Iranian Journal of Finance

Print ISSN 2676-6337
Online ISSN 2676-6345

Forecasting Value at Risk (VaR) and Expected Shortfall (ES) in the Tehran Stock Exchange Index Using Stateful Recurrent Neural Network (Stateful-RNN) Models

Ali Namaki*

*Corresponding author, Associate Prof., Department of Markets and Financial Institutions, Faculty of Accounting and Financial Sciences, College of Management, University of Tehran, Tehran, Iran. (Email: alinamaki@ut.ac.ir)

Ali Emarloo

MSc. Student. Department of Markets and Financial Institutions, Faculty of Accounting and Financial Sciences, College of Management, University of Tehran, Tehran, Iran. (Email: emarlooali.ar10@ut.ac.ir)

Iranian Journal of Finance, 2026, Vol. 10, No.2, pp. 85-127.
Publisher: Iran Finance Association
doi: https://doi.org/10.66224/ijf.2026.580106.1575
Article Type: Original Article

Received: January 14, 2026
Received in revised form: February 09, 2026
Accepted: March 05, 2026
Published online: April 01, 2026

Abstract

The increasing importance of financial risk measurement, driven by regulatory requirements and market uncertainties, necessitates precise forecasting tools for Value at Risk (VaR) and Expected Shortfall (ES). This study introduces three novel models based on Stateful Recurrent Neural Networks (RNNs) and Feedforward Neural Networks (FNNs) for predicting VaR and ES. These models are evaluated against traditional econometric models, including

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, using a dataset of the Tehran Exchange Divedend and Price Index (TEDPIX) spanning more than three decades. This research presents a comprehensive comparison between Stateful and Non-Stateful RNN models, demonstrating that Stateful RNNs significantly enhance forecasting performance. First, the data is standardized before being fed into machine learning models and then used as input for the models. The results indicate strong predictive performance, while a reduction in the FZ0 loss score highlights the superior accuracy of these models. In the empirical study, the proposed models are applied to the Tehran Exchange Divedend and Price Index (TEDPIX). The models' performance is assessed using the FZ0 loss function and backtesting methods, such as DQ regression tests and DES tests. The findings reveal that Stateful RNN-based models outperform traditional econometric models in terms of both accuracy and stability. In particular, the STRNN-II and STRNN-III models consistently rank highest, demonstrating their superior ability to capture the complex dynamics of financial risk. Moreover, these models do not rely on strong distributional assumptions, offering greater flexibility than traditional parametric models like GARCH.

Keywords: Risk models, Value-at-Risk, Expected shortfall, Machine learning, Neural networks

JEL Classification: G21, C53, C45, C38

Introduction

In recent decades, with the rapid growth of financial markets and the increasing complexity of investment instruments, risk management has become a fundamental topic in the field of finance and economics. Financial crises, such as the 2008 financial crisis, have revealed the crucial role of systematic and unsystematic risks in financial market instability. These crises demonstrated that traditional risk assessment models are inadequate for forecasting and managing market volatility, prompting extensive efforts to improve risk evaluation methodologies. Measuring risk has gained prominence in finance due to the widespread damage that shocks from market crashes can inflict on the economy. Value at Risk (VaR), since its introduction in the RiskMetrics model by J.P. Morgan in 1989, has been widely used by financial institutions to measure market risk (McNeil et al., 2015).

Value at Risk (VaR) refers to the worst expected return of an asset over a predetermined time horizon at a given significance level (α). However, VaR has been criticized for disregarding the magnitude of losses beyond the VaR threshold and its inability to optimally capture risk, especially when dealing with asymmetric and skewed return distributions. As an alternative to VaR, Expected Shortfall (ES), proposed by Artzner (1997) and Artzner et al. (1999), is considered a coherent and desirable risk measure. ES at level α represents the expected return below the VaR threshold at the same significance level. Following the 2007–2008 financial crisis, the Basel III Accord (Basel Committee on Banking Supervision, 2010) recommended replacing VaR with ES as the primary risk metric. These measures have been widely adopted in the financial industry, particularly as tools for risk management, as they help estimate an asset's potential losses at a given risk level and contribute to more efficient risk allocation.

Classical risk forecasting methods, such as GARCH models and parametric distribution-based models, have been widely utilized in both academic literature and practical applications (So & Philip, 2006; Degiannakis et al., 2013). These models generally rely on specific statistical assumptions about the distribution of returns, which may not align with the real, nonlinear nature of financial data. Furthermore, GARCH models and other traditional approaches are limited in their ability to capture complex dependencies and nonlinear patterns in financial time series data, thereby lacking sufficient accuracy in predicting financial market behavior (Hartz et al., 2006).

In recent years, machine learning-based models have gained significant popularity in financial applications. One of the main driving factors is their independence from simplifying assumptions, which are often inherent in traditional models. Among machine learning techniques, neural networks have become especially popular due to their capacity to learn and predict complex patterns within data. By leveraging generative neural networks to capture the statistical features of input data, these models can produce more accurate outputs. As such, generative neural networks can represent the inherent dependency structure in asset returns (Arian et al., 2022).

Recurrent Neural Networks (RNNs), due to their ability to process time series data and learn long-term dependencies, have been widely used in financial forecasting tasks. Particularly, Stateful-RNN models are well-suited for forecasting Value at Risk (VaR) and Expected Shortfall (ES) in financial markets. The main reasons for employing these models include: RNNs are specifically designed to model time series data by maintaining a memory of past information through their internal hidden layers. In the financial context—such as risk forecasting using VaR and ES measures—this ability to understand temporal dependencies is critically important. However, the effectiveness of this memory mechanism heavily depends on whether the RNN is implemented in a stateful or stateless manner (Qiu, Lazar, & Nakata, 2024).

Traditional models like GARCH require specific distributional assumptions about asset returns to forecast VaR and ES. These assumptions may not hold in the complex realities of financial markets. In contrast, stateful RNN models, by adopting a nonparametric approach, do not require such assumptions and can more accurately model complex data patterns (Qiu, Lazar, & Nakata, 2024).

In a stateless RNN, the network's hidden state is reset after processing each batch of input data. This means the model does not retain any information between batches during training. This design simplifies the training process and avoids unintended long-term dependencies that may cause instability; however, it limits the model's ability to understand long-term temporal structures in the data (Qiu, Lazar, & Nakata, 2024). In the context of financial risk modeling, this memory reset prevents the model from adequately capturing significant events or regimes (such as market crashes) that have lasting effects on volatility and risk. In the referenced study, three stateless RNN models are implemented, named RNN-I, RNN-II, and RNN-III, which structurally mirror their stateful counterparts but with the stateful feature turned off (i.e., stateful=0). These models begin training from scratch with each new batch and retain no prior memory (Qiu, Lazar, & Nakata, 2024).

Conversely, a stateful RNN preserves its hidden state across different batches during both training and prediction phases (stateful=1). This persistence allows the network to remember long-term dependencies in the data and establishes a temporal continuity in learning. As a result, the model can correctly interpret and forecast gradual changes in the risk structure that evolve over time (Qiu, Lazar, & Nakata, 2024).

In this thesis, three recurrent neural network models are introduced and analyzed in both stateful and stateless configurations. These models are as follows:

STRNN-I: A combination of a stateful RNN and a feedforward neural network (FNN) for the joint prediction of VaR and ES from squared returns.

STRNN-II: To enhance its ability to capture nonlinear relationships, this model applies a transformation (square root of absolute values) to the RNN's hidden state before feeding it into the FNN.

STRNN-III: A hybrid model that combines the outputs of the previous two models via element-wise summation to boost signal strength. These models are specifically designed to retain memory over time, thereby improving their ability to model time-dependent risk structures, such as volatility clustering or regime shifts.

Considering the importance of risk forecasting in financial markets, this research aims to forecast Value at Risk (VaR) and Expected Shortfall (ES) for indices in the Tehran Stock Exchange using stateful Recurrent Neural Network (Stateful-RNN) models. Subsequently, the performance of these models will be evaluated and compared with that of classical risk forecasting models, such as GARCH. Finally, a comparison between the performance of stateful and stateless neural network models will be conducted.

Literature Review

Value-at-Risk (VaR) has become a cornerstone risk measure in finance since its popularization by J.P. Morgan's RiskMetrics in the late 1980s. VaR represents the worst expected loss over a given horizon at a specified confidence level (e.g. 95% or 99%). Despite its widespread adoption by financial institutions and regulators in the 1990s, VaR faced criticism for failing to account for the severity of losses beyond the threshold (it is a *quantile* of the loss distribution) and for lacking certain theoretical properties. Notably, VaR is not a *coherent* risk measure in the sense of Artzner et al. (1999), failing to satisfy subadditivity in some cases. (Qiu, Lazar, & Nakata, 2024)

ES is a coherent risk measure and intuitively provides a fuller picture of tail risk by answering the question: "If the loss exceeds the VaR level, how much might we lose on average?" After the financial crisis of 2007–2008, the shortcomings of VaR became evident as many institutions underestimated extreme losses. This led regulators (the Basel Committee under the Basel III framework) to advocate replacing VaR with ES for bank capital calculations. Since then, accurate forecasting of both VaR and ES has been a critical focus for risk management. (Qiu, Lazar, & Nakata, 2024)

Parametric variance-covariance methods were among the first systematic approaches to VaR forecasting. J.P. Morgan's RiskMetrics (1994/1996) introduced a simple but influential model that assumes asset returns are normally distributed, with volatility updated using an exponentially weighted moving average (EWMA). Under RiskMetrics, VaR is calculated as a multiple of the current volatility. The appeal of this method was its simplicity and ease of implementation across many assets. However, its drawbacks became quickly apparent. The normal distribution assumption severely underestimates tail risk, given that financial returns typically exhibit excess kurtosis (fat tails). In practice, far more extreme moves occur than the normal VaR would predict (a phenomenon sometimes dubbed "the failure of the normality assumption"). Moreover, the RiskMetrics volatility model (EWMA) captures volatility clustering to some extent but cannot account for sudden jumps or leverage effects (the asymmetric response of volatility to negative returns). These limitations spurred the development of more flexible time-series models for volatility and distribution. (Abad, Benito, & López, 2014)

Quantile regression is a well-known approach for estimating risk measuresEngle and Manganelli (2004) already demonstrated CAViaR's efficacy on test datasets, and subsequent studies provided a deeper assessment. For example, Bao et al. (2006) compared various CAViaR specifications and traditional VaR models on out-of-sample forecasts; they found that CAViaR performed well in relatively stable periods but that its performance could deteriorate during highly volatile periods – presumably because the linear quantile specification may not fully capture sudden regime shifts or nonlinearities during crises. Polanski and Stoja (2009) also compared multiple CAViaR forms. They did not find a single CAViaR specification that universally dominated the others, highlighting that model risk (choice of specification) remains a concern. On the other hand, refinements of the CAViaR approach showed promise. Huang et al. (2009) applied CAViaR to oil markets. They stressed the importance of incorporating oil-specific dynamics, and Chen et al. (2011) compared the original Engle-Manganelli models with others, finding that a Threshold CAViaR (which introduces a structural break or threshold in the VaR updating equation) tended to perform best for extreme tails (e.g., for 1% VaR). These findings suggest that allowing nonlinearity or regime-switching in quantile models can improve forecasts during turbulent times.

