FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Bc. Jan Hanousek

Dense Zeros

Department of Probability and Mathematical Statistics

Supervisor of the master thesis: doc. RNDr. Michal Pešta, Ph.D.
Study programme: Financial and Insurance Mathematics

Prague 2024

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

I would like to dedicate this thesis to those who have been my constant support throughout this journey.
Firstly, my gratitude goes to my supervisor, doc. RNDr. Michal Pešta, Ph.D., for their great mentorship. Their expertise, encouragement, and balance between helpful guidance and providing me with free space to explore, have been an invaluable help in shaping this project from conception to completion.
To my friends, whose unwavering camaraderie and ability to lift my spirits even during the toughest of times, I owe a huge thank you. Your laughter and joy have been not only a safe place for relaxation, but also a source of strength and motivation that I will always cherish.
Finally, this thesis is dedicated to my loving parents. Their unconditional love, unwavering belief in me, and the comfort of their support have provided the foundation upon which I have been able to achieve this goal. Thank you all.

Title: Dense Zeros

Author: Bc. Jan Hanousek

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Michal Pešta, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This research focuses on a special type of time series data where a significant proportion of values is zero. The aim is to develop a statistical model that accurately captures the behavior of such data. By exploring existing theories on GARCH and MEM models, new models together with derivation of important theoretical properties are proposed. To assess their effectiveness, they are tested on real-world data. This evaluation reveals that each model has its own strengths and weaknesses. The overall results are promising, proving the models' validity and real-world applicability, opening doors for further exploration in this area.

Keywords: stochastic processes, time series, conditional heteroscedasticity models, GARCH, MEM, dependent zeros, non-negative observations, sparse positive observations

Název práce: Husté nuly

Autor: Bc. Jan Hanousek

Katedra: Katedra matematické pravděpodobnosti a statistiky

Vedoucí diplomové práce: doc. RNDr. Michal Pešta, Ph.D., katedra matematické pravděpodobnosti a statistiky

Abstrakt: Tento výzkum se zaměřuje na speciální typ dat časových řad, kde je významný podíl hodnot rovných nule. Cílem je vytvořit statistický model, který přesně zachycuje chování těchto dat. Prostřednictvím zkoumání stávajících teorií o GARCH a MEM modelech jsou navrženy nové modely spolu s odvozením jejich důležitých teoretických vlastností. Pro posouzení jejich účinnosti jsou tyto modely testovány na reálných datech. Toto hodnocení odhaluje, že každý model má své vlastní silné a slabé stránky. Celkové výsledky jsou však nadějné, prokazují platnost modelů a jejich využitelnost v praxi a otevírají dveře pro další výzkum v této oblasti.

Klíčová slova: stochastické procesy, časové řady, modely podmíněné heteroskedasticity, GARCH, MEM, závislé nuly, nezáporná pozorování, řídká kladná pozorování

Contents

Introduction

The analysis of time series data is an important part of probability and statistics, especially in fields like finance and insurance. According to an observed evolution of time series it is possible to forecast the future development. The precision of these forecasts increases with the complexity of the models involved in this task; however, creating too elaborate structures may lead to overfitting and poor performance in longer time horizons. Within this context, the presence of dense zeros, where a significant proportion of observations are zero, poses a unique challenge that requires specialized modeling techniques which are the main topic of this thesis.

The concept of dense zeros in time series data refers to scenarios where a significant proportion of observations are zero and the rest are strictly positive. By delving into the intricacies of dense zeros and, subsequently, into time series where zero-inflated series are prevalent, it is possible to enhance the reliability and applicability of statistical analyses in real-world settings.

To illustrate this concept in practice, we may imagine a time series of daily observed insurance claims on assets such that their occurence is not very common, e.g. transport ships or rare diseases insurance claims. Since these do not happen every day (every time period t), in the observed time series there are many "true zero" observations, i.e. values equal to zero which have not been created by rounding small non-zero values or by other artifical means. It also implies that there cannot be negative values in the series. Therefore, this situation falls into the framework of dense zeros. Naturally, already existing models such as GARCH, MEM, ARMA and others may be called for duty in this case. Nevertheless this specific definition of the problem gives rise to new approaches which can take into account the significant probability of having zero values.

Hence, the goal of this text is to create a model suitable for dealing with these particular scenarios. There already exist various models with numerous properties in the literature and the author's aim is to take inspiration from these findings and propose brand new ways of how to handle the dense-zeros time series effectively while keeping their complexity at reasonable levels.

The structure of this thesis reflects a comprehensive exploration of time series analysis with dense zeros. It begins with a theoretical background that provides an overview of current literature on GARCH models and Multiplicative Error Models (MEMs), introducing the recent findings in these statistical approaches. The subsequent chapters introduce the author's new theory, presenting five new models (DZ-EGARCH1, DZ-EGARCH2, DZ-OGARCH, DZ-GMEM, DZ-NMEM21) specifically designed to address the challenges of dense zeros. These models come with an innovative perspective on modeling zero-inflated series with non-negative values. The practical part of the thesis then focuses on evaluating these models on real-life data. The results, models' performance and implications for future research are to be discussed as well.

1. GARCH Models and MEMs

This introductory chapter focuses on providing an overview of existing theory of two main categories of stochastic models – the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models and Multiplicative Error Models (MEMs). These two families have been proven to exhibit very satisfying theoretical properties as well as to provide meaningful predictions and forecasts in various, not only financial or economic, scenarios.

The first section defines the GARCH models, states their properties and also presents a handful of modifications and extensions stemming from the base model. The second section devoted to MEMs is structured analogously. Additionally, notation and general framework built in these two sections is going to be used throughout the rest of the text.

Since we are not going to implement these models in their already existing form but rather create their modifications designed specifically to fit in the concept of dense zeros, this chapter is not supposed to cover all the findings connected with the mentioned families of models. Hence, the brief summary provided below serves primarily as a theoretical foundation which is going to be utilized in the following text.

1.1 Introductory Overview

This chapter provides a foundation for the analysis of time series data in the context of probability and statistics. We introduce key concepts such as random processes, stationarity, covariance, and functions that measure dependence within a time series over time.

1.1.1 Random Processes and Time Series

A random process, denoted by $\{X_t : t \in T\}$, is a collection of random variables indexed by an index set T, typically representing time. Each element X_t represents the outcome at time t. The set T can be discrete (e.g., hourly temperatures) or continuous (e.g., stock prices over a day). The complete probabilistic behavior of the process is determined by the joint probability distribution of any finite collection of these random variables. Formally, for any finite subset $t_1, t_2, \dots, t_n \in T$, there exists a joint probability distribution function:

$$F_{X_{t_1},X_{t_2},\dots,X_{t_n}}(x_1, x_2, \dots, x_n) = P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n) \tag{1.1}$$

where P denotes the probability measure.

A time series is a specific realization of a random process. It is an ordered sequence of observations collected over time, representing the values of the random variable X_t at different time points. Let $T = \{t_1, \dots, t_N\}$ be a finite index set, then a time series is a particular outcome denoted by $\{y_{t_1}, \dots, y_{t_N}\}$ where y_{t_i} is the observed value of the random variable X_{t_i} at time t_i, $i = 1, \dots, N$.

1.1.2 Stationarity

Stationarity is a crucial property for analyzing time series data. A time series is considered stationary if its statistical properties are constant over time. Formally, a stationary time series satisfies the following conditions, depending on the type of stationarity.

1) Weak Stationarity:

   - The mean $\mu_t = E(X_t)$ and variance $\sigma_t^2 = \mathrm{Var}(X_t)$ are constant for all times $t \in T$,
   - The covariance between any X_t, X_s, defined as

$$\mathrm{cov}(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)] \tag{1.2}$$

     for every t, $s \in T$, depends only on the difference $|t - s|$.

2) Strict Stationarity:

   - The joint probability distribution of any finite collection of random variables $X_{t_1}, \ldots, X_{t_n}$ is equal to the joint probability distribution of $X_{t_1+k}, \ldots, X_{t_n+k}$ for all times $t_i \in T$ and all time lags k.

It can be easily shown that under the assumption of existence and finiteness of first and second moments, the strict stationarity implies weak stationarity. In practice, weak stationarity is the more commonly used condition. Stationarity allows us to develop models and make predictions based on past observations with the assumption that the future will behave similarly.

1.1.3 Autocovariance and Autocorrelation

Further, the function defined in 1.2 is called the autocovariance function because it provides covariances at different time points for different pairs of random variables of the same random process.

There also exists the autocorrelation function which can be described as a normalized autocovariance function. It is defined as

$$\rho_{t,s} = \mathrm{cor}(X_t, X_s), \tag{1.3}$$

where

$$\mathrm{cor}(X_t, X_s) = \frac{\mathrm{cov}(X_t, X_s)}{\sqrt{\mathrm{Var}(X_t)\,\mathrm{Var}(X_s)}}, \tag{1.4}$$

which can be reduced to $\rho_{t,s} = \mathrm{cov}(X_t, X_s)/\sigma_t^2$ if the weak stationarity holds.

1.2 GARCH Models

The fundamentals of GARCH models were established by Bollerslev [1986] whose work expanded upon the ARCH models introduced by Engle [1982]. These models have continually evolved and improved, especially because they have proved very useful in modelling financial time series.

The primary strength of GARCH models lies in their ability to capture heteroscedasticity and varying volatility over time. Despite their successful integration of thick tails and volatility clustering, GARCH models may not fully accommodate additional assumptions such as asymmetry (leverage effect) or non-negativity, as mentioned in Degiannakis and Xekalaki [2004]. However, numerous modifications have been proposed to address these limitations, some of which will be discussed subsequently.

Definitions, notation, and other references in this section draw from the aforementioned literature, as well as from Cipra [2013] and Tsay [2010]. Specific literature will typically be cited in particular instances, such as when introducing a new source or highlighting unique and significant information.

1.2.1 Fundamentals of GARCH Models

Since we are trying to model time series, we denote by t the time index and by T the set of all possible values of t. For our purposes, it is sufficient to limit ourselves to the case of $T \subseteq \mathbb{Z}$; or, in other words, to models with discrete time.

In addition, we want to model financial and/or insurance time series that have been shown to have specific properties. For example, volatility tends to occur in "clusters", i.e. low (or high) levels of volatility are usually expected to follow a period of previously low (or high) volatility. This effect is known as volatility clustering.

The leverage effect is another common phenomenon. This means that the level of volatility changes differently depending on whether the values of the observed variable go up or down. In finance, this often means that a fall in prices is often followed by an increase in volatility (investors may become uncertain about future developments). In the insurance industry, for example, this can happen when there is a natural disaster, the number of claims rises rapidly and volatility increases for several days or weeks.

