

Modelling and Forecasting Financial and Economic Time Series Using Neural Networks and Classical Statistical Models

Der Fakultät für Wirtschaftswissenschaften der
Universität Paderborn
zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
-Doctor rerum politicarum-
vorgelegte Dissertation
von

M.Sc. Shujie Li

(2026)

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Prof. Dr. Yuanhua Feng for giving me the opportunity to pursue my PhD research and for his continuous guidance and support throughout my doctoral studies.

I am also grateful to Prof. Dr. Thomas Gries, Prof. Dr. André Uhde, and Dr. Christian Peitz for serving on my doctoral committee and for their valuable comments and constructive suggestions on this work.

Additionally, I would like to extend my special thanks to my colleague Mr. Dominik Schulz for his constant help and encouragement. I am also thankful to my colleagues Mr. Oliver Kojo Ayensu and Mr. Aliyu Abubakar Musa for their support and for the many insightful discussions.

My deepest gratitude also goes to my husband Sebastian, my daughter Lia Jolie, and my son Lennart for their love, patience, and unconditional support. I am also deeply grateful to my parents and parents-in-law, as well as my grandparents for their guidance and support throughout my life. Without their encouragement and understanding, this work would not have been possible.

I would also like to express my gratitude to everyone who has accompanied and supported me throughout my doctoral research and my life.

Last but not least, I dedicate this work to my grandfather and grandmother, who are no longer with us but remain forever in my heart.

Abstract

Accurate modeling of economic and financial time series is essential for forecasting. It helps detect hidden patterns and gain a better understanding of the underlying process. One of the key contributions of this dissertation is the integration of a feedforward neural network (FFNN) with traditional approaches to improve forecasting performance. Neural networks are considered universal function approximators because they are fully data-driven and require no prior knowledge of the data. However, empirical results show that they may perform poorly on raw data. To address this limitation, the series is decomposed into a trend component and a trend-stationary error process. Both deterministic and stochastic trends are considered. The former is estimated nonparametrically via local polynomial regression, while the latter is captured through a differencing filter. The resulting trend-adjusted process is then modeled with a FFNN.

Traditional models are applied to forecast the conditional mean in this thesis, including random walk variants, autoregressive moving average (ARMA), fractionally integrated ARMA (FARIMA) and their semiparametric extensions, namely semiparametric ARMA (Semi-ARMA) and semiparametric FARIMA (Semi-FARIMA). For random walk variants, the differencing filter accounts for the stochastic trend, while Semi-ARMA and Semi-FARIMA apply ARMA and FARIMA models to the detrended process obtained from local polynomial regression.

Volatility estimation plays a key role in risk management, particularly for risk measures like value at risk (VaR) and expected shortfall (ES). An advanced type of recurrent neural networks (RNNs) is used for conditional volatility forecasting and its performance is compared with exponential GARCH (EGARCH) family models across multiple distributions in terms of VaR and ES forecasting. This comparison provides insights into the relative performance of the neural network model and traditional volatility models for risk measurement. Moreover, semiparametric extensions of the EGARCH family models are also employed for volatility estimation. To capture complex dynamics in both the mean and variance simultaneously, this dissertation further proposes a dual-trend and dual long-memory framework. Within this framework, semiparametric models are jointly used to estimate both the first and second moments. This framework is able to capture trend dynamics and long-memory behavior while leaving no systematic patterns in the residuals.

Contents

List of Abbreviations	III
List of Figures	VI
List of Tables	VIII
List of papers	IX
Chapter 1 Introduction	1
1.1 Conventional estimation methods	1
1.1.1 Parametric estimation	2
1.1.2 Nonparametric estimation	3
1.1.3 Semiparametric estimation	5
1.2 Neural networks in time series	7
1.3 Summary of contents	11
Chapter 2 Forecasting economic growth	13
2.1 Introduction	13
2.2 Nonparametric regression approach	15
2.2.1 Local polynomial fitting	15
2.2.2 IPI-algorithm for bandwidth selection	17
2.3 Trend stationary residuals analysis	19
2.3.1 Trend stationary ARMA	19
2.3.2 Trend stationary NNAR	19
2.4 Random walk variants	20
2.5 Forecasting approaches	21
2.5.1 j -step ahead point forecasts for trend-stationary models	21
2.5.2 j -step ahead point forecasts for random walk models	22
2.5.3 Prediction interval estimation	23
2.6 Model evaluation	24
2.7 Data and empirical analysis	25
2.8 Conclusion	31
Chapter 3 Comparing the estimation of VaR and ES	53
3.1 Introduction	53
3.2 Model specifications	55
3.2.1 GARCH and FIGARCH	55
3.2.2 EGARCH and FIEGARCH	57

3.2.3	Log-GARCH and FLog-GARCH	59
3.2.4	MLog-GARCH and FIMLog-GARCH	61
3.2.5	MEGARCH and FIMEGARCH	62
3.2.6	Long short-term memory network	63
3.3	Distributional assumptions	65
3.3.1	Normal distribution	66
3.3.2	Conditional t -distribution	67
3.3.3	Generalized error distribution	67
3.3.4	Average laplace distribution	68
3.3.5	Skewed extensions	69
3.4	Application and backtesting of VaR and ES	71
3.4.1	Calculating VaR and ES	71
3.4.2	Unconditional coverage test	72
3.4.3	Conditional coverage test	73
3.4.4	Traffic light test	74
3.4.5	Loss functions	76
3.5	Data and empirical analysis	77
3.6	Conclusion	86
Chapter 4	Dual-trend and dual long-memory modelling	98
4.1	Introduction	98
4.2	Semi-FARIMA with persistent volatility	100
4.3	Semiparametric long-memory volatility models	101
4.3.1	Semi-FIGARCH	103
4.3.2	Semi-FIEGARCH	104
4.3.3	Semi-FIMLog-GARCH	105
4.3.4	Semi-FIMEGARCH	106
4.3.5	Stationary solutions for EGARCH-type models	106
4.4	Proposed estimation procedure	109
4.4.1	Nonparametric estimation algorithm	109
4.4.2	MLE for estimating the parametric models	111
4.4.3	Properties of the estimation of parametric models	112
4.5	Data and application	115
4.6	Conclusion	120
Chapter 5	Concluding Remarks	121
	Bibliography	123

List of Abbreviations

ACF	autocorrelation function
Adam	adaptive moment estimation
AIC	Akaike information criterion
ALD	asymmetric Laplace distribution
AMISE	asymptotic mean integrated squared error
ANN	artificial neural network
APARCH	asymmetric power ARCH
AR	autoregressive
ARCH	autoregressive conditional heteroskedasticity
ARIMA	autoregressive integrated moving average
ARMA	autoregressive moving average
BIC	Bayesian information criterion
BPTT	backpropagation through time
EGARCH	exponential GARCH
EGF	exponential GARCH family
ES	expected shortfall
ETS	exponential smoothing
FARIMA	fractionally integrated ARMA
FFNN	feedforward neural network
FIAPARCH	fractionally integrated APARCH
FIEGARCH	fractionally integrated EGARCH
FIGARCH	fractionally integrated GARCH
FILog-GARCH	fractionally integrated Log-GARCH
FIMLog-GARCH	fractionally integrated MLog-GARCH
FLF	firm's loss function
FRED	Federal Reserve Economic Data (Federal Reserve Bank of St. Louis)
GARCH	generalized autoregressive conditional heteroskedasticity

GDP	gross domestic product
GED	generalized error distribution
GNP	gross national product
GRU	gated recurrent unit
i.i.d.	independent and identically distributed
IPI	iterative plug-in
LAVE	locally adaptive volatility estimator
LB	Ljung–Box test
LLNNAR	local linear NNAR
LLR	local linear regression
Log-GARCH	logarithmic GARCH
LSE	least squares estimation
LSTM	long short-term memory
MA	moving average
MAE	mean absolute error
MASE	mean absolute scaled error
MdAE	median absolute error
MEGARCH	modified EGARCH
MGF	moment generating function
MISE	mean integrated squared error
MLE	maximum likelihood estimation
MLog-GARCH	modulus asymmetric Log-GARCH
MLP	multi-layer perceptron
MSE	mean squared error
NNAR	autoregressive neural network
PDF	probability density function
POF	proportion of failures
QMLE	quasi-maximum likelihood estimation
RLL	random walk with local linear drift
RMSE	root mean squared error
RMSSE	root mean squared scaled error
RNN	recurrent neural network
RNNAR	random walk with NNAR drift
RSL	random walk with linear drift
RWC	random walk with constant drift
Semi-ARMA	semiparametric ARMA
Semi-FARIMA	semiparametric FARIMA

Semi-FIAPARCH	semiparametric FIAPARCH
Semi-FIEGARCH	semiparametric FIEGARCH
Semi-FILog-GARCH	semiparametric FILog-GARCH
Semi-GARCH	semiparametric GARCH
SEMIFAR	semiparametric fractional autoregressive
VaR	value at risk
WAD	weighted absolute deviation

