DeepVol: volatility forecasting from high-frequency data with dilated causal convolutions

Fernando Moreno-Pino & Stefan Zohren

Published online: 05 Sep 2024.

Submit your article to this journal ⬀

Article views: 2021

View related articles ⬀

View Crossmark data ⬀

Citing articles: 3 View citing articles ⬀

DeepVol: volatility forecasting from high-frequency data with dilated causal convolutions

FERNANDO MORENO-PINO*†‡ and STEFAN ZOHREN†§

†Oxford-Man Institute of Quantitative Finance, University of Oxford, Oxford, UK
‡Signal Processing and Learning Group, Universidad Carlos III de Madrid, Madrid, Spain
§Machine Learning Research Group, University of Oxford, Oxford, UK

Volatility forecasts play a central role among equity risk measures. Besides traditional statistical models, modern forecasting techniques based on machine learning can be employed when treating volatility as a univariate, daily time-series. Moreover, econometric studies have shown that increasing the number of daily observations with high-frequency intraday data helps to improve volatility predictions. In this work, we propose DeepVol, a model based on Dilated Causal Convolutions that uses high-frequency data to forecast day-ahead volatility. Our empirical findings demonstrate that dilated convolutional filters are highly effective at extracting relevant information from intraday financial time-series, proving that this architecture can effectively leverage predictive information present in high-frequency data that would otherwise be lost if realised measures were precomputed. Simultaneously, dilated convolutional filters trained with intraday high-frequency data help us avoid the limitations of models that use daily data, such as model misspecification or manually designed handcrafted features, whose devise involves optimising the trade-off between accuracy and computational efficiency and makes models prone to lack of adaptation into changing circumstances. In our analysis, we use two years of intraday data from NASDAQ-100 to evaluate the performance of DeepVol. Our empirical results suggest that the proposed deep learning-based approach effectively learns global features from high-frequency data, resulting in more accurate predictions compared to traditional methodologies and producing more accurate risk measures.

Keywords: Volatility forecasting; Realised volatility; High-frequency data; Deep learning; Dilated causal convolutions

1. Introduction

In recent years, the assessment of portfolios' risk through volatility measures has garnered significant attention (Brownlees and Gallo 2010), leading to the growing adoption of volatility conditional portfolios (Harvey *et al.* 2018). Various studies have reported overall gains in their Sharpe ratios (Moreira and Muir 2017) and reductions in the likelihood of observing extreme heavy-tailed returns when using them (Harvey *et al.* 2018). Consequently, there has been widespread interest in the development of volatility forecasting models.

In volatility forecasting, as in many other forecasting problems in economics and finance, the variable of interest is latent. This is exemplified in figure 1, where we employ a rolling window to estimate the unobserved latent volatility process from the observed returns. To address this complexity,

an unbiased estimator of this latent variable must be chosen as a proxy measure. Within the scope of volatility forecasting, the squared returns of an asset over a specific period of time is one of the most obvious realised volatility measures that can be interpreted as a conditionally unbiased estimator of the true unobserved conditional variance of the asset over that same time span (Patton 2011).

In a context where volatility forecasting methods offer considerable advantages, with an extensive literature pointing out that intraday volatility forecasts are important for improving the understanding of the risk involved in trading strategies (Bates 2019), pricing derivatives and the development of quantitative strategies (Engle and Sokalska 2012), as well as for risk management and trading applications (Stroud and Johannes 2014), among others, numerous studies have explored the challenges of intraday and daily volatility prediction. Delving into the latter, practitioners commonly rely on classic methodologies, with a heavy reliance on models that provide forecasts of daily volatility from daily returns. The Generalized Autoregressive Conditional Heteroskedasticity
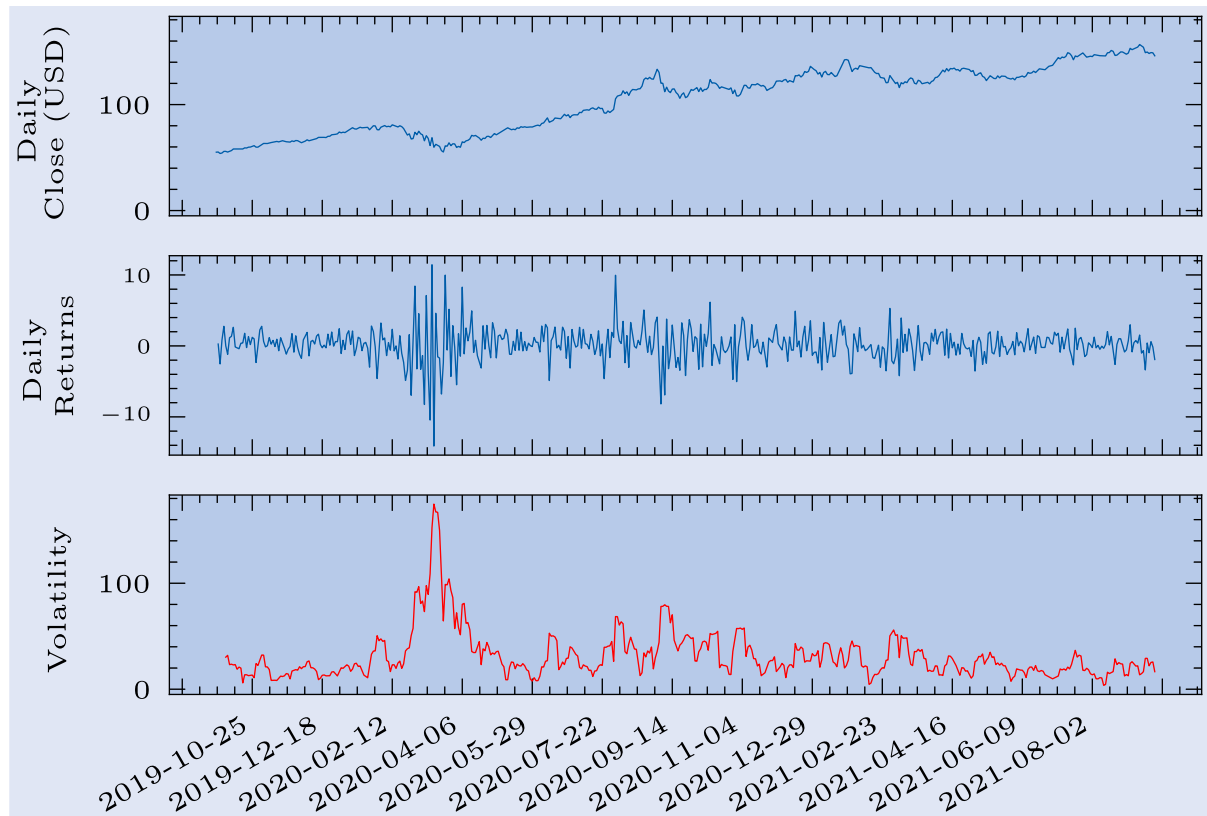
Figure 1. Apple's daily data. The top row shows the price trend, while the second row depicts the associated daily returns. In the bottom row, an estimation of the unobserved latent volatility process is calculated using a 5-day moving window over the daily returns. Notice that this method of estimating the volatility serves just as an approximation to the underlying latent volatility process, providing insights into the dynamic nature of market volatility.

(GARCH) model (Bollerslev 1986) and traditional stochastic volatility models (Andersen *et al.* 2006b, Calvet *et al.* 2006) are predominant in this regard. These models leverage volatility spillover effects and past volatility along with daily squared returns as the driving variables for predicting day-ahead volatility. However, their effectiveness is constrained by their inability to effectively leverage high-frequency data (Andersen *et al.* 2003, 2006b, Engle and Patton 2007) and their inefficacy to analysing simultaneously multiple assets in high-dimensionality.

Recent advancements have addressed some of these challenges by incorporating realised measures as predictors for realised volatility, thereby enhancing the prediction accuracy of classic models (Hansen *et al.* 2012, Patton and Zhang 2022). Realised measures, non-parametric estimators of an asset price variation over a time interval, extract and summarise information embedded in high-frequency data (Andersen *et al.* 2010). Moreover, ongoing research continues to introduce novel methodologies for computing realised measures. E.g.Pascalau and Poirier (2023) proposes a new approach to estimate and forecast realised volatility that contrast to the usual calendar approach.

However, methodologies that take advantage of realised measures require pre-processing steps to use them, as they cannot directly model the complex relations exhibited by intraday financial data. One concrete example of this pre-processing is the procedure followed to obtain the realised measures themselves, which summarises vectors of daily intraday high-frequency data into single scalars to avoid

managing the microstructure noise associated with the former. In contrast, our approach uses raw high-frequency data as input to the model, eliminating the need for data preprocessing and mitigating associated drawbacks, such as information loss during the aggregation of intraday data into daily realised measures. Our proposed model is capable of directly handling the microstructure noise linked to higher intraday data's sampling frequencies.

Among the methodologies employing realised measures, the HEAVY model (Shephard and Sheppard 2010) is of special appeal among industry practitioners (Karanasos *et al.* 2022, Papantonis *et al.* 2023, Yuan *et al.* 2022). HEAVY is based on insights from the ARCH architecture, with superior performance over other classical benchmarks, as shown in section 5. Nevertheless, the inability of realised measures-based models, such as HEAVY, Realised GARCH (Hansen *et al.* 2012), or HAR (Corsi 2009) to use unprocessed raw high-frequency data directly as input, exposes them to several disadvantages. Firstly, the dependence on the realised measures for day-ahead volatility forecasting artificially limits the amount of information these architectures use, which is not the case when using raw intraday data, as proven in section 5. Furthermore, some of the most used realised measures of volatility lack robustness to microstructure noise (Baars 2014), implying that the trained models may be based on biased data. Finally, methodologies based on realised measures often rely on manually designed handcrafted features, as the realised measures design themselves, formulated to optimise the trade-off between accuracy and

increasing computational costs. These issues, coupled with common model misspecification in classical model-based approaches, undermines their reported performances.

Here, we use Deep Neural Networks (DNN) (LeCun *et al.* 2015) to take advantage of the abundance of high-frequency data without prejudice, preventing the constraints of models based on realised measures in the context of day-ahead volatility forecasting. Despite the success of these DNN architectures in different areas, such as healthcare (Shamshirband *et al.* 2021), image recognition (Kaur and Gandhi 2020, Chen *et al.* 2021), time-series modelling and forecasting (Li *et al.* 2019, Jiménez Rama *et al.* 2023, Moreno-Pino *et al.* 2024b), and text analytics (Conneau *et al.* 2016), they have not been widely adopted for the problem of volatility forecasting, leading to a large gap between modern machine learning models and those applied in the volatility framework. Among DNN-based models, Recurrent Neural Networks (RNN) (Rumelhart *et al.* 1985) and Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber 1997) are the most popular approaches with regard to time-series forecasting (Lim and Zohren 2021). Furthermore, the addition of the attention mechanism (Bahdanau *et al.* 2014) into these base architectures allowed them to focus on the most relevant input data while producing predictions, making them especially prominent in fields such as Natural Language Processing (NLP). These advances also lead to the appearance of Transformer models (Vaswani *et al.* 2017), which were initially introduced for NLP, and later used for the problem of time-series forecasting (Li *et al.* 2019, Moreno-Pino *et al.* 2023). These models are also applied in the context of financial time-series through diverse variations (Su 2021, Scalzo *et al.* 2021, Arroyo *et al.* 2022, Lin *et al.* 2022, Christensen *et al.* 2023, Song *et al.* 2023, Zhu *et al.* 2023, Arroyo *et al.* 2024, Ghosh *et al.* 2024, Souto and Moradi 2024). More specifically, regarding volatility forecasting, various studies have delved into the application of deep-learning architectures, such as LSTM (Yu and Li 2018), Convolutional Neural Networks (CNN) (Vidal and Kristjanpoller 2020), Graph Neural Networks (GNN) (Chen and Robert 2021, Zhang *et al.* 2022), Transformer models (Ramos-Pérez *et al.* 2021), and NLP-based word embedding techniques (Rahimikia and Poon 2020, Rahimikia *et al.* 2021). Furthermore, models combining traditional volatility forecasting methods with deep-learning techniques can be found in the literature (Kim and Won 2018, Mademlis and Dritsakis 2021), as well as other approaches using DNN as calibration methods for implying volatility surfaces (Horvath *et al.* 2019), proving how neural network-based approaches work as complex pricing function approximators.

We capitalise on the increased availability of high-frequency data. In this work, we employ a Dilated Causal Convolutions (DCC)-based model. This architecture, initially proposed as a fully probabilistic model for audio generation (Oord *et al.* 2016), with equivalents for image-related problems (Van Oord *et al.* 2016), can handle lengthy sequences of data effectively without a significant increase in the number of parameters thanks to the use of dilated connections. In the literature, other works use DCC in the context of realised volatility forecasting. More specifically, Reisenhofer *et al.* (2022) propose a model based on

dilated convolutions, strongly inspired by the well-known Heterogeneous Autoregressive (HAR) model (Corsi 2009). However, their approach does not use unprocessed raw intraday high-frequency data as input. Conversely, it still bases its predictions on the pre-computed daily realised variance, therefore requiring pre-processing steps to obtain the indispensable realised measures for forecasting the one-step-ahead volatility. This, in our judgment, does not fully explore the capabilities of DCC-based methodologies of exploiting a more dynamic representation of the intraday data. Hence, models adopting DCC-based approaches that operate from daily data still succumb to the limitations enumerated previously.

Motivated by the improved performance of classical methods that employ realised measures (Hansen *et al.* 2012, Shephard and Sheppard 2010), we propose using DCCs to bypass the estimation of these non-parametric estimators of assets' variance, aiming to tackle the volatility forecasting problem from a data-driven perspective. The proposed model, DeepVol, offers several advantages for volatility forecasting. Primarily, it does not require any pre-processing steps, as the model directly uses raw high-frequency data as input. Furthermore, DeepVol is not bounded to static realised measures whose use may be counter-productive, i.e. the optimal realised measure to use may vary depending on the traded assets' liquidity. Instead, through the attention mechanism and internal non-linearities, DeepVol intelligently performs the required transformations over the input data to maximise the accuracy of the predictions, combining relevant intraday datapoints and merging them for each day's volatility forecast, dinamically adapting to different scenarios. Moreover, through the use of dilated convolutions, DeepVol's large receptive field easily processes long sequences of high-frequency data, enabling the model to exponentially increase its input window without triggering an unrestrained increase in the model's complexity while performing the predictions. While DeepVol's operation does not directly produce any type of realised measure —meaning the model does not explicitly construct an ex-post estimate of the returns variation— its aggregation of high-frequency data resembles how realised measures condense intraday data into daily statistics. Unlike the computation of these realised measures, DeepVol undertakes this aggregation in a data-driven manner, diverging from traditional approaches used for computing realised measures of volatility. DeepVol's ability to dynamically select and weight the most relevant time-steps from the conditioning range at each time instant stands out as its primary distinction from traditional methods computing realised measures, allowing DeepVol to hierarchically integrate the most relevant high-frequency data into the predictions. We perform extensive experiments to show the effectiveness of the proposed architecture, which consistently outperforms the base models used by practitioners.