These and other "specialities" were the motivation for creation of non-linear time series models with conditional expected value μ_t and conditional variance σ_t^2. These are conditioned on some known past information as

$$\mu_t = E(y_t|\Omega_{t-1}), \quad \sigma_t^2 = h_t = \text{Var}(y_t|\Omega_{t-1}). \tag{1.5}$$

We denote by Ω_{t-1} the information set (σ-field) of all information available at time $t-1$. The information set is usually based on the past observations $(y_{t-1}, y_{t-2}, \dots)$ and on the past errors $(e_{t-1}, e_{t-2}, \dots)$.

The error process $\{e_t\}_{t\in T}$ is usually written as $e_t = \sigma_t\varepsilon_t$ where $\{\varepsilon_t\}_{t\in T}$ denotes a sequence of independent, identically distributed (iid) variables with zero mean, unit variance and Normal distribution; that is, $\varepsilon_t \sim N(0,1)$ for all possible $t \in T$.[1]

Having set the fundamentals, we may now define one of the most famous non-linear time series models, the GARCH model.

Definition 1. *Generalized Autoregressive Conditional Heteroscedasticity model of orders* $(p,\, q)$ *is defined as*

$$y_t = \mu_t + e_t, \quad e_t = \sigma_t\varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \tag{1.6}$$

[1] Normality of ε_t is often assumed, but not needed in the general case.

We assume that, for all t, ε_t are independent identically distributed (iid) random variables with $E(\varepsilon_t) = 0$ and $\mathrm{Var}(\varepsilon_t) = 1$, that μ_t is conditional expected values as defined in (1.5) and that for all $i = 1, \ldots, p$ and $j = 1, \ldots, q$ it holds that

$$\alpha_0 > 0, \quad \alpha_i \geq 0, \quad \beta_j \geq 0.$$

Moreover, the GARCH(p, q) process is weakly stationary with $E(e_t) = 0$, $\mathrm{cov}(e_t, e_s) = 0$ for $t \neq s$ and

$$\mathrm{Var}(e_t) = \frac{\alpha_0}{1 - \sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i)} \tag{1.7}$$

if and only if

$$\sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i) < 1. \tag{1.8}$$

This theorem together with its proof may be found in Bollerslev [1986].

Since rather than the original and most general form of GARCH(p, q) models, as in (1.6), we are interested in the particular version GARCH(1, 1) with modifications to suit the dense-zeros scenario, we are going to discuss further properties only for the case where $p = q = 1$.

1.2.2 Properties and Estimation of GARCH(1, 1)

Let us have GARCH(1, 1) model with zero conditional mean value given by

$$y_t = e_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{1.9}$$

under the condition that

$$\alpha_0 > 0, \quad \alpha_1 \geq 0, \quad \beta_1 \geq 0, \quad \alpha_1 + \beta_1 < 1.$$

The last inequality is a necessary and sufficient condition for weak stationarity of the process.

According to Bollerslev [1986], it is also possible to show that the $2m$-th moment of this process can be expressed recursively as

$$E(e_t^{2m}) = \frac{a_m \left[\sum_{n=0}^{m-1} a_n^{-1} E(e_t^{2n}) \alpha_0^{m-n} \binom{m}{m-n} \eta(\alpha_1, \beta_1, n)\right]}{1 - \eta(\alpha_1, \beta_1, m)}, \tag{1.10}$$

where

$$a_0 = 1, \quad a_j = \prod_{i=1}^{j} (2j - 1), \quad j = 1, 2, \ldots$$

and where we assume

$$\eta(\alpha_1, \beta_1, m) = \sum_{j=0}^{m} \binom{m}{j} a_j \alpha_1^j \beta_1^{m-j} < 1. \tag{1.11}$$

Example 1 (Kurtosis of GARCH(1,1) model)**.** Recall the assumptions on the sequence $\{\varepsilon_t\}_{t \in T}$ stating that, for all times t,

$$E(\varepsilon_t) = 0, \quad \mathrm{Var}(\varepsilon_t) = 1.$$

Moreover, let us denote

$$E(\varepsilon_t^4) = K_\varepsilon + 3$$

where K_ε is the excess kurtosis of ε_t. [2] Therefore, if $E(\sigma_t^4)$ exists,

$$\begin{aligned} \text{Var}(y_t) = \text{Var}(e_t) &= \frac{\alpha_0}{1-\alpha_1-\beta_1}, \\ E(y_t^4) = E[E(e_t^4|\Omega_{t-1})] &= E(\sigma_t^4 E(\varepsilon_t^4)) = E(\sigma_t^4)(K_\varepsilon + 3). \end{aligned} \tag{1.12}$$

As proven in Tsay [2010], the excess kurtosis of y_t (given it exists) is

$$\begin{aligned} K_y = \frac{E(y_t^4)}{(E(y_t^2))^2} - 3 &= \frac{(K_\varepsilon + 3)(1-(\alpha 1 + \beta_1)^2)}{1 - 2\alpha_1^2 - (\alpha_1+\beta_1)^2 - K_\varepsilon \alpha_1^2} - 3 \\ &= \frac{K_\varepsilon - K_\varepsilon(\alpha_1+\beta_1) + 6\alpha_1^2 + 3K_\varepsilon\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1+\beta_1)^2 - K_\varepsilon\alpha_1^2}. \end{aligned} \tag{1.13}$$

We may now distinguish between two cases.

(i) Firstly, assume that $\varepsilon_t \sim N(0,1)$. Then its excess kurtosis $K_\varepsilon = 0$ and we obtain

$$K_y^g = \frac{6\alpha_1^2}{1-2\alpha_1^2-(\alpha_1+\beta_1)^2}. \tag{1.14}$$

It is clear now that the kurtosis of y_t exists if $1 - 2\alpha_1^2 - (\alpha_1+\beta_1)^2 > 0$. Additionally, in this Gaussian case, it is visible that $K_y^g = 0$ if and only if $\alpha_1 = 0$, in which case the given GARCH(1,1) model does not include heavy tails.

(ii) On the other hand, if ε_t does not follow Normal distribution, the excess kurtosis of y_t becomes, after rearranging and using (1.14),

$$K_y = \frac{K_\varepsilon + K_y^g + \frac{5}{6}K_\varepsilon K_y^g}{1 - \frac{1}{6}K_\varepsilon K_y^g}. \tag{1.15}$$

This result comes from Bai, Russell, and Tiao [2003]; supposing the kurtosis exists, it holds for all GARCH models, and, subsequently, for its many modifications (e.g. for ARCH(1) model obtained by setting $\beta_1 = 0$).

Finally, we examine predictions of volatility in GARCH(1, 1) model. Let us denote by $\hat{\sigma}_t^2(t-1)$ the forecasted value of σ_t^2 based on Ω_{t-1}, the information available up to the time $t-1$. Clearly

$$\hat{\sigma}_t^2(t-1) = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1\sigma_{t-1}^2. \tag{1.16}$$

Next, we can write

$$\begin{aligned} \sigma_{t+1}^2 &= \alpha_0 + (\alpha_1\sigma_t^2 - \alpha_1\sigma_t^2) + \alpha_1 e_t^2 + \beta_1\sigma_t^2 \\ &= \alpha_0 + (\alpha_1+\beta_1)\sigma_t^2 + \alpha_1\sigma_t^2(\varepsilon_t^2 - 1) \end{aligned} \tag{1.17}$$

[2] Excess kurtosis is defined as (kurtosis - 3) and is used to compare kurtosis of the given (sequence of) random variables against the kurtosis of a normal distribution which is equal to three.

and use that $E(\varepsilon_t^2 - 1|\Omega_{t-1}) = 0$ to obtain

$$\hat{\sigma}_{t+1}^2(t-1) = \alpha_0 + (\alpha_1 + \beta_1)\hat{\sigma}_t^2(t-1). \tag{1.18}$$

Generally, this means that

$$\hat{\sigma}_{t+\tau}^2(t) = \alpha_0 + (\alpha_1 + \beta_1)\hat{\sigma}_{t+\tau-1}^2(t), \quad \tau = 1, 2, \ldots \tag{1.19}$$

and by recursive substitution this leads to

$$\hat{\sigma}_{t+\tau}^2(t) = \frac{\alpha_0(1 - (\alpha_1 + \beta_1)^{\tau-1})}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{\tau-1}\hat{\sigma}_{t+1}^2(t) \to \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \tag{1.20}$$

for $\tau \to \infty$. Hence, as the prediction horizon grows, forecasts of volatility converge to the unconditional variance of the prediction errors e_t.

1.2.3 Modifications of GARCH Models

The original GARCH model addresses many features of financial time series. However, its abilities are limited. If we want to include more complex behaviour in the model, we can choose from its many modifications, which usually add additional terms to the original formula (1.6).

Let us now present some of such modifications which are often used when working with real-world data.

(i) **IGARCH** – Integrated GARCH model of orders p, q is defined as in (1.6) under the condition that

$$\sum_{i=1}^{\max(p,q)} (\alpha_1 + \beta_1) = 1. \tag{1.21}$$

This results in so called "persistence in variance" – as opposed to original GARCH where the predictions of volatility converge with increasing time horizon to unconditional variance, in IGARCH the current information is significant for predictions in all future times.

Hence, the formula (1.19) can be simplified into

$$\hat{\sigma}_{t+\tau}^2(t) = \alpha_0 + \hat{\sigma}_{t+\tau-1}^2(t), \quad \tau = 1, 2, \ldots \tag{1.22}$$

to obtain, recursively,

$$\hat{\sigma}_{t+\tau}^2(t) = (\tau - 1)\alpha_0 + \hat{\sigma}_{t+1}^2(t), \quad \tau = 1, 2, \ldots \tag{1.23}$$

This proves the previous claim of persistence and also shows that the predictions follow a straight line with its slope equal to α_0.

(ii) **GJR-GARCH** – Glosten-Jagannathan-Runkle GARCH is a model addressing the assymetric leverage effect, i.e. positive disturbances may have a different effect on volatility than negative ones. It was set in the work of Glosten, Jagannathan, and Runkle [1993] and its most prominent form is

$$y_t = \mu_t + e_t, \quad e_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + (\alpha_1 + \gamma_1)e_{t-1}\mathbb{1}_{t-1}^-, \tag{1.24}$$

where

$$\mathbb{1}_{t-1}^{-} = \begin{cases} 1 & \text{for } e_t < 0, \\ 0 & \text{for } e_t \geq 0. \end{cases} \tag{1.25}$$

Such model specifies a parametric form for conditional heteroscedasticity, directly distinguishing between de facto two different models; one for negative errors and one for positive errors.

The threshold, however, does not have to be set to zero as in (1.25), allowing for further modifications of various threshold GARCH (TGARCH) models.