List of Figures

2.1	Point forecasts and 95% prediction intervals from the combined model M126 for the UK, India, Mexico, China, the USA, and Vietnam.	28
2.2	Point forecasts and 95% prediction intervals from the combined model M126 for France, Japan, Turkey, Germany, Korea and Israel.	29
2.3	Point forecasts and 95% prediction intervals from the combined model M126 for the Netherlands, Canada, Australia, Switzerland and Indonesia.	30
A2.1	Point forecasts and 95% prediction intervals from all single models for Indian.	32
A2.2	Point forecasts and 95% prediction intervals from all single models for the UK.	33
A2.3	Point forecasts and 95% prediction intervals from all single models for the USA.	34
A2.4	Point forecasts and 95% prediction intervals from all single models for Mexico.	35
A2.5	Point forecasts and 95% prediction intervals from all single models for China.	36
A2.6	Point forecasts and 95% prediction intervals from all single models for Vietnam.	37
A2.7	Point forecasts and 95% prediction intervals from all single models for France.	38
A2.8	Point forecasts and 95% prediction intervals from all single models for Japan.	39
A2.9	Point forecasts and 95% prediction intervals from all single models for Turkey.	40
A2.10	Point forecasts and 95% prediction intervals from all single models for Germany.	41
A2.11	Point forecasts and 95% prediction intervals from all single models for Korea.	42
A2.12	Point forecasts and 95% prediction intervals from all single models for Israel.	43
A2.13	Point forecasts and 95% prediction intervals from all single models for the Netherlands.	44
A2.14	Point forecasts and 95% prediction intervals from all single models for Canada.	45
A2.15	Point forecasts and 95% prediction intervals from all single models for Australia.	46

A2.16	Point forecasts and 95% prediction intervals from all single models for Switzerland.	47
A2.17	Point forecasts and 95% prediction intervals from all single models for Indonesia.	48
3.1	Basic LSTM unit following Hewamalage et al. (2021).	64
3.2	The best performing LSTM and GARCH-type models across all distributions for the DAX Index.	80
A3.1	Best performing LSTM and GARCH-type models across all distributions for the DJI index.	87
A3.2	The best performing LSTM and GARCH-type models across all distributions for the EURO STOXX 50 index.	88
A3.3	The best performing LSTM and GARCH-type models across all distributions for the S&P 500 index.	89
A3.4	The best performing LSTM and GARCH-type models across all distributions for the HSI index.	90
A3.5	The best performing LSTM and GARCH-type models across all distributions for the Nikkei 225 index.	91
A3.6	The best performing LSTM and GARCH-type models across all distributions for the NASDAQ index.	92
A3.7	The best performing LSTM across all distributions for the FTSE 100 index.	93
4.1	Dual-trend and dual long-memory example - ABCOMP.	118
4.2	Dual-trend and dual long-memory example - Mortgage.	118
4.3	Dual-trend and dual long-memory example - PPIPLOG.	119
4.4	Dual-trend and dual long-memory example - IPMINE.	119
4.5	Dual-trend and dual long-memory example - WGUK.	120

List of Tables

2.1	Summary of MASE across countries for single models.	26
A2.1	Summary of MASE across countries for all models.	49
3.1	Backtesting results of top-performing LSTM and GARCH-type models. .	79
3.2	Summary of traffic light tests for the best models by distribution. . . .	82
3.3	Coverage tests and loss functions for distribution-wise best models across all datasets.	83
A3.1	Training epochs with WAD values across indices and distributions. . . .	94
4.1	Estimated bandwidths, coefficients of the selected mean models and the LB test statistics.	116
4.2	Estimated bandwidth, coefficients of the selected volatility models and the LB test statistics.	117

List of Papers

Chapter two:

Li, Shujie and Yuanhua Feng. “Forecasting economic growth with traditional methods and a simple neural network model.” CIE Working Paper No. 172, Paderborn University, 2026.

Available at: <https://ideas.repec.org/p/pdn/ciepap/172.html>

Chapter three:

Li, Shujie. “Comparing the estimation of Value at Risk and Expected Shortfall with LSTM and EGARCH family members.” CIE Working Paper No. 173, Paderborn University, 2026.

Available at: <https://ideas.repec.org/p/pdn/ciepap/173.html>

Chapter four:

Li, Shujie and Yuanhua Feng. “Dual-trend and dual long-memory time series modelling.” CIE Working Paper No. 174, Paderborn University, 2026.

Available at: <https://ideas.repec.org/p/pdn/ciepap/174.html>

Introduction

This dissertation consists of three studies on modeling economic and financial time series. The first study develops novel hybrid neural network-based models to capture complex dynamics in the data, and further improves forecasting accuracy through model combination. The second study conducts a comprehensive comparison between EGARCH family models and a long short-term memory (LSTM) network for tail risk estimation. Various backtesting methods are applied to provide insights for firms and regulators. The third study proposes a novel model to capture both trend and long-memory features in the first and second moments, ensuring that the residuals are approximately uncorrelated, which is crucial for valid statistical inference. Together, these studies provide guidance for understanding and modeling complex time series behavior.

1.1 Conventional estimation methods

The classical statistical methodologies for modeling economic and financial time series include parametric, nonparametric, and semiparametric estimation methods. In parametric estimation, a fixed function form is pre-specified, whereas nonparametric estimation allows the data to determine the functional relationship. Semiparametric estimation combines both approaches, which is a concept known as decomposition. The following sections provide a detailed description of each approach in the context of time series modelling.

1.1.1 Parametric estimation

In classical time series analysis, the conditional mean is commonly modeled using the ARMA (Box and Jenkins, 1976) model and homoskedastic innovations are assumed. The foundation of this framework lies in the work of Yule (1927), in which an autoregressive (AR) method was first proposed, and an AR(2) model was used to model annual sunspot numbers. This model was later generalized by Walker (1931), who extended the AR(2) to an AR(p) and derived the Yule-Walker equations. In the 19th century, the moving average (MA) was used as an intuitive smoothing device. The Bank of England applied it as early as 1833 to smooth its bullion reserves (Klein, 2005). The theoretical foundation of the MA process is established by Wold (1938), who referred it as a type of stochastic process. The combination of the AR and MA components results in the ARMA model. The error term, ϵ_t , of the ARMA model is assumed to be homoscedastic, i.e. $\text{var}(\epsilon_t | \mathcal{F}_{t-1}) = \sigma^2$. However, many economic and financial time series violate this assumption and exhibit volatility clustering where periods of high volatility tend to persist. This phenomenon had been long recognized in literature (Mandelbrot et al., 1963; Fama, 1965). However, conditional variance was not explicitly modeled until Engle (1982) proposed the autoregressive conditional heteroskedasticity (ARCH) to model time-varying volatility. The breakthrough of the ARCH is changing the constant conditional variance, $\text{var}(\epsilon_t | \mathcal{F}_{t-1}) = \sigma^2$ to a time-varying process, $\text{var}(\epsilon_t | \mathcal{F}_{t-1}) = \sigma_t^2$. Bollerslev (1986) later introduced the most well-known extension of the ARCH model, the GARCH model, which incorporates both lagged conditional variances and lagged squared residuals. In the following decades, a vast literature of GARCH-type models has emerged to address specific empirical features. Nelson (1991) proposed the EGARCH model to capture the leverage effects, where negative shocks have a larger impact on volatility than positive shocks. Ding et al. (1993) introduced the asymmetric power ARCH (APARCH), which includes power transformations and asymmetry simultaneously. The logarithmic GARCH (Log-GARCH) model was introduced in early studies such as Geweke (1986), Milhøj (1987) and Pantula (1986). Over the past decade, interest in the Log-GARCH has been renewed, and it has been further studied by Sucarrat and Escribano (2010), Francq et al. (2013) and Francq and Sucarrat (2018). Feng et al. (2023) addressed the near-zero or zero return issue in the Log-GARCH model using a modulus-log transformation and then proposed the modulus asymmetric Log-GARCH (MLog-GARCH) model.

Long memory is another important feature of economic and financial time series and it refers to the strong persistence or lasting dependence in the observed series (Banerjee and Urga, 2005). A general definition of long memory is provided by McLeod and Hipel

(1978) and is given by, $\lim_{n \rightarrow \infty} \sum_{j=-n}^n |\gamma_j| = \infty$, γ_j is the autocovariance function at lag j . This phenomenon was first mentioned in time series from hydrology, geophysics, climatology and other natural science (Hurst, 1951; Hurst, 1957; Mandelbrot and Wallis, 1968; Mandelbrot, 1972; McLeod and Hipel, 1978). The most well known long memory time series is the yearly minimal water levels of the Nile River from the years 1007 to 1206. The persistence observed in river flows is known as the Hurst effect. This phenomenon of slowly decaying autocorrelations is also found in economic or financial time series (Banerjee and Urga, 2005). Granger and Joyeux (1980) introduced the FARIMA model, which enables modeling time series with long memory. Hosking (1981) summarized the essential properties of the FARIMA model and presented an approximate maximum likelihood method.