This paper presents three main contributions. Firstly, it empirically demonstrates the advantages of DCCs for forecasting realised volatility using high-frequency data, providing a data-driven solution that consistently outperforms classical methodologies. Further, the proposed model overcomes the limitations of classical methods, such as model misspecification or their inability to directly use intraday data to perform
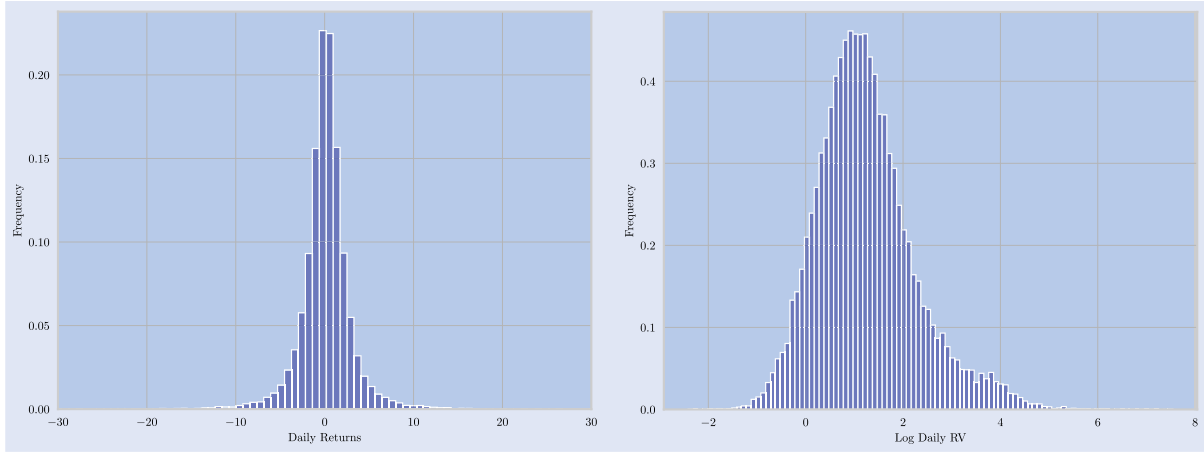
Figure 2. **Left:** Distribution of daily returns. **Right:** Distribution of realised volatility (in logs).

the forecast. Secondly, we offer an analysis showing how DeepVol optimises the balance between extracting signals from high-frequency data and minimising the microstructure noise inherent to higher sampling frequencies. The reported results are consistent with studies validating this trade-off for constructing realised measures. Thirdly, the proposed volatility forecasting model generates appropriate risk measures through its predictions in an out-of-sample forecasting task, both in low and high volatility regimes. Moreover, we evaluate the proposed model's generalisation capabilities on out-of-distribution stocks, demonstrating DeepVol's capabilities to transfer learning as it performs accurate predictions into data distributions not observed during the training phase.

The structure of the paper is as follows. Section 2 details the dataset used, while section 3 contains a brief overview on volatility forecasting, describing the baselines used for benchmarking purposes and the metrics that will be utilised for model comparison. Section 4 presents the proposed model, which is empirically evaluated in section 5. Finally, section 6 summarises the findings and provides concluding remarks.

2. Data and model inputs

2.1. Data

To demonstrate the effectiveness of the proposed architecture, we use high-frequency data from various stocks in the U.S. market for fitting the models. More specifically, Deep-Vol and the baseline architectures are trained and tested using two years of NASDAQ-100 data. We set the start of our sample at September 30, 2019, with September 30, 2021 being the last date. High-frequency data of different sampling frequencies (granularities), i.e. 1, 5, 15, 30, and 60 minutes, is used in our analysis. Tables A1 and A2 in Appendix provide the summary statistics (mininum, maximum, mean, standard deviation, median, skewness, and kurtosis) for the daily close-to-close returns and the realised variance, see equation (2). From these tables, we can see that all of the returns exhibit excess kurtosis, indicating that the distribution of the returns is fat-tailed. Further, figure 2 presents the estimated densities of the daily returns and the daily logarithmic realised volatility, where a departure from normality is evident. Finally, figures 3 and 4 show the cross-sectional average

for both returns and realised volatility, along with different percentiles.

DeepVol does directly perform its prediction from raw high-frequency data, unlike the baseline models, which prior to training require the estimation of daily statistics. The analyses conducted in this work are based on financial returns, which allow us to transform the original assets' price trend into a quasi-stationary process:

$$r_{i,t} = \log\left(\frac{p_{i,t}}{p_{i-1,t}}\right), \tag{1}$$

where $p_{i,t}$ is the last price of an asset in the ith interval on day t, and $r_{i,t}$ is the return over this interval, at the specified sampling frequency, i.e. 1, 5, 15, 30 or 60 minutes.

We have chosen not to include overnight information in our analysis, which aligns with previous studies (Engle and Sokalska 2012, Zhang *et al.* 2022). This decision is also supported by recent research indicating a lack of clear methods for integrating overnight session data into daily volatility assessments (Pascalau and Poirier 2023).

2.1.1. Baselines: data preparation. The benchmark models used in this work are divided into two categories. Firstly, we consider methods that solely use daily returns to perform day-ahead volatility forecasts. Secondly, we examine methods that take advantage of realised measures in the forecasts.

Regarding models depending exclusively on daily returns, these are obtained through an analogous procedure to the one followed to retrieve intraday returns through equation (1), but using daily returns instead of intraday data. Concerning methods utilising realised measures, and in consonance with other studies (Shephard and Sheppard 2010, Hansen *et al.* 2012, Harvey *et al.* 2018), we focus on the realised variance for the scope of this work, which is a proxy measure of the volatility that has been shown to be less noisy than daily squared returns (Andersen and Bollerslev 1998, Barndorff-Nielsen and Shephard 1999). It is obtained as follows:

$$RV_t = \sum_{i=1}^{I} r_{i,t}^2, \tag{2}$$

where $r_{i,t}$ is the ith intraday return for day t, see equation (1). Previuous works have shown that the sum of
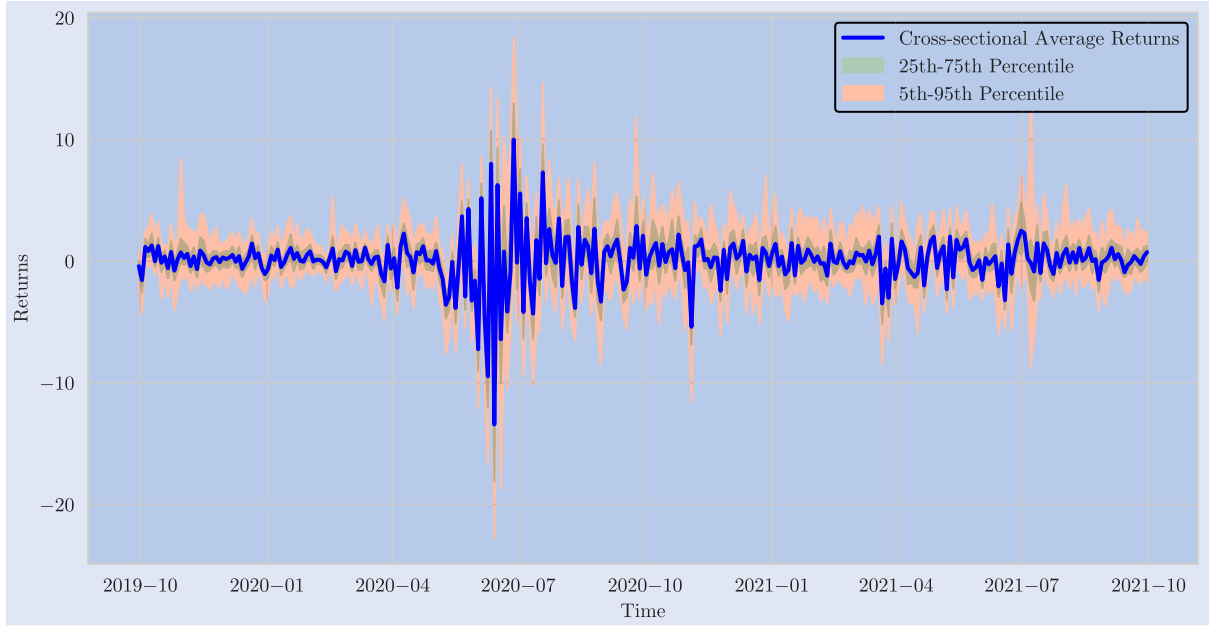
Figure 3. The blue curve represents the cross-sectional average of daily returns across the analysed stocks, with the green area covering the 25th percentile to the 75th percentile, and the red area covering the 5th percentile to the 95th percentile.
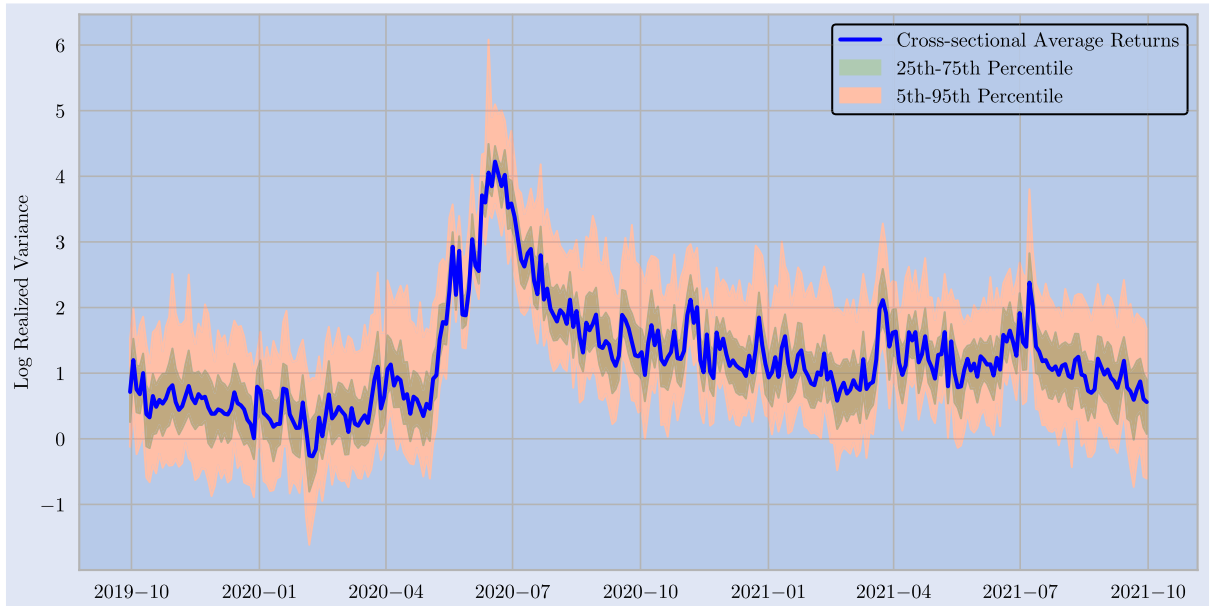
Figure 4. The blue curve represents cross-sectional average of daily realised volatility across the analysed stocks, with the green area covering the 25th percentile to the 75th percentile, and the red area covering the 5th percentile to the 95th percentile.

the squared intraday returns is a consistent estimator the unsobservable latent integrated variance of stocks (Andersen *et al.* 2001, Barndorff-Nielsen and Shephard 2002).

Various works (Andersen *et al.* 2001, Ait-Sahalia *et al.* 2005) study the usage of different sampling frequencies to compute the realised variance through equation (2). Selecting a specific intraday's data sampling frequency to compute the realised volatility (e.g. 5 or 30 min) involves the optimisation of a trade-off: while we aim to maximise the number of datapoints used, higher sampling frequencies entail an increase of the microstructure noise, which we want to minimise. We use 5-minutes intraday returns to compute the realised variance through equation (2), as this sampling frequency is usually accepted as the optimal value (Andersen *et al.* 2001, Bandi and Russell 2006).

2.1.2. DeepVol: data preparation. As previously discussed, DeepVol differs from classical methods by directly utilising raw high-frequency data without the need for pre-processing. Notably, DeepVol employs a rolling window approach for producing day-ahead volatility forecasts, σ_t^2, using a window of past intraday high-frequency returns, $r_{i,t}$, as input data. In section 5, we explore the optimal window size, referred to hereafter as the receptive field. This parameter determines the number of past days used for predicting day-ahead volatility, essentially indicating how far in the past the first element of the time series that influences the model's prediction lies. DeepVol is confined to use a specific receptive field, e.g. the previous day's high-frequency data. This non-recursive architecture reduces the input data length required by the model, which translates into

faster training in comparison to purely autoregressive architectures. Additionally, we examine the optimal intraday data sampling frequency. Finally, we should remark that DeepVol's operative differs from most state-of-the-art forecasting architectures, which produce predictions at the same granularity (sampling frequency) as the model's input data. Therefore, DeepVol must effectively learn the necessary relationships between high-frequency data and daily volatility, implicitly performing this time-domain transformation.

3. Baseline models and metrics

3.1. Baseline models

Most of the models commonly used for volatility forecasting can be traced back to Autoregressive Conditional Heteroscedastic (ARCH) models (Engle 1982). This family of models assume volatility clustering (Cont 2007), i.e. large shocks in prices tend to cluster together. ARCH-based models evolved into the well-known Generalized Autoregressive Conditional Heteroscedastic (GARCH) model (Bollerslev 1986), which is still widely utilised among industry participants. A GARCH(p, q) process is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{3}$$

where ω is the model's bias, q is the number of lags (order) of the observed volatility, σ_t^2; and p is the number of lags of the innovations, ε_t. In turn, the returns of prices are related to the innovations by:

$$r_t = \mu + \varepsilon_t, \tag{4}$$

where μ is the expected return (usually set to zero), and the volatility is related to these innovations by means of the residuals, e_t:

$$\varepsilon_t = \sigma_t e_t, \quad e_t \sim \mathcal{N}(0, 1). \tag{5}$$

The model's parameters $\{\mu, \omega, \alpha, \beta\}$ can be estimated by performing maximum-likelihood estimation of the joint distribution $f(\varepsilon_1, \ldots, \varepsilon_T; \{\mu, \omega, \alpha, \beta\})$. The simplest GARCH model consists on a GARCH(1, 1) process where $\sigma_t = 1$ and $\mu = 0$. Several variations leading to new architectures to address the volatility forecasting problem have been developed from the GARCH model. Here, we select some of them for benchmarking purposes. The integrated GARCH (IGARCH) model (Engle and Bollerslev 1986), modifies the design of the previous model to grant a longer memory in the autocorrelation of the squared returns, allowing the model to react in a more persistent way to the impact of past squared shocks. Also, IGARCH imposes the following restriction on the model's parameters:

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1, \tag{6}$$

which makes the resulting process a weakly stationary one, since the mean, variance, and autocovariance are finite and constant over time. The idea behind the IGARCH model motivated the development of the Fractionally Integrated GARCH (FIGARCH) process (Baillie *et al.* 1996), which is able to capture long-term volatility persistence and clustering features. To do so, it integrates a fractional difference operator (lag operator) L into the conditional variance:

$$\sigma_t^2 = \omega + \left[1 - \beta L - \phi L (1 - L)^d \right] \varepsilon_t^2 + \beta \sigma_t^2, \tag{7}$$

where $0 < d < 1$ is known as the fractional differencing parameters. The FIGARCH model has been widely used thanks to its ability to capture the volatility's persistence and integrate it into its predictions (Cochran *et al.* 2012, Biage 2019). Threshold ARCH (TARCH) models (Rabemananjara and Zakoian 1993) are also used for benchmarking purposes. The main difference compared to previous methodologies is that TARCH models divide the distribution of the innovations into disjoint intervals, which are later approximated by a linear function on the conditional standard deviation (Zakoian 1994). TARCH models are therefore capable of separately considering the influence of positive and negative innovations:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j}^2 \mathbb{I}_{\varepsilon_{t-j} < 0}, \tag{8}$$

where $\mathbb{I}_{(\cdot)}$ is the indicator function. The main characteristic of TARCH and other threshold-based approaches, such as TGARCH (Park *et al.* 2009), is their ability to detect abrupt disruptions in the time-series through the indicator function, which may be replaced with a continuous function if a smoother transition is desired.