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family of models (Engle, 1982; Bollerslev, 1986) provided a more rigorous framework for modeling time-varying volatility. In a typical GARCH(1,1) model, today's variance is a function of yesterday's variance and yesterday's squared return, allowing volatility to persist over time. By the late 1990s, it had become common to use GARCH models to forecast VaR: one would forecast the next-day volatility using GARCH and then compute VaR by assuming a distribution for standardized returns. If one assumes normal

GARCH innovations, the VaR is like an updated RiskMetrics (with a more sophisticated volatility estimate); if one assumes Student-t innovations, heavier tails are captured. A vast literature examined GARCH-based VaR forecasting. For example, So and Philip (2006) and Hartz et al. (2006) showed that incorporating heavier-tailed distributions into GARCH (e.g., using a Student-t error term) improved VaR forecasts, capturing more extreme losses.

Following the 2007–2008 financial crisis, Expected Shortfall (ES) has become increasingly prominent in the financial sector. A substantial body of research has focused on the simultaneous estimation of Value at Risk (VaR) and ES. For instance, Taylor (2008) introduced conditional autoregressive expectile (CARE) models to evaluate VaR and ES jointly. Nevertheless, Xu et al. (2016) point out that parametric CARE models face difficulties in determining a specific parametric structure. Using the asymmetric Laplace distribution, Taylor (2019) developed a semiparametric method with a novel scoring function to model risk measures jointly. Subsequently, Gerlach and Wang (2020) advanced the CAViaR frameworks by integrating a realized measure into the dynamic models. Despite these advancements, CAViaR models operate under the assumption that VaR and ES share identical dynamics.

Taylor (2019) proposed a semiparametric joint VaR–ES model based on the Asymmetric Laplace (AL) distribution. This approach exploits the fact that an AL distribution can be used to generate quantile and expectile estimates by appropriate likelihood (or scoring) functions. Taylor (2019) also leveraged the Fissler-Ziegel (2016) scoring function, which provides a theoretically consistent way to jointly estimate VaR and ES by optimizing a single objective. While this semiparametric method showed strong performance in empirical tests, a potential inefficiency noted later was its assumption of an additive impact on ES. In reality, volatility changes might scale ES multiplicatively. To address this, Taylor (2022) proposed a framework featuring a dynamic multiplicative Omega ratio factor for the simultaneous prediction of VaR and ES. Additionally, Gerlach and Wang (2020) expanded upon CAViaR frameworks by embedding a realized measure into the dynamic equations. Nevertheless, an inherent limitation of CAViaR models is their assumption of identical dynamics for both VaR and ES.

Another influential contribution is by Patton, Ziegel, and Chen (2019), who proposed dynamic semiparametric models for joint VaR and ES within the framework of Generalized Autoregressive Score (GAS) models. GAS (Creal et al., 2013; Harvey, 2013) is a class of models that update parameters over time using the score of the predictive likelihood (or some scoring rule). Patton et al. adapted this idea to the VaR–ES context by using a joint scoring function for VaR and ES, enabling time-varying updates that respond to tail losses. Their models do not require specifying a full distribution of returns; instead, they specify how VaR and ES evolve in response to feedback from tail events. The authors found these models performed well in forecasting both VaR and ES across a range of assets. However, they also noted practical challenges: the estimation could be sensitive to starting values and the choice of optimization method, occasionally leading to convergence or robustness issues. This underscores that adding complexity (joint modeling, nonlinearity) can improve fit but at the cost of a more difficult estimation surface.

Parameter estimation for these models often relies on specific loss functions. A well-known example is the quantile loss function introduced by Koenker and Bassett, Jr. (1978), which is commonly utilized for quantile regression and VaR estimation within the CAViaR framework. While the original CAViaR model focuses exclusively on VaR, Taylor (2008) developed the conditional autoregressive expectile (CARE) model. This approach employs asymmetric least squares (ALS) regression and the expectile loss function to produce joint forecasts for both VaR and ES. Building on the FZ loss function established by Fissler and Ziegel (2016), Taylor (2019) introduced the asymmetric Laplace (AL) log-score function to estimate both risk metrics simultaneously.

In the last decade, there has been an explosion of interest in applying machine learning (ML) techniques to financial risk forecasting. Machine learning methods, encompassing a broad class of algorithms from support vector machines to neural networks and decision trees, offer two main potential advantages: they can capture complex nonlinear relationships that traditional models might miss, and they often make fewer a priori assumptions about data distribution. As computing power and data availability have grown, ML approaches have become feasible for large-scale risk management applications. One driving factor is the lack of reliance on strict assumptions – unlike parametric models that assume (for example) a specific distribution or functional form, ML models can theoretically learn the patterns from data.

Within the domain of time series prediction, machine learning models have demonstrated considerable efficacy. A substantial segment of the academic literature concentrates on deploying these techniques to forecast financial returns, consistently documenting measurable improvements in predictive accuracy. To illustrate, Patel et al. (2015) utilized machine learning frameworks to forecast the directional movements of stocks and stock index. In a related comparative study of linear and nonlinear models for forecasting asset returns, Gu et al. (2020) established that these methods enhance asset pricing performance. Their analysis determined that neural networks delivered the superior outcome among all evaluated methodologies. Taking a different approach, Saha et al. (2021) integrated event-specific features into deep neural networks to predict stock price movements. However, machine learning is actively applied in other complex domains, particularly derivatives pricing and volatility modeling (see Ye and Zhang, 2019; Zhang, 2020; Ivașcu, 2021; Vrontos et al., 2021; Lu et al., 2022).

The forecasting of VaR and ES has increasingly incorporated machine learning techniques. To improve VaR prediction, Khan (2011) integrated the heterogeneous autoregressive model into a Support Vector Machine framework. Subsequently, Shim et al. (2012) introduced a semiparametric Support Vector Quantile regression approach for VaR estimation. In related developments, one-day ahead VaR estimates are constructed by coupling GARCH-type models with long short-term memory (LSTM) and bidirectional LSTM (BiLSTM) neural networks, utilizing the Parametric and Filtered Historical Simulation (FHS) method. Wu and Yan (2019) formulated a conditional quantile model that leverages LSTM neural networks for predicting

VaR.

Kakade and colleagues (2022) introduce a novel hybrid framework that integrates LSTM and BiLSTM neural networks with GARCH-type model forecasts via an ensemble technique. This approach is designed to predict one-day-ahead Value-at-Risk (VaR) estimates at 95% and 99% confidence levels, employing the Parametric (PAR) and Filtered Historical Simulation (FHS) methods. The predictive capabilities of conventional GARCH, exponential GARCH (eGARCH), and threshold GARCH (tGARCH) specifications are merged with LSTM networks to encapsulate diverse attributes of the latent volatility process. Their analysis reveals a marked enhancement in both the quality and precision of VaR forecasts generated by the hybrid models relative to all benchmark alternatives, as evaluated across a range of loss functions and coverage assessments. The proposed FHS-BiLSTM-HYBRID model—an FHS-based hybrid that fuses the BiLSTM network with three distinct GARCH-type models—emerges as the top performer, registering the smallest values for each loss function considered. Conventional and GARCH-type models exhibit limitations in capturing volatility dynamics during crises, leading to suboptimal VaR predictions. In contrast, the FHS method consistently proves superior to all other methodologies examined for VaR estimation.

Ormaniec et al. (2022) introduce a novel Value-at-Risk (VaR) estimator that uses an LSTM neural network, concluding that the proposed machine learning model outperforms GARCH models when applied to real market data. Cont et al. (2022) present a new data-driven methodology that employs a Generative Adversarial Network (GAN) architecture to estimate tail risk. Chronopoulos et al. (2023) apply a deep quantile estimator based on neural networks to forecast VaR. Nevertheless, a significant research gap persists in the joint forecasting of VaR and Expected Shortfall (ES) through neural network approaches.

Research Methodology

In the subsequent sections, we introduce two widely-used risk metrics: Value at Risk (VaR) and Expected Shortfall (ES). We also describe the FZ0 scoring function and the architectures of Feedforward Neural Networks (FNNs) and Recurrent Neural Networks (RNNs). Additionally, this article proposes three novel implementations of stateful machine learning models for forecasting VaR and ES, and presents the benchmark models used for comparison.

Risk measures

VaR arguably is the most commonly applied risk measure within financial institutions. Take a portfolio composed of risky assets with a fixed holding horizon Δt, and denote by $F_L(l) = p(L \leq l)$ the df of the corresponding loss distribution. The aim is to construct a statistic derived from F_L that captures the riskiness of holding the portfolio over the interval Δt. The concept behind VaR is to swap the “maximum loss” for “the greatest loss that is not surpassed with a specified high probability.” For a chosen confidence level $\alpha \in (0,1)$, the VaR of a portfolio with loss L at level α is defined as the smallest number l for which the probability that L exceeds l is no more than 1 − α. Formally,

$$\begin{aligned} VaR_\alpha = VaR_\alpha\,(L) & \\ &= \inf\{l \,\epsilon\, \mathbb{R}: p(L > l) \leq 1 - \alpha \\ &= \inf\{l \,\epsilon\, \mathbb{R}: F_L(l) \geq \alpha\}\} \end{aligned} \quad (1)$$

Because quantiles play a central role in risk management, we restate their precise definition.

(i) For an increasing function T: $\mathbb{R} \rightarrow \mathbb{R}$, the generalized inverse of T is defined by

$T^{\leftarrow}(y) := \inf\{x \in \mathbb{R}: T(x) \geq y\}$, where the infimum of the empty set is taken to be ∞.