Numerous studies also have documented the presence of long memory in the autocorrelations of squared or absolute financial returns (De Lima et al., 1994; Dacorogna et al., 1993; Harvey, 1993). These works indicate that the autocorrelations of return volatilities decay slowly in a hyperbolic fashion, rather than the exponential decay in the ARCH models. Motivated by these findings, Baillie et al. (1996) proposed the fractionally integrated GARCH (FIGARCH) model. Bollerslev and Mikkelsen (1996) developed the fractionally integrated EGARCH (FIEGARCH) model to account for both long memory in volatility and the leverage effects of positive and negative return innovations. Tse (1998) extended the APARCH model to the fractionally integrated APARCH (FIAPARCH) to model the conditional heteroscedasticity of yen-dollar exchange rate. The fractionally integrated Log-GARCH (FILog-GARCH) was developed by Feng et al. (2020a). Feng et al. (2023) extended the MLog-GARCH model to its long-memory counterpart, the fractionally integrated MLog-GARCH (FIMLog-GARCH) model.

The parametric models in this thesis are estimated using R packages `fEGarch` (Schulz et al., 2025) for GARCH-type models and `fracdiff` (Maechler, 2024) for the FARIMA model.

1.1.2 Nonparametric estimation

Parametric estimation requires to assume a specific functional form in advance without knowing the underlying true shape of the data. In contrast, nonparametric methods are more flexible and can capture more complex patterns in the data. Here, the relationship is driven by the data itself rather than a pre-assumed functional form. Commonly used nonparametric techniques include kernel smoothing and local regression. Both methods

use observations from a small neighborhood around each point. This is the concept of smoothing data, which can be traced back to the late 19th century. One of the earliest references is Woolhouse (1870), who proposed a method based on local quadratic fitting. Henderson (1916) proposed the adjusted moving average. Whittaker and Robinson (1924) and Macaulay (1931) also made early contributions in this area. In the 1950s and 1960s, the modern smoothing (kernel-based smoothing) methods were developed for density (Rosenblatt, 1956; Parzen, 1962) and regression estimation. Kernel regression is a special case of local polynomial regression in which the polynomial degree is 0. One of the most well known kernel-based regression estimators is the Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964), which is equivalent to a local constant estimator. In this framework, the unknown function is estimated by a constant within a small neighborhood of the target point. Recognizing the limitations of the local constant approximation, the more general local polynomial regression regained attention in the late 1970s with significant contributions from Stone (1977), Cleveland (1979), Katkovnik (1979), Fan and Gijbels (1994) and Loader (1996). One advantage of local polynomial is that it automatically corrects the boundary effect (Fan and Gijbels, 1992; Hastie and Loader, 1993; Ruppert and Wand, 1994). Previous studies on the statistic properties of local polynomial regression under short range dependence includes Fan et al. (1996), Masry and Fan (1997), Härdle and Tsybakov (1997) and Härdle et al. (1998). The case of long range dependence has been examined in Hall and Hart (1990), Csörgő and Mielniczuk (1995), Beran and Feng (2002c) and Beran and Feng (2002b).

Selecting an appropriate smoothing parameter, or bandwidth h , is a crucial task in non-parametric regression. The bandwidth governs the balance between bias and variance in the estimated function, $\hat{m}(\cdot)$, and influences the estimation of the dependence structure of the (trend-adjusted) residuals (Beran, 2017). If the bandwidth is too small, the estimated function may fit random fluctuations instead of the underlying pattern. If the bandwidth is too large, it may smooth out true features of the function. The two extreme cases of the bandwidth h are $h = 0$ and $h = \infty$. The former corresponds to no smoothing, yielding $\hat{m}(\cdot) = Y_t$, where Y_t is the observed value at time t . The latter corresponds to complete smoothing with $\hat{m}(\cdot) \equiv \bar{Y}$. Early studies on smoothing parameter selection were carried out by Wahba (1975) and Rice (1984) for smoothing splines and kernel estimators, respectively. Their approaches rely on cross-validation which minimizes the sum of squares of leave-one-out residuals to obtain the optimal bandwidth. Hall and Marron (1991) noted that a major limitation of cross-validation-based bandwidth estimators is their high variability. This is supported by simulations in their study, which shows a poor convergence rate of $O_p(n^{-1/10})$. Gasser et al. (1991) introduced a data-driven iterative

plug-in algorithm (IPI) to determine the optimal bandwidth h_A . They showed that this plug-in estimator is effective with a convergence rate of $O_p(n^{-1/2})$ for smooth functions. h_A is derived from the large-sample approximation to the mean integrated squared error (MISE). The plug-in method estimates the unknown quantities and substitutes them into the expression for h_A .

The R package `smoots` and `esemifar` (Feng et al., 2022; Letmathe et al., 2024) are utilized for the nonparametric estimation under both short- and long-range dependence error structures, respectively.

1.1.3 Semiparametric estimation

Semiparametric estimation essentially decomposes a model into a parametric component and a nonparametric component. Such decomposition has a long history and the idea can be traced back to the early 1900s. According to Persons (1919), four types of components in a time series are identified as trend, cycle, seasonal variations and the irregular fluctuations. Any given time series can be described by one or more of these components (Dagum and Dagum, 2006). In 1971, Anderson showed that a classical model contains a systematic part or trend and a random part or irregular component (Anderson, 2011). Dagum (2010) further summarized that the trend estimation generally follows two main categories including a deterministic trend and a stochastic trend. The deterministic trend can be represented by a fixed smooth function, such as a polynomial or a logistic curve. The stochastic trend can be transformed into a stationary process through a homogeneous filter, most commonly a difference filter. Zhang and Qi (2005) mentioned that trend modeling should depend on the nature of the nonstationarity of the time series. A linear or polynomial trend is suitable if the trend is deterministic, whereas if a series has a stochastic trend, random walk processes and related extensions may be more appropriate. Since it is difficult to distinguish the trend type in practice, both deterministic and stochastic trends are considered in this thesis. To estimate the deterministic trend more flexibly, we use local polynomial regression instead of polynomial regression, as it does not require a fixed functional form and is estimated completely nonparametrically.

The iterative plug-in (IPI) algorithm proposed by Gasser et al. (1991) was developed for semiparametric regression with different types of time series errors exhibiting both short- and long-range dependencies (Beran and Feng, 2002a; Beran and Feng, 2002b; Beran and Feng, 2002c; Feng et al., 2016; Feng et al., 2020b; Letmathe et al., 2024). We assume that the time series $Y_t, t = 1, \dots, n$ can be decomposed into a nonstationary mean component

$m(x_t)$, which is a deterministic trend, and a stochastic term ϵ_t , such that

$$Y_t = m(x_t) + \epsilon_t, \quad (1.1)$$

where $m(\cdot)$ is a nonparametric smooth mean function which is at least twice differentiable on $[0,1]$, and $x_t = t/n \in (0,1]$ is the rescaled time. The stochastic component ϵ_t has an autocovariance function $\gamma_\epsilon(k)$ ($k \in \mathbb{Z}$) and $\gamma_\epsilon(0) = \sigma_\epsilon^2$. ϵ_t is assumed to be second order stationary with zero mean. We distinguish three different types of dependence structures of the error term ϵ_t :

1) i.i.d.:

The entire variance arises from the current innovations, not from any past dependence. In this case, we have $\sum_{k=-\infty}^{\infty} \gamma_\epsilon(k) = \sigma_\epsilon^2$.

2) Short memory:

The autocovariances decay fast enough so that the total sum is still finite, that is $0 < \sum_{k=-\infty}^{\infty} |\gamma_\epsilon(k)| < \infty$. The error process exhibits short memory. In this case, the error term can be assumed to follow an ARMA process.

3) Long-memory:

The error process exhibits long memory and the total sum of the autocovariances diverges, such that $\sum_{k=-\infty}^{\infty} |\gamma_\epsilon(k)| = \infty$. In this case, the error term can be assumed to follow an FARIMA process.

In Chapter 2, both deterministic and stochastic trends are considered. To model the deterministic component, local polynomial regression is applied. The resulting trend-adjusted series is then modeled using an ARMA model. In the case of stochastic trends, random walks with various drift specifications are employed. In Chapter 4, the detrended series is assumed to exhibit long-memory features, corresponding to the third case.