Volatility usually exhibits asymmetric characteristics. This property has led to the development of different asymmetric ARCH type models. For example, the Asymmetric Power ARCH (APARCH) model (Ding *et al.* 1993) assumes a parametric form for the conditional heteroskedasticity's powers. It defines the variance dynamics as follows:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i \left(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}\right)^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta, \tag{9}$$

where we now also have to estimate $\delta > 0$, and γ. APARCH models nest many other volatility frameworks that can be obtained by imposing restrictions on the APARCH model's parameters. A similar idea leads to the Asymmetric GARCH model (AGARCH) (Engle and Ng 1993), which captures the asymmetry in the volatility by using an impact curve associated with the α_i parameter:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \left(\varepsilon_{t-i} - \gamma_i\right)^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \tag{10}$$

Most of the mentioned models, like IGARCH or APARCH, impose restrictions on the parameters in practice, as equation (6) states. These restrictions are lifted in the Exponential GARCH (EGARCH) model (Nelson 1991), which is

defined as:

$$\ln \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \left(|\epsilon_{t-i}| + \gamma_i \epsilon_{t-i} \right) + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2. \quad (11)$$

As evidenced by the definition above, the EGARCH model integrates one powerful volatility clustering assumption into its architecture: negative shocks at time $t-1$ produce a stronger impact on the value of the volatility at time t than positive shocks do, allowing for asymmetric effects between positive and negative asset returns. This asymmetry is known in the volatility forecasting literature as leverage effect (Bouchaud *et al.* 2001).

All the methods described previously operate through the usage of daily returns. However, as mentioned above, more recent proposals have included the usage of realised measures obtained from the high-frequency data as additional input features for daily volatility forecasting. Among these methodologies, the High-Frequency-Based Volatility (HEAVY) model (Shephard and Sheppard 2010), has shown superior forecasting capabilities (Shephard and Sheppard 2010, Noureldin *et al.* 2012, Sheppard and Xu 2019). Formally, the model is defined as follows:

$$\text{var} \left(r_t \mid \mathcal{F}_{t-1}^{\text{HF}} \right) = \sigma_t^2 = \omega + \alpha RM_{t-1} + \beta \sigma_{t-1}^2,$$
$$\mathbb{E} \left(\text{RM}_t \mid \mathcal{F}_{t-1}^{\text{HF}} \right) = \mu_t = \omega_R + \alpha_R RM_{t-1} + \beta_R \mu_{t-1}, \quad (12)$$

where r_t denotes daily returns, RM_t denotes daily realised measures, and $\mathcal{F}_{t-1}^{\text{HF}}$ denotes the high-frequency data utilised to obtain these realised measures. In previous equation, the restrictions $\{\omega, \alpha \geq 0, \beta \in [0, 1)\}$ are imposed on the variation of the returns, and $\{\omega_R, \alpha_R, \beta_R \geq 0, \alpha_R + \beta_R \in [0, 1)\}$ on the realised measures' evolution, while observing the following variables:

$$r_t = \sqrt{\sigma_t^2} z_t,$$
$$x_t = \mu_t z_{RV,t}^2, \quad (13)$$

with:

$$\begin{pmatrix} z_t \\ z_{RV,t} \end{pmatrix} \sim \mathcal{N}(0, I). \quad (14)$$

Equation (12) shows that HEAVY consists of two parts. While σ_t^2 explains the development of the unobserved conditional variance, μ_t is responsible for explaining the development of the realised measures. The HEAVY model is clearly motivated by GARCH methodologies, which makes it simple to understand while reporting additional gains in the performance. For further details about it, like parameters inference, we refer the readers to Shephard and Sheppard (2010).

Further, we include both the Realized GARCH (Hansen *et al.* 2012) and Realized EGARCH (Hansen and Huang 2016) models, which also consider high-frequency/intraday data as predictors to estimate daily volatility. The Realized GARCH

model is defined as:

$$\text{var} \left(r_t \mid \mathcal{F}_{t-1}^{\text{HF}} \right) = \sigma_t^2 = \omega + \alpha RM_{t-1} + \beta \sigma_{t-1}^2,$$
$$\mathbb{E} \left(\text{RM}_t \mid \mathcal{F}_{t-1}^{\text{HF}} \right) = \xi + \phi \sigma_t^2 + \tau_1 z_t + \tau_2 \left(z_t^2 - 1 \right) + u_t, \quad (15)$$

with u_t being independent and following a standard normal distribution. Additionally, following the leverage function proposed in Hansen *et al.* (2012), we include z_t and z_{t-1}^2 as a way to incorporate leverage effects, generating an asymmetric response in volatility to return shocks.

The main difference of the Realized-GARCH model with respect to the HEAVY model is that the conditional variance at day t, that is, σ_t^2, is also used as a regressor for estimating the realised measure at time t, which entails that both α and β parameters are also based on the equation explaining the realised measure evolution. This second equation also depends on the random variable z_t, which is not observed directly, as stated in equation (14).

The Realized EGARCH model, on the other hand, considers the log transformation and integrates the leverage effect into the conditional variance part, which results in:

$$\text{var}(r_t \mid \mathcal{F}_{t-1}^{\text{HF}}) = \exp(\omega + \alpha \log(RM_{t-1}) + \beta \log(h_{t-j})$$
$$+ \tau_1 z_t + \tau_2 (z_t^2 - 1)),$$
$$\log(RM_t) = \xi + \phi \log(h_t) + u_t, \quad (16)$$

with u_t being normally distributed and independent of z_t when $t \neq s$.

We also include an ARFIMA model (Granger 1980, Granger and Joyeux 1980) as baseline, which is defined as:

$$\varphi(L)(1 - L)^d \left(RM_t - \mu \right) = \theta(L) \epsilon_t, \quad (17)$$

where $\varphi(L) = 1 - \sum_{i=1}^{p} \varphi_i L^i$ and $\theta(L) = 1 - \sum_{j=1}^{q} \theta_j L^j$ are the autoregressive and moving average lag polynomials. The fractional differencing parameter d is responsible of capturing the long-memory properties and ϵ_t is the error term.

We additionally consider the Heterogeneous Autoregressive model (HAR), originally proposed by Corsi (2009) as a response to the drawbacks of ARFIMA models. The HAR model has a autoregressive structure and includes averages of daily realised variances over different time horizons. Corsi (2009) proposes using aggregation periods of one, five, and twenty two for computing the realised variances, see equation (2), then define the HAR model as:

$$\mathbb{E} \left(\text{RM}_t \mid \mathcal{F}_{t-1}^{\text{HF}} \right) = \beta_0 + \beta_1 RM_{t-1} + \beta_2 RM_{t-1}^{(w)}$$
$$+ \beta_3 RM_{t-1}^{(m)} + \epsilon_t, \quad (18)$$

where the parameters $\{\beta_0, \beta_1, \beta_2, \beta_3\}$ are usually estimated with ordinary least squares (OLS) or weighted least squares (WLS) (Patton and Sheppard 2015). Weekly $RM_{t-1}^{(w)}$ and monthly $RM_{t-1}^{(m)}$ realised measures are given by:

$$RM_t^{(w)} = \frac{1}{5} \left(RM_{t-1}^{(d)} + RM_{t-2}^{(d)} + RM_{t-3}^{(d)} + RM_{t-4}^{(d)} + RM_{t-5}^{(d)} \right)$$
$$RM_t^{(m)} = \frac{1}{22} \left(RM_{t-1}^{(d)} + RM_{t-2}^{(d)} + \cdots + RM_{t-21}^{(d)} + RM_{t-22}^{(d)} \right)$$
$$(19)$$

Finally, we include as baselines a MLP (Rumelhart *et al.* 1986) and a LSTM (Hochreiter and Schmidhuber 1997) that operate directly over the high-frequency data, as done in previous research papers (Zhang *et al.* 2022, Masini *et al.* 2023).

3.2. Evaluation metrics

In this section, we define a series of metrics that will be used to assess the day-ahead volatility forecast of our proposed architecture against previously defined baseline models. The Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) constitute two of the most common error functions to evaluate the performance of volatility forecasting architectures. While a number of articles focus entirely on those two metrics to report performance (Shen *et al.* 2021, Izzeldin *et al.* 2019), we complement them with the usage of the Symmetric Mean Absolute Percentage Error (SMAPE). This relative error measure has both a lower and upper bound, contrary to the Mean Absolute Percentage Error (MAPE), and it is scale independent. Also, the Maximum Error (ME) is used to illustrate which models produce the more significant inaccuracies: poor performance adapting to new regimes, as volatility shocks, lead certain models to substantial momentary discrepancies between the forecast and the actual volatility, which leads to an increase in the ME. We complement the ME with the Median Absolute Error (MedAE), an outliers-robust metric. Lastly, we include the Quasi Log-Likelihood (QLIKE), which has proven to be a noise robust loss function in the volatility proxy. Both the QLIKE and the RMSE will be used as loss functions to optimise the model's parameters during section 5, while the rest of the metrics will be used to assess the models' performance. We summarise the definitions of all metrics below.

$$\ell_{MAE}(\sigma_t^2, \hat{\sigma}_t^2) = \frac{1}{T} \sum_{t=1}^{T} \left| \sigma_t^2 - \hat{\sigma}_t^2 \right|,$$

$$\ell_{rmse}(\sigma_t^2, \hat{\sigma}_t^2) = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\sigma_t^2 - \hat{\sigma}_t^2 \right)^2},$$

$$\ell_{SMAPE}(\sigma_t^2, \hat{\sigma}_t^2) = \frac{1}{T} \sum_{t=1}^{T} \frac{\left| \sigma_t^2 - \hat{\sigma}_t^2 \right|}{(\sigma_t^2 + \hat{\sigma}_t^2)/2},$$

$$\ell_{ME}(\sigma_t^2, \hat{\sigma}_t^2) = \max \left(\left| \sigma_t^2 - \hat{\sigma}_t^2 \right| \right),$$

$$\ell_{MedAE}(\sigma_t^2, \hat{\sigma}_t^2) = median(\left| \sigma_t^2 - \hat{\sigma}_t^2 \right|),$$

$$\ell_{QLIKE}(\sigma_t^2, \hat{\sigma}_t^2) = \frac{1}{T} \sum_{t=1}^{T} \log \left(\hat{\sigma}_t^2 \right) + \frac{\sigma_t^2}{\hat{\sigma}_t^2},$$

where $\hat{\sigma}_t^2$ and σ_t^2 represent the volatility forecast and the volatility proxy measure, respectively, with T the total amount of rolling forecasts.

3.3. Model confidence set

When assessing the predictions of various models for a specific task, a lower value of the loss function for one model in comparison to another implies superior forecasting performance of the former over the latter. However, drawing such conclusions based on a single loss function or a single sample may not be reliable. Therefore, in addition to employing multiple loss functions in our evaluation, as mentioned in section 3.1, we also utilise statistical tests to determine whether a particular model outperforms another.

Various works have focussed on developing frameworks to determine whether a particular model outperforms another. Among them, the Diebold-Mariano (DM) test (Diebold and Mariano 2002) and the Superior Predictive Ability (SPA) test (Hansen 2005) are popular and reliable approaches. However, the Model Confidence Set (MCS) method (Hansen *et al.* 2011) has several advantages over them. For instance, the MCS method has fewer assumptions and does not require the selection of a benchmark, making it suitable for scenarios where identifying a clear baseline is challenging. Additionally, it allows for multiple optimal models, which contrasts with the DM test, which is only applicable for pairwise comparisons between two models. In our study, where we evaluate 16 models in parallel (see section 5), the MCS test is more appropriate. Therefore, we adopt MCS test in our empirical evaluation, following the approach of Hansen *et al.* (2011) and Zhang *et al.* (2022), among others.

The MCS test aims to identify a subset of models $\mathcal{M}^*$ with significantly superior performance from a set of model candidates $\mathcal{M}_0$, at a specified level of confidence. It operates through an elimination process that is based on sequentially testing the null hypothesis:

$$H_{0,\mathcal{M}} : \mathbb{E}(\Delta L_{ij,t}) = 0, \quad \text{for all } i,j \in \mathcal{M} \subset \mathcal{M}_0, \quad (20)$$

where $\Delta L_{ij,t}$ represents the difference in loss between models i and j at time t, as measured by a specific loss function L, see section 3.2.

The MCS procedure involves a series of significance tests aimed at identifying objects significantly inferior to others in $\mathcal{M}_0$ for elimination. The set of superior objects is defined by:

$$\mathcal{M}^* = \{i \in \mathcal{M}_0 : \mathbb{E}(\Delta L_{ij,t}) \leq 0 \text{ for all } j \in \mathcal{M}_0\}. \quad (21)$$

Through an equivalence test, $\delta_{\mathcal{M}}$, and elimination rule, $e_{\mathcal{M}}$, we can test the hypothesis H_0 for any $\mathcal{M} \subset \mathcal{M}_0$, identifying the model that is to be removed from out set of candidates in case H_0 is rejected. After repeating the two tests, we obtain a set of 'surviving' objects; the model confidence set. For further details, we refer the readers to Hansen *et al.* (2011).

4. Model

4.1. Problem definition

Considering a set of assets, $\Delta \in \mathbb{R}^d$, where $d \in \mathbb{N}$ denotes the dimension of the input vector, with $T \in \mathbb{N}$ days' intraday

high-frequency data associated to them, $\{\mathbf{r}_t^{1:J}\}_{t=1}^T$, where $\mathbf{r}_t^{1:J} = (r_t^1, r_t^2, \ldots, r_t^J)$ are the intraday returns of the tth day, with T being referred to as receptive field, and with $J \in \mathbb{N}$ the length of each day's intraday data, our goal is to forecast the day-ahead realised volatility:

$$\hat{\sigma}_{T+1}^2 = f_\theta \left(r_{t=1}^1, r_{t=1}^2, \ldots, r_{t=1}^J, r_{t=2}^1, \ldots, r_{t=T}^1, \ldots, r_{t=T}^J \right), \tag{22}$$

where $f_\theta : \mathbb{R}^d \to \mathbb{R}^m, m \in \mathbb{N}$, is a function implemented through a Dilated Causal Convolutions (DCC)-based neural network, with $\theta \in \Theta$ being the learnable parameters of the model from a set $\Theta \in \mathbb{R}^n$, for some $n \in \mathbb{N}$.