(ii) For a distribution function F, the generalized inverse $F^{\leftarrow}$ is called the quantile function of F. For α ∈ (0, 1), the α-quantile of F is given by

$$q_\alpha(F) := F^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R}: F(x) \geq \alpha\} \quad (3)$$

For a random variable X with distribution function F, the notation

$q_\alpha(X) := q_\alpha(F)$ is commonly used when F is continuous and strictly increasing; this simplifies to $q_\alpha(F) = F^{-1}(\alpha)$, where F^{-1} denotes the ordinary inverse of F.

VaR for normal and t loss distributions

Assume that the loss distribution F_L follows a normal distribution with mean μ and variance σ^2. For a chosen confidence level α ∈ (0, 1). Then

$$VaR_\alpha = \mu + \sigma\phi^{-1}(\alpha) \quad (3)$$

where ϕ represents the distribution function of the standard normal and $\phi^{-1}(\alpha)$ is the α-quantile of ϕ. Naturally, these findings can be extended to any location-scale family. The Student t loss distribution serves as another highly practical example of this. Assume a loss variable, L, where the standardized term (L − μ)/σ ollows a standard t distribution characterized by ν degrees of freedom. We can denote this specific distribution as L ∼ t (ν,μ, σ^2). It should be noted that for ν > 2, the moments are defined as E(L) = μ and var(L) = $\nu\sigma^2/(\nu-2)$. Consequently, σ does not act as the actual standard deviation for this distribution. From this, we obtain

$$VaR_\alpha = \mu + \sigma t_\nu^{-1}(\alpha) \quad (4)$$

where TV denotes the df of a standard t distribution.

Expected shortfall (Conditional Value at Risk)

Although ES and VaR are intimately linked, the specific advantages and limitations of these two risk measures remain subjects of ongoing discussion within the risk-management community. For a loss L with $E(|L|) < \infty$ and df F_L, the ES at confidence level α ∈ (0, 1) is defined as

$$ES_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F_L)du, \quad (5)$$

where $q_u(F_L) = F_L^{\leftarrow}(u)$is the quantile function of F_L. The condition E(|L|) < ∞ ensures that the integral in (5) is well defined. By definition, ES is related to VaR by

$$ES_{\alpha} = \frac{1}{1-\alpha}\int_{\alpha}^{1} VaR_{u}(L)du, \tag{6}$$

Rather than anchoring to a specific confidence level α, this approach computes the mean VaR across all quantiles $u \geq \alpha$, thereby exploring further into the tail of the loss distribution. Naturally, ES_{α} is entirely determined by the distribution of the loss variable L, maintaining the inherent condition that $ES_{\alpha} \geq VaR_{\alpha}$.

Suppose that the loss distribution F_L is normal with mean μ and variance σ^2. Fix α ∈ (0, 1). Then

$$ES_{\alpha} = \mu + \sigma\frac{\phi(\phi^{-1}(\alpha))}{1-\alpha} \tag{7}$$

where φ is the density of the standard normal distribution. Assume the loss L is such that standardized variable $\tilde{L}$ = (L−μ)/σ follows a standard t distribution with ν degrees of freedom, with the condition that ν > 1. Applying the same logic as in the preceding example - which remains valid for any location-scale family - we obtain ES_{α} = μ+σ $ES_{\alpha}(\tilde{L})$. The Expected Shortfall for the standard t distribution can be computed directly through integration, giving

$$ES_{\alpha}(\tilde{L}) = \frac{g_{v}(t_{v}^{-1}(\alpha))}{1-\alpha}\left(\frac{v+(t_{v}^{-1}(\alpha))^{2}}{v-1}\right) \tag{8}$$

Because ESα can be interpreted as the mean of all losses that meet or exceed VaRα, it is responsive to the magnitude of extreme losses beyond the VaR threshold.

FZ0 loss score function

In this section, we present the FZ0 loss function, a commonly used tool in the literature for model estimation. Fissler and Ziegel (2016) introduced a new loss function, termed FZ, designed for the simultaneous estimation of Value at Risk (VaR) and Expected Shortfall (ES), which is defined as follows:

$$L_{FZ}(y, v, e; \alpha, G_1, G_2) = (1\{y \le v\} - \alpha)\left(G_1(v) - G_1(y) + \frac{1}{\alpha} G_2(e) v\right) - G_2(e)\left(\frac{1}{\alpha} 1\{y \le v\} y - e\right) g_2(e) \tag{9}$$

Where G_1 is a weakly increasing function, G_2 is a strictly increasing and strictly positive function. The condition $g_2(e) = G_2$ holds, the variable e denotes the Expected Shortfall (ES), and v represents the Value at Risk (VaR).

Estimation of VaR and ES Using the FZ Loss Function

The Value at Risk (VaR) and Expected Shortfall (ES) estimates are derived by minimizing the FZ loss function, as shown below:

$$(VaR_t^\alpha, ES_t^\alpha) = \arg\min E_{t-1}[L_{FZ}(y_t, v, e; \alpha, G_1, G_2)] \tag{10}$$

Patton et al. (2019) examined a particular version of this loss function, known as the FZ0 loss function. This version is obtained by applying the following constraints on G_1 and G_2:

$$G_1(x) = 0$$

$$G_2(x) = -\frac{1}{x}$$

As a result, the **FZ0 loss function** is defined as follows:

$$L_{FZ0}(y, v, e; \alpha) = -\frac{1}{\alpha e} 1\{y \le v\}(v - y) + \frac{v}{e} + \log(-e) - 1 \tag{11}$$

In this study, the FZ0 loss function is employed in the following two stages:

Training phase: Neural network models are trained with this loss function to jointly predict Value-at-Risk (VaR) and Expected Shortfall (ES) values.

Performance evaluation phase: The same loss function is used to evaluate the out-of-sample performance of the neural network models.

Feedforward Neural Network (FNN) and Recurrent Neural Network (RNN) Frameworks

Artificial Neural Networks (ANNs)—commonly referred to simply as neural networks—are computational frameworks designed to identify significant patterns within data by processing information in high-dimensional spaces. These models employ techniques such as dropout layers and regularization methods to address the issue of overfitting.

In this section, two widely used neural network architectures are discussed:

- Feedforward Neural Network (FNN)
- Recurrent Neural Network (RNN)

Both of these models serve as foundational architectures for the models proposed in this thesis. Artificial neural networks represent a nonlinear modeling approach and are theoretically recognized as universal approximators for smooth prediction functions (Hornik et al., 1989).

Structure of Feedforward Neural Networks (FNN)

Among various types of neural networks, Feedforward Neural Networks (FNNs) are composed of three main components:

Input Layer: This layer consists of the initial variables (predictors) that serve as inputs to the network.

Hidden Layers: These intermediate layers process and transform the input variables. Each hidden layer comprises a set of nodes (neurons) that receive information from the previous layer, apply an activation function, and then transmit the processed signals to the subsequent layer.

Output Layer: The function of this layer is to aggregate the information processed through the hidden layers and produce the final prediction.

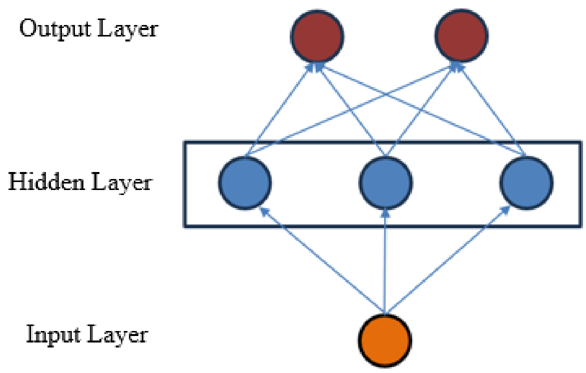


Figure 1. The structure of the FNN model

The hidden layers play a crucial role in learning complex patterns. They contain neurons that interconnect various layers, enabling the flow and transformation of information across multiple stages of the network. Figure 1 illustrates the structure of the FNN model employed in this study:

As illustrated in Figure 2, this network consists of a single hidden layer with three neurons (nodes). Each neuron receives information linearly from all the inputs in the input layer. Then, each neuron applies an activation function $f(.)$ to the weighted sum of the input signals. The output of neuron i in the hidden layer is expressed as follows:

$$x_i^{(1)} = f\left(\theta_{i,0}^{(0)} + \sum_{j=1}^{1} z_j\, \theta_{i,j}^{(0)}\right) \tag{12}$$

Where, z_j denotes the raw inputs to the network, $\theta_{i,j}^{(0)}$are the weight parameters between the inputs and neuron ii in the hidden layer, $f(.)$ is an activation function used to model complex relationships in the data. Finally, the outputs of the neurons in the hidden layer are linearly aggregated and passed to the output layer to produce the final prediction:

$$\theta_0^{(1)} + \sum_{j=1}^{3} x_j^{(1)}\, \theta_j^{(1)} \tag{13}$$

This structure enables the neural network to model complex relationships between input and output variables. (Qiu, Lazar, & Nakata, 2024)

Structure of Recurrent Neural Networks (RNNs)

Linear regression-based methods, including traditional models such as Autoregressive (AR), Moving Average (MA), and the combined Autoregressive Moving Average (ARMA), have been widely employed in the statistical modeling of time series for forecasting target variables. However, in comparison to these traditional linear models, Recurrent Neural Networks (RNNs)-first introduced by Elman (1990)- demonstrate superior performance in capturing complex nonlinear dynamics and long-term serial dependencies in time series data. The general structure of a Recurrent Neural Network (RNN) is defined as follows:

$$\begin{gathered} h_t = \Delta(\mu x_t + w h_{t-1+b}),\ h_0 = 0 \\ y_t | h_t \sim p(y_t | h_t), t = 1,2, ... \end{gathered} \tag{14}$$

Where, h_t represents the hidden state at time t, which retains information from past inputs. x_t denotes the input at time t. μ and w are weight coefficients that determine the influence of the current input and the previous hidden state, respectively. b is the bias term. Δ is an activation function, typically defined as a linear function. y_t is the model output at time t, predicted based on the conditional probability distribution $p(y_t|h_t)$. This recursive structure enables the model to capture and analyze sequential dependencies in time series data. The objective of learning in this model is to estimate the optimal conditional distribution $p(y_t|h_t)$. Furthermore, if the activation function Δ in the above equation is assumed to be linear and the input x_t is taken as the squared returns, the RNN process follows a form analogous to a conditional volatility process. Compared to Feed-Forward Neural Networks (FNNs), Recurrent Neural Networks (RNNs) possess the capability to utilize their internal state (or memory) to process sequences of inputs. As a result, RNNs can more effectively retain and incorporate crucial past information into their decision-making, thereby enhancing predictive performance. (Qiu, Lazar, & Nakata, 2024)

