with the expected value given as in Politis and White (2004), and Patton et al. (2009) with 1000 bootstrap samples.

observed analysis that implied volatility is the most useful in the left tail when the US dollar strengthens.

5. Discussion

5.1. Loss differentials over time

If the usefulness of implied volatility is driven by a few events or is not stable over time, one cannot rely on our recommendations set forth in the previous sections. In Fig. 5, the black line corresponds to the loss under the QR-RV model and the red line to the loss under the QR-IV model. The initial years are used for estimation purposes, and no predictions are made, as depicted by the horizontal lines at the beginning.

Several observations may be inferred based on Fig. 5. For the period until approximately the sixth month of 2014, the differences in losses are not that visible. This period corresponds to a low-volatility regime on the EUR/USD exchange rate market (see Figs. 2 and 3). With the increase in market volatility at the end of 2014, the differences between the losses widen, most visibly for the most extreme quantiles, $\tau = 0.01$ and $1-\tau = 0.99$. Steep changes are visible in the losses, and they occur when there is a hit event, i.e. when the returns move below (above) the left (right) tail of the predicted VaR. However, the overall performance does not seem to be driven by a few events and is stable over time.

5.2. The role of market uncertainty

The results in the previous section suggest that the differences between losses from the QR-RV and QR-IV models differ mostly during periods of higher market uncertainty. In this section, we stratify the losses according to both measures of market uncertainty: (i) the realized volatility, RV_t; and (ii) the daily implied volatility, IV_t^D. In Table 7, we report the average values of the loss differential, $FZ^{(0)}_{t,QR-RV} - FZ^{(0)}_{t,QR-IV}$, for days when the given uncertainty measure is below the historical 90$^{\text{th}}$ percentile or above. The larger the average difference is, the worse the performance of the QR-RV (DB-RV) model in comparison to QR-IV (DB-IV), which utilizes implied volatilities. This stratification is important, as periods of high market uncertainty tend to be accompanied by large return swings in both tails of the return distribution; that is, during such periods, accurate management of market risk matters the most.

In Table 7 we report average loss differentials with corresponding standard errors. Stratifying losses according to the realized variance (first six columns) leads to a small difference between average losses, regardless of whether we compare results between the QR or between the DB models. If anything, it appears that during days with higher realized variance, implied volatility models performed worse. However, after stratifying losses according to implied volatility (last six columns), we observe much larger differences. Regardless of the quantile of interest, during days with higher implied volatility (with daily maturity), losses were smaller for the QR-IV and DB-IV models. These results suggest that when forward-looking uncertainty expectations are large, models that use implied volatilities perform much better.

5.3. Alternative loss functions

In this section, we discuss how our results differ after we evaluate the ES predictions against two alternative loss functions. First, we use that from Fissler and Ziegel (2016), Fissler et al. (2015) in the version of Taylor (2019),

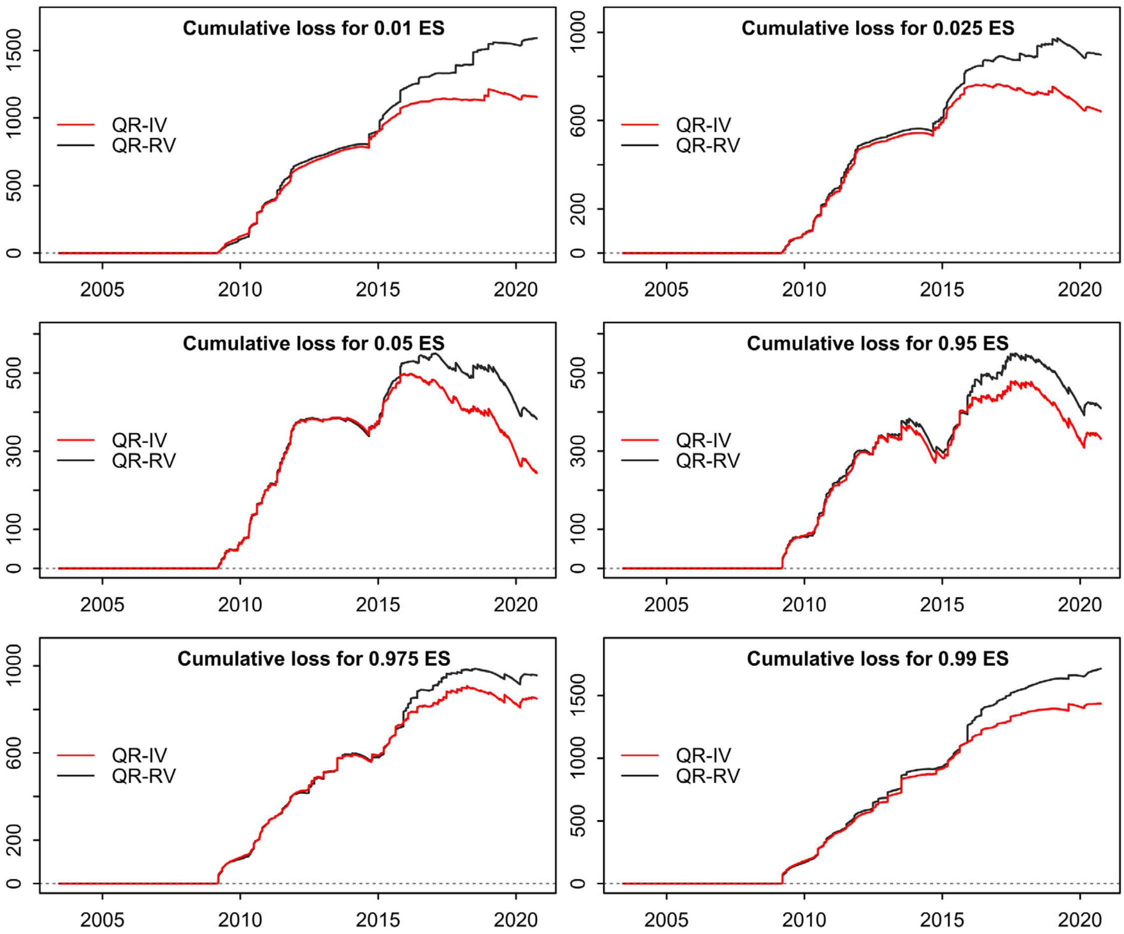

Fig. 5. Cumulative loss of FZ^0.

Table 7
Stratified loss differentials.

