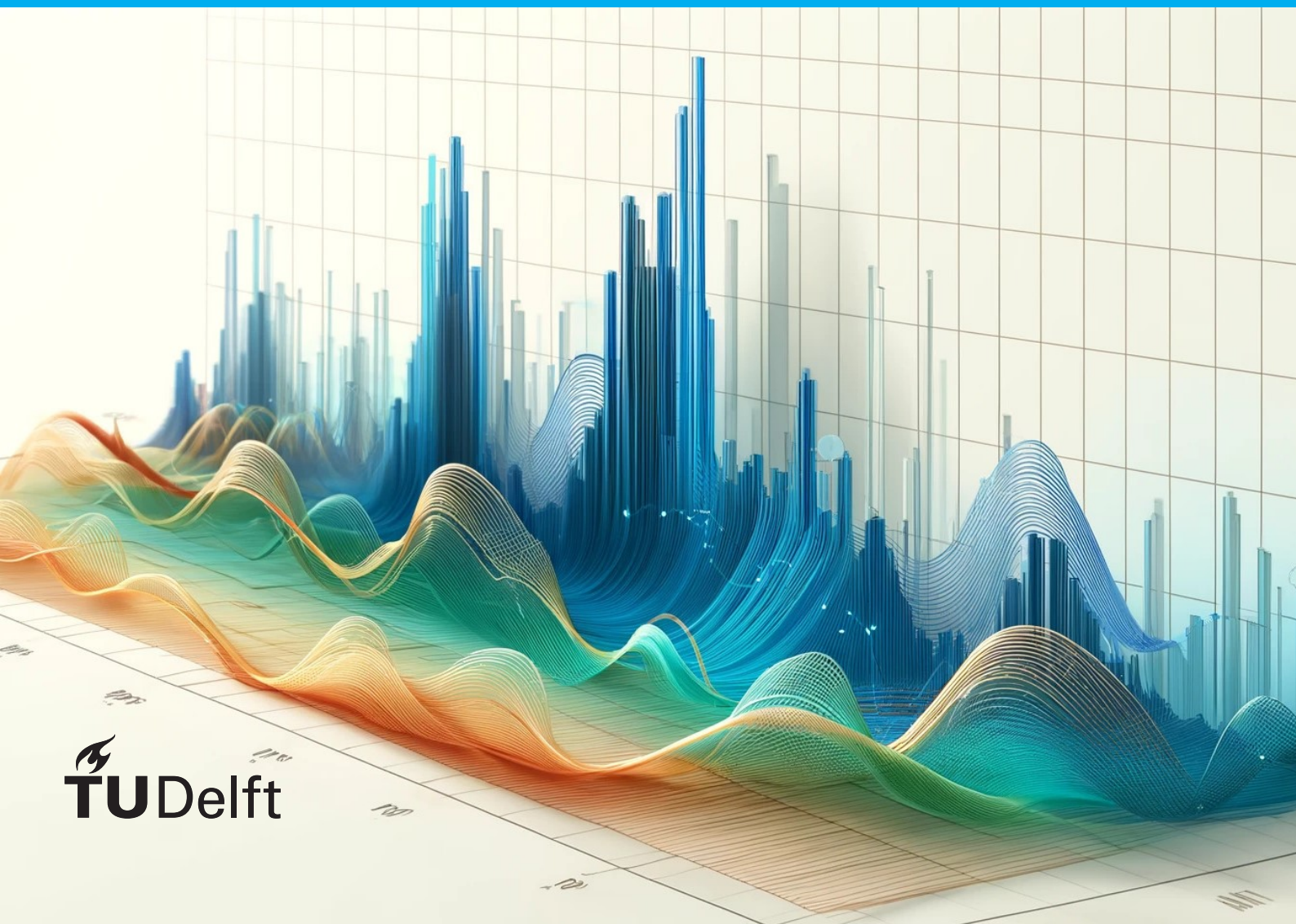
A universal method for statistical inference of low-frequency time series

Fernando De Diego Ávila



TU Delft

A universal method for statistical inference of low-frequency time series

by

Fernando De Diego Ávila

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Wednesday August 28, 2024 at 10:00 AM.

Student number:	5846625	
Project duration:	December 1, 2023 – August 28, 2024	
Thesis committee:	Prof. dr. A. Derumigny,	TU Delft, supervisor
	Prof. dr. F. Fang,	TU Delft, supervisor
	Prof. dr. D. Kurowicka,	TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Abstract

Statistical inference of low-frequency time series is a challenge present in various fields, such as financial risk management and weather forecasting. Practical difficulties arise due to the scarcity of non-overlapping observations. The "direct method", which directly uses the available low-frequency data to construct estimators, often results in inaccurate estimations.

In this thesis, we propose a novel "simulation-based method" for statistical inference of low-frequency time series that result from the aggregation of a higher-frequency time series over a period of time. We start by estimating the distribution of this higher-frequency process. We then simulate a large number of paths from this estimated distribution. By independently aggregating each simulated path, we generate corresponding low-frequency data. This provides us with a large simulated dataset of the low-frequency process, which enables us to apply estimation procedures and bypass the limitations posed by the shortage of original low-frequency data.

We also provide a theoretical framework and propose three families of estimators constructed from the estimated higher-frequency distribution, analyzing their properties under additional assumptions. Through a comprehensive simulation study, we compare the simulation-based method with the traditional direct method across different scenarios and objectives. While our study focuses on the marginal distributions of low-frequency processes, the simulation-based method's applicability extends to joint distributions across multiple time points. This research offers a robust method for parameter estimation when faced with limited low-frequency data.

Acknowledgements

I would like to express my profound gratitude to Professor Alexis Derumigny for his exceptional guidance throughout this project, from start to finish. Your commitment and the considerable time you invested in our weekly meetings have been invaluable, and I deeply appreciate your assistance. Similarly, I am thankful to Professor Fang Fang, whose expertise effectively connected theoretical concepts with their practical applications in the financial world, providing insightful feedback and support. I would like to extend my gratitude to all the professors and teachers I have encountered throughout my life, many of whom taught me more than mere concepts, imparting values and perspectives.

This journey would not have been possible without the support and love from my family, who believed in me even through my own doubts. Their encouragement and faith in my capabilities kept me motivated through the highs and lows of this challenging journey. I could not have asked for more, they are the best. Alongside my family, in the Netherlands, I was fortunate to forge friendships with people from all around the world, who provided me with support and care. I am deeply thankful to them. Additionally, I am grateful to my friends from Chile who have always supported me, and to all the wonderful people I met during my Master's program, each of whom, through brief or lasting interactions, contributed significantly to my personal and professional development.

To everyone who has been a part of this journey: Thank you. Your support, wisdom, and friendship have been incredibly meaningful to me. [1]

Fernando De Diego Ávila
Delft, August 2024

[1] The cover image for this thesis was created using the AI tool DALL-E, based on the prompt "Cover for a MSc thesis involving time series analysis". I highly recommend exploring this innovative tool for its creative potential.

Contents

1

Introduction

The challenge of estimating parameters of the distribution of low-frequency stochastic processes is present in various fields, such as risk management and weather forecasting. These random processes are characterized by a low number of non-overlapping observations over long time spans. Examples include the yearly returns of a portfolio of assets and the decennial rainfall in the Netherlands. Unfortunately, practical challenges arise when estimating parameters from available low-frequency data. Specifically, the number of non-overlapping observations is often too limited to yield reliable estimates. Moreover, using overlapping data, in which the calculation periods for each value intersect, introduces autocorrelation, leading to biased estimates. Therefore, it is essential to explore the most effective ways to utilize the available data to capture the behavior of low-frequency processes and accurately estimate the parameters of their distribution.

Low-frequency processes can sometimes be represented as the temporal aggregation of higher-frequency processes. For instance, the yearly absolute or log-returns of a portfolio of assets are equal to the sum of the daily returns of the portfolio over a period of one year. When such a relationship holds, it becomes attractive to use higher-frequency data, which is naturally more abundant than its related low-frequency data, to obtain properties of the low-frequency distribution. Several authors have studied this relationship, as presented in the comprehensive survey by Silvestrini and Veredas [20]. In this survey, the authors discuss the true models of aggregated ARIMA and ARMA-GARCH processes, as well as other types of processes. Unfortunately, such direct results are often unavailable for non-linear models with higher complexity.

In a similar line of investigation, some authors study the accuracy of model estimation for aggregated processes when their aggregated distribution is known. For instance, Teles and Sousa [21] examine the estimation accuracy of temporal aggregation of ARMA models by calculating the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) of the estimated parameters relative to the true parameters. They conclude that the accuracy of the estimated parameters for the aggregated series is usually poor, being good or acceptable only in the case of short aggregation periods and simple models. This conclusion makes sense, as the available data for estimation decreases when considering longer aggregation periods, and complex models usually require a large amount of data to avoid over-fitting. On the contrary, they mention that the estimation accuracy of the basic (higher-frequency) processes is at least acceptable for all models. Naturally, this conclusion encourages exploiting disaggregated data for the inference on aggregated variables.

Some studies also focus on comparing the performance of low- and higher-frequency models for applications, particularly in forecasting. In the literature, direct methods for forecasting multi-period (low-frequency) processes involve fitting models directly using multi-period data. Conversely, iterated methods use one-period (higher-frequency) models that are iterated forward by the number of periods considered by the multi-period processes. In this latter case, the forecasts of the low-frequency processes are obtained by aggregating the forecasts of the higher-frequency processes. Lütkepohl [13] conducts a survey on methods for forecasting aggregated time series, concluding that forecasting the disaggregated process and then aggregating the forecasts usually achieves a lower Mean Squared Error (MSE) than forecasting the aggregate process directly, as the latter approach ignores the disaggregated information. In other words, utilizing higher-frequency data for forecasting is often beneficial.

An important application in financial risk management involving low-frequency returns is the construction of the Default Risk Charge (DRC) model outlined by the Basel Committee on Banking Supervision (BCBS) [2], which is part of the minimum capital requirements for market risk. Under the Internal Model Approach (IMA), the DRC is defined as the 99.9% Value-at-Risk (VaR) of the one-year total loss distribution of Trading Book positions subject to issuer risk. Banks are restricted to use equity prices or Credit Default Spreads (CDSs) to represent their "asset returns". Furthermore, the one-year forecasting horizon implies that the calibration should be based on yearly returns. Given the low-frequency nature of these returns, the amount of non-overlapping yearly (low-frequency) returns available to calibrate the model is very limited, even when considering long time spans. For instance, a decade of data would only correspond to ten non-overlapping data points, which is insufficient for statistical modeling.

A number of studies have attempted to address the challenges in estimating multi-period Value-at-Risk. Ruiz & Nieto [17] presented a comprehensive survey in 2023 comparing direct and iterated methods to forecast multi-period VaR of portfolios. Multi-period VaR can be calculated by the temporal aggregation of one-period returns, and consequently, both direct and iterated methods for forecasting can be considered. The authors state that the scarce literature comparing these alternatives tends to favor iterated methods over direct methods. Since the DRC is defined as the VaR measure, and the one-year return distribution is obtained from multi-period returns, the insights gathered in their study can be considered for the calculation of the DRC.

In this thesis, we present a universal method for statistical inference of low-frequency processes that result from the aggregation of a higher-frequency process over a period of time. We put special focus on the marginal distributions of these processes, which are crucial for various financial applications in risk management, such as calibrating the DRC model for the one-year horizon. We compare two methods for estimating the parameters of the distribution of low-frequency processes. The first method, which we call the "direct method", is analogous to the direct method for forecasting. It directly utilizes the available low-frequency observations to construct the estimators. The direct method is the conventional approach for parameter estimation and, unfortunately, may suffer from the limitation of having a small amount of non-overlapping data. Moreover, when using overlapping data, each observation depends on the previous one(s), potentially leading to biased estimators.

We propose an alternative approach, which we call the "simulation-based method". Similar to the iterated method for forecasting, the simulation-based method aims to address situations where the small amount of available low-frequency data is insufficient to construct reliable direct estimators for the parameters of the distribution of low-frequency processes. To effectively utilize the higher-frequency data, we first estimate the distribution of the higher-frequency process and then simulate a large number of higher-frequency paths from this distribution. Afterwards, we independently aggregate each path to generate a large sample of low-frequency data. Finally, from this data, we estimate the parameters of the true distribution of the low-frequency process. The simulation-based method can be used to estimate parameters of the marginal distributions of the low-frequency process, as well as parameters of joint distributions across multiple time points.

Our simulation-based method can be seen as a model-based data augmentation technique. Data augmentation techniques are used to increase the diversity of a dataset without collecting new data. In the context of time series, these techniques include noise injection, reflection with respect to the y-axis (also known

as flipping), and more advanced techniques in which models are fit to available data [24]. Naturally, data augmentation techniques are task-dependent, as certain techniques might not produce appropriate datasets for some tasks. For this thesis, the possibility of precisely estimating higher-frequency models, and the known relationship between low- and higher-frequency data, allow the generation of a large low-frequency dataset from simulated higher-frequency data, bypassing the shortage of data at the low-frequency scale.

This thesis is organized as follows: Chapter 2 provides an overview of the theory behind the topic, covering the basics of random variables and probability measures, as well as properties of estimators for parameters of probability distributions. Additionally, it introduces several time series models for describing time-dependent data and copula models for explaining dependencies between multiple time series. Chapter 3 presents the mathematical framework for this work, including relevant notation and assumptions. It reintroduces the motivation for the study with more mathematical rigor and introduces the direct and simulation-based methods for parameter estimation. Three families of estimators derived from the estimated distribution of the higher-frequency process are then introduced. Their main properties are analyzed, and new results on their consistency, rate of convergence, and asymptotic distribution are provided where possible. Examples of estimators that belong to these families are also included. Chapter 4 compares the performance of the direct and simulation-based methods across different scenarios and objectives. It begins with an introduction to the experimental framework, outlining the main assumptions and metrics used to evaluate the accuracy of the estimators, followed by various parameter estimations using both direct and simulation-based methodologies. Theoretical and simulation results are provided throughout these sections[1]. Finally, Chapter 5 concludes the thesis, summarizing the key findings.

[1]The code is available at:
https://github.com/fernandodediegoavila/A-universal-method-for-statistical-inference-of-low-frequency-time-series

2

Preliminaries

This chapter aims at introducing the required fundamentals for the material discussed in the upcoming chapters. First, in Section 2.1 we introduce basic definitions regarding continuity, random variables, probability measures, and stochastic processes. Afterwards, in Section 2.2 we provide results on consistency, rates of convergence, and asymptotic distributions of estimators, mainly M-estimators. Finally, in Section 2.3, we introduce copulas and the Copula-GARCH model, which is one of the models used in our simulations.

2.1. On random variables and probability measures

Throughout this thesis, we rely heavily on understanding how sequences of random variables and sequences of probability measures can converge. The goal of this section is to introduce these concepts and the primary results associated with them. We start by introducing metric spaces and continuous functions defined on them. Later, we discuss random variables on arbitrary probability spaces, and various ways in which sequences of random variables can converge. We also introduce concepts such as inner and outer probability, which help to extend the notion of convergence in probability of sequences of random variables. We continue by defining probability measures, and explaining two ways in which sequences of probability measures can converge. We also present the definition of tightness in these measures and the powerful Continuous Mapping Theorem. Afterwards, we generalize the notion of sequences of probability measures by allowing them to be random. We end this section by introducing stochastic processes and time series, which help to model time-dependent phenomena.

The definitions of metrics and metric spaces (Definition 1) and empirical probability measures (Example 1) are based on Dudley [8]. The concepts of inner & outer integral (Definition 3), inner & outer probability (Definition 5), and tightness (Definition 10) are extracted from van der Vaart & Wellner [23]. The definitions of convergence of sequences of random variables (Definition 4), and stochastic o and O notations (Definition 6) are based on van der Vaart [22]. The definition of probability measures (Definition 7), the definition of weak convergence of sequences of probability measures (Definition 8), and the Continuous Mapping Theorem (Theorem 1) are adapted from Billingsley [3]. Lastly, the content in subsection 2.1.5 is based on Brockwell & Davis [4].

2.1.1. Metric spaces and continuity

Let us start by defining the concept of metric, which is a mathematical formalization of our everyday understanding of distance.

Definition 1 (Metric and metric space)**.** Given a set E, a metric for E is a function d_E from $E \times E$ into $\mathbb{R}_+$ such that

1. for all x, y in E, $d_E(x, y) = 0$ if and only if $x = y$,
2. for all x, y in E, $d_E(x, y) = d_E(y, x)$ (symmetry), and
3. for all x, y, z in E, $d_E(x, z) \leq d_E(x, y) + d_E(y, z)$ (triangle inequality).

If these conditions are satisfied, (E, d_E) is called a metric space.

Continuity in metric spaces captures the concept that small changes in the input lead to small changes in the output, while sequential continuity ensures that sequences converging in their domain are mapped to sequences converging in their codomain. These concepts are fundamental in mathematical analysis and are useful to prove properties of functions defined throughout this thesis. We thus provide a formal definition of both concepts.

Definition 2 (Continuity and sequential continuity)**.** Let (E_1, d_{E_1}) and (E_2, d_{E_2}) be metric spaces. A function $f : E_1 \to E_2$ is continuous at $x_0 \in E_1$ if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that $d_{E_1}(x, x_0) < \delta$ implies that $d_{E_2}(f(x), f(x_0)) < \epsilon$. If f is continuous at all $x \in E_1$, then we say that f is continuous on E_1 or simply continuous. Similarly, a function $f : E_1 \to E_2$ is sequentially continuous at $x_0 \in E_1$ if, for $\{x_n\}_{n\in\mathbb{N}}$ in E_1, the condition $d_{E_1}(x_n, x_0) \to 0$ implies that $d_{E_2}(f(x_n), f(x_0)) \to 0$.

Remark 1. In the case of metric spaces, continuity and sequential continuity are equivalent.

2.1.2. Random variables

Having defined metric spaces and continuity, let us shift our focus to random variables in a probability space. Let $(\Omega, \mathscr{A}, \mathbf{P})$ be an arbitrary probability space. A random variable is a measurable map $X : \Omega \to E$ from the sample space Ω into a measurable space $(E, \mathscr{E})$. Unless specified, throughout this thesis we will consider E to be a metric space with metric d_E and $\mathscr{E}$ the Borel σ-algebra related to d_E. We represent the Lebesgue integral of X with respect to $\mathbf{P}$, which is often called the expectation of X, as

$$\mathbb{E}[X] := \int_\Omega X \, d\mathbf{P}.$$

This integral can be extended to non-measurable maps by defining the inner and outer integrals.

Definition 3 (Inner & outer integral)**.** For an arbitrary probability space $(\Omega, \mathscr{A}, \mathbf{P})$ and $T : \Omega \to \mathbb{R}$ an arbitrary map, the inner integral of T with respect to $\mathbf{P}$ is defined as

$$\mathbb{E}_*[T] = \sup\{\mathbb{E}[U] : U \leq T, U : \Omega \to \mathbb{R} \text{ measurable and } \mathbb{E}[U] \text{ exists}\}.$$

Similarly, the outer integral of T with respect to $\mathbf{P}$ is defined as

$$\mathbb{E}^*[T] = \inf\{\mathbb{E}[U] : U \geq T, \Omega \to \mathbb{R} \text{ measurable and } \mathbb{E}[U] \text{ exists}\}.$$

We proceed to discuss the convergence of random variables. Three of the most common ways of convergence of random variables are **P**-almost sure convergence, convergence in probability, and convergence in distribution, which are presented in the following definition.

Definition 4 (Convergence of sequences of random variables)**.** Let (E, d_E) be a metric space, let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of E-valued random variables, and let X be another E-valued random variable. We say that:

- $\{X_n\}_{n\in\mathbb{N}}$ converges to X **P**-almost surely (or strongly) if $\mathbf{P}(d_E(X_n, X) \to 0) = 1$. We denote it $X_n \to X$ **P**-a.s.
- $\{X_n\}_{n\in\mathbb{N}}$ converges to X in probability if, for all $\epsilon > 0$, it holds that $\mathbf{P}(d_E(X_n, X) > \epsilon) \to 0$. We denote it $X_n \xrightarrow{\mathbf{P}} X$.
- $\{X_n\}_{n\in\mathbb{N}}$ converges to X in distribution (or weakly) if, for every measurable set $A \subseteq E$, we have $\mathbf{P}(X_n \in A) \to \mathbf{P}(X \in A)$. We denote it $X_n \rightsquigarrow X$.

Remark 2. It can be proven that convergence **P**-almost surely implies convergence in probability and convergence in distribution.

Remark 3. Almost sure convergence and convergence in probability require that $\{X_n\}_{n\in\mathbb{N}}$ and X are defined on the same probability space $(\Omega, \mathscr{A}, \mathbf{P})$, contrary to weak convergence.

In some occasions, we also talk about convergence in inner or outer probability, which extend the concept of convergence in probability. It is thus crucial to define what inner and outer probability are.

Definition 5 (Inner & outer probability)**.** For an arbitrary probability space $(\Omega, \mathscr{A}, \mathbf{P})$ and a subset B of Ω, the inner probability of B is defined as

$$\mathbf{P}_*(B) := \sup\{\mathbf{P}(A) : A \subseteq B, A \in \mathscr{A}\}.$$

Similarly, the outer probability of a subset B of Ω is defined as

$$\mathbf{P}^*(B) := \inf\{\mathbf{P}(A) : A \supseteq B, A \in \mathscr{A}\}.$$

With these definitions, convergence in inner and outer probability of sequences of random variables are defined as in Definition 4, but with **P** being replaced by the terms $\mathbf{P}_*$ and $\mathbf{P}^*$, respectively.

We end this subsection by introducing the stochastic o and O notations.

Definition 6 (Stochastic o and O notations)**.** The notation $o_{\mathbf{P}}(1)$ represents a sequence of random variables $\{R_n\}_{n\in\mathbb{N}}$ that converges to zero in probability, i.e. $R_n \xrightarrow{\mathbf{P}} 0$. On the other hand, the expression $O_{\mathbf{P}}(1)$ refers to a sequence $\{R_n\}_{n\in\mathbb{N}}$ that is bounded in probability. Specifically, for every $\epsilon > 0$, there exists a natural number n' and a constant $M > 0$ such that $\mathbf{P}(|R_n| > M) < \epsilon$ for all $n' \geq n$. Now, let $\{R_n\}_{n\in\mathbb{N}}, \{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ be sequences of random variables. Then,

$$X_n = o_{\mathbf{P}}(R_n) \text{ means } X_n = Y_n R_n \text{ and } Y_n \xrightarrow{\mathbf{P}} 0;$$

$$X_n = O_{\mathbf{P}}(R_n) \text{ means } X_n = Y_n R_n \text{ and } Y_n = O_{\mathbf{P}}(1).$$

As for the convergence of sequences of random variables, we can extend these concepts to inner and outer probabilities by replacing **P** by $\mathbf{P}_*$ and $\mathbf{P}^*$, respectively. In such cases, we write $o_{*\mathbf{P}}$ and $O_{*\mathbf{P}}$ when referring to inner probability, and $o^*_{\mathbf{P}}$ and $O^*_{\mathbf{P}}$ when talking about outer probability.

2.1.3. Probability measures

Let us now introduce the concept of probability measures on metric spaces, which allows us to quantify the probabilities of events in spaces that incorporate notions of distance and continuity.

Definition 7 (Probability measure)**.** Let (E, d_E) be a metric space and $\mathscr{E}$ its Borel σ-algebra, i.e. the σ-algebra generated by the open sets of E with respect to the metric d_E. A probability measure on $(E, \mathscr{E})$ is a non-negative, countably additive set function P satisfying $P(E) = 1$.

Remark 4. Probability measures can be defined on any measurable space, that is, any set E that has a σ-algebra $\mathscr{E}$. However, throughout this thesis, we only consider E to be a space equipped with a metric d_E, and $\mathscr{E}$ to be the Borel σ-algebra generated by this metric.

For a class of functions $\mathscr{F}$ on a metric space, let us denote the space of all uniformly bounded, real-valued functions on $\mathscr{F}$ as $\ell^\infty(\mathscr{F})$. That is, $\ell^\infty(\mathscr{F})$ is the set of all functions $z : \mathscr{F} \to \mathbb{R}$ such that

$$\|z\|_{\mathscr{F}} := \sup_{f \in \mathscr{F}} \left| z(f) \right| < \infty.$$

This is a metric space with respect to the metric $\mathrm{d}_{\mathscr{F}}(z_1, z_2) = \|z_1 - z_2\|_{\mathscr{F}}$ for $z_1, z_2 \in \ell^\infty(\mathscr{F})$. To avoid confusion, note that each operator $z \in \ell^\infty(\mathscr{F})$ takes a function $f \in \mathscr{F}$ as its argument and maps this function to a real number.

In the field of probability theory, different types of convergence of sequences of probability measures are well established. In this thesis, we are mainly interested in weak convergence and convergence in $\ell^\infty(\mathscr{F})$, which are presented in the following definitions.

Definition 8 (Weak convergence of sequences of probability measures)**.** Let (E, d_E) be a metric space and $\mathscr{E}$ be the Borel σ-algebra associated to it. Let P be a probability measure on $(E, \mathscr{E})$, and $\{P_n\}_{n \in \mathbb{N}}$ a sequence of probability measures on the same space. If

$$P_n f = \int_E f \, dP_n \xrightarrow[n \to \infty]{} \int_E f \, dP = Pf$$

for every bounded, continuous real function f on E, we say that $\{P_n\}_{n \in \mathbb{N}}$ converges weakly to P and write $P_n \rightsquigarrow P$.

Remark 5. If X is a random variable and $\{X_n\}_{n \in \mathbb{N}}$ a sequence of random variables such that $X_n \rightsquigarrow X$, then the induced measures $P_n := \mathbf{P} \circ X_n^{-1}$ converge weakly to the induced measure $P := \mathbf{P} \circ X^{-1}$.

Definition 9 ($\ell^\infty(\mathscr{F})$-convergence of sequences of probability measures)**.** Let (E, d_E) be a metric space and $\mathscr{E}$ be the Borel σ-algebra associated to it. Let P be a probability measure on $(E, \mathscr{E})$, $\{P_n\}_{n \in \mathbb{N}}$ a sequence of probability measures on $(E, \mathscr{E})$, and $\mathscr{F}$ a class of functions on the same space. We say that $\{P_n\}_{n \in \mathbb{N}}$ converges to P in $\ell^\infty(\mathscr{F})$ if

$$\|P_n - P\|_{\mathscr{F}} = \sup_{f \in \mathscr{F}} \left| P_n f - Pf \right| \to 0.$$

In such a case, we write $P_n \xrightarrow{\ell^\infty(\mathscr{F})} P$.

We now introduce the concept of tightness of probability measures and measurable maps, which is a fundamental concept in probability theory, especially in the context of weak convergence, since it is a required condition for several results to hold.

Definition 10 (Tightness)**.** Let (E, d_E) be a metric space and $\mathscr{E}$ the Borel σ-algebra associated to it. A probability measure P on $(E, \mathscr{E})$ is tight if for every $\epsilon > 0$ there exists a compact set $K \subseteq E$ with $P(K) \geq 1 - \epsilon$. A measurable map $X : \Omega \to E$ is called tight if its law $P_X = \mathbf{P} \circ X^{-1}$ is tight, while a sequence $\{X_n\}_{n \in \mathbb{N}}$ of measurable

maps is uniformly tight if K can be chosen the same for every n. Finally, a sequence $\{X_n\}_{n\in\mathbb{N}}$ is asymptotically tight if for every $\epsilon > 0$ there exists a compact set $K \subseteq E$ such that

$$\liminf_{n\to\infty} \mathbf{P}_* \left(X_n \in K^\delta \right) \geq 1 - \epsilon, \quad \text{for every } \delta > 0,$$

where $K^\delta := \{x \in E : d_E(x, K) < \delta\}$ is called the δ-enlargement around K.

Remark 6. Asymptotic tightness is a weaker condition than tightness.

Remark 7. The concept of tightness should not be immediately generalized to arbitrary σ-algebras without considering additional conditions.

To end this section, we introduce the powerful Continuous Mapping Theorem, a fundamental result in probability theory that extends the concept of convergence when a continuous map is applied. This theorem is crucial in analyzing the behavior of sequences of random variables and their transformations. To do so, suppose that h is a map from a metric space (E_1, d_{E_1}) to another metric space (E_2, d_{E_2}). Let $\mathscr{E}_1$ and $\mathscr{E}_2$ be the Borel σ-algebras of E_1 and E_2, respectively. If h is $\mathscr{E}_1/\mathscr{E}_2$ measurable, then each probability measure P on $(E_1, \mathscr{E}_1)$ induces a probability measure $P \circ h^{-1}$ on $(E_2, \mathscr{E}_2)$ defined by $\left(P \circ h^{-1}\right)(A) = P\left(h^{-1}(A)\right)$. In other words,

$$\int_{E_2} f(y)\, d\left(P \circ h^{-1}\right)(y) = \int_{E_1} f(h(x))\, dP(x), \tag{2.1}$$

for any $f : E_2 \to \mathbb{R}$ such that one of the two integrals exists (and therefore both).

Theorem 1 (Continuous Mapping Theorem)**.** *Let h be a continuous map from a metric space (E_1, d_{E_1}) to another metric space (E_2, d_{E_2}), and let $\mathscr{E}_1$ and $\mathscr{E}_2$ be the Borel σ-algebras of E_1 and E_2, respectively. Let $\{P_n\}_{n\in\mathbb{N}}$ be a (non-random) sequence of probability measures on $(E_1, \mathscr{E}_1)$. Then $P_n \rightsquigarrow P$ implies $P_n \circ h^{-1} \rightsquigarrow P \circ h^{-1}$.*

2.1.4. Convergence of random sequences of probability measures

In statistical applications, we often encounter sequences of probability measures that are non-deterministic. These random sequences of probability measures $\{P_n\}_{n\in\mathbb{N}} = \left\{P_n^{(\omega)}\right\}_{n\in\mathbb{N}}$ are such that each probability measure P_n in the sequence depends explicitly on events $\omega \in \Omega$ from a probability space $(\Omega, \mathscr{A}, \mathbf{P})$. This dependency introduces variability into the sequence, and allows to consider occasions in which the probability measures calculated are uncertain by themselves.

Example 1 (Empirical probability measures)**.** Let (E, d_E) be a metric space and P a probability measure on $(E, \mathscr{E})$. Let $X_1, \ldots, X_n$ be a random sample from P. We define the empirical probability measure corresponding to the sample $X_1, \ldots, X_n$ as

$$\mathbb{P}_n^{\text{emp}}(A) := \frac{1}{n} \sum_{i=1}^{n} 1_A(X_i), \quad \text{for } A \subseteq E.$$

For each $n \in \mathbb{N}$, let us draw a random element X_n from P, add it to the sample $X_1, \ldots, X_{n-1}$, and calculate its empirical probability measure $\mathbb{P}_n^{\text{emp}}$. Then, the sequence $\left\{\mathbb{P}_n^{\text{emp}}\right\}_{n\in\mathbb{N}}$ is a random sequence of probability measures.