RNN-Based Models

In this section, we introduce three innovative implementations of both stateful and stateless RNN-based architectures tailored specifically for predicting Value-at-Risk (VaR) and Expected Shortfall (ES). During the training phase, these models are configured with stateful=1 to preserve memory and hidden states across consecutive batches of input data. By carrying this information forward from prior batches, the network becomes highly adept at capturing extended long-term dependencies. Furthermore, in contrast to traditional GARCH-type frameworks, this technique does not require strict assumptions about the probability density function of the underlying distribution $F(r_t \mid I_{t-1})$. (Qiu, Lazar, & Nakata, 2024)

STRNN-I Model

To estimate the relationship between the target variables and the auxiliary variable, the STRNN-I model is defined as follows:

$$[v_t, e_t] = FNN(h_t)$$
$$h_t = RNN(h_{t-1}, Y_{t-1}^2, stateful) \tag{15}$$

Where, Y_{t-1}^2 represents the squared return at time $t-1$, h_t is the hidden state within the RNN architecture. stateful is a binary variable that takes the value 1 if the model is stateful and 0 otherwise. The RNN layer is configured in a stateful manner to prevent the hidden state from being reset after processing each batch of data. This allows the model to maintain and transfer information across batches, enhancing its ability to capture long-term dependencies in the data.

STRNN-II Model

To improve the model's ability to capture nonlinear dependencies, an additional layer - specified by the following equation - is integrated between the FNN and the RNN. The STRNN-II model is formulated as:

$$[v_t, e_t] = FNN(h_t)$$
$$k_t = \sqrt{(abs(h_t))} \tag{16}$$
$$h_t = RNN(h_{t-1}, Y_{t-1}^2, stateful)$$

Where, abs(.) denotes the absolute value of the expression inside the parentheses.

When a linear activation function Δ\DeltaΔ is employed within the RNN equation, the structure of the RNN exhibits characteristics similar to those of a GARCH(1,1) model, particularly when the input variable consists of squared returns. Furthermore, this equation brings the STRNN-II model closer in structure to the GARCH-FZ models. This enhanced resemblance makes it significantly easier to interpret how these stateful models successfully capture and predict both Value-at-Risk (VaR) and Expected Shortfall (ES) (Qiu, Lazar, & Nakata, 2024).

STRNN-III Model

Finally, we consider a hybrid model that integrates both STRNN-I and STRNN-II, thereby leveraging both approaches to estimate Value-at-Risk (VaR) and Expected Shortfall (ES).

This model is defined as follows:

$$[v_t, e_t] = -abs([v_t^1, e_t^1] \oplus [v_t^2, e_t^2])$$
$$[v_t^2, e_t^2] = FNN(k_t) \tag{17}$$
$$[v_t^1, e_t^1] = FNN(h_t)$$

$$k_t = \sqrt{\left(abs(h_t)\right)}$$

$$h_t = RNN(h_{t-1}, Y_{t-1}^2, stateful)$$

Where the operation ⊕ denotes element-wise addition, abs(.) refers to the absolute value of the expression within the parentheses. This model combines the characteristics of STRNN-I and STRNN-II and leverages integrated information from both models to enhance the prediction accuracy of Value-at-Risk (VaR) and Expected Shortfall (ES).

We adopt the notation SRNN-VE to denote these stateful RNN-based models used for measuring VaR and ES. These models also incorporate a Feedforward Neural Network (FNN) structure. The notations FNN() and RNN() represent the FNN and RNN components, respectively, within these models. In all three stateful RNN-based models introduced above, the number of nodes is set to 1, and the activation function in both the FNN and RNN layers is set to linear. To mitigate overfitting, we apply early stopping and dropout. To enable a more comprehensive comparison between stateful and stateless models, we implemented three stateless models in addition to the three stateful models presented above. These models are denoted as RNN-I, RNN-II, and RNN-III. The formulations of these models are analogous to the equations defined above, except that the stateful parameter is set to 0 in the equations for these three models. We utilize the TensorFlow API for implementing the Dense layer in the FNN structure:

(tensorflow.keras.layers.Dense). We also use the TensorFlow API to implement the SimpleRNN layer in the RNN structure (tensorflow.keras.layers.SimpleRNN). We monitor validation error throughout training via TensorFlow's EarlyStopping callback: (tensorflow.keras.callbacks.EarlyStopping.). The Dropout technique is applied by specifying a dropout rate directly within TensorFlow's RNN layer: (tensorflow.keras.layers.SimpleRNN (dropout=0.2)).

GARCH Models

GARCH models, initially proposed by Engle (1982) and subsequently extended by Bollerslev (1986), are widely used in financial applications and have gained considerable prominence. Within the family of GARCH-type models, the standard GARCH(1,1) specification represents the most elementary and frequently adopted formulation, which is expressed as follows:

$$\begin{aligned} r_t &= \mu + u_t \\ u_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \beta_0 + \beta_1 u_{t-1}^2 + \beta_2 \sigma_{t-1}^2 \end{aligned} \tag{18}$$

Where, σ_t^2 denotes the daily variance of returns rtr_trt, μ represents the mean return, u_t captures the shock or innovation component of the model, σ_t sigma_tσt is the conditional standard deviation, ϵ_t is an independent random variable following a standard normal distribution. The parameters β_0, β_1, and β_2 determine the dynamics of conditional variance in the model. This model captures the serial dependence in the volatility of financial returns and is extensively used for modeling and forecasting financial market volatility. In parametric GARCH frameworks, the distribution of ϵ_t is specified as a model input and used to estimate parameters via the likelihood function. Two frequently adopted distributions for ϵ_t are the standard normal distribution and the Student's t-distribution. Empirical analyses of financial return series indicate that the Student's t-distribution frequently outperforms the normal distribution, owing to its enhanced ability to capture heavy-tailed returns (Billio & Pelizzon, 2000). Building on the GARCH models described, daily Value-at-Risk (VaR) and Expected Shortfall (ES) are derived using the following equations:

$$VaR_{t+1}^{\alpha} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{q}_{t+1}(\alpha) \tag{19}$$

$$ES_{t+1}^{\alpha} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{s}(\alpha) \tag{20}$$

Where, $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ represent the conditional mean and conditional standard deviation forecasts of r_{t+1}, respectively, based on information available up to time t. $\hat{s}(\alpha)$ is defined as $\mathbb{E}[\epsilon_{t+1}|\epsilon_{t+1} < \hat{q}_{t+1}(\alpha)]$ The quantile $\hat{q}_{t+1}(\alpha)$ is defined as:

$$\hat{q}_{t+1}(\alpha) = \inf\{\epsilon_{t+1}: \alpha \leq F(\epsilon_{t+1})\} \tag{21}$$

where F represents the cumulative distribution function (CDF) of the standardized residuals ϵ. The GARCH model provides the conditional volatility forecast $\hat{\sigma}_{t+1}$, which is derived from historical returns. To estimate Value at Risk (VaR) at confidence level α, one computes a linear function of the forecasted α-quantile of the standardized residuals, $\hat{q}_{t+1}(\alpha)$. Clearly, both VaR and Expected Shortfall (ES) estimators depend on the assumed distribution of the standardized residuals. GARCH-type models are parametric in nature and typically impose a specific distribution on the error term—most commonly the standard normal distribution. However, such rigid assumptions often make it difficult to capture the distinctive statistical properties of financial returns accurately. In this study, three GARCH-type models serve as benchmarks: (1) a GARCH model with standard normally distributed residuals (GCH-N), (2) a GARCH model with Hansen's (1994) skewed t-distribution (GCH-Skt), and (3) a GARCH model in which the distribution is estimated via the empirical distribution function (EDF), offering a nonparametric alternative (GCH-EDF).

Dynamic Quantile (DQ) Test

The Dynamic Quantile (DQ) test is a statistical test used to evaluate the adequacy of quantile forecasts. This test was introduced by Engle and Manganelli (2004) within the framework of the Conditional Autoregressive Value-at-Risk (CAViaR) model. The main objective of the DQ test is to assess whether the quantile forecasts of a time series model accurately represent the conditional distribution of the data. This test is particularly important in risk management and financial econometrics, where quantile forecasts are widely used in models such as Value-at-Risk (VaR). The DQ test helps determine whether the predicted quantile levels are statistically valid in terms of their dynamic properties—that is, whether the actual exceedance rate over time aligns with the expected rate, and whether any systematic deviations are present. (Patton, Ziegel, & Chen, 2019). Financial institutions rely on quantile-based measures (e.g., VaR) to estimate potential losses. The DQ test ensures that these estimates are statistically reliable. Many financial models generate quantile forecasts for risk assessment. The DQ test helps validate these forecasts and ensures their accuracy. Traditional backtesting methods, such as the Kupiec (1995) proportion-of-failures test and the Christoffersen (1998) independence test, evaluate only the unconditional coverage of quantile forecasts. The DQ test, in contrast, evaluates conditional coverage by checking for dynamic misspecification. The DQ test examines whether the realized values of a time series fall within the estimated quantile forecast range at the expected frequency. The hypotheses are defined as follows:

- Null Hypothesis (H_0): The quantile forecasts are conditionally correct, meaning that the realized values fall within the quantile range at the expected rate without systematic deviations.
- Alternative Hypothesis (H_1): The quantile forecasts are misspecified, meaning that the probability of exceedance deviates from the expected level systematically.

If H_0 is rejected, it suggests that the quantile model does not adequately capture the conditional distribution of the data. The DQ regression test with one lag is defined by equation 22. This test provides valuable insights into the reliability of the models under analysis for risk estimation. (Patton, Ziegel, & Chen, 2019)

$$H\ it_{\alpha,t} = w_0 + w_1 H\ it_{\alpha,t-1} + w_2 v_{t-1} + u_t \tag{22}$$

where $H\ it_{\alpha,t}$ is defined as:

$$H\,it_{\alpha,t} = \mathbb{1}\{Y_t \leq v_t\} - \alpha \tag{23}$$

and u_t denotes the regression residual.