The semiparametric approach is also employed to model the variance. GARCH model variants assume a constant unconditional variance. In practice, however, this assumption does not hold for many financial and economic series, which exhibit unconditional time heteroskedasticity (i.e., time-varying long-term variance). Nonstationarity in financial returns has been studied by Mikosch and Stărică (1998), who indicated that for a GARCH(1,1) model with, $\alpha_1 + \beta_1 \approx 1$, nonstationarity is often implied. Different approaches have been proposed to address this issue by defining the volatility model as a function of past values and time. For instance, the piecewise GARCH model is introduced

by (Mikosch and Stărică, 1998). Mercurio and Spokoiny (2004) proposed a locally adaptive volatility estimator (LAVE) which is a local time homogeneous volatility model. It makes no assumptions about the parametric form of the processes. Härdle et al. (2005) introduced a time heterogeneous stochastic volatility model by defining an interval of homogeneity for a model. Feng (2004) introduced a slowly changing scale function $\sigma(x_t)$ into the GARCH model to address the issue of nonstationary returns. The function is defined as follows,

$$r_t = \sigma(x_t)\sigma_t\eta_t, \quad \eta_t \sim \text{i.i.d.}(0, 1), \quad t = 1, \dots, n, \quad (1.2)$$

where $x_t = t/n$ is rescaled time. The aim of this approach is to account for the unconditional time heteroskedasticity and fit a GARCH model to the resulting stationary process. This model is referred to the semiparametric GARCH (Semi-GARCH) model. Letmathe et al. (2022) extended the Semi-GARCH framework to long-memory GARCH-type models, including semiparametric FIEGARCH (Semi-FIEGARCH), semiparametric FILog-GARCH (Semi-FILog-GARCH) and semiparametric FIAPARCH (Semi-FIAPARCH).

There is also evidence that long memory exists in both conditional mean and conditional variance (Kasman et al., 2009; Conrad and Karanasos, 2005b; Ural and Küçüközmen, 2011). Recently, Peitz et al. (2023) proposed a dual long-memory FARIMA-FILog-GARCH model and provided evidence for trend-stationary dual long-memory time series. Feng et al. (2023) introduced a trend-stationary dual long-memory FARIMA with a novel FIMLog-GARCH specification. To the best of our knowledge, the dual-trend and dual long-memory method has not been studied yet. After examining several economic time series, we find evidence of this phenomenon, which motivates us to propose the dual-trend and dual long-memory model to jointly model the first and second moments of a process.

1.2 Neural networks in time series

Traditional time series models mainly aim to capture the autocorrelation structure within the data and generally require the series to be stationary. The well-known Box and Jenkins approach (Box and Jenkins, 1976) also refers to the ARMA class of model, which is widely recognized in academic and professional fields. Numerous studies have demonstrated its ability to effectively capture the dynamics of many time series. However, the Box and Jenkins approach fails to capture the nonlinear dynamics. Neural networks have been proposed as an alternative for capturing nonlinear patterns in statistical modeling

and have gained popularity in recent years. Since the 2010s, the dramatic increase in computing power has made it possible to run complex networks on standard computers. However, the use of neural networks for prediction is not a recent development. McCulloch and Pitts (1943) and Hebb (2005) established the mathematical and conceptual basis of neural networks. Rosenblatt (1957) proposed one of the first artificial neural network (ANN) models, the single-layer perceptron, which is considered as the first trainable ANN. Minsky and Papert (1969) showed that the problems solvable by Rosenblatt's perceptron are very limited. Werbos (1974) and Werbos (1988) formulated the backpropagation concept and demonstrated that neural networks trained with this method outperform traditional statistical approaches such as the Box-Jenkins models. With the significant increase in computational power during the late 1980s, and the growing popularity of the backpropagation algorithm, research on neural networks accelerated rapidly. In 1987, Lapedes and Farber (1987) proposed one of the earliest framework employing FFNNs to capture nonlinear signal processing dynamics. FFNNs are also known as multi-layer perceptrons (MLPs), which are composed of fully connected layers in which information flows only from the input layer to the output layer without any feedback connections.

Cheng and Titterton (1994) compared different FFNNs with traditional statistical methods and demonstrated strong associations of the FFNNs with discriminant analysis, regression, as well as between supervised neural networks and cluster analysis. Warner and Misra (1996) conducted a comparative study between regression analysis and FFNNs in terms of notation and implementation. They also indicated situations in which neural networks are preferred to parametric regression models. A comparison between FFNNs and logistic regression was carried out by Schumacher et al. (1996) and Vach et al. (1996). Wong et al. (1997) provided a systematic survey of journal publications on the use of neural networks in business between 1988 and 1995, while Vellido et al. (1999) conducted it for the period of 1992-1998. Zhang and Huang (1995) reviewed FFNNs application in manufacturing. Wong and Selvi (1998) focused on financial applications of FFNNs, and Krycha and Wagner (1999) reviewed their use in management science. Paliwal and Kumar (2009) conducted a comparative study of FFNNs and statistical approaches for prediction and classification tasks across various areas, including accounting and finance, health and medicine, engineering and manufacturing, marketing.

Zhang et al. (1998) summarized the main features that make FFNNs appealing for forecasting task. Unlike the traditional models, neural networks are data-driven and self-adaptive, requiring no prior knowledge of the underlying data structure. This makes them particularly useful when the true relationship among variables is unclear. For this reason, they are considered as a class of multivariate, nonlinear and nonparametric statisti-

cal methods (White, 1989b; Ripley, 1993; Cheng and Titterton, 1994; Ripley, 1994). Zhang et al. (1998) also noted that FFNNs can often correctly make prediction for the unseen part of a population even if the sample data contain noise. Additionally, although neural networks are inherently nonlinear models, they are capable of capturing both linear and nonlinear relationships (Zhang et al., 1998). Building on these attractive properties, the autoregressive neural network (NNAR) extends FFNN to time series forecasting by incorporating the classical AR models. NNAR applies a nonlinear activation function to the AR model to capture both linear and nonlinear behaviors in the series (White, 1989a; Qi and Zhang, 2001; Chen and Chang, 2009). Faraway and Chatfield (1998) suggested to use traditional statistical modeling techniques to design an effective neural network specification, such as by determining the number of lagged variables to include as inputs. In chapter two, NNAR based models are employed to forecast GDP in the selected countries. The number of lags is determined based on the optimal order of the AR model, which is selected using the Akaike information criterion (AIC).

Another class of ANNs is RNNs. Unlike FFNNs, where information flows only in one direction, RNNs contain cycles that allow information to be fed back into the network. By incorporating these recurrent connections, RNNs extend the functionality of FFNNs and enable the model to account for both current inputs and information from previous time steps (Bengio et al., 1994). A large number of studies based on RNNs have been published in the literature, Salehinejad et al. (2017) summarized the major advances in RNNs research from 1990 to 2017 (see Table 1 in Salehinejad et al., 2017). Over the past decade, RNNs have been successfully applied to various problem, including language modeling (Graves, 2013; Pascanu and Bengio, 2013; Mikolov et al., 2012; Mikolov et al., 2012; Sutskever et al., 2011), online handwritten recognition (Graves et al., 2008), speech recognition (Graves et al., 2013) and time series modeling (Yunita et al., 2025; Zhang and Xiao, 2000; Jahan and Sajal, 2018; Zhang et al., 2018). Usually, RNNs are constructed using one of the three most popular cells, Elman (Elman, 1990), LSTM (Hochreiter and Schmidhuber, 1997) or gated recurrent unit (GRU) cells (Cho et al., 2014). Developing robust RNN models for time series forecasting remains difficult. There is little guidance on which cell structures are best suited for different types of data (Khaldi et al., 2023). Hewamalage et al. (2021) provided an empirical study comparing various existing RNN architectures with traditional univariate forecasting benchmarks, such as autoregressive integrated moving average (ARIMA) and exponential smoothing (ETS). Their results suggest that RNNs can be competitive in many scenarios, although they are not universally superior. The well known challenge of training RNNs is learning long-term dependencies in data. In RNNs, the hidden state at each time step is conditioned

on its previous state, which enables the network to store and process past information over long periods of time (Mikolov et al., 2014). However, the memory generated by the recurrent connections can be heavily limited by the training algorithms used for RNNs, such as backpropagation through time (BPTT) and truncated BPTT. Gradient vanishing or exploding is a consequence of BPTT with certain activation functions like *tanh* or *sigmoid*. Some advanced variants, such as LSTM networks, are specifically designed to address these limitations. It is one of the most popular and efficient methods to reduce the effect of vanishing or exploding gradient (Hochreiter and Schmidhuber, 1997). Instead of storing past information in hidden units using activation functions like *sigmoid* or *tanh*, LSTM networks use gates to keep, update, or discard in the hidden state (Hochreiter and Schmidhuber, 1997). LSTM-based time series modeling can be found in (Hu et al., 2020; Sahoo et al., 2019; Agarwal et al., 2024; Shelatkar et al., 2020; Bukhari et al., 2020). Despite their advantages, LSTM are computationally more complex than simple RNNs. According to Mikolov et al. (2014), a typical LSTM requires about four times the number of parameters of a simple RNN. In Chapter 3, LSTM is applied for predicting volatility of returns and subsequently VaR and ES are calculated. The results are then compared with traditional EGARCH-type models.