(iii) **EGARCH** – Exponential GARCH model was first presented by Nelson [1991] and today, more convenient, but still equivalent, forms of the original model exist (see e.g. Cipra [2013] or Huang, Wang, and Yao [2008]). Let us show an alternative form used by Tsay [2010]

$$\begin{aligned} y_t &= \mu_t + e_t, \\ e_t &= \sigma_t \varepsilon_t, \\ \ln(\sigma_t^2) &= \alpha_0 + \sum_{i=1}^{s} \alpha_i \frac{|e_{t-i}| + \gamma_i e_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{m} \beta_j \ln(\sigma_{t-j}^2). \end{aligned} \tag{1.26}$$

In this case, if the error e_{t-i} is positive (an "upward" movement of the process), the log-volatility changes according to

$$\alpha_i(1+\gamma_i)|\varepsilon_{t-i}|,$$

whereas for negative e_{t-i} (a "downward" movement) the change is proportional to

$$\alpha_i(1-\gamma_i)|\varepsilon_{t-i}|.$$

The leverage effect is thus represented by the parameter γ_i. Moreover, when applied to financial time series, where we assume that negative disturbances affect volatility more than positive ones, we would expect $\gamma_i < 0$.

Finally, the use of the logarithm of the conditional variance ensures positivity and we do not need to impose any further positivity constraints on the coefficients.

(iv) **OGARCH** – Overresponse GARCH model was created in order to capture the overresponse in markets caused by spells of positive and/or negative shocks. First presented in the work of Liu and Morimune [2005], from which the following takes inspiration, it can be defined (in a slightly modified form to fit our notation) as

$$\begin{aligned} y_t &= \mu_t + e_t, \\ e_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \exp(\theta \gamma_{t-1}) e_{t-1}^2 + \beta \sigma_{t-1}, \end{aligned} \tag{1.27}$$

assuming $\theta \in \mathbb{R}$, $\alpha_0 > 0$ and $\alpha_1, \beta \geq 0$. Most importantly, $\gamma_{t-1} \in \mathbb{N}$ is the number of days in the spells of shocks. Formally,

$$\gamma_{t-1} = i \iff \operatorname{sign}(e_{t-1}) = \cdots = \operatorname{sign}(e_{t-i}) = -\operatorname{sign}(e_{t-i-1}). \tag{1.28}$$

It follows that, at every time period, one of only two values can be attained, namely

$$\gamma_{t-1} = \begin{cases} 1 & \text{if } \operatorname{sign}(e_{t-1}) = -\operatorname{sign}(e_{t-2}), \\ \gamma_{t-2}+1 & \text{if } \operatorname{sign}(e_{t-1}) = \operatorname{sign}(e_{t-2}). \end{cases} \tag{1.29}$$

Moreover, if we define $p = P(\varepsilon_t) > 0$ and $q = 1 - p$, then the distribution of γ_{t-1} in (1.27) is

$$P(\gamma_{t-1} = \gamma) = q^\gamma p + qp^\gamma, \quad \gamma \geq 1. \tag{1.30}$$

Since $\operatorname{sign}(e_{t-i}) = \operatorname{sign}(\varepsilon_{t-i})$ for any t and i, then e_{t-i} and ε_{t-i} are interchangable in the previous three equations, meaning the focus may be directed exclusively to the sign, and subsequently on the underlying distribution, of the random innovations ε_t.

In conclusion, this model competes with the EGARCH model – the exponential term in (1.27) is always positive (and if $\theta = 0$, the GARCH(1,1) model is obtained), both models behave similarly, but OGARCH adds the effect of increasingly longer periods spent in positive or negative shocks.

1.3 Multiplicative Error Models (MEMs)

Multiplicative error models have been studied for more than twenty five years. The first appearance can be traced back to Engle and Russell [1998], where a "new statistical model for the analysis of data which arrive at irregular intervals" was proposed. It focused on the expected duration between events and was called the Autoregressive Conditional Duration (ACD) model.

It proved useful in many scenarios and soon a new class of models, MEMs, was established within the general framework set by the ACD models. Their main feature is that they are (usually) non-negative by definition, which makes them well suited to the purpose of this paper. However, unlike the GARCH models, the notation and formulations of the model are not as uniform and, apart from the most general form, there are several alternative structures of the model in the literature, hidden behind the same name of MEM.

Therefore, we will first define the most general form of MEM, briefly discuss its properties, and then focus on some specific modifications. The following text is inspired by Engle [2002], Engle and Gallo [2006], Brownlees, Cipollini, and Gallo [2011] and Cipollini and Gallo [2022].

1.3.1 General Formulation of MEM

Definition 2. *Let us consider a non-negative time series $\{y_t\}_{t\in\mathbb{N}}$, a sequence of iid random variables $\{\varepsilon_t\}_{t\in\mathbb{N}}$, and let us denote by Ω_{t-1} the set of information available at time $t-1$. Then the process $\{y_t\}$ follows a MEM if we may express it as*

$$y_t = \mu_t \varepsilon_t. \tag{1.31}$$

It is assumed that

$$\mu_t|\Omega_{t-1} = \mu(\boldsymbol{\theta}, \Omega_{t-1}), \tag{1.32}$$

$$\varepsilon_t|\Omega_{t-1} \sim D^+(1, \sigma^2), \tag{1.33}$$

where $D^+(\mu, \sigma^2)$ *denotes a distribution with mean* μ*, variance* σ^2 *and probability density function defined over a non-negative support.*

Remark 1. The specification (1.32) means that, conditionally on Ω_{t-1}, μ_t is deterministically given by the known information and a vector of parameters θ. It follows that

$$E(x_t|\Omega_{t-1}) = \mu_t \cdot 1 = \mu_t, \tag{1.34}$$

$$\text{Var}(x_t|\Omega_{t-1}) = \text{Var}(\mu_t, \varepsilon_t|\Omega_{t-1}) = \mu_t^2 \, \text{Var}(\varepsilon_t|\Omega_{t-1}) = \mu_t^2\sigma^2. \tag{1.35}$$

We can see that MEM is generally defined as the product of a scale factor, which can evolve in different ways depending on the exact specification of μ_t in (1.32), and of an error term with unit mean. As noted in Remark 1, the scale factor represents the conditional expectation of the process. Its specification, together with the specification of the conditional distribution of ε_t, allows for tailoring the model to be adapted to the given situation.

In the following, the assumptions will become more concrete. It has been agreed in the literature that one of the most useful yet not too complex modifications of MEM is such that its mean, as defined in (1.32), resembles a GARCH process, usually of the first orders, and that the underlying conditional distribution of ε_t is from the family of gamma densities, similarly to the original case of the ACD models proposed by Engle and Russell [1998].[3] These parametrizations are now going to be explored in more detail.

1.3.2 Properties and Estimation of MEM

Considering firstly the mean value μ_t, it is convenient to specify it conditionally on Ω_{t-1} as

$$\mu_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_1 \mu_{t-1} \tag{1.36}$$

while assuming $\alpha_1 + \beta_1 < 1$ to ensure weak stationarity of the process (as discussed in Engle [2002]). Further restrictions on the parameters may be imposed to have fully non-negative values or other required properties. This model, proposed e.g. by Brownlees, Cipollini, and Gallo [2011], can be called as baseline MEM.[4]

Supposing, additionally, that the process $\{y_t\}$ is mean-stationary, meaning the unconditional mean is not dependent on time, i.e. $E(y_t) = E(\mu_t) = \mu$, it follows then from (1.36) after taking expectation of both sides that

$$\alpha_0 = \mu(1 - \alpha_1 - \beta_1). \tag{1.37}$$

[3]Despite the fact that some works, e.g. Engle [2002], elaborate on the case of exponential distribution or other specifically defined densities, the upcoming section tries to provide a brief insight into the more general case.

[4]This parametrization allows for further modifications. For instance, adding $\gamma_1 x_{t-1} \mathbb{1}_{\{x_t<0\}}$ to (1.36) indicates incorporation of possible assymetry in case of negative observations.

Therefore, there is one less parameter to be estimated because α_0 can be calculated via (1.37) after obtaining α_1 and β_1 and after estimating μ by the unconditional (sample) mean $\bar{y}$.

Secondly, the distribution of ε_t has still not been specified. In some cases and applications, it is not even necessary, since some inference can be done without it. Most often the (quasi) maximum likelihood is used in the MEM framework. For instance, Brownlees, Cipollini, and Gallo [2011] express the conditional probability distribution function (pdf) of ε_t by $f_\varepsilon(\varepsilon_t|\Omega_{t-1})$. It follows that

$$f_x(x_t|\Omega_{t-1}) = f_\varepsilon\left(x_t\mu_t^{-1}|\Omega_{t-1}\right)\mu_t^{-1} \tag{1.38}$$

and that the log-likelihood function used in the maximization process is

$$l_T = \sum_{t=1}^{T} l_t = \sum_{t=1}^{T}\left(\ln\left[f_\varepsilon(\varepsilon_t|\Omega_{t-1})\right] + \ln[\varepsilon_t] - \ln[y_t]\right). \tag{1.39}$$

However, it can be beneficial to restrict ourselves to a more specific distribution. Cipollini and Gallo [2022], Engle and Gallo [2006] and Engle [2002] suggest $Gamma(\phi, \phi)$ as the conditional distribution of ε_t. In that case, $E(\varepsilon_t|\Omega_{t-1}) = 1$ (unit mean distribution) and $\text{Var}(\varepsilon_t|\Omega_{t-1}) = \sigma^2 = 1/\phi$. This implies

$$f(y_t|\Omega_{t-1}) = \Gamma(\phi)^{-1}\phi^\phi y_t^{\phi-1}\exp\left(\frac{-\phi y_t}{\mu_t}\right). \tag{1.40}$$

Should the goal now be to estimate the parameters defining μ_t only, the simplified log-likelihood would be of the form

$$l = C - \phi\sum_{t-1}^{T}\left(\ln[\mu_t] + y_t\mu_t^{-1}\right). \tag{1.41}$$

The constant $C \in \mathbb{R}$ can be omitted in the maximizing procedure and the variable ϕ is also not relevant; hence, the first order conditions must satisfy

$$\sum_{t=1}^{T}\left(\frac{y_t - \mu_t}{\mu_t^2}\right)\frac{\partial\mu_t}{\partial\theta}, \tag{1.42}$$

where θ is the vector of parameters needed to estimate μ_t. For instance, in model given by (1.36), $\theta = (\alpha_0,\, \alpha_1,\, \beta_1)^T$.

Moreover, Engle [2002] proposes a so-called $(p,\, q)$ mean specification for this model, namely

$$\mu_t = \alpha_0 + \sum_{j=1}^{p}\alpha_j y_{t-j} + \sum_{j=1}^{q}\beta_j\mu_{t-j} + \gamma^T z_t, \tag{1.43}$$

where z_t is a $k \times 1$ vector of predetermined variables. This specification can be considered a generalization of (1.36). Under the condition that

$$\sum_{j=1}^{p}\alpha_j + \sum_{j=1}^{q}\beta_j < 1,$$

Engle [2002] claims that "Corrollary to Lee and Hansen [1994]", which was first proposed in Engle and Russell [1998], can be utilized to show that "even mildly explosive models may be estimated consistently by QMLE".