Dynamic Expected Shortfall (DES) Regression Test

The Dynamic Expected Shortfall (DES) regression test is a statistical method used to assess the accuracy of Expected Shortfall (ES) forecasts in risk management. Expected Shortfall (ES)—also known as Conditional Value at Risk (CVaR)—is one of the most widely used risk measures, estimating the average expected loss in the worst-case scenarios beyond a specified quantile (i.e., the VaR level). The DES regression test examines whether ES forecasts accurately reflect the dynamics of tail risk or exhibit systematic deviations over time. This test has been introduced as part of broader financial risk evaluation frameworks. It is particularly important under the Basel III regulations, which mandate the use of Expected Shortfall (ES) in place of Value at Risk (VaR) for determining market risk capital requirements. (Patton, Ziegel, & Chen, 2019)

The DES test evaluates whether Expected Shortfall (ES) forecasts appropriately reflect the conditional expected losses beyond the Value-at-Risk (VaR) threshold. The hypotheses of the test are defined as follows:

- Null Hypothesis (H_0): The ES model is correctly specified, meaning that the ES forecasts accurately account for tail risk in the distribution.
- Alternative Hypothesis (H_1): The ES model is misspecified; that is, the ES estimates fail to adjust to past exceedances and exhibit systematic bias appropriately.

If the null hypothesis H_0 is rejected, it indicates that the ES model cannot adequately capture extreme losses. The DES test operates within a conditional regression framework, examining whether past exceedances and other risk factors systematically influence the ES forecasts. (Patton, Ziegel, & Chen, 2019).

The following equation defines the DES regression test:

$$\lambda_{e,t}^{S} = b_0 + b_1 \lambda_{e,t-1}^{S} + b_2 e_{t-1} + u_{e,t} \tag{24}$$

where $\lambda_{e,t}^{S}$ is defined as:

$$\lambda_{e,t}^{S} = \mathbb{1}\{Y_t \leq v_t\} \frac{Y_t}{e_t} - 1 \tag{25}$$

and u_t denotes the regression residual.

Data description

The dataset used in this study consists of the daily closing prices of Tehran Exchange Divedend and Price Index (TEDPIX). The time series spans from June 29, 1992 (8 Tir 1371) to February 9, 2025 (21 Bahman 1403).

This time series is divided into two segments:

- The first 80% of the data is used for model parameter estimation,
- The remaining 20% is used to evaluate and compare model performance.

In this study, we use the squared daily log-returns of the Tehran Exchange Divedend and Price Index as the input data for the RNN-based models. The daily log-returns are computed using the following formula:

$$r_t = \log(P_t) - \log(P_{t-1}) \tag{26}$$

Results

In this section, the performance of the SRNN models will be compared with that of the GARCH models. First, the descriptive statistics of the data will be presented and analyzed. Subsequently, the results obtained from forecasting the two variables, namely Value-at-Risk (VaR) and Expected Shortfall (ES), will be presented. Finally, the performance of the models will be assessed using the

DQ (Dynamic Quantile) and DES (Dynamic Expected Shortfall) tests. Additionally, the performance of the machine learning-based models will be compared with each other using the FZ0 loss function.

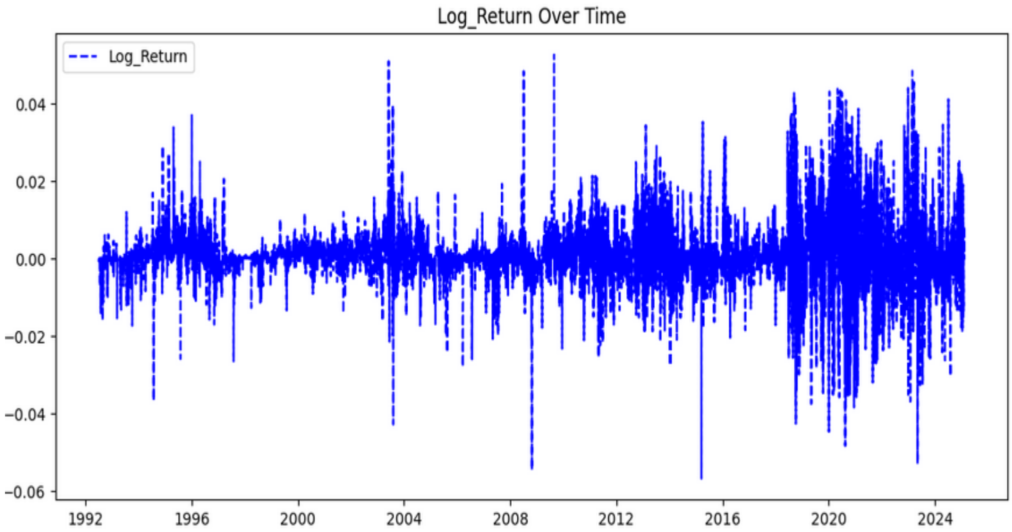
Description of Research Samples

The table below presents the descriptive statistics for the time series of daily returns of the Tehran Exchange Divedend and Price Index (TEDPIX) for the entire sample period.

Table 1. Descriptive Statistics

Statistic	Value
Total count	7878
In-sample	6302
Out-of-sample	1576
Mean	0.110039
StdDev	0.816052
Skew	0.351634
Kurt	7.216230
VaR (0.01)	-2.503297
ES (0.01)	-3.204795
VaR (0.025)	-1.556199
ES (0.025)	-2.443378
VaR (0.05)	-0.985285
ES(0.05)	-1.831971
VaR (0.1)	-0.590678
ES (0.1)	-1.296669

The mean of the daily returns is positive, indicating a slight upward trend in the asset price. The standard deviation of the index reflects the degree of return volatility. The relatively low standard deviation indicates moderate fluctuations across different days. The positive skewness suggests that extreme positive returns have occurred more frequently than extreme negative returns, implying a long right tail in the distribution. The high kurtosis (greater than 3) indicates the presence of fat tails, meaning that the probability of observing extremely large returns (both positive and negative) is higher than that of a standard normal distribution. The volatility of the Tehran Exchange Divedend and Price Index returns is relatively moderate, indicating a reasonable level of risk for this asset. The positive skewness implies that strong upward movements are more common and dominant than severe downward movements. The high kurtosis indicates the presence of extreme events and abnormal fluctuations, underscoring the need for effective risk management strategies. The chart of the logarithmic returns for the entire time series is presented as follows:



The chart of the Tehran Exchange Divedend and Price Index (TEDPIX) returns, after splitting the data into in-sample and out-of-sample sets and after applying the Min-Max scaling normalization method, is presented in Figures 3 and 4:

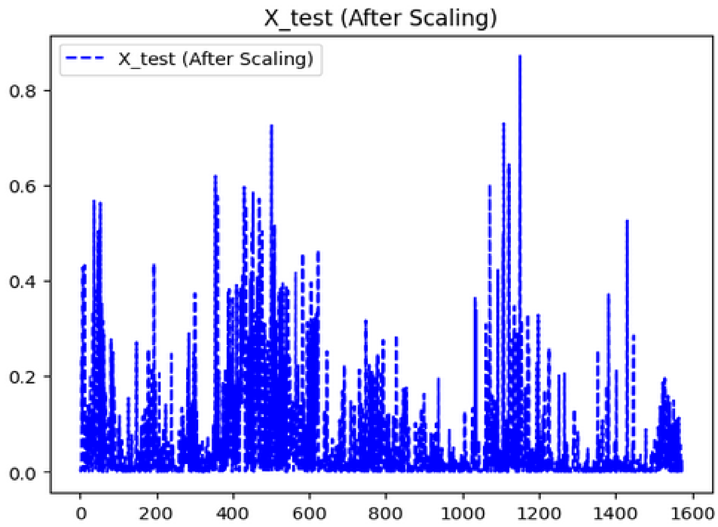


Figure 2. Scaled Returns of the Tehran Exchange Divedend and Price Index (Out-of-Sample Data)

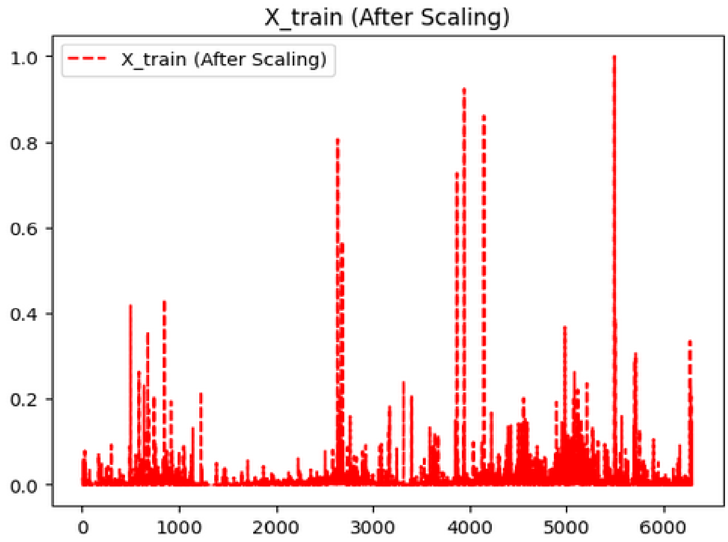


Figure 3. Scaled Returns of the Tehran Exchange Divedend and Price Index (In-Sample Data)

Forecasting Using RNN Models

The training time for each of the machine learning models is presented in the table below:

Table 5. Time cost of training

Model name	Time cost
RNN-I	2 min
RNN-II	3 min
RNN-III	2 min
STRNN-I	5 min
STRNN-II	4 min
STRNN-III	5 min

This table presents the training times required for the stateful and non-stateful RNN models applied to the Tehran Exchange Divedend and Price Index (TEDPIX). The training process was conducted to estimate Value-at-Risk (VaR) and Expected Shortfall (ES) at the 1% confidence level. The training process was executed on a laptop equipped with a 14-core processor running at 2.30 GHz (specifically, an Intel Core i7-12700H).

In the following section, the performance of the three stateful RNN models and the three non-stateful RNN models is evaluated based on the daily returns of the Tehran Exchange Divedend and Price Index (TEDPIX).