Neural networks are universal function approximators, which implies that they are capable of capturing both seasonality and trend in a time series. Many researchers use neural networks to directly model series with trend or seasonal patterns. Sharda and Patil (1992) tested 88 seasonal time series from the M1 competition (Makridakis et al., 1982) and indicated that simple FFNNs perform as well as the Box-Jenkins method and additional seasonal removal method is not necessary. Similar study can be found in Tang and Fishwick (1993), Nam and Schaefer (1995). However, Nelson et al. (1999) showed that neural networks benefit from deseasonalization, similar to traditional statistical models. Zhang and Qi (2005) claimed that neural networks are not able to handle raw data well, especially when the data contains trend or seasonal patterns. Therefore, detrending or deseasonalization can be used as remedies to reduce forecasting errors. Typically, the trend component is estimated and removed before modeling the remaining components. Qi and Zhang (2008) tested FFNNs on various trend patterns, including linear, non-linear, deterministic, stochastic and breaking trend and they found that their neural network models performed well on the differenced stationary U.S. gross national product (GNP) series. In Chapter 2, we remove the trend for the neural network using local polynomial regression and differencing to achieve stationarity.

The well-known limitations of neural networks should not be overlooked. The training process of neural network models requires a large amount of data available. The architec-

ture and hyperparameters need to be carefully selected to avoid over- and under-fitting. Moreover, they are often criticized for being “black boxes” due to lower interpretability. Furthermore, the computational cost increases dramatically as the network becomes more complex. Therefore, future research could investigate more computationally efficient and interpretable neural network models. Additionally, traditional models like ARIMA and ETS have gained popularity not only due to their high predictive accuracy, but also because they are fully automatic, efficient, and robust. Those features make them suitable for non-expert users. In this sense, neural networks still have a long way to go.

The neural network models in this thesis are trained using the R package `forecast` (Hyndman and Khandakar, 2008), as well as `keras` Chollet et al. (2017) and `tensorflow` (Allaire and Tang, 2017).

1.3 Summary of contents

The aim of chapter 2 is to improve the forecasting quality through various model combinations. Two types of trends are considered, including deterministic and stochastic trends. The former is estimated using a local polynomial estimation while the latter is addressed through differencing. For the deterministic trend, a local polynomial regression is combined with either an ARMA model or an NNAR model. For the stochastic trend, random walks with different drifts are considered, such as constant, linear, local linear and NNAR drifts. In total, six individual models are developed, and their combinations yield 63 different models. The results show that, on average, the newly proposed NNAR-based models outperform other individual models in terms of gross domestic product (GDP) forecasting. Additionally, a single model alone rarely delivers the most accurate forecasts, whereas combining models significantly improves forecasting performance in most cases. Moreover, model combination is less risky than relying on only one model.

Chapter 3 employs an advanced RNN architecture—LSTM network—for forecasting VaR and ES. Its performance is compared with that of recently proposed EGARCH family models. The EGARCH family models are estimated under various symmetric distributions as well as their skewed versions. The subsequent computation of VaR and ES from both LSTM method and EGARCH-type models is carried out under all considered distributions. The backtesting is conducted using traffic light test, conditional and unconditional coverage tests. Model performance is evaluated based on their weighted absolute deviation (WAD). The results indicate that the LSTM model provides the most

satisfactory performance compared with EGARCH family models in most cases. However, several EGARCH family members remain strong competitors, particularly under the asymmetric distributions, and might be preferred for certain indices. Additionally, multiple loss functions are also calculated to evaluate models from a different perspective.

In chapter 4, the dual-trend and dual long-memory model is proposed to capture both nonstationarity and long-memory features in the first and second moments. The trend components in both the mean and variance levels are estimated nonparametrically. The parametric models, including FARIMA (or ARMA), are utilized on the detrended series, and three volatility models from the EGARCH family are applied on the scale-adjusted processes. Additionally, FIGARCH is included as a benchmark model in the final stage of the analysis. A four-step estimation procedure is proposed to estimate the dual-trend and dual long-memory model. The results demonstrate that the proposed model can efficiently capture the process exhibiting trend and long-memory characteristics in both the mean and variance levels with no significant autocorrelation in the residuals.

Concluding Remarks

The aim of this thesis is to model the conditional mean, the conditional variance, or both, of economic and financial time series. Both traditional statistic models and neural networks are applied for this purpose. The R packages `smoots` (Feng et al., 2022) and `esemifar` (Letmathe et al., 2024) are used to estimate the time-varying unconditional mean and unconditional variance, respectively. The subsequent parametric models are estimated using the `fracdiff` (Maechler, 2024) and `fEGarch` (Schulz et al., 2025) packages. The simple FFNN is implemented with the `forecasts` package (Hyndman and Khandakar, 2008), while the LSTM model is constructed via the `keras` package (Chollet et al., 2017) built on tensorflow (Allaire and Tang, 2017).

Chapter 2 aims to forecast the log-GDP for the selected countries. Four models are adopted from Fritz et al. (2024). In addition, two NNAR-based models are proposed, which incorporate deterministic and stochastic trends, respectively. The deterministic trend is estimated using local linear regression, while the stochastic trend component is addressed by applying first order differencing. The NNAR model is then applied to the resulting stationary series. In total, six models and their possible combinations are examined, resulting in a total of 63 models. Model combination is performed using equal weights. The results show that the two newly proposed NNAR-based models, RNNAR and LLNNAR, are the two best-performing single models on average. The combined model M126, which incorporates the RWC, RSL and LLNNAR models, produces the lowest average MASE across all countries. This study shows that combining models is generally effective for improving forecast accuracy. While a single model may perform well in some cases, model combination is still less risky than relying on only one model. A limitation of this study is that only a very simple FFNN model with a single neuron in

both input and hidden layers is used. Further studies will explore more complex neural network architectures to better capture the nonlinearity in the data.

Chapter 3 employs a more sophisticated RNN architecture, namely the LSTM, to forecast VaR and ES for eight indices across the Asian, European, and U.S. markets. The results are primarily evaluated using the traffic light test and the selection measure, WAD. The LSTM-based results are compared with the members of the EGARCH family. Overall, the LSTM consistently outperforms EGARCH-class models, with the exception of the Nikkei 225 index, where the Log-GARCH model slightly outperforms the LSTM. Although the LSTM is superior, the small difference in performance indicates that EGARCH-type models continue to be strong competitors. Moreover, all EGARCH-type models perform better under asymmetric distributional assumptions, which confirms that financial returns exhibit asymmetric features. It should be noted that the high computational cost of neural networks cannot be ignored. For future studies, it is worthwhile to investigate more computationally efficient neural network architectures. Additionally, parametric volatility models can be extended to semi-parametric frameworks to better capture the underlying dynamics of the processes.

Chapter 4 proposes the dual-trend and dual long-memory model to simultaneously model trend and long-memory features in both the mean and the variance. A Semi-FARIMA or Semi-ARMA is utilized on the mean level to model the trend and conditional mean, while a semiparametric volatility model is employed to estimate the conditional variance with a slowly changing scale. FIMEGARCH, FIMLog-GARCH, FIEGARCH and FIGARCH models are applied in the final step to the stationary returns. The results demonstrate that the proposed models successfully identify trend behavior and long-range dependence in both moments. The LB test indicates that there is no evidence of remaining autocorrelation in the residuals, which is essential for valid statistical inference. A simulation study can be carried out to investigate the consistency and convergence rate of the proposed model in future studies.

The main contribution of this work lies in integrating neural networks with traditional models to better capture complex data dynamics. A systematic comparison between the sophisticated neural network model and parametric volatility models is carried out for risk measure estimation, with evaluations from perspectives of both investors and regulators. Furthermore, a newly proposed method enables researchers and practitioners to diagnose hidden patterns and better understand the underlying process.

Bibliography

- [1] Abad, P., Muela, S. B. and Martín, C. L. (2015). The role of the loss function in value-at-risk comparisons. *The Journal of Risk Model Validation* 9, 1.
- [2] Agarwal, H., Mahajan, G., Shrotriya, A. and Shekhawat, D. (2024). Predictive data analysis: Leveraging RNN and LSTM techniques for time series dataset. *Procedia Computer Science* 235, 979–989.
- [3] Allaire, J. J. and Tang, Y. (2017). *tensorflow: R Interface to TensorFlow*. R package version 2.20.0.
- [4] Anderson, T. W. (2011). *The statistical analysis of time series*. John Wiley & Sons.
- [5] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical finance* 9, 203–228.
- [6] Azoff, E. M. (1994). *Neural network time series forecasting of financial markets*. John Wiley & Sons, Inc.
- [7] Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 74, 3–30.
- [8] Baillie, R. T., Han, Y. W. and Kwon, T.-G. (2002). Further long memory properties of inflationary shocks. *Southern Economic Journal* 68, 496–510.
- [9] Banerjee, A. and Urga, G. (2005). Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics* 129, 1–34.