These parameters fully specify the corresponding volatility forecast. Therefore, we aim to obtain the set of optimal parameters $\hat{\theta} \in \Theta$ that minimises the difference between the forecasted volatility, $\hat{\sigma}_t^2$, and the volatility's proxy measure σ_t^2 for the considered assets:

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \mathcal{L} \left(f_\theta(\Delta), \sigma_t^2(\Delta) \right), \tag{23}$$

where $\mathcal{L}$ is the selected metric for evaluating the forecast accuracy.

4.2. Dilated causal convolutions

Our volatility forecasting proposal, DeepVol, uses DCCs as a technique to integrate the high-frequency information into the realised volatility prediction. The deployment of such an architecture allows the usage of a large receptive field that can handle lengthy sequences of data effectively without a significant increase in the number of parameters thanks to the use of dilated connections, yielding improved computational efficiency.

The proposed architecture consists of L convolutional layers. The convolution operation performed by the first layer, between the input sequences x and the kernel k, can be defined as follows:

$$F^{(l=1)}(t) = \left(x *_d k^{(l=1)} \right)(t) = \sum_{\tau=0}^{s-1} k_\tau^{(l=1)} \cdot x_{t-d\tau}, \tag{24}$$

being d the dilation factor and k the filter, with size $s \in \mathbb{Z}$. For each of the rest lth layers, we can define the convolution operation as:

$$F^{(l)}(t) = \left(F^{(l-1)} *_d k^{(l)} \right)(t) = \sum_{\tau=0}^{s-1} k_\tau^{(l)} \cdot F_{t-d\tau}^{l-1}(t). \tag{25}$$

As previous equations state, the inner product performed by the dilated causal convolutions is based on entries that are a fixed number of steps apart from each other, contrary to CNN and Causal-CNN, which operate with consecutive entries. Furthermore, each of the layers in this hierarchical structure defines the kernel operation as an affine function acting between layers:

$$k^{(l)} : \mathbb{R}^{N_l} \longrightarrow \mathbb{R}^{N_{l+1}}, \quad 1 \le l \le L. \tag{26}$$

We should mention that, through the introduction of residual connections, initially proposed by He *et al.* (2016), certain layers of the model can deviate from this traditional sequential flow of information. In standard sequential architectures, the output of layer l typically serves as the input for layer $l + 1$, as stated in equation (26). However, residual connections introduce an exception by connecting layers that are not directly adjacent, thereby creating parallel paths for information flow. This bypasses some intermediate layers and allows for the direct transmission of information between non-adjacent layers. By introducing additional paths for information flow, residual connections enable the usage of deeper models with larger receptive fields and help alleviate challenges associated with training deep networks, such as the vanishing gradient problem (Hochreiter *et al.* 2001). The complete operative of the proposed model can be defined as follows:

$$\sigma_{T+1}^2(\{\mathbf{r}_t^{1:J}\}_{t=1}^T) = \alpha_0 + \sum_{l=1}^L \alpha_l \, \sigma_{ReLu}(F^{(l)}(\{\mathbf{r}_t^{1:J}\}_{t=1}^T)), \tag{27}$$

where $\sigma_{ReLu} : \mathbb{R} \mapsto \mathbb{R}$ is the selected non-linearity and $\{\alpha_0, \ldots, \alpha_l, \ldots, \alpha_L\}$ is a set of weights applied to the convolutional operations. Further, the attention mechanism —we use the classic attention mechanism (Bahdanau *et al.* 2014), which contrast with self-attention (Vaswani *et al.* 2017), as used in Transformer models, where the computational cost scales quadratically in the length of the input sequence, see Moreno-Pino *et al.* (2024a, 2024b)— weights each layer's inputs to reflect their importance at each time-step. Figure 5 illustrates how the dilated causal convolutions used by the model operate, where a receptive field with previous T days' intraday data is processed through a hierarchy of these dilated convolutions to forecast the day-ahead realised volatility.

DeepVol, like any deep-feed-forward neural network, is approximating the volatility's unknown function through sample pairs of input and output data (x, y). Formally speaking, DeepVol is approximating some function $f_\theta(\cdot)$ which is not available in closed form by finding the optimal model's parameters $\hat{\theta}$ derived from the best function approximation $f_\theta^*(\cdot)$.

5. Experiments

In this section, DeepVol's volatility forecasts will be evaluated and compared with the baseline models described in section 3.1. For benchmarking purposes, we will utilise the metrics described in section 3.2. Besides classic out-of-sample forecast comparisons, we perform different experiments to present some additional insights into the inner workings of DeepVol. We validate DeepVol's behaviour while varying the intraday data sampling frequency, studying the discrepancy in model behaviour when trained on different granularity regimes. In close relation to this, we also explore the usage of different receptive field sizes and how this affects the model performance. Further, we analyse the inclusion of realised measures as an extra input to the model, studying if its addition can improve the forecasting accuracy. Finally, we analyse the generalisation capabilities of our model on out-of-distribution-stocks. Besides the models
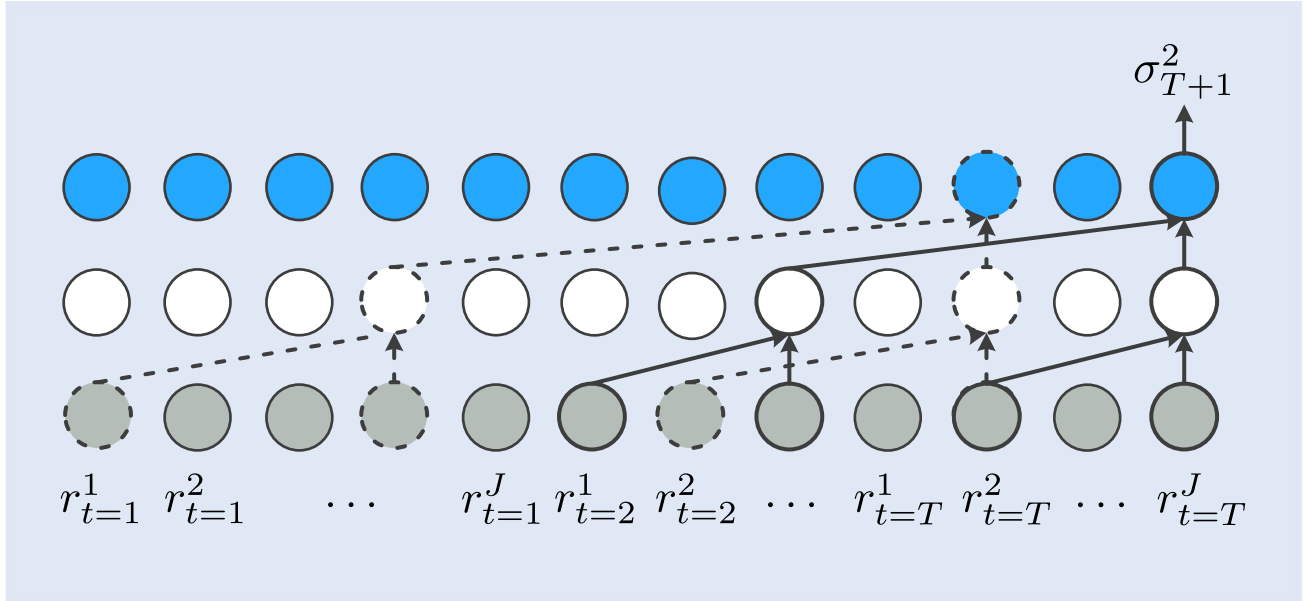
Figure 5. DeepVol's operation over the high-frequency data through dilated causal convolutions. The dilation factor grows exponentially, allowing an increase in the receptive field without increasing the model's complexity.

Table 1. DeepVol's hyperparameters for out-of-sample forcasting.

Hyperparameter	Optimal Value
Learning Rate	$1e - 3$
Batch Size	512
Number of Epochs	1000
Early Stopping (Patience)	50
Optimizer	Adam
Kernel Size	3
Conditioning Range	1
Sampling Frequency	5 mins
Num. Blocks	2
Num. Layers	6
Residual Channels	32
Dilation Channels	64
Skip Channels	128
End Channels	64
Loss Function	QLIKE

presented in section 3.1, we also include a martingale process for comparison purposes.

5.1. Training details

We use the Quasi Log-Likelihood as loss function to train the model parameters. We choose Adaptive Moment Estimation Algorithm (ADAM) (Kingma and Ba 2014) as optimiser, even though different experiments were conducted exploring the usage of Averaged Stochastic Gradient Descent (ASGD) (Kingma and Ba 2014) and Limited Memory BFGS (L-BFGS) (Liu and Nocedal 1989). While the usage of these optimisers usually entails smoother predictions, the reported performance declined with respect to ADAM, hence, they were not considered further. Early stopping is used during the training process through the validation set. DeepVol is

implemented in Pytorch-Lightning (Falcon 2019), and the experiments are conducted using a NVIDIA Titan Xp GPU.

Through the use of validation sets —we detail how the data is divided into train/validation/test sets in each experiment—, we assess a range of hyperparameters to identify the ones that perform best on the validation data. These hyperparameters, together with their optimal value for the out-of-sample forecast experiment, are summarised in table 1, we refer the readers to Kingma and Ba (2014), Goodfellow *et al.* (2016), Prechelt (1998), Oord *et al.* (2016), He *et al.* (2016), Mikolov *et al.* (2013), Bahdanau *et al.* (2014), and Vaswani *et al.* (2017) for further details on the employed methodologies and the validation of their hyperparameters. Further, in section 5.2, we present a detailed analysis of this validation process, focussing specifically on different sampling frequencies and receptive field sizes, having selected previously the optimal values for the remaining hyperparameters. We believe that examining these two parameters offers an insightful comparison with previous studies that investigated the same parameters while computing realised measures from high-frequency data (Andersen *et al.* 2001, Bandi and Russell 2006). Finally, we also validate the number of hidden layers and neurons per layer for the MLP, and the number of layers and units for the LSTM.

5.2. Receptive field and sampling frequency analysis

Besides being two hyperparameters in our model, the receptive field size and the intraday sampling frequency can shed light on the inner workings of DeepVol when analysed further. As mentioned in section 2.1.1, a number of studies have validated the optimal intraday data sampling frequency for computation of the realised measures from high-frequency data (Andersen *et al.* 2006a, Corradi and Distaso 2006, Hansen and Lunde 2006), commonly concluding that using a granularity of 5 or 10 minutes minimises the microstructure noise effect while maximising the usage of high-frequency information.

Table 2. Receptive field and sampling frequency study.

Sampling Frequency	Receptive Field	MAE	RMSE	SMAPE	QLIKE	ME	MedAE
1 min	1	4.096	8.462	0.287	342.313	71.749	2.396
5 min	1	**3.903**	8.457	**0.279**	**340.779**	71.779	2.008
	2	4.429	9.495	0.308	367.209	**70.036**	1.756
	3	4.054	**8.379**	0.285	343.359	70.457	2.334
15 min	1	3.993	8.436	0.283	343.412	70.915	2.185
	2	4.651	10.437	0.312	365.893	70.926	1.836
	3	5.817	9.520	0.323	362.235	72.338	4.577
	5	6.736	10.217	0.336	366.192	72.240	5.594
30 min	1	4.259	9.699	0.318	689.633	75.793	1.632
	2	4.503	10.140	0.327	789.326	79.345	1.724
	3	4.473	9.931	0.324	784.843	75.059	1.705
	5	4.705	10.802	0.326	833.966	83.416	1.676
	10	4.732	11.084	0.327	853.981	85.656	**1.591**
60 min	1	4.988	10.516	0.297	366.509	82.207	2.402
	2	5.616	12.596	0.324	709.082	99.828	2.178
	3	5.441	12.615	0.319	688.996	99.612	2.017
	5	5.520	12.456	0.326	714.871	95.763	2.071
	10	4.997	11.186	0.322	706.917	87.096	1.927

In this section, we study this same trade-off in the proposed deep-learning architecture, analysing the effect that using different sampling frequencies and receptive field sizes has on the model performance. To do so, we use the NASDAQ-100 dataset described in section 2.1, splitting it into three main folds. The training set contains 12 months of intraday data, from September 30, 2019, to September 30, 2020. The validation set contains three months of data, from October 01, 2020, to December 31, 2020. Finally, the test set contains nine months of data, from January 1, 2021, to September 30, 2021. Metrics introduced in section 3 are used to analyse the models' performance.

Table 2 collects the results of an analysis whose main objective is not only to select the hyperparameters values with the best performance on the validation set, but also to evaluate if DeepVol's performance is robust to increasing the receptive field or modifying the sampling frequency. Firstly, it is interesting to remark that using intraday data from one day with a sampling frequency of 5-minutes proves to be optimal. This scenario reports the best results with regard to all the considered metrics but the MedAE. Further, this result is particularly relevant considering that it aligns with previous literature suggesting an optimal granularity around 5 minutes, like Andersen *et al.* (2001), Bandi and Russell (2006), or Andersen and Bollerslev (1997), where it is affirmed that 'the five-minute horizon is short enough that the accuracy of the continuous record asymptotics underlying our realised volatility measures work well, and long enough that the confounding influences from market microstructure frictions are not overwhelming'. Even so, it's important to note that our observation of 5-minute intervals appearing optimal may be influenced by our task of predicting realised variance computed from intraday data sampled at 5-minute intervals. Therefore, we want to clarify that we do not propose a 5-minute frequency as universally optimal for volatility forecasting. Instead, in the specific context of predicting realised

volatility in our study, a 5-minute granularity proves effective. In real-world scenarios, practitioners should properly validate both the sampling frequency and the conditioning range, taking into account the specific volatility proxy being used.

Secondly, the increment of the receptive field leads to a degradation of performance. This indicates that, for the proposed architecture and the selected proxy measure of volatility, all the relevant information for forecasting the day-ahead volatility can be obtained from the previous day high-frequency data and increasing the receptive field yields more conservative predictions that deteriorate its performance.

Thirdly, the best performance in terms of MedAE is obtained when using a 30 minutes sampling frequency, together with a receptive field of ten days. This result can be directly related to the hypothesis previously mentioned: a longer receptive field leads to a more conservative forecast, resulting in a lower Median Absolute Error. In this scenario, the model is less prone to forecast volatility jumps, a behaviour commonly associated with integrating momentum indicators; however, this specific setup leads to a deterioration in all the other metrics. The growth in the receptive field size prevents the model from forecasting more drastic changes in the presence of volatility shocks, leading to more conservative predictions compared to those reported when using just the previous day's intraday data.

5.3. Out-of-sample forecast

This section aims to provide an out-of-sample performance comparison between the proposed model and some classical methodologies widely used in the finance industry. As mentioned in previous section, the NASDAQ-100 dataset is spitted into three main folds. The results presented in this section are computed over the test set, containing nine months

Table 3. Out-of-sample forecast: experiments results for the NASDAQ-100 dataset.