2. Getting Dense with Zeros

Building upon the established challenge of dense zeros in time series data, this chapter dives into the heart of our proposed solutions. We present five novel models specifically tailored to effectively model these zero-inflated series. Each model adopts a distinct approach to address the presence of excess zeros while simultaneously capturing the dependence structure within the non-zero observations.

In the following sections, each model will be rigorously examined. We will begin by formally defining the model structure, followed by a brief exploration of its theoretical properties. Next, we will outline the estimation procedures necessary to fit the model to a given time series. Finally, suitable methods for generating forecasts using the estimated model are to be described, allowing for informed predictions into the future behavior of the series despite the prevalence of zeros.

2.1 DZ-EGARCH

Let us name the first model as DZ-EGARCH, where DZ stands for "dense zeros" modification. It is based on the model 1.26. The main goal of this model will be to incorporate possible assymetry and leverage. It is expected it will react not only to the change of the sign of the disturbances, but also to the transition between zeros and the following spike of positive values.

With all of the above, let us define two slightly different versions of this type called DZ-EGARCH1 and DZ-EGARCH2.

(i) DZ-EGARCH1:

$$\begin{aligned} Y_t &= \sigma_t \varepsilon_t \\ \ln(\sigma_t^2) &= \alpha_0 + \alpha_1 \frac{Y_{t-1} + \gamma Y_{t-1}}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2) \\ &= \alpha_0 + \alpha_1 \frac{Y_{t-1}(1+\gamma)}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2); \end{aligned} \tag{2.1}$$

(ii) DZ-EGARCH2:

$$\begin{aligned} Y_t &= \sigma_t \varepsilon_t \\ \ln(\sigma_t^2) &= \alpha_0 + \alpha_1 \frac{Y_{t-1} + \gamma Y_{t-1} \mathbb{1}_{\{Y_{t-2}>0\}}}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2) \\ &= \alpha_0 + \alpha_1 \frac{Y_{t-1}(1+\gamma \mathbb{1}_{\{Y_{t-2}>0\}})}{\sigma_{t-1}} + \beta \ln(\sigma_{t-1}^2) \end{aligned} \tag{2.2}$$

assuming in both cases that α_0, α_1, $\beta > 0$, $\alpha_1 + \beta < 1$ and $\gamma > -1$.

The first proposed version closely aligns with the classical definition established earlier. In contrast, the second version incorporates the tuning parameter

γ only when the observation preceding the penultimate observation (at time $t-2$) is positive. This approach balances the benefit of incorporating information from the more distant past with maintaining a parsimonious model structure by summarizing this past information into a single constant term applicable across all time steps t. Furthermore, a resemblance can be drawn to the OGARCH model, as both incorporate a term that reacts solely to non-zero historical values, essentially capturing the impact of a recent shock. The underlying assumption is that the DZ-EGARCH2 model will exhibit superior performance compared to the first version.

Following this exposition, a detailed examination of the parameterization employed in both proposed versions is now to be clarified. For ε_t, suppose it is defined as

$$\varepsilon_t = \nu_t b_t. \tag{2.3}$$

Here, $\{\nu_t\}_{t\in\mathbb{Z}}$ is an iid sequence and every ν_t follows a half normal distribution, meaning that $\nu_t = |U|$ where $U \sim N\left(0, \frac{1}{1-p}\right)$. Secondly, $\{b_t\}_{t\in\mathbb{Z}}$ is a Bernoulli variable, with alternative (Bernoulli) distribution with probability of success $1-p$, so that $b_t \sim Be(1-p)$, $p \in (0,1)$. This implies $P(\varepsilon_t = 0) = P(b_t = 0) = p$.[1]

It then follows that

$$f_\varepsilon(y) = p\mathbb{1}_{\{y=0\}} + (1-p)f_{\varepsilon|\varepsilon>0}(y)\mathbb{1}_{\{y>0\}} \tag{2.4}$$

and the density is with respect to measure $\lambda_+ + \delta_0$, where λ_+ is Lebesgue measure on $(0,\infty)$ and δ_0 is Dirac measure at 0. In our particular case it holds

$$f_{\varepsilon|\varepsilon>0}(y) = \sqrt{\frac{2(1-p)}{\pi}} \exp\left\{-\frac{1}{2}(1-p)y^2\right\} \tag{2.5}$$

for $y > 0$. Moreover, $E\nu_t^2 = EU^2 = \frac{1}{1-p}$ and consequently

$$E\varepsilon_t^2 = 1. \tag{2.6}$$

2.1.1 Estimation

Let us denote by $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta, \gamma)^T$ the vector of unknown parameters in the "volatility equations" (2.1) and (2.2). Working with unconditional density would be laborous and even unnecessary because we work with time series where we can observe past values (past information). Thus, it is not difficult to obtain the conditional density of Y_t given the past information included in observations $\{y_{t-1}, y_{t-2} \dots y_1\}$ as

$$f_{Y_t|y_{t-1},\dots,y_1}(y_t, \boldsymbol{\theta}) = [p]^{\mathbb{1}_{\{y_t=0\}}} \left[\frac{1-p}{\sigma_t} f_{\varepsilon|\varepsilon>0}\left(\frac{y_t}{\sigma_t}\right)\right]^{\mathbb{1}_{\{y_t>0\}}}. \tag{2.7}$$

[1] This setup is almost identical to the one with the so-called *gray noise* presented in Hanousek [2022]. The difference lies in the direct specification of half normal distribution which makes the framework cohere with related literature.

The corresponding conditional likelihood of the (not necessarily fully observed) sample of M values denoted as $\boldsymbol{y}_M$ is

$$\begin{aligned} L(\boldsymbol{y}_M, \boldsymbol{\theta}) &= \prod_{t=1}^{M} f_{Y_t|y_{t-1},\ldots,y_1}(y_t, \boldsymbol{\theta}) \\ &= \prod_{t=1}^{M} [p]^{\mathbb{1}_{\{y_t=0\}}} \left[\frac{1-p}{\sigma_t} f_{\varepsilon|\varepsilon>0}\left(\frac{y_t}{\sigma_t}\right)\right]^{\mathbb{1}_{\{y_t>0\}}} . \end{aligned} \tag{2.8}$$

Taking the logarithm and substituting (2.5) yields the conditional log likelihood. Its maximization is equivalent to solving

$$\min_{\boldsymbol{\theta}} \sum_{t=1}^{M} \left(\frac{y_t^2}{\tilde{\sigma}_t^2} + \log[\tilde{\sigma}_t^2]\mathbb{1}_{\{y_t>0\}} \right). \tag{2.9}$$

Here, for $t \geq 1$ and the version DZ-EGARCH1,

$$\ln(\tilde{\sigma}_t^2) = \alpha_0 + \alpha_1 \frac{y_{t-1}(1+\gamma)}{\tilde{\sigma}_{t-1}} + \beta \ln(\tilde{\sigma}_{t-1}^2) \tag{2.10}$$

while setting $\tilde{\sigma}_0 = y_0 = y_1$.

In case of the version DZ-EGARCH2,

$$\ln(\tilde{\sigma}_t^2) = \alpha_0 + \alpha_1 \frac{y_{t-1}(1+\gamma \mathbb{1}_{y_{t-2}>0})}{\tilde{\sigma}_{t-1}} + \beta \ln(\tilde{\sigma}_{t-1}^2) \tag{2.11}$$

while setting $\tilde{\sigma}_0 = y_0 = y_1$ and disabling the term with the indicator for $t \in \{1, 2\}$.

The resulting arguments of the minima $\hat{\alpha_0}$, $\hat{\alpha_1}$, $\hat{\beta}$, $\hat{\gamma}$ are the corresponding quasi-maximum likelihood estimators of the model's respective parameters.

To obtain the estimate $\hat{p}$ of p, the proportion of (true) zeros in the observed data can be calculated. Hence,

$$\hat{p} = \frac{1}{M} \sum_{t=1}^{M} \mathbb{1}_{\{y_t=0\}}. \tag{2.12}$$

According to the general theory discussed in the literature on quasi-maximum likelihood, especially in Hudecová and Pešta [2024] which focuses on very similar scenario as presented above, the above described estimates are meaningful and theoretically justifiable. Moreover, under additional assumptions, also asymptotically normal. Therefore, they should provide a good basis for the following predictions and inference.

2.1.2 Consistency of the Estimates

Nevertheless it is important to provide at least a brief proof that the estimates are reasonable, at least asymptotically. One of the most crucial properties is consistency.[2]

[2] In extension, another useful property of an estimate is asymptotic normality, useful to validate e.g. Wald tests of parameters. However, in this thesis, such methods are not utilised and therefore further properties of our estimates are not explored.

Consistency of the QML estimates of GARCH processes is approached in numerous ways in the literature. Fundamentals were established already by Lee and Hansen [1994] or Lumsdaine [1996]. Subsequent results were then published by and valuable sources of information have been found in Berkes, Horvath, and Kokoszka [2003] and Francq and Zakoïan [2004], as well as Preminger and Storti [2014] which explored the ordinary GARCH(1,1).

We begin with concepts appearing in the work of Preminger and Storti [2014]. The model in our case satisfies the following conditions stated in the mentioned paper:

Theorem 1 (Consistency of the QML Estimates)**.** *Consider a time series model of the EGARCH family. Suppose the process of innovations $\{\varepsilon_t\}_{t\in T}$ follows the structure as in (2.3). Define the following set of assumptions:*

(A1) The vector of true parameters $\boldsymbol{\theta}$ belongs to a compact set

$$\Theta \equiv \{\alpha_0 > 0,\, \alpha_1 > 0,\, \beta > 0,\, \gamma > -1,\, \alpha_1 + \beta < 1\};$$

(A2) $E[\ln(\alpha_0 \varepsilon_t^2 + \beta)] < 0$;

(A3) $E|\varepsilon_t|^{2s} < \infty$ *for some* $s > 0$;

(A4) $\lim_{r\to 0} r^{-(1+v)} P(\varepsilon_t^2 \leq r) < \infty$ *for some* $v > 0$.

Then

(i) if the considered model satisfies the assumptions (A1)-(A4), the QML estimate $\hat{\boldsymbol{\theta}}_M$ of $\boldsymbol{\theta}$ based on a sample of M observations is consistent, i.e. $\hat{\boldsymbol{\theta}}_M \xrightarrow{a.s.} \boldsymbol{\theta}$ for $M \to \infty$;

(ii) models DZ-EGARCH1 and DZ-EGARCH2 as defined in (2.1) and (2.2), respectively, satisfy the assumptions (A1)-(A4) and thus the estimates obtained by the optimization problem (2.9) are consistent.