STRNN-I Model

he chart illustrating the estimation of Value-at-Risk (VaR) and Expected Shortfall (ES) using the STRNN-I model is presented as follows:

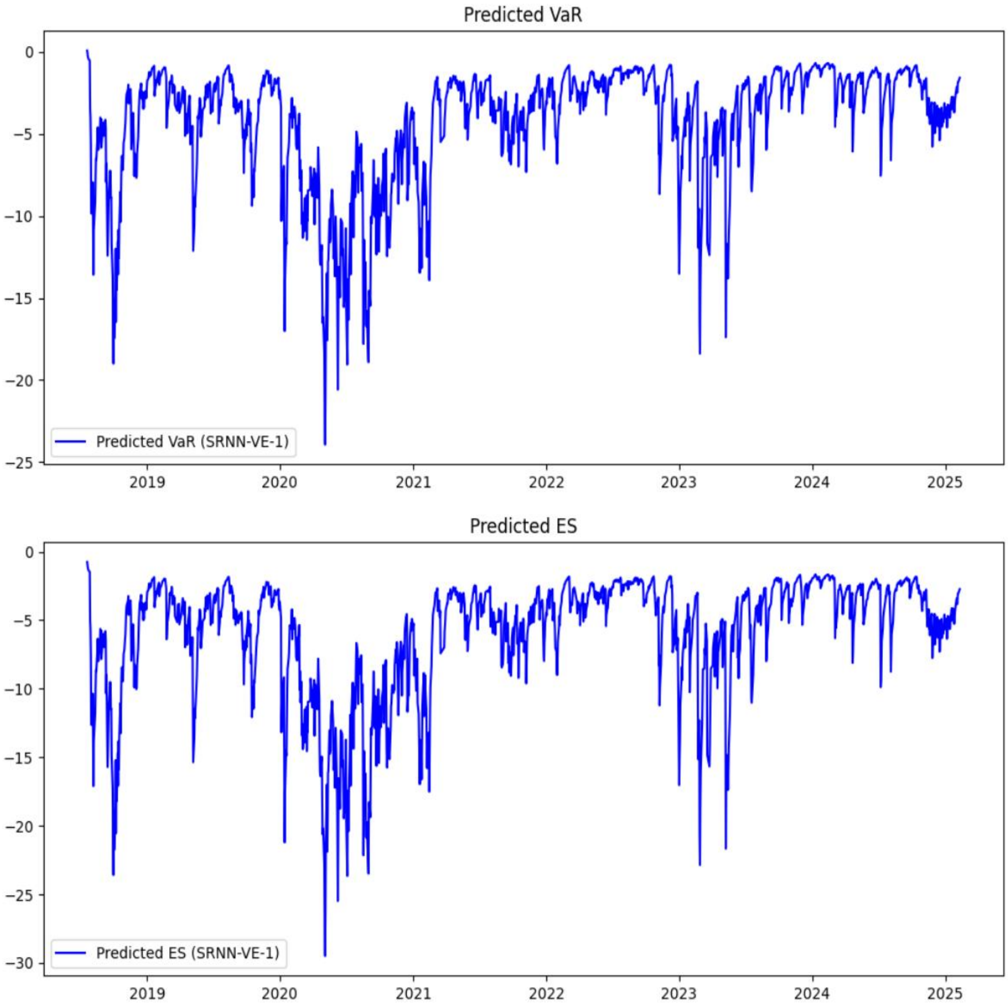


Figure 4. Forecasting Value-at-Risk (VaR) and Expected Shortfall (ES) of the Tehran Exchange Divedend and Price Index Using the SRNN-VE-1 Model

STRNN-II Model

The chart illustrating the estimation of Value-at-Risk (VaR) and Expected Shortfall (ES) using the STRNN-II model is presented as follows:

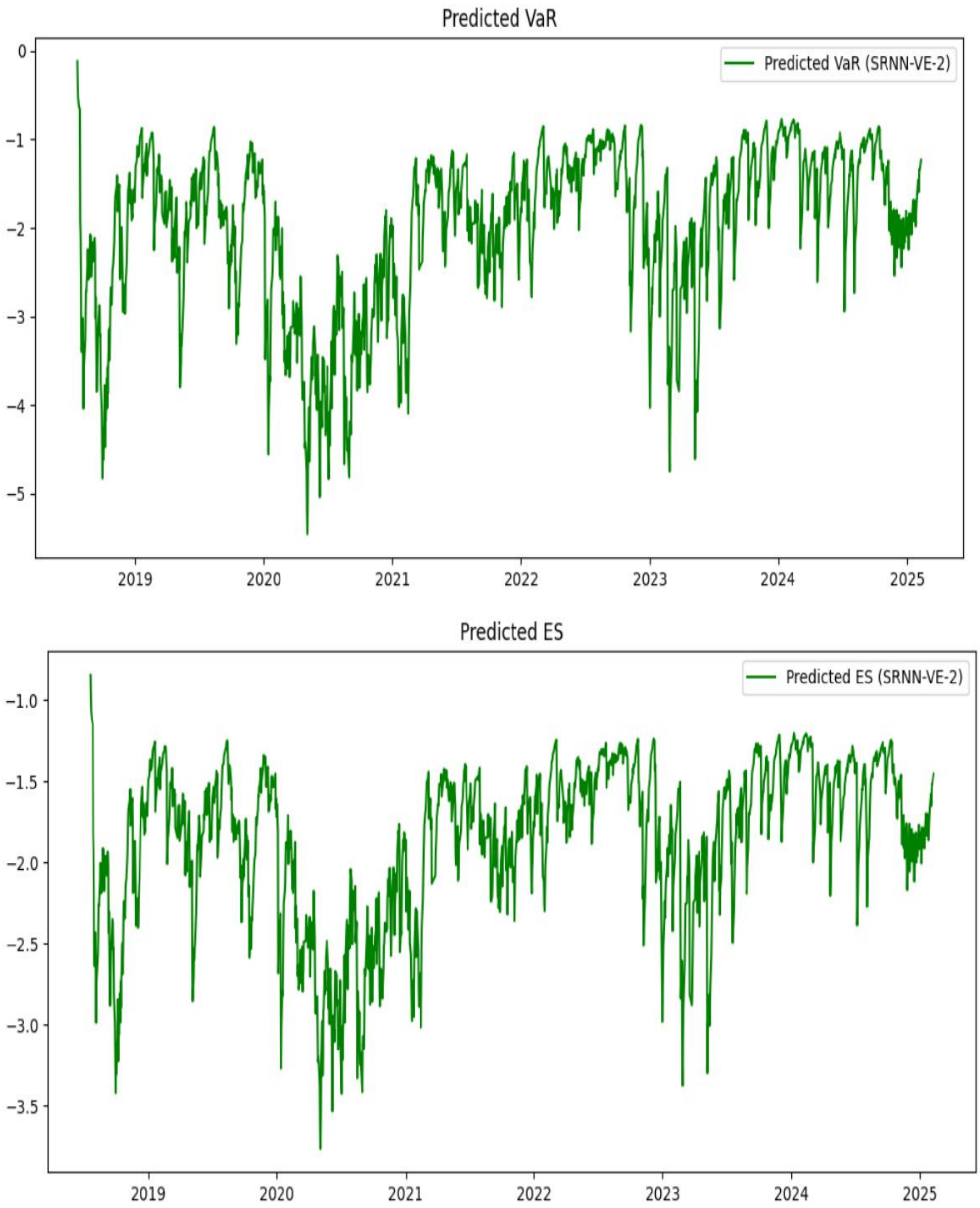


Figure 5. Forecasting Value-at-Risk (VaR) and Expected Shortfall (ES) of the Tehran Exchange Divedend and Price Index Using the SRNN-VE-2 Model

STRNN-III Model

The chart illustrating the estimation of Value-at-Risk (VaR) and Expected Shortfall (ES) using the STRNN-III model is presented as follows:

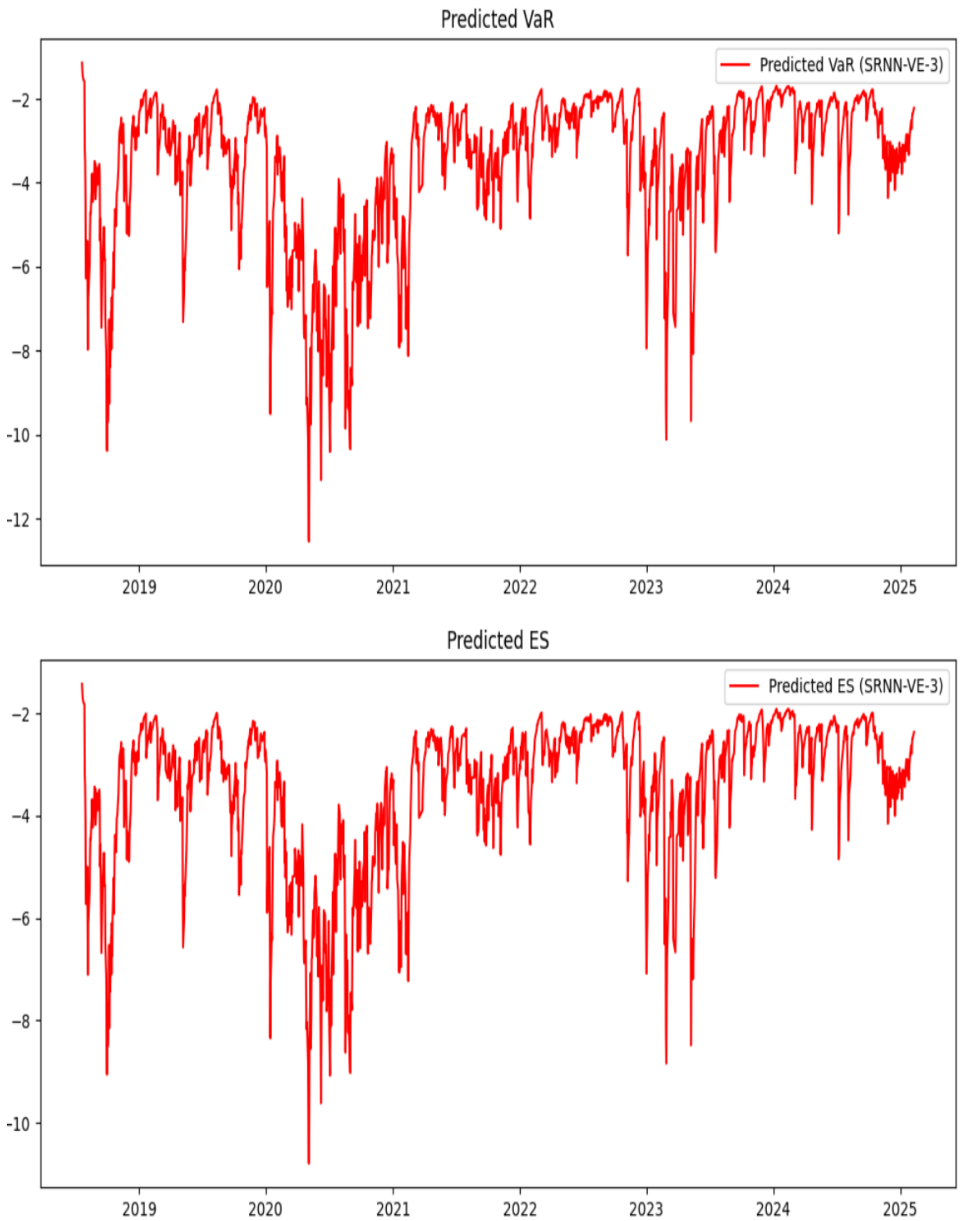


Figure 6. Forecasting Value-at-Risk (VaR) and Expected Shortfall (ES) of the Tehran Exchange Divedend and Price Index Using the SRNN-VE-3 Model

Research Questions

In this section, we address the research questions using the results obtained from the machine learning models and the classical econometric GARCH model. (1) Do stateful Recurrent Neural Network (Stateful-RNN) models perform better than the traditional econometric GARCH model in forecasting the Value-at-Risk (VaR) and Expected Shortfall (ES) of the Tehran Stock Exchange indices? (2) Do stateful Recurrent Neural Network (Stateful-RNN) models outperform non-stateful RNN models in forecasting the Value-at-Risk (VaR) and Expected Shortfall (ES) of the Tehran Stock Exchange indices?