	Stratification based on realized variance						Stratification based on implied volatility					
	0.01	0.025	0.05	0.95	0.975	0.99	0.01	0.025	0.05	0.95	0.975	0.99
Panel A: Loss differentials QR-RV vs. QR-IV												
Bellow 90% percentile												
Average loss differential	0.154	0.087	0.048	0.035	0.045	0.109	0.036	0.030	0.016	0.003	0.008	0.045
Standard deviation	[0.06]	[0.03]	[0.02]	[0.02]	[0.03]	[0.06]	[0.02]	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]
Above 90% percentile → high-uncertainty period												
Average loss differential	0.053	0.068	0.022	−0.060	−0.052	−0.065	1.084	0.562	0.308	0.225	0.276	0.502
Standard deviation	[0.06]	[0.03]	[0.03]	[0.04]	[0.05]	[0.07]	[0.45]	[0.20]	[0.11]	[0.14]	[0.20]	[0.44]
Panel B: Loss differentials DB-RV vs. DB-IV												
Bellow 90% percentile												
Average loss differential	0.098	0.089	0.052	0.033	0.037	0.099	0.069	0.044	0.026	−0.001	0.001	0.043
Standard deviation	[0.07]	[0.03]	[0.02]	[0.02]	[0.03]	[0.05]	[0.03]	[0.02]	[0.01]	[0.01]	[0.02]	[0.03]
Above 90% percentile → high-uncertainty period												
Average loss differential	0.070	0.052	0.071	−0.047	−0.061	0.023	0.323	0.447	0.291	0.243	0.256	0.514
Standard deviation	[0.04]	[0.03]	[0.03]	[0.04]	[0.05]	[0.10]	[0.51]	[0.18]	[0.11]	[0.13]	[0.21]	[0.40]

Notes: The values correspond to average loss differentials under the FZ^0 loss function between the QR-RV and QR-IV (DB-RV and DB-IV) models. Positive values suggest that the implied volatility model performs better. The values in brackets are standard deviation of losses.

which leads to positive values of the loss function:

$$FZG_t = (L_t - \tau) VaR_{R_t}(\tau) - L_t R_t + \frac{ES_{R_t}(\tau)}{1 + e^{ES_{R_t}(\tau)}} \left(ES_{R_t}(\tau) - VaR_{R_t}(\tau) + \frac{L_t (VaR_{R_t}(\tau) - R_t)}{\tau} \right) + \log\left(\frac{2}{1 + e^{ES_{R_t}(\tau)}} \right) \tag{32}$$

Detailed results are reported in Table 8. Our main conclusions hold under this loss function as well. That is, the implied volatility models present a safer choice in that (i) they are more often members of the set of superior models, (ii) the losses are lower than under models without implied volatility, and (iii) the ranking of the models is also better. Comparing the losses of the best-performing combination forecasts that do not use implied

Table 8
Average values of the FZG loss function, model confidence set, and average ranks.

	0.01 ES			0.025 ES			0.05 ES			LTR	0.95 ES			0.975 ES			0.99 ES			RTR
Model	FZG	M*	R	FZG	M*	R	FZG	M*	R		FZG	M*	R	FZG	M*	R	FZG	M*	R	
Panel A: Historical distribution-based expected shortfall forecasting models																				
SSK	1.291		38	0.889		38	0.724		38	38.000	0.637		37	0.719		38	0.924		38	37.667
JSU	1.144		37	0.848		37	0.708		37	37.000	0.642		38	0.706		37	0.847		37	37.333
Panel B: Dynamic latent volatility expected shortfall forecasting models																				
EGARCH-JSU	0.622		36	0.601		36	0.545		36	36.000	0.483		32	0.460		28	0.334		25	28.333
EGARCH-SSK	0.609		35	0.597		35	0.542		35	35.000	0.489		35	0.465		29	0.333		21	28.333
Panel C: Dynamic latent volatility expected shortfall forecasting models with realized volatility																				
EGARCH-RV-SSK	0.587		31	0.582		31	0.534		31	31.000	0.475		24	0.442		21	0.316		11	18.667
EGARCH-RV-JSU	0.577		29	0.574		29	0.529		29	29.000	0.480		30	0.444		24	0.314		9	21.000
Panel D: Quantile regression approximate expected shortfall forecasting models with realized volatility																				
QR-RV	0.314	‡	11	0.366		13	0.402		13	12.333	0.400	‡	10	0.406	‡	16	0.333	‡	21	15.667
QR-RV-CJ	0.272	‡	5	0.350		9	0.400		12	8.667	0.399	‡	9	0.416	‡	19	0.324	‡	16	14.667
QR-RV-CQ	0.321	‡	13	0.347	‡	8	0.399		11	10.667	0.400	‡	10	0.410	‡	18	0.356	‡	29	19.000
QR-RV-SV	0.283	‡	6	0.361		11	0.388		8	8.333	0.401	‡	13	0.392	‡	10	0.304	‡	7	10.000
QR-RV-SJ	0.315	‡	12	0.362		12	0.395		9	11.000	0.398	‡	8	0.397	‡	13	0.333	‡	21	14.000
Panel E: Dimitriadis and Bayer's expected shortfall forecasting models with realized volatility																				
DB-RV	0.453		24	0.474		27	0.479		22	24.333	0.475		24	0.506		36	0.333	‡	21	27.000
DB-RV-CJ	0.447		22	0.449		22	0.485		23	22.333	0.474		23	0.499		33	0.350	‡	28	28.000
DB-RV-SV	0.416		20	0.473		26	0.499		24	23.333	0.479		29	0.486		31	0.390	‡	34	31.333
DB-RV-CQ	0.393		16	0.449		22	0.512		28	22.000	0.493		36	0.488		32	0.393	‡	35	34.333
DB-RV-SJ	0.415		19	0.461		24	0.506		27	23.333	0.468		22	0.502		34	0.381	‡	33	29.667
Panel F: Dynamic latent volatility expected shortfall forecasting models with implied volatility																				
EGARCH-IV-SSK	0.608		34	0.589		34	0.537		34	34.000	0.476		26	0.442		21	0.321		14	20.333
EGARCH-IV-JSU	0.591		32	0.584		32	0.535		32	32.000	0.484		34	0.450		26	0.325		17	25.667
Panel G: Quantile regression approximate expected shortfall forecasting models with implied volatility																				
QR-IV	0.313	‡	10	0.346		7	0.354		7	8.000	0.370	‡	3	0.348	‡	1	0.311	‡	8	4.000
QR-IV-RV	0.307	‡	8	0.313	‡	6	0.346	‡	6	6.667	0.369	‡	2	0.362	‡	7	0.284	‡	4	4.333
QR-IV-CJ	0.186	‡	2	0.290	‡	1	0.336	‡	1	1.333	0.372	‡	6	0.360	‡	6	0.289	‡	5	5.667
QR-IV-CQ	0.148	‡	1	0.304	‡	4	0.345	‡	5	3.333	0.380		7	0.348	‡	1	0.297	‡	6	4.667
QR-IV-SV	0.249	‡	4	0.293	‡	2	0.336	‡	1	2.333	0.370	‡	3	0.357	‡	4	0.256	‡	1	2.667
QR-IV-SJ	0.248	‡	3	0.308	‡	5	0.337	‡	3	3.667	0.364	‡	1	0.349	‡	3	0.281	‡	2	2.000
Panel H: Dimitriadis and Bayer's expected shortfall forecasting models with implied volatility																				
DB-IV	0.383	‡	15	0.416		17	0.423		17	16.333	0.437		20	0.416		19	0.327	‡	19	19.333
DB-IV-CJ	0.460		25	0.419		19	0.433		20	21.333	0.419		16	0.380	‡	8	0.380	‡	32	18.667
DB-IV-SV	0.514		27	0.401		15	0.422		16	19.333	0.425		17	0.401	‡	14	0.316	‡	11	14.000
DB-IV-CQ	0.379	‡	14	0.418		18	0.420		15	15.667	0.433		19	0.389		9	0.362	‡	30	19.333
DB-IV-SJ	0.407		18	0.409		16	0.425		18	17.333	0.418		15	0.395	‡	12	0.319	‡	13	13.333
Panel I: Combination forecasts																				
EGARCH realized volatility	0.582		30	0.578		30	0.531		30	30.000	0.477		28	0.443		23	0.315	‡	10	20.333
EGARCH implied volatility	0.601		33	0.586		33	0.536		33	33.000	0.480		30	0.446		25	0.323	‡	15	23.333
QR realized volatility	0.309	‡	9	0.357		10	0.397		10	9.667	0.400	‡	10	0.405	‡	15	0.332	‡	20	15.000
QR implied volatility	0.286	‡	7	0.298	‡	3	0.338	‡	4	4.667	0.371		5	0.357	‡	4	0.282	‡	3	4.000
DB realized volatility	0.444		21	0.465		25	0.499		24	23.333	0.483		32	0.502		34	0.376	‡	31	32.333
DB implied volatility	0.499		26	0.421		20	0.429		19	21.667	0.431		18	0.407		17	0.348	‡	27	20.667
All realized volatility and historical models	0.526		28	0.512		28	0.499		24	26.667	0.476		26	0.483		30	0.410		36	30.667
All realized volatility	0.401		17	0.440		21	0.459		21	19.667	0.448		21	0.451		27	0.346	‡	26	24.667
All implied volatility	0.451		23	0.398		14	0.405		14	17.000	0.414		14	0.393		11	0.325	‡	17	14.000