-
- [10] Basel Committee on Banking Supervision (2016). *Minimum capital requirements for market risk*. Bank for International Settlements, Basel.
 - [11] Basel Committee on Banking Supervision (2017). *Basel III: Finalising post-crisis reforms*. Bank for International Settlements, Basel.
 - [12] Bengio, Y., Simard, P. and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks* 5, 157–166.
 - [13] Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 659–672.
 - [14] Beran, J. (2017). *Statistics for long-memory processes*. Routledge.
 - [15] Beran, J. and Feng, Y. (2001). Local polynomial estimation with a FARIMA-GARCH error process. *Bernoulli* 7, 733–758.
 - [16] Beran, J. and Feng, Y. (2002a). Iterative plug-in algorithms for SEMIFAR models—definition, convergence, and asymptotic properties. *Journal of Computational and Graphical statistics* 11, 690–713.
 - [17] Beran, J. and Feng, Y. (2002b). Local polynomial fitting with long-memory, short-memory and antipersistent errors. *Annals of the Institute of Statistical Mathematics* 54, 291–311.
 - [18] Beran, J. and Feng, Y. (2002c). SEMIFAR models—a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis* 40, 393–419.
 - [19] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 31, 307–327.
 - [20] Bollerslev, T. and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of econometrics* 73, 151–184.
 - [21] Boubacar Maïnassara, Y., Esstafa, Y. and Saussereau, B. (2021). Estimating FARIMA models with uncorrelated but non-independent error terms. *Statistical Inference for Stochastic Processes* 24, 549–608.

-
- [22] Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- [23] Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *Journal of Time Series Analysis* 17, 247–270.
- [24] Bukhari, A. H., Raja, M. A. Z., Sulaiman, M., Islam, S., Shoaib, M. and Kumam, P. (2020). Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting. *IEEE Access* 8, 71326–71338.
- [25] Caporin, M. (2002). FIGARCH models: stationarity, estimation methods and the identification problem. *Universita Ca’Foscari di Venezia, Italy*.
- [26] Chen, Y.-h. and Chang, F.-J. (2009). Evolutionary artificial neural networks for hydrological systems forecasting. *Journal of Hydrology* 367, 125–137.
- [27] Cheng, B. and Titterton, D. M. (1994). Neural networks: A review from a statistical perspective. *Statistical science*, 2–30.
- [28] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- [29] Chollet, F., Allaire, J. and Kalinowski, T. (2017). *R Interface to Keras*. R package version 2.16.0.
- [30] Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, 841–862.
- [31] Chryssolouris, G., Lee, M. and Ramsey, A. (1996). Confidence interval prediction for neural network models. *IEEE Transactions on neural networks* 7, 229–232.
- [32] Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association* 74, 829–836.
- [33] Conrad, C. and Karanasos, M. (2005a). Dual long memory in inflation dynamics across countries of the Euro area and the link between inflation uncertainty and macroeconomic performance. *Studies in Nonlinear Dynamics & Econometrics* 9.

-
- [34] Conrad, C. and Karanasos, M. (2005b). On the inflation-uncertainty hypothesis in the USA, Japan and the UK: a dual long memory approach. *Japan and the world Economy* 17, 327–343.
- [35] Costanzino, N. and Curran, M. (2018). A simple traffic light approach to backtesting expected shortfall. *Risks* 6, 2.
- [36] Costanzino, N. and Curran, M. (2015). *Backtesting general spectral risk measures with application to expected shortfall*. Tech. rep. SSRN.
- [37] Csörgő, S. and Mielniczuk, J. (1995). Nonparametric regression under long-range dependent normal errors. *The Annals of Statistics*, 1000–1014.
- [38] Dacorogna, M. M., Müller, U. A., Nagler, R. J., Olsen, R. B. and Pictet, O. V. (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance* 12, 413–438.
- [39] Dagum, E. B. (2010). Time series modeling and decomposition. *Statistica* 70, 433–457.
- [40] Dagum, E. B. and Dagum, C. (2006). Stochastic and deterministic trend models. *Statistica* 66, 269–280.
- [41] De Lima, P. J., Breidt, F. J. and Crato, N. (1994). *Modeling long-memory stochastic volatility*. Johns Hopkins University, Department of Economics.
- [42] Ding, Z. and Granger, C. W. (1996). Modeling volatility persistence of speculative returns: a new approach. *Journal of econometrics* 73, 185–215.
- [43] Ding, Z., Granger, C. W. and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance* 1, 83–106.
- [44] Dybowski, R. and Roberts, S. J. (2001). Confidence intervals and prediction intervals for feed-forward neural networks. Cambridge University Press.
- [45] Elman, J. L. (1990). Finding structure in time. *Cognitive science* 14, 179–211.
- [46] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, 987–1007.

-
- [47] Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business* 38, 34–105.
- [48] Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American statistical Association* 87, 998–1004.
- [49] Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The annals of Statistics*, 196–216.
- [50] Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, 2008–2036.
- [51] Fan, J. and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *Journal of the American Statistical Association* 89, 560–570.
- [52] Fan, J. and Gijbels, I. (2018). *Local polynomial modelling and its applications*. Routledge.
- [53] Fan, J., Gijbels, I., Hu, T.-C. and Huang, L.-S. (1996). A study of variable bandwidth selection for local polynomial regression. *Statistica Sinica*, 113–127.
- [54] Faraway, J. and Chatfield, C. (1998). Time series forecasting with neural networks: a comparative study using the air line data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 47, 231–250.
- [55] Feng, Y. (2004). Simultaneously modeling conditional heteroskedasticity and scale change. *Econometric Theory* 20, 563–596.
- [56] Feng, Y., Beran, J., Letmathe, S. and Ghosh, S. (2020a). *Fractionally integrated Log-GARCH with application to value at risk and expected shortfall*. Universität Paderborn, Center for International Economics.
- [57] Feng, Y., Beran, J. and Yu, K. (2007). Modelling financial time series with SEMIFAR–GARCH model. *IMA Journal of Management Mathematics* 18, 395–412.
- [58] Feng, Y., Forstinger, S. and Peitz, C. (2016). On the iterative plug-in algorithm for estimating diurnal patterns of financial trade durations. *Journal of Statistical Computation and Simulation* 86, 2291–2307.

-
- [59] Feng, Y., Gries, T. and Fritz, M. (2020b). Data-driven local polynomial for the trend and its derivatives in economic time series. *Journal of Nonparametric Statistics* 32, 510–533.
- [60] Feng, Y., Gries, T. and Letmathe, S. (2023). *FIEGARCH, modulus asymmetric FILog-GARCH and trend-stationary dual long memory time series*. Universität Paderborn, Center for international Economics.
- [61] Feng, Y., Gries, T., Letmathe, S. and Schulz, D. (2022). The smoots Package in R for Semiparametric Modeling of Trend Stationary Time Series. *R Journal* 14.
- [62] Feng, Y. and Peitz, C. (2019). Semiparametric GARCH models for value-at-risk and expected shortfall: an object-driven procedure. *Journal of Risk*.
- [63] Feng, Y. and Zhou, C. (2015). Forecasting financial market activity using a semi-parametric fractionally integrated Log-ACD. *International Journal of Forecasting* 31, 349–363.
- [64] Field, P. (2003). Introduction to modern risk management, a history. *Incisive RWG, Haymarket House, London*.
- [65] Francq, C. and Sucarrat, G. (2018). An exponential chi-squared QMLE for log-GARCH models via the ARMA representation. *Journal of Financial Econometrics* 16, 129–154.
- [66] Francq, C., Wintenberger, O. and Zakoian, J.-M. (2013). GARCH models without positivity constraints: Exponential or log GARCH? *Journal of Econometrics* 177, 34–46.
- [67] Fritz, M., Forstinger, S., Feng, Y. and Gries, T. (2024). Forecasting economic growth by combining local linear and standard approaches. *Journal of Applied Statistics*, 1–19.
- [68] Gasser, T., Kneip, A. and Köhler, W. (1991). A flexible and fast method for automatic smoothing. *Journal of the american statistical association* 86, 643–652.
- [69] Geweke, J. (1986). Comment on: modelling the persistence of conditional variances. *Econometric Reviews* 5, 57–61.