To answer these questions, we use two evaluation methods: the Dynamic Quantile (DQ) test and the Dynamic Expected Shortfall (DES) regression test for assessing model performance.

The evaluation results at the 1% significance level are presented in the following table.

Table 6. Dynamic Quantile (DQ) Test over the out-of-sample period

Dynamic Quantile (DQ) Test	
Models	P-Value
GCH-N	0.0000
GCH-Skt	0.0023
GCH-EDF	0.0000
RNN-I	0.0000
RNN-II	0.0000
RNN-III	0.0000
STRNN-I	0.0625
STRNN-II	0.0999
STRNN-III	0.0390

Table 7. FZ0 loss score function over the out-of-sample period

FZ0 loss score function		
Models	FZ0 score	Ranking
RNN-I	7.1079	6
RNN-II	5.4453	4
RNN-III	6.0592	5
STRNN-I	4.2673	3
STRNN-II	3.5621	2
STRNN-III	3.0492	1

Table 8. Dynamic Expected Shortfall (DES) Regression Test over the out-of-sample period

Dynamic Expected Shortfall (DES) Regression Test	
Models	P-Value
GCH-N	0.0033
GCH-Skt	0.0334
GCH-EDF	0.0035
RNN-I	0.0000
RNN-II	0.0000
RNN-III	0.0000
STRNN-I	0.0000
STRNN-II	0.0000
STRNN-III	0.0387

To answer questions (1) and (2), we employ two evaluation tests: the Dynamic Quantile (DQ) test and the Dynamic Expected Shortfall (DES) test to assess model performance. As shown in the table (6), in the DQ test, all three models — STRNN-I, STRNN-II, and STRNN-III — demonstrate acceptable performance in forecasting VaR, since their p-values exceed the 1% significance level. In contrast, the traditional GARCH models and the non-stateful RNN models, with p-values below the 1% significance level, reject the null hypothesis of the DQ test, indicating that their performance on the DQ test is unsatisfactory.

Moreover, as can be seen from the table (8), in the DES test, two models - GCH-Skt and STRNN-III - exhibit acceptable performance in forecasting ES, as their p-values exceed the 1% significance level. On the other hand, the remaining traditional GARCH models and non-stateful RNN models, with p-values below the 1% significance level, reject the null hypothesis of the DES test, indicating that their performance on the DES test is unsatisfactory.

As shown in the table (7), between the SRNN and RNN models, the SRNN models exhibit better performance in comparison to the RNN models, as indicated by their lower FZ0 loss function values. Furthermore, among the SRNN models, the STRNN-III model demonstrates the best performance on the FZ0 loss function test, with the STRNN-II model ranking second, only slightly behind. Considering the greater complexity of the STRNN-III model and its lower loss function value, it can be concluded that the STRNN-III model provides the best performance in forecasting both the Value-at-Risk (VaR) and Expected Shortfall (ES) of the Tehran Stock Exchange indices.

Conclusion

In this study, the efficiency of stateful Recurrent Neural Networks (Stateful-RNN) in the joint estimation of Value-at-Risk (VaR) and Expected Shortfall (ES) has been examined, and by identifying the weaknesses of traditional econometric models, new approaches have been proposed. The key findings, practical implications, and future research directions are summarized below. In this research, three novel models based on Stateful-RNN — namely, STRNN-I, STRNN-II, and STRNN-III — have been introduced, and their performance has been compared with traditional econometric models such as GARCH. The results indicate that the proposed models exhibit superior performance in risk forecasting.

The distinctive features of stateful RNN models are as follows:

1. The ability to learn long-term dependencies and retain information from the past, which non-stateful models lack.

2. The absence of a need for strict distributional assumptions, which are required in GARCH models, provides greater flexibility for modeling.

3. Improved forecasting accuracy through reducing prediction errors and achieving a stronger correlation with the actual risk values.

As demonstrated in the previous section, the results of the empirical evaluation of machine learning models on the returns of the Tehran Stock Exchange All-Share Index indicate that the SRNN models outperform both the traditional econometric GARCH models and the non-stateful RNN models in forecasting Value-at-Risk (VaR) and Expected Shortfall (ES). One of the most important reasons for the superior performance of machine learning models is that, unlike the GARCH model, these models are not dependent on restrictive distributional assumptions. This study has successfully applied machine learning to financial risk management by introducing stateful Recurrent Neural

Network (RNN) models for joint forecasting of Value-at-Risk (VaR) and Expected Shortfall (ES). Compared to traditional methods, these models demonstrate higher accuracy, greater flexibility, and enhanced capability in learning the complex patterns of financial markets.

The advantages of using machine learning methods can be summarized as follows:

- More accurate forecasts compared to traditional econometric models, such as GARCH
- Elimination of the need for strict assumptions regarding the distribution of financial returns, unlike parametric models
- Faster adaptability to market changes and nonlinear financial behaviors due to the memory structure of stateful RNN models

Given the increasing complexity of financial markets, advanced machine learning-based models, such as stateful RNNs, are expected to play a key role in financial risk management and in providing more accurate forecasts of financial market volatility. This study paves the way for future research in artificial intelligence and financial risk management and can serve as a foundation for developing more innovative models in this domain.

Research Limitations

1. Lack of Using Multivariate Models

In this study, risk forecasting has been conducted in a univariate manner, meaning that each asset has been analyzed independently. However, financial markets typically exhibit strong correlations among assets, and risk management is often performed at the portfolio level rather than solely for a specific asset.

2. Dependence on Optimal Parameter Selection

The performance of machine learning models, including RNN, is highly dependent on the optimal tuning of hyperparameters, such as the number of hidden nodes, the learning rate, and the dropout rate. In this study, standard optimization techniques have been applied. However, future research could benefit from the use of more advanced methods such as Bayesian Optimization or Reinforcement Learning for optimal parameter tuning.

3. Lack of Examination of Crisis Market Conditions

While this study has analyzed data over a long-term period (30 years), certain crises (such as the COVID-19 pandemic) have created unique risk patterns. These conditions were not separately examined in this research.

4. Execution of Machine Learning Models on CPU Instead of GPU

In this study, machine learning models have been executed on a Central Processing Unit (CPU) rather than a Graphics Processing Unit (GPU). GPUs, due to their parallel processing capabilities, can significantly accelerate deep learning model execution, especially when training neural networks on large datasets. Using a CPU has increased model training and processing times and limited the ability to explore more complex architectures and higher-frequency data.

Declaration of Conflicting Interests

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

Funding

The authors received no financial support for the research, authorship and, or publication of this article.

References

Abad, P., Benito, S., & López, C. (2014). A comprehensive review of Value at Risk methodologies. *The Spanish Review of Financial Economics*, *12*(1), 15–32. https://doi.org/10.1016/j.srfe.2013.06.001

Acerbi, C., & Tasche, D. (2002a). Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes*, *31*(2), 379–388. https://doi.org/10.1111/1468-0300.00091

Acerbi, C., & Tasche, D. (2002b). On the coherence of expected shortfall. *Journal of Banking & Finance*, *26*(7), 1487–1503. https://doi.org/10.1016/s0378-4266(02)00283-2

Aggarwal, C. C. (2023). *Neural networks and deep learning: A textbook*. Springer International Publishing.

Alberg, D., Shalit, H., & Yosef, R. (2008). Estimating stock market volatility using asymmetric GARCH models. *Applied Financial Economics*, *18*(15), 1201–1208. https://doi.org/10.1080/09603100701604225

Aliyev, F., Ajayi, R., & Gasim, N. (2020). Modelling asymmetric market volatility with univariate GARCH models: Evidence from Nasdaq-100. *The Journal of Economic Asymmetries*, *22*(e00167), e00167. https://doi.org/10.1016/j.jeca.2020.e00167

Almeida, D. de, & Hotta, L. K. (2014). The leverage effect and the asymmetry of the error distribution in GARCH-based models: The case of Brazilian market-related series. *Pesquisa Operacional*, *34*(2), 237–250. https://doi.org/10.1590/0101-7438.2014.034.02.0237

Ardia, D., Bluteau, K., Boudt, K., & Catania, L. (2018). Forecasting risk with Markov-switching GARCH models:A large-scale performance study. *International Journal of Forecasting*, *34*(4), 733–747. https://doi.org/10.1016/j.ijforecast.2018.05.004

Arian, H., Moghimi, M., Tabatabaei, E., & Zamani, S. (2022). Encoded Value-at-Risk: A machine learning approach for portfolio risk measurement. *Mathematics and Computers in Simulation*, *202*, 500–525. https://doi.org/10.1016/j.matcom.2022.07.015

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance. An International Journal of Mathematics, Statistcs and Financial Economics*, *9*(3), 203–228. https://doi.org/10.1111/1467-9965.00068

BenSaïda, A., Boubaker, S., Nguyen, D. K., & Slim, S. (2018). Value-at-risk under market shifts through highly flexible models. *Journal of Forecasting*, *37*(8), 790–804. https://doi.org/10.1002/for.2503

Billio, M., & Pelizzon, L. (2000). Value-at-Risk: a multivariate switching regime approach. *Journal of Empirical Finance*, *7*(5), 531–554. https://doi.org/10.1016/s0927-5398(00)00022-0

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Bucevska, V. (2013). An empirical evaluation of GARCH models in Value-at-risk estimation: Evidence from the Macedonian stock exchange. *Business Systems Research Journal*, *4*(1), 49–64. https://doi.org/10.2478/bsrj-2013-0005

Chen, H., Wan, Q., & Wang, Y. (2014). Refined Diebold-Mariano test methods for the evaluation of wind power forecasting models. *Energies*, *7*(7), 4185–4198. https://doi.org/10.3390/en7074185

Chronopoulos, I., Raftapostolos, A., & Kapetanios, G. (2024). Forecasting value-at-risk using deep neural network quantile regression. *Journal of Financial Econometrics*, *22*(3), 636–669. https://doi.org/10.1093/jjfinec/nbad014

Cont, R., Cucuringu, M., Xu, R., & Zhang, C. (2022). Tail-GAN: Learning to simulate tail risk scenarios. In *arXiv [q-fin.RM]*. https://doi.org/10.48550/ARXIV.2203.01664

Crouhy, M., Galai, D., & Mark, R. (2014). *The essentials of risk management, second edition* (2nd ed.). McGraw-Hill Professional.