Notes: Values in the table correspond to the average FZG loss of the ES model in a given row. The symbols † and ‡ denote models that are in the superior set of models M*, as indicated by the model confidence set testing procedure († for 90% confidence, and ‡ for 75% confidence), while their loss is below the average loss across all 38 models. R denotes the rank of the model, and LTR and RTR denote the average rank across left or right tails of the distribution, respectively.

volatilities to those that use them still leads to forecast improvements of 7.9% ($\tau = 0.01$), 19.5% ($\tau = 0.025$), 17.5% ($\tau = 0.05$), 7.8% ($1-\tau = 0.95$), 13.6% ($1-\tau = 0.975$) and 11.6% ($1-\tau = 0.99$).[18]

What differs is the best-performing model; in the previous analysis, it was the QR-IV model. Now, the QR-IV-SV model tends to perform best, while among the models without implied volatility, the analogous QR-RV-SV performs very well. Clearly, under different loss functions, the general conclusions do not vary, but in reality, one cannot rely on a single model. Even though combination forecasts never lead to predictions with the lowest loss, they are certainly a safer bet, offering predictions that are among the most accurate overall.

The second loss function is the AL log score, as introduced by Taylor (2019):

$$AL_t = -\log\left(\frac{\tau - 1}{ES_{R_t}}\right) - \frac{(R_t - VaR_t)(\tau - L_t)}{\tau ES_{R_t}} \tag{33}$$

Detailed results can be found in Table 9. As before, implied volatility models tend to lead to lower losses, although notable differences in loss values are found for the left-tail predictions, i.e. euro weakening, but not so much for those for the right tail. Comparing the best-performing realized and implied volatility combination forecasts (i.e. non-implied vs. implied volatility models) shows reductions in losses for the implied volatility models of 9.1% ($\tau = 0.01$), 5.6% ($\tau = 0.025$), 4.1% ($\tau = 0.05$), 1.7% ($1-\tau = 0.95$), 1.7% ($1-\tau = 0.975$) and 3.6% ($1-\tau = 0.99$).[19] However,

18 In absolute terms, the improvements correspond to 0.023, 0.058, 0.059 for 0.01, 0.025, 0.05 in the left tail, and to 0.029, 0.049, 0.033 for 0.95, 0.975, 0.99 in the right tail of the return distribution.

19 In absolute terms, the improvements correspond to 0.125, 0.070, 0.046 for 0.01, 0.025, 0.05 in the left tail, and to 0.020, 0.023, 0.054 for 0.95, 0.975, 0.99 in the right tail of the return distribution.

Table 9
Average values of the AL loss function, model confidence set, and average ranks.

	0.01 ES			0.025 ES			0.05 ES			LTR	0.95 ES			0.975 ES			0.99 ES			RTR
Model	AL	M*	R	AL	M*	R	AL	M*	R		AL	M*	R	AL	M*	R	AL	M*	R	
Panel A: Historical distribution-based expected shortfall forecasting models																				
SSK	1.636		36	1.447		36	1.287		37	36.333	1.262		37	1.428		37	1.625		36	36.667
JSU	1.651		38	1.450		37	1.288		38	37.667	1.266		38	1.432		38	1.625		36	37.333
Panel B: Dynamic latent volatility expected shortfall forecasting models																				
EGARCH-JSU	1.515		19	1.312		16	1.174		15	16.667	1.178		11	1.330		9	1.543		11	10.333
EGARCH-SSK	1.516		20	1.312		16	1.174		15	17.000	1.178		11	1.331		10	1.543		11	10.667
Panel C: Dynamic latent volatility expected shortfall forecasting models with realized volatility																				
EGARCH-RV-SSK	1.540		30	1.336		31	1.201		30	30.333	1.198		29	1.356		26	1.571		17	24.000
EGARCH-RV-JSU	1.538		28	1.337		32	1.203		32	30.667	1.200		33	1.359		33	1.571		17	27.667
Panel D: Quantile regression approximate expected shortfall forecasting models with realized volatility																				
QR-RV	1.535		25	1.321		23	1.177		17	21.667	1.187	‡	18	1.342	‡	16	1.577		22	18.667
QR-RV-CJ	1.559		32	1.335		30	1.182		21	27.667	1.192	‡	21	1.346	‡	18	1.599		29	22.667
QR-RV-CQ	1.641		37	1.393		35	1.223		35	35.667	1.191	‡	20	1.353	‡	23	1.607		32	25.000
QR-RV-SV	1.564		33	1.327		26	1.189		23	27.333	1.192	‡	21	1.357		30	1.599		29	26.667
QR-RV-SJ	1.595		34	1.346		34	1.194		25	31.000	1.193	‡	23	1.355		25	1.595		27	25.000
Panel E: Dimitriadis and Bayer's expected shortfall forecasting models with realized volatility																				
DB-RV	1.503		14	1.316		21	1.187		22	19.000	1.193	‡	23	1.337	‡	13	1.603		31	22.333
DB-RV-CJ	1.528		23	1.326		25	1.196		26	24.667	1.203		34	1.353	‡	23	1.608		33	30.000
DB-RV-SV	1.526		22	1.312		16	1.209		34	24.000	1.205		35	1.366		34	1.616		35	34.667
DB-RV-CQ	1.518		21	1.649		38	1.228		36	31.667	1.197		27	1.346	‡	18	1.639		38	27.667
DB-RV-SJ	1.512		17	1.319		22	1.204		33	24.000	1.206		36	1.349	‡	22	1.613		34	30.667
Panel F: Dynamic latent volatility expected shortfall forecasting models with implied volatility																				
EGARCH-IV-SSK	1.533		24	1.331		27	1.196		26	25.667	1.199		31	1.356		26	1.567		14	23.667
EGARCH-IV-JSU	1.537		27	1.333		29	1.196		26	27.333	1.198		29	1.356		26	1.567		14	23.000
Panel G: Quantile regression approximate expected shortfall forecasting models with implied volatility																				
QR-IV	1.393	‡	3	1.238	‡	2	1.133	‡	2	2.333	1.161	‡	1	1.306	‡	1	1.485	‡	1	1.000
QR-IV-RV	1.446	‡	8	1.267	‡	8	1.146	‡	8	8.000	1.164	‡	3	1.312	‡	3	1.500	‡	2	2.667
QR-IV-CJ	1.495	‡	12	1.289	‡	12	1.156	‡	12	12.000	1.172	‡	8	1.320	‡	6	1.515	‡	6	6.667
QR-IV-CQ	1.479	‡	9	1.270	‡	10	1.153	‡	10	9.667	1.166	‡	4	1.333	‡	11	1.534	‡	9	8.000
QR-IV-SV	1.512	‡	17	1.303		14	1.160	‡	13	14.667	1.177	‡	10	1.329	‡	8	1.529	‡	8	8.667
QR-IV-SJ	1.483	‡	10	1.298	‡	13	1.160	‡	13	12.000	1.171	‡	7	1.323	‡	7	1.523	‡	7	7.000
Panel H: Dimitriadis and Bayer's expected shortfall forecasting models with implied volatility																				
DB-IV	1.409	‡	5	1.232	‡	1	1.134	‡	3	3.000	1.169	‡	6	1.310	‡	2	1.511	‡	4	4.000
DB-IV-CJ	1.627		35	1.244	‡	5	1.143	‡	6	15.333	1.193		23	1.366		34	1.579		23	26.667
DB-IV-SV	1.408	‡	4	1.276		11	1.154		11	8.667	1.185	‡	16	1.357		30	1.589		26	24.000
DB-IV-CQ	1.445	‡	7	1.244	‡	5	1.145	‡	7	6.333	1.185	‡	16	1.370		36	1.571	‡	17	23.000
DB-IV-SJ	1.484	‡	11	1.269		9	1.150		9	9.667	1.183	‡	13	1.347	‡	21	1.573	‡	21	18.333
Panel I: Combination forecasts																				
EGARCH realized volatility	1.539		29	1.337		32	1.202		31	30.667	1.199		31	1.357		30	1.571	‡	17	26.000
EGARCH implied volatility	1.535		25	1.332		28	1.196		26	26.333	1.197		27	1.356		26	1.567	‡	14	22.333
QR realized volatility	1.541		31	1.322		24	1.179		19	24.667	1.188	‡	19	1.346	‡	18	1.583		25	20.667
QR implied volatility	1.425	‡	6	1.260	‡	7	1.141	‡	5	6.000	1.167	‡	5	1.318	‡	5	1.513	‡	5	5.000
DB realized volatility	1.502		13	1.312		16	1.190		24	17.667	1.194		26	1.344	‡	17	1.597		28	23.667
DB implied volatility	1.377	‡	1	1.238	‡	2	1.134	‡	3	2.000	1.175	‡	9	1.334	‡	12	1.537	‡	10	10.333
All realized volatility and historical models	1.504		15	1.313		20	1.179		19	18.000	1.183	‡	13	1.338	‡	14	1.563	‡	13	13.333
All realized volatility	1.507		16	1.307		15	1.178		18	16.333	1.184	‡	15	1.339	‡	15	1.579		23	17.667
All implied volatility	1.377	‡	1	1.240	‡	4	1.132	‡	1	2.000	1.163	‡	2	1.316	‡	4	1.509	‡	3	3.000