As their deterministic counterpart, random sequences of probability measures can converge to a limit, either deterministic or random. A deterministic limit indicates that, although the sequence is random, it stabilizes to a fixed measure as n increases, regardless of the realization of the process. Conversely, a random limit suggests that the variability remains and the limit depends on the realization itself. In the following definitions, we introduce $\mathbf{P}$-almost sure weak convergence and $\ell^\infty(\mathscr{F})$-convergence of random sequences of probability measures, extending the already known definitions for deterministic sequences.

Definition 11 (**P**-almost sure weak convergence of random sequences of probability measures)**.** Let $\{P_n\}_{n\in\mathbb{N}} := \left\{P_n^{(\omega)}\right\}_{n\in\mathbb{N}}$ be a random sequence of probability measures on $(E,\mathscr{E})$ from the probability space $(\Omega,\mathscr{A},\mathbf{P})$. We define **P**-almost sure weak convergence (**P**-a.s. weak convergence) to the random probability measure $P := P^{(\omega)}$ on $(E,\mathscr{E})$ if

$$\mathbf{P}\left(P_n \rightsquigarrow P\right) = \mathbf{P}\left(\omega\in\Omega : P_n^{(\omega)} \rightsquigarrow P^{(\omega)}\right) = 1$$

holds, and we write $P_n \rightsquigarrow P$ **P**-a.s.

Definition 12 (**P**-almost sure $\ell^\infty(\mathscr{F})$-convergence of random sequences of probability measures)**.** Let $\{P_n\}_{n\in\mathbb{N}} := \left\{P_n^{(\omega)}\right\}_{n\in\mathbb{N}}$ be a random sequence of probability measures on $(E,\mathscr{E})$ from the probability space $(\Omega,\mathscr{A},\mathbf{P})$. We define **P**-almost sure convergence in $\ell^\infty(\mathscr{F})$-sense to the random probability measure $P := P^{(\omega)}$ on $(E,\mathscr{E})$ if

$$\mathbf{P}\left(P_n \xrightarrow{\ell^\infty(\mathscr{F})} P\right) = \mathbf{P}\left(\omega\in\Omega : P_n^{(\omega)} \xrightarrow{\ell^\infty(\mathscr{F})} P^{(\omega)}\right) = 1$$

holds, and we write $P_n \xrightarrow{\ell^\infty(\mathscr{F})} P$ **P**-a.s.

Remark 8. When the limit probability measure P is non-random, we have by definition that $P^{(\omega_1)} = P^{(\omega_2)}$ for all $\omega_1,\omega_2\in\Omega$.

To become more familiar with the concept of **P**-almost sure weak convergence of probability measures, we introduce the following simple example regarding the convergence of normal distributions.

Example 2 (**P**-almost sure weak convergence of normal distribution)**.** Let $X_1, X_2, \ldots$ be independent draws from a (true) normal distribution $P_X^{\text{true}} := N(\mu,1)$. For each $n\in\mathbb{N}$, let us consider the sample mean $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ as an approximation of the true mean, and define $P_{X,n} := N\left(\overline{X}_n,1\right)$. By the strong law of large numbers, $\overline{X}_n$ converges **P**-almost surely to μ, i.e. $\mathbf{P}\left(\overline{X}_n \to \mu\right) = 1$. On the other hand, if $\{a_n\}_{n\in\mathbb{N}}$ is a sequence of numbers converging to a, then the cumulative distribution function $\Phi(x-a_n)$ of $N(a_n,1)$ converges to $\Phi(x-a)$ as $a_n\to a$, which is equivalent to $N(a_n,1)\rightsquigarrow N(a,1)$ in the unidimensional case. Thus, we have that $\mathbf{P}\left(P_{X,n}\rightsquigarrow P_X^{\text{true}}\right) = 1$, giving the **P**-almost sure weak convergence of the estimated distribution $P_{X,n}$ to the true distribution P_X^{true}.

As will be seen in the subsequent chapters, the concept of random sequences of probability measures is particularly useful when building estimators for probability measures based on sample paths of a stochastic process.

2.1.5. Stochastic processes & time series

On certain occasions, our interest lies in collections of random variables from the same probability space that are indexed by a specific set. One such occasion arises when the variables are indexed by a set of time points and we aim to understand the evolution of phenomena over time. To address this, we explore the fundamental concepts of stochastic processes and time series. These frameworks are essential for modeling and interpreting the behavior of time-dependent phenomena effectively.

Definition 13 (Stochastic process)**.** Let $\mathscr{T}$ be a set. A stochastic process is a family of random variables $\{X_t\}_{t\in\mathscr{T}}$ defined on a probability space $(\Omega,\mathscr{A},\mathbf{P})$. The functions $\left\{X_\cdot^{(\omega)},\omega\in\Omega\right\}$ in $\mathscr{T}$ are known as the realizations or sample-paths of the process $\{X_t\}_{t\in\mathscr{T}}$.

Special cases of stochastic processes are time series, in which the set $\mathscr{T}$, called index or parameter set, is a set of time points. Let us formally define this concept.

Definition 14 (Time series)**.** A time series is a stochastic process in which the index set $\mathcal{T}$ is a set of time points. Discrete-time series are those in which the set of times $\mathcal{T}$ is a discrete set, while continuous-time series are those in which $\mathcal{T}$ is continuous (i.e., $\mathcal{T} = (t_a, t_b) \subseteq \mathbb{R}$).

Remark 9. For discrete-time series, we often take $\mathcal{T} = \mathbb{N}$ or $\mathcal{T} = \mathbb{Z}$.

Modeling time-dependent phenomena as time series allows for uncertainty in their nature. In this way, we treat every observed value at time t as a specific realization of the random variable X_t. In this thesis, we focus exclusively on discrete-time series, and therefore we will simply refer to them as time series.

We now provide the definitions of the distribution functions of stochastic processes.

Definition 15 (Distribution function of stochastic processes with $\mathcal{T} \subseteq \mathbb{R}$)**.** Let $\{X_t\}_{t\in\mathcal{T}\subseteq\mathbb{R}}$ be a stochastic process and let

$$\mathrm{T} := \left\{\mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathcal{T}^n : t_1 < t_2 < \ldots < t_n,\ n = 1, 2, \ldots\right\}.$$

Then, the (finite-dimensional) distribution functions of $\{X_t\}_{t\in\mathcal{T}}$ are the functions $\{F_{\mathbf{t}}(\cdot)\}_{\mathbf{t}\in\mathrm{T}}$ defined for $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ by

$$F_{\mathbf{t}}(\mathbf{x}) = \mathbf{P}\left(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \ldots, X_{t_n} \leq x_n\right), \quad \mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n.$$

To gain insights into the dependence of the random variables that compose a stochastic process, we define the autocovariance function of a stochastic process, which extends the concept of covariance matrix for a finite number of random variables.

Definition 16 (Autocovariance function)**.** Let $\{X_t\}_{t\in\mathcal{T}}$ be a process such that $Var[X_t] < \infty$ for all $t \in \mathcal{T}$. Then, the autocovariance function $\gamma_X(\cdot,\cdot)$ of $\{X_t\}_{t\in\mathcal{T}}$ is defined by

$$\gamma_X(r, s) = Cov[X_r, X_s] = \mathbb{E}[(X_r - \mathbb{E}[X_r])(X_s - \mathbb{E}[X_s])], \quad r, s \in \mathcal{T}.$$

With the autocovariance function, we can introduce the fundamental concept of weak stationarity.

Definition 17 (Weak stationarity)**.** Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stochastic process with index set $\mathbb{Z} = \{0, \pm1, \pm2, \ldots\}$ and let $m \in \mathbb{R}$ be a constant. This process is said to be weakly stationary if

1. $\mathbb{E}\left[|X_t|^2\right] < \infty$ for all $t \in \mathbb{Z}$,
2. $\mathbb{E}[X_t] = m$ for all $t \in \mathbb{Z}$, and
3. $\gamma_X(r, s) = \gamma_X(r + t, s + t)$ for all $r, s, t \in \mathbb{Z}$.

If $\{X_t\}_{t\in\mathbb{Z}}$ is weakly stationary, then $\gamma_X(r, s) = \gamma_X(r - s, 0)$ for all $r, s \in \mathbb{Z}$, and so we redefine the covariance function of a weakly stationary process as

$$\gamma_X(h) := \gamma_X(h, 0) = Cov[X_{t+h}, X_t], \quad \text{for all } t, h \in \mathbb{Z}.$$

The value h is often referred to as lag. The autocorrelation function of $\{X_t\}_{t\in\mathbb{Z}}$ is defined analogously as the function whose value at lag h is

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)} = Corr[X_{t+h}, X_t], \quad \text{for all } t, h \in \mathbb{Z}.$$

Finally, we define another important sense of stationarity, which is usually called strict stationarity.

Definition 18 (Strict stationarity)**.** The process $\{X_t\}_{t\in\mathbb{Z}}$ is said to be strictly stationary if the joint distributions of $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ and $(X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h})$ are the same for all positive integers k and for all $t_1, t_2, \ldots, t_k, h \in \mathbb{Z}$. This is equivalent to the statement that $(X_1, X_2, \ldots, X_k)$ and $(X_{1+h}, X_{2+h}, \ldots, X_{k+h})$ have the same joint distributions for all positive integers k and integers h.

Remark 10. A strictly stationary process with finite second moments is weakly stationary.

Remark 11. If X_{t_1} and X_{t_2} are independent for $t_1 \neq t_2$ and all X_t for $t \in \mathbb{Z}$ are identically distributed, then $\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary.

2.2. Properties of estimators

In this section, we state several results on consistency, rate of convergence, and asymptotic distribution of various estimators. The definition of M-estimators, the convergence in probability of M-estimators (Theorems 2 and 3), the definition of Hadamard-differentiable maps (Definition 19), the Delta Method (Theorem 4), the second and third results on rate of convergence of M-estimators (Theorem 6 and Corollary 1), the first and second results on asymptotic normality of M-estimators (Theorems 7 and 8), and the asymptotic distribution of empirical quantiles (Corollary 2) are based on van der Vaart [22]. The definition of asymptotic M-estimators, and the strong consistency of asymptotic M-estimators (Lemmas 1 and 2) are based on Chafaı̈ & Concordet [5]. The first result on rate of convergence of M-estimators (Theorem 5), the normality under regular maps (Lemma 3), the definition of covering and bracketing numbers (Definitions 20 and 21), and the third result on asymptotic normality of M-estimators (Theorem 9) are extracted from van der Vaart & Wellner [23]. Finally, the rate of convergence of empirical quantiles (Theorem 10) is based on Serfling [18]. For the proofs and deeper discussions on the results we refer to those references.

Let $(\Omega, \mathscr{A}, \mathbf{P})$ be a probability space and (Θ, μ) a parameter space, which we assume to be a metric space. Suppose we are interested in estimating the distribution of a random variable X from a sample $X_1, \ldots, X_n$. One method to achieve this is by considering a statistical model $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ for the distribution and finding an estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ that maximizes a criterion function $M_n(\theta)$ over Θ, i.e.,

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\arg\max}\, M_n(\theta). \tag{2.2}$$

An estimator of this type is called a M-estimator. In a similar way, we say that $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ is a sequence of asymptotic M-estimators if and only if

$$\overline{\lim_{n\to\infty}}\left(\sup_{\Theta} M_n - M_n(\hat{\theta}_n)\right) = 0 \quad \mathbf{P}\text{-a.s.} \tag{2.3}$$

Note that when the maximum in (2.2) exists then it is equal to the supremum, and thus the parenthesis in (2.3) is exactly zero. In other words, M-estimators are a special case of asymptotic M-estimators.

2.2.1. Consistency

Consistency of estimators is a desirable property in statistical inference, representing the guarantee that as the sample size tends to infinity, the estimator converges to the true parameter value. This property is essential for assessing the reliability and accuracy of inference based on estimated parameters. Since there are various types of convergence of random variables, there exist corresponding types of consistency of estimators.

We first state two lemmas about the strong consistency (i.e. convergence **P**-almost surely to the true value) of asymptotic M-estimators.

Lemma 1. *Assume that for any neighborhood U of $\theta_0 \in \Theta$, for any sequence $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ in U^c, there exists a sequence $\{\theta'_n\}_{n\in\mathbb{N}}$ in Θ such that*

$$\varliminf_{n\to\infty} \left(M_n\left(\theta'_n\right) - M_n\left(\hat{\theta}_n\right)\right) > 0 \quad \mathbf{P}\text{-a.s.}$$

Then, any asymptotic M-estimators sequence $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ is strongly consistent.

Lemma 2. *Assume that Θ is compact and that there exists a map $a^* : \Theta \to \Theta$ such that for any $\theta \neq \theta_0$ in Θ, there exists a neighborhood U_θ of θ such that*

$$\varliminf_{n\to\infty} \inf_{U_\theta} \left(M_n\left(a^*\right) - M_n\right) > 0 \quad \mathbf{P}\text{-a.s.}$$

Then, any asymptotic M-estimators sequence $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ is strongly consistent.

Similarly, the following theorem provides straightforward conditions for convergence in probability of M-estimators. This notion of convergence in probability offers a slightly weaker guarantee compared to strong consistency, since it can be proven that **P**-almost sure convergence implies convergence in probability, but not vice versa. Nonetheless, it remains a crucial property, particularly in cases where strong consistency may not be achievable due to the complexity of the underlying model or the estimator.

Theorem 2. *Let $\{M_n\}_{n\in\mathbb{N}}$ be random functions and let M be a fixed function of θ such that for every $\epsilon > 0$*

$$\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \xrightarrow{\mathbf{P}} 0,$$

$$\sup_{\theta:\mu(\theta,\theta_0)\geq\epsilon} M(\theta) < M(\theta_0).$$

Then any sequence of estimators $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ with $M_n\left(\hat{\theta}_n\right) \geq M_n(\theta_0) - o_{\mathbf{P}}(1)$ converges in probability to θ_0.

We finally present a useful result for the common case when M_n and M are of the form

$$M_n(\theta) = \mathbb{P}_n^{\text{emp}} m_\theta := \frac{1}{n}\sum_{i=1}^{n} m_\theta(X_i), \quad M(\theta) = P m_\theta := \int m_\theta \, dP,$$

for m_θ known real-valued functions and a probability measure P. To this end, let us recall that a function $f : T \to \mathbb{R}$ from a topological space T is upper-semicontinuous at x_0 if and only if $\varlimsup_{x\to x_0} f(x) \leq f(x_0)$.

Theorem 3. *Let $\theta \mapsto m_\theta(x)$ be upper-semicountinuous for **P**-almost all x and assume that for every sufficiently small ball $U \subseteq \Theta$ the function $x \mapsto \sup_{\theta\in U} m_\theta(x)$ is measurable and satisfies*

$$P \sup_{\theta\in U} m_\theta < \infty. \tag{2.4}$$

Let $\Theta_0 := \left\{\theta_0 \in \Theta : P m_{\theta_0} = \sup_{\theta\in\Theta} P m_\theta\right\}$, which we assume not empty. Then, for any estimator $\hat{\theta}_n$ such that $M_n\left(\hat{\theta}_n\right) \geq M_n(\theta_0) - o_{\mathbf{P}}(1)$ for some $\theta_0 \in \Theta_0$, for every $\epsilon > 0$ and every compact set $K \subseteq \Theta$,

$$\mathbf{P}\left(\mu\left(\hat{\theta}_n, \Theta_0\right) \geq \epsilon \wedge \hat{\theta}_n \in K\right) \to 0.$$

As will be seen later, consistency of other estimators may still be proved, for instance when the functions defining them are sufficiently regular.

2.2.2. Convergence rates

We now present results concerning the rate of convergence of estimators. Obtaining the rate of convergence is important to assess the speed at which an estimator approaches the true parameter value as the sample size increases. Naturally, a faster rate of convergence implies a more accurate estimation, which is definitely a desirable property.

As with consistency, most results are exclusive for M-estimators. However, we also provide conditions to estimate the rate of convergence of sufficiently regular maps. To do so, let us first introduce the concept of Hadamard-differentiability, which will be linked to the useful Delta Method.

Definition 19 (Hadamard-differentiable map)**.** Let $\mathbb{D}$ and $\mathbb{V}$ be normed vector spaces. A map $\phi : \mathbb{D}_\phi \mapsto \mathbb{V}$, defined on a subset $\mathbb{D}_\phi$ of $\mathbb{D}$ that contains z is called Hadamard-differentiable at z if there exists a continuous, linear map $\phi'_z : \mathbb{D} \mapsto \mathbb{V}$ such that

$$\left\| \frac{\phi(z+th_t) - \phi(z)}{t} - \phi'_z(h) \right\|_{\mathbb{V}} \to 0, \quad \text{as } t \downarrow 0, \text{ for every } h_t \to h. \tag{2.5}$$

Remark 12. The above definition requires that $\phi'_z : \mathbb{D} \mapsto \mathbb{V}$ exists as a map on the entire space $\mathbb{D}$. If ϕ'_z only exists on a subset $\mathbb{D}_0 \subseteq \mathbb{D}$ and the sequences $h_t \to h$ are restricted to converge to limits $h \in \mathbb{D}_0$, then ϕ is called Hadamard-differentiable tangentially to $\mathbb{D}_0$.

It can be proven that Hadamard-differentiability is equivalent to the difference in condition (2.5) converging uniformly to zero for h in compact subsets of $\mathbb{D}$. Notably, weak convergence in metric spaces is intimately linked to compact sets. Therefore, the notion of Hadamard-differentiability emerges as the appropriate form of differentiability to be explored in the following theorem.

Theorem 4 (Delta Method)**.** *Let $\mathbb{D}$ and $\mathbb{V}$ be normed linear spaces. Let $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \mapsto \mathbb{V}$ be Hadamard-differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_0 \subseteq \mathbb{D}$. Let $\{(\Omega_n, \mathscr{A}, \mathbf{P})\}_{n\in\mathbb{N}}$ a sequence of probability spaces, and let $T_n : \Omega_n \mapsto \mathbb{D}_\phi$ be maps such that $r_n(T_n - \theta) \rightsquigarrow T$ for some sequence of numbers $\{r_n\}_{n\in\mathbb{N}}$ such that $r_n \to \infty$ and a random element T that takes its values in $\mathbb{D}_0$. Then $r_n\left(\phi(T_n) - \phi(\theta)\right) \rightsquigarrow \phi'_\theta(T)$. If ϕ'_θ is defined and continuous on the whole space $\mathbb{D}$, then we also have $r_n\left(\phi(T_n) - \phi(\theta)\right) = \phi'_\theta(r_n(T_n - \theta)) + o_{\mathbf{P}}(1)$.*

The Delta Method implies that estimators derived from estimators with known rate of convergence through Hadamard-differentiable maps inherit their rate of convergence.

The following result addresses the interesting case when the estimator belongs to the family of M-estimators.

Theorem 5. *Let $\{M_n\}_{n\in\mathbb{N}}$ be stochastic processes indexed by a semimetric space (Θ, μ), i.e. μ satisfies conditions 2 and 3 in Definition 1. Let the notation $\lesssim$ read as "bounded above up to a universal constant", and let $M : \Theta \to \mathbb{R}$ be a deterministic function, such that for every θ in a neighborhood of θ_0,*

$$M(\theta) - M(\theta_0) \lesssim -\mu^2(\theta, \theta_0).$$

Suppose that, for every n and sufficiently small δ, the centered process $\{M_n - M\}_{n\in\mathbb{N}}$ satisfies

$$\mathbb{E}^* \left[\sup_{\mu(\theta,\theta_0)<\delta} |(M_n - M)(\theta) - (M_n - M)(\theta_0)| \right] \lesssim \frac{\phi_n(\delta)}{\sqrt{n}},$$

for functions ϕ_n such that $\delta \mapsto \frac{\phi_n(\delta)}{\delta^\alpha}$ is decreasing for some $\alpha < 2$ (not depending on n). Let

$$r_n^2 \phi_n\left(\frac{1}{r_n}\right) \leq \sqrt{n},$$

for every n. If the sequence $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ satisfies $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - O_{\mathbf{P}}(r_n^{-2})$ and converges in outer probability to θ_0, then $r_n \mu(\hat{\theta}_n, \theta_0) = O_{\mathbf{P}}^*(1)$. If the displayed conditions are valid for every θ and δ, then the condition that $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ is consistent is unnecessary.

In the common case where the criterion function is of the form $M_n(\theta) := \mathbb{P}_n^{\text{emp}} m_\theta$ and $M(\theta) := P m_\theta$, there are several results available to obtain the rates of convergence. We present a theorem and a corollary that do not require i.i.d. samples, as the draws considered in this work are neither necessarily independent nor identically distributed.

Theorem 6. *Assume that for fixed constants C and $\alpha > \beta$, for every n, and for every sufficiently small $\delta > 0$,*

$$\sup_{\mu(\theta,\theta_0)<\delta} P(m_\theta - m_{\theta_0}) \leq -C\delta^\alpha,$$
$$E^*\left[\sup_{\mu(\theta,\theta_0)<\delta} \left|\mathbb{G}_n(m_\theta - m_{\theta_0})\right|\right] \leq C\delta^\beta,$$

where $\mathbb{G}_n$ denotes the empirical process defined as $\mathbb{G}_n m_\theta := \sqrt{n}(\mathbb{P}_n^{emp} m_\theta - P m_\theta)$. If the sequence $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ satisfies $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - O_{\mathbf{P}}\left(n^{\frac{\alpha}{2\beta-2\alpha}}\right)$ and converges in outer probability to θ_0, then $n^{\frac{1}{2\alpha-2\beta}} \mu(\hat{\theta}_n, \theta_0) = O_{\mathbf{P}}^*(1)$.

In the simple case where there exists a Lipschitz condition on the maps $\theta \mapsto m_\theta$, it is possible to prove that the rate of convergence is $O(\sqrt{n})$, as stated in the following corollary.

Corollary 1. *For each θ in an open subset of Euclidean space, let $x \mapsto m_\theta(x)$ be a measurable function such that, for every θ_1 and θ_2 in a neighborhood of θ_0 and a measurable function $\dot{m}$ such that $P\dot{m}^2 < \infty$,*

$$\left|m_{\theta_1}(x) - m_{\theta_2}(x)\right| \leq \dot{m}(x) \left\|\theta_1 - \theta_2\right\|.$$

Furthermore, suppose that the map $\theta \mapsto P m_\theta$ admits a second-order Taylor expansion at the point of maximum θ_0 with non-singular second derivative. If $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - O_{\mathbf{P}}(n^{-1})$ then $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_{\mathbf{P}}(1)$, provided that $\hat{\theta}_n \xrightarrow{\mathbf{P}} \theta_0$.

2.2.3. Asymptotic distribution

As was the case for consistency and rate of convergence, the asymptotic distribution of estimators plays an important role in statistical inference, providing insights into the behavior of the estimator as the sample size grows. It characterizes the variability of the estimator around the true value, and allows the construction of confidence intervals and tests of hypotheses.

We start by presenting a useful result to prove that normality is retained under sufficiently regular maps.

Lemma 3. *Let $X := \{X_t\}_{t\in T} \in \ell^\infty(T)$ be a tight Borel-measurable Gaussian map and let $\mathbb{V}$ be a Banach space. Then $\phi(X)$ is normally distributed for every continuous, linear map $\phi : \ell^\infty(T) \to \mathbb{V}$.*

Remark 13. Recall that the process X is called a Gaussian process if and only if $(X_{t_1}, \ldots, X_{t_k})$ is multivariate normally distributed for every $k \in \mathbb{N}$ and finite set $t_1, \ldots, t_k$ in T.

The following theorem addresses the asymptotic distribution of general M-estimators.

Theorem 7. *Let $\{M_n\}_{n\in\mathbb{N}}$ be stochastic processes indexed by an open subset Θ of Euclidean space and $M:\Theta\to\mathbb{R}$ a deterministic function. Assume that $\theta\mapsto M(\theta)$ is twice continuously differentiable at a point of maximum $\theta_0\in\Theta$ with non-singular second-derivative matrix V. Suppose that*

$$\begin{aligned}&r_n(M_n-M)(\tilde\theta_n)-r_n(M_n-M)(\theta_0)\\&=(\tilde\theta_n-\theta_0)'Z_n+o_{\mathbf{P}}^*\left(\|\tilde\theta_n-\theta_0\|+r_n\|\tilde\theta_n-\theta_0\|^2+r_n^{-1}\right),\end{aligned}$$

for every random sequence $\{\tilde\theta_n\}_{n\in\mathbb{N}}$ such that $\tilde\theta_n=\theta_0+o_{\mathbf{P}}^*(1)$, a uniformly tight sequence of random vectors $\{Z_n\}_{n\in\mathbb{N}}$, and a sequence of numbers $\{r_n\}_{n\in\mathbb{N}}$. If the sequence $\{\hat\theta_n\}_{n\in\mathbb{N}}$ converges in outer probability to θ_0 and satisfies $M_n(\hat\theta_n)\geq\sup_\theta M_n(\theta)-o_{\mathbf{P}}(r_n^{-2})$ for every n, then

$$r_n(\hat\theta_n-\theta_0)=-V^{-1}Z_n+o_{\mathbf{P}}^*(1).$$

If it is known that the sequence $\{r_n(\hat\theta_n-\theta_0)\}_{n\in\mathbb{N}}$ is uniformly tight, then the displayed condition needs to be verified for sequences $\{\tilde\theta_n\}_{n\in\mathbb{N}}$ such that $\tilde\theta_n=\theta_0+O_{\mathbf{P}}^*(r_n^{-1})$ only.

The following two theorems work for criterion functions of the form $M_n(\theta):=\mathbb{P}_n^{\text{emp}}m_\theta$ and $M(\theta):=Pm_\theta$.

Theorem 8. *For each θ in an open subset of Euclidean space let $x\mapsto m_\theta(x)$ be a measurable function such that $\theta\mapsto m_\theta(x)$ is differentiable at θ_0 for $\mathbf{P}$-almost every x with derivative $\dot m_{\theta_0}(x)$ and such that, for every θ_1 and θ_2 in a neighborhood of θ_0 and a measurable function $\dot m$ with $P\dot m^2<\infty$,*

$$|m_{\theta_1}(x)-m_{\theta_2}(x)|\leq\dot m(x)\|\theta_1-\theta_2\|.$$

Furthermore, assume that the map $\theta\mapsto Pm_\theta$ admits a second-order Taylor expansion at a point of maximum θ_0 with non-singular symmetric second derivative matrix V_{θ_0}. If $M_n(\hat\theta_n)\geq\sup_\theta M_n(\theta)-o_{\mathbf{P}}(n^{-1})$ and $\hat\theta_n\xrightarrow{\mathbf{P}}\theta_0$, then

$$\sqrt{n}(\hat\theta_n-\theta_0)=-V_{\theta_0}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\dot m_{\theta_0}(X_i)+o_{\mathbf{P}}(1).$$

In particular, the sequence $\{\sqrt{n}(\hat\theta_n-\theta_0)\}_{n\in\mathbb{N}}$ is asymptotically normal with mean zero and covariance matrix

$$V_{\theta_0}^{-1}P\dot m_{\theta_0}\dot m_{\theta_0}^TV_{\theta_0}^{-1}.$$

Before stating the next theorem, we introduce the concepts of covering and bracketing numbers. Let $(\mathscr{F},\|\cdot\|)$ be a subset of a normed space of real functions $f:\mathscr{X}\to\mathbb{R}$ on some set.

Definition 20 (Covering number)**.** The covering number $N(\epsilon,\mathscr{F},\|\cdot\|)$ is the minimal number of balls $\{g:\|g-f\|<\epsilon\}$ of radius ϵ needed to cover the set $\mathscr{F}$. The centers of the balls need not belong to $\mathscr{F}$, but they should have finite norms. The entropy (without bracketing) is the logarithm of the covering number.

Definition 21 (Bracketing number)**.** Given two functions l and u, the bracket $[l,u]$ is the set of all functions f with $l\leq f\leq u$. An ϵ-bracket is a bracket $[l,u]$ with $\|u-l\|<\epsilon$. The bracketing number $N_{[]}(\epsilon,\mathscr{F},\|\cdot\|)$ is the minimum number of ϵ-brackets needed to cover $\mathscr{F}$. The entropy with bracketing is the logarithm of the bracketing number. The upper and lower bounds u and l of the brackets need not belong to $\mathscr{F}$ but are assumed to have finite norms.

For the following theorem, assume that either the uniform entropy or the bracketing integrals of the classes $\mathscr{M}_\delta = \{m_\theta - m_{\theta_0} : \mu(\theta,\theta_0) < \delta\}$ are uniformly bounded as δ tends to zero. Denoting the $L_2(P)$-norm of a function f with respect to the probability measure P as $\|f\|_{P,2} = \left(P|f|^2\right)^{\frac{1}{2}}$, these conditions can be written as

$$\int_0^\infty \sup_{\delta<\delta_0} \sup_Q \sqrt{\log N\left(\epsilon \|M_\delta\|_{Q,2}, \mathscr{M}_\delta, L_2(Q)\right)}\, d\epsilon < \infty \tag{2.6}$$

for the uniform entropy, and

$$\int_0^\infty \sup_{\delta<\delta_0} \sqrt{\log N_{[]}\left(\epsilon \|M_\delta\|_{P,2}, \mathscr{M}_\delta, L_2(P)\right)}\, d\epsilon < \infty \tag{2.7}$$

for the bracketing integral, where M_δ is an envelope function of $\mathscr{M}_\delta$, i.e. a function such that $|m(x)| \le M_\delta(x) < \infty$ for every $m \in \mathscr{M}_\delta$ and x in the domain of m.