Proof. Let us begin by examining the assumptions (A1)-(A4). These assumptions closely resemble those in the research paper of Silvennoinen and Teräsvirta [2021]. They ensure that the parameters do not explode and are sufficient conditions for consistency. The assumptions can be easily verified. For example, (A2) is true because the model supposes $\alpha_1 + \beta < 1$ and we have shown (2.6); a sufficient condition for (A4) is that the density of innovations is bounded, which is also the case, because the distribution is bounded by zero on the left side and tends to zero (asymptotically) on the other.

Further, let us focus now on the "non-zero" part of the log likelihood derived from (2.8) and denote

$$l_M(\boldsymbol{y}_M, \tilde{\boldsymbol{\theta}}_M) = -\frac{1}{2}\sum_{t=1}^{M}\left(\frac{y_t^2}{\tilde{\sigma}_t^2} + \log[\tilde{\sigma}_t^2]\mathbb{1}_{\{y_t>0\}}\right). \tag{2.13}$$

Therefore, the situation is equivalent to the one described in Berkes, Horvath, and Kokoszka [2003], in particular in Lemmas 5.4 and 5.5 and Theorems 4.1 and 4.4, while bearing in mind the dense-zeros modification which has been incorporated by the modified log likelihood inspired by Hudecová and Pešta [2024]. It follows that as the number of training observations approaches the infinity, the corresponding sample and theoretical log likelihoods tend to each other; moreover, the log likelihood function has only one extreme among all possible estimates $\hat{\boldsymbol{\theta}}$ which is $\boldsymbol{\theta}$. Hence, for the parameter estimates obtained by QMLE from the observed sample of M observations it holds that $\hat{\boldsymbol{\theta}}_M \xrightarrow{a.s.} \boldsymbol{\theta}$ as M tends to infinity. □

2.1.3 Prediction

An outline of the prediction procedure based on the observed sample $y_1, \ldots, y_M$. The goal is to predict $\hat{y}_{M+h}$, $h = 2, \ldots, H$. Prediction of $\hat{y}_{M+1}$ is trivial because it follows straight from the definition of the model.

(i) Estimate $\hat{p} = \frac{1}{M}\sum_{t=1}^{M} \mathbb{1}_{\{y_t=0\}}$. This is the only parameter needed to be estimated in order to create sequences of b_t and ν_t.

(ii) Create the sequence $\{\hat{b}_{M+h}\}_{h=M+2}^{H}$.

(iii) It is now sufficient to create only those predictions of $\hat{\nu}_{M+h}$ for which the corresponding $\hat{b}_{M+h} \neq 0$. Create such sequence using the estimated distribution $N\left(0, \frac{1}{1-\hat{p}}\right)$.

(iv) Combine the estimated sequences to create $\{\hat{\varepsilon}_{M+h}\}_{h=M+2}^{H}$.

(v) Using only the positive cases of $y_1, \ldots, y_M$, estimate $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}, \hat{\gamma})^T$ as outlined in (2.9), (2.10) and (2.11).

(vi) Create a sequence $\{\hat{\sigma}_{M+h}\}_{h=M+2}^{H}$.

(vii) Create the desired sequence of predicted values $\{\hat{y}_{M+h}\}_{h=M+2}^{H}$.

2.2 DZ-OGARCH

The second model, DZ-OGARCH, is inspired by OGARCH as defined in (1.27). Its original purpose is to incorporate information about spells of shocks. Because in our scenario there cannot be negative values, it will then reflect the length of the time interval since the last zero observation.

Hence, the definition of DZ-OGARCH is

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \exp(\delta\gamma_{t-1}) y_{t-1} + \beta\sigma_{t-1}, \end{aligned} \tag{2.14}$$

where

$$\gamma_{t-1} = \begin{cases} 1 & \text{if } y_{t-1} = 0, \\ \gamma_{t-2} + 1 & \text{if } y_{t-1} > 0. \end{cases} \tag{2.15}$$

The innovations ε_t are defined as in (2.3) and the restrictions for the parameter values are that $\alpha_0 > 0$, $\alpha_1 > 0$, $\beta > 0$ and $\alpha_1 + \beta < 1$.

2.2.1 Estimation

Since the underlying process is still a modification of GARCH framework with half-normal "dense zeros" modification of distribution of the innovation process, we may again proceed similarly as in the DZ-EGARCH case with quasi-maximum likelihood.

The final minimization problem can be formulated as

$$\min_{\boldsymbol{\theta}} \sum_{t=1}^{M} \left(\frac{y_t^2}{\tilde{\sigma}_t^2} + \log[\tilde{\sigma}_t^2] \mathbb{1}_{\{y_t > 0\}} \right). \tag{2.16}$$

Here, for $t \geq 1$,

$$\tilde{\sigma}_t^2 = \alpha_0 + \alpha_1 \exp(\delta \gamma_{t-1}) y_{t-1} + \beta \tilde{\sigma}_{t-1} \tag{2.17}$$

while setting $\tilde{\sigma}_0 = y_0 = y_1$.

Naturally, the resulting estimates may be influenced by the changes in the definition of the model equations (2.14) and (2.15). Nevertheless, the aforementioned (asymptotic) properties of the estimates will be assumed to hold and possible deviations will be discussed in the third part of this text.

2.2.2 Prediction

An outline of the prediction procedure based on the observed sample $y_1, \ldots, y_M$. The goal is to predict $\hat{y}_{M+h}$, $h = 2, \ldots, H$.

(i) Estimate $\hat{p} = \frac{1}{M} \sum_{t=1}^{M} \mathbb{1}_{\{y_t = 0\}}$. This is the only parameter needed to be estimated in order to create sequences of b_t and ν_t.

(ii) Create the sequence $\{\hat{b}_{M+h}\}_{h=M+2}^{H}$.

(iii) It is now sufficient to create only those predictions of $\hat{\nu}_{M+h}$ for which the corresponding $\hat{b}_{M+h} \neq 0$. Create such sequence using the estimated distribution $N\left(0, \frac{1}{1-\hat{p}}\right)$.

(iv) Combine the estimated sequences to create $\{\hat{\varepsilon}_{M+h}\}_{h=M+2}^{H}$.

(v) Using only the positive cases of $y_1, \ldots, y_M$, estimate $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}, \hat{\delta})^T$ as outlined in (2.16) and (2.17).

(vi) Create a sequence $\{\hat{\sigma}_{M+h}\}_{h=M+2}^{H}$.

(vii) Create the desired sequence of predicted values $\{\hat{y}_{M+h}\}_{h=M+2}^{H}$.

2.3 DZ-GMEM

As the third candidate, a modification of a baseline MEM is being proposed. Having the condition of non-negativity already included in the original definition, it is important to complete the DZ condition by ensuring a non-trivial possibility of producing zero values. To achieve that, in this case a Gamma version of the innovation process ε_t^G will be applied.

The DZ-GMEM takes the form of

$$\begin{aligned} y_t &= \mu_t \varepsilon_t^G \\ \mu_t &= \alpha_0 + \alpha_1 y_{t-1} + \beta \mu_{t-1} \end{aligned} \tag{2.18}$$

where $\alpha_0 > 0$, $\alpha_1 > 0$, $\beta > 0$ and $\alpha_1 + \beta < 1$. We define the process

$$\varepsilon_t^G = \xi_t b_t, \tag{2.19}$$

where a Gamma distribution of $\xi_t \sim \Gamma(a_1, a_2) = \Gamma\left(1 - p, \frac{1}{(1-p)^2}\right)$ is assumed. Since in literature there is usually the case of exponential distribution discussed, we decide to extend it to a more general Gamma distribution. The parametrization is chosen so that similar ideas as in the previous framework with half-normal distribution can be applied.[3]

2.3.1 Estimation

Again, it is assumed (and will be validated in the following chapter) that the already mentioned general (asymptotic) theory, as proposed in Hudecová and Pešta [2024], still holds. Then it can be shown that by using quasi-MLE approach, the problem of maximizing quasi-maximum likelihood is equivalent to minimizing

$$\min_{\boldsymbol{\theta}} \sum_{t=1}^{M} \left(\frac{y_t^2}{\tilde{\mu}_t^2} + \log[\tilde{\mu}_t^2] \mathbb{1}_{\{y_t > 0\}} \right). \tag{2.20}$$

Here, for $t \geq 1$,

$$\tilde{\mu}_t = \alpha_0 + \alpha_1 y_{t-1} + \beta \tilde{\mu}_{t-1} \tag{2.21}$$

while setting $\tilde{\mu}_0 = y_0 = y_1$.

2.3.2 Prediction

An outline of the prediction procedure based on the observed sample $y_1, \ldots, y_M$. The goal is to predict $\hat{y}_{M+h}$, $h = 2, \ldots, H$.

(i) Estimate $\hat{p} = \frac{1}{M} \sum_{t=1}^{M} \mathbb{1}_{\{y_t = 0\}}$. This is the only parameter needed to be estimated in order to create sequences of b_t and ξ_t.

(ii) Create the sequence $\{\hat{b}_{M+h}\}_{h=M+2}^{H}$.

[3] The parametrization implies that $E(\xi_t) = a_1 a_2 = \frac{1}{1-p}$.

(iii) It is now sufficient to create only those predictions of $\hat{\xi}_{M+h}$ for which the corresponding $\hat{b}_{M+h} \neq 0$. Create such sequence using the estimated distribution $\Gamma\left(1-\hat{p},\, \frac{1}{(1-\hat{p})^2}\right)$.

(iv) Combine the estimated sequences to create $\{\hat{\varepsilon}^G_{M+h}\}_{h=M+2}^{H}$.

(v) Using only the positive cases of $y_1, \ldots, y_M$, estimate $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta})^T$ as outlined in (2.20) and (2.21).

(vi) Create a sequence $\{\hat{\mu}_{M+h}\}_{h=M+2}^{H}$.

(vii) Create the desired sequence of predicted values $\{\hat{y}_{M+h}\}_{h=M+2}^{H}$.

2.4 DZ-NMEM21

The last member on the candidate list is DZ-NMEM21 model. Similarly as in the previous case, the original model is a multiplicative error model. However, this time it includes one more term looking one step farther into history – analogously as in case of DZ-EGARCH2, the idea is to present another version of a MEM but with a bit more complex structure. Moreover, not the Gamma innovations ε_t^G, but the Normal innovations ε_t as defined in (2.3) are implemented.

The DZ-NMEM21 is therefore defined as

$$
\begin{aligned}
y_t &= \mu_t \varepsilon_t \\
\mu_t &= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta \mu_{t-1}
\end{aligned}
\tag{2.22}
$$

where $\alpha_0 > 0$, $\alpha_1 > 0$, $\alpha_2 > 0$, $\beta > 0$ and $\alpha_1 + \alpha_2 + \beta < 1$.