Degiannakis, S., Floros, C., & Dent, P. (2013). Forecasting value-at-risk and expected shortfall using fractionally integrated models of conditional volatility: International evidence. *International Review of Financial Analysis*, *27*, 21–33. https://doi.org/10.1016/j.irfa.2012.06.001

Delbaen, F. (2002). Coherent risk measures on general probability spaces, in *Advances in Finance and Stochastics* (pp. 1–37). Springer Berlin Heidelberg.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics: A Publication of the American Statistical Association*, *13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Duffie, D., & Pan, J. (1997). An overview of value at risk. *The Journal of Derivatives*, *4*(3), 7–49. https://doi.org/10.3905/jod.1997.407971

Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1016/0364-0213(90)90002-e

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, *50*(4), 987. https://doi.org/10.2307/1912773

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics: A*

Publication of the American Statistical Association, *22*(4), 367–381. https://doi.org/10.1198/073500104000000370

Fissler, T., & Ziegel, J. F. (2015). Higher order elicitability and Osband's principle. *arXiv [Math.ST]*. https://doi.org/10.48550/ARXIV.1503.08123

Fong Chan, K., & Gray, P. (2006). Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting*, *22*(2), 283–300. https://doi.org/10.1016/j.ijforecast.2005.10.002

Gerlach, R., & Wang, C. (2020). Semi-parametric dynamic asymmetric Laplace models for tail risk forecasting, incorporating realized measures. *International Journal of Forecasting*, *36*(2), 489–506. https://doi.org/10.1016/j.ijforecast.2019.07.003

Geron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Guermat, C., & Harris, R. D. F. (2002). Forecasting value-at-risk, allowing for time variation in the variance and kurtosis of portfolio returns. *International Journal of Forecasting*, *18*(3), 409–419. https://doi.org/10.1016/s0169-2070(01)00122-4

Hansen, B. E. (1994). Autoregressive Conditional Density Estimation. *International Economic Review*, *35*(3), 705. https://doi.org/10.2307/2527081

Hartz, C., Mittnik, S., & Paolella, M. (2006). Accurate value-at-risk forecasting based on the normal-GARCH model. *Computational Statistics & Data Analysis*, *51*(4), 2295–2312. https://doi.org/10.1016/j.csda.2006.09.017

Holton, G. A. (2003). *Value-at-risk: Theory and practice* (1st ed.). Academic Press.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks: The Official Journal of the International Neural Network Society*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Huang, D., Yu, B., Fabozzi, F. J., & Fukushima, M. (2009). CAViaR-based forecast for oil price risk. *Energy Economics*, *31*(4), 511–518. https://doi.org/10.1016/j.eneco.2008.12.006

Hutchinson, J. M., Lo, A. W., & Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance*, *49*(3), 851. https://doi.org/10.2307/2329209

Ivașcu, C.-F. (2021). Option pricing using Machine Learning. *Expert Systems with*

Applications, *163*(113799), 113799. https://doi.org/10.1016/j.eswa.2020.113799

Jorion, P. (2000). *Value at risk: The new benchmark for managing financial risk* (2nd ed.). McGraw-Hill.

Kakade, K., Jain, I., & Mishra, A. K. (2022). Value-at-Risk forecasting: A hybrid ensemble learning GARCH-LSTM based approach. *Resources Policy*, *78*(102903), 102903. https://doi.org/10.1016/j.resourpol.2022.102903

Khan, M. K., Kim, S.-K., & Alghathbar, K. (2011). Cryptanalysis and security enhancement of a 'more efficient & secure dynamic ID-based remote user authentication scheme.' *Computer Communications*, *34*(3), 305–309. https://doi.org/10.1016/j.comcom.2010.02.011

Koenker, R., & Bassett, G., join(' '. (1978). Regression Quantiles. *Econometrica: Journal of the Econometric Society*, *46*(1), 33. https://doi.org/10.2307/1913643

Lu, X., Ma, F., Xu, J., & Zhang, Z. (2022). Oil futures volatility predictability: New evidence based on machine learning models. *International Review of Financial Analysis*, *83*(102299), 102299. https://doi.org/10.1016/j.irfa.2022.102299

Lucas, A., & Zhang, X. (2016). Score-driven exponentially weighted moving averages and Value-at-Risk forecasting. *International Journal of Forecasting*, *32*(2), 293–302. https://doi.org/10.1016/j.ijforecast.2015.09.003

Mandelbrot, B. B. (1997). The variation of certain speculative prices. In *Fractals and Scaling in Finance* (pp. 371–418). Springer New York.

Mariano, R. S., & Preve, D. (2012). Statistical tests for multiple forecast comparison. *Journal of Econometrics*, *169*(1), 123–130. https://doi.org/10.1016/j.jeconom.2012.01.014

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative Risk Management: Concepts, techniques and tools - revised edition*. Princeton University Press.

Ormaniec, W., Pitera, M., Safarveisi, S., & Schmidt, T. (2022). Estimating value at risk: LSTM vs. GARCH. In *arXiv [q-fin.RM]*. https://doi.org/10.48550/ARXIV.2207.10539

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications*, *42*(1), 259–268. https://doi.org/10.1016/j.eswa.2014.07.040

Patton, A. J., Ziegel, J. F., & Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and Value-at-Risk). *Journal of Econometrics*, *211*(2), 388–413. https://doi.org/10.1016/j.jeconom.2018.10.008

Qiu, Z., Lazar, E., & Nakata, K. (2024). VaR and ES forecasting via recurrent neural network-based stateful models. *International Review of Financial Analysis*, *92*(103102), 103102. https://doi.org/10.1016/j.irfa.2024.103102

Roccioletti, S. (2015). *Backtesting value at risk and expected shortfall* (1st ed.). Springer Gabler.

Saha, S., Gao, J., & Gerlach, R. (2021). Stock movement prediction on ex-divedend day using event-specific features and machine learning techniques—*2021 International Joint Conference on Neural Networks (IJCNN)*.

Shim, J., Kim, Y., Lee, J., & Hwang, C. (2012). Estimating value at risk with semiparametric support vector quantile regression. *Computational Statistics*, *27*(4), 685–700. https://doi.org/10.1007/s00180-011-0283-z

So, M. K. P., & Yu, P. L. H. (2006). Empirical analysis of GARCH models in value at risk estimation. *Journal of International Financial Markets Institutions and Money*, *16*(2), 180–197. https://doi.org/10.1016/j.intfin.2005.02.001

Sollis, R. (2009). Value at risk: a critical overview. *Journal of Financial Regulation and Compliance*, *17*(4), 398–414. https://doi.org/10.1108/13581980911004370

Taylor, J. W. (2007). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, *6*(2), 231–252. https://doi.org/10.1093/jjfinec/nbn001

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics: A Publication of the American Statistical Association*, *37*(1), 121–133. https://doi.org/10.1080/07350015.2017.1281815

Tsay, R. S. (2010). *Analysis of financial time series: Tsay/financial time series 3E* (3rd ed.). Wiley-Blackwell.

Vrontos, S. D., Galakis, J., & Vrontos, I. D. (2021). Implied volatility directional forecasting: a machine learning approach. *Quantitative Finance*, *21*(10), 1687–1706. https://doi.org/10.1080/14697688.2021.1905869

Wang, J., Wang, S., Lv, M., & Jiang, H. (2024). Forecasting VaR and ES by using deep quantile regression, GANs-based scenario generation, and heterogeneous market hypothesis. *Financial Innovation*, *10*(1). https://doi.org/10.1186/s40854-023-00564-5

Wu, Q., & Yan, X. (2019). Capturing deep tail risk via sequential learning of quantile dynamics. *Journal of Economic Dynamics & Control*, *109*(103771), 103771. https://doi.org/10.1016/j.jedc.2019.103771

Xu, Q., Liu, X., Jiang, C., & Yu, K. (2016). Nonparametric conditional autoregressive expectile model via neural network with applications to estimating financial risk: Q. XUET AL. *Applied Stochastic Models in Business and Industry*, *32*(6), 882–908. https://doi.org/10.1002/asmb.2212

Ye, T., & Zhang, L. (2019). Derivatives pricing via machine learning. *Journal of Mathematical Finance*, *09*(03), 561–589. https://doi.org/10.4236/jmf.2019.93029

Zhang, L. (2020). A general framework of derivatives pricing. *Journal of Mathematical Finance*, *10*(02), 255–266. https://doi.org/10.4236/jmf.2020.102016

Zhang, N., Su, X., & Qi, S. (2023). An empirical investigation of multiperiod tail risk forecasting models. *International Review of Financial Analysis*, *86*(102498), 102498. https://doi.org/10.1016/j.irfa.2023.102498

Bibliographic information of this paper for citing:

Namaki, Ali & Emarloo, Ali (2026). Forecasting Value at Risk (VaR) and Expected Shortfall (ES) in the Tehran Stock Exchange Index Using Stateful Recurrent Neural Network (Stateful-RNN) Models. *Iranian Journal of Finance*, 10(2), 85-127.