Method	MAE	RMSE	SMAPE	QLIKE	ME	MedAE
martingale	5.180	11.410	0.324	747.480	96.654	**1.614***
TARCH	4.849	10.320	0.301	351.310	71.659	2.804
IGARCH	5.008	10.534	0.302	351.702	72.048	2.797
FIGARCH	4.631	10.356	0.294	349.245	71.050	2.460
APARCH	4.730	10.096	0.299	349.974	**70.088***	2.859
AGARCH	4.833	10.304	0.324	351.217	71.215	2.819
EGARCH	4.793	10.180	0.300	348.614	70.615*	2.917
HEAVY	4.565	10.239	0.292	343.490	72.404	2.545
Real-GARCH	4.598	10.062	0.299	344.282	70.896	2.488
Real-EGARCH	4.686	10.103	0.299	345.739	71.061	2.512
ARFIMA	4.620	10.345	0.296	344.800	71.901	2.480
HAR	4.587	10.306	0.293	343.652	72.116	2.451
MLP	4.820	10.607	0.307	350.994	73.302	2.939
LSTM	4.441	10.032	0.295	345.465	72.650	2.570
DeepVol-RM	4.383	9.823	0.284	342.433*	72.016	2.405
DeepVol	**3.903***	**8.457***	**0.279***	**340.779***	71.779	2.008

Notes: An asterisk (*) indicates models that are included in the MCS at the 5% significance level.

of data, from January 1, 2021, to September 30, 2021. The validation set is used both to identify the optimal hyperparameters and to monitor the early stopping criteria. For this specific study, a sampling frequency of 5-minutes for the intraday data and a receptive field of one day (using t-day intraday's data to predict the realised volatility for day $t + 1$) are utilised, as they are found to be optimal in section 5.2. Additionally, table 1 provides a summary of the remaining hyperparameters' values.

Table 3 summarises the out-of-sample forecast performance of the different models we evaluate. Notice that, in addition to the baselines included in section 3.1, we introduce a variation of the proposed model, termed DeepVol-RM. Inspired by the HAR model (Corsi 2009), DeepVol-RM utilises realised measures computed over the previous twenty-two days as input instead of high-frequency data. The primary objective of incorporating this variation of DeepVol is to empirically verify whether there may exist predictive information in the high-frequency data that might be overlooked when computing a realised measure. We doing so by comparing the prediction accuracy of a DCC-based model fed with raw high-frequency data (DeepVol) against a DCC-based model that receives realised volatility as input (DeepVol-RM). It can be seen that the proposed architecture, DeepVol, improves upon the baseline results for most metrics, with the exception of ME and MedAE. Peculiarly, for the later, the martingale process proves to provide the best results. Considering that the MadAE is an outlier-robust error function, this behaviour is not surprising, as the martingale process is the most conservative among the evaluated strategies. For this particular error metric, DeepVol is the second-best in performance terms.

As mentioned before, the evaluated baseline methodologies operate in a recurrent manner, utilising past data, while DeepVol uses just the previous day's intraday information for the day-ahead prediction. Considering these facts, DeepVol's good performance with respect to the MedAE is especially surprising, as noisier behaviour could be expected due to the lack of recurrence. Furthermore, some of the baseline models

evaluated, as the HEAVY model, integrate momentum indicators into their architecture, something that we do not explicitly model in DeepVol. Consequently, DeepVol's accurate predictions in terms of MAE and RMSE are particularly interesting considering how the model maintains a low MedAE. In conclusion, the proposed architecture shows robustness in the presence of volatility shocks and avoids an escalation on the ME and MedAE as unstable methods would report.

Concerning DeepVol-RM, which leverages the DeepVol architecture but utilises realised measures from previous days instead of intraday data, it reports the second best results, outperforming the HAR model, from which it took inspiration, and the LSTM and MLP models, which directly utilise intraday data. Its superior performance over HAR is not surprising, as DeepVol-RM benefits from weighting the importance of all realised measures from the previous twenty-two days at inference time, whereas HAR can just aggregates this information, as reflected in equation (19). This observation is consistent with the findings reported in Reisenhofer *et al.* (2022), where a similar implementation of a dilated causal convolutions-based network is discussed. Additionally, DeepVol's superior performance compared to the LSTM model suggests that DCC-based models are more effective than recurrent ones in handling high-frequency data for volatility forecasting.

Table 4 extends the results of table 3, displaying the improvement/degradation for each evaluated method relative to a basic martingale process and the HEAVY model. For the former, we aim to report how much improvement each model provides over the most basic modelling of the problem. For the latter, and considering that the HEAVY model is the best performer among the classic models used as baselines, slightly outperforming the HAR model —as also found in Vis *et al.* (2017), where they conclude that HEAVY models are generally superior to HAR-type models—, a direct comparison with it is particularly insightful for analysing DeepVol's performance. It can be seen that DeepVol thoroughly outperforms HEAVY concerning the MAE, RMSE, SMAPE,

Table 4. Out-of-sample forecast: percentage of improvement/degradation over the martingale process and the HEAVY model, for each of the evaluated models.

Method	MAE	RMSE	SMAPE	QLIKE	ME	MedAE
	Improvement over martingale (%)					
martingale	–	–	–	–	–	–
TARCH	6.398	9.553	7.099	53.001	25.860	− 73.730
IGARCH	3.320	7.677	6.790	52.948	25.458	− 73.296
FIGARCH	10.598	9.238	9.259	53.277	26.490	− 52.410
APARCH	8.687	11.516	7.716	53.179	**27.486**	− 77.138
AGARCH	6.699	9.693	0.000	53.013	26.319	− 74.659
EGARCH	7.471	10.780	7.407	53.361	26.940	− 80.731
HEAVY	11.873	10.263	9.877	54.047	25.089	− 57.677
Real-GARCH	11.236	11.814	7.716	53.941	26.650	− 51.673
Real-EGARCH	9.537	11.455	7.716	53.746	26.479	− 55.638
ARFIMA	10.811	9.334	8.642	53.872	25.610	− 53.656
HAR	11.448	9.676	9.568	54.025	25.387	− 51.859
MLP	6.950	7.038	5.247	53.043	24.160	− 82.094
LSTM	14.266	12.077	10.494	53.900	25.107	− 60.408
DeepVol-RV	15.386	13.908	12.346	54.188	25.491	− 49.009
DeepVol	**24.653**	**25.881**	**13.889**	**54.410**	25.736	**− 24.411**
	Improvement over HEAVY (%)					
martingale	− 13.472	− 11.437	− 10.959	− 117.613	− 33.492	**36.579**
TARCH	− 6.212	− 0.791	− 3.082	− 2.277	1.029	− 10.181
IGARCH	− 9.704	− 2.881	− 3.425	− 2.391	0.492	− 9.906
FIGARCH	− 1.446	− 1.143	− 0.685	− 1.675	1.870	3.340
APARCH	− 3.614	1.397	− 2.397	− 1.888	**3.120**	− 12.342
AGARCH	− 5.871	− 0.635	− 10.959	− 2.250	1.642	− 10.771
EGARCH	− 4.995	0.576	− 2.740	− 1.492	2.471	− 14.621
HEAVY	–	–	–	–	–	–
Real-GARCH	− 0.723	1.729	− 2.397	− 0.231	2.083	3.808
Real-EGARCH	− 2.651	1.328	− 2.397	− 0.655	1.855	1.293
ARFIMA	− 1.205	− 1.035	− 1.370	− 0.381	0.695	2.550
HAR	− 0.482	− 0.654	− 0.342	− 0.047	0.398	3.690
MLP	− 5.586	− 3.594	− 5.137	− 2.185	− 1.240	− 15.486
LSTM	2.716	2.022	0.685	− 0.322	0.024	− 1.733
DeepVol-RV	3.987	4.062	2.740	0.308	0.536	5.497
DeepVol	**14.502**	**17.404**	**4.452**	**0.789**	0.863	21.097

Table 5. Robustness study.

	Realised Measures Window	MAE	RMSE	SMAPE	QLIKE	ME	MedAE
DeepVol	1	4.130	8.720	0.286	345.150	72.292	2.193
+	2	6.804	10.036	0.335	369.529	69.309	5.728
RV	3	6.899	10.008	0.340	371.762	69.903	6.577
	1	**3.903**	8.457	**0.279**	**340.779**	71.779	2.008
DeepVol	2	4.429	9.495	0.308	592.600	78.690	**1.756**
	3	4.054	**8.379**	0.285	343.359	**70.457**	2.334

Notes: DeepVol + RV merges DeepVol's predictions with the realised variance through a linear layer and additional non-linearities.

and MedAE errors, while the differences in QLIKE and ME are tighter. As previously mentioned, we consider particularly interesting DeepVol' ability to overperform the rest of the models while proving a robust noise behaviour, avoiding an escalation in the ME and MedAE while performing more accurate predictions.

5.4. Robustness study

The analysis in section 5.2 has demonstrated that employing a receptive field of one day and a sampling frequency of 5-minutes yields the most accurate results for forecasting day-ahead realised volatility. Moreover, we aim to explore potential benefits of incorporating realised measures into our methodology. The rationale behind this notion is that integrating information from realised measures of previous days as additional input would enable DeepVol to observe a broader window of past data, thereby complementing high-frequency data with additional historical information from the time series. Note that this study differs from the variation of Deep-Vol previously introduced, DeepVol-RM, which forecasts volatility solely based on realised measures from previous

Table 6. Out-of-sample-stocks forecast.

Method	MAE	RMSE	SMAPE	QLIKE	ME	MedAE
martingale	9.673	35.235	0.324	2142.795	341.457	2.169
TARCH	8.525	28.178	0.295	893.795	282.096	3.236
IGARCH	9.208	27.753	0.312	947.409	279.080	3.982
FIGARCH	7.805	26.752	0.299	899.955	267.533	3.581
APARCH	8.179	26.749	0.297	896.063	**265.910***	3.557
AGARCH	7.928	26.486	0.295	893.682	269.577	3.191
EGARCH	8.180	26.767	0.297	897.432	277.022	3.530
HEAVY	8.315	26.322	0.294*	**874.409***	277.780	2.158
Real-GARCH	8.402	26.523	0.297	890.894	890.894	2.301
Real-EGARCH	8.231	26.102	0.296	888.909	279.693	2.341
ARFIMA	7.993	25.001	0.295	888.031	278.882	2.446
HAR	8.184	25.042	0.294*	889.471	279.302	2.510
MLP	8.902	27.636	0.305	902.342	289.023	2.790
LSTM	7.998	26.941	0.296	898.907	898.907	2.494
DeepVol-RV	8.514	26.832	0.296	889.880	282.271	2.583
DeepVol	**7.288***	**23.396***	**0.292***	894.283	275.255	**1.927***

Notes: Generalisation Study: experiments results for the NASDAQ-100 dataset. An asterisk (*) indicates models that are included in the MCS at the 5% significance level.

days, inspired by the approach proposed in Corsi (2009), and does not directly use intraday data as raw input to the DCC-based model.

To incorporate the realised measures into DeepVol, we make a slight adjustment to its architecture by adding a linear output as a final layer. This layer combines the results of the dilated convolutions performed over the high-frequency data, which has a sampling frequency of 5-minutes and a receptive field of 1 day, with the realised measures from the previous N days. The various values of N evaluated can be seen in the column labelled 'Realised Measures Window' in table 5. Each component of these two is weighted by its corresponding terms.

The reported results are shown in table 5, where it can be noted that our DNNs-based proposal does not benefit from the inclusion of realised measures as an additional input feature. Adding the past realised measures results in an analogous behaviour to increasing the receptive field, highlighting again that DeepVol is especially efficient in utilising recent high-frequency data for volatility forecasting, not requiring a more extended lookback window to do so.

5.5. *Generalisation and transfer learning analysis*

Previous experiments used all NASDAQ-100 tickers during training, validation, and testing, preserving a portion of the dataset's dates for the out-of-sample forecast. In this section, in addition, we split the dataset into two folds in the cross-section. During training and validation, just half of the tickers are used, while the other half is utilised for testing. For training purposes, we use the first four months of data corresponding to the first half of tickers, that is, from September 30, 2019, through January 31, 2020. Further, we keep the last of these months, that is, January 01 to January 31, 2020, as validation set for hyperparameters selection. The model is later tested on the remainder tickers, using data from February 01, 2020, through September 30, 2021. This set-up allows us to evaluate the quality of the models' forecasts during the volatility shocks provoked by the

COVID-19 crisis, which started in February, 2020. In this scenario, we can evaluate our model's generalisation capabilities, predicting the day-ahead volatility for tickers that were not previously available. As done in section 5.3, a 5-minutes sampling frequency and a receptive field of one day are used after validating their values analogously to section 5.2. The results of this out-of-sample-stocks forecast study are collected in table 6.

In these conditions, DeepVol still report the best MAE, RMSE, and SMAPE results, while the HEAVY model reports a better QLIKE than the rest of the evaluated methods. Concerning the MeadAE, DeepVol reports the best results, immediately followed by the martingale process, which still outperforms the rest of the baseline models. Furthermore, DeepVol-RM's performance has worsened in comparison to section 5.3, presumably due to the fact that in this new environment of volatility shocks, past information contained in the realised measures used as input is less informative. These results, which are in general similar to the obtained ones in the out-of-sample forecast study of section 5.5, confirm that DeepVol still shows a conservative behaviour in this new forecast scenario, proving its generalisation capabilities to transfer learning from training to test, presumably learning global features of the data that allow the model to perform well on out-of-distribution data. As in section 5.3, table 7 reports the improvement/degradation for each evaluated method with respect to a basic martingale process and the HEAVY model on the test set tickers.

Finally, figures 6 to 9 exemplify how the evaluated models generalise and transfer learning from the training tickers to the test distribution. Model forecasts are juxtaposed with the daily squared returns, allowing a direct comparison between forecasts from DeepVol and the baselines. Note that classical methodologies yield smoother predictions, a phenomenon especially visible in the HEAVY model as it integrates a momentum indicator. This conservative behaviour, while advantageous for stable conditions, clearly poses a disadvantage in terms of slower adaptation to volatility shocks, as seen in the sharp movements during the COVID-19 crisis in

Table 7. Out-of-sample-stocks forecast: percentage of improvement/degradation over the martingale process and the HEAVY model. for each of the evaluated models.