Theorem 9. *For each θ in an open subset of Euclidean space, let m_θ be a measurable function such that $\theta \mapsto Pm_\theta$ is twice continuously differentiable at a point of maximum θ_0, with non-singular second derivative matrix V. Let the conditions (2.6) or (2.7) hold. Assume that for some continuous function ϕ, such that $\phi^2(\delta) \ge P^* M_\delta^2$ and such that $\delta \mapsto \frac{\phi(\delta)}{\delta^\alpha}$ is decreasing for some $\alpha < 2$, and for every $\eta > 0$,*

$$\lim_{\delta\downarrow 0} \frac{P^* M_\delta^2 \{M_\delta > \eta\delta^{-2}\phi^2(\delta)\}}{\phi^2(\delta)} = 0,$$

$$\lim_{\epsilon\downarrow 0} \limsup_{\delta\downarrow 0} \sup_{\|h-g\|<\epsilon, \|h\|\vee\|g\|\le K} \frac{P\left(m_{\theta_0+\delta g} - m_{\theta_0+\delta h}\right)^2}{\phi^2(\delta)} = 0,$$

$$\lim_{\delta\downarrow 0} \frac{P\left(m_{\theta+\delta g} - m_{\theta_0+\delta h}\right)^2}{\phi^2(\delta)} = \mathbb{E}\left[\left(G(g) - G(h)\right)^2\right],$$

for all K and some zero-mean Gaussian process G such that $G(g) = G(h)$ $\mathbf{P}$-almost surely only if $h = g$. Then there exists a version of G (i.e. a process G^* such that $\mathbf{P}\left(G(g) = G^*(g), \forall g\right) = 1$) with bounded, uniformly continuous sample paths on compacta. Define r_n as the solution of $r_n^2 \phi\left(\frac{1}{r_n}\right) = \sqrt{n}$. If $\hat\theta_n$ nearly maximizes the map $\theta \mapsto \mathbb{P}_n^{emp} m_\theta$ for every n and converges in outer probability to θ_0, then the sequence $\{r_n(\hat\theta_n - \theta_0)\}_{n\in\mathbb{N}}$ converges in distribution to the unique maximizer $\hat h$ of the process $h \mapsto G(h) + \frac{1}{2}h'Vh$.

2.2.4. Asymptotic distribution of quantiles

Finally, to end this section we address the rate of convergence and asymptotic distribution of empirical quantiles, which are estimators for the true quantiles of a distribution. Let F be the underlying distribution function of a random sample $X_1, X_2, \ldots, X_n$ and $\mathbb{F}_n$ its empirical distribution function defined as

$$\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \le t).$$

The order statistics $X_{n(1)} \le X_{n(2)} \le \ldots \le X_{n(n)}$ of the sample are its values positioned in increasing order. Since the sample is random, the order statistics are also random and depend on each event.

Let $F^{-1} : (0,1) \to \mathbb{R}$ be the generalized inverse of F given by

$$F^{-1}(p) = \inf\{x : F(x) \ge p\},$$

which is also called the quantile function of F. We define the empirical quantile function $\mathbb{F}_n^{-1}$ as

$$\mathbb{F}_n^{-1}(p) = X_{n(i)}, \quad \text{for } p \in \left(\frac{i-1}{n}, \frac{i}{n}\right].$$

The following corollary shows that, under certain regularity conditions, the convergence rate of the empirical quantile is $O(\sqrt{n})$. Moreover, the scaled error of the empirical quantile is asymptotically normally distributed.

Corollary 2. *Fix $0 < p < 1$. If F is differentiable at $F^{-1}(p)$ with positive derivative $f(F^{-1}(p))$, then*

$$\sqrt{n}\left(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\right) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\mathbb{1}\left(X_i \leq F^{-1}(p)\right) - p}{f\left(F^{-1}(p)\right)} + o_{\mathbf{P}}(1).$$

Consequently, the sequence $\left\{\sqrt{n}\left(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\right)\right\}_{n\in\mathbb{N}}$ is asymptotically normal with mean zero and variance

$$\frac{p(1-p)}{\left[f\left(F^{-1}(p)\right)\right]^2}.$$

On the other hand, the following theorem by Bahadur, provides a proper rate of convergence of the empirical quantile in Big-O notation.

Theorem 10. *Let $0 < p < 1$. Suppose that F is twice differentiable at $F^{-1}(p)$, with $F'\left(F^{-1}(p)\right) = f\left(F^{-1}(p)\right) > 0$. Then,*

$$\mathbb{F}_n^{-1}(p) = F^{-1}(p) + \frac{p - \mathbb{F}_n\left(F^{-1}(p)\right)}{f\left(F^{-1}(p)\right)} + R_n,$$

where, with probability one, $R_n = O\left(n^{-\frac{3}{4}}\left(\log(n)\right)^{\frac{3}{4}}\right)$. That is, there exists a set $\Omega_0 \in \Omega$ such that $\mathbf{P}(\Omega_0) = 1$ and for each $\omega \in \Omega_0$, there exists a constant $B(\omega)$ such that

$$|R_n(\omega)| \leq B(\omega)\, n^{-\frac{3}{4}}\left(\log(n)\right)^{\frac{3}{4}}$$

for all sufficiently large n. In this particular theorem, the constants $B(\omega)$ may be chosen not to depend on ω.

2.3. Copulas and Copula-GARCH model

Let $(\Omega, \mathscr{A}, \mathbf{P})$ be an arbitrary probability space, and let us consider a set of real-valued random variables $X_1, \ldots, X_d$ with marginal cumulative distribution functions $F_j(x_j) = \mathbf{P}(X_j \leq x_j)$ for $j = 1, \ldots, d$, and joint cumulative distribution function (CDF) given by $H(x_1, \ldots, x_d) = \mathbf{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$. When modeling the distribution of these random variables, it is common to encounter situations where the marginals F_j are relatively easy to describe, while an explicit expression for the joint distribution H may be difficult to obtain [12]. In such cases, the use of copulas to capture the dependency structure is convenient and often used in practice. Thus, in this section we introduce this powerful concept following Durante & Sempi [9]. Subsequently, to introduce the Copula-GARCH model, we provide the definitions of ARMA(p,q) and GARCH(p,q) processes (Definitions 26 and 27), which are based on Shumway & Stoffer [19], and we define ARMA(p_1,q_1)-GARCH(p_2,q_2) processes (Definition 28) following Ghani & Rahim [10].

For ease of notation, during this section we consider the concept of random vectors. A d-dimensional

random vector $\mathbf{X}$ is a measurable mapping from Ω into $\mathbb{R}^d$. Therefore, a random vector can be represented as $\mathbf{X} = (X_1, \ldots, X_d)$, where $X_1, \ldots, X_d$ are one-dimensional random variables. We also denote the unit interval as $\mathbb{I} := [0,1]$.

Definition 22. For every $d \geq 2$, a d-dimensional copula (a d-copula) is a d-dimensional CDF concentrated on $\mathbb{I}^d$ whose univariate marginals are uniformly distributed on $\mathbb{I}$. An equivalent statement that takes into account the properties of distribution functions can be given. In particular, a function $C : \mathbb{I}^d \to \mathbb{I}$ is a d-copula if and only if the following hold:

- $C(u_1, \ldots, u_d) = 0$ if $u_j = 0$ for at least one index $j \in \{1, \ldots, d\}$;
- C has uniform margins, i.e. $C(1, \ldots, 1, u_j, 1 \ldots, 1) = u_j$; and
- C is d-increasing, i.e. for every hyper-rectangle $A = \Pi_{j=1}^{d} [a_j, b_j] \subseteq \mathbb{I}^d$, it holds that $\int_A \mathrm{d}C(\mathbf{u}) \geq 0$.

The versatility of copulas is not immediately apparent if we do not introduce the main results behind them. To this end, we provide an extremely useful result regarding the definition of multivariate cumulative distribution functions from copulas and univariate distribution functions.

Theorem 11. *Let $F_1, \ldots, F_d$ be univariate CDFs and let C be any d-copula. Then, the function $H : \mathbb{R}^d \to \mathbb{I}$ defined as*

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \tag{2.8}$$

for $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, is a d-dimensional distribution function with margins given by $F_1, \ldots, F_d$.

As mentioned in [9], this result suggests an approach to build multivariate models, which consists of first defining the marginal cumulative distribution functions with great flexibility (such as by considering different families of distributions), and then choosing a copula to link the marginals in a common model. This two-stage procedure is applied when estimating the parameters of a parametric copula-based stochastic model, i.e. a model where the marginals $F_j = F_{\boldsymbol{\theta}_j}$ and the copula $C = C_{\boldsymbol{\alpha}}$ depend on parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$, respectively. In such ocassions, the univariate marginals are estimated independently, while the copula parameters $\boldsymbol{\alpha}$ are estimated from Equation 2.8 either using the already estimated univariate parameters for the marginals or by replacing them by their empirical counterparts.

We now present Sklar's theorem, which is one of the main results regarding copulas. Durante & Sempi [9] provide several proofs of it.

Theorem 12 (Sklar's theorem)**.** *Let a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ be given on an arbitrary probability space $(\Omega, \mathscr{A}, \mathbf{P})$, let $H(\mathbf{x}) := \mathbf{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$ be the joint CDF of $\mathbf{X}$, and let $F_j(x_j) = \mathbf{P}(X_j \leq x_j)$ be its marginals $(j = 1, \ldots, d)$. Then, there exists a d-copula $C = C_{\mathbf{X}}$ such that, for every $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$,*

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{2.9}$$

If the marginals $F_1, \ldots, F_d$ are continuous, then the copula C is unique.

If H is absolutely continuous, then the density function of H is given by

$$h(\mathbf{x}) = c(F_1(x_1), \ldots, F_d(x_d)) f_1(x_1) \cdots f_d(x_d),$$

for $\mathbf{P}$-almost all $\mathbf{x} \in \mathbb{R}^d$, where

$$f_j(x) = \frac{d}{dx} F_j(x), \quad j = 1, 2, \ldots, d,$$

are the derivatives of $F_1,\dots,F_d$, and

$$c(\mathbf{u}) = \frac{\partial^d}{\partial u_1 \dots \partial u_d} C(u_1,\dots,u_d)$$

is the density function of the copula C.

We now state a lemma related to Sklar's theorem that provides a formula for the copula when the marginals are continuous.

Lemma 4. *Under the assumptions of Sklar's theorem, if $F_1,\dots,F_d$ are continuous, then there exists a unique copula C associated with $\mathbf{X}$ that is the CDF of the random vector $(F_1 \circ X_1,\dots,F_d \circ X_d)$. This copula is determined, for every $\boldsymbol{u} = (u_1,\dots,u_d) \in \mathbb{I}^d$, via the formula*

$$C(\mathbf{u}) = H\left(F_1^{-1}(u_1),\dots,F_d^{-1}(u_d)\right),$$

where, for $j \in \{1,\dots,d\}$, F_j^{-1} is the generalized inverse of F_j.

The previous lemma is crucial to define elliptical copulas, and subsequently to define the Gaussian and Student's t-copula, which are of interest in later chapters. We say that a random vector $\mathbf{X} = (X_1,\dots,X_d)$ has an elliptical distribution if it can be expressed in the form

$$\mathbf{X} \overset{d}{=} \mu + R\mathbf{A}\mathbf{U}, \tag{2.10}$$

where $\mu \in \mathbb{R}^d$, $\mathbf{A} \in \mathbb{R}^{d\times k}$ with $\Sigma := \mathbf{A}\mathbf{A}^T \in \mathbb{R}^{d\times d}$ and $rank(\Sigma) = k \le d$, $\mathbf{U}$ is a d-dimensional random vector uniformly distributed on $\{\mathbf{u} \in \mathbb{R}^d : u_1^2 + \dots + u_d^2 = 1\}$ and R is a positive random variable independent of $\mathbf{U}$.

We can now give the definition of elliptical copulas.

Definition 23 (Elliptical copula)**.** An elliptical copula is any copula that can be obtained from an elliptical distribution using the inversion method shown in Lemma 4.

A key aspect of general elliptical copulas is that there are methods available both for simulating them and for estimating the distribution of R. Two specific cases of elliptical copulas are the Gaussian copula and the Student's t-copula, the latter of which is of great importance in this thesis.

Definition 24 (Gaussian copula)**.** The Gaussian copula is the copula of an elliptical random vector $\mathbf{X}$ that follows a Gaussian distribution. This vector can be expressed as $\mathbf{X} \overset{d}{=} \mu + \mathbf{A}\mathbf{Z}$, where $\mathbf{Z}$ is a d-dimensional random vector whose independent components have univariate standard Gaussian law, and $\mathbf{A}$, Σ and μ are defined as in Equation (2.10). We say that $\mathbf{X} \sim N_d(\mu,\Sigma)$. The bivariate Gaussian copula can be expressed semi-analytically as

$$C_\rho^{\mathrm{Ga}}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} ds \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) dt,$$

where $\rho \in (-1,1)$ and Φ is the standard Gaussian CDF.

Definition 25 (Student's t-copula)**.** The Student's t-copula is the copula of an elliptical random vector $\mathbf{X}$ that follows a multivariate Student's t-distribution. This vector can be expressed as $\mathbf{X} \overset{d}{=} \mu + \Sigma^{\frac{1}{2}}\sqrt{W}\mathbf{Z}$, where $\mathbf{Z} \sim N_d(\mathbf{0},\mathbf{I}_d)$ and Σ is a positive definite matrix. In addition, W and $\mathbf{Z}$ are independent, and W follows an inverse Gamma distribution with parameters $\left(\frac{\nu}{2},\frac{\nu}{2}\right)$. In Chapter 4, the degrees-of-freedom of a bivariate Student's

t-copula are estimated. This copula can be expressed as

$$C^{\mathrm{t}}_{\rho,\nu}(u,v) = \mathbf{t}_{\rho,\nu}\left(t_\nu^{-1}(u), t_\nu^{-1}(v)\right),$$

where $\rho \in (-1,1)$ is the correlation and $\nu > 1$ are the degrees-of-freedom. Moreover, $\mathbf{t}_{\rho,\nu}$ denotes the bivariate Student's t-distribution with zero mean, correlation matrix with off-diagonal element equal to ρ, and ν degrees-of-freedom. On the other hand, t_ν^{-1} denotes the inverse of the standard t-distribution.

The use of copula models can be considered when multivariate stochastic models are built for random phenomena. Durante & Sempi [9] provide a few real-world situations that can be described by such models. These include Portfolio Management, in which each random variable can represent the return of an asset constituting a portfolio of investments at a certain time, and Hydrology, in which each random variable may represent quantities related to an environmental event (such as the duration and intensity of a storm).

2.3.1. Copula-GARCH models

Let us introduce the Copula-GARCH model, which is popular to model multivariate processes that depend on each other. To do so, we start by defining ARMA(p,q) processes, which are used to model the conditional mean of a process when the conditional variance is constant.

Definition 26 (ARMA(p,q) process)**.** Let $\{W_t\}_{t\in\mathbb{Z}}$ be a collection of uncorrelated random variables with mean zero and variance $\sigma_W^2 > 0$. A process $\{X_t\}_{t\in\mathbb{Z}}$ is an Autoregressive Moving Average process of order p,q, or ARMA(p,q), if it is stationary and

$$X_t = \gamma_0 + \gamma_1 X_{t-1} + \ldots + \gamma_p X_{t-p} + W_t + \eta_1 W_{t-1} + \ldots + \eta_q W_{t-q},$$

with $\gamma_p, \eta_q \neq 0$. The parameters p and q are called the autoregressive and the moving average orders, respectively. If X_t has mean zero, then $\gamma_0 = 0$; otherwise, for non-zero mean μ we set $\gamma_0 = \mu\left(1-\gamma_1-\ldots-\gamma_p\right)$.

Note that this type of process, as well as subsequent ones, provides a way to model the behavior of stochastic processes with unknown distributions. To this end, the parameters of the process must be estimated, typically through Maximum Likelihood Estimation or Least Squares Estimation.

In many cases, the assumption of constant conditional variance is not valid. The GARCH(p,q) model addresses this by allowing changes in volatility, such as clustered periods of high volatility.

Definition 27 (GARCH(p,q) process)**.** A process $\{X_t\}_{t\in\mathbb{Z}}$ is a Generalized Autoregressive Conditionally Heteroscedastic process of order p,q, or GARCH(p,q), if it satisfies

$$\begin{aligned} X_t &= \sigma_t Z_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2, \end{aligned}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ is an i.i.d. process with mean zero and variance one.

The ARMA(p_1,q_1)-GARCH(p_2,q_2) process is a combination of ARMA(p_1,q_1) and GARCH(p_2,q_2) processes. In such a process, the mean is modelled as an ARMA(p_1,q_1) process, and the variance is modelled as a GARCH(p_2,q_2) process.

Definition 28 (ARMA(p_1, q_1)-GARCH(p_2, q_2) process)**.** A process $\{X_t\}_{t\in\mathbb{Z}}$ is ARMA(p_1, q_1)-GARCH(p_2, q_2) if

$$
\begin{aligned}
X_t &= \mu_t + \epsilon_t, \\
\mu_t &= \gamma_0 + \gamma_1 X_{t-1} + \ldots + \gamma_{p_1} X_{t-p_1} + \eta_1 \epsilon_{t-1} + \ldots + \eta_{q_1} \epsilon_{t-q_1}, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_{p_2} \epsilon_{t-p_2}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_{q_2} \sigma_{t-q_2}^2, \\
\epsilon_t &= \sigma_t Z_t,
\end{aligned}
$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ is an i.i.d. process with mean zero and variance one.

Finally, the Copula-GARCH model is an approach designed for analyzing the dependence structures between multiple series. When fitting a Copula-GARCH model to d-dimensional data, each available time series is modeled individually using an ARMA(p_1, q_1)-GARCH(p_2, q_2) process. This step ensures that both the means and volatilities of each series are adequately captured. Common ways to fit ARMA(p_1, q_1)-GARCH(p_2, q_2) processes to data involve Maximum Likelihood Estimation when the errors are assumed to be normal, and Quasi-Maximum Likelihood Estimation in other cases. After these components are modeled, the residuals $\boldsymbol{\epsilon}_t$ from the mean equations are calculated. These residuals are then standardized using the conditional standard deviation from the GARCH model, producing the innovations $\mathbf{Z}_t$. Subsequently, these innovations are transformed into a uniform scale (as required by the copula model) by getting their pseudo-observations. The pseudo-observations of a d-dimensional sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are defined as $z_{i,j} = \frac{R_{i,j}}{n+1}$, where $R_{i,j}$ is the rank of $Z_{i,j}$ within the univariate path $Z_{1,j}, \ldots, Z_{n,j}$ for $j \in \{1, \ldots, d\}$. Finally, a selected copula is fitted to the pseudo-observations $(z_{1,1}, \ldots, z_{1,d}), \ldots, (z_{n,1}, \ldots, z_{n,d})$ to model the dependency structure between the series. Among the methods used to fit copulas we find Maximum Likelihood Estimation and Kendall's tau Inversion.

Conversely, to simulate multivariate data from a Copula-GARCH model, the process begins by drawing independent samples from the specified copula, which captures the dependencies between the innovations of different time series. These samples are then transformed back to the scale of the original innovations by using the normalized inverse of the assumed marginal distributions. That is, the quantile functions of each marginal distribution are applied to the corresponding simulated data and, in case these distributions do not have mean zero and variance one, the resulting observations are normalized. Finally, the innovations are input into the specific ARMA(p_1, q_1)-GARCH(p_2, q_2) process to generate the time series.

In Section 4.6, we generate bi-variate data using an ARMA(1, 1)-GARCH(1, 1) process for both components of the data, with Student's t-distributions for the marginals, and with a Student's t-copula to model the dependence between the coordinates of the innovations. Afterwards, we fit another Student's t-copula to the aggregated data.

3

Estimators for low-frequency processes

The goal of this chapter is to introduce estimators for parameters of the distribution of a low-frequency stochastic process **Y**, hereafter referred to as the "low-frequency distribution". The estimators presented are constructed from estimations of the distribution of a related higher-frequency stochastic process **X**, which from now on we will call the "higher-frequency distribution". Based on their characteristics, we will explore the consistency, rates of convergence, and asymptotic distribution of the estimators. Section 3.1 establishes the motivation and assumptions behind this research, along with most of the notation to be utilized throughout the thesis. Subsequently, in Section 3.2, we introduce the proposed estimators. Section 3.3 explores properties associated with estimators constructed via sufficiently regular maps, while Section 3.4 examines the utilization of M-estimators. Lastly, Section 3.5 is dedicated to the special case when the estimator is a Minimum Distance Estimator.

3.1. Our setting

Let $(\Omega, \mathscr{A}, \mathbf{P})$ be an arbitrary probability space and consider a discrete-time stochastic process $X = \left(X_t^{(\omega)}\right)_{t\in\mathbb{Z}}$ in $\mathbb{R}^{\mathbb{Z}} = \{(\ldots, x_{-1}, x_0, x_1, \ldots) \,|\, x_t \in \mathbb{R}, t \in \mathbb{Z}\}$, the space of doubly-infinite sequences that take real values at each time point. Occasionally, linear transformations involving the components of these processes are of interest, such as its non-overlapping aggregation over $T \in \mathbb{N}$ consecutive values

$$\left(\ldots, \sum_{i=1}^{T} X_{-T+i}, \sum_{i=1}^{T} X_i, \sum_{i=1}^{T} X_{T+i}, \ldots\right).$$

This transformed sequence, which we denote $Y = \left(Y_t^{(\omega)}\right)_{t\in\mathbb{Z}}$, also constitutes a stochastic process in $\mathbb{R}^{\mathbb{Z}}$. Given that the time intervals between consecutive points in Y are longer compared to those in X, we refer to Y as a low-frequency stochastic process and X as a higher-frequency stochastic process (relative to Y). A practical scenario illustrating the relevance of both processes is when X represents daily returns of a single asset, with each time point corresponding to a financial day. In this context, Y denotes the aggregated returns of the

same asset over a span of T consecutive financial days. For instance, T could be set to 70 (financial days) for quarterly returns and 250 for yearly returns.

Naturally, these stochastic processes have values in infinite-dimensional vector spaces and cannot be fully observed. Therefore, in practice, only finite truncations of them are observed. For instance, from X we could observe $n \in \mathbb{N}$ consecutive periods of length T (a path of length nT) starting from X_1, which yields the truncation

$$(X_1, X_2, \ldots, X_{nT}).$$

Similarly, from Y, we could observe n aggregated returns associated to n consecutive and non-overlapping periods of length T. Starting from Y_1, this would lead to the truncation

$$(Y_1, Y_2, \ldots, Y_n).$$

It is worth noting that non-overlapping aggregation and truncation (over a multiple of T) of consecutive values are commutative operations. This means that aggregating X into Y and truncating the series to obtain $(Y_1, Y_2, \ldots, Y_n)$ yields the same result as truncating the series X into $(X_1, X_2, \ldots, X_{nT})$ and aggregating these values to obtain $(Y_1, Y_2, \ldots, Y_n)$.

As a concrete example of a situation in which we observe only a truncation of a stochastic process, consider the returns of an asset over time. In such instances, analyzing the complete process is impossible. The historical data, although extensive, does not fully capture the infinite-dimensional stochastic process over time due to the missing data or the finite existence of the assets themselves. Therefore, in practice, the process is truncated to include, at most, the period during which the asset data are available. Moreover, analysts and researchers often truncate its history further, focusing on specific, manageable intervals, typically recent years or distinct periods of market behavior. This focus makes the problem more tractable and the data more manageable for computational and analytical methods. Additionally, truncating the history helps to avoid issues of non-stationarity, since, in practice, the distribution evolves over time. However, if the time interval is small enough, the process can be seen as nearly stationary. In other words, a stationary model may be a good fit in the short- to medium-term.

Now, let $\mathbf{X} = \left(\mathbf{X}_t^{(\omega)}\right)_{t\in\mathbb{Z}}$ be a discrete-time stochastic process in $\left(\mathbb{R}^d\right)^{\mathbb{Z}} = \left\{(\ldots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \ldots) \,|\, \mathbf{x}_t \in \mathbb{R}^d, t \in \mathbb{Z}\right\}$, the space of doubly-infinite sequences that take values in $\mathbb{R}^d$ at each time point. We denote $\mathbb{P}_{\mathbf{X}}^{\text{true}}$ the law of $\mathbf{X}$, which is the probability measure on $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ defined by $\mathbb{P}_{\mathbf{X}}^{\text{true}} := \mathbf{P} \circ \mathbf{X}^{-1}$, also called the push-forward measure of $\mathbf{X}$. To define the process $\mathbf{Y} = \left(\mathbf{Y}_t^{(\omega)}\right)_{t\in\mathbb{Z}}$ from $\mathbf{X}$, let $\psi_T : \left(\mathbb{R}^d\right)^{\mathbb{Z}} \to \left(\mathbb{R}^d\right)^{\mathbb{Z}}$ be the aggregation function that takes a sequence in $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ and returns its doubly-infinite non-overlapping temporal aggregation over a period of length $T \in \mathbb{N}$, i.e.

$$\begin{aligned} \psi_T : \left(\mathbb{R}^d\right)^{\mathbb{Z}} &\to \left(\mathbb{R}^d\right)^{\mathbb{Z}} \\ (\mathbf{x}_t)_{t\in\mathbb{Z}} &\mapsto \left(\sum_{i=1}^{T} \mathbf{x}_{T(t-1)+i}\right)_{t\in\mathbb{Z}}. \end{aligned} \tag{3.1}$$

We define the stochastic process $\mathbf{Y}$ as the transformation of $\mathbf{X}$ under ψ_T, that is, $\mathbf{Y} := \psi_T(\mathbf{X})$. We denote $\mathbb{P}_{\mathbf{Y}}^{\text{true}}$ as the probability measure induced on $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ by ψ_T and $\mathbb{P}_{\mathbf{X}}^{\text{true}}$, such that $\mathbb{P}_{\mathbf{Y}}^{\text{true}} := \mathbb{P}_{\mathbf{X}}^{\text{true}} \circ \psi_T^{-1}$. The process $\mathbf{X}$ can represent the values of the daily returns of d assets that conform a portfolio, and the process $\mathbf{Y}$ their non-overlapping aggregation over T consecutive financial days. That is, $\mathbf{Y}$ represents the returns of these d assets over non-overlapping periods of length T. As we will discuss, the same analysis regarding the aggregation and truncation of the series can be applied, resulting in the commutativity of these operations.

For every $n \in \mathbb{N}$ and fixed $T \in \mathbb{N}$, let us define the finite-dimensional versions of the non-overlapping

aggregation function as

$$\begin{aligned}\psi_T : \left(\mathbb{R}^d\right)^{nT} &\to \left(\mathbb{R}^d\right)^n \\ (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{nT}) &\mapsto \left(\sum_{i=1}^{T} \mathbf{x}_i, \sum_{i=1}^{T} \mathbf{x}_{T+i}, \ldots, \sum_{i=1}^{T} \mathbf{x}_{n(T-1)+i}\right).\end{aligned}$$

Remark 14. We use the same symbol ψ_T for aggregation functions on different spaces, since they represent the same concept. The exact domain and image of these functions are given by the context.

Likewise, for each $n \in \mathbb{N}$, let us define the truncation operators $\tau_n : \left(\mathbb{R}^d\right)^{\mathbb{Z}} \to \left(\mathbb{R}^d\right)^n$ as

$$\begin{aligned}\tau_n : \left(\mathbb{R}^d\right)^{\mathbb{Z}} &\to \left(\mathbb{R}^d\right)^n \\ (\mathbf{x}_t)_{t\in\mathbb{Z}} &\mapsto (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n).\end{aligned}$$

With these operators, we have that $\psi_T(\tau_{nT}(\mathbf{X})) = \tau_n\left(\psi_T(\mathbf{X})\right)$, which shows the commutativity of these operations (up to dimensional considerations). Figure 3.1 depicts how the truncated low-frequency process can be derived in two different ways, summarizing the previous discussion.

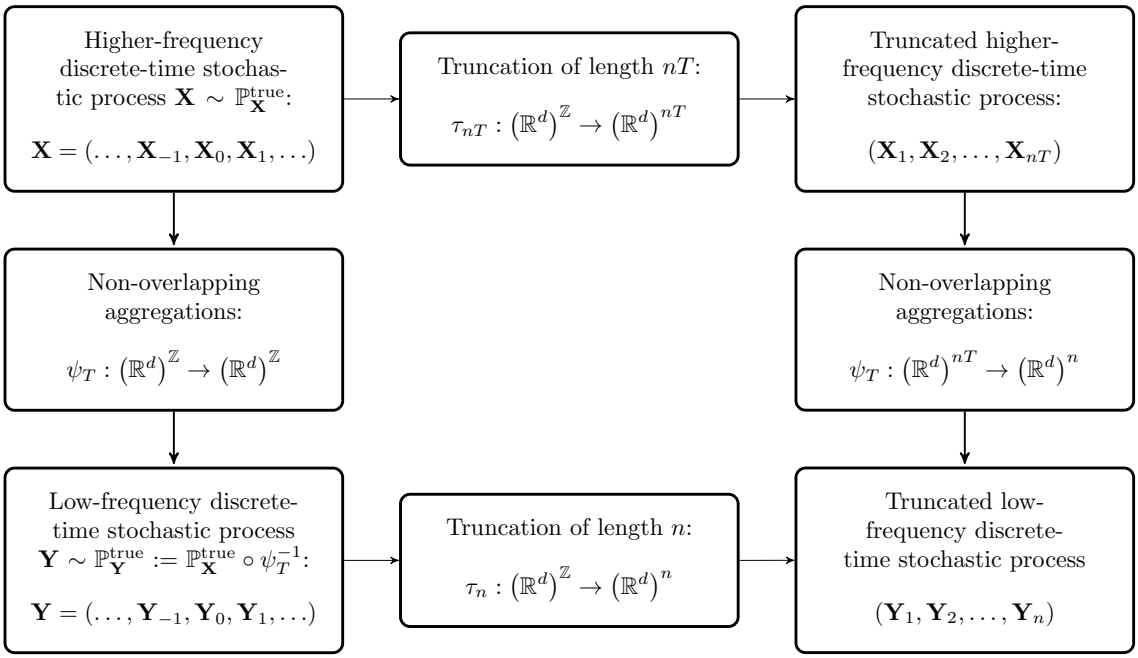


Figure 3.1: Diagram of the commutativity of aggregation and truncation operations in discrete-time stochastic processes.

This figure depicts the transformation of a higher-frequency stochastic process, $\mathbf{X}$, into a truncation of the low-frequency process $\mathbf{Y}$. It details two primary operations: non-overlapping aggregation and truncation. The first approach, shown at the top of the diagram, begins by truncating $\mathbf{X}$ to capture its initial nT values, followed by the non-overlapping aggregation over periods of length T of these values. Alternatively, the second approach, depicted at the bottom, starts by applying non-overlapping aggregation across the entire stochastic process $\mathbf{X}$ to derive $\mathbf{Y}$, which is then truncated to its first n values. The diagram clearly demonstrates that whether aggregation precedes truncation (ψ_T followed by τ_n) or truncation precedes aggregation (τ_{nT} followed by ψ_T), the resultant truncation of the low-frequency process $\mathbf{Y}$ remains the same, showing the commutativity of these operations (up to dimensional considerations).