Notes: Values in the table correspond to the average AL loss of the ES model in a given row. The symbols † and ‡ denote models that are in the superior set of models M*, as indicated by the model confidence set testing procedure († for 90% confidence, and ‡ for 75% confidence), while their loss is below the average loss across all 38 models. R denotes the rank of the model, and LTR and RTR denote the average rank across left or right tails of the distribution, respectively.

similarly to the case under the $FZ^{(0)}$ loss function, the QR-IV model is to be preferred, and among non-implied volatility models, the EGARCH-JSU model performs the best. As before, we recommend the use of forecasts that combine forecasts across all implied volatility models, including the DB model.

5.4. Reducing expected shortfall: An economic application

This section demonstrates an application of our model for purpose of EUR/USD hedging. In our scenario, a US-based investor seeks to hedge against excessive expected shortfall in the left tail of the return distribution.[20] The investor hedges her position every day using her forecasts of the expected shortfall for a given τ. The selected hedge ratio ($H_t(L)$) depends on her accepted maximum expected shortfall at that τ (denoted as L) and on the predicted expected shortfall ($ES_{R_t}(\tau)$). The hedge ratio is given as:

$$H_t(L) = \begin{cases} \frac{ES_{R_t}(\tau)-L}{ES_{R_t}(\tau)} & : \frac{ES_{R_t}(\tau)-L}{ES_{R_t}(\tau)} > 0 \\ 0 & : \frac{ES_{R_t}(\tau)-L}{ES_{R_t}(\tau)} \leq 0 \end{cases} \tag{34}$$

According to Eq. (34), we hedge only if the predicted expected shortfall suggests risks above the accepted risk level given by L. The overall position of the investor can be expressed as a sum of a long position in EUR, assumed to be 1 (i.e. its size is irrelevant), and a time-varying opposing short position that depends on $H_t(L)$. The larger

20 As an example, we might use a US-based retailer who sells products to Europe on a daily basis and receives payments in euros that need to be converted back to US dollars on a daily basis as well. We assume that the pricing of products can be adjusted at the end of a day; therefore, the total amount of expected daily revenue is assumed to be fixed and is thus irrelevant in this example. Because such a producer works under specific profitability margins, she is interested in hedging against extreme price movements.

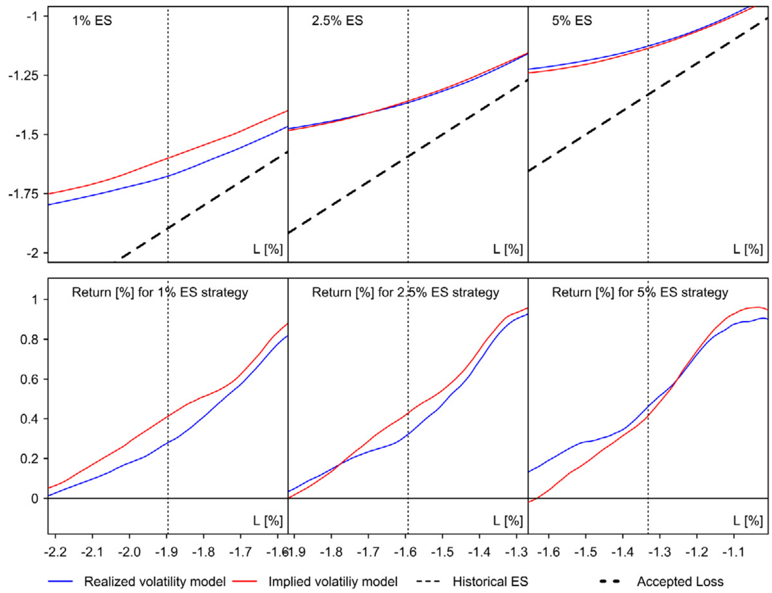

Fig. 6. Evaluation of a trading strategy for 1%, 2.5%, and 5% predicted ES.
Note: The x-axis represents the accepted loss. The vertical dashed line corresponds to the historical expected shortfall at the given quantile (0.01, 0.025, and 0.05). The first rows in the subfigures show the expected shortfall to which the investor is exposed after using either the strategy of prediction from a model that uses predictions from combination forecasts that use only realized measures (blue line), or a strategy that also uses implied volatilities as covariates (red line). Red and blue lines above the dashed diagonal show a reduction in the expected shortfall. The second row shows the costs of the hedging, where the y-axis corresponds to daily annualized average percentage returns achieved by using these strategies. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the predicted expected shortfall is, the larger the size of the hedging position. The overall return from this hedging strategy is given by:

$$HR_t(L) = R_t \times (1 - H_t(L)) \tag{35}$$

We calculate $HR_t(L)$ for two trading strategies. The investor uses either forecasts of expected shortfall that do not use implied volatility (the QR-RV model forecasts) or, in the second case, predictions from models that utilize implied volatilities (QR-IV model forecasts).