2.4.1 Estimation

Again, it is assumed (and will be validated in the following chapter) that the already mentioned general (asymptotic) theory, as proposed in Hudecová and Pešta [2024], still holds. Then it can be shown that by using quasi-MLE approach, the problem of maximizing quasi-maximum likelihood is equivalent to minimizing

$$
\min_{\boldsymbol{\theta}} \sum_{t=1}^{M} \left(\frac{y_t^2}{\tilde{\mu}_t^2} + \log[\tilde{\mu}_t^2] \mathbb{1}_{\{y_t>0\}} \right). \tag{2.23}
$$

Here, for $t \geq 1$,

$$
\tilde{\mu}_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta \tilde{\mu}_{t-1} \tag{2.24}
$$

while setting $\tilde{\mu}_0 = y_0 = y_1$.

2.4.2 Prediction

An outline of the prediction procedure based on the observed sample $y_1, \dots, y_M$. The goal is to predict $\hat{y}_{M+h}$, $h = 2, \dots, H$.

(i) Estimate $\hat{p} = \frac{1}{M}\sum_{t=1}^{M} \mathbb{1}_{\{y_t=0\}}$. This is the only parameter needed to be estimated in order to create sequences of b_t and ν_t.

(ii) Create the sequence $\{\hat{b}_{M+h}\}_{h=M+2}^{H}$.

(iii) It is now sufficient to create only those predictions of $\hat{\nu}_{M+h}$ for which the corresponding $\hat{b}_{M+h} \neq 0$. Create such sequence using the estimated distribution $N\left(0, \frac{1}{1-\hat{p}}\right)$.

(iv) Combine the estimated sequences to create $\{\hat{\varepsilon}_{M+h}\}_{h=M+2}^{H}$.

(v) Using only the positive cases of $y_1, \dots, y_M$, estimate $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta})^T$ as outlined in (2.23) and (2.24).

(vi) Create a sequence $\{\hat{\sigma}_{M+h}\}_{h=M+2}^{H}$.

(vii) Create the desired sequence of predicted values $\{\hat{y}_{M+h}\}_{h=M+2}^{H}$.

---

3. Practical Part

The third part of this text aims to apply the newly created models to a real-world scenario. The task is to collect data with the dense zeros property, train and tune the models based on the data, and then evaluate their performance using various metrics. The chapter concludes with a discussion of the results, the parameter estimation methods and the overall performance of the proposed models.

3.1 Meeting the Dataset

The dataset used to evaluate the models consists of uninsured material damage claims to cars provided by the Czech Insurance Bureau. It contains daily observations of such claims from the 1st of January 2015 to the 31st of December 2019, a total of 1826 observed non-negative values.

The dataset is divided into two parts, the out-of-time observations and the in-time observations. The latter consist of data between the 1st of March 2015 and the 31st of May 2019, while the rest of the data belong to the out-of-time sample, which is discarded for the training and estimation procedures in order to reduce possible biases and/or seasonality related to the beginning and end of the analysed time series.

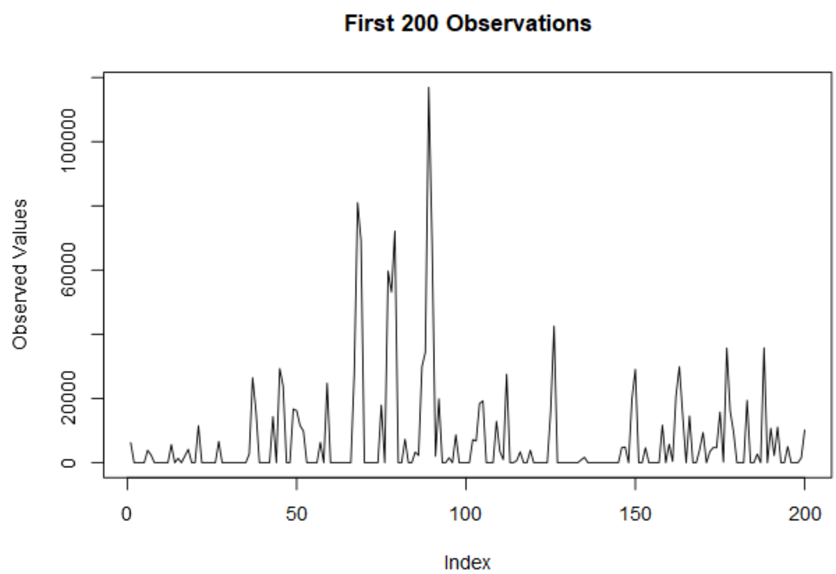

Figure 3.1: First 200 Values of the In-Time Dataset

Figure 3.1 provides an initial glimpse into the characteristics of the time series data. The figure reveals a consistent pattern throughout the data: a high density of zero values, a concentration of positive observations below a threshold of 45,000, and occasional outliers exceeding double this value. Numerical characteristics of the complete in-time sample (positive values only) can be found in Table 3.1.

Ideally, the proposed models should exhibit similar behavior by capturing this inherent data distribution. An overly conservative model, lacking the ability to capture these extreme spikes, would inadequately forecast low-probability, high-impact loss scenarios, potentially leading to under-reserves and insufficient risk preparedness. Conversely, an excessively dynamic model could generate overly pessimistic predictions, resulting in over-reserving and potential profit loss due to unnecessarily high reserve allocations.

5th quant.	25th quant.	Median	Mean	75th quant.	95th quant.
$7.49 \cdot 10^2$	$3.73 \cdot 10^3$	$9.13 \cdot 10^3$	$1.48 \cdot 10^4$	$2.06 \cdot 10^4$	$4.48 \cdot 10^4$

Table 3.1: Chosen Quantiles and Mean of the True Positive Observations

It should also be noted that the time series exhibit no signs of nonstationarity. Using the Augmented Dickey-Fuller (ADF) test, which tests the null hypothesis H_0 that there is a unit root in the time series against the alternative hypothesis H_A that the time series is stationary (or that all roots are less than one in the complex plane), we prove that the time series is indeed stationary at an a priori set significance level $\alpha = 0.05$. Thus, it is reasonable to proceed with parameter estimation.

3.2 Methodology

The computational methods used for estimation, prediction and evaluation need to be suited for working with time series and with models with random elements. There are 1553 observations available (after discarding the out-of-time sample), which is enough for asymptotic results to hold and for implementation of splitting methods used in evaluation of the models.

As the main evaluation procedure, a growing-window forward-validation with $K = 6$ splits has been chosen. It is a modification of the classical K-fold cross-validation which divides the dataset into K splits. In each k-th iteration, where $k = 1, \dots, K-1$, the first k splits are used for training and the following $(k+1)$-st split is used as a test split (against which the current predictions are compared). There is enough empirical evidence to suggest this method is better suited for dealing with various types of time series than the ordinary K-fold cross-validation, see for example Schnaubelt [2019]. For each iteration, the corresponding estimates of $\hat{p}$ and of the vector of parameters $\hat{\boldsymbol{\theta}}$ are calculated using a fitting optimization procedure.[1]

Secondly, in every k-th iteration of the growing-window forward-validation procedure, a total of $J = 1000$ realizations of the forecast innovation sequences $\{\hat{\varepsilon}_{M+h}\}_{h=M+2}^{H}$ and $\{\hat{\varepsilon}^G_{M+h}\}_{h=M+2}^{H}$, respectively, have been created.

It has been set that $M = \lfloor 1553/K \rfloor = 258$ denotes the size of each split and $H = M = 258$ denotes the number of predictions computed for each j-th realization of the innovation sequence where $j = 1, \dots, J$.

It follows that there have been calculated $(K-1) \cdot J = 5000$ total simulations for each of the proposed models. In order to compare them to each other, several criteria have been set and evaluated for every model separately, namely

[1] Constrained optimization algorithm with Nelder-Mead optimization method has been chosen to complete this task.

(i) the mean of mean absolute deviations (MAE's) over all realizations,[2]

(ii) the mean sum of predicted observations over all realizations,

(iii) the 5th percentile, 25th percentile, median, 75th percentile and 95th percentile over all realizations.

Values from (ii) and (iii) are to be compared with their counterparts computed from the original dataset.

The main results are summarized in the following section.

3.3 Results and Comparison

This section presents an evaluation of the models' performance and a comparative analysis. Key numerical metrics and graphical illustrations are included. Figure 3.2 depicts, for each of the candidate models, a single realization of predicted data for the final split based on information from the initial five splits (refer to the Methodology section for details). These may be compared to the true observed reality in the last split as presented in the right downmost picture in the same figure.

	EGARCH1	EGARCH2	OGARCH	GMEM	NMEM21
Mean MAE	23 665	11 824	$6.05 \cdot 10^9$	119 041	10 848
Mean sum	$5.55 \cdot 10^6$	$2.28 \cdot 10^6$	$1.56 \cdot 10^{12}$	$3.00 \cdot 10^7$	$1.85 \cdot 10^6$
5th perc.	$2.93 \cdot 10^3$	$1.60 \cdot 10^3$	$3.96 \cdot 10^0$	$1.30 \cdot 10^2$	$6.76 \cdot 10^2$
25th perc.	$1.50 \cdot 10^4$	$8.06 \cdot 10^3$	$2.04 \cdot 10^1$	$5.74 \cdot 10^3$	$3.66 \cdot 10^3$
Median	$3.30 \cdot 10^4$	$1.71 \cdot 10^4$	$4.59 \cdot 10^1$	$3.34 \cdot 10^4$	$9.26 \cdot 10^3$
75th perc.	$6.24 \cdot 10^4$	$2.96 \cdot 10^4$	$9.97 \cdot 10^1$	$1.28 \cdot 10^5$	$2.09 \cdot 10^4$
95th perc.	$1.51 \cdot 10^5$	$5.19 \cdot 10^4$	$1.44 \cdot 10^3$	$7.54 \cdot 10^5$	$5.60 \cdot 10^4$

Table 3.2: Comparison of the Models

Our analysis begins with Table 3.2. Although all models possess a sound theoretical foundation, their suitability for handling dense zeros varies.

DZ-OGARCH demonstrates the lowest performance across all evaluation criteria. Notably, the model consistently generates very low values, with even the 75th percentile remaining low. Furthermore, the 95th percentile fails to reach significant heights. Despite this, both its Mean Absolute Error (MAE) and mean sum are unexpectedly high. This discrepancy can be attributed to the exponential function inherent to the model's definition. As the window of a shock event

[2]MAE was chosen as the evaluation metric because it does not penalize big differences between true and predicted values as much as MSE or RMSE and it works even with "true zeros" in observed data as opposed to e.g. MAPE.