In this thesis, we assume that the higher-frequency process $\mathbf{X}$ is strictly stationary (Definition 18). Under this assumption, the marginals $\mathbf{X}_t$ of $\mathbf{X}$ are identically distributed across all time points $t \in \mathbb{Z}$. This assumption

implies the strict stationarity of $\mathbf{Y}$ as well, leading to the marginals $\mathbf{Y}_t$ of $\mathbf{Y}$ also being identically distributed for every $t \in \mathbb{Z}$. We refer to the distribution of the marginals $\mathbf{X}_t$ of $\mathbf{X}$ as the "higher-frequency marginal distribution", denoted $\mathbb{P}^{\text{true}}_{\mathbf{X}_t}$, where the sub-index t clarifies that it refers to the marginal distribution and not to the distribution of the whole process $\mathbf{X}$. Similarly, we refer to the distribution of the marginals $\mathbf{Y}_t$ of $\mathbf{Y}$ as the "low-frequency marginal distribution", and denote it by $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$.

We aim for the marginal higher-frequency distribution $\mathbb{P}^{\text{true}}_{\mathbf{X}_t}$ to have the same behaviour as $\mathbb{P}^{\text{true}}_{\mathbf{X}}$ when the latter is applied to a single, arbitrary margin $\mathbf{X}_t$. To achieve this, we can define $\mathbb{P}^{\text{true}}_{\mathbf{X}_t}$ using $\mathbb{P}^{\text{true}}_{\mathbf{X}}$ and a projection operator π that maps sequences in $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ to their values at time zero, i.e.

$$\pi : \left(\mathbb{R}^d\right)^{\mathbb{Z}} \to \mathbb{R}^d$$
$$(\mathbf{x}_t)_{t\in\mathbb{Z}} \mapsto \mathbf{x}_0.$$

Given an arbitrary time $t \in \mathbb{Z}$, and considering the marginal $\mathbf{X}_t$ of $\mathbf{X}$, stationarity of $\mathbf{X}$ implies that $\mathbf{X}_t$ and $\mathbf{X}_0$ have the same distribution. Therefore,

$$\int_{(\mathbb{R}^d)^{\mathbb{Z}}} \mathbf{X}_t \, \mathrm{d}\mathbb{P}^{\text{true}}_{\mathbf{X}} = \int_{(\mathbb{R}^d)^{\mathbb{Z}}} \mathbf{X}_0 \, \mathrm{d}\mathbb{P}^{\text{true}}_{\mathbf{X}} = \int_{(\mathbb{R}^d)^{\mathbb{Z}}} \pi(\mathbf{X}) \, \mathrm{d}\mathbb{P}^{\text{true}}_{\mathbf{X}},$$

where the last equality follows from the definition of π. Since π is a measurable map, from Equation (2.1) we have that $\mathbb{P}^{\text{true}}_{\mathbf{X}}$ induces a measure $\mathbb{P}^{\text{true}}_{\mathbf{X}} \circ \pi^{-1}$ on $\mathbb{R}^d$ which precisely satisfies

$$\int_{(\mathbb{R}^d)^{\mathbb{Z}}} \pi(\mathbf{X}) \, \mathrm{d}\mathbb{P}^{\text{true}}_{\mathbf{X}} = \int_{\mathbb{R}^d} \mathbf{X}_0 \, \mathrm{d}\mathbb{P}^{\text{true}}_{\mathbf{X}} \circ \pi^{-1}. \tag{3.2}$$

Consequently, it is logical to define $\mathbb{P}^{\text{true}}_{\mathbf{X}_t}$ as $\mathbb{P}^{\text{true}}_{\mathbf{X}_t} := \mathbb{P}^{\text{true}}_{\mathbf{X}} \circ \pi^{-1}$ to obtain that

$$\int_{(\mathbb{R}^d)^{\mathbb{Z}}} \mathbf{X}_t \, \mathrm{d}\mathbb{P}^{\text{true}}_{\mathbf{X}} = \int_{\mathbb{R}^d} \mathbf{X}_0 \, \mathrm{d}\mathbb{P}^{\text{true}}_{\mathbf{X}_t} = \int_{\mathbb{R}^d} \mathbf{X}_t \, \mathrm{d}\mathbb{P}^{\text{true}}_{\mathbf{X}_t}.$$

Analogously, we define $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ as $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t} := \mathbb{P}^{\text{true}}_{\mathbf{Y}} \circ \pi^{-1}$.

3.1.1. Goal

The primary goal of this thesis is to provide a methodology for constructing estimators for parameters of the low-frequency marginal distribution $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$, which is crucial for various financial applications in risk management. For instance, when $\mathbf{Y}$ represents the yearly returns of d assets, accurately estimating this marginal distribution is essential for calculating the Value-at-Risk in a specific year, such as the upcoming one. To develop and implement these estimators, we assume availability of an observed path of length nT from the higher-frequency process $\mathbf{X}$, which corresponds to a path of length n from its associated low-frequency process $\mathbf{Y}$, where $n, T \in \mathbb{N}$.

We compare two methods for estimating the parameters of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ from the available observations. The first one, which we call the "direct method", is the conventional approach for parameter estimation of the low-frequency marginal distribution, and consists of directly using the available low-frequency observations to construct the estimators. Unfortunately, in several applications, this approach suffers from the limitation of having insufficient data. Moreover, if the underlying process is stationary, but not independent (which is usually the case), then each observed time point might be dependent on the previous one(s), leading to biased estimators for the marginals. We refer to this method as the "direct method", because the estimators are derived directly from the observed realization, and we name "direct estimators" such estimators.

We propose an alternative approach, which we call the "simulation-based method". We begin by estimating the higher-frequency distribution $\mathbb{P}_{\mathbf{X}}^{\text{true}}$ from the higher-frequency sample path, which is naturally more extensive than the low-frequency path. We denote this estimated higher-frequency distribution as $\hat{\mathbb{P}}_{\mathbf{X},n}$. For each $\omega \in \Omega$, we define $\hat{\mathbb{P}}_{\mathbf{Y},n}^{(\omega)} := \hat{\mathbb{P}}_{\mathbf{X},n}^{(\omega)} \circ \psi_T^{-1}$ and consider this to be an estimation of the true distribution $\mathbb{P}_{\mathbf{Y}}^{\text{true}}$. We can generate paths of length $k \in \mathbb{N}$ from $\hat{\mathbb{P}}_{\mathbf{Y},n}$ by sampling kT values from $\hat{\mathbb{P}}_{\mathbf{X},n}$ and applying non-overlapping aggregation over T consecutive data points. Taking advantage of this, we proceed by simulating a large number $m \in \mathbb{N}$ of independent paths of length T from the estimated distribution $\hat{\mathbb{P}}_{\mathbf{X},n}$. These paths are then aggregated into m independent values, forming a sample from the distribution $\hat{\mathbb{P}}_{\mathbf{Y}_t,n}$, which we use to estimate the desired parameters of $\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}$. This approach is named the "simulation-based method" because it involves estimating the parameters of the low-frequency marginal distribution by simulating multiple independent values from its estimated distribution, $\hat{\mathbb{P}}_{\mathbf{Y}_t,n}$. Estimators derived through this approach are referred to as "simulation-based estimators". Figure 3.2 provides a flowchart that summarizes the direct and simulation-based methods. Note that, since $\hat{\mathbb{P}}_{\mathbf{X},n}$ originates from the $\mathbb{R}^d$-valued stochastic process $\mathbf{X}$ defined on the probability space $(\Omega, \mathscr{A}, \mathbf{P})$, then it also has a random nature. Thus, when considering different sample sizes nT for $n \in \mathbb{N}$, the sequence of estimated distributions $\{\hat{\mathbb{P}}_{\mathbf{X},n}\}_{n\in\mathbb{N}}$ forms a random sequence of probability measures on $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ from $(\Omega, \mathscr{A}, \mathbf{P})$. Similarly, $\{\hat{\mathbb{P}}_{\mathbf{Y},n}\}_{n\in\mathbb{N}}$, like $\{\hat{\mathbb{P}}_{\mathbf{X},n}\}_{n\in\mathbb{N}}$, is a random sequence of probability measures on $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ from $(\Omega, \mathscr{A}, \mathbf{P})$.

Some experiments conducted in Chapter 4 involve both non-parametric and parametric estimations. In the non-parametric approach, we estimate the parameters of $\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}$ without assuming the underlying distribution for $\mathbf{Y}_t$, such as estimating the mean of $\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}$ as the sample mean of the data. Conversely, the parametric approach involves fitting a predefined distribution to $\mathbf{Y}_t$ based on the data, and then estimating the parameter values. For example, one first assumes the low-frequency marginal is normally distributed, with mean and variance equal to the sample mean and sample variance of the low-frequency data, respectively, and then estimates the quantiles of $\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}$ from the quantiles of normal distributions. We distinguish between direct parametric and direct non-parametric estimators, as well as between simulation-based parametric and simulation-based non-parametric estimators, when applicable.

The simulation-based method intends to tackle the common challenge that the available low-frequency data is not enough to construct reliable direct estimators for the parameters of the low-frequency distribution, such as its standard deviation and quantiles. Naturally, it is expected that the simulation-based method outperforms the direct method for small n and large T, where the number of points nT to define $\hat{\mathbb{P}}_{\mathbf{X},n}$ is enough to estimate $\mathbb{P}_{\mathbf{X}}^{\text{true}}$, but the amount of values n for the direct method is insufficient to construct estimators for the parameters of $\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}$. Of course, the performance of the simulation-based method depends on the estimation of the higher-frequency distribution. Therefore, it remains important to appropriately model the time dependence of $\mathbf{X}$ to construct accurate models for higher-frequency data.

A specific real-world application that motivates the use of the simulation-based method is the estimation of parameters of the distribution of the yearly returns of an asset. If we only have the history of daily returns for a total of ten years, then we have at most ten non-overlapping yearly returns, which is far from sufficient to construct a reliable estimator. Using, for example, monthly-overlapping yearly returns can increase the data volume, but this approach by design leads to autocorrelation in the estimates, potentially biasing the results. In this context, the simulation-based method can be useful. It would involve using ten years of daily returns to construct an estimator $\hat{\mathbb{P}}_{\mathbf{X},n}$ of $\mathbb{P}_{\mathbf{X}}^{\text{true}}$, sampling a large number m of independent paths of yearly length T from $\hat{\mathbb{P}}_{\mathbf{X},n}$, aggregating them to obtain an independent sample of size m from $\hat{\mathbb{P}}_{\mathbf{Y}_t,n}$ representing yearly returns, and then using these aggregated values to estimate the necessary parameters.

Although our primary goal involves estimating parameters of the low-frequency marginal distribution, and the experiments detailed in Chapter 4 are designed with this objective, the simulation-based method can be further applied to estimate parameters of the joint distributions of low-frequency processes, thereby capturing the dependency between its values. For instance, to estimate the autocorrelation between two consecutive values of $\mathbf{Y}$, we can generate m independent paths of length $2T$ from $\hat{\mathbb{P}}_{\mathbf{X},n}$, and independently aggregate these paths over periods of length T. This process yields an independent sample of m pairs of values from which the autocorrelation can be estimated. Therefore, in the remainder of this chapter, we keep these generalization capabilities of the simulation-based method and develop a framework to construct

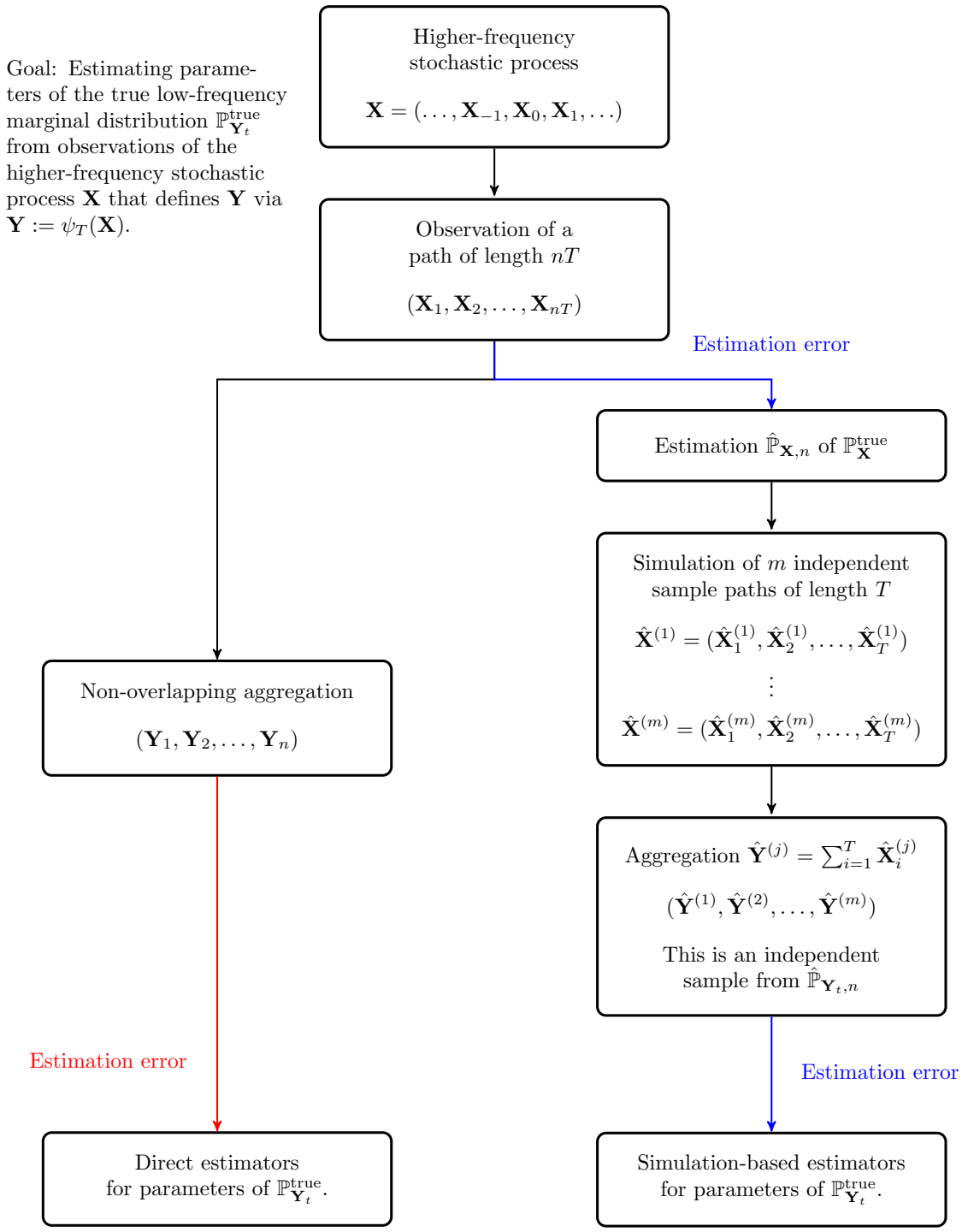


Figure 3.2: Overview of the direct and simulation-based methods for parameter estimation.

This flowchart illustrates the methodology for estimating parameters of the low-frequency marginal distribution from higher-frequency observations. The process begins by observing a path of length nT from the higher-frequency stochastic process $\mathbf{X}$. To the left, the direct method involves non-overlapping aggregation over periods of length T, resulting in a sample path of length n from $\mathbf{Y}$. From this, we construct the direct estimators for parameters of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$. To the right, the simulation-based method starts by estimating $\hat{\mathbb{P}}_{\mathbf{X},n}$ from $\mathbf{X}$, followed by simulating m independent paths of length T, and independently aggregating each one of them. These aggregations produce a sample of m independent values from $\hat{\mathbb{P}}_{\mathbf{Y}_t,n}$, from which the simulation-based estimators are built. The colored lines represent the main sources of error to estimate parameters of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$. The simulation-based method is advantageous when the red line in the left side of the chart represents a higher estimation error than the sum of the errors represented by the two blue lines in the right side.

estimators for parameters of $\mathbb{P}_{\mathbf{Y}}^{\text{true}}$ based on estimations of $\mathbb{P}_{\mathbf{X}}^{\text{true}}$. As previously done, we can obtain finite-dimensional distributions for specific time points of $\mathbf{Y}$ by projection.

3.1.2. The space $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$

Given that the processes $\mathbf{X}$ and $\mathbf{Y}$ are defined in $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$, we are interested in defining a suitable metric for it. For $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^d$ and $\|\cdot\|$ a norm on $\mathbb{R}^d$, let $b(\boldsymbol{\alpha}, \boldsymbol{\beta}) = 1 \wedge \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|$ and let us define the metric ρ on $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ by

$$\rho\left(\mathbf{x}, \mathbf{y}\right) = \sum_{t=-\infty}^{\infty} \frac{b\left(\mathbf{x}_t, \mathbf{y}_t\right)}{2^{|t|}}.$$

The following proposition states that, for an element $\mathbf{x}$ and a sequence $\{\mathbf{x}^{(n)}\}_{n\in\mathbb{N}}$ in $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$, $\mathbf{x}^{(n)}$ tends to $\mathbf{x}$ under ρ if and only if $\mathbf{x}_t^{(n)}$ tends to $\mathbf{x}_t$ under b for every time point $t \in \mathbb{Z}$.

Proposition 1. *Let $\mathbf{x} \in \left(\mathbb{R}^d\right)^{\mathbb{Z}}$ and $\{\mathbf{x}^{(n)}\}_{n\in\mathbb{N}}$ be a sequence in $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$. Then, $\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) \to 0$ if and only if $b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right) \to 0$ for every $t \in \mathbb{Z}$.*

Proof. We prove the left to right implication by contrapositive. Let us suppose that there exists a $t' \in \mathbb{Z}$ such that $b\left(\mathbf{x}_{t'}^{(n)}, \mathbf{x}_{t'}\right) \not\to 0$ as $n \to \infty$. For such t', there exists an $\epsilon > 0$ such that $b\left(\mathbf{x}_{t'}^{(n)}, \mathbf{x}_{t'}\right) > \epsilon$ for all $n \in \mathbb{N}$. Then,

$$\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) = \sum_{t=-\infty}^{\infty} \frac{b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right)}{2^{|t|}} > \frac{\epsilon}{2^{|t'|}} > 0$$

for all $n \in \mathbb{N}$, which implies that $\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) \not\to 0$. Thus, it must hold that $\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) \to 0 \implies b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right) \to 0$ for every $t \in \mathbb{Z}$.

For the opposite implication, let us suppose that $b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right) \to 0$ for every $t \in \mathbb{Z}$. Note that, for $k \in \mathbb{N}$,

$$\begin{aligned}
\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) &= \sum_{t=-\infty}^{\infty} \frac{b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right)}{2^{|t|}} \\
&= \sum_{t=-\infty}^{-(k+1)} \frac{b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right)}{2^{|t|}} + \sum_{t=-k}^{k} \frac{b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right)}{2^{|t|}} + \sum_{t=k+1}^{\infty} \frac{b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right)}{2^{|t|}} \\
&\leq \sum_{t=-k}^{k} \frac{b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right)}{2^{|t|}} + 2 \sum_{t=k+1}^{\infty} \frac{1}{2^t},
\end{aligned}$$

where we used that $b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right) \leq 1$. Now, let $\epsilon > 0$ be arbitrary. Since the sum $\sum_{t=0}^{\infty} \frac{1}{2^t}$ converges to a constant, we have that there exists $k' \in \mathbb{N}$ such that $2\sum_{t=k'+1}^{\infty} \frac{1}{2^t} = \sum_{t=k'}^{\infty} \frac{1}{2^t} < \frac{\epsilon}{2}$. Then,

$$\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) < \sum_{t=-k'}^{k'} \frac{b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right)}{2^{|t|}} + \frac{\epsilon}{2}.$$

Since $\sum_{t=-k}^{k} \frac{b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right)}{2^{|t|}}$ is a finite sum of terms each of which converges to zero, then this sum also converges to zero. Hence, there exists an $n' \in \mathbb{N}$ such that $\sum_{t=-k}^{k} \frac{b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right)}{2^{|t|}} < \frac{\epsilon}{2}$ for $n \geq n'$. We conclude that for every $\epsilon > 0$

there exists an $n' \in \mathbb{N}$ such that $\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) < \epsilon$ for $n \geq n'$, and so $\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) \to 0$. □

It is not difficult to verify that the metric space $\left(\left(\mathbb{R}^d\right)^{\mathbb{Z}}, \rho\right)$ is separable and complete. This can be done by using the same arguments as the ones in Billingsley [3] for $\mathbb{R}^{\infty}$. In summary, one countable and dense subset of $\left(\left(\mathbb{R}^d\right)^{\mathbb{Z}}, \rho\right)$ consists of the points having only finitely many non-zero coordinates, with each of them being rational, which gives the separability. On the other hand, completeness follows from the fact that for a Cauchy sequence $\left\{\mathbf{x}^{(n)}\right\}_{n\in\mathbb{N}}$ in $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ and $t \in \mathbb{Z}$, the sequence $\left\{\mathbf{x}_t^{(n)}\right\}_{n\in\mathbb{N}}$ in $\mathbb{R}^d$ is also Cauchy, and thus, converges to some $\mathbf{x}_t \in \mathbb{R}^d$ (by completeness of $\mathbb{R}^d$). Then, $\mathbf{x}^{(n)}$ converges to the sequence $\mathbf{x}$ composed of the limits $\mathbf{x}_t$ for every $t \in \mathbb{Z}$.

We now prove that the aggregation function ψ_T is a continuous function on the metric space $\left(\left(\mathbb{R}^d\right)^{\mathbb{Z}}, \rho\right)$. This allows us to derive properties of the estimators for parameters of $\mathbb{P}_{\mathbf{Y}}^{\text{true}}$ from the properties of the estimated distributions $\hat{\mathbb{P}}_{\mathbf{X},n}$. To prove the continuity, we show that the sequence is sequentially continuous (recall Definition 2), and use that continuity and sequential continuity are equivalent in metric spaces.

Lemma 5. *The function ψ_T is continuous on $\left(\left(\mathbb{R}^d\right)^{\mathbb{Z}}, \rho\right)$.*

Proof. Let $\mathbf{x} \in \left(\mathbb{R}^d\right)^{\mathbb{Z}}$ be arbitrary, and suppose that for a sequence $\left\{\mathbf{x}^{(n)}\right\}_{n\in\mathbb{N}}$ in $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$, we have $\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) \to 0$. Note that

$$\begin{aligned}
\rho\left(\psi_T\left(\mathbf{x}^{(n)}\right), \psi_T(\mathbf{x})\right) &= \sum_{t=-\infty}^{\infty} \frac{1 \wedge \left\|\psi_T\left(\mathbf{x}^{(n)}\right)_t - \psi_T(\mathbf{x})_t\right\|}{2^{|t|}} \\
&= \sum_{t=-\infty}^{\infty} \frac{1 \wedge \left\|\sum_{i=1}^{T} \mathbf{x}^{(n)}_{T(t-1)+i} - \sum_{i=1}^{T} \mathbf{x}_{T(t-1)+i}\right\|}{2^{|t|}} \\
&= \sum_{t=-\infty}^{\infty} \frac{1 \wedge \left\|\sum_{i=1}^{T} \left(\mathbf{x}^{(n)}_{T(t-1)+i} - \mathbf{x}_{T(t-1)+i}\right)\right\|}{2^{|t|}} \\
&\leq \sum_{t=-\infty}^{\infty} \frac{1 \wedge \left(\left\|\mathbf{x}^{(n)}_{T(t-1)+1} - \mathbf{x}_{T(t-1)+1}\right\| + \ldots + \left\|\mathbf{x}^{(n)}_{Tt} - \mathbf{x}_{Tt}\right\|\right)}{2^{|t|}} \\
&\leq \sum_{t=-\infty}^{\infty} \frac{\left(1 \wedge \left\|\mathbf{x}^{(n)}_{T(t-1)+1} - \mathbf{x}_{T(t-1)+1}\right\|\right) + \ldots + \left(1 \wedge \left\|\mathbf{x}^{(n)}_{Tt} - \mathbf{x}_{Tt}\right\|\right)}{2^{|t|}} \\
&= \sum_{t=-\infty}^{\infty} \frac{b\left(\mathbf{x}^{(n)}_{T(t-1)+1}, \mathbf{x}_{T(t-1)+1}\right) + \ldots + b\left(\mathbf{x}^{(n)}_{Tt}, \mathbf{x}_{Tt}\right)}{2^{|t|}}.
\end{aligned}$$

Since $\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) \to 0$, by Propostion 1 we have that $b\left(\mathbf{x}_t^{(n)}, \mathbf{x}_t\right) \to 0$ for every $t \in \mathbb{Z}$. Let $\epsilon > 0$ be arbitrary. Proceeding as before, for $k \in \mathbb{N}$ we have that

$$\rho\left(\psi_T\left(\mathbf{x}^{(n)}\right), \psi_T(\mathbf{x})\right) \leq \sum_{t=-k}^{k} \frac{b\left(\mathbf{x}^{(n)}_{T(t-1)+1}, \mathbf{x}_{T(t-1)+1}\right) + \ldots + b\left(\mathbf{x}^{(n)}_{Tt}, \mathbf{x}_{Tt}\right)}{2^{|t|}} + \sum_{t=k}^{\infty} \frac{T}{2^t},$$

and since $\sum_{t=0}^{\infty} \frac{T}{2^t}$ is a convergent sequence, there exists a $k' \in \mathbb{N}$ such that $\sum_{t=k'}^{\infty} \frac{T}{2^t} < \frac{\epsilon}{2}$. Given that the first sum on the right is a finite sum of terms converging to zero, it also converges to zero, and so there exists an $n' \in \mathbb{N}$ such that, for every $n \geq n'$,

$$\sum_{t=-k}^{k} \frac{b\left(\mathbf{x}^{(n)}_{T(t-1)+1}, \mathbf{x}_{T(t-1)+1}\right) + \ldots + b\left(\mathbf{x}^{(n)}_{Tt}, \mathbf{x}_{Tt}\right)}{2^{|t|}} < \frac{\epsilon}{2}.$$

Thus, for every $\epsilon > 0$ there exists an $n' \in \mathbb{N}$ such that $\rho\left(\psi_T\left(\mathbf{x}^{(n)}\right), \psi_T(\mathbf{x})\right) < \epsilon$ for every $n \geq n'$. We then conclude that $\rho\left(\mathbf{x}^{(n)}, \mathbf{x}\right) \to 0 \implies \rho\left(\psi_T\left(\mathbf{x}^{(n)}\right), \psi_T(\mathbf{x})\right) \to 0$, and so ψ_T is continuous on $\left(\left(\mathbb{R}^d\right)^{\mathbb{Z}}, \rho\right)$. □

The following lemma shows that **P**-almost sure weak convergence of the estimated laws of $\mathbf{X}$, $\left\{\hat{\mathbb{P}}_{\mathbf{X},n}\right\}_{n\in\mathbb{N}}$, implies **P**-almost sure weak convergence of the estimated laws of $\mathbf{Y}$, $\left\{\hat{\mathbb{P}}_{\mathbf{Y},n}\right\}_{n\in\mathbb{N}}$.

Lemma 6. *Suppose that* $\hat{\mathbb{P}}_{\mathbf{X},n} \rightsquigarrow \mathbb{P}_{\mathbf{X}}^{true}$ **P***-a.s. Then, it holds that* $\hat{\mathbb{P}}_{\mathbf{Y},n} \rightsquigarrow \mathbb{P}_{\mathbf{Y}}^{true}$ **P***-a.s.*

Proof. Let $\omega \in \Omega$ be arbitrary. By the continuous mapping theorem, we have that $\hat{\mathbb{P}}_{\mathbf{X},n}^{(\omega)} \rightsquigarrow \mathbb{P}_{\mathbf{X}}^{\text{true}}$ implies $\hat{\mathbb{P}}_{\mathbf{X},n}^{(\omega)}\psi^{-1} \rightsquigarrow \mathbb{P}_{\mathbf{X}}^{\text{true}}\psi^{-1}$, which by definition is equivalent to $\hat{\mathbb{P}}_{\mathbf{Y},n}^{(\omega)} \rightsquigarrow \mathbb{P}_{\mathbf{Y}}^{\text{true}}$. Thus, if $\mathbf{P}\left(\hat{\mathbb{P}}_{\mathbf{X},n} \rightsquigarrow \mathbb{P}_{\mathbf{X}}^{\text{true}}\right) = 1$, then $\mathbf{P}\left(\hat{\mathbb{P}}_{\mathbf{Y},n} \rightsquigarrow \mathbb{P}_{\mathbf{Y}}^{\text{true}}\right) = 1$. That is, $\hat{\mathbb{P}}_{\mathbf{X},n} \rightsquigarrow \mathbb{P}_{\mathbf{X}}^{\text{true}}$ **P**-a.s. implies $\hat{\mathbb{P}}_{\mathbf{Y},n} \rightsquigarrow \mathbb{P}_{\mathbf{Y}}^{\text{true}}$ **P**-a.s. □

3.2. Estimators

Now, let us assume we have a parametric family $(\mathbb{P}_\theta)_{\theta\in\Theta}$ of probability measures on $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ and that there exists a $\theta^{\text{true}} \in \Theta$ such that $\mathbb{P}_{\mathbf{Y}}^{\text{true}} = \mathbb{P}_{\theta^{\text{true}}}$. We shall assume that the parameter space (Θ, μ) is a metric space. By Theorem 1.2 in Billingsley [3], an arbitrary probability measure $\mathbb{P}$ on the metric space $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ can be identified with the mapping

$$f \mapsto \mathbb{P}f := \int_{\left(\mathbb{R}^d\right)^{\mathbb{Z}}} f \, \mathrm{d}\mathbb{P}$$

for $f \in \mathscr{F}_{bu}$, the class of bounded and uniformly continuous functions on $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$. Clearly, this mapping is linear in $f \in \mathscr{F}_{bu}$. Also, $\mathbb{P}$ can be seen as an element of $\ell^\infty(\mathscr{F}_{bu})$, the space of all uniformly bounded, real functions on $\mathscr{F}_{bu}$.

Let $\mathscr{F}$ be a class of functions on $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ and $\vartheta : \ell^\infty(\mathscr{F}) \to \Theta$ such that $\vartheta(\mathbb{P}_{\mathbf{Y}}^{\text{true}}) = \theta^{\text{true}}$. We define an estimator $\hat{\theta}_n$ of θ^{true} as $\hat{\theta}_n := \vartheta\left(\hat{\mathbb{P}}_{\mathbf{Y},n}\right)$ and distinguish three main (possibly overlapping) cases, namely:

1. ϑ is regular. More specifically, for consistency we assume continuity of the map in the sense that for all $\epsilon > 0$ there exists a $\delta > 0$ such that if $\|\mathbb{P} - \mathbb{Q}\|_{\mathscr{F}} < \delta$, then $\mu(\vartheta(\mathbb{P}), \vartheta(\mathbb{Q})) < \epsilon$. On the other hand, to obtain the rate of convergence and the asymptotic distribution, we suppose Hadamard-differentiability (to apply the Delta Method).