To compare the results, we consider the accepted risk level L over a range of possible values around the historical ES (dashed vertical line in Figs. 6 and 7).[21] For a given accepted risk level L, we calculate hedge ratio $H_t(L)$ and trade accordingly. We evaluate the resulting series of realized returns $HR_t(L)$ via two measures.

First, we find the ES of $HR_t(L)$ over the predicted horizon. In Fig. 6 (for $\tau = 0.01, 0.025$, and 0.05) and Fig. 7 (for $1-\tau = 0.95, 0.975$, and 0.99), the x-axis corresponds to the accepted risk level L. The blue (QR-RV) and red (QR-IV) lines correspond to the realized ES. Because we enter an opposing position only when predicted ES falls above the accepted risk level L, hedging should lead to realized ES that is lower than the accepted risk level L. Ideally, we want to observe that the application of the hedging strategy leads to a large reduction of the ES; i.e. the red and blue lines should lie above (below) the sloped dashed line in Fig. 6 (Fig. 7). In that regard, Figs. 6 and 7 show that hedging was successful and that the reduction is larger the higher the accepted risk level, irrespective of whether one uses ES forecasts from the QR-RV or QR-IV models. Slightly lower realized ES is achieved via $QR-IV$.

Second, we calculate the average annualized returns $HR_t(L)$, which are on the y-axis of the bottom part of Figs. 6 and 7, while the x-axis again corresponds to the accepted loss L. The particular returns are not that relevant, and their value depends on the sample period. In our empirical application, these returns are actually positive, meaning that hedging leads to small profits. We are interested in the difference between returns using the hedging strategies based on the QR-RV and QR-IV ES forecasts. We observe that returns tend to be higher if hedging is conducted via QR-IV (except $\tau = 0.05$).

Overall, this exercise demonstrates that ES predictions can improve risk management, regardless of whether the predictions are based on the QR-RV or the QR-IV model. However, under the QR-IV model we observe greater reduction of the realized ES but also higher returns (lower hedging costs).[22]

21 The first out-of-sample observation is $T_1 = 1501$, and we have $T = 4522$ observations and thus $n = 3022$ out-of-sample observations. The historical ES (dashed vertical line in Figs. 6 and 7) is given as an average of returns that fall below a given unconditional quantile $\tau \ll 0.5$, and it is found as $(n\tau)^{-1}\sum_{t=T_1}^{T} R_t \times I(R_t < Q_{T_1,T}(\tau))$, where $Q_{T_1,T}(\tau)$ is the unconditional quantile of returns over the out-of-sample period.

22 As the size of the hedging position depends on the predicted ES, the realized returns might also be affected by trading costs. However, the EUR/USD exchange rate is among the most liquid assets in the world, and the bid–ask spread is only 0.5 pips. Given the average exchange rate over the out-of-sample period of 1.23, this amounts to approximately 0.0041%. Under such fixed percentage costs, the return

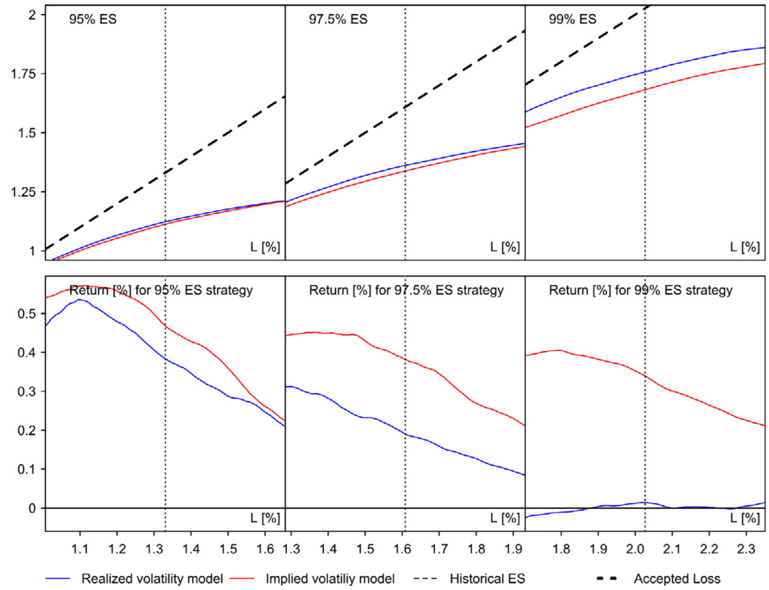

Fig. 7. Evaluation of a trading strategy for 95%, 97.5%, and 99% predicted ES.
Note: The x-axis represents the accepted loss. The vertical dashed line corresponds to the historical expected shortfall at the given quantile (0.95, 0.975, and 0.99). The first rows in the subfigures show the expected shortfall to which the investor is exposed after using either the strategy of prediction from a model that uses predictions from combination forecasts that use only realized measures (blue line), or a strategy that also uses implied volatilities as covariates (red line). Red and blue lines below the dashed diagonal show a reduction in the expected shortfall. The second row shows the costs of the hedging, where the y-axis corresponds to daily annualized average percentage returns achieved by using these strategies. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.5. Increasing forecasting horizon: Five-, 10-, and 22-day-ahead market risk forecasts

Our detailed results are based on the day-ahead forecasts of VaR and ES. As we increase the forecasting horizon, the expectations embedded in implied volatilities should be more uncertain and, consequently, we expect models with implied volatilities to be less useful. In this exercise, we use multiple specifications for QR and DM models only, which directly predicted five-, 10-, and 22-day-ahead returns. For evaluation purposes, we use the FZ^0 loss function and compare the accuracy between the best-performing realized volatility model against the best-performing implied volatility model.

We find that with increasing forecast horizon, the benefits of implied volatility forecasts become increasingly uncertain. Specifically, under the five-day-ahead forecasting framework, forecast improvements are 1.6% ($\tau = 0.01$), 1.3% ($\tau = 0.025$), 2.8% ($\tau = 0.05$), 2.4% ($1 - \tau = 0.95$), 2.1% ($1 - \tau = 0.975$), and 2.7% ($1 - \tau = 0.99$). For the 10-day-ahead forecasts, we observe 1.3% ($\tau = 0.01$), 2.6% ($\tau = 0.025$), 2.5% ($\tau = 0.05$), 0.2% ($1 - \tau = 0.95$), 0.1% ($1 - \tau = 0.975$), and 3.0% ($1 - \tau = 0.99$). Finally, for the 22-day-ahead forecasts, improvements were found only in the right tail where we previously identified the increased importance of implied volatility models, i.e. 1.4% ($\tau = 0.01$), 1.6% ($\tau = 0.025$), 1.1% ($\tau = 0.05$), −0.1% ($1 - \tau = 0.95$), 0.3% ($1 - \tau = 0.975$), and −1.7% ($1 - \tau = 0.99$).