Method	MAE	RMSE	SMAPE	QLIKE	ME	MedAE
	Improvement over martingale (%)					
martingale	–	–	–	–	–	–
TARCH	11.863	20.030	9.068	58.288	17.385	− 49.161
IGARCH	4.804	21.235	3.671	55.786	18.268	− 83.565
FIGARCH	19.315	24.075	7.711	58.001	21.650	− 65.089
APARCH	15.440	24.083	8.267	58.183	**22.125**	− 63.973
AGARCH	18.043	24.829	9.066	58.294	21.051	− 47.105
EGARCH	15.443	24.032	8.452	58.119	18.871	− 62.737
HEAVY	14.037	25.295	9.200	**59.193**	18.649	0.534
Real-GARCH	13.138	24.725	8.390	58.424	17.994	− 6.076
Real-EGARCH	14.906	25.920	8.698	58.516	18.088	− 7.920
ARFIMA	17.366	29.045	9.007	58.557	18.326	− 12.760
HAR	15.392	28.928	9.315	58.490	18.203	− 15.711
MLP	7.969	21.566	5.922	57.889	15.356	− 28.619
LSTM	17.315	23.539	8.698	58.050	16.976	− 14.973
DeepVol-RV	11.980	23.850	8.698	58.471	17.333	− 19.076
DeepVol	**24.660**	**33.599**	**9.932**	58.266	19.388	**11.165**
	Improvement over HEAVY (%)					
martingale	− 16.330	− 33.859	− 10.132	− 145.056	− 22.924	− 0.537
TARCH	− 2.529	− 7.048	− 0.145	− 2.217	− 1.554	− 49.962
IGARCH	− 10.741	− 5.434	-6.090	− 8.349	− 0.468	− 84.551
FIGARCH	6.140	− 1.632	− 1.640	− 2.921	3.689	− 65.975
APARCH	1.632	− 1.622	− 1.028	− 2.476	**4.273**	− 64.854
AGARCH	4.660	-0.623	− 0.148	− 2.204	2.953	− 47.895
EGARCH	1.624	− 1.691	− 0.824	− 2.633	0.273	− 63.611
HEAVY	–	–	–	–	–	–
Real-GARCH	− 1.046	− 0.762	− 0.892	− 1.885	− 0.805	− 6.646
Real-EGARCH	1.010	0.837	− 0.552	− 1.658	− 0.689	− 8.500
ARFIMA	3.873	5.020	− 0.213	− **1.558**	− 0.397	− 13.366
HAR	1.575	4.864	0.127	− 1.723	− 0.548	− 16.332
MLP	− 7.060	− 4.991	− 3.610	− 3.195	− 4.047	− 29.310
LSTM	3.812	− 2.350	− 0.552	− 2.802	− 2.056	− 15.591
DeepVol-RV	− 2.393	− 1.936	− 0.552	− 1.769	− 1.617	− 19.716
DeepVol	**12.357**	**11.116**	**0.806**	− 2.273	0.909	**10.688**

Figure 6. Out-of-sample-stocks: HEAVY's and DeepVol's forecast on PYPL. Green dotted vertical lines mark the forecast start.

2020 and 2021, evident in the associated figures. All the evaluated models reacted to the bigger of these shocks, the 2020 stock market crash, starting in February 2020, in one way or another. However, during the minor shocks that followed that year, baseline predictions are almost negligible with the exception of IGARCH in figure 9. HEAVY and EGARCH exhibit an invariable behaviour in this turbulent environment, showing a lack of adaptability to changing conditions.

Finally, we should remark that DeepVol requires just one day of intraday data to perform the out-of-sample volatility forecasting, unlike classical methodologies that operate recursively and require a sufficiently long window of past data. This places DeepVol in an advantaged position in situations of low data availability, such as the inclusion of new tickers in the stock market, as it does not require a long horizon of historical to perform its predictions.

Figure 7. Out-of-sample-stocks: HEAVY's and DeepVol's forecast on QCOM. Green dotted vertical lines mark the forecast start.

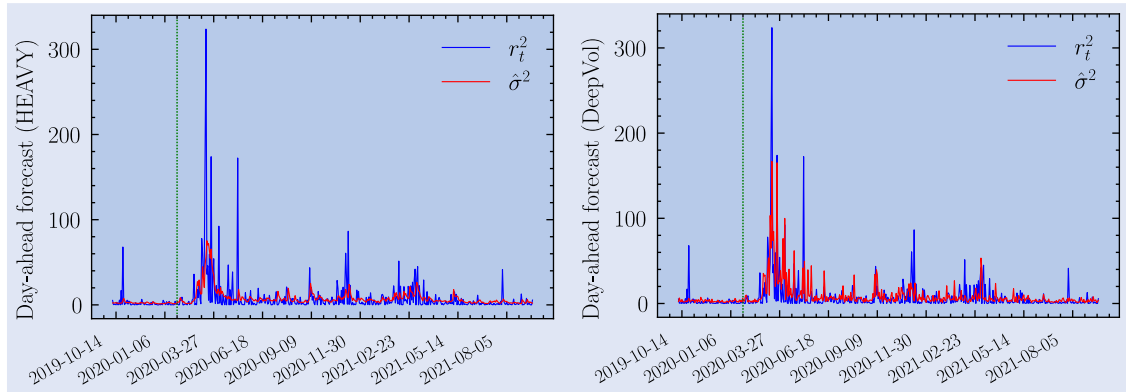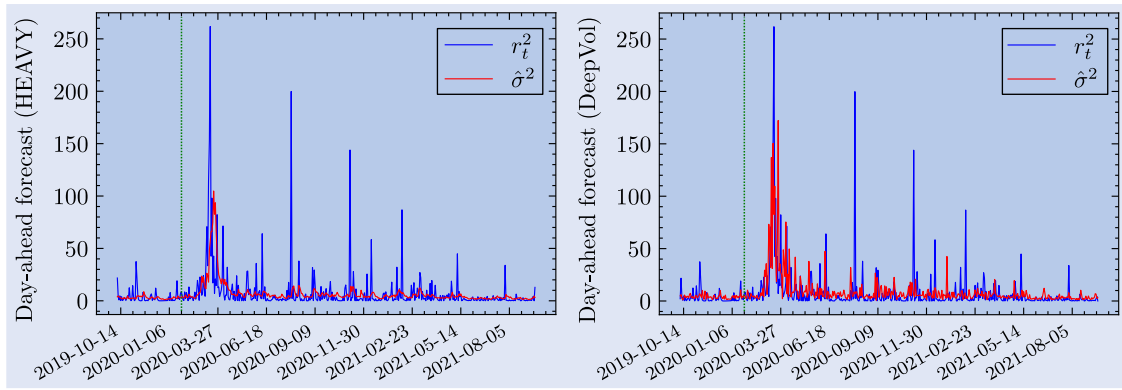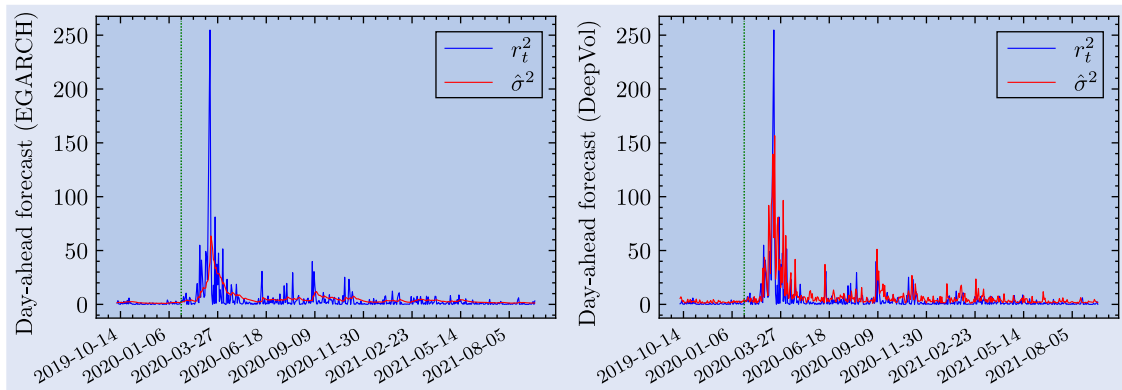Figure 8. Out-of-sample-stocks: EGARCH's and DeepVol's forecast on MSFT. Green dotted vertical lines mark the forecast start.
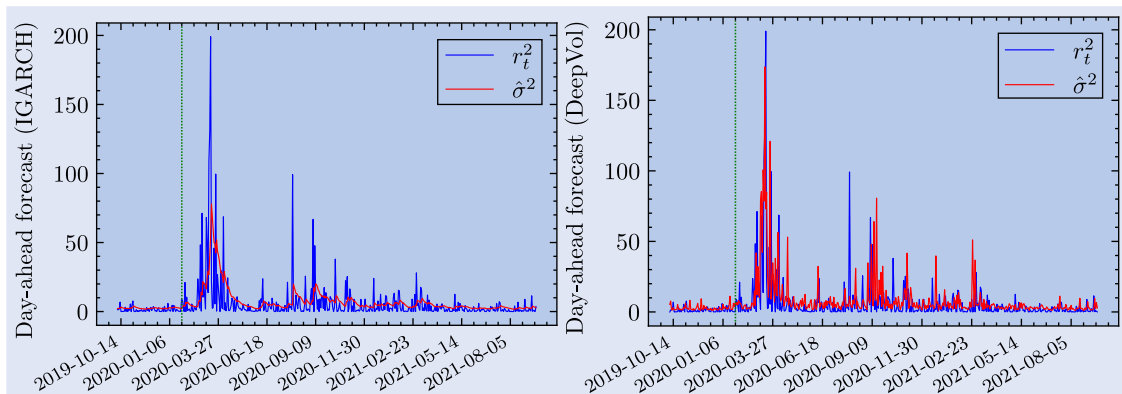
Figure 9. Out-of-sample-stocks: IGARCH's and DeepVol's forecast on AAPL. Green dotted vertical lines mark the forecast start.

6. Conclusions

In this paper, we propose a deep learning model based on hierarchies of Dilated Causal Convolutions —termed DeepVol— to forecast day-ahead realised volatility from high-frequency data. Our model takes advantage of the automatic feature extraction inherent to Deep Neural Networks to bypass the estimation of the realised measures, tackling the problem of volatility forecasting from a pure data-driven perspective. Moreover, the usage of dilated convolutions enables DeepVol to exponentially increase its input window without a significant increase in the number of parameters.

While DeepVol's operation does not directly produce any type of realised measure —the model does not explicitly

construct an ex-post estimate of the returns variation— its aggregation of high-frequency data resembles how realised measures condense intraday data into daily statistics. Reported results show how DeepVol's predictions significantly improve the baseline models performance, proving that the proposed data-driven approach avoids the limitations of classical methods, such as model misspecification or the usage of hand-crafted noisy realised measures, by taking advantage of the abundance of high-frequency data.

Several findings from the experiments are worth highlighting with regard to the usage of Dilated Causal Convolutions for the day-ahead realised volatility forecasting. Firstly, DeepVol generally outperforms traditional autoregressive architectures. Notably, DeepVol's performance remains robust even in high volatility regimes, as evidenced by the reported

Median Absolute Error (MedAE) and Mean Error (ME) metrics in section 5, which indicate quick adaptation to volatility shocks while maintaining some conservatism in their presence. These results align with existing literature, indicating that deep-learning-based volatility forecasting architectures (Ramos-Pérez *et al.* 2021, Zhang *et al.* 2022) and hybrid models (Baek and Kim 2018, Kim and Won 2018) consistently outperform classical methodologies. Further, the comparison between DeepVol-RM, which exclusively utilises the realised variance of previous days as input, and DeepVol itself, sheds more light on why DeepVol outperforms realised-measures-based models in the considered market regimes, suggesting that the model operating on intraday data is capable of leveraging information from the high-frequency returns that would be lost if realised measures were directly used. Lastly, our analysis of the receptive field size and sampling frequency in section 5.2 reveals that, for the proposed method and in the analysed market conditions, data from the previous trading day contains enough information for predicting the day-ahead realised volatility with high accuracy and therefore, increasing the receptive field of DeepVol does not generally lead to better performance. Further, we find that, in this market regime and for the proxy measure of the volatility employed, a sampling frequency of 5-minutes strikes a balance between capturing intraday information and minimising noise, echoing previous studies analysing this same trade-off in the context of estimation of realised measures from high-frequency data.

Finally, consideration should be given to tailoring all hyperparameters to specific stocks. For instance, adjusting parameters such as the receptive field size might facilitate more conservative predictions if desired in certain contexts, and tailoring the optimal sampling frequency of the intraday data could significantly enhance the model's accuracy. However, in this paper, our aim is to demonstrate the efficacy of a single overarching model capable of learning from a diverse set of stocks and generalising and transferring learned insights to new stocks. Therefore, practitioners should consider fine-tuning the model's parameters per stock depending on the context.

In summary, DeepVol offers improved forecasting accuracy compared to baseline methods, proving that the model quickly adapts to volatility shocks while demonstrating noise robustness. These results are especially relevant considering that experiments were conducted in high volatility regimes, such as the 2020 stock crisis caused by the COVID-19 pandemic, a context in which DeepVol shows its ability to extract universal features and transfer learning to out-of-distribution data during the out-of-sample-stocks forecasts.

The empirical results collected in this paper suggest that models based on Dilated Causal Convolutions should be carefully considered in the context of volatility forecasting and as a result can play a key role in the valuation of financial derivatives, risk management, and portfolio construction.

Acknowledgments

Funding

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Ait-Sahalia, Y., Mykland, P.A. and Zhang, L., How often to sample a continuous-time process in the presence of market microstructure noise. *Rev. Financ. Stud.*, 2005, **18**(2), 351–416.

Andersen, T.G. and Bollerslev, T., Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *J. Finance*, 1997, **52**(3), 975–1005.

Andersen, T.G. and Bollerslev, T., Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *Int. Econ. Rev. (Philadelphia)*, 1998, **39**(4), 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Ebens, H., The distribution of realized stock return volatility. *J. Financ. Econ.*, 2001, **61**(1), 43–76.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P., Modeling and forecasting realized volatility. *Econometrica*, 2003, **71**(2), 579–625.

Andersen, T.G., Bollerslev, T. and Meddahi, N., Market microstructure noise and realized volatility forecasting. Unpublished paper: Department of Economics, Duke University, 2006a.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F. and Diebold, F.X., Volatility and correlation forecasting. *Handb. Econ. Forecast.*, 2006b, **1**, 777–878.

Andersen, T.G., Bollerslev, T. and Diebold., F.X., Parametric and nonparametric volatility measurement. In *Handbook of Financial Econometrics: Tools and Techniques*, pp. 67–137, 2010 (Elsevier).

Arroyo, A., Scalzo, B., Stanković, L. and Mandic, D.P., Dynamic portfolio cuts: A spectral approach to graph-theoretic diversification. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5468–5472, 2022 (IEEE).

Arroyo, A., Cartea, A., Moreno-Pino, F. and Zohren, S., Deep attentive survival analysis in limit order books: Estimating fill probabilities with convolutional-transformers. *Quant. Finance*, 2024, **24**(1), 35–57.

Baars, B., *HEAVY and Realized (E) GARCH Models*, 2014 (GlobeEdit).

Baek, Y. and Kim, H.Y., ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert. Syst. Appl.*, 2018, **113**, 457–480.

Bahdanau, D., Cho, K. and Bengio, Y., Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.

Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O., Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econom.*, 1996, **74**(1), 3–30.

Bandi, F.M. and Russell, J.R., Separating microstructure noise from volatility. *J. Financ. Econ.*, 2006, **79**(3), 655–692.

Barndorff-Nielsen, O.E. and Shephard, N., *Non-Gaussian OU Based Models and Some of Their Uses in Financial Economics*, 1999 (Nuffield College Oxford).

Barndorff-Nielsen, O.E. and Shephard, N., Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B: Stat. Methodol.*, 2002, **64**(2), 253–280.

Bates, D.S., How crashes develop: Intradaily volatility and crash evolution. *J. Finance*, 2019, **74**(1), 193–238.

Biage, M., Analysis of shares frequency components on daily value-at-risk in emerging and developed markets. *Phys. A: Stat. Mech. Appl.*, 2019, **532**, 121798.

Bollerslev, T., Generalized autoregressive conditional heteroskedasticity. *J. Econom.*, 1986, **31**(3), 307–327.