2. $\hat{\theta}_n$ is an M-estimator. To make it fit with our framework, we define $M_n : \Theta \to \mathbb{R}$ as $M_n(\theta) := L\left(\hat{\mathbb{P}}_{\mathbf{Y},n}; \theta\right)$ for a fixed function $L(\cdot;\theta) : \ell^\infty(\mathscr{F}) \to \mathbb{R}$. Then, $\vartheta(\cdot) := \underset{\theta\in\Theta}{\arg\max}\, L(\cdot;\theta)$.

3. $\hat{\theta}_n$ is a Minimum Distance Estimator, defined by

$$\vartheta(\mathbb{Q}) := \underset{\theta\in\Theta}{\arg\min}\, \mathrm{d}(\mathbb{Q}, \mathbb{P}_\theta) \tag{3.3}$$

for a probability measure $\mathbb{Q}$ on $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$ and a distance function d between probability measures on $\left(\mathbb{R}^d\right)^{\mathbb{Z}}$. This is the parameter $\theta \in \Theta$ that makes $\mathbb{P}_\theta$ as close as possible to $\mathbb{Q}$ according to the distance d. Note that since $\mathbb{P}_{\mathbf{Y}}^{\text{true}} = \mathbb{P}_{\theta^{\text{true}}}$ (the model is well-specified), we have

$$\vartheta\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}\right) = \vartheta\left(\mathbb{P}_{\theta^{\text{true}}}\right) = \underset{\theta\in\Theta}{\arg\min}\, \mathrm{d}\left(\mathbb{P}_{\theta^{\text{true}}}, \mathbb{P}_\theta\right) = \theta^{\text{true}}.$$

Remark 15. Generally speaking, we would like the process $M_n(\theta) := L\left(\hat{\mathbb{P}}_{\mathbf{Y},n};\theta\right)$ to converge to the deterministic function $M(\theta) := L\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}};\theta\right)$.

Remark 16. The third case is clearly a special case of the second for $L(\cdot;\theta) := -\mathrm{d}(\cdot,\mathbb{P}_\theta)$. Thus,

$$M_n(\theta) := -d\left(\hat{\mathbb{P}}_{\mathbf{Y},n},\mathbb{P}_\theta\right).$$

We will prove the consistency of these estimators under different assumptions. Moreover, the rates of convergence and asymptotic distributions of the estimators will be identified when possible.

3.3. Regular estimators

In the following section, we derive properties of estimators for parameters of the low-frequency distribution when these estimators are constructed from estimations of the higher-frequency distribution through sufficiently regular maps. By "sufficiently regular maps", we mean either continuous or Hadamard-differentiable maps, depending on the desired characteristic.

We start by proving the strong consistency of the estimator for continuous ϑ. To that end, let us define the map $\varphi : \ell^\infty(\mathscr{F}) \to \ell^\infty(\mathscr{F})$ as

$$\varphi(\mathbb{P})(f) := \left(\mathbb{P}\circ\psi_T^{-1}\right)(f) = \int f(y)\; d\mathbb{P}\circ\psi_T^{-1}(y) = \int f\left(\psi_T(x)\right)\; d\mathbb{P}(x) = \mathbb{P}\left(f\left(\psi_T\right)\right),$$

for $f \in \mathscr{F}$ and $\mathbb{P} \in \ell^\infty(\mathscr{F})$, provided the integrals exist. Note that when $\mathbb{P}$ is a probability measure, the map φ outputs the probability measure induced by ψ_T and $\mathbb{P}$, yielding $\varphi\left(\hat{\mathbb{P}}_{\mathbf{X},n}\right) = \hat{\mathbb{P}}_{\mathbf{Y},n}$ and $\varphi\left(\mathbb{P}_{\mathbf{X}}^{\text{true}}\right) = \mathbb{P}_{\mathbf{Y}}^{\text{true}}$. By further applying the map ϑ, we obtain $\hat{\theta}_n = \vartheta\left(\varphi\left(\hat{\mathbb{P}}_{\mathbf{X},n}\right)\right)$ and $\theta^{\text{true}} = \vartheta\left(\varphi\left(\mathbb{P}_{\mathbf{X}}^{\text{true}}\right)\right)$, which is a convenient way to represent our estimator and the true value of the parameter as a function of higher-frequency probability measures.

The map φ is linear and continuous, as shown in the following lemma.

Lemma 7. *The map φ is linear and continuous.*

Proof. The map φ is linear, since for $f \in \mathscr{F}$, $\mathbb{P}_1, \mathbb{P}_2 \in \ell^\infty(\mathscr{F})$ and $\alpha, \beta \in \mathbb{R}$, it holds that

$$\begin{aligned}
\varphi\left(\alpha\mathbb{P}_1 + \beta\mathbb{P}_2\right)(f) &= \int f\left(\psi_T(x)\right)\; d\left(\alpha\mathbb{P}_1 + \beta\mathbb{P}_2\right)(x) \\
&= \alpha\int f\left(\psi_T(x)\right)\; d\mathbb{P}_1(x) + \beta\int f\left(\psi_T(x)\right)\; d\mathbb{P}_2(x) \\
&= \alpha\varphi(\mathbb{P}_1)(f) + \beta\varphi(\mathbb{P}_2)(f).
\end{aligned}$$

Similarly, φ is continuous since, if $\mathbb{P}_n \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{P}$, then for every $f \in \mathscr{F}$ it holds that

$$\left|\varphi(\mathbb{P}_n)(f) - \varphi(\mathbb{P})(f)\right| = \left|\mathbb{P}_n\left(f\left(\psi_T\right)\right) - \mathbb{P}\left(f\left(\psi_T\right)\right)\right| \le \|\mathbb{P}_n - \mathbb{P}\|_{\mathscr{F}} \to 0,$$

and thus, $\varphi(\mathbb{P}_n) \xrightarrow{\ell^\infty(\mathscr{F})} \varphi(\mathbb{P})$. □

We provide conditions for the strong convergence of $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ to θ^{true} when ϑ is continuous and the estimations of the higher-frequency law converge $\mathbf{P}$-almost surely to the true value in $\ell^\infty(\mathscr{F})$.

Theorem 13 (Strong consistency of $\hat{\theta}_n$ for continuous ϑ)**.** *Let $\hat{\mathbb{P}}_{\mathbf{X},n} \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{P}_{\mathbf{X}}^{true}$ $\mathbf{P}$-a.s. and suppose that ϑ is continuous. Then, $\hat{\theta}_n \to \theta^{true}$ $\mathbf{P}$-a.s.*

Proof. Let us suppose that $\hat{\mathbb{P}}_{\mathbf{X},n} \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{P}_{\mathbf{X}}^{\text{true}}$ $\mathbf{P}$-almost surely. That is,

$$\mathbf{P}\left(\hat{\mathbb{P}}_{\mathbf{X},n} \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{P}_{\mathbf{X}}^{\text{true}}\right) := \mathbf{P}\left(\omega : \hat{\mathbb{P}}_{\mathbf{X},n}^{(\omega)} \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{P}_{\mathbf{X}}^{\text{true}}\right) = 1.$$

Since φ and ϑ are both continuous, we have that $\vartheta\circ\varphi : \ell^\infty(\mathscr{F}) \to \Theta$ is a continuous map. Take $\omega \in \Omega$ such that $\hat{\mathbb{P}}_{\mathbf{X},n}^{(\omega)} \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{P}_{\mathbf{X}}^{\text{true}}$. Then, by sequential continuity of $\vartheta\circ\varphi$ we have that $\vartheta\circ\varphi\left(\hat{\mathbb{P}}_{\mathbf{X},n}^{(\omega)}\right) \to \vartheta\circ\varphi\left(\mathbb{P}_{\mathbf{X}}^{\text{true}}\right)$ in Θ. That is,

$$\mu\left(\hat{\theta}_n^{(\omega)}, \theta^{\text{true}}\right) = \mu\left(\vartheta\circ\varphi\left(\hat{\mathbb{P}}_{\mathbf{X},n}^{(\omega)}\right), \vartheta\circ\varphi\left(\mathbb{P}_{\mathbf{X}}^{\text{true}}\right)\right) \to 0.$$

Therefore, for every $\omega \in \Omega$ such that $\hat{\mathbb{P}}_{\mathbf{X},n}^{(\omega)} \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{P}_{\mathbf{X}}^{\text{true}}$, it follows that $\hat{\theta}_n^{(\omega)} \to \theta^{\text{true}}$. Consequently, we conclude that

$$\mathbf{P}\left(\hat{\theta}_n \to \theta^{\text{true}}\right) := \mathbf{P}\left(\omega : \hat{\theta}_n^{(\omega)} \to \theta^{\text{true}}\right) = 1,$$

getting the result desired. □

The following proposition states that φ is not only linear and continuous, but also Hadamard-differentiable. This property holds for linear and continuous functions in general.

Lemma 8. *The map φ is Hadamard-differentiable.*

Proof. Take $t > 0$, $\mathbb{P}, \mathbb{Q} \in \ell^\infty(\mathscr{F})$ and a net $\{\mathbb{Q}_t\}_t$ in $\ell^\infty(\mathscr{F})$ such that $\|\mathbb{Q}_t - \mathbb{Q}\|_{\mathscr{F}} \to 0$ as $t \to 0$. By linearity,

$$\varphi(\mathbb{P} + t\mathbb{Q}_t) = \varphi(\mathbb{P}) + t\varphi(\mathbb{Q}_t),$$

and therefore

$$\frac{\varphi(\mathbb{P} + t\mathbb{Q}_t) - \varphi(\mathbb{P})}{t} - \varphi(\mathbb{Q}_t) = 0.$$

Also, since φ is continuous we have that $\left\|\varphi(\mathbb{Q}_t) - \varphi(\mathbb{Q})\right\|_{\mathscr{F}} \to 0$ as $\|\mathbb{Q}_t - \mathbb{Q}\|_{\mathscr{F}} \to 0$. Then,

$$\left\|\frac{\varphi(\mathbb{P} + t\mathbb{Q}_t) - \varphi(\mathbb{P})}{t} - \varphi(\mathbb{Q})\right\|_{\mathscr{F}} \leq \left\|\frac{\varphi(\mathbb{P} + t\mathbb{Q}_t) - \varphi(\mathbb{P})}{t} - \varphi(\mathbb{Q}_t)\right\|_{\mathscr{F}} + \left\|\varphi(\mathbb{Q}_t) - \varphi(\mathbb{Q})\right\|_{\mathscr{F}} \to 0,$$

when $t \to 0$ for $\mathbb{Q}_t \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{Q}$. Thus, choosing $\varphi'_{\mathbb{P}}(\mathbb{Q}) = \varphi(\mathbb{Q})$ (which is linear and continuous) we get the proper derivative. □

We can now state the weak convergence of the scaled errors $\left\{r_n\left(\hat{\theta}_n - \theta^{\text{true}}\right)\right\}_{n\in\mathbb{N}}$ for the same rates of convergence $\{r_n\}_{n\in\mathbb{N}}$ of the scaled errors $\left\{r_n\left(\hat{\mathbb{P}}_{\mathbf{X},n} - \mathbb{P}_{\mathbf{X}}^{\text{true}}\right)\right\}_{n\in\mathbb{N}}$ provided ϑ is Hadamard-differentiable.

Theorem 14 (Weak convergence of $\left\{r_n\left(\hat{\theta}_n - \theta^{\text{true}}\right)\right\}_{n\in\mathbb{N}}$ for Hadamard-differentiable ϑ)**.** *Suppose there exists a sequence of numbers $\{r_n\}_{n\in\mathbb{N}}$, with $r_n \to \infty$, and a random element $\mathbb{T}$ that takes its values in $\ell^\infty(\mathscr{F})$ such*

that $r_n\left(\hat{\mathbb{P}}_{\mathbf{X},n} - \mathbb{P}_{\mathbf{X}}^{true}\right) \rightsquigarrow \mathbb{T}$. *Suppose further that the map* ϑ *is Hadamard-differentiable at* $\mathbb{P}_{\mathbf{Y}}^{true}$ *with derivative* $\vartheta'_{\mathbb{P}_{\mathbf{Y}}^{true}}$. *Then,* $r_n\left(\hat{\theta}_n - \theta^{true}\right) \rightsquigarrow \vartheta'_{\mathbb{P}_{\mathbf{Y}}^{true}}\left(\varphi\left(\mathbb{T}\right)\right)$.

Proof. Since φ is Hadamard-differentiable with derivative $\varphi'_{\mathbb{P}} = \varphi$ at an arbitrary $\mathbb{P} \in \ell^\infty(\mathscr{F})$, by the Delta Method (Theorem 4) we have that $r_n\left(\varphi\left(\hat{\mathbb{P}}_{\mathbf{X},n}\right) - \varphi\left(\mathbb{P}_{\mathbf{X}}^{\text{true}}\right)\right) \rightsquigarrow \varphi\left(\mathbb{T}\right)$. Then, since ϑ is Hadamard-differentiable at $\varphi\left(\mathbb{P}_{\mathbf{X}}^{\text{true}}\right) =: \mathbb{P}_{\mathbf{Y}}^{\text{true}}$, we can apply the Delta Method once again to get

$$r_n\left(\hat{\theta}_n - \theta^{\text{true}}\right) = r_n\left(\vartheta\left(\varphi\left(\hat{\mathbb{P}}_{\mathbf{X},n}\right)\right) - \vartheta\left(\varphi\left(\mathbb{P}_{\mathbf{X}}^{\text{true}}\right)\right)\right) \rightsquigarrow \vartheta'_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}(\varphi(\mathbb{T})).$$

□

Finally, we are interested in the asymptotic distribution of the estimators for the parameters of the low-frequency distribution . In general, what is really proven is the weak convergence of the sequence of scaled errors $\left\{r_n\left(\hat{\theta}_n - \theta^{\text{true}}\right)\right\}_{n\in\mathbb{N}}$ to a known probability distribution. If we assume that our estimator $\hat{\theta}_n$ is built from the estimator $\hat{\mathbb{P}}_{\mathbf{Y},n}$ through a Hadamard-differentiable map ϑ, and that the scaled error of $\hat{\mathbb{P}}_{\mathbf{X},n}$ has asymptotic normal behavior, we can prove that the scaled error of $\hat{\theta}_n$ is also asymptotically normally distributed, as shown in the following theorem.

Theorem 15 (Asymptotic normality for Hadamard-differentiable ϑ)**.** *Suppose there exists a sequence of numbers* $\{r_n\}_{n\in\mathbb{N}}$, *with* $r_n \to \infty$, *and a tight Borel-measurable Gaussian map* $\mathbb{T}$ *that takes its values in* $\ell^\infty(\mathscr{F})$ *such that* $r_n\left(\hat{\mathbb{P}}_{\mathbf{X},n} - \mathbb{P}_{\mathbf{X}}^{true}\right) \rightsquigarrow \mathbb{T}$. *Suppose further that the parameter space* Θ *is a Banach space and that the map* ϑ *is Hadamard-differentiable at* $\mathbb{P}_{\mathbf{Y}}^{true}$ *with derivative* $\vartheta'_{\mathbb{P}_{\mathbf{Y}}^{true}}$. *Then* $\left\{r_n\left(\hat{\theta}_n - \theta^{true}\right)\right\}_{n\in\mathbb{N}}$ *converges weakly to* $\vartheta'_{\mathbb{P}_{\mathbf{Y}}^{true}}\left(\varphi\left(\mathbb{T}\right)\right)$, *which is normally distributed.*

Proof. By Theorem 14, we have that

$$r_n\left(\hat{\theta}_n - \theta^{\text{true}}\right) \rightsquigarrow \vartheta'_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}} \circ \varphi\left(\mathbb{T}\right).$$

Since both $\vartheta'_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}$ and φ are linear and continuous, then its composition $\vartheta'_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}} \circ \varphi$ is linear and continuous. Thus, given that $\mathbb{T}$ is a tight Borel-measurable Gaussian map into $\ell^\infty(\mathscr{F})$ and $\vartheta'_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}} \circ \varphi$ is a continuous, linear map into the Banach space Θ, by Lemma 3 we have that $\vartheta'_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}\left(\varphi\left(\mathbb{T}\right)\right)$ is normally distributed. □

We now provide two concrete examples of estimators built from regular maps.

Example 3 (Mean of the low-frequency marginal distribution)**.** In this example, we are interested in the properties of the mean of the low-frequency marginal distribution, i.e. the mean $\theta^{\text{true}} \in \Theta := \mathbb{R}^d$ of $\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}$. Let us assume that the projection function π belongs to $\mathscr{F}$, and recall from Equation (3.2) that, for arbitrary $t \in \mathbb{Z}$,

$$\begin{aligned}
\theta^{\text{true}} &:= \mathbb{E}_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}\left[\mathbf{Y}_t\right] \\
&= \int_{\mathbb{R}^d} \mathbf{Y}_t \, \mathrm{d}\mathbb{P}_{\mathbf{Y}_t}^{\text{true}} \\
&= \int_{(\mathbb{R}^d)^{\mathbb{Z}}} \pi\left(\mathbf{Y}\right) \, \mathrm{d}\mathbb{P}_{\mathbf{Y}}^{\text{true}} \\
&= \mathbb{P}_{\mathbf{Y}}^{\text{true}} \pi.
\end{aligned}$$

Therefore, in order to have $\theta^{\text{true}} = \vartheta\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}\right)$, we define ϑ as $\vartheta(\mathbb{P}) := \mathbb{P}\pi$ for $\mathbb{P} \in \ell^\infty(\mathscr{F})$. This map is linear, since for $\alpha, \beta \in \mathbb{R}$ and $\mathbb{P}, \mathbb{Q} \in \ell^\infty(\mathscr{F})$,

$$\begin{aligned}
\vartheta\left(\alpha\mathbb{P} + \beta\mathbb{Q}\right) &= \left(\alpha\mathbb{P} + \beta\mathbb{Q}\right)\pi \\
&= \int \pi \, d\left(\alpha\mathbb{P} + \beta\mathbb{Q}\right) \\
&= \alpha \int \pi \, d\mathbb{P} + \beta \int \pi \, d\mathbb{Q} \\
&= \alpha\mathbb{P}\pi + \beta\mathbb{Q}\pi \\
&= \alpha\vartheta(\mathbb{P}) + \beta\vartheta(\mathbb{Q}).
\end{aligned}$$

Moreover, the map is continuous. This can be easily seen by defining μ as the metric induced by the ℓ^∞-norm and noting that, since $\pi \in \ell^\infty(\mathscr{F})$, then

$$\mu(\vartheta(\mathbb{P}), \vartheta(\mathbb{Q})) = \|\vartheta(\mathbb{P}) - \vartheta(\mathbb{Q})\|_\infty \leq \|\mathbb{P} - \mathbb{Q}\|_{\mathscr{F}}.$$

Thus, the map ϑ is Lipschitz continuous. Given that ϑ is linear and continuous, then it is also Hadamard-differentiable, and the derivative at an arbitrary $\mathbb{P}$ is ϑ itself, i.e. $\vartheta'_{\mathbb{P}}(\mathbb{Q}) = \vartheta(\mathbb{Q})$.

Now, let us suppose we have a sequence of estimated higher-frequency distributions $\left\{\hat{\mathbb{P}}_{\mathbf{X},n}\right\}_{n\in\mathbb{N}}$ such that $\hat{\mathbb{P}}_{\mathbf{X},n} \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{P}_{\mathbf{X}}^{\text{true}}$ **P**-almost surely. Given that ϑ is continuous, by Theorem 13, we find that $\hat{\theta}_n \to \theta^{\text{true}}$ **P**-almost surely, demonstrating strong consistency of the estimator. Further assume there is a sequence of numbers $\{r_n\}_{n\in\mathbb{N}}$, with $r_n \to \infty$, and a random element $\mathbb{T}$ that takes its values in $\ell^\infty(\mathscr{F})$ such that $r_n\left(\hat{\mathbb{P}}_{\mathbf{X},n} - \mathbb{P}_{\mathbf{X}}^{\text{true}}\right) \rightsquigarrow \mathbb{T}$. Then, by Theorem 14, $\left\{r_n\left(\hat{\theta}_n - \theta^{\text{true}}\right)\right\}_{n\in\mathbb{N}}$ converges weakly to $\vartheta\left(\varphi(\mathbb{T})\right)$, establishing that the rate of convergence of $\left\{\hat{\theta}_n\right\}_{n\in\mathbb{N}}$ is also given by $\{r_n\}_{n\in\mathbb{N}}$. Lastly, considering that $\Theta = \mathbb{R}^d$ is a Banach space under the infinity norm, if $\mathbb{T}$ is a tight Borel-measurable Gaussian map, then by Theorem 15, $\vartheta\left(\varphi(\mathbb{T})\right)$ is normally distributed.

Example 4 (Covariance matrix of the low-frequency marginal distribution)**.** We are now interested in the covariance matrix of the low-frequency marginal distribution $\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}$, which is formed by the covariance between the coordinates of $\mathbf{Y}_t \in \mathbb{R}^d$ for an arbitrary but unique time point $t \in \mathbb{Z}$. This is distinct from the autocovariance between different time points, which could also be of interest. Given the stationarity assumption, the covariance matrix remains constant across different time points.

As in the previous example, we assume that the projection function π belongs to $\mathscr{F}$. By the definition of the covariance matrix, the true parameter $\theta^{\text{true}} \in \Theta := \mathbb{R}^{d\times d}$ is given by

$$\theta^{\text{true}} := \mathbb{E}_{\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}}\left[\mathbf{Y}_0\mathbf{Y}_0^T\right] - \mathbb{E}_{\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}}\left[\mathbf{Y}_0\right]\mathbb{E}_{\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}}\left[\mathbf{Y}_0\right]^T.$$

The defining property of the marginal distribution allows us to exchange the marginal distribution $\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}$ in the sub-indices of the expectations for the joint distribution $\mathbb{P}_{\mathbf{Y}}^{\text{true}}$. Moreover, since $\pi(\mathbf{Y}) = \mathbf{Y}_0$, we have

$$\theta^{\text{true}} = \mathbb{E}_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}\left[\pi(\mathbf{Y})\,\pi(\mathbf{Y})^T\right] - \mathbb{E}_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}\left[\pi(\mathbf{Y})\right]\mathbb{E}_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}\left[\pi(\mathbf{Y})\right]^T.$$

Let us define the map Π as

$$\begin{aligned}
\Pi : \left(\mathbb{R}^d\right)^{\mathbb{Z}} &\to \mathbb{R}^{d\times d} \\
(\mathbf{x}_t)_{t\in\mathbb{Z}} &\to \mathbf{x}_0\mathbf{x}_0^T,
\end{aligned}$$

and suppose that $\Pi \in \mathscr{F}$. With this map, we can rewrite the true parameter as

$$\theta^{\text{true}} = \mathbb{E}_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}[\Pi(\mathbf{Y})] - \mathbb{E}_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}[\pi(\mathbf{Y})]\,\mathbb{E}_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}[\pi(\mathbf{Y})]^T,$$

and thus

$$\theta^{\text{true}} = \mathbb{P}_{\mathbf{Y}}^{\text{true}}\Pi - \left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}\pi\right)\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}\pi\right)^T =: \vartheta\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}\right),$$

which defines the map ϑ.

Clearly, for $\alpha \in \mathbb{R}$ and $\mathbb{P} \in \ell^\infty(\mathscr{F})$, $\alpha\vartheta(\mathbb{P})$ is not necessarily equal to $\vartheta(\alpha\mathbb{P})$, and thus, the map is not linear. However, it is continuous. To see this, let us remember that $\ell^\infty(\mathscr{F})$ is a metric space with respect to $d_{\mathscr{F}}$, and therefore continuity is equivalent to sequential continuity. Let us take a sequence $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ in $\ell^\infty(\mathscr{F})$ such that $\mathbb{P}_n \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{P}$. Then, for all $f \in \mathscr{F}$, it holds that $\mathbb{P}_n f \to \mathbb{P} f$. In particular, for Π and π, we have that $\mathbb{P}_n\Pi \to \mathbb{P}\Pi$ and $\mathbb{P}_n\pi \to \mathbb{P}\pi$. Thus,

$$\lim_{n\to\infty} \mathbb{P}_n\Pi - (\mathbb{P}_n\pi)(\mathbb{P}_n\pi)^T = \lim_{n\to\infty} \mathbb{P}_n\Pi - \left(\lim_{n\to\infty}\mathbb{P}_n\pi\right)\left(\lim_{n\to\infty}\mathbb{P}_n\pi\right)^T = \mathbb{P}\Pi - (\mathbb{P}\pi)(\mathbb{P}\pi)^T.$$

We then conclude that $\lim_{n\to\infty}\vartheta(\mathbb{P}_n) = \vartheta(\mathbb{P})$, proving the desired continuity of ϑ.

The map ϑ is also Hadamard-differentiable. To see this, take $\mathbb{P}, \mathbb{Q}_t \in \ell^\infty(\mathscr{F})$ and note that

$$\begin{aligned}
\vartheta(\mathbb{P}+t\mathbb{Q}_t) - \vartheta(\mathbb{P}) &= (\mathbb{P}+t\mathbb{Q}_t)\Pi - ((\mathbb{P}+t\mathbb{Q}_t)\pi)((\mathbb{P}+t\mathbb{Q}_t)\pi)^T - \mathbb{P}\Pi + (\mathbb{P}\pi)(\mathbb{P}\pi)^T \\
&= \mathbb{P}\Pi + t\mathbb{Q}_t\Pi - (\mathbb{P}\pi + t\mathbb{Q}_t\pi)(\mathbb{P}\pi + t\mathbb{Q}_t\pi)^T - \mathbb{P}\Pi + (\mathbb{P}\pi)(\mathbb{P}\pi)^T \\
&= t\mathbb{Q}_t\Pi - t(\mathbb{P}\pi)(\mathbb{Q}_t\pi)^T - t(\mathbb{Q}_t\pi)(\mathbb{P}\pi)^T - t^2(\mathbb{Q}_t\pi)(\mathbb{Q}_t\pi)^T.
\end{aligned}$$

Then,

$$\frac{\vartheta(\mathbb{P}+t\mathbb{Q}_t) - \vartheta(\mathbb{P})}{t} = \mathbb{Q}_t\Pi - (\mathbb{P}\pi)(\mathbb{Q}_t\pi)^T - (\mathbb{Q}_t\pi)(\mathbb{P}\pi)^T - t(\mathbb{Q}_t\pi)(\mathbb{Q}_t\pi)^T.$$

Thus, if we take $\vartheta'_{\mathbb{P}}(\mathbb{Q}) = \mathbb{Q}\Pi - (\mathbb{P}\pi)(\mathbb{Q}\pi)^T - (\mathbb{Q}\pi)(\mathbb{P}\pi)^T$, we have that

$$\begin{aligned}
\left\| \frac{\vartheta(\mathbb{P}+t\mathbb{Q}_t) - \vartheta(\mathbb{P})}{t} - \vartheta'_{\mathbb{P}}(\mathbb{Q}) \right\| &= \left\| (\mathbb{Q}_t - \mathbb{Q})\Pi + (\mathbb{P}\pi)((\mathbb{Q}-\mathbb{Q}_t)\pi)^T + ((\mathbb{Q}-\mathbb{Q}_t)\pi)(\mathbb{P}\pi)^T - t(\mathbb{Q}_t\pi)(\mathbb{Q}_t\pi)^T \right\| \\
&\leq \|\mathbb{Q}_t - \mathbb{Q}\|_{\mathscr{F}} + \|\mathbb{P}\|_{\mathscr{F}}\|\mathbb{Q}-\mathbb{Q}_t\|_{\mathscr{F}} + \|\mathbb{Q}-\mathbb{Q}_t\|_{\mathscr{F}}\|\mathbb{P}\|_{\mathscr{F}} + |t|\,\|\mathbb{Q}_t\|_{\mathscr{F}}^2 \\
&\to 0
\end{aligned}$$

as $t \downarrow 0$, for every $\mathbb{Q}_t \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{Q}$.

Since ϑ is continuous and Hadamard-differentiable, a similar analysis as in the case of the mean can be conducted. If we assume that we have a sequence $\left\{\hat{\mathbb{P}}_{\mathbf{X},n}\right\}_{n\in\mathbb{N}}$ such that $\hat{\mathbb{P}}_{\mathbf{X},n} \xrightarrow{\ell^\infty(\mathscr{F})} \mathbb{P}_{\mathbf{X}}^{\text{true}}$ $\mathbf{P}$-almost surely, then by Theorem 13, we find that $\left\{\hat{\theta}_n\right\}_{n\in\mathbb{N}}$ is strongly consistent. Furthermore, if we suppose that there exists a sequence of numbers $\{r_n\}_{n\in\mathbb{N}}$ with $r_n \to \infty$ and a random element $\mathbb{T}$ that takes its values in $\ell^\infty(\mathscr{F})$ such that

$r_n\left(\hat{\mathbb{P}}_{\mathbf{X},n} - \mathbb{P}_{\mathbf{X}}^{\text{true}}\right) \rightsquigarrow \mathbb{T}$, then by Theorem 14, we find that $\left\{r_n\left(\hat{\theta}_n - \theta^{\text{true}}\right)\right\}_{n\in\mathbb{N}}$ converges weakly to

$$\vartheta'_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}\left(\varphi\left(\mathbb{T}\right)\right) = \varphi\left(\mathbb{T}\right)\Pi - \left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}\pi\right)\left(\varphi\left(\mathbb{T}\right)\pi\right) - \left(\varphi\left(\mathbb{T}\right)\pi\right)\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}\pi\right)^T.$$

Finally, considering that $\Theta = \mathbb{R}^{d\times d}$ is a Banach space under the Frobenius norm, if $\mathbb{T}$ is a tight Borel-measurable Gaussian map, then by Theorem 15 we conclude that $\vartheta'_{\mathbb{P}_{\mathbf{Y}}^{\text{true}}}\left(\varphi\left(\mathbb{T}\right)\right)$ is normally distributed.

3.4. M-estimators

Let us now consider the case when $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ is a sequence of M-estimators. This includes common cases such as Maximum Likelihood Estimators and certain types of Minimum Distance Estimators. Given the generality of the conditions imposed on the estimators, we may only invoke some of the results established in Chapter 2, as specific conclusions are limited.

Strong consistency of general M-estimators can be proven using Lemmas 1 and 2, while convergence in probability of $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ to θ^{true} can be demonstrated by using Theorem 2. Similarly, readers can apply Theorem 5 to obtain the rate of convergence in the general case when $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ is a sequence of M-estimators of the proposed form. Finally, to address the asymptotic distribution of the estimators, readers can refer to Theorem 7.