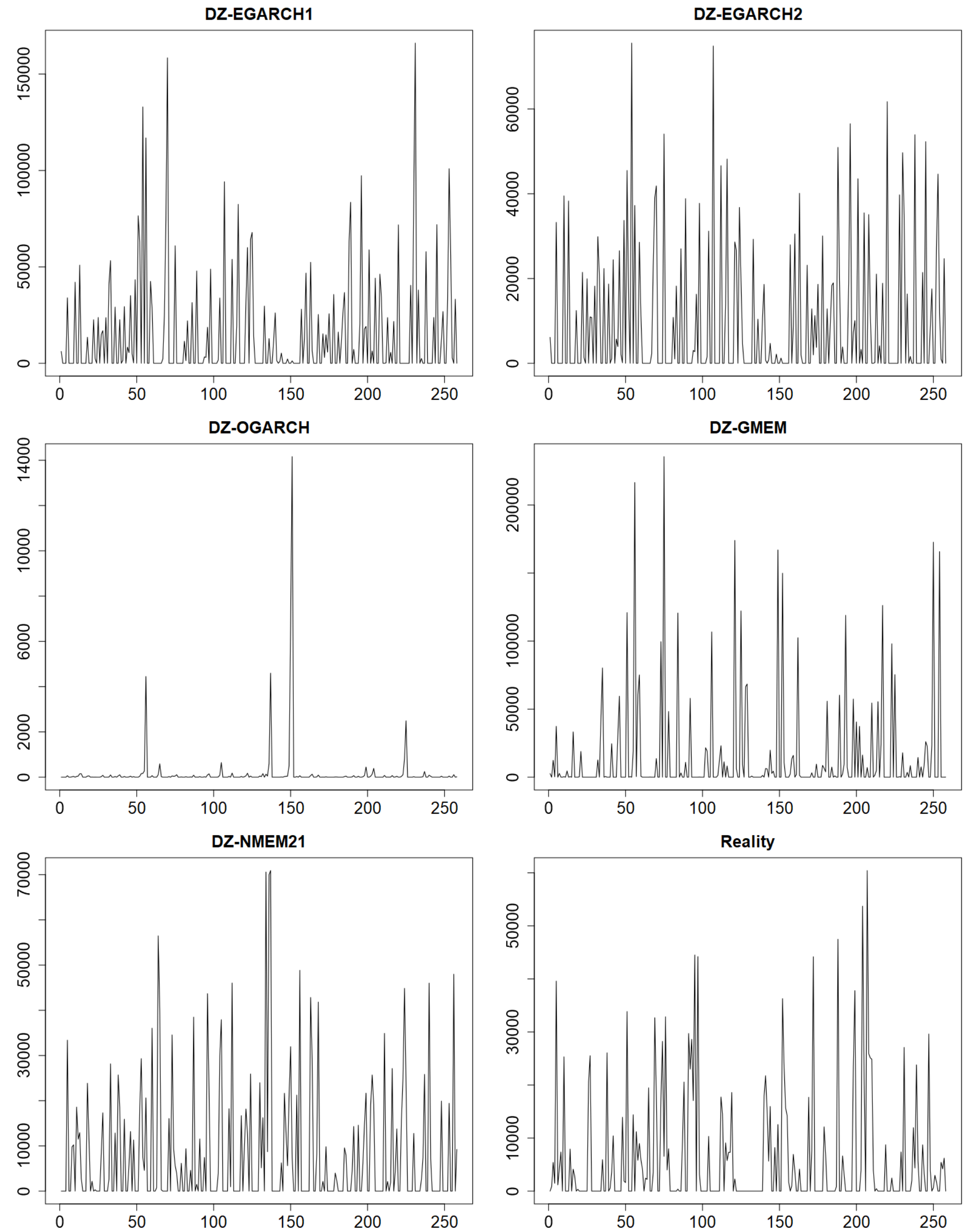

Figure 3.2: One Realization of Predictions from Last Splits Compared to Reality

expands, the exponential term rapidly grows, resulting in sharp spikes. This behavior is likely balanced by the overall low level of most remaining values, leading to the observed disparity.

In contrast, DZ-EGARCH1 tells a much different story. It achieves a desirable balance between low mean MAE and mean sum, while simultaneously capturing the empirical quantiles of the real data effectively. This model allows for occasional high-value spikes, as demonstrated in its visualisation, resulting in a very good overall fit. However, it seems to produce values which exceed the original data by a non-negligible amount which might be a considerable drawback in some real-life applications.

Next, DZ-EGARCH2 emerges as the leading performer among the three models discussed thus far. Sharing much of its structure with its predecessor DZ-EGARCH1, it also includes a slightly more distant past in its structure represented by the parameter γ, allowing for incorporating more information in the computations. Having the second lowest MAE and making its empirical percentiles follow the real data very closely, it becomes a very favourable contestant.

Finally, even though the last two multiplicative error models are from the same family, they are not twins for sure. The Gamma sibling performs poorly, especially because attaining values too high in comparison to the observed reality. DZ-NMEM21, conversely, learned from the mistakes of its brother and its forecast seems to follow the distribution and pattern of reality very well. The lowest mean MAE, reasonable empirical percentiles and also visual resemblance to real data indicate this model as being one of the most plausible options.

	EGARCH1	EGARCH2	OGARCH	GMEM	NMEM21
$\hat{\alpha}_0$	9.190	9.191	351.745	574.278	277.700
$\hat{\alpha}_1$	0.067	0.041	0.946	0.071	0.076
$\hat{\alpha}_2$	—	—	—	—	0.012
$\hat{\beta}$	0.512	0.512	0.054	0.929	0.911
$\hat{\gamma}$	—	16.091	—	—	—
$\hat{\delta}$	—	—	1.999	—	—

Table 3.3: Estimated Parameter Values Taken from the Last (Largest) Split

3.4 Discussion

Building on the established properties and performance of the candidate models, this section delves deeper into their behavior, particularly focusing on models with the most disparate performance: DZ-OGARCH on one side and DZ-EGARCH2 with DZ-NMEM21 on the other. The following aims to discuss the reasons behind their observed behaviors and identify potential areas for future research.

3.4.1 Exponential Loss of DZ-OGARCH

DZ-OGARCH exhibits a notably poor performance compared to other models. While the original results of OGARCH model by Liu and Morimune [2005] were promising, it shall not be forgotten that its evaluation employed data with distinct characteristics. In particular, the data in Liu and Morimune [2005] lacked the dense zero structure present in our study and represented a financial time series, where leverage effects due to shock persistence were expected. These factors justified the inclusion of the exponential term in the original OGARCH model.

However, in our case with dense zeros, this very term appears to be the source of the undesirable behavior. As shown in Table 3.3, the estimated value of the term in the exponential, $\hat{\delta}$, reaches its upper bound, and $\hat{\alpha}_0$ is also relatively high compared to the empirical percentiles of predicted values in Table 3.2. This combination leads to a phenomenon of near-constant very low values, revolving around $\hat{\alpha}_0$, which can unexpectedly surge (due to the exponential term) following a sequence of positive predictions.

Thus, despite its intriguing theoretical foundation, the applicability of DZ-OGARCH seems limited to specific scenarios that differ from the one examined in this thesis. However, if the dense-zeros conditions were relaxed to allow for negative observations, DZ-OGARCH might prove itself to be of much better performance. In addition, there is a possibility that a more thorough simulation and computation studies would reveal that the parameters should be estimated in a different way, or, possibly, that innovations should be defined in a more complex manner. All these possibilities offer numerous options for more detailed investigations in a future study.

3.4.2 Success Echos Through History

On the other hand, two models outperform the rest across all observed metrics – DZ-EGARCH2 and DZ-NMEM21.

Firstly, let us note that DZ-EGARCH2 shares a significant portion of its internal structure with DZ-EGARCH1, and their estimated parameter values (see Table 3.3) are highly similar. The key difference lies in the inclusion of $\hat{\gamma}$ as a "past information carrier," which influences the model's outcome only when the penultimate datapoint is positive. Nevertheless, DZ-EGARCH2 not only demonstrates a superior fit to the data compared to its sibling, but also presents a more practical choice for real-world applications due to its relative simplicity.

Secondly, DZ-NMEM21 also succeeded in a comparable way as the previously discussed model. Again, it shares similar structure with DZ-GMEM as well as the estimated parameters $\hat{\alpha}_1$ and $\hat{\beta}$. The striking difference in the quality of output is caused mainly by including the "historical" term $\alpha_2 y_{t-2}$ and marginally also by changing the distribution of innovations to Normal.[3]

Ultimately, both models showcase the positive effect of including recent history and Normal innovations. Additionally, DZ-EGARCH2 avoids the traditional approach of increasing complexity through structures like GARCH(p, q), where $p \geq 1$ and $q \geq 1$ with at least one inequality strict. Instead, it proposes and successfully validates an alternative approach with a simpler structure with the term γ as described above. A more thorough examination of this alternative approach in various scenarios, particularly in comparison to more elaborate models, could be a rewarding option for further research.

[3] By standalone simulation testing, the biggest effect on performance had indeed the inclusion of the extra term $\alpha_2 y_{t-2}$, proved e.g. by reducing the mean MAE from 119 041 in the case of DZ-GMEM to 33 304. The subsequent change of the innovation distribution then lowered the mean MAE to the value shown in Table 3.2.

3.4.3 Limitations and Future Research

It should be noted that all the findings are related to the particular dataset which has been used in the practical part. It is well possible, also because of solid theoretical background of all the candidate models, that in altered dense-zeros scenarios, the results and the final rankings of the proposed models might get different.

However, it was not desired to create a model which would maximally explain the current data and provide the best possible fit. This approach would lead to overfitting and would not be beneficial for further applications.

In contrast, the main goal was to explore the new ground delimited by the concept of dense zeros and bring fresh ideas on how to design suitable and theoretically justifiable methods of tackling this specific problem. Should the need arise, the several propositions and techniques offered in this thesis can always be modified for the given real-life situation which is one of the most important benefit of this author's work.

Conclusion

The thesis focuses on the dense zeros framework within financial and insurance time series. It begins by establishing a solid theoretical foundation, exploring the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models and Multiplicative Error Models (MEMs). These models serve as the basis for understanding the dynamics of time series data, laying the fundamentals for the following development of novel approaches which should address the issue of dense zeros.

Building upon this theoretical foundations, the thesis introduces five new models specifically designed to handle the presence of excess zeros in time series data. These models, namely DZ-EGARCH1, DZ-EGARCH2, DZ-OGARCH, DZ-GMEM and DZ-NMEM21, offer innovative solutions to effectively model zero-inflated series while capturing the underlying dependence structure within the non-zero observations. Each model is presented with a rigorous definition, outline of its structure and derivation of its properties and estimation and prediction procedures. Creation and further validation of these models is the main contribution of this text.