5.6. Results summary

Our results confirm conclusions from the literature that find it beneficial to combine RV and IV measures in a single model (Giot, 2005; Jeon & Taylor, 2013; Kambouroudis et al., 2016; Kim & Ryu, 2015; Leiss & Nax, 2018; Siu, 2018). Our results also suggest that using only implied volatility leads to the most accurate models, regardless of whether we use the approximate ES forecast with QRs or the joint VaR and ES framework of Dimitriadis and Bayer (2019). There are theoretical concerns that IV is upward-biased and that the volatility risk premium contained in it is time varying and rises, especially in periods of high volatility, which makes it less useful for tail-event prediction (Andersen et al., 2020; Bams et al., 2017). Contrary to these assumptions, we found that this bias is beneficial for forecasting ES. The models containing implied volatility outperform purely realized volatility models, especially during periods of extremely high implied volatility. This suggests that the implied volatility premium contains at least some additional forward-looking information, and its time-varying property helps improve ES forecasts. However, as Slim et al. (2020) claim, the

can be expressed as $R_t \times (1 - H_t) - c \times H_t$, where $c > 0$ is the cost [%]. Assuming a slightly higher cost at $c = 0.005\%$, the results are qualitatively nearly identical, a small shift to larger ES (higher risk) but still far from the accepted levels, while we observe a small downward shift in returns.

volatility risk premium and resulting benefits for tail-event predictions could differ across markets.

6. Conclusion

Although the concept of expected shortfall, or equivalently, conditional value at risk, cannot be called new (Rockafellar et al., 2000), its recommendation by the Basel Committee on Banking Supervision (2019) in the Basel III Accord has spurred substantial academic and research interest in its estimation, backtesting, and forecasting. While the measure resolves some of the issues encountered with value at risk, the previously widely used risk metric, it introduces some problems of its own, such as nonelicitability, which makes research into new approaches for its use both useful and necessary.

In this paper, we explored various models for the estimation of the value at risk of EUR/USD returns at different quantiles, which were then aggregated into an approximation for the desired expected shortfall (Couperier & Leymarie, 2020; Lyócsa, Todorova, & Vŷrost, 2021) or used to predict expected shortfall directly Dimitriadis and Bayer (2019). The essential components in our models include various volatility estimates, following a simple concept whereby extreme returns from the tails of the return probability distribution are usually encountered in times of market uncertainty, which is captured well by measures of market volatility. Our paper focused on the performance of models augmented by forward-looking implied volatility obtained for option contracts, as opposed to the historical high-frequency realized volatility estimator.

There are compelling theoretical reasons why implied volatility might have different and possibly superior performance in the modeling of expected shortfall. Other volatility measures, such as realized volatility or latent volatility obtained from GARCH models, are inherently backward looking, as they are based on historical observations. In contrast, implied volatility is forward looking, as it is formed from individual expectations of market participants. As implied volatility may be observed directly during active trading, it may be interpreted as market consensus on the level of future price variation of a given asset. Implied volatility thus has the potential to introduce additional information that may also be exploited in the forecasting of the desired expected shortfall measure.

To explore the usefulness of implied volatility, we ran tests on a battery of econometric models to provide an expected shortfall estimate. To ensure that the effect of implied volatility is not a mere coincidence tied to a specific model specification, we explored several different classes of models: (i) historical simulation-based ES forecasts, (ii) exponential GARCH models with flexible distributional assumptions, (iii) exponential GARCH models with realized or implied volatilities in respective variance equations, (iv) a QR-based model (QR-RV) with various extensions that utilized components of the price variation (e.g. continuous, jump components, volatility measurement uncertainty, positive and negative semivariances, and signed jump components), (v) similar specifications under the joint VaR and ES regression framework of Dimitriadis and Bayer (2019) (DM-RV), and (vi) combination forecasts. For the last three classes, the individual specifications were used in their basic form but were also augmented by the information on implied volatility (QR-IV and DM-IV models).

Our main results from the in-sample study suggested that when the dollar strengthens (the left tail of the EUR/USD return distribution), the results from models based on realized (DM-RV) and implied volatilities (DM-IV) are clearly distinct, and implied volatility seems to be more informative. However, we found very little practical use in combining the two measures into a single model, a sign that information on past volatility is already incorporated in implied volatility. For dollar weakening, the implied volatility models seemed to have similar explanatory power as baseline models with realized variances. In the out-of-sample setting, we first explored whether the ES predictions are well-specified by means of the strict ES backtesting framework of Bayer and Dimitriadis (2022). We found that most of the models led to unbiased forecasts of ES. Additionally, using the selected loss function, we showed that the implied volatility models tended to be superior and provided evidence of an asymmetric effect, whereas the implied volatility-augmented models were more useful in periods where the US dollar strengthened. We also performed a battery of pairwise weak dominance tests of Ziegel et al. (2020) and confirmed that implied volatility models should be preferred when the US dollar is strengthening, a result that held under a wide variety of loss functions.

For robustness, we addressed several aspects to verify the validity of our main results. First, we evaluated the usefulness of implied volatility over time to assess the stability of its performance. Second, in line with our motivation, we explored the performance of the baseline model and the implied volatility–augmented model, particularly in times of high market uncertainty, providing evidence that in such market conditions, the incorporation of implied volatility into forecasting models is highly beneficial. Third, we demonstrated that the results are robust to the choice of the loss function used to evaluate the precision of the ES predictions.

We also presented an economic application in which the investor is interested in reducing her expected shortfall. A hedging trading strategy was proposed and evaluated. We showed that irrespective of the preferred model (one without or with implied volatility), she can reduce her exposure to excessive risks. However, employing models that use implied volatility leads to less expensive hedging. We thus provided an example of a case where the superior statistical performance of an ES model translates into improved risk management. Finally, we showed that as the forecasting horizon increases to five-, 10-, and 22-day-ahead market returns, the benefits of the implied volatility models become uncertain.

Although the results presented in the paper compellingly support the use of implied volatility for ES prediction, our findings are restricted to only one currency pair, EUR/USD, the most important exchange rate market with the highest traded volume. Given the efficiency and liquidity of the FX market, it is likely that our results hold for other major currency pairs. However, we expect that the usefulness of IV likely differs from what we found for the EUR/USD exchange rate. Demand for protection might not be imbalanced between FX pairs, creating different hedging costs. We should also mention that not all FX pairs enjoy enough liquidity, which is needed for the implied volatility to represent an efficient estimate of the market's consensus. It follows that our research is relevant primarily for liquid derivative markets.

Data and code availability

In our research, the dataset primarily consists of information sourced from Bloomberg's paid database. Due to the licensing agreements with Bloomberg, we are legally bound to adhere to confidentiality terms. These terms explicitly prohibit the public sharing of this data, ensuring compliance with contractual obligations and maintaining the integrity of the confidentiality agreements in place.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Acerbi, C., & Tasche, D. (2002). Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes*, *31*(2), 379–388.