-
- [70] Granger, C. W. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis* 1, 15–29.
- [71] Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.
- [72] Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H. and Schmidhuber, J. (2008). A novel connectionist system for unconstrained handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence* 31, 855–868.
- [73] Graves, A., Mohamed, A.-r. and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee, 6645–6649.
- [74] Hall, P. and Hart, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Processes and Their Applications* 36, 339–351.
- [75] Hall, P. and Marron, J. S. (1991). Lower bounds for bandwidth selection in density estimation. *Probability Theory and Related Fields* 90, 149–173.
- [76] Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of econometrics* 81, 223–242.
- [77] Härdle, W., Tsybakov, A. and Yang, L. (1998). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference* 68, 221–245.
- [78] Härdle, W. K., Teyssière, G. and Spokoiny, V. (2005). *Adaptive estimation for a time inhomogeneous stochastic-volatility model*. Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät.
- [79] Harvey, A. (1993). Long Memory in Stochastic Volatility, London School of Economics. *Statistics Preprint Series. LSESPS10*.
- [80] Hastie, T. and Loader, C. (1993). Local regression: Automatic kernel carpentry. *Statistical Science*, 120–129.
- [81] Hebb, D. O. (2005). *The organization of behavior: A neuropsychological theory*. Original work published 1949. Psychology press.

-
- [82] Henderson, R. (1916). Note on graduation by adjusted average. *Transactions of the actuarial society of America* 17, 43–8.
- [83] Hendry, D. F. and Clements, M. P. (2004). Pooling of forecasts. *The Econometrics Journal* 7, 1–31.
- [84] Hewamalage, H., Bergmeir, C. and Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting* 37, 388–427.
- [85] Hibon, M. and Evgeniou, T. (2005). To combine or not to combine: selecting among forecasts and their combinations. *International journal of forecasting* 21, 15–24.
- [86] Hill, T., O'Connor, M. and Remus, W. (1996). Neural network models for time series forecasts. *Management science* 42, 1082–1092.
- [87] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation* 9, 1735–1780.
- [88] Hosking, J. R. (1981). Fractional differencing. *Biometrika* 68, 165–176.
- [89] Hu, J., Wang, X., Zhang, Y., Zhang, D., Zhang, M. and Xue, J. (2020). Time series prediction method based on variant LSTM recurrent neural network. *Neural Processing Letters* 52, 1485–1500.
- [90] Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American society of civil engineers* 116, 770–799.
- [91] Hurst, H. E. (1957). A suggested statistical model of some time series which occur in nature. *Nature* 180, 494–494.
- [92] Hyndman, R. J. (2023). *fma: Data Sets and Functions with "Forecasting: Methods and Applications"*. R package version 2.5.
- [93] Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- [94] Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* 27, 1–22.

-
- [95] Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting* 22, 679–688.
- [96] Jahan, I. and Sajal, S. (2018). Stock price prediction using recurrent neural network (RNN) algorithm on time-series data. *2018 Midwest instruction and computing symposium*. MSRP.
- [97] John, J. and Draper, N. R. (1980). An alternative family of transformations. *Journal of the Royal Statistical Society Series C: Applied Statistics* 29, 190–197.
- [98] Karanasos, M., Psaradakis, Z. and Sola, M. (2004). On the Autocorrelation Properties of Long-Memory GARCH Processes. *Journal of Time Series Analysis* 25, 265–282.
- [99] Kasman, A., Kasman, S. and Torun, E. (2009). Dual long memory property in returns and volatility: Evidence from the CEE countries’ stock markets. *Emerging Markets Review* 10, 122–139.
- [100] Katkovnik, V. Y. (1979). LINEAR AND NON-LINEAR METHODS OF NON-PARAMETRIC REGRESSION-ANALYSIS. *Avtomatika*, 35–46.
- [101] Khaldi, R., El Afia, A., Chiheb, R. and Tabik, S. (2023). What is the best RNN-cell structure to forecast each time series behavior? *Expert Systems with Applications* 215, 119140.
- [102] Klein, J. L. (2005). Statistical visions in time. *Cambridge Books*.
- [103] Krycha, K. A. and Wagner, U. (1999). Applications of artificial neural networks in management science: a survey. *Journal of Retailing and Consumer Services* 6, 185–203.
- [104] Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives* 3, 73–84.
- [105] Lapedes, A. and Farber, R. (1987). *Nonlinear signal processing using neural networks: Prediction and system modelling*. Tech. rep.
- [106] Letmathe, S., Beran, J. and Feng, Y. (2024). An extended exponential SEMIFAR model with application in R. *Communications in Statistics-Theory and Methods* 53, 7914–7926.

-
- [107] Letmathe, S., Feng, Y. and Uhde, A. (2022). Semiparametric GARCH models with long memory applied to Value at Risk and Expected Shortfall. *Journal of Risk*.
- [108] Ling, S. and Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association* 92, 1184–1194.
- [109] Loader, C. R. (1996). Local likelihood density estimation. *The Annals of Statistics* 24, 1602–1618.
- [110] Lopez, J. (1998). Testing your risk tests. *Financial Survey* 20, 18–20.
- [111] Lopez, J. A. (1999). Methods for evaluating value-at-risk estimates. *Economic Review-Federal Reserve Bank of San Francisco*, 3.
- [112] Macaulay, F. R. (1931). *The smoothing of time series*. National Bureau of Economic Research New York.
- [113] Maechler, M. (2024). *fracdiff: Fractionally Differenced ARIMA aka ARFIMA(p, d, q) Models*. R package version 1.5-3.
- [114] Makridakis, A. A., Carbone, R, Fildes, R, Hibon, M, Lewadowski, R, Newton, J and Winkler, R (1982). The Accuracy of Extrapolation (Time Series) Methods: Results of. *Journal of Forecasting (pm—1986)* 50, 2.
- [115] Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International journal of forecasting* 38, 1346–1364.
- [116] Mandelbrot, B. (1972). Statistical methodology for nonperiodic cycles: from the covariance to R/S analysis. *Annals of Economic and Social Measurement, Volume 1, Number 3*. Nber, 259–290.
- [117] Mandelbrot, B. et al. (1963). The variation of certain speculative prices. *Journal of business* 36, 394.
- [118] Mandelbrot, B. B. and Wallis, J. R. (1968). Noah, Joseph, and operational hydrology. *Water resources research* 4, 909–918.

-
- [119] Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics* 24, 165–179.
- [120] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5, 115–133.
- [121] McLeod, A. I. and Hipel, K. W. (1978). Preservation of the rescaled adjusted range: 1. A reassessment of the Hurst Phenomenon. *Water Resources Research* 14, 491–508.
- [122] Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models.
- [123] Mikolov, T., Joulin, A., Chopra, S., Mathieu, M. and Ranzato, M. (2014). Learning longer memory in recurrent neural networks. arXiv preprint arXiv:1412.7753.
- [124] Mikolov, T., Sutskever, I., Deoras, A., Le, H.-S., Kombrink, S. and Cernocky, J. (2012). Subword language modeling with neural networks. Preprint.
- [125] Mikosch, T. and Stărică, C. (1998). *Change of structure in financial time series, long range dependence and the GARCH model*. Chalmers Tekniska Högskola/Göteborgs Universitet. Department of Mathematics.
- [126] Milhøj, A. (1987). *A Multiplicative Parameterization of ARCH Models*. Research Memorandum. Copenhagen, Denmark: University of Copenhagen, Institute of Statistics.
- [127] Minsky, M. and Papert, S. (1969). An introduction to computational geometry. *Cambridge tiass., HIT* 479, 104.
- [128] Missiakoulis, S. (1983). Sargan densities which one? *Journal of Econometrics* 23, 223–233.
- [129] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications* 9, 141–142.
- [130] Nam, K. and Schaefer, T. (1995). Forecasting international airline passenger traffic using neural networks. *The Logistics and Transportation Review* 31, 239–252.

-
- [131] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, 347–370.
- [132] Nelson, M., Hill, T., Remus, W. and O’Connor, M. (1999). Time series forecasting using neural networks: Should the data be deseasonalized first? *Journal of forecasting* 18, 359–367.
- [133] Paliwal, M. and Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert systems with applications* 36, 2–17.
- [134] Pantula, S. G. (1986). Modeling the persistence of conditional variances: a comment. *Econometric Reviews* 5, 79–97.
- [135] Papadopoulos, G., Edwards, P. J. and Murray, A. F. (2001). Confidence estimation methods for neural networks: A practical comparison. *IEEE transactions on neural networks* 12, 1278–1287.
- [136] Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics* 33, 1065–1076.
- [137] Pascanu, R. and Bengio, Y. (2013). Revisiting natural gradient for deep networks. arXiv preprint arXiv:1301.3584.
- [138] Peitz, C., Feng, Y. and Siddiqui, S. (2023). *A FARIMA-FI-Log-GARCH for modeling dual long memory time series in different fields*. Unpublished manuscript.
- [139] Peitz, C., Feng, Y. and Siddiqui, S. (upcoming). *A few useful members of the EGARCH-family with short- or long-memory in volatility*. Unpublished manuscript.
- [140] Persons, W. M. (1919). *Indices of business conditions: an index of general business conditions*. Harvard University Press.
- [141] Qi, M. and Zhang, G. P. (2008). Trend time-series modeling and forecasting with neural networks. *IEEE Transactions on neural networks* 19, 808–816.
- [142] Qi, M. and Zhang, G. P. (2001). An investigation of model selection criteria for neural network time series forecasting. *European journal of operational research* 132, 666–680.