In the specific cases where the criterion functions have the form

$$M_n\left(\theta\right) := \left(\mathbb{P}_n^{\text{emp}}\right)_{\mathbf{Y}} m_\theta, \tag{3.4}$$

where $\left(\mathbb{P}_n^{\text{emp}}\right)_{\mathbf{Y}}$ is the empirical distribution function associated with a sample path $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$ of $\mathbf{Y}$, the convergence in probability of $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ to θ^{true} can be proven using Theorem 3. On the other hand, to obtain the rate of convergence, readers can apply Theorem 6 or Corollary 1; the latter requiring a Lipschitz condition on m_θ. Lastly, to get the asymptotic distribution of the estimators, readers can refer to Theorems 8 and 9.

One particularly important case of criterion functions of the form (3.4) involves Maximum Likelihood Estimators (MLEs), in which such functions correspond to the negative log-likelihoods. MLEs are arguably among the most widely used estimators in practical applications. Many computational implementations for parameter estimation across various fields rely heavily on MLEs due to their effectiveness and simplicity.

Predicting the properties of the estimators in advance is challenging due to the general nature of the proposed framework. However, as extensively explored by van der Vaart [22] and further developed in collaboration with Wellner [23], M-estimators, as a broad category, exhibit characteristics that have been rigorously studied. A significant amount of research has been dedicated to exploring their properties and applications.

3.4.1. M-estimators with pseudo-observations

Recall from Section 2.3.1 that one of the methods for fitting copulas to pseudo-observations is Maximum Likelihood Estimation. In such cases, the criterion functions, given by the negative log-likelihood of the pseudo-observations, are not related to the empirical probability measure of the sample itself but rather to that of its pseudo-observations. Therefore, the form (3.4) does not capture these cases. However, some conclusions can still be drawn. Poignard & Fermanian [15] explore the properties of M-estimators with pseudo-observations. They prove consistency and provide the convergence rate of general (penalized) M-estimators based on pseudo-observations under certain regularity assumptions. Additionally, they apply these results to

Gaussian copulas, as well as to general Elliptical copulas and mixtures of copula models.

3.5. Minimum Distance Estimators

In the upcoming section, we consider the case where ϑ takes the form specified in Equation (3.3), resulting in $\hat{\theta}_n$ being a type of Minimum Distance Estimator. The family of Minimum Distance Estimators (MDEs) was introduced by Wolfowitz [25] in 1957. These versatile estimators possess some interesting properties. For instance, Parr & Schucany [14] compared the results of MDEs with the results of other M-estimators, and concluded that MDEs are more robust under model specification and easier to implement than competitive alternatives. In addition to robustness and simplicity, MDEs possess the invariance property of MLEs, as was shown by Drossos and Philippou [7]. In other words, for a transformation $u : \Theta \to \Lambda \subseteq \mathbb{R}^k$, if $\hat{\theta} \in \Theta$ is the MDE of $\theta \in \Theta$, then $\hat{\lambda} = u(\hat{\theta})$ is the MDE of $u(\theta)$.

Since this type of estimator is a specific instance of those discussed in the preceding section, our analysis relies on the theorems established there. The results presented in this section presume that d is a proper distance. However, in practical contexts, alternative discrepancy functions between probability measures might be preferred, such as Maximum Mean Discrepancy. In such occasions, the theorems provided in this section do not hold, and proper adaptations should be made.

Using Lemma 6 and Theorem 2, we can establish the consistency of $\hat{\theta}_n$ when d is defined as a strong distance. In this context, a "strong distance" is a distance function that ensures $\mathrm{d}(\mathbb{P}, \mathbb{Q}) \to 0$ whenever $\mathbb{P} \rightsquigarrow \mathbb{Q}$.

Theorem 16 (Consistency of $\hat{\theta}_n$ as MDE)**.** *Let Θ be compact, d be a distance function that is stronger than weak convergence, and $M_n(\theta) := -\mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_\theta\right)$, $M(\theta) := -\mathrm{d}\left(\mathbb{P}_{\mathbf{Y}}^{true}, \mathbb{P}_\theta\right)$. Suppose that $M(\theta)$ is continuous on Θ and uniquely maximized at θ^{true}. Further, suppose that $\hat{\mathbb{P}}_{\mathbf{X},n} \rightsquigarrow \mathbb{P}_{\mathbf{X}}^{true}$ $\mathbf{P}$-a.s. Then $\hat{\theta}_n \xrightarrow{\mathbf{P}} \theta^{true}$ $\mathbf{P}$-a.s.*

Proof. As mentioned in van der Vaart [22], for a compact set Θ and continuous function M, uniqueness of θ^{true} as maximizer implies that

$$\sup_{\theta : \mu(\theta, \theta^{\text{true}}) \geq \epsilon} M(\theta) < M\left(\theta^{\text{true}}\right).$$

Now, using triangle inequality and symmetry of the distance d,

$$M_n(\theta) - M(\theta) = \mathrm{d}\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}, \mathbb{P}_\theta\right) - \mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_\theta\right) \leq \mathrm{d}\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}, \hat{\mathbb{P}}_{\mathbf{Y},n}\right)$$
$$M(\theta) - M_n(\theta) = \mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_\theta\right) - \mathrm{d}\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}, \mathbb{P}_\theta\right) \leq \mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_{\mathbf{Y}}^{\text{true}}\right) = \mathrm{d}\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}, \hat{\mathbb{P}}_{\mathbf{Y},n}\right).$$

Thus, it holds that $|M_n(\theta) - M(\theta)| \leq d\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}, \hat{\mathbb{P}}_{\mathbf{Y},n}\right)$. Taking the supremum over Θ,

$$\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \leq \mathrm{d}\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}, \hat{\mathbb{P}}_{\mathbf{Y},n}\right).$$

By Lemma 6, $\hat{\mathbb{P}}_{\mathbf{X},n} \rightsquigarrow \mathbb{P}_{\mathbf{X}}^{\text{true}}$ $\mathbf{P}$-a.s. implies $\hat{\mathbb{P}}_{\mathbf{Y},n} \rightsquigarrow \mathbb{P}_{\mathbf{Y}}^{\text{true}}$ $\mathbf{P}$-a.s., and since d is stronger than weak convergence, we have that $\mathrm{d}\left(\mathbb{P}_{\mathbf{Y}}^{\text{true}}, \hat{\mathbb{P}}_{\mathbf{Y},n}\right) \to 0$ $\mathbf{P}$-almost surely as $n \to \infty$. This gives us

$$\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \to 0 \quad \mathbf{P}\text{-a.s.},$$

which in turn implies the convergence in probability of the supremum (recall Remark 2). The condition

$M_n\left(\hat{\theta}_n\right) \geq M_n\left(\theta^{\text{true}}\right) - o_P(1)$ follows directly from the definition of $\hat{\theta}_n = \underset{\theta\in\Theta}{\arg\min}\, \mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_\theta\right)$, since

$$\mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_{\hat{\theta}_n}\right) \leq \mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_{\theta^{\text{true}}}\right)$$

implies that

$$M_n\left(\hat{\theta}_n\right) = -\mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_{\hat{\theta}_n}\right) \geq -\mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_{\theta^{\text{true}}}\right) = M_n\left(\theta^{\text{true}}\right).$$

Thus, by Theorem 2, we conclude that $\hat{\theta}_n \xrightarrow{\mathbf{P}} \theta^{\text{true}}$. □

Remark 17. Note that, if $\mathbb{P}_\theta$ is identifiable and well-specified, then the uniqueness of θ^{true} as maximizer is guaranteed. The measure $\mathbb{P}_\theta$ is identifiable if for every $\theta_1, \theta_2 \in \Theta$ it holds that $\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2} \implies \theta_1 = \theta_2$. By properties of distances, we have that $M \leq 0$, and for $\theta \in \Theta$

$$M(\theta) = 0 \iff \mathrm{d}(\mathbb{P}^{\text{true}}_{\mathbf{Y}}, \mathbb{P}_\theta) = \mathrm{d}(\mathbb{P}_{\theta^{\text{true}}}, \mathbb{P}_\theta) = 0 \iff \mathbb{P}_\theta = \mathbb{P}_{\theta^{\text{true}}} \implies \theta = \theta^{\text{true}},$$

which proves that if $M(\theta) = 0$, then θ must be θ^{true}.

The following theorem is a modification of Theorem 5 for the case in which d is a distance.

Theorem 17 (Rate of convergence of $\hat{\theta}_n$ as MDE)**.** *Suppose that $\hat{\theta}_n \xrightarrow{\mathbf{P}^*} \theta^{true}$ and $-\mathrm{d}\left(\mathbb{P}_{\theta^{true}}, \mathbb{P}_\theta\right) \lesssim -\mu^2\left(\theta, \theta^{true}\right)$ for every θ in a neighborhood of θ^{true}. Assume further that, for every n and sufficiently small δ,*

$$\mathbb{E}^*\left[\sup_{\mu(\theta,\theta^{true})<\delta} \left|\mathrm{d}\left(\mathbb{P}_{\theta^{true}}, \mathbb{P}_\theta\right) + \mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_{\theta^{true}}\right) - \mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_\theta\right)\right|\right] \lesssim \frac{\delta^\alpha}{\sqrt{n}}$$

with $\alpha < 2$ (not depending on n). Then, $n^{\frac{1}{4-2\alpha}}\, \mu\left(\hat{\theta}_n, \theta^{true}\right) = O^*_{\mathbf{P}}(1)$.

Proof. Recall that, since the model is well-specified, it holds that $\mathbb{P}^{\text{true}}_{\mathbf{Y}} = \mathbb{P}_{\theta^{\text{true}}}$, so

$$M(\theta) - M\left(\theta^{\text{true}}\right) = -\mathrm{d}\left(\mathbb{P}^{\text{true}}_{\mathbf{Y}}, \mathbb{P}_\theta\right) + \mathrm{d}\left(\mathbb{P}^{\text{true}}_{\mathbf{Y}}, \mathbb{P}_{\theta^{\text{true}}}\right) = -\mathrm{d}\left(\mathbb{P}_{\theta^{\text{true}}}, \mathbb{P}_\theta\right) \lesssim -\mu^2\left(\theta, \theta^{\text{true}}\right)$$

for every θ in a neighborhood of θ^{true}. On the other hand,

$$\begin{aligned}
&(M_n - M)(\theta) - (M_n - M)\left(\theta^{\text{true}}\right) \\
&= -\mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_\theta\right) + \mathrm{d}\left(\mathbb{P}^{\text{true}}_{\mathbf{Y}}, \mathbb{P}_\theta\right) + \mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_{\theta^{\text{true}}}\right) - \mathrm{d}\left(\mathbb{P}^{\text{true}}_{\mathbf{Y}}, \mathbb{P}_{\theta^{\text{true}}}\right) \\
&= \mathrm{d}\left(\mathbb{P}_{\theta^{\text{true}}}, \mathbb{P}_\theta\right) + \mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_{\theta^{\text{true}}}\right) - \mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_\theta\right).
\end{aligned}$$

Then, for every $n \in \mathbb{N}$ and sufficiently small δ, we have that

$$\mathbb{E}^*\left[\sup_{\mu(\theta,\theta^{\text{true}})<\delta} \left|(M_n - M)(\theta) - (M_n - M)\left(\theta^{\text{true}}\right)\right|\right] \lesssim \frac{\delta^\alpha}{n}.$$

Note that $\phi_n(\delta) = \delta^\alpha$ is such that $\delta \mapsto \frac{\phi_n(\delta)}{\delta^\alpha}$ is constant (decreasing) in δ. Taking $r_n = n^{\frac{1}{4-2\alpha}}$, we have

$$r_n^2 \phi_n\left(\frac{1}{r_n}\right) = n^{\frac{2}{4-2\alpha}} n^{-\frac{\alpha}{4-2\alpha}} = \sqrt{n}.$$

Since by definition of M_n we have that $M_n\left(\hat{\theta}_n\right) \geq M_n\left(\theta^{\text{true}}\right)$, and $\left\{\hat{\theta}_n\right\}_{n\in\mathbb{N}}$ converges in outer probability to θ^{true}, by Theorem 5 we conclude that $n^{\frac{1}{4-2\alpha}} \mu\left(\hat{\theta}_n, \theta^{\text{true}}\right) = O^*_{\mathbf{P}}(1)$. □

Unlike for consistency and rate of convergence, we cannot apply directly any of the theorems listed in Section 2.2.3 to get the asymptotic distribution. This is due to the fact that $M(\theta) = -\mathrm{d}\left(\hat{\mathbb{P}}_{\mathbf{Y},n}, \mathbb{P}_\theta\right)$ is in general not twice continuously differentiable for a distance function d. However, when considering smoother discrepancy functions instead of distance functions, Theorem 7 might be useful.

4

Applications

In this chapter, our goal is to construct and test estimators for parameters of the low-frequency marginal distribution $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$. We examine the application of direct estimators, which are obtained directly from the low-frequency data, as well as simulation-based estimators, which are derived from data simulated from the estimated marginal distribution $\hat{\mathbb{P}}_{\mathbf{Y}_t,n}$. Section 4.1 establishes the framework of the different experiments carried on, while Sections 4.2-4.6 are dedicated to the study of estimators for different parameters of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$.

4.1. Experimental framework

The purpose of this section is to detail our experimental framework, which includes the general procedure carried out in all experiments, the choice of distributions for $\mathbb{P}^{\text{true}}_{\mathbf{X}}$, the assumptions for constructing $\hat{\mathbb{P}}_{\mathbf{X},n}$, and the metrics employed to evaluate the accuracy of estimators.

In every experiment, we draw a sample path $\mathbf{X}_1,\dots,\mathbf{X}_{nT}$ from $\mathbb{P}^{\text{true}}_{\mathbf{X}}$, from which we construct a sample path $\mathbf{Y}_1,\dots,\mathbf{Y}_n$ from $\mathbb{P}^{\text{true}}_{\mathbf{Y}}$ by aggregating over non-overlapping periods of length T. These samples represent the available information and follow the true distributions of the observed processes. From $\mathbf{Y}_1,\dots,\mathbf{Y}_n$ we construct direct estimators for parameters of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$, while from $\mathbf{X}_1,\dots,\mathbf{X}_{nT}$ we derive an estimation $\hat{\mathbb{P}}_{\mathbf{X},n}$ of $\mathbb{P}^{\text{true}}_{\mathbf{X}}$. From this estimation, we simulate m independent sample paths of length T,

$$\begin{gathered}\hat{\mathbf{X}}_1^{(1)},\dots,\hat{\mathbf{X}}_T^{(1)}\\ \vdots\\ \hat{\mathbf{X}}_1^{(m)},\dots,\hat{\mathbf{X}}_T^{(m)},\end{gathered}$$

and construct a sample $\hat{\mathbf{Y}}_1,\dots,\hat{\mathbf{Y}}_m \overset{\text{i.i.d.}}{\sim} \hat{\mathbb{P}}_{\mathbf{Y}_t,n}$ by aggregating each sample path. Finally, from this last sample, we build estimators for parameters of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$. As mentioned in Chapter 3, this final step might involve assuming

a parametric form for the distribution for $\hat{\mathbb{P}}_{\mathbf{Y}_t,n}$ based on $\hat{\mathbf{Y}}_1,\ldots,\hat{\mathbf{Y}}_m$ before deriving the estimators. In the specific cases in which the marginals are univariate, we denote the samples without bold characters, such as $X_1,\ldots,X_{nT}$ for the higher-frequency sample.

Our analysis includes simple examples in which the marginals $\mathbf{X}_t$ are independent and identically distributed, as well as examples where the marginals $\mathbf{X}_t$ of $\mathbf{X}$ are dependent on each other. In the former, a sample path of length nT from $\mathbb{P}^{\text{true}}_{\mathbf{X}}$ can be interpreted as nT independent draws from $\mathbb{P}^{\text{true}}_{\mathbf{X}_t}$. Additionally, because the sums that define $\mathbf{Y}$ are non-overlapping, if the marginals $\mathbf{X}_t$ are i.i.d., then the marginals $\mathbf{Y}_t$ of $\mathbf{Y}$ are also i.i.d. In such cases, the non-overlapping aggregation of m sample paths of length T from $\mathbb{P}^{\text{true}}_{\mathbf{X}}$ yields a sample of size m from $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$, which can be seen as a sample path of length m from $\mathbb{P}^{\text{true}}_{\mathbf{Y}}$.

We consider four different distributions for $\mathbf{X}$, which we refer to as the settings of the experiments. The first two settings are based on Gaussian independent and identically distributed data, while the other two are based on dependent time series data. These settings were purposely separated from the sections that define the estimators, since one single setting can give rise to several estimators for the parameters of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$, many of which we do not explore and could be subjects for future research. Moreover, the distributions $\mathbb{P}^{\text{true}}_{\mathbf{X}},\mathbb{P}^{\text{true}}_{\mathbf{Y}}$, $\hat{\mathbb{P}}_{\mathbf{X},n}$ and $\hat{\mathbb{P}}_{\mathbf{Y},n}$ of each setting can be described without the need to introduce estimators for parameters of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$.

4.1.1. Univariate normal setting

In this setting, all marginals of $\mathbf{X}$ are independent and identically distributed $X_t \sim N(\mu,\sigma^2) =: \mathbb{P}^{\text{true}}_{\mathbf{X}_t}$. Since the aggregated variables are i.i.d., the true distribution of Y_t is $N(T\mu,T\sigma^2) =: \mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$. We denote the mean, standard deviation and variance of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ as $\lambda := T\mu$, $\nu := \sigma\sqrt{T}$ and $\nu^2 = T\sigma^2$, respectively.

We consider three possible cases within this setting:

(C1) The mean μ is unknown, but the variance σ^2 is known. We define $\hat{\sigma}^2_n = \sigma^2$, and estimate μ as the sample mean of $X_1,\ldots,X_{nT}$, given by

$$\hat{\mu}_n = \frac{1}{nT}\sum_{i=1}^{nT} X_i. \tag{4.1}$$

Since the sample is i.i.d., we have that $\hat{\mu}_n \sim N\left(\mu, \frac{\sigma^2}{nT}\right)$.

(C2) The mean μ is known, but the variance σ^2 is unknown. We define $\hat{\mu}_n = \mu$, and estimate σ^2 by the sample variance of $X_1,\ldots,X_{nT}$, given by

$$\hat{\sigma}^2_n = \frac{1}{nT-1}\sum_{i=1}^{nT}\left(X_i-\hat{\mu}_n\right)^2, \tag{4.2}$$

where $\hat{\mu}_n$ is the sample mean defined in Equation (4.1). It is well-known that the sample variance of an i.i.d. sample $X_1,\ldots,X_{nT} \overset{\text{i.i.d.}}{\sim} N(\mu,\sigma^2)$ is such that

$$\frac{(nT-1)\,\hat{\sigma}^2_n}{\sigma^2} \sim \chi^2_{nT-1}.$$

Given that the expectation and variance of a chi-squared distribution with $nT-1$ degrees of freedom are $nT-1$ and $2(nT-1)$, respectively, then $\mathbb{E}\left[\hat{\sigma}^2_n\right] = \sigma^2$ and $Var\left[\hat{\sigma}^2_n\right] = \frac{2\sigma^4}{nT-1}$.

Let

$$g(k) = \sqrt{\frac{2}{k-1}}\left(\frac{\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right)}\right), \quad k > 1.$$

We define the sample standard deviation as the square root of $\hat{\sigma}_n^2$ and denote it $\hat{\sigma}_n$. Since $\frac{(nT-1)\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2_{nT-1}$, we have that its square root follows a chi distribution with $nT-1$ degrees-of-freedom, i.e. $\frac{\hat{\sigma}_n\sqrt{nT-1}}{\sigma} \sim \chi_{nT-1}$. The expectation and variance of a chi random variable R_k with k degrees-of-freedom are

$$\mathbb{E}[R_k] = \sqrt{2}\left(\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}\right) = \sqrt{k}g(k+1) \quad \text{and} \quad Var[R_k] = k - \mathbb{E}[R_k]^2 = k\left(1 - g(k+1)^2\right),$$

respectively. Therefore,

$$\mathbb{E}[\hat{\sigma}_n] = \sigma g(nT) \quad \text{and} \quad Var[\hat{\sigma}_n] = \sigma^2\left(1 - g(nT)^2\right).$$

(C3) Both the mean μ and the variance σ^2 are unknown. We estimate them as the sample mean and sample variance of $X_1, \ldots, X_{nT}$, respectively.

For all cases, we define our estimation of $\mathbb{P}^{\text{true}}_{\mathbf{X}_t}$ as $\hat{\mathbb{P}}_{\mathbf{X}_t,n} := N\left(\hat{\mu}_n, \hat{\sigma}_n^2\right)$. In Sections 4.2, 4.3, and 4.4.1, we test the performance of the estimators taking $\mu = 2$ and $\sigma^2 = 4$.

Remark 18. The factor $g(n)$ tends to 1 as n grows, and thus, the expectation and variance of $\hat{\sigma}_n$ converge to σ and zero, respectively. This can be proved by using Stirling's formula for the Gamma function (see Abramowitz & Stegun [1]):

$$\Gamma(z) \sim \exp(-z)z^{z-\frac{1}{2}}\sqrt{2\pi}\left(1 + \frac{1}{12z} + \frac{1}{288z^2} - \frac{139}{51840z^3} - \frac{571}{2488320z^4} + \ldots\right), \tag{4.3}$$

as $z \to +\infty$. Note first that

$$\frac{\exp\left(-\frac{n}{2}\right)}{\exp\left(-\frac{n-1}{2}\right)} = \exp\left(-\frac{1}{2}\right),$$

and

$$\frac{\left(\frac{n}{2}\right)^{\frac{n}{2}-\frac{1}{2}}}{\left(\frac{n-1}{2}\right)^{\frac{n-1}{2}-\frac{1}{2}}} = \sqrt{\frac{n-1}{2}}\left(\frac{\frac{n}{2}}{\frac{n-1}{2}}\right)^{\frac{n-1}{2}} = \sqrt{\frac{n-1}{2}}\left(\frac{\frac{n-1}{2}+\frac{1}{2}}{\frac{n-1}{2}}\right)^{\frac{n-1}{2}} = \sqrt{\frac{n-1}{2}}\left(1 + \frac{1}{n-1}\right)^{\frac{n-1}{2}}.$$

Therefore, using Equation (4.3) with the choices $z = \frac{n}{2}$ and $z = \frac{n-1}{2}$, we get

$$g(n) = \sqrt{\frac{2}{n-1}}\left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right) = \exp\left(-\frac{1}{2}\right)\left(1 + \frac{1}{n-1}\right)^{\frac{n-1}{2}}\left(\frac{1 + \frac{1}{6n} + O\left(\frac{1}{n^2}\right)}{1 + \frac{1}{6(n-1)} + O\left(\frac{1}{n^2}\right)}\right).$$

Recall the Taylor expansion $\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$ as $x \to 0$. Substituting $x = \frac{1}{n-1}$ gives

$$\begin{aligned}\left(1+\frac{1}{n-1}\right)^{\frac{n-1}{2}} &= \exp\left(\frac{n-1}{2}\ln\left(1+\frac{1}{n-1}\right)\right)\\ &= \exp\left(\frac{n-1}{2}\left(\frac{1}{n-1} - \frac{1}{2(n-1)^2} + O\left(\frac{1}{n^3}\right)\right)\right)\\ &= \exp\left(\frac{1}{2} - \frac{1}{4(n-1)} + O\left(\frac{1}{n^2}\right)\right) = \exp\left(\frac{1}{2} - \frac{1}{4n} + O\left(\frac{1}{n^2}\right)\right).\end{aligned}$$

Therefore,

$$g(n) = \exp\left(-\frac{1}{4n} + O\left(\frac{1}{n^2}\right)\right)\left(\frac{1+\frac{1}{6n}+O\left(\frac{1}{n^2}\right)}{1+\frac{1}{6(n-1)}+O\left(\frac{1}{n^2}\right)}\right) = \left(1-\frac{1}{4n}+O\left(\frac{1}{n^2}\right)\right)\left(\frac{1+\frac{1}{6n}+O\left(\frac{1}{n^2}\right)}{1+\frac{1}{6(n-1)}+O\left(\frac{1}{n^2}\right)}\right),$$

by the expansion $\exp(x) = 1 + x + O(x^2)$. We now use the expansion $\frac{1}{1+x} = 1 - x + O(x^2)$ when $x \to 0$, with $x = \frac{1}{6(n-1)} + O\left(\frac{1}{n^2}\right)$, which gives

$$\begin{aligned}g(n) &= \left(1-\frac{1}{4n}+O\left(\frac{1}{n^2}\right)\right)\left(1+\frac{1}{6n}+O\left(\frac{1}{n^2}\right)\right)\left(1-\frac{1}{6(n-1)}+O\left(\frac{1}{n^2}\right)\right)\\ &= 1 - \frac{1}{4n} + \frac{1}{6n} - \frac{1}{6(n-1)} + O\left(\frac{1}{n^2}\right)\\ &= 1 - \frac{1}{4n} + O\left(\frac{1}{n^2}\right).\end{aligned}$$

Then,

$$\mathbb{E}[\hat{\sigma}_n] = \sigma g(nT) = \sigma\left(1-\frac{1}{4nT}\right) + O\left(\frac{1}{n^2}\right) \quad \text{and} \quad Var[\hat{\sigma}_n] = \sigma^2\left(1-g(nT)^2\right) = \sigma^2\left(\frac{1}{2nT}\right) + O\left(\frac{1}{n^2}\right).$$

Finally, we can get a bound for g using Wendel's inequality (see Qi & Guo [16]):

$$\left(\frac{\frac{n-1}{2}}{\frac{n}{2}}\right)^{1-\frac{1}{2}} \leq \frac{\Gamma\left(\frac{n}{2}\right)}{\left(\frac{n-1}{2}\right)^{\frac{1}{2}}\Gamma\left(\frac{n-1}{2}\right)} \leq 1,$$

for $n > 1$, and therefore

$$\sqrt{1-\frac{1}{n}} \leq g(n) \leq 1.$$

Thus, $g(n)$ approximates 1 from below.

4.1.2. Bivariate normal setting

In this setting, all marginals of $\mathbf{X}$ are independent and identically distributed $\mathbf{X}_t \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}) =: \mathbb{P}^{\text{true}}_{\mathbf{X}_t}$, where $\boldsymbol{\mu} := (\mu_1, \mu_2) \in \mathbb{R}^2$ denotes the mean of $\mathbf{X}_t$ and $\Sigma \in \mathbb{R}^{2\times 2}$ its covariance matrix. Adopting the notation $\mathbf{X}_t =$

$(X_{t,1}, X_{t,2})$, the entries of the matrix are given by

$$\Sigma = \begin{pmatrix} Cov(X_{t,1}, X_{t,1}) & Cov(X_{t,1}, X_{t,2}) \\ Cov(X_{t,2}, X_{t,1}) & Cov(X_{t,2}, X_{t,2}) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where σ_1^2 is the variance of $X_{t,1}$, σ_2^2 the variance of $X_{t,2}$, and ρ the correlation coefficient between $X_{t,1}$ and $X_{t,2}$. By definition of the multivariate normal distribution, we know that $X_{t,1} \sim N(\mu_1, \sigma_1^2)$ and $X_{t,2} \sim N(\mu_2, \sigma_2^2)$. We denote the covariance between $X_{t,1}$ and $X_{t,2}$ as $\sigma_{1,2}$.

Let $\mathbf{Y} := \psi_T(\mathbf{X})$. As in the univariate case, the i.i.d. condition implies that $\mathbf{Y}_t := (Y_{t,1}, Y_{t,2})$ follows the distribution $N_2(T\boldsymbol{\mu}, T\boldsymbol{\Sigma})$. Thus its covariance matrix is given by

$$T\boldsymbol{\Sigma} = \begin{pmatrix} T\sigma_1^2 & T\rho\sigma_1\sigma_2 \\ T\rho\sigma_1\sigma_2 & T\sigma_2^2 \end{pmatrix},$$

while the distribution of the components $Y_{t,1}$ and $Y_{t,2}$ are $N(T\mu_1, T\sigma_1^2)$ and $N(T\mu_2, T\sigma_2^2)$, respectively. From the covariance matrix of $\mathbf{Y}_t$, it can be seen that the covariance between $Y_{t,1}$ and $Y_{t,2}$ is equal to $\rho(T\sigma_1^2)^{1/2}(T\sigma_2^2)^{1/2}$. Therefore, the correlation between $Y_{t,1}$ and $Y_{t,2}$ is

$$Corr[Y_{t,1}, Y_{t,2}] := \frac{Cov[Y_{t,1}, Y_{t,2}]}{\sqrt{Var[Y_{t,1}]}\sqrt{Var[Y_{t,2}]}} = \frac{\rho(T\sigma_1^2)^{1/2}(T\sigma_2^2)^{1/2}}{(T\sigma_1^2)^{1/2}(T\sigma_2^2)^{1/2}} = \rho,$$

which is the same as the one between $X_{t,1}$ and $X_{t,2}$.

The equivalence of the correlation between the components of $\mathbf{Y}_t$ and the correlation between the components of $\mathbf{X}_t$ holds not only when the marginals of $\mathbf{X}$ are i.i.d. bivariate normally distributed but also in every case in which the marginals of $\mathbf{X}$ are independent and have the same covariance and marginal variances, as proven in the following proposition.