Aloud, M., Fasli, M., Tsang, E., Dupuis, R., & Olsen, R. (2013). Stylized facts of trading activity in the high frequency FX market: An empirical study. *Journal of Finance and Investment Analysis*, *2*(4), 145–183.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, *89*(4), 701–720.

Andersen, T. G., Dobrev, D., & Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, *169*(1), 75–93.

Andersen, T. G., Fusari, N., & Todorov, V. (2020). The pricing of tail risk and the equity premium: evidence from international option markets. *Journal of Business & Economic Statistics*, *38*(3), 662–678.

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

Bams, D., Blanchard, G., & Lehnert, T. (2017). Volatility measures and value-at-risk. *International Journal of Forecasting*, *33*(4), 848–863.

Barendse, S., Kole, E., & van Dijk, D. (2023). Backtesting value-at-risk and expected shortfall in the presence of estimation error. *Journal of Financial Econometrics*, *21*(2), 528–568.

Barndorff-Nielsen, O. E., Kinnebrock, S., & Shephard, N. (2008). *Measuring downside risk-realised semivariance*: *CREATES research paper 2008-42*.

Barrodale, I., & Roberts, F. (1978). An efficient algorithm for discrete l_1 linear approximation with linear constraints. *SIAM Journal on Numerical Analysis*, *15*(3), 603–611.

Basak, S., & Shapiro, A. (2001). Value-at-risk-based risk management: optimal policies and asset prices. *The Review of Financial Studies*, *14*(2), 371–405.

Basel Committee on Banking Supervision (1996). *Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements*. Bank for International Settlements.

Basel Committee on Banking Supervision (2010). *Basel III: A global regulatory framework for more resilient banks and banking systems*. Bank for International Settlements.

Basel Committee on Banking Supervision (2019). *Minimum capital requirements for market risk*. Bank for International Settlements.

Basel Committee on Banking Supervision (2023). *MAR - Calculation of RWA for market risk*. Bank for International Settlements.

Bayer, S., & Dimitriadis, T. (2022). Regression-based expected shortfall backtesting. *Journal of Financial Econometrics*, *20*(3), 437–471.

Becker, R., & Clements, A. E. (2008). Are combination forecasts of S&P 500 volatility statistically superior? *International Journal of Forecasting*, *24*(1), 122–133.

Becker, R., Clements, A. E., & White, S. I. (2006). On the informational efficiency of S&P500 implied volatility. *The North American Journal of Economics and Finance*, *17*(2), 139–153.

Becker, R., Clements, A. E., & White, S. I. (2007). Does implied volatility provide any information beyond that captured in model-based volatility forecasts? *Journal of Banking & Finance*, *31*(8), 2535–2549.

Bentes, S. R. (2015). A comparative analysis of the predictive power of implied volatility indices and GARCH forecasted volatility. *Physica A. Statistical Mechanics and its Applications*, *424*, 105–112.

Blair, B. J., Poon, S. H., & Taylor, S. J. (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. http://dx.doi.org/10.1016/S0304-4076(01)00068-9, URL: https://www.sciencedirect.com/science/article/pii/S0304407601000689.

Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, *192*(1), 1–18.

Busch, T., Christensen, B. J., & Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, *160*(1), 48–57.

Candila, V., Gallo, G. M., & Petrella, L. (2020). Using mixed-frequency and realized measures in quantile regression. Available at SSRN 3722927.

Canina, L., & Figlewski, S. (1993). The informational content of implied volatility. *The Review of Financial Studies*, *6*(3), 659–681.

Charoenwong, C., Jenwittayaroje, N., & Low, B. S. (2009). Who knows more about future currency volatility? *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *29*(3), 270–295.

Choi, P., & Nam, K. (2008). Asymmetric and leptokurtic distribution for heteroscedastic asset returns: the SU-normal distribution. *Journal of Empirical finance*, *15*(1), 41–63.

Chong, J. (2004). Value at risk from econometric models and implied from currency options. *Journal of Forecasting*, *23*(8), 603–620.

Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, *50*(2), 125–150.

Christoffersen, P., & Mazzotta, S. (2005). The accuracy of density forecasts from foreign exchange options. *Journal of Financial Econometrics*, *3*(4), 578–605.

Chun, D., Cho, H., & Ryu, D. (2019). Forecasting the KOSPI200 spot volatility using various volatility measures. *Physica A. Statistical Mechanics and its Applications*, *514*, 156–166.

Clements, A., Liao, Y., & Tang, Y. (2022). Moving beyond volatility index (VIX): Harnessing the term structure of implied volatility. *Journal of Forecasting*, *41*(1), 86–99.

Corlu, C. G., & Corlu, A. (2015). Modelling exchange rate returns: which flexible distribution to use? *Quantitative Finance*, *15*(11), 1851–1864.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Couperier, O., & Leymarie, J. (2020). Backtesting expected shortfall via multi-quantile regression.

Covrig, V., & Low, B. S. (2003). The quality of volatility traded on the over-the-counter currency market: A multiple horizons study. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *23*(3), 261–285.

Dacorogna, M., Mller, U., Olsen, R., & Pictet, O. (2001). *Defining efficiency in heterogeneous markets*. Taylor & Francis.