Bouchaud, J.-P., Matacz, A. and Potters, M., Leverage effect in financial markets: The retarded volatility model. *Phys. Rev. Lett.*, 2001, **87**(22), 228701.

Brownlees, C.T. and Gallo, G.M., Comparison of volatility measures: A risk management perspective. *J. Financ. Econom.*, 2010, **8**(1), 29–56.

Calvet, L.E., Fisher, A.J. and Thompson, S.B., Volatility comovement: A multifrequency approach. *J. Econom.*, 2006, **131**(1-2), 179–215.

Chen, Q. and Robert, C.-Y., Multivariate realized volatility forecasting with graph neural network. arXiv preprint arXiv:2112.09015, 2021.

Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S. and Miao, Y., Review of image classification algorithms based on convolutional neural networks. *Remote. Sens. (Basel)*, 2021, **13**(22), 4712.

Christensen, K., Siggaard, M. and Veliyev, B., A machine learning approach to volatility forecasting. *J. Financ. Econom.*, 2023, **21**(5), 1680–1727.

Cochran, S.J., Mansur, I. and Odusami, B., Volatility persistence in metal returns: A FIGARCH approach. *J. Econ. Bus.*, 2012, **64**(4), 287–305.

Conneau, A., Schwenk, H., Barrault, L. and Lecun, Y., Very deep convolutional networks for natural language processing. 2016, **2**(1). arXiv preprint arXiv:1606.01781.

Cont, R., Volatility clustering in financial markets: Empirical facts and agent-based models. In *Long Memory in Economics*, edited by G. Teyssière and A. P. Kirman, pp. 289–309, 2007 (Springer: Berlin, Heidelberg).

Corradi, V. and Distaso, W., Semi-parametric comparison of stochastic volatility models using realized measures. *Rev. Econ. Stud.*, 2006, **73**(3), 635–667.

Corsi, F., A simple approximate long-memory model of realized volatility. *J. Financ. Econom.*, 2009, **7**(2), 174–196.

Diebold, F.X. and Mariano, R.S., Comparing predictive accuracy. *J. Bus. Econ. Stat.*, 2002, **20**(1), 134–144.

Ding, Z., Granger, C.W.J. and Engle, R.F., A long memory property of stock market returns and a new model. *J. Empir. Finance*, 1993, **1**(1), 83–106.

Engle, R.F., Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econom. J. Econom. Soc.*, 1982, **50**(4), 987–1007.

Engle, R.F. and Bollerslev, T., Modelling the persistence of conditional variances. *Econom. Rev.*, 1986, **5**(1), 1–50.

Engle, R.F. and Ng, V.K., Measuring and testing the impact of news on volatility. *J. Finance*, 1993, **48**(5), 1749–1778.

Engle, R.F. and Patton, A.J., What good is a volatility model? In *Forecasting Volatility in the Financial Markets*, pp. 47–63, 2007 (Butterworth-Heinemann).

Engle, R.F. and Sokalska, M.E., Forecasting intraday volatility in the us equity market. multiplicative component garch. *J. Financ. Econom.*, 2012, **10**(1), 54–83.

Falcon, W., The PyTorch Lightning Team, PyTorch Lightning. 3. https://github.com/Lightning-AI/lightning, 2019.

Ghosh, B.P., Bhuiyan, M.S., Das, D., Nguyen, T.N., Mahmud Jewel, R., Mia, M.T., Cao, D.M. and Shahid, R., Deep learning in stock market forecasting: Comparative analysis of neural network architectures across NSE and NYSE. *J. Comput. Sci. Tech. Stud.*, 2024, **6**(1), 68–75.

Goodfellow, I., Bengio, Y. and Courville, A., *Deep Learning*, 2016 (MIT Press).

Granger, C.W.J., Long memory relationships and the aggregation of dynamic models. *J. Econom.*, 1980, **14**(2), 227–238.

Granger, C.W.J. and Joyeux, R., An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, 1980, **1**(1), 15–29.

Hansen, P.R., A test for superior predictive ability. *J. Bus. Econ. Stat.*, 2005, **23**(4), 365–380.

Hansen, P.R. and Huang, Z., Exponential GARCH modeling with realized measures of volatility. *J. Bus. Econ. Stat.*, 2016, **34**(2), 269–287.

Hansen, P.R. and Lunde, A., Realized variance and market microstructure noise. *J. Bus. Econ. Stat.*, 2006, **24**(2), 127–161.

Hansen, P.R., Lunde, A. and Nason, J.M., The model confidence set. *Econometrica*, 2011, **79**(2), 453–497.

Hansen, P.R., Huang, Z. and Shek, H.H., Realized GARCH: A joint model for returns and realized measures of volatility. *J. Appl. Econom.*, 2012, **27**(6), 877–906.

Harvey, C.R., Hoyle, E., Korgaonkar, R., Rattray, S., Sargaison, M. and Van Hemert, O., The impact of volatility targeting. *J. Portfolio Manage.*, 2018, **45**(1), 14–33.

He, K., Zhang, X., Ren, S. and Sun, J., Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 770–778, 2016.

Hochreiter, S. and Schmidhuber, J., Long short-term memory. *Neural. Comput.*, 1997, **9**(8), 1735–1780.

Hochreiter, S., Bengio, Y., Frasconi, P. and Schmidhuber, J., Gradient flow in recurrent nets: The difficulty of learning long-term dependencies, 2001.

Horvath, B., Muguruza, A. and Tomas, M., Deep learning volatility. arXiv preprint arXiv:1901.09647, 2019.

Izzeldin, M., Hassan, M.K., Pappas, V. and Tsionas, M., Forecasting realised volatility using ARFIMA and HAR models. *Quant. Finance*, 2019, **19**(10), 1627–1638.

Jiménez Rama, Ó., Moreno-Pino, F., Ramírez, D. and Olmos, P.M., Interpretable Spectral Variational AutoEncoder (ISVAE) for time series clustering. arXiv e-prints arXiv–2310, 2023.

Karanasos, M., Yfanti, S. and Hunter, J., Emerging stock market volatility and economic fundamentals: The importance of US uncertainty spillovers, financial and health crises. *Ann. Oper. Res.*, 2022, **313**(2), 1077–1116.

Kaur, T. and Gandhi, T.K., Deep convolutional neural networks with transfer learning for automated brain image classification. *Mach. Vis. Appl.*, 2020, **31**(3), 20.

Kim, H.Y. and Won, C.H., Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert. Syst. Appl.*, 2018, **103**, 25–37.

Kingma, D.P. and Ba, J., Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

LeCun, Y., Bengio, Y. and Hinton, G., Deep Learning. *Nature*, 2015, **521**(7553), 436–444.

Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X. and Yan, X., Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Adv. Neural. Inf. Process. Syst.*, 2019, **32**.

Lim, B. and Zohren, S., Time-series forecasting with deep learning: A survey. *Philoso. Trans. R. Soc. A*, 2021, **379**(2194), 20200209.

Lin, Y., Lin, Z., Liao, Y., Li, Y., Xu, J. and Yan, Y., Forecasting the realized volatility of stock price index: A hybrid model integrating CEEMDAN and LSTM. *Expert. Syst. Appl.*, 2022, **206**, 117736.

Liu, D.C. and Nocedal, J., On the limited memory method for large scale optimization: Mathematical Programming B, 1989.

Mademlis, D.K. and Dritsakis, N., Volatility forecasting using hybrid GARCH neural network models: The case of the Italian stock market. *Int. J. Econ. Financ. Issues*, 2021, **11**(1), 49.

Masini, R.P., Medeiros, M.C. and Mendes, E.F., Machine learning advances for time series forecasting. *J. Econ. Surv.*, 2023, **37**(1), 76–111.

Mikolov, T., Chen, K., Corrado, G. and Dean, J., Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.

Moreira, A. and Muir, T., Volatility-managed portfolios. *J. Finance*, 2017, **72**(4), 1611–1644.

Moreno-Pino, F., Olmos, P.M. and Artés-Rodríguez, A., Deep autoregressive models with spectral attention. *Pattern. Recognit.*, 2023, **133**, 109014.

Moreno-Pino, F., Arroyo, A., Waldon, H., Dong, X. and Cartea, Á., Rough transformers for continuous and efficient time-series modelling. arXiv preprint arXiv:2403.10288, 2024a.

Moreno-Pino, F., Arroyo, Á., Waldon, H., Dong, X. and Cartea, Á., Rough transformers: Lightweight continuous-time sequence modelling with path signatures. arXiv preprint arXiv:2405.20799, 2024b.

Nelson, D.B., Conditional heteroskedasticity in asset returns: A new approach. *Econom. J. Econom. Soc.*, 1991, **59**(2), 347–370.

Noureldin, D., Shephard, N. and Sheppard, K., Multivariate high-frequency-based volatility (HEAVY) models. *J. Appl. Econom.*, 2012, **27**(6), 907–933.

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. and Kavukcuoglu, K., Wavenet: A generative model for raw audio. arXiv preprint arXiv:1609.03499, 2016.

Papantonis, I., Rompolis, L. and Tzavalis, E., Improving variance forecasts: The role of realized variance features. *Int. J. Forecast.*, 2023, **39**(3), 1221–1237.

Park, J.A., Baek, J.S. and Hwang, S.Y., Persistent-threshold-GARCH processes: Model and application. *Stat. Probab. Lett.*, 2009, **79**(7), 907–914.

Pascalau, R. and Poirier, R., Increasing the information content of realized volatility forecasts. *J. Financ. Econom.*, 2023, **21**(4), 1064–1098.

Patton, A.J., Volatility forecast comparison using imperfect volatility proxies. *J. Econom.*, 2011, **160**(1), 246–256.

Patton, A.J. and Sheppard, K., Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Rev. Econom. Stat.*, 2015, **97**(3), 683–697.

Patton, A.J. and Zhang, H., Bespoke realized volatility: Tailored measures of risk for volatility prediction. Available at SSRN 4315106, 2022.

Prechelt, L., Early stopping-but when?" In *Neural Networks: Tricks of the Trade*, edited by G. B. Orr and K. R. Müller, pp. 55–69, 1998 (Springer: Berlin, Heidelberg).

Rabemananjara, R. and Zakoian, J.-M., Threshold ARCH models and asymmetries in volatility. *J. Appl. Econom.*, 1993, **8**(1), 31–49.

Rahimikia, E. and Poon, S.-H., Big data approach to realised volatility forecasting using HAR model augmented with limit order book and news. Available at SSRN 3684040, 2020.

Rahimikia, E., Zohren, S. and Poon, S.-H., Realised volatility forecasting: Machine learning via financial word embedding. arXiv preprint arXiv:2108.00480, 2021.

Ramos-Pérez, E., Alonso-González, P.J. and Núñez-Velázquez, J.J., Multi-transformer: A new neural network-based architecture for forecasting S&P volatility. *Mathematics*, 2021, **9**(15), 1794.

Reisenhofer, R., Bayer, X. and Hautsch, N., HARNet: A convolutional neural network for realized volatility forecasting. arXiv preprint arXiv:2205.07719, 2022.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., Learning internal representations by error propagation. Technical Report. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., Learning representations by back-propagating errors. *Nature*, 1986, **323**(6088), 533–536.

Scalzo, B., Arroyo, A., Stanković, L. and Mandic, D.P., Non-stationary portfolios: Diversification in the spectral domain. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5155–5159, 2021 (IEEE).

Shamshirband, S., Fathi, M., Dehzangi, A., Chronopoulos, A.T. and Alinejad-Rokny, H., A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *J. Biomed. Inform.*, 2021, **113**, 103627.

Shen, Z., Wan, Q. and Leatham, D.J., Bitcoin return volatility forecasting: A comparative study between GARCH and RNN. *J. Risk Financ. Manag.*, 2021, **14**(7), 337.

Shephard, N. and Sheppard, K., Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. *J. Appl. Econom.*, 2010, **25**(2), 197–231.

Sheppard, K. and Xu, W., Factor high-frequency-based volatility (HEAVY) models. *J. Financ. Econom.*, 2019, **17**(1), 33–65.

Song, Y., Tang, X., Wang, H. and Ma, Z., Volatility forecasting for stock market incorporating macroeconomic variables based on GARCH-MIDAS and deep learning models. *J. Forecast.*, 2023, **42**(1), 51–59.

Souto, H.G. and Moradi, A., Introducing NBEATSx to realized volatility forecasting. *Expert. Syst. Appl.*, 2024, **242**, 122802.

Stroud, J.R. and Johannes, M.S., Bayesian modeling and forecasting of 24-hour high-frequency volatility. *J. Am. Stat. Assoc.*, 2014, **109**(508), 1368–1384.

Su, J.-B., How to promote the performance of parametric volatility forecasts in the stock market? a neural networks approach. *Entropy*, 2021, **23**(9), 1151.

Van Oord, A., Kalchbrenner, N. and Kavukcuoglu, K., 'Pixel recurrent neural networks. In *International Conference on Machine Learning*, pp. 1747–1756, 2016 (PMLR).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., Attention is all you need. *Adv. Neural. Inf. Process. Syst.*, 2017, **30**.

Vidal, A. and Kristjanpoller, W., Gold volatility prediction using a CNN-LSTM approach. *Expert. Syst. Appl.*, 2020, **157**, 113481.

Vis, W., Barendse, S. and Gresnigt, F., Volatility forecasting with realized measures: HEAVY vs. HAR, 2017.

Yu, S. and Li, Z., Forecasting stock price index volatility with LSTM deep neural network. In *Recent Developments in Data Science and Business Analytics*, pp. 265–272, 2018 (Springer).

Yuan, H., Li, G. and Wang, J., High-frequency-based volatility model with network structure. arXiv preprint arXiv:2204.12933, 2022.

Zakoian, J.-M., Threshold heteroskedastic models. *J. Econ. Dyn. Control*, 1994, **18**(5), 931–955.

Zhang, C., Zhang, Y., Cucuringu, M. and Qian, Z., Volatility forecasting with machine learning and intraday commonality. arXiv preprint arXiv:2202.08962, 2022.

Zhu, H., Bai, L., He, L. and Liu, Z., Forecasting realized volatility with machine learning: Panel data perspective. *J. Empir. Finance*, 2023, **73**, 251–271.

Appendix. Data description

Table A1. Summary statistics for the NASDAQ stocks' daily returns over the considered time span.