Proposition 2. *Let $\mathbf{Y}$ be a random variable that takes values in $(\mathbb{R}^2)^{\mathbb{Z}}$ and can be characterized as $\mathbf{Y} := \psi_T(\mathbf{X})$ for a higher-frequency random variable $\mathbf{X}$ with values in $(\mathbb{R}^2)^{\mathbb{Z}}$ and an aggregation period $T \in \mathbb{N}$. Let us suppose that all marginals of $\mathbf{X}$ are independent to each other and have the same covariance $\sigma_{1,2}$ and marginal variances σ_1, σ_2. Then, every marginal $\mathbf{X}_t$ of $\mathbf{X}$ and every marginal $\mathbf{Y}_t$ of $\mathbf{Y}$ have correlation $\rho := \frac{\sigma_{1,2}}{\sigma_1\sigma_2}$.*

Proof. Let $t \in \mathbb{Z}$ be arbitrary. The correlation between $X_{t,1}$ and $X_{t,2}$ is given by

$$\rho_{X_{t,1}, X_{t,2}} := \frac{Cov[X_{t,1}, X_{t,2}]}{\sqrt{Var[X_{t,1}]}\sqrt{Var[X_{t,2}]}} = \frac{\sigma_{1,2}}{\sigma_1\sigma_2} =: \rho,$$

while the correlation between $Y_{t,1}$ and $Y_{t,2}$ reads

$$\rho_{Y_{t,1}, Y_{t,2}} := \frac{Cov[Y_{t,1}, Y_{t,2}]}{\sqrt{Var[Y_{t,1}]}\sqrt{Var[Y_{t,2}]}}. \tag{4.4}$$

Using the bilinearity of the covariance, we obtain that

$$\begin{aligned} Cov\left[Y_{t,1}, Y_{t,2}\right] &= Cov\left[\sum_{i=1}^{T} X_{T(t-1)+i,1}, \sum_{i=1}^{T} X_{T(t-1)+i,2}\right] \\ &= \sum_{i=1}^{T}\sum_{k=1}^{T} Cov\left[X_{T(t-1)+i,1}, X_{T(t-1)+k,2}\right]. \end{aligned}$$

Now, for $i \neq k$, $X_{T(t-1)+i,1}$ and $X_{T(t-1)+k,2}$ are independent and thus their covariance is zero, yielding

$$Cov\left[Y_{t,1}, Y_{t,2}\right] = \sum_{i=1}^{T} Cov\left[X_{T(t-1)+i,1}, X_{T(t-1)+i,2}\right] = \sum_{t=1}^{T} \sigma_{1,2} = T\sigma_{1,2}.$$

On the other hand, by the independence of each marginal $\mathbf{X}_t$,

$$Var\left[Y_{t,j}\right] = Var\left[\sum_{i=1}^{T} X_{T(t-1)+i,j}\right] = \sum_{i=1}^{T} Var\left[X_{T(t-1)+i,j}\right] = \sum_{i=1}^{T} \sigma_j^2 = T\sigma_j^2,$$

for $j = 1,2$. Thus, replacing these values in Equation (4.4), we get

$$\rho_{Y_{t,1},Y_{t,2}} = \frac{T\sigma_{1,2}}{\sqrt{T\sigma_1^2}\sqrt{T\sigma_2^2}} = \frac{\sigma_{1,2}}{\sigma_1\sigma_2} =: \rho,$$

which proves the assertion. □

We consider two possible cases within this setting:

(C4) The mean $\boldsymbol{\mu} := (\mu_1, \mu_2)$ and marginal variances σ_1^2, σ_2^2 are known, but the variance-covariance matrix $\boldsymbol{\Sigma}$ of $\mathbb{P}_{\mathbf{X}_t}^{\text{true}}$ is unknown. We define $\hat{\mu}_{n,1} := \mu_1$, $\hat{\mu}_{n,2} := \mu_2$, $\hat{\boldsymbol{\mu}}_n := (\hat{\mu}_{n,1}, \hat{\mu}_{n,2})$, $\hat{\sigma}_{n,1}^2 := \sigma_1^2$ and $\hat{\sigma}_{n,2}^2 := \sigma_2^2$, and we estimate $\boldsymbol{\Sigma}$ as

$$\hat{\boldsymbol{\Sigma}}_n := \begin{pmatrix} \hat{\sigma}_{n,1}^2 & \hat{r}_n\hat{\sigma}_{n,1}\hat{\sigma}_{n,2} \\ \hat{r}_n\hat{\sigma}_{n,1}\hat{\sigma}_{n,2} & \hat{\sigma}_{n,2}^2 \end{pmatrix}, \tag{4.5}$$

where

$$\hat{r}_n := \frac{\sum_{i=1}^{nT}\left(X_{i,1} - \hat{\mu}_{n,1}\right)\left(X_{i,2} - \hat{\mu}_{n,2}\right)}{\sqrt{\sum_{i=1}^{nT}\left(X_{i,1} - \hat{\mu}_{n,1}\right)^2}\sqrt{\sum_{i=1}^{nT}\left(X_{i,2} - \hat{\mu}_{n,2}\right)^2}}. \tag{4.6}$$

(C5) The mean $\boldsymbol{\mu} := (\mu_1, \mu_2)$, marginal variances σ_1^2, σ_2^2, and variance-covariance matrix $\boldsymbol{\Sigma}$ are unknown. We define $\hat{\mu}_{n,1}$ and $\hat{\mu}_{n,2}$ as the sample means of $X_{1,1}, \ldots, X_{nT,1}$ and $X_{1,2}, \ldots, X_{nT,2}$, respectively, and define $\hat{\boldsymbol{\mu}}_n := (\hat{\mu}_{n,1}, \hat{\mu}_{n,2})$. Similarly, we define $\hat{\sigma}_{n,1}^2$ and $\hat{\sigma}_{n,2}^2$ as the sample variances of $X_{1,1}, \ldots, X_{nT,1}$ and $X_{1,2}, \ldots, X_{nT,2}$, respectively. We use these estimators in Equation (4.6) to define $\hat{r}_n$, and we estimate $\boldsymbol{\Sigma}$ as in Equation (4.5).

For both cases, we estimate $\mathbb{P}_{\mathbf{X}_t}^{\text{true}}$ by $\hat{\mathbb{P}}_{\mathbf{X}_t,n} := N_2\left(\hat{\boldsymbol{\mu}}_n, \hat{\boldsymbol{\Sigma}}_n\right)$. In Section 4.5, we test the performance of the estimators using $\mu_1 = 2$, $\mu_2 = 2$, $\sigma_1^2 = 4$, $\sigma_2^2 = 4$, and $\rho = 0.5$.

4.1.3. Univariate time series setting

In this setting, the higher-frequency process $\mathbf{X}$ is an AR(2)-GARCH(1, 1) process, with standard normal innovations. Following Definition 28, the process is defined by the equations

$$\begin{aligned} X_t &= \mu_t + \epsilon_t \\ \mu_t &= \gamma_0 + \gamma_1 X_{t-1} + \gamma_2 X_{t-2} \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ \epsilon_t &= \sigma_t Z_t, \end{aligned}$$

where $\{Z_t\}_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} N(0,1)$. Due to the complexity of the model, the true distribution of $\mathbf{Y}_t$ remains unknown.

Within this setting, we consider two possibles cases:

(C6) The parameters of $\mathbb{P}_{\mathbf{X}}^{\text{true}}$ are unknown, and we assume that $\mathbf{X}$ is an AR(2)-GARCH(1, 1) process. Since the true process is indeed AR(2)-GARCH(1, 1), we have that the model assumed is well-specified.

(C7) The parameters of $\mathbb{P}_{\mathbf{X}}^{\text{true}}$ are unknown, and we assume that $\mathbf{X}$ is an AR(1)-GARCH(1, 1) process. Given that not all AR(2)-GARCH(1, 1) processes are included in the family of AR(1)-GARCH(1, 1) processes, the model assumed for $\mathbf{X}$ is misspecified.

We then estimate the parameters of these models via Maximum Likelihood Estimation for $\hat{\mathbb{P}}_{\mathbf{X},n}$. In Section 4.4.2, we test the performance of our estimators using $\gamma_0 = 0.5, \gamma_1 = 0.5, \gamma_2 = 0.3$ for the mean, and $\alpha_0 = 0.1$, $\alpha = 0.5$, $\beta_1 = 0.3$ for the variance of the true model.

4.1.4. Bivariate time series setting

In this setting, the higher-frequency process $\mathbf{X}$ can be described by a bivariate Copula-GARCH model, in which the mean and volatility form a bivariate ARMA(1, 1)-GARCH(1, 1) process with Student's t-marginal distributions for each component of the innovations, and with a Student's t-copula to model the dependency between these components.

Following Definition 28, each component of the process $\mathbf{X}$ is defined by the equations

$$\begin{aligned} X_{t,j} &= \mu_{t,j} + \epsilon_{t,j} \\ \mu_{t,j} &= \gamma_0 + \gamma_1 X_{t-1,j} + \eta_1 \epsilon_{t-1,j} \\ \sigma_{t,j}^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1,j}^2 + \beta_1 \sigma_{t-1,j}^2 \\ \epsilon_{t,j} &= \sigma_{t,i} Z_{t,j}, \end{aligned}$$

for $j \in \{1,2\}$. The innovations $\{(Z_{t,1}, Z_{t,2})\}_{t\in\mathbb{Z}}$ are generated by sampling an i.i.d. process $\{(U_{t,1}, U_{t,2})\}_{t\in\mathbb{Z}}$ from a Student's t-copula $C_{r,d}^{\text{t}}$ that is then transformed and normalized as $Z_{t,j} = \sqrt{\frac{\kappa-2}{\kappa}}\, t_\kappa^{-1}(U_{t,j})$, where κ are the degrees-of-freedom of the marginal distributions. In this way, the processes $\{Z_{t,1}\}_{t\in\mathbb{Z}}$ and $\{Z_{t,2}\}_{t\in\mathbb{Z}}$ are both i.i.d. with mean zero and variance one. As in the previous sub-section, the true distribution of $\mathbf{Y}_t$ remains unknown due to the complexity of the model.

To estimate the behaviour of $\mathbf{X}$, we first fit ARMA(1, 1)-GARCH(1, 1) models for the mean and the variance of each component of $\mathbf{X}$, considering Student's t-marginal distributions, and we calculate their residuals $(\hat{\epsilon}_{1,1}, \hat{\epsilon}_{1,2}), \ldots, (\hat{\epsilon}_{nT,1}, \hat{\epsilon}_{nT,2})$. Afterwards, we standardize the residuals based on the conditional standard deviation derived from the GARCH model, getting the innovations. We then compute the pseudo-observations of these innovations and fit a Student's t-copula via Maximum Likelihood Estimation. Since the true law of $\mathbf{X}$ belongs to the model class, the model is said to be well-specified.

In Section 4.6, we test the performance of the estimators using the parameters $\gamma_0 = 0.5$, $\gamma_1 = 0.6$, $\eta_1 = 0.4$ for the mean, and $\alpha_0 = 0.1$, $\alpha_1 = 0.5$, $\beta_1 = 0.3$ for the variance of the true model. The marginal distributions selected are Student's t-distributions with $\kappa = 5$ degrees-of-freedom. Furthermore, the correlation r and degrees-of-freedom d of the copula are set to $r = 0.7$ and $d = 0.5$, respectively.

4.1.5. Metrics

For every experiment, we conduct a simulation study using two parameter sets for the length n of the observed path of $\mathbf{Y}$ and the aggregation period T. The first set includes $n \in \{5, \ldots, 10\}$ and $T = 250$, while the second set is formed by $n \in \{35, \ldots, 40\}$ and $T = 70$. In the context of financial returns, the first set could represent five to ten years of yearly returns, while the second set could correspond to approximately nine to ten years of quarterly returns. Under normal settings, we sample $m = 10.000$ independent values from $\hat{\mathbb{P}}_{\mathbf{Y}_t,n}$ to construct the simulation-based estimators, and we repeat the process $K = 1.000$ times. Time series settings are more computationally intensive, so we adjust the parameters to $m = 1.000$ and $K = 100$.

For each estimator $\hat{p}$ for a parameter p, we report, where feasible, the expectation and variance of its error, defined as $\hat{p} - p$. Moreover, we evaluate two main metrics to assess the accuracy of these estimators. Before introducing the metrics, let us define the Mean Absolute Error (MAE) of $\hat{p}$ over K iterations as

$$MAE_K(\hat{p}) := \frac{\sum_{i=1}^{K} |\hat{p}_i - p|}{K},$$

where $\hat{p}_1, \ldots, \hat{p}_K$ represent K realizations of $\hat{p}$, each derived from a single iteration of the procedure used to construct the estimator. Similarly, let $\hat{q}_1, \ldots, \hat{q}_K$ be K realizations of another estimator $\hat{q}$ for the parameter p.

The metrics studied are:

- Mean Absolute Error over True Value: We calculate the ratio between $MAE_k(\hat{p})$ and the true value p of the parameter, i.e. $\frac{MAE_K(\hat{p})}{p}$. This metric provides an indication of the relative size of the Mean Absolute Error in contrast to the true value of the parameter, enabling comparisons of the errors of estimators across different scales.

  We also estimate the 95% confidence intervals of this quotient. To do so, we estimate the variance of the absolute error $|\hat{p} - p|$ as $\hat{s}^2 = \frac{\sum_{i=1}^{K} |\hat{p}_i - p|^2}{K-1}$, from which we estimate the variance of $MAE_k(\hat{p})$ as $\frac{\hat{s}^2}{K}$. We then estimate the 95% confidence intervals as

  $$\left[\frac{MAE_K(\hat{p})}{p} - \Phi(0.975) \sqrt{\frac{\hat{s}^2}{K}}, \; \frac{MAE_K(\hat{p})}{p} + \Phi(0.975) \sqrt{\frac{\hat{s}^2}{K}} \right],$$

  where Φ is the CDF of the standard normal distribution.

- Ratio of MAEs: We calculate the ratio $\frac{MAE_K(\hat{p})}{MAE_K(\hat{q})}$, providing a measure of $\hat{p}$'s performance relative to $\hat{q}$ in terms of their Mean Absolute Errors. Values lower than 1 indicate $\hat{p}$ performs better than $\hat{q}$, whereas values greater than 1 indicate worse performance.

  We also estimate the 95% confidence intervals of the ratio using the results of Derumigny et al. [6]. Assuming $\hat{p}_1, \ldots, \hat{p}_K$ and $\hat{q}_1, \ldots, \hat{q}_K$ are i.i.d., and both $\mathbb{E}[|\hat{p}_1 - p|], \mathbb{E}[|\hat{q}_1 - p|] < \infty$ with $\mathbb{E}[|\hat{q}_1 - p|] > 0$,

we have that

$$\sqrt{K}\left(\frac{MAE_K(\hat{p})}{MAE_K(\hat{q})} - \frac{\mathbb{E}\left[\left|\hat{p}-p\right|\right]}{\mathbb{E}\left[\left|\hat{q}-p\right|\right]}\right) \overset{K\to\infty}{\rightsquigarrow} N(0,M),$$

where

$$M := \frac{Var\left[\left|\hat{p}-p\right|\right]}{\mathbb{E}\left[\left|\hat{q}-p\right|\right]^2} + \frac{\mathbb{E}\left[\left|\hat{p}-p\right|\right]^2 Var\left[\left|\hat{q}-p\right|\right]}{\mathbb{E}\left[\left|\hat{q}-p\right|\right]^4} - 2\frac{Cov\left[\left|\hat{p}-p\right|,\left|\hat{q}-p\right|\right]\mathbb{E}\left[\left|\hat{p}-p\right|\right]}{\mathbb{E}\left[\left|\hat{q}-p\right|\right]^3}. \tag{4.7}$$

We estimate M by $\hat{M}$, replacing the expectations, variances, and covariance in the equation with their sample counterparts. Using $\hat{M}$, we calculate the standard deviation of $\frac{MAE_K(\hat{p})}{MAE_K(\hat{q})}$ as $\sqrt{\frac{\hat{M}}{K}}$, and we estimate the 95% confidence intervals as

$$\left[\frac{MAE_K(\hat{p})}{MAE_K(\hat{q})} - \Phi(0.975)\sqrt{\frac{\hat{M}}{K}}, \frac{MAE_K(\hat{p})}{MAE_K(\hat{q})} + \Phi(0.975)\sqrt{\frac{\hat{M}}{K}}\right].$$

4.2. Mean

Let us consider cases (C1) and (C3) within the univariate normal setting. The objective of this section is to compare the performance of direct and simulation-based estimators for the mean λ of the low-frequency marginal distribution $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$.

As a direct estimator for λ, we consider the sample mean of $Y_1,\dots,Y_n$, given by

$$\hat{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i. \tag{4.8}$$

Proposition 3. *The error of $\hat{\lambda}_n$ is normally distributed with expectation zero and variance $\frac{\nu^2}{n}$.*

Proof. Observe that

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n}\sum_{k=1}^{T} X_{T(i-1)+k} \sim N\left(nT\mu, nT\sigma^2\right),$$

and thus $\hat{\lambda}_n \sim N\left(\lambda, \frac{\nu^2}{n}\right)$. From this, it follows that the error of this estimator satisfies $\hat{\lambda}_n - \lambda \sim N\left(0, \frac{\nu^2}{n}\right)$. □

We propose a simulation-based estimator for λ, defined by

$$\hat{\lambda}_{n,m} = \frac{1}{m}\sum_{i=1}^{m} \hat{Y}_i. \tag{4.9}$$

Proposition 4. *The error of $\hat{\lambda}_{n,m}$ has expectation zero and variance $\nu^2\left(\frac{1}{n}+\frac{1}{m}\right)$.*

Proof. Since the sample $\hat{X}_1,\dots,\hat{X}_{mT}$ is i.i.d. we have that $\hat{Y}_1,\dots,\hat{Y}_m \overset{\text{i.i.d.}}{\sim} N\left(T\hat{\mu}_n, T\hat{\sigma}^2_n\right) =: \hat{\mathbb{P}}_{\mathbf{Y}_t,n}$ conditionally on $\hat{\mu}_n$ and $\hat{\sigma}^2_n$. Thus, conditionally on $\hat{\mu}_n$ and $\hat{\sigma}^2_n$, the estimator $\hat{\lambda}_{n,m}$ is distributed as $N\left(T\hat{\mu}_n, \frac{T\hat{\sigma}^2_n}{m}\right)$.

To calculate the unconditional mean and the unconditional variance of $\hat{\lambda}_{n,m}$, we use the Law of Total Expectation and the Law of Total Variance, which read

$$\mathbb{E}\left[\hat{\lambda}_{n,m}\right] = \mathbb{E}\left[\mathbb{E}\left[\hat{\lambda}_{n,m}|\hat{\mu}_n, \hat{\sigma}_n^2\right]\right] \tag{4.10}$$

and

$$Var\left[\hat{\lambda}_{n,m}\right] = \mathbb{E}\left[Var\left[\hat{\lambda}_{n,m}|\hat{\mu}_n, \hat{\sigma}_n^2\right]\right] + Var\left[\mathbb{E}\left[\hat{\lambda}_{n,m}|\hat{\mu}_n, \hat{\sigma}_n^2\right]\right], \tag{4.11}$$

respectively. Using the conditional expectation on Equation (4.10), we get

$$\mathbb{E}\left[\hat{\lambda}_{n,m}\right] = \mathbb{E}\left[T\hat{\mu}_n\right] = \lambda.$$

Similarly, using the conditional mean and conditional variance on Equation (4.11), we obtain

$$\begin{aligned} Var\left[\hat{\lambda}_{n,m}\right] &= \mathbb{E}\left[\frac{T\hat{\sigma}_n^2}{m}\right] + Var\left[T\hat{\mu}_n\right] \\ &= \frac{T\sigma^2}{m} + \frac{T^2\sigma^2}{nT} \\ &= \nu^2\left(\frac{1}{n} + \frac{1}{m}\right). \end{aligned}$$

Therefore the error of the estimator, given by $\hat{\lambda}_{n,m} - \lambda$, has expectation zero and variance $\nu^2\left(\frac{1}{n} + \frac{1}{m}\right)$. □

Remark 19. It can be seen from the Law of Total Expectation and the Law of Total Variance that the previous result holds for cases (C1) and (C3). This is because the expectation of a constant is the constant itself, and the sample variance is an unbiased estimator of the variance, which lead to

$$\mathbb{E}\left[\frac{T\hat{\sigma}_n^2}{m}\right] = \frac{T\sigma^2}{m} = \frac{\nu^2}{m}$$

in both cases.

Remark 20. Assuming normality for the higher-frequency marginal distribution $\mathbb{P}_{\mathbf{X}_t}^{\text{true}}$ implies that $\hat{\mu}_n$ and $\hat{\lambda}_n - \lambda$ are normally distributed. Nevertheless, if we do not assume normality, the i.i.d. assumption still provides the same results on the expectation and variance of these estimators.

Comparing Propositions 3 and 4, it is clear that the simulation-based estimator exhibits a higher variance than the direct estimator. As evidenced in Proposition 4, by the Law of Total Variance, the variance of the simulation-based estimator $\hat{\lambda}_{n,m}$ can be decomposed into two non-zero terms, one of which accounts for the full variance of the direct estimator $\hat{\lambda}_n$. More specifically, we have

$$Var\left[\hat{\lambda}_{n,m}\right] > Var\left[T\hat{\mu}_n\right] = \frac{\nu^2}{n} = Var\left[\hat{\lambda}_n\right].$$

In general, equalities like $Var\left[T\hat{\mu}_n\right] = Var\left[\hat{\lambda}_n\right]$, in which the variance of the direct estimator for the parameter of $\mathbb{P}_{\mathbf{Y}_t}^{\text{true}}$ is just the variance of T times the estimator for the same parameter of $\mathbb{P}_{\mathbf{X}_t}^{\text{true}}$, do not hold. However, the i.i.d. nature of the sample $Y_1, \ldots, Y_n$ and the fact that the sample mean depends linearly on each Y_i, yields the equality in this case.

Figure 4.1 illustrates the ratio $\frac{MAE_{1.000}(\hat{\lambda}_{n,m})}{MAE_{1.000}(\hat{\lambda}_n)}$. For both cases, and for every combination of T and n tested, the curves lie between 99.5% and 101%, indicating that, in terms of MAE, both estimators perform similarly

when considering $m = 10.000$ simulated points. However, the simulation-based estimator incurs a higher computational burden because of estimating $\hat{\mathbb{P}}_{\mathbf{X}_t,n}$ and the high number of samples drawn from this distribution. Consequently, employing the simulation-based estimator for the mean of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ is disadvantageous when $\mathbf{X}$ consists of independent and identically distributed time points.

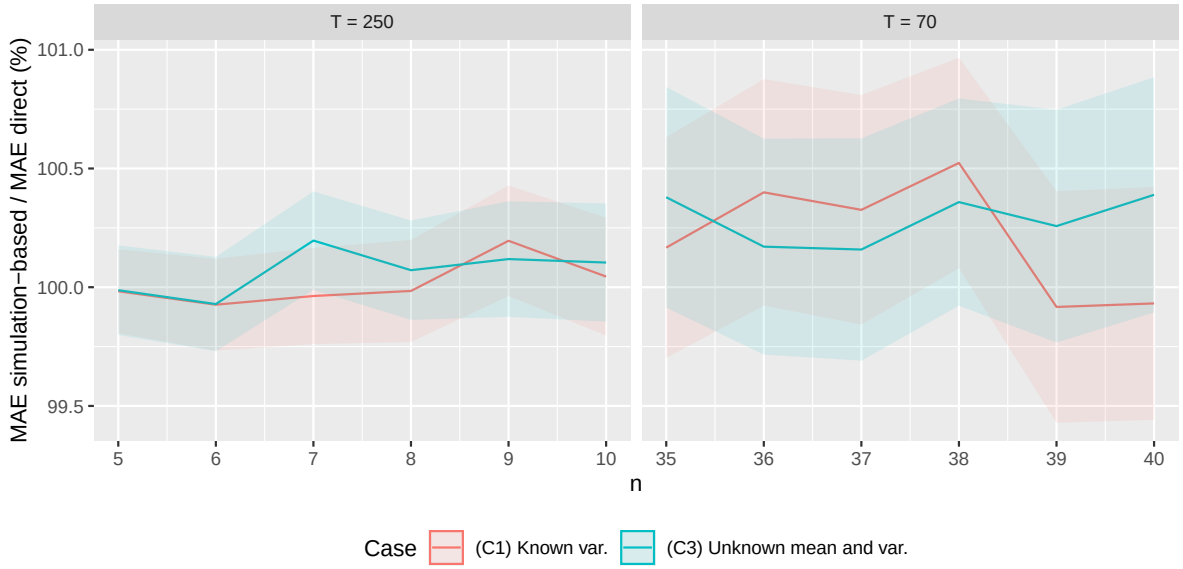


Figure 4.1: Ratio $\frac{MAE_{1.000}(\hat{\lambda}_{n,m})}{MAE_{1.000}(\hat{\lambda}_n)}$ for Cases (C1) and (C3), and for $T = 250$ (left) and $T = 70$ (right).

This graph illustrates the performance of the simulation-based and direct estimators for the mean λ of the low-frequency marginal distribution $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$, under two different aggregation periods ($T = 250$ and $T = 70$). Each panel plots the ratio of Mean Absolute Errors (MAEs) of the simulation-based estimator to the direct estimator against different low-frequency sample sizes n. The number of iterations used to estimate the MAEs was $K = 1.000$. The red lines represent Case (C1), in which the true mean of $\mathbf{X}_t$ is unknown but the true variance is known, while the teal lines represent Case (C3), in which both the mean and the variance of $\mathbf{X}_t$ are unknown (and so estimated). The shaded areas represent the 95% confidence intervals of the estimations. All curves lie between 99.5% and 101%, indicating that both estimators perform similarly in terms of MAE when considering $m = 10.000$ simulated values. However, the simulation-based estimator incurs a higher computational burden, and is therefore disadvantageous when $\mathbf{X}$ consists of independent and identically distributed time points.

4.3. Variance

Let us consider cases (C2) and (C3) within the univariate normal setting. The goal of this section is to compare the performance of direct and simulation-based estimators for the variance v^2 of the low-frequency marginal distribution $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$.

As a direct estimator for v^2, we consider the sample variance of $Y_1, \ldots, Y_n$, given by

$$\hat{v}^2_n = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \hat{\lambda}_n\right)^2, \tag{4.12}$$

where $\hat{\lambda}_n$ is the direct estimator for λ defined in Equation (4.8).

Proposition 5. *The error of $\hat{v}^2_n$ has expectation zero and variance $\frac{2v^4}{n-1} = \frac{2T^2\sigma^4}{n-1}$.*

Proof. Similarly to $\hat{\sigma}^2_n$, the sample variance $\hat{v}^2_n$ of a sample $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} N\left(\lambda, v^2\right)$ is such that

$$\frac{(n-1)\,\hat{v}^2_n}{v^2} \sim \chi^2_{n-1}.$$

Therefore, $\mathbb{E}\left[\hat{v}^2_n\right] = v$ and $Var\left[\hat{v}^2_n\right] = \frac{2v^4}{n-1}$. We conclude that the error $\hat{v}^2_n - v^2$ has expectation zero and variance $\frac{2v^4}{n-1}$. □

We propose a simulation-based estimator for v^2, defined by

$$\hat{v}^2_{n,m} = \frac{1}{m-1} \sum_{i=1}^{m} \left(\hat{Y}_i - \hat{\lambda}_{n,m}\right)^2, \tag{4.13}$$

where $\hat{\lambda}_{n,m}$ is the simulation-based estimator for λ defined in Equation (4.9).

Proposition 6. *The error of $\hat{v}^2_{n,m}$ has expectation zero and variance*

$$2v^4\left(\frac{2}{(m-1)(nT-1)} + \frac{1}{m-1} + \frac{1}{nT-1}\right) = 2T^2\sigma^4\left(\frac{2}{(m-1)(nT-1)} + \frac{1}{m-1} + \frac{1}{nT-1}\right).$$

Proof. Conditionally on $\hat{\mu}_n$ and $\hat{\sigma}^2_n$, we have that

$$\frac{(m-1)\,\hat{v}^2_{n,m}}{T\hat{\sigma}^2_n} \sim \chi^2_{m-1},$$

and therefore

$$\mathbb{E}\left[\hat{v}^2_{n,m} | \hat{\mu}_n, \hat{\sigma}^2_n\right] = T\hat{\sigma}^2_n, \quad Var\left[\hat{v}^2_{n,m} | \hat{\mu}_n, \hat{\sigma}^2_n\right] = \frac{2T^2\hat{\sigma}^4_n}{m-1}.$$

To derive its unconditional expectation and variance, we use the expectation and variance of $\hat{\sigma}^2_n$. By the Law of Total Expectation, we have

$$\mathbb{E}\left[\hat{v}^2_{n,m}\right] = \mathbb{E}\left[\mathbb{E}\left[\hat{v}^2_{n,m} | \hat{\mu}_n, \hat{\sigma}^2_n\right]\right] = \mathbb{E}\left[T\hat{\sigma}^2_n\right] = v^2,$$

which proves that the error $\hat{v}^2_{n,m} - v^2$ has expectation zero. On the other hand, by the Law of Total Variance, we get

$$\begin{aligned} Var\left[\hat{v}^2_{n,m}\right] &= \mathbb{E}\left[Var\left[\hat{v}^2_{n,m}|\hat{\mu}_n, \hat{\sigma}^2_n\right]\right] + Var\left[\mathbb{E}\left[\hat{v}^2_{n,m}|\hat{\mu}_n, \hat{\sigma}^2_n\right]\right] \\ &= \mathbb{E}\left[\frac{2T^2\hat{\sigma}^4_n}{m-1}\right] + Var\left[T\hat{\sigma}^2_n\right] \\ &= \frac{2T^2}{m-1}\mathbb{E}\left[\hat{\sigma}^4_n\right] + \frac{2v^4}{nT-1}. \end{aligned}$$

To derive $\mathbb{E}\left[\hat{\sigma}^4_n\right]$, we use the relation $\mathbb{E}\left[\hat{\sigma}^4_n\right] = Var\left[\hat{\sigma}^2_n\right] + \left(\mathbb{E}\left[\hat{\sigma}^2_n\right]\right)^2$, which yields

$$\begin{aligned} Var\left[\hat{v}^2_{n,m}\right] &= \frac{2T^2}{m-1}\left(Var\left[\hat{\sigma}^2_n\right] + \left(\mathbb{E}\left[\hat{\sigma}^2_n\right]\right)^2\right) + \frac{2v^4}{nT-1} \\ &= \frac{2T^2}{m-1}\left(\frac{2\sigma^4}{nT-1} + \sigma^4\right) + \frac{2v^4}{nT-1} \\ &= 2v^4\left(\frac{2}{(m-1)(nT-1)} + \frac{1}{m-1} + \frac{1}{nT-1}\right). \end{aligned}$$

The variance of $\hat{v}^2_{n,m} - v^2$ is the same as the variance of $\hat{v}^2_{n,m}$, yielding the result desired. □

Remark 21. Note that we use the sample mean to define $\hat{\sigma}^2_n$. Therefore, when our goal is to estimate the variance, the location parameter of the estimated distribution is irrelevant and cases (C2) and (C3) lead to the same analysis.

From Proposition 5, we have that the error of $\hat{v}^2_n$ has expectation zero and variance $\frac{2v^4}{n-1}$, while from Proposition 6, we get that $\hat{v}^2_{n,m}$ has expectation zero and variance $2v^4\left(\frac{2}{(m-1)(nT-1)} + \frac{1}{m-1} + \frac{1}{nT-1}\right)$. For fixed $T > 1$, a lower bound for m in order for the simulation-based estimator to have lower variance than the direct estimator is

$$\frac{2}{(m-1)(nT-1)} + \frac{1}{m-1} + \frac{1}{nT-1} < \frac{1}{n-1} \iff m > \frac{(n-1)(nT+1)}{n(T-1)} + 1. \tag{4.14}$$

In contrast to the case of the mean, the comparison between the two estimators clearly shows that the simulation-based estimator achieves a lower variance than the direct estimator when $m \gg n$ and $T > 1$. This reduction is mainly because the variance of the direct estimator is usually high when the sample size n of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ is small, while the variance of the simulation-based estimator depends mostly on the variance of the higher-frequency estimator, which can be low due to the factor nT.