The validation itself is conducted in the practical part of the thesis using real-world data from the insurance field. The evaluation process involves meeting the dataset, defining the methodology, presenting results and opening a detailed discussion. The results and comparisons obtained from this evaluation reveal the strengths and limitations of the models, not evading the fact that some models may have proven to be actually of very bad performance. Nevertheless, the flexibility offered by the different strengths and weaknesses of the models, implied by various changes in their inner structure and working with historical observations, is one of the most significant contributions of this text.

In conclusion, this thesis represents a significant contribution to the field of time series analysis with dense zeros. By bridging the gap between theoretical concepts and practical applications, the research not only expands the existing knowledge base but also creates opportunities for further exploration and research in this area. The underlying conditions may be made more complex, the models might get suited for slightly different scenarios and more elaborate estimation and optimization methods can be utilized – however, that is left for future research and for the interested reader (or even the author himself) to explore further.

Bibliography

Xuezheng Bai, Jeffrey R. Russell, and George C. Tiao. Kurtosis of garch and stochastic volatility models with non-normal innovations. *Journal of Econometrics*, 114(2):349–360, June 2003. ISSN 0304-4076. doi: 10.1016/s0304-4076(03)00088-5.

Istvan Berkes, Lajos Horvath, and Piotr Kokoszka. Garch processes: structure and estimation. *Bernoulli*, 9(2), April 2003. ISSN 1350-7265. doi: 10.3150/bj/1068128975.

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

Christian T. Brownlees, Fabrizio Cipollini, and Giampiero M. Gallo. Multiplicative error models. *SSRN Electronic Journal*, 2011. ISSN 1556-5068. doi: 10.2139/ssrn.1852285.

Fabrizio Cipollini and Giampiero M. Gallo. Multiplicative error models: 20 years on. *Econometrics and Statistics*, May 2022. ISSN 2452-3062. doi: 10.1016/j.ecosta.2022.05.005.

Tomáš Cipra. *Finanční ekonometrie*. Ekopress, 2 edition, 2013. ISBN 978-80-86929-93-4.

Stavros Degiannakis and Evdokia Xekalaki. Autoregressive conditional heteroscedasticity (arch) models: A review. *Quality Technology and Quantitative Management*, 1(2):271–324, jan 2004. doi: 10.1080/16843703.2004.11673078.

Robert Engle. New frontiers for arch models. *Journal of Applied Econometrics*, 17(5):425–446, September 2002. ISSN 1099-1255. doi: 10.1002/jae.683.

Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987, jul 1982. doi: 10.2307/1912773.

Robert F. Engle and Giampiero M. Gallo. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1–2):3–27, March 2006. ISSN 0304-4076. doi: 10.1016/j.jeconom.2005.01.018.

Robert F. Engle and Jeffrey R. Russell. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66(5):1127, September 1998. ISSN 0012-9682. doi: 10.2307/2999632.

Christian Francq and Jean-Michel Zakoïan. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4), August 2004. ISSN 1350-7265. doi: 10.3150/bj/1093265632.

Lawrence R. Glosten, Ravi Jagannathan, and David E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, December 1993. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1993.tb05128.x.

Jan Hanousek. Dependent zeros. Bachelor thesis, Faculty of Mathematics and Physics, Charles University, 2022.

Da Huang, Hansheng Wang, and Qiwei Yao. Estimating garch models: when to use what? *The Econometrics Journal*, 11(1):27–38, March 2008. ISSN 1368-423X. doi: 10.1111/j.1368-423x.2008.00229.x.

Šárka Hudecová and Michal Pešta. Quasi-likelihood estimation in volatility models for semi-continuous time series. 2024.

Sang-Won Lee and Bruce E. Hansen. Asymptotic theory for the garch(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, 10(1):29–52, March 1994. ISSN 1469-4360. doi: 10.1017/s0266466600008215.

Qingfeng Liu and Kimio Morimune. A modified garch model with spells of shocks. *Asia-Pacific Financial Markets*, 12(1):29–44, March 2005. ISSN 1573-6946. doi: 10.1007/s10690-006-9011-z.

Robin L. Lumsdaine. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in igarch(1,1) and covariance stationary garch(1,1) models. *Econometrica*, 64(3):575, May 1996. ISSN 0012-9682. doi: 10.2307/2171862.

Daniel B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347, March 1991. ISSN 0012-9682. doi: 10.2307/2938260.

Arie Preminger and Giuseppe Storti. Least squares estimation for garch (1,1) model with heavy tailed errors. 2014.

Matthias Schnaubelt. A comparison of machine learning model validation schemes for non-stationary time series data. *FAU Discussion Papers in Economics*, (11/2019), 2019. ISSN 1867-6707.

Annastiina Silvennoinen and Timo Teräsvirta. Consistency and asymptotic normality of maximum likelihood estimators of a multiplicative time-varying smooth transition correlation garch model. *Econometrics and Statistics*, August 2021. ISSN 2452-3062. doi: 10.1016/j.ecosta.2021.07.008.

Ruey S. Tsay. *Analysis of financial time series*. Wiley, 2010. ISBN 9780470414354.

List of Figures

List of Tables

List of Abbreviations

ACD . . . . . Autoregressive Conditional Duration
ADF . . . . . Augmented Dickey-Fuller
ARMA . . . . Autoregressive Moving Average
a.s. . . . . . . Almost Surely
cor . . . . . . Correlation
cov . . . . . . Covariance
DZ . . . . . . Dense Zeros
EGARCH . . Exponential GARCH
GARCH . . . Generalized Autoregressive Conditional Heteroscedasticity
GJR-GARCH Glosten-Jagannathan-Runkle GARCH
GMEM . . . Gamma MEM
IGARCH . . Integrated GARCH
K . . . . Kurtosis
MAE . . . . . Mean Absolute Error
MAPE . . . . Mean Absolute Percentage Error
MEM . . . . . Multiplicative Error Model
MLE . . . . . Maximum Likelihood Estimate
MSE . . . . . Mean Square Error
NMEM . . . Normal MEM
OGARCH . . Overresponse GARCH
QML . . . . . Quasi-Maximum Likelihood
RMSE . . . . Root Mean Square Error
TGARCH . . Threshold GARCH
Var . . . . . . Variance

A. Attachments

A.1 Major Parts of the Source Code

The following code shows the main parts of the optimization, estimation and prediction procedures for the model DZ-EGARCH1. (The in-time sample from the original dataset was stored in variable ”train”.) Other models’ codes are analogous.

```
# Initialization of variables
train.all = train[ ,2]
train.pos = train[which(train$Amount>0),2]
l = length(train.all)
K = 6
J = 1000
sample.size = floor(l/K)
paramNames = c("alpha0","alpha1","beta","gamma")
parDf = setNames(
    data.frame(matrix(
    ncol = length(paramNames),
    nrow = (K-1)
    )),
    paramNames)
y.hatDf = setNames(
    data.frame(matrix(ncol = (K-1), nrow = sample.size)),
    1:(K-1))
phat = rep(0,(K-1))
temp.mae = rep(0,(K-1)*J)
temp.sum = rep(0, (K-1)*J)
realizations = vector()

# Beginning of the main cycle
set.seed(128)
for (k in 1:(K-1)) {
setTxtProgressBar(pb,k)
train.all.k = train.all[1:(k*sample.size)]
test.k = train.all[(k*sample.size+1):((k+1)*sample.size)]
train.pos.k = train.all.k[which(train.all.k>0)]
zeros.count = length(which(train.all.k == 0))
phat[k] = (1/length(train.all.k))*zeros.count


# Estimate hattheta
garch_filter = function(vP, y) {
    alpha0 = vP[1]
    alpha1 = vP[2]
    beta = vP[3]
    gamma = vP[4]
```

```
    sigma_t = rep(0,length(y))
    for (i in 1:length(y)) {
        if (i == 1) {
            sigma_t[i] = y[i]
        } else {
            sigma_t[i] = sqrt(exp(alpha0 +
                (alpha1*(y[i-1]*(1+gamma))/sigma_t[i-1]) +
                (beta*log((sigma_t[i-1])^2))
                ))
        }
    }
    return(sigma_t)
}

# Function to minimize in QMLE
minim.problem1 = function(vP, y) {
    alpha0 = vP[1]
    alpha1 = vP[2]
    beta = vP[3]
    gamma = vP[4]
    sigma_t = garch_filter(vP, y)
    mnm = sum(y^2/sigma_t^2+log((sigma_t)^2))
    return(mnm)
}

# Optimization - Initialization
library(alabama)
const.func = function(vP, y) {
    alpha0 = vP[1]
    alpha1 = vP[2]
    beta = vP[3]
    gamma = vP[4]
    # set constraints
    h = rep(NA, 5)
    h[1] = alpha0  # alpha0 > 0
    h[2] = alpha1  # alpha1 > 0
    h[3] = beta  # beta > 0
    h[4] = -alpha1-beta+1  # alpha0 + beta < 1
    h[5] = gamma+1  # gamma > -1
    return(h)
}
optim_method = c(
    "Nelder-Mead",
    "BFGS",
    "CG",
    "SANN"
)
```

```
# Optimization - Main (some parts omitted)
m_coptim = constrOptim.nl(par = vP0,
                            y = train.pos.k,
                            fn = minim.problem1,
                            hin = const.func,
                            control.outer = list(
                                method = optim_method[1],
                                trace = FALSE
                            )
)
alpha0.hat = m_coptim$par[1]
alpha1.hat = m_coptim$par[2]
beta.hat = m_coptim$par[3]
gamma.hat = m_coptim$par[4]
parDf[k, ] = c(alpha0.hat, alpha1.hat, beta.hat, gamma.hat)

# Cycle 2 - seqs ####
# --> Omitted

# Cycle 3 - preds ####
for (h in 2:m) {
    sigma.hat[h] = sqrt(exp(alpha0.hat
                            + alpha1.hat*(
                                (y.hat[h-1]*(1+gamma.hat))/
                                sigma.hat[h-1]
                                )
                            + beta.hat*log(
                                (sigma.hat[h-1])^2
                                )))
    y.hat[h] = sigma.hat[h]*eps[h]
}
y.hat.pos = y.hat[which(y.hat>0)]
temp.mae[((k-1)*J+j)] = mae(y.hat,test.k)
temp.sum[((k-1)*J+j)] = sum(y.hat)
realizations = append(realizations,y.hat.pos)

if (k == (K-1) & j == J) {
    dzegarch.mae = mean(temp.mae)
    dzegarch.meansum = mean(temp.sum)
    dzegarch.quantsum = quantile(temp.sum)
    dzegarch.quantreals = quantile(realizations,
        probs = seq(0,1,0.05))
    close(pb)
}
}
```