Symbol	Daily Returns						
	Mean	Std	Median	Minimum	Maximun	Skewness	Kurtosis
AAL	−0.16	5.79	−0.34	−29.07	34.56	0.73	8.33
AAPL	0.27	2.69	0.27	−14.11	11.44	−0.37	5.17
ADBE	0.19	2.76	0.38	−15.89	16.31	−0.15	7.98
ADI	0.09	2.9	0.13	−18.26	13.64	−0.59	8.31
ADP	0.04	2.57	0.16	−16.55	11.07	−0.75	9.45
ADSK	0.23	2.94	0.36	−12.79	15.05	−0.45	5.16
ALGN	0.34	4.04	0.21	−19.35	30.23	1.18	13.17
ALXN	0.15	2.82	−0.04	−8.79	25.52	2.37	20.63
AMAT	0.17	3.53	0.16	−22.74	12.06	−1.0	7.51
AMD	0.37	3.56	0.21	−15.82	15.24	−0.04	3.25
AMGN	0.06	2.2	−0.04	−8.71	10.74	0.57	4.82
AMZN	0.2	2.24	0.21	−8.21	7.59	−0.03	1.82
ASML	0.22	2.85	0.19	−19.06	11.43	−1.1	8.41
ATVI	0.17	2.28	0.21	−11.25	8.64	−0.18	3.3
AVGO	0.16	2.98	0.33	−22.2	13.08	−1.63	14.23
BIDU	0.23	2.78	0.06	−11.12	12.98	0.59	4.49
BIIB	0.13	3.6	−0.05	−10.14	36.53	∼	∼
BKNG	0.04	3.0	0.09	−11.95	17.4	0.34	6.42
BMRN	0.09	3.53	0.24	−43.58	11.28	−6.18	74.46
CDNS	0.23	2.69	0.44	−13.43	12.53	−0.28	4.58
CERN	0.06	1.83	0.14	−6.28	6.89	0.02	2.08
CHKP	0.06	1.99	0.1	−10.07	13.2	0.44	9.72
CHTR	0.14	2.07	0.15	−15.53	9.43	−0.91	12.57
CMCSA	0.06	2.22	0.11	−8.89	12.17	0.17	5.19
COST	0.09	1.68	0.08	−6.66	9.45	0.78	7.63
CSCO	−0.01	2.45	0.08	−11.84	12.52	−0.39	7.45
CSX	0.1	2.66	0.19	−16.87	15.23	−0.54	10.73
CTAS	0.1	3.0	0.24	−18.51	13.72	−0.96	9.43
CTSH	0.1	2.76	0.04	−18.9	14.01	−0.39	10.71
CTXS	0.1	2.36	0.25	−13.87	14.06	−0.04	8.88
DLTR	−0.02	2.91	0.09	−17.33	13.16	−0.97	12.22
EA	0.13	2.12	0.21	−9.32	10.06	−0.15	4.65
EBAY	0.09	2.26	0.14	−9.64	8.36	−0.39	2.13
EXPE	0.0	4.44	0.13	−31.79	21.88	−1.32	13.8
FAST	0.15	2.43	0.09	−11.75	15.83	0.82	9.48
FB	0.13	2.68	0.22	−15.61	9.95	−0.49	5.18
FISV	0.02	2.72	0.13	−18.27	11.81	−0.75	8.71
GILD	−0.01	2.18	−0.11	−9.03	9.27	0.39	3.33
HAS	−0.06	3.3	0.14	−20.77	19.61	−1.03	12.91
HSIC	0.02	2.57	0.13	−10.14	9.77	−0.21	2.77
IDXX	0.19	2.44	0.4	−12.74	11.46	−0.4	4.51
ILMN	0.06	2.74	0.3	−11.69	10.19	−0.84	3.39
INCY	0.05	2.49	−0.13	−9.86	8.37	0.14	1.53
INTC	0.0	3.13	0.12	−20.04	17.87	−0.92	12.9
INTU	0.12	2.73	0.22	−15.69	18.3	0.23	10.0
ISRG	0.14	2.65	0.23	−15.32	15.85	−0.25	8.53
JBHT	0.08	2.39	0.13	−15.44	12.33	−0.63	9.26
JD	0.35	2.9	0.41	−13.5	11.33	−0.17	2.55
KHC	0.1	2.58	0.19	−14.51	17.81	0.72	12.62
KLAC	0.16	3.49	0.13	−14.49	15.89	0.15	5.13
LRCX	0.23	3.75	0.24	−20.36	18.39	−0.42	6.77
LULU	0.19	3.03	0.4	−24.4	11.46	−1.88	15.17
MAR	0.03	3.78	0.15	−16.48	17.81	0.22	4.51
MCHP	0.13	3.59	0.09	−22.67	14.51	−0.67	8.33
MDLZ	0.02	1.96	0.04	−12.14	10.78	−0.02	11.03
MELI	0.35	3.44	0.53	−13.68	17.9	−0.3	4.01
MNST	0.16	2.06	0.24	−8.28	8.39	0.09	4.4
MSFT	0.16	2.53	0.22	−15.96	13.52	−0.44	8.76
MU	0.17	3.44	0.16	−22.13	12.3	−0.8	6.56
MYL	−0.06	2.82	0.0	−10.34	12.87	−0.12	2.19
NFLX	0.22	2.78	0.13	−12.48	10.95	−0.3	2.88

Table A1. Continued.

Symbol	Daily Returns						
	Mean	Std	Median	Minimum	Maximun	Skewness	Kurtosis
NTAP	0.09	3.04	0.2	− 12.02	12.88	− 0.18	3.09
NTES	0.2	2.66	0.12	− 12.11	8.68	− 0.3	2.09
NVDA	0.34	3.39	0.33	− 20.3	16.22	− 0.59	6.18
NXPI	0.12	3.48	0.34	− 20.47	15.44	− 0.61	7.62
ORLY	0.05	2.49	0.01	− 16.76	12.15	− 0.95	12.74
PAYX	0.04	2.76	0.15	− 21.65	16.48	− 1.12	17.44
PCAR	0.09	2.09	0.17	− 7.95	13.34	0.66	7.34
PEP	0.03	2.11	0.09	− 12.13	12.18	− 0.64	14.88
PYPL	0.26	3.07	0.21	− 17.99	13.19	− 0.1	6.52
QCOM	0.23	3.07	0.02	− 16.18	14.14	0.0	5.29
REGN	0.17	2.54	0.11	− 10.91	11.27	0.23	2.38
ROST	0.04	3.3	0.17	− 22.15	15.67	− 0.43	10.52
SBUX	0.08	2.63	0.04	− 18.17	13.86	− 0.5	11.51
SIRI	0.0	2.54	0.15	− 16.35	9.69	− 1.21	8.74
SNPS	0.2	2.5	0.45	− 14.15	10.91	− 0.52	5.59
SWKS	0.21	3.28	0.3	− 20.07	15.04	− 0.38	6.72
TMUS	0.17	2.32	0.28	− 11.85	11.06	− 0.13	7.63
TSLA	0.87	5.22	0.82	− 23.62	18.19	− 0.49	3.72
TTWO	0.17	2.47	0.25	− 12.51	9.32	− 0.78	4.46
TXN	0.09	2.56	0.17	− 12.78	12.7	− 0.09	5.05
UAL	− 0.21	5.94	− 0.27	− 36.31	22.85	− 0.61	7.27
ULTA	0.04	3.81	0.09	− 24.86	21.75	− 0.1	12.39
VRSK	0.09	2.22	0.11	− 11.24	12.67	− 0.11	8.79
VRSN	0.05	2.3	0.19	− 13.05	12.0	0.09	8.0
VRTX	0.1	2.75	0.1	− 23.29	10.15	− 1.9	17.24
WBA	− 0.07	2.82	− 0.14	− 11.63	11.89	− 0.08	3.16
WDAY	0.1	3.12	0.36	− 13.66	11.86	− 0.26	3.26
WDC	− 0.01	4.21	0.0	− 22.91	16.34	− 1.11	5.89
WLTW	0.04	2.32	0.05	− 9.66	14.01	0.15	7.75
WYNN	0.02	4.99	− 0.21	− 28.0	24.56	− 0.38	6.47
XEL	0.02	2.25	0.13	− 13.57	10.37	− 0.83	10.46
XLNX	0.14	2.86	− 0.11	− 11.32	13.19	0.4	3.15

Table A2. Summary statistics for the NASDAQ stocks' daily realised variance over the considered time span.

Symbol	Daily Realised Variance						
	Mean	Std	Median	Minimum	Maximun	Skewness	Kurtosis
AAL	22.6	42.86	10.31	1.13	411.94	4.97	31.55
AAPL	4.04	6.83	1.83	0.26	49.23	4.06	18.58
ADBE	4.4	6.47	2.48	0.26	47.92	4.08	19.18
ADI	5.13	9.85	2.44	0.33	87.12	5.17	30.21
ADP	4.8	10.7	1.89	0.31	102.61	5.35	33.47
ADSK	6.5	9.43	3.84	0.37	85.45	4.52	25.56
ALGN	11.24	16.42	6.36	1.72	132.13	4.41	23.6
ALXN	6.18	7.92	3.81	0.32	63.54	4.16	20.04
AMAT	5.5	8.94	2.88	0.64	66.71	4.61	23.61
AMD	7.61	8.72	4.83	1.08	64.31	3.55	14.47
AMGN	4.39	8.34	2.13	0.38	62.02	4.79	24.59
AMZN	3.42	4.72	2.0	0.15	41.64	4.01	21.0
ASML	3.53	5.48	1.79	0.47	36.73	4.06	18.13
ATVI	4.83	7.66	2.53	0.57	67.18	4.76	26.15
AVGO	5.22	10.89	2.22	0.48	96.23	5.02	28.45
BIDU	5.19	5.17	3.47	0.67	36.34	3.01	10.53
BIIB	7.01	19.78	2.99	0.03	302.31	11.42	159.51
BKNG	9.13	11.49	5.76	0.68	88.17	3.66	16.04
BMRN	7.64	10.22	4.68	1.33	74.02	4.24	20.59
CDNS	6.61	24.02	2.9	0.8	403.4	14.65	236.91
CERN	4.54	8.79	1.95	0.39	69.47	4.56	23.22
CHKP	5.45	13.48	2.34	0.59	176.49	8.44	90.08
CHTR	5.2	9.85	2.34	0.43	85.43	5.04	30.63
CMCSA	4.35	10.02	1.76	0.44	81.18	5.2	28.69
COST	2.55	5.55	1.03	0.18	44.94	4.98	27.4
CSCO	3.63	8.27	1.45	0.35	71.13	5.23	29.96
CSX	4.68	9.75	2.03	0.36	85.58	4.88	27.26
CTAS	8.09	16.02	4.01	0.76	195.18	6.77	63.57
CTSH	10.71	106.99	2.15	0.34	1892.0	17.45	304.39
CTXS	5.23	9.42	2.8	0.25	74.6	4.7	25.05
DLTR	5.79	12.69	2.75	0.66	119.42	6.05	42.69
EA	4.7	7.22	2.66	0.4	61.63	4.5	23.88
EBAY	4.17	5.87	2.56	0.22	48.64	4.02	19.12
EXPE	11.83	22.11	5.94	0.63	194.6	5.33	34.43
FAST	4.98	9.67	2.1	0.57	76.47	4.46	22.05
FB	4.33	6.2	2.69	0.41	52.23	4.59	24.9
FISV	5.47	13.59	2.41	0.31	191.08	9.38	113.63
GILD	4.53	10.12	1.71	0.48	86.76	5.73	37.25
HAS	7.9	12.87	4.26	0.69	92.6	4.36	20.86
HSIC	7.83	12.86	4.32	0.5	112.02	4.53	24.85
IDXX	9.79	16.12	4.73	0.91	128.64	4.4	23.07
ILMN	6.65	10.38	3.62	0.8	77.88	4.21	19.46
INCY	9.71	56.18	4.2	1.22	986.19	16.88	290.59
INTC	4.7	11.21	1.84	0.35	99.29	5.63	35.5
INTU	5.57	9.48	2.74	0.56	81.0	4.62	24.89
ISRG	6.33	10.29	3.4	0.8	98.68	4.98	30.94
JBHT	7.5	14.74	3.6	0.71	135.71	5.58	38.12
JD	6.3	7.6	4.08	0.67	85.1	5.18	40.26
KHC	4.4	8.79	2.05	0.47	72.58	5.37	31.64
KLAC	6.61	9.28	3.85	0.87	75.85	4.4	23.24
LRCX	6.99	9.56	3.98	0.72	70.87	3.91	17.89
LULU	6.68	11.81	3.94	0.53	139.13	6.53	57.39
MAR	13.45	51.38	4.68	0.3	715.86	10.62	128.32
MCHP	6.6	9.93	3.65	0.45	90.88	4.54	25.74
MDLZ	3.36	8.73	1.17	0.32	81.08	5.64	35.21
MELI	13.81	14.73	9.05	1.16	106.06	3.52	15.27
MNST	4.48	9.43	1.93	0.52	99.82	5.95	43.66
MSFT	3.76	8.12	1.43	0.09	71.93	5.14	30.36
MU	6.53	11.06	3.52	0.84	97.46	5.22	31.78
MYL	10.03	15.66	5.42	1.21	114.53	4.22	19.84
NFLX	5.91	7.47	3.41	0.8	62.96	4.29	22.29
NTAP	7.24	17.28	3.46	0.76	250.72	9.97	127.58
NTES	6.74	7.0	4.52	1.21	57.85	3.72	17.52
NVDA	6.69	10.66	3.55	0.71	105.72	4.88	31.05

(*Continued*).

Table A2. Continued.

Symbol	Daily Realised Variance						
	Mean	Std	Median	Minimum	Maximun	Skewness	Kurtosis
NXPI	7.54	13.29	3.83	0.75	138.6	5.85	42.58
ORLY	7.36	19.65	2.91	0.62	293.07	10.54	142.63
PAYX	5.64	15.43	1.72	0.29	148.48	5.91	40.68
PCAR	4.69	7.75	2.19	0.59	62.82	4.16	20.28
PEP	3.42	10.31	1.12	0.16	120.15	7.23	63.74
PYPL	5.6	8.49	3.19	0.27	58.1	4.24	20.26
QCOM	5.01	7.66	2.82	0.35	60.25	4.52	22.97
REGN	7.26	9.1	4.53	1.08	72.22	4.32	22.3
ROST	6.99	13.47	3.48	0.55	109.05	5.13	30.07
SBUX	4.38	10.26	1.51	0.17	83.66	5.06	28.21
SIRI	6.31	9.3	3.82	1.17	93.44	5.13	33.54
SNPS	6.13	8.58	3.71	0.8	59.56	4.14	19.0
SWKS	6.35	10.77	3.44	0.92	110.83	5.41	36.91
TMUS	4.23	7.63	1.98	0.19	66.62	4.86	28.33
TSLA	16.15	22.21	8.25	0.98	152.42	3.18	11.75
TTWO	5.99	8.52	3.63	0.84	77.57	4.82	28.95
TXN	3.96	7.48	1.86	0.24	72.12	5.29	33.65
UAL	22.93	57.24	8.2	0.57	575.89	6.52	49.71
ULTA	11.11	21.24	5.38	1.21	156.61	4.35	19.75
VRSK	10.73	73.47	2.28	0.46	1269.28	16.21	273.97
VRSN	6.4	12.2	2.8	0.72	118.17	4.96	30.98
VRTX	9.19	47.49	3.43	0.77	823.47	16.25	275.06
WBA	5.73	12.23	3.01	0.65	155.82	7.74	78.06
WDAY	7.21	10.28	4.71	0.94	118.09	6.03	49.52
WDC	9.03	15.67	4.52	1.46	132.72	4.78	25.64
WLTW	7.34	11.88	3.62	0.64	96.86	4.1	20.33
WYNN	16.28	35.4	7.5	1.03	360.04	6.04	43.25
XEL	4.63	12.95	1.62	0.35	170.39	8.36	90.58
XLNX	4.94	6.77	2.97	0.74	57.79	4.69	25.99