Dacorogna, M. M., Müller, U. A., Pictet, O. V., & Olsen, R. B. (1997). Modelling short-term volatility with GARCH and HARCH models. Available at SSRN 36960.
Day, T. E., & Lewis, C. M. (1992). Stock market volatility and the information content of stock index options. *Journal of Econometrics*, *52*(1–2), 267–287.
Dimitriadis, T., & Bayer, S. (2019). A joint quantile and expected shortfall regression framework.
Du, Z., & Escanciano, J. C. (2017). Backtesting expected shortfall: accounting for tail risk. *Management Science*, *63*(4), 940–958.
Embrechts, P., Liu, H., & Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, *66*(4), 936–949.
Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., & Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, *2*(1), 25–48.
Escanciano, J. C., & Lobato, I. N. (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics*, *151*(2), 140–149.
Ferreira, J. T. S., & Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions. *Journal of the American Statistical Association*, *101*(474), 823–829.
Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and osband's principle. *The Annals of Statistics*, *44*(4), 1680–1707.
Fissler, T., Ziegel, J. F., & Gneiting, T. (2015). Expected shortfall is jointly elicitable with value at risk-implications for backtesting. arXiv preprint arXiv:1507.00244.
Francq, C., et al. (2019). QML inference for volatility models with covariates. *Economic Theory*, *35*(1), 37–72.
Garman, M. B., & Kohlhagen, S. W. (1983). Foreign currency option values. *Journal of International Money and Finance*, *2*(3), 231–237.
Gerlach, R., Naimoli, A., & Storti, G. (2020). Time-varying parameters realized GARCH models for tracking attenuation bias in volatility dynamics. *Quantitative Finance*, *20*(11), 1849–1878.
Giot, P. (2005). Implied volatility indexes and daily value at risk models. *The Journal of Derivatives*, *12*(4), 54–64.
Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.
Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762.
Gurrola, P. (2007). Capturing fat-tail risk in exchange rate returns using SU curves: A comparison with the normal mixture and skewed student distributions. *The Journal of Risk*, *10*(2), 73.
Han, H., & Park, M. D. (2013). Comparison of realized measure and implied volatility in forecasting volatility. *Journal of Forecasting*, *32*(6), 522–533.
Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.
Haugom, E., Ray, R., Ullrich, C. J., Veka, S., & Westgaard, S. (2016). A parsimonious quantile regression model to forecast day-ahead value-at-risk. *Finance Research Letters*, *16*, 196–207.
Hoga, Y. (2021). The uncertainty in extreme risk forecasts from covariate-augmented volatility models. *International Journal of Forecasting*, *37*(2), 675–686.
Horváth, R., Lyócsa, Š., & Baumöhl, E. (2018). Stock market contagion in central and eastern europe: unexpected volatility and extreme co-exceedance. *The European Journal of Finance*, *24*(5), 391–412.
Hull, J., & Basu, S. (2018). *Options, futures, and other derivatives*. Pearson, URL: https://books.google.cz/books?id=GeOswwEACAAJ.
Jeon, J., & Taylor, J. W. (2013). Using CAViaR models with implied volatility for value-at-risk estimation. *Journal of Forecasting*, *32*(1), 62–74.
Johnson, N. L. (1949a). Bivariate distributions based on simple translation systems. *Biometrika*, *36*(3/4), 297–304.
Johnson, N. L. (1949b). Systems of frequency curves generated by methods of translation. *Biometrika*, *36*(1/2), 149–176.
Jorion, P. (1995). Predicting volatility in the foreign exchange market. *The Journal of Finance*, *50*(2), 507–528.
Kambouroudis, D. S., McMillan, D. G., & Tsakou, K. (2016). Forecasting stock return volatility: A comparison of GARCH, implied volatility, and realized volatility models. *Journal of Futures Markets*, *36*(12), 1127–1163.
Kim, J. S., & Ryu, D. (2015). Are the KOSPI 200 implied volatilities useful in value-at-risk models? *Emerging Markets Review*, *22*, 43–64.
Koenker, R. (2020). quantreg: Quantile regression. URL: https://CRAN.R-project.org/package=quantreg, R package version 5.55.
Koenker, R., & d'Orey, V. (1994). A remark on algorithm AS 229: Computing dual regression quantiles and regression rank scores. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *43*(2), 410—414.
Kourtis, A., Markellos, R. N., & Symeonidis, L. (2016). An international comparison of implied, realized, and GARCH volatility forecasts. *Journal of Futures Markets*, *36*(12), 1164–1193.
Le, T. H. (2020). Forecasting value at risk and expected shortfall with mixed data sampling. *International Journal of Forecasting*, *36*(4), 1362–1379.
Leiss, M., & Nax, H. H. (2018). Option-implied objective measures of market risk. *Journal of Banking & Finance*, *88*, 241–249.
Liang, C., Wei, Y., & Zhang, Y. (2020). Is implied volatility more informative for forecasting realized volatility: An international perspective. *Journal of Forecasting*, *39*(8), 1253–1276.
Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, *187*(1), 293–311.
Löser, R., Wied, D., & Ziggel, D. (2018). New backtests for unconditional coverage of expected shortfall. *Journal of Risk*, *21*(4), 1–21.
Lyócsa, Š., Plíhal, T., & Vỳrost, T. (2021). FX market volatility modelling: Can we use low-frequency data? *Finance Research Letters*, *40*, Article 101776.
Lyócsa, Š., & Stašek, D. (2021). Improving stock market volatility forecasts with complete subset linear and quantile HAR models. *Expert Systems with Applications*, Article 115416.
Lyócsa, Š., Todorova, N., & Vỳrost, T. (2021). Predicting risk in energy markets: Low-frequency data still matter. *Applied Energy*, *282*, Article 116146.
Lyócsa, Š., Vỳrost, T., & Baumöhl, E. (2019). Return spillovers around the globe: A network approach. *Economic Modelling*, *77*, 133–146.
Meng, X., & Taylor, J. W. (2020). Estimating value-at-risk and expected shortfall using the intraday low and range data. *European Journal of Operational Research*, *280*(1), 191–202.
Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3–46). NBER.
Müller, U. A., Dacorogna, M. M., Davé, R. D., Pictet, O. V., Olsen, R. B., & Ward, J. R. (1993). *Fractals and intrinsic time: A challenge to econometricians*: *Unpublished manuscript 130*, Olsen & Associates, Zürich.
Nadarajah, S., Zhang, B., & Chan, S. (2014). Estimation methods for expected shortfall. *Quantitative Finance*, *14*(2), 271–291.
Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.
Pati, P. C., Barai, P., & Rajib, P. (2018). Forecasting stock market volatility and information content of implied volatility index. *Applied Economics*, *50*(23), 2552–2568.
Patton, A., Politis, D. N., & White, H. (2009). Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White. *Econometric Reviews*, *28*(4), 372–375.
Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics*, *97*(3), 683–697.
Patton, A. J., Ziegel, J. F., & Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, *211*(2), 388–413.
Plíhal, T., & Lyócsa, Š. (2021). Modeling realized volatility of the EUR/USD exchange rate: Does implied volatility really matter? *International Review of Economics & Finance*, *71*, 811–829.
Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*(428), 1303–1313.
Politis, D. N., & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, *23*(1), 53–70.
Pong, S., Shackleton, M. B., Taylor, S. J., & Xu, X. (2004). Forecasting currency volatility: A comparison of implied volatilities and AR (FI) MA models. *Journal of Banking & Finance*, *28*(10), 2541–2563.

Poon, S. H., Blair, B., & Taylor, S. (2001). Forecasting S&P 100 volatility: Using high frequency data and implied volatility. *Journal of Econometrics*, *105*, 5–26.

Rockafellar, R. T., Uryasev, S., et al. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, *2*, 21–42.

Siu, Y. W. (2018). Volatility forecast by volatility index and its use as a risk management tool under a value-at-risk approach. *Review of Pacific Basin Financial Markets and Policies*, *21*(02), Article 1850010.

Slim, S., Dahmene, M., & Boughrara, A. (2020). How informative are variance risk premium and implied volatility for value-at-risk prediction? International evidence. *The Quarterly Review of Economics and Finance*, *76*, 22–37.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics*, *37*(1), 121–133.

Taylor, J. W. (2020). Forecast combinations for value at risk and expected shortfall. *International Journal of Forecasting*, *36*(2), 428–441.

Taylor, S. J., Yadav, P. K., & Zhang, Y. (2010). The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks. *Journal of Banking & Finance*, *34*(4), 871–881.

Timmermann, A. (2006). Forecast combinations. In *Handbook of economic forecasting, vol. 1* (pp. 135–196). Elsevier.

Wang, Y. H., & Wang, Y. Y. (2016). The information content of intraday implied volatility for volatility forecasting. *Journal of Forecasting*, *35*(2), 167–178.

Wayne, W. Y., Lui, E. C., & Wang, J. W. (2010). The predictive power of the implied volatility of options traded OTC and on exchanges. *Journal of Banking & Finance*, *34*(1), 1–11.

Xu, X., & Taylor, S. J. (1995). Conditional volatility and the informational efficiency of the PHLX currency options market. *Journal of Banking & Finance*, *19*(5), 803–821.

Ziegel, J. F., Krüger, F., Jordan, A., & Fasciati, F. (2020). Robust forecast evaluation of expected shortfall. *Journal of Financial Econometrics*, *18*(1), 95–120.