-
- [143] Refenes, A.-P. (1994). *Neural networks in the capital markets*. John Wiley & Sons, Inc.
- [144] Rice, J. (1984). Bandwidth choice for nonparametric regression. *The annals of Statistics*, 1215–1230.
- [145] Ripley, B. D. (1993). Statistical Aspects of Neural Networks. *Networks and Chaos-Statistical and Probabilistic Aspects*. Vol. 50. Monographs on Statistics and Applied Probability. Chapman & Hall, 40–123.
- [146] Ripley, B. D. (1994). Neural networks and related methods for classification. *Journal of the Royal Statistical Society: Series B (Methodological)* 56, 409–437.
- [147] Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* 47, 67–84.
- [148] Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton: (Project Para)*. Cornell Aeronautical Laboratory.
- [149] Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics* 27, 832–837.
- [150] Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The annals of statistics*, 1346–1370.
- [151] Sahoo, B. B., Jha, R., Singh, A. and Kumar, D. (2019). Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophysica* 67, 1471–1481.
- [152] Salehinejad, H., Sankar, S., Barfett, J., Colak, E. and Valaee, S. (2017). Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*.
- [153] Sarma, M., Thomas, S. and Shah, A. (2003). Selection of Value-at-Risk models. *Journal of Forecasting* 22, 337–358.
- [154] Schulz, D., Feng, Y., Peitz, C. and Ayensu, O. K. (2025). *fEGarch: SM/LM EGARCH and GARCH, VaR/ES Backtesting Dual LM Extensions*. R package version 1.0.1.
- [155] Schumacher, M., Roßner, R. and Vach, W. (1996). Neural networks and logistic regression: Part I. *Computational Statistics & Data Analysis* 21, 661–682.

-
- [156] Sharda, R. and Patil, R. B. (1992). Connectionist approach to time series prediction: an empirical test. *Journal of Intelligent Manufacturing* 3, 317–323.
- [157] Shelatkar, T., Tondale, S., Yadav, S. and Ahir, S. (2020). Web traffic time series forecasting using ARIMA and LSTM RNN. *ITM Web of Conferences*. Vol. 32. EDP Sciences, 03017.
- [158] Smith, J. and Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford bulletin of economics and statistics* 71, 331–355.
- [159] Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting* 36, 75–85.
- [160] Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of forecasting* 23, 405–430.
- [161] Stone, C. J. (1977). Consistent nonparametric regression. *The annals of statistics*, 595–620.
- [162] Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics* 34, 2449–2495.
- [163] Sucarrat, G. and Escribano, Á. (2010). *The Power Log-GARCH Model*. Tech. rep. Universidad Carlos III de Madrid, Departamento de Economía.
- [164] Sun, M., Feng, Z., Li, Z., Gu, W. and Gu, X. (2024). Enhancing financial risk management through lstm and extreme value theory: A high-frequency trading volume approach. *Journal of Computer Technology and Software* 3.
- [165] Sutskever, I., Martens, J. and Hinton, G. E. (2011). Generating text with recurrent neural networks. *Proceedings of the 28th international conference on machine learning (ICML-11)*, 1017–1024.
- [166] Tang, Z. and Fishwick, P. A. (1993). Feedforward neural nets as models for time series forecasting. *ORSA journal on computing* 5, 374–385.
- [167] Terui, N. and Van Dijk, H. K. (2002). Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting* 18, 421–438.

-
- [168] Tse, Y. K. (1998). The conditional heteroscedasticity of the yen-dollar exchange rate. *Journal of Applied Econometrics* 13, 49–55.
- [169] Ural, M. and Küçüközmen, C. C. (2011). Analyzing the dual long memory in stock market returns. *Ege Academic Review* 11, 19–28.
- [170] Vach, W., Roßner, R. and Schumacher, M. (1996). Neural networks and logistic regression: Part II. *Computational Statistics & Data Analysis* 21, 683–701.
- [171] Vellido, A., Lisboa, P. J. and Vaughan, J (1999). Neural networks in business: a survey of applications (1992–1998). *Expert Systems with applications* 17, 51–70.
- [172] Vilmarest, J. de and Werge, N. (2024). An adaptive volatility method for probabilistic forecasting and its application to the M6 financial forecasting competition. *International Journal of Forecasting*.
- [173] Wahba, G. (1975). Smoothing noisy data with spline functions. *Numerische mathematik* 24, 383–393.
- [174] Walker, G. T. (1931). On periodicity in series of related terms. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 131, 518–532.
- [175] Warner, B. and Misra, M. (1996). Understanding neural networks as statistical tools. *The american statistician* 50, 284–293.
- [176] Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359–372.
- [177] Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. *PhD thesis, Committee on Applied Mathematics, Harvard University, Cambridge, MA*.
- [178] Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural networks* 1, 339–356.
- [179] White, H. (1989a). An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. *Proceedings of the international joint conference on neural networks*. Vol. 2. Washington, DC, 451–455.

-
- [180] White, H. (1989b). Learning in artificial neural networks: A statistical perspective. *Neural computation* 1, 425–464.
- [181] Whittaker, E. T. and Robinson, G. (1924). *The calculus of observations: a treatise on numerical mathematics*. Blackie and Son limited.
- [182] Wold, H. (1938). A study in the analysis of stationary time series. PhD thesis. Almqvist & Wiksell.
- [183] Wong, B. K., Bodnovich, T. A. and Selvi, Y. (1997). Neural network applications in business: A review and analysis of the literature (1988–1995). *Decision Support Systems* 19, 301–320.
- [184] Wong, B. K. and Selvi, Y. (1998). Neural network applications in finance: A review and analysis of literature (1990–1996). *Information & management* 34, 129–139.
- [185] Woolhouse, W. (1870). Explanation of a new method of adjusting mortality tables, with some observations upon Mr. Makeham’s modification of Gompertz’s theory. *Journal of the Institute of Actuaries* 15, 389–410.
- [186] Yamai, Y. and Yoshida, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance* 29, 997–1015.
- [187] Yule, G. (1927). On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer’s Sunspot Numbers. *Philosophical Transactions of the Royal Society of London Series A* 226, 267–298.
- [188] Yunita, A., Pratama, M. I., Almuzakki, M. Z., Ramadhan, H., Akhir, E. A. P., Mansur, A. B. F. and Basori, A. H. (2025). Performance analysis of neural network architectures for time series forecasting: A comparative study of RNN, LSTM, GRU, and hybrid models. *MethodsX* 15, 103462.
- [189] Zhang, B., Wu, J.-L. and Chang, P.-C. (2018). A multiple time series-based recurrent neural network for short-term load forecasting. *Soft Computing* 22, 4099–4112.
- [190] Zhang, C.-X., Li, J., Huang, X.-F., Zhang, J.-S. and Huang, H.-C. (2022). Forecasting stock volatility and value-at-risk based on temporal convolutional networks. *Expert Systems with Applications* 207, 117951.

- [191] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50, 159–175.
- [192] Zhang, G. P. and Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European journal of operational research* 160, 501–514.
- [193] Zhang, G., Patuwo, B. E. and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International journal of forecasting* 14, 35–62.
- [194] Zhang, H.-C. and Huang, S. (1995). Applications of neural networks in manufacturing: a state-of-the-art survey. *The International Journal of Production Research* 33, 705–728.
- [195] Zhang, J.-S. and Xiao, X.-C. (2000). Predicting chaotic time series using recurrent neural network. *Chinese Physics Letters* 17, 88.

Declaration

I hereby declare that I prepared this thesis entirely on my own and have not used outside sources without declaration in the text. Any concepts or quotations applicable to these sources are clearly attributed to them. This thesis has not been submitted in the same or substantially similar version, not even in part, to any other authority for grading and has not been published elsewhere.

Erklärung

Hiermit versichere ich, durch eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbstständig und ohne unerlaubte Hilfe Dritter angefertigt habe. Alle Stellen, die inhaltlich oder wörtlich aus Veröffentlichungen stammen, sind kenntlich gemacht. Diese Arbeit lag nach meinem Informationsstand in gleicher oder ähnlicher Weise noch keiner Prüfungsbehörde vor und wurde bisher nicht veröffentlicht.

Paderborn, 15 April 2026

A handwritten signature in black ink, reading "Shujie Li". The script is cursive and fluid.

M.Sc. Shujie Li