Propositions 5 and 6 also show that a major contributor to the variances of the errors of $\hat{v}^2_n$ and $\hat{v}^2_{n,m}$ is the aggregation period T. Specifically, T has a quadratic effect on the variances, while the effects of n and m are just inversely proportional. This is specially significant in the case of the direct estimator, where a high aggregation period might be difficult to manage due to a small amount of low-frequency data n. However, in the case of the simulation-based estimator, the only quadratic term in the variance is $\frac{2T^2\sigma^4}{m-1}$. Despite the term $2T^2\sigma^4$ being potentially large, m can be increased arbitrarily, allowing $\frac{2T^2\sigma^4}{m-1}$ to be minimized as much as desired (within computational considerations).

Figure 4.2 displays the ratios $\frac{MAE_{1,000}(\hat{v}^2_n)}{v^2}$ and $\frac{MAE_{1,000}(\hat{v}^2_{n,m})}{v^2}$. The results significantly favor the simulation-based estimator over the direct estimator. This was expected from Equation (4.14), which yields a lower bound of $m = 10.04$ for the pair $(n,T) = (10,250)$ and $m = 40.58$ for $(n,T) = (40,70)$ in order for the simulation-based estimator to have lower variance than the direct one. The differences between cases (C2) and (C3) are not substantial. As previously mentioned, this is likely due to the fact that the location parameter of the estimated higher-frequency distribution is not relevant, since we still use the sample mean $\hat{\mu}_n$ to estimate $\hat{\sigma}^2_n$.

Figure 4.3 presents the ratio $\frac{MAE_{1.000}(\hat{v}^2_{n,m})}{MAE_{1.000}(\hat{v}^2_n)}$ for cases (C2) and (C3) and for both sets of parameters. Unlike the results for the mean, the MAE for the simulation-based estimator is only 5-7.5% of that for the direct estimator for $T = 250$. Similarly, for $T = 70$, the MAE of the simulation-based estimator is 12.5-15% of the MAE of the direct estimator, which underscores the superiority of the simulation-based method. These findings align with those shown in Figure 4.2. A notable observation is the reduced relative effectiveness of the simulation-based method for larger n and smaller T, even though nT remains relatively similar (e.g. comparing $(n,T) = (10,250)$ and $(n,T) = (35,70)$). This arises not from a decline in performance of the simulation-based estimator, but rather an improvement in the direct estimator for higher n and lower T. Despite this observation, the primary conclusion from these experiments is the substantial enhancement in the performance of the simulation-based estimator compared to the direct estimator across both sets of parameters and cases tested.

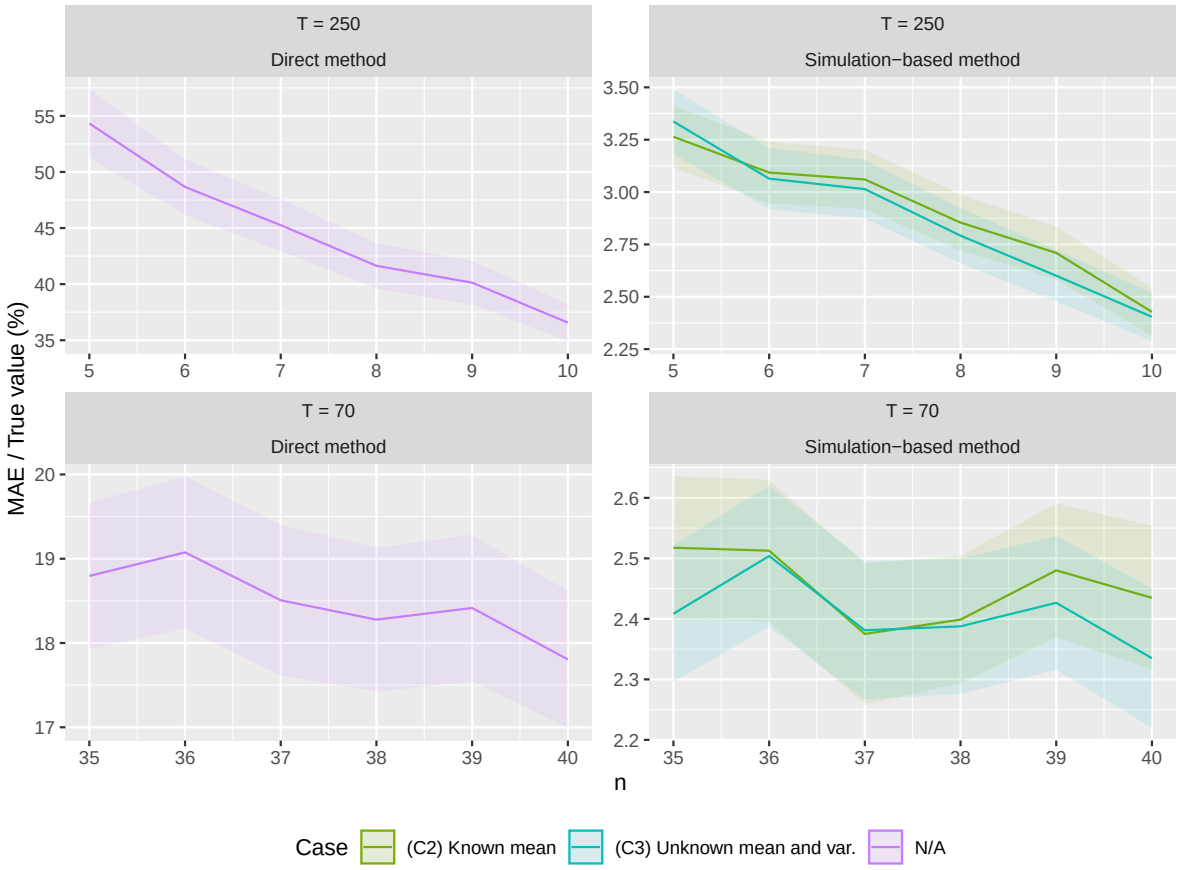


Figure 4.2: Ratios $\frac{MAE_{1.000}(\hat{v}^2_n)}{v^2}$ (left column) and $\frac{MAE_{1.000}(\hat{v}^2_{n,m})}{v^2}$ (right column) for Cases (C2) and (C3), and for $T = 250$ (top) and $T = 70$ (bottom).

This figure illustrates the performance of the simulation-based and direct estimators for the variance v^2 of the low-frequency marginal distribution $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$, under two different aggregation periods ($T = 250$ and $T = 70$). The left panels illustrate the ratios between the MAE of the direct estimator and its true value , while the right panels show the ratios between the MAE of the simulation-based estimator and its true value . The number of iterations used to estimate the MAEs was $K = 1.000$. The green lines represent Case (C2), in which the true mean of $\mathbf{X}_t$ is known but the true variance is unknown, while the teal lines represent Case (C3), in which both the mean and the variance of $\mathbf{X}_t$ are unknown (and thus estimated). The shaded areas represent the 95% confidence intervals of the estimations. The simulation-based method has considerably better results than the direct method for both $T = 250$ and $T = 70$. Note that the availability of the true mean does not significantly impact the accuracy of the simulation-based method, i.e. the differences between Cases (C2) and (C3) are not substantial.

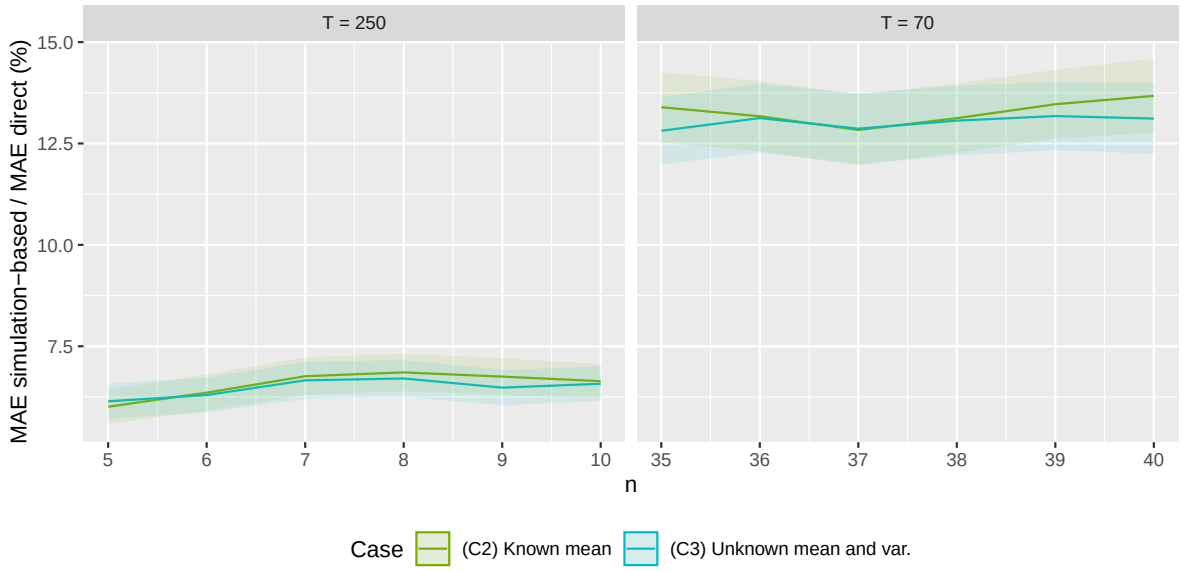


Figure 4.3: Ratio $\frac{MAE_{1.000}(\hat{v}^2_{n,m})}{MAE_{1.000}(\hat{v}^2_n)}$ for Cases (C2) and (C3), and for $T = 250$ (left) and $T = 70$ (right).

4.4. Quantiles

4.4.1. Univariate normal setting

Let us consider cases (C1), (C2) and (C3) within the univariate normal setting. The goal of this section is to compare the performance of direct and simulation-based estimators for the quantiles of the low-frequency marginal distribution $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$.

Let q_α be the α-quantile of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ for $\alpha \in (0,1)$, i.e. $q_\alpha = \left(F^{\text{true}}_{\mathbf{Y}_t}\right)^{-1}(\alpha)$, where $F^{\text{true}}_{\mathbf{Y}_t}$ denotes the CDF of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$. By properties of the normal distribution, q_α can be expressed as $q_\alpha = \lambda + \Phi^{-1}(\alpha)\nu$, where Φ is the standard normal CDF. Moreover, for $x \in \mathbb{R}$, it holds that $F^{\text{true}}_{\mathbf{Y}_t}(x) = \Phi\left(\frac{x-\lambda}{\nu}\right)$, and its derivative with respect to x is given by

$$\left(F^{\text{true}}_{\mathbf{Y}_t}\right)'(x) = \frac{1}{\nu}\phi\left(\frac{x-\lambda}{\nu}\right), \tag{4.15}$$

where ϕ is the standard normal density function. Since the formula for ϕ is known, we can use it in Equation (4.15) to get

$$\left(F^{\text{true}}_{\mathbf{Y}_t}\right)'(x) = \frac{1}{\nu\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\lambda}{\nu}\right)^2\right).$$

Note that

$$\left(F^{\text{true}}_{\mathbf{Y}_t}\right)'(q_\alpha) = \frac{1}{\nu\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\Phi^{-1}(\alpha)\right)^2\right).$$

As a direct non-parametric estimator for q_α we consider the α-empirical quantile of the sample $Y_1,\dots,Y_n$, which is given by

$$\hat{q}_{\alpha,\text{NP}} = Y_{n(i)}, \quad \text{for } \alpha \in \left(\frac{i-1}{n}, \frac{i}{n}\right],$$

where $Y_{n(i)}$ is the i-th order statistic of the sample.

Proposition 7. *The error of the direct non-parametric estimator $\hat{q}_{\alpha,NP}$ has expectation $O\left(n^{-\frac{3}{4}}\left(\log(n)\right)^{\frac{3}{4}}\right)$ and variance $\frac{2\pi\alpha(1-\alpha)\nu^2}{n}\exp\left(\left(\Phi^{-1}(\alpha)\right)^2\right) + O\left(n^{-\frac{5}{4}}\left(\log(n)\right)^{\frac{3}{4}}\right)$. Moreover, the scaled error satisfies*

$$\sqrt{n}\left(\hat{q}_{\alpha,NP} - q_\alpha\right) \rightsquigarrow N\left(0,\, 2\pi\alpha(1-\alpha)\nu^2\exp\left(\left(\Phi^{-1}(\alpha)\right)^2\right)\right)$$

as $n \to \infty$.

Proof. From Theorem 10,

$$\hat{q}_{\alpha,\text{NP}} - q_\alpha = L_n + R_n,$$

where $L_n = \frac{\alpha - \mathbb{F}_n(q_\alpha)}{\left(F^{\text{true}}_{\mathbf{Y}_t}\right)'(q_\alpha)}$ and $R_n = O\left(n^{-\frac{3}{4}}\left(\log(n)\right)^{\frac{3}{4}}\right)$, with $\mathbb{F}_n(q_\alpha) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\left(Y_i \le q_\alpha\right)$.

Given that $\mathbb{1}\left(Y_i \leq q_\alpha\right) \overset{\text{i.i.d.}}{\sim} Ber\left(\alpha\right)$, then

$$\mathbb{E}\left[\mathbb{F}_n\left(q_\alpha\right)\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{1}\left(Y_i \leq q_\alpha\right)\right] = \alpha$$

and

$$Var\left[\mathbb{F}_n\left(q_\alpha\right)\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var\left[\mathbb{1}\left(Y_i \leq q_\alpha\right)\right] = \frac{\alpha\left(1-\alpha\right)}{n}.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\hat{q}_{\alpha,\text{NP}} - q_\alpha\right] &= \mathbb{E}\left[L_n\right] + \mathbb{E}\left[R_n\right] = \frac{\alpha - \mathbb{E}\left[\mathbb{F}_n\left(q_\alpha\right)\right]}{\left(F^{\text{true}}_{\mathbf{Y}_t}\right)'\left(q_\alpha\right)} + O\left(n^{-\frac{3}{4}}\left(\log\left(n\right)\right)^{\frac{3}{4}}\right) \\ &= O\left(n^{-\frac{3}{4}}\left(\log\left(n\right)\right)^{\frac{3}{4}}\right).\end{aligned}$$

On the other hand,

$$Var\left[\hat{q}_{\alpha,\text{NP}} - q_\alpha\right] = Cov\left[L_n + R_n, L_n + R_n\right] = Var\left[L_n\right] + 2Cov\left[L_n, R_n\right] + Var\left[R_n\right].$$

The first term in the right yields

$$Var\left[L_n\right] = \frac{Var\left[\mathbb{F}_n\left(q_\alpha\right)\right]}{\left(\left(F^{\text{true}}_{\mathbf{Y}_t}\right)'\left(q_\alpha\right)\right)^2} = \frac{\alpha\left(1-\alpha\right)}{n\left(\left(F^{\text{true}}_{\mathbf{Y}_t}\right)'\left(q_\alpha\right)\right)^2},$$

while the third term can be bounded by

$$Var\left[R_n\right] \leq \mathbb{E}\left[R_n^2\right] = O\left(n^{-\frac{3}{2}}\left(\log\left(n\right)\right)^{\frac{3}{2}}\right).$$

The second term can be derived from the definition of covariance, resulting in

$$Cov\left[L_n, R_n\right] = \mathbb{E}\left[\left(L_n - \mathbb{E}\left[L_n\right]\right)\left(R_n - \mathbb{E}\left[R_n\right]\right)\right] \leq \mathbb{E}\left[\left|L_n\right|\left|R_n - \mathbb{E}\left[R_n\right]\right|\right].$$

Then,

$$\begin{aligned}Cov\left[L_n, R_n\right] &\lesssim n^{-\frac{3}{4}}\left(\log\left(n\right)\right)^{\frac{3}{4}}\mathbb{E}\left[\left|L_n\right|\right] \\ &\leq n^{-\frac{3}{4}}\left(\log\left(n\right)\right)^{\frac{3}{4}}\sqrt{\mathbb{E}\left[L_n^2\right]} \\ &= n^{-\frac{3}{4}}\left(\log\left(n\right)\right)^{\frac{3}{4}}\sqrt{Var\left[L_n\right]} \\ &\lesssim n^{-\frac{5}{4}}\left(\log\left(n\right)\right)^{\frac{3}{4}},\end{aligned}$$

and therefore

$$Var\left[\hat{q}_{\alpha,\text{NP}} - q_\alpha\right] = \frac{\alpha\left(1-\alpha\right)}{n\left(\left(F^{\text{true}}_{\mathbf{Y}_t}\right)'\left(q_\alpha\right)\right)^2} + O\left(n^{-\frac{5}{4}}\left(\log\left(n\right)\right)^{\frac{3}{4}}\right)$$

$$= \frac{2\pi\alpha\,(1-\alpha)\,\nu^2}{n} \exp\left(\left(\Phi^{-1}(\alpha)\right)^2\right) + O\left(n^{-\frac{5}{4}}\left(\log(n)\right)^{\frac{3}{4}}\right).$$

Finally, given that $\left(F^{\text{true}}_{\mathbf{Y}_t}\right)'$ is positive, by Corollary 2 we have that the scaled error $\sqrt{n}(\hat{q}_{\alpha,\text{NP}} - q_\alpha)$ is asymptotically normal with mean zero and variance

$$\frac{\alpha(1-\alpha)}{\left(\left(F^{\text{true}}_{\mathbf{Y}_t}\right)'(q_\alpha)\right)^2} = 2\pi\alpha(1-\alpha)\nu^2 \exp\left(\left(\Phi^{-1}(\alpha)\right)^2\right).$$

□

Now, let us assume that $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ belongs to the family of normal distributions $\{\mathbb{P}_\theta\}_{\theta\in\Theta} := \{N(\lambda,\nu^2) \,|\, \lambda\in\mathbb{R}, \nu\in(0,\infty)\}$, and estimate $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ as $\hat{\mathbb{P}}^{\text{dir}}_{\mathbf{Y}_t,n} := N(\hat{\lambda}_n, \hat{\nu}^2_n)$, where $\hat{\lambda}_n$ is the sample mean of $Y_1,\ldots,Y_n$ given in Equation (4.8) and $\hat{\nu}^2_n$ is its sample variance defined in Equation (4.12). Note that, since $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ is indeed normal, the model assumed is well-specified.

We define a direct parametric estimator for q_α as the α-quantile of $\hat{\mathbb{P}}^{\text{dir}}_{\mathbf{Y}_t,n}$, i.e.

$$\hat{q}_{\alpha,\text{P}} = \hat{\lambda}_n + \Phi^{-1}(\alpha)\,\hat{\nu}_n.$$

Proposition 8. *Consider case (C3). The error of the direct parametric estimator has expectation* $\nu\Phi^{-1}(\alpha)\left(g(n)-1\right)$ *and variance*

$$\nu^2\left(\frac{1}{n} + \left(\Phi^{-1}(\alpha)\right)^2\left(1-g(n)^2\right)\right).$$

Proof. The error of $\hat{q}_{\alpha,\text{P}}$ is given by

$$\begin{aligned}\hat{q}_{\alpha,\text{P}} - q_\alpha &= \hat{\lambda}_n + \Phi^{-1}(\alpha)\,\hat{\nu}_n - \left(\lambda + \Phi^{-1}(\alpha)\,\nu\right)\\ &= \left(\hat{\lambda}_n - \lambda\right) + \Phi^{-1}(\alpha)\left(\hat{\nu}_n - \nu\right).\end{aligned} \tag{4.16}$$

Consequently, the error is mostly influenced by the errors of $\hat{\lambda}_n$ and $\hat{\nu}_n$. By linearity of the expectation,

$$\begin{aligned}\mathbb{E}\left[\hat{q}_{\alpha,\text{P}} - q_\alpha\right] &= \mathbb{E}\left[\hat{\lambda}_n - \lambda\right] + \mathbb{E}\left[\Phi^{-1}(\alpha)\left(\hat{\nu}_n - \nu\right)\right]\\ &= \Phi^{-1}(\alpha)\left(\mathbb{E}\left[\hat{\nu}_n\right] - \nu\right).\end{aligned}$$

As in the case of $\hat{\sigma}_n$, we have that $\frac{\hat{\nu}_n\sqrt{n-1}}{\nu} \sim \chi_{n-1}$, and thus $\mathbb{E}[\hat{\nu}_n] = \nu g(n)$ and $Var[\hat{\nu}_n] = \nu^2\left(1-g(n)^2\right)$. Therefore, the error of $\hat{q}_{\alpha,\text{NP}}$ has expectation $\nu\Phi^{-1}(\alpha)\left(g(n)-1\right)$.

To derive the variance of the error, let us recall that, under the normality assumption, the sample mean $\hat{\lambda}_n$ and the sample standard deviation $\hat{\nu}_n$ are independent. Thus, we can decompose the variance to get

$$\begin{aligned}Var\left[\hat{q}_{\alpha,\text{P}} - q_\alpha\right] &= Var\left[\hat{\lambda}_n - \lambda\right] + Var\left[\Phi^{-1}(\alpha)\left(\hat{\nu}_n - \nu\right)\right]\\ &= \frac{\nu^2}{n} + \left(\Phi^{-1}(\alpha)\right)^2 Var\left[\hat{\nu}_n\right]\end{aligned}$$

$$= \nu^2 \left(\frac{1}{n} + \left(\Phi^{-1}(\alpha)\right)^2 \left(1 - g(n)^2\right) \right),$$

yielding the result desired. □

Remark 22. In case (C1), the error of the direct parametric estimator is $\hat{q}_{\alpha,\mathrm{P}} - q_\alpha = \hat{\lambda}_n - \lambda$. The expectation of the error is zero, and its variance is $\frac{\nu^2}{n}$. In case (C2), the error of $\hat{q}_{\alpha,\mathrm{P}}$ is $\hat{q}_{\alpha,\mathrm{P}} - q_\alpha = \Phi^{-1}(\alpha)(\hat{\nu}_n - \nu)$. Thus, the expectation of the error is $\nu\Phi^{-1}(\alpha)\big(g(n) - 1\big)$, which is the same as in case (C3). The variance, however, is $\nu^2 \left(\Phi^{-1}(\alpha)\right)^2 \left(1 - g(n)^2\right)$, lacking the $\frac{1}{n}$ factor present in case (C3).

As a simulation-based non-parametric estimator for q_α we consider the empirical quantile of the sample $\hat{Y}_1, \ldots, \hat{Y}_m$, given by

$$\hat{Q}_{\alpha,\mathrm{NP}} = \hat{Y}_{m(i)}, \quad \text{for } \alpha \in \left(\frac{i-1}{m}, \frac{i}{m} \right],$$

where $\hat{Y}_{m(i)}$ is the i-th order statistic of the sample $\hat{Y}_1, \ldots, \hat{Y}_m$.

Proposition 9. *Consider case (C3) an let $q_{\alpha,m}$ be the empirical quantile of an i.i.d. standard normal sample of length m. The error of the simulation-based non-parametric estimator $\hat{Q}_{\alpha,NP}$ has expectation*

$$\nu\Phi^{-1}(\alpha)\left(g(nT) - 1\right) + O\left(m^{-\frac{3}{4}} \left(\log(m)\right)^{\frac{3}{4}}\right)$$

and variance

$$\nu^2 \left(\frac{1}{n} + \frac{2\pi\alpha(1-\alpha)}{m} \exp\left(\left(\Phi^{-1}(\alpha)\right)^2 \right) + \left(\Phi^{-1}(\alpha)\right)^2 \left(1 - g(nT)^2\right) \right) + O\left(m^{-\frac{3}{4}} \left(\log(m)\right)^{\frac{3}{4}}\right).$$

Proof. Given $\hat{\mu}_n$ and $\hat{\sigma}_n$, $\hat{Y}_1, \ldots, \hat{Y}_m \overset{\text{i.i.d.}}{\sim} N\left(T\hat{\mu}_n, T\hat{\sigma}_n^2\right)$. Thus, for $i \in \{1, \ldots, m\}$, we can write $\hat{Y}_i = T\hat{\mu}_n + \hat{\sigma}_n\sqrt{T}Z_i$, where $Z_1, \ldots, Z_m \overset{\text{i.i.d.}}{\sim} N(0,1)$. This in turn implies that the non-parametric simulation-based estimator can be written in terms of $q_{\alpha,m}$ as

$$\hat{Q}_{\alpha,\mathrm{NP}} = T\hat{\mu}_n + \hat{\sigma}_n\sqrt{T}q_{\alpha,m}.$$

Using the aforementioned, the expectation of the non-parametric estimator $\hat{Q}_{\alpha,\mathrm{NP}}$ is given by

$$\begin{aligned} \mathbb{E}\left[\hat{Q}_{\alpha,\mathrm{NP}}\right] &= \mathbb{E}\left[T\hat{\mu}_n + \hat{\sigma}_n\sqrt{T}q_{\alpha,m}\right] \\ &= T\mathbb{E}\left[\hat{\mu}_n\right] + \sqrt{T}\mathbb{E}\left[\hat{\sigma}_n\right]\mathbb{E}\left[q_{\alpha,m}\right] \\ &= \lambda + \nu g(nT)\,\mathbb{E}\left[q_{\alpha,m}\right], \end{aligned}$$

where we also utilized the independence between $\hat{\sigma}_n$ and $q_{\alpha,m}$. Then, the error of $\hat{Q}_{\alpha,\mathrm{NP}}$ has expectation

$$\mathbb{E}\left[\hat{Q}_{\alpha,\mathrm{NP}} - q_\alpha\right] = \nu\left(g(nT)\,\mathbb{E}\left[q_{\alpha,m}\right] - \Phi^{-1}(\alpha)\right).$$

On the other hand, by first using the independence between $\hat{\mu}_n$ and $\hat{\sigma}_n$ due to the normality assumption,

and then the independence between $\hat{\sigma}_n$ and $q_{\alpha,m}$, we derive

$$\begin{aligned}
Var\left[\hat{Q}_{\alpha,\text{NP}}\right] &= Var\left[T\hat{\mu}_n + \sqrt{T}\hat{\sigma}_n q_{\alpha,m}\right] \\
&= T^2 Var\left[\hat{\mu}_n\right] + TVar\left[\hat{\sigma}_n q_{\alpha,m}\right] \\
&= \frac{\nu^2}{n} + T\left(\mathbb{E}\left[\left(\hat{\sigma}_n q_{\alpha,m}\right)^2\right] - \left(\mathbb{E}\left[\hat{\sigma}_n q_{\alpha,m}\right]\right)^2\right) \\
&= \frac{\nu^2}{n} + T\mathbb{E}\left[\hat{\sigma}_n^2\right]\mathbb{E}\left[q_{\alpha,m}^2\right] - T\mathbb{E}\left[\hat{\sigma}_n\right]^2\mathbb{E}\left[q_{\alpha,m}\right]^2 \\
&= \frac{\nu^2}{n} + \nu^2\mathbb{E}\left[q_{\alpha,m}^2\right] - \nu^2 g\left(nT\right)^2\mathbb{E}\left[q_{\alpha,m}\right]^2 \\
&= \frac{\nu^2}{n} + \nu^2\left(Var\left[q_{\alpha,m}\right] + \mathbb{E}\left[q_{\alpha,m}\right]^2\right) - \nu^2 g\left(nT\right)^2\mathbb{E}\left[q_{\alpha,m}\right]^2.
\end{aligned}$$

We then have that the error of $\hat{Q}_{\alpha,\text{NP}}$ has variance

$$Var\left[\hat{Q}_{\alpha,\text{NP}} - q_\alpha\right] = \nu^2\left(\frac{1}{n} + Var\left[q_{\alpha,m}\right] + \mathbb{E}\left[q_{\alpha,m}\right]^2\left(1 - g\left(nT\right)^2\right)\right).$$

To derive the moments of $q_{\alpha,m}$, we use Theorem 10 to get that $q_{\alpha,m}$ can be expressed as

$$q_{\alpha,m} = \Phi^{-1}\left(\alpha\right) + L_m + R_m,$$

where $L_m = \frac{\alpha - \mathbb{F}_m(q_\alpha)}{\phi(q_\alpha)}$ and $R_m = O\left(m^{-\frac{3}{4}}\left(\log\left(m\right)\right)^{\frac{3}{4}}\right)$, with ϕ the density of $N\left(0,1\right)$ and $\mathbb{F}_m\left(q_\alpha\right) = \frac{1}{m}\sum_{i=1}^m \mathbb{1}\left(Z_i \leq q_\alpha\right)$. By similar computations to the ones used in Proposition 7,

$$\mathbb{E}\left[q_{\alpha,m}\right] = \Phi^{-1}\left(\alpha\right) + O\left(m^{-\frac{3}{4}}\left(\log\left(m\right)\right)^{\frac{3}{4}}\right)$$

and

$$\begin{aligned}
Var\left[q_{\alpha,m}\right] &= \frac{\alpha\left(1-\alpha\right)}{m\left(\phi\left(\Phi^{-1}\left(\alpha\right)\right)\right)^2} + O\left(m^{-\frac{5}{4}}\left(\log\left(m\right)\right)^{\frac{3}{4}}\right) \\
&= \frac{2\pi\alpha\left(1-\alpha\right)}{m}\exp\left(\left(\Phi^{-1}\left(\alpha\right)\right)^2\right) + O\left(m^{-\frac{5}{4}}\left(\log\left(m\right)\right)^{\frac{3}{4}}\right).
\end{aligned}$$

Replacing the expectation and variance of $q_{\alpha,m}$ in the values obtained for the error of $\hat{Q}_{\alpha,\text{NP}}$, we get

$$\begin{aligned}
\mathbb{E}\left[\hat{Q}_{\alpha,\text{NP}} - q_\alpha\right] &= \nu\left(g\left(nT\right)\left(\Phi^{-1}\left(\alpha\right) + O\left(m^{-\frac{3}{4}}\left(\log\left(m\right)\right)^{\frac{3}{4}}\right)\right) - \Phi^{-1}\left(\alpha\right)\right) \\
&= \nu\Phi^{-1}\left(\alpha\right)\left(g\left(nT\right) - 1\right) + O\left(m^{-\frac{3}{4}}\left(\log\left(m\right)\right)^{\frac{3}{4}}\right)
\end{aligned}$$

and

$$\begin{aligned}
&Var\left[\hat{Q}_{\alpha,\text{NP}} - q_\alpha\right] \\
&= \nu^2\left(\frac{1}{n} + \frac{2\pi\alpha\left(1-\alpha\right)}{m}\exp\left(\left(\Phi^{-1}\left(\alpha\right)\right)^2\right) + O\left(m^{-\frac{5}{4}}\left(\log\left(m\right)\right)^{\frac{3}{4}}\right) + \left(1 - g\left(nT\right)^2\right)\left(\left(\Phi^{-1}\left(\alpha\right)\right)^2 + O\left(m^{-\frac{3}{4}}\left(\log\left(m\right)\right)^{\frac{3}{4}}\right)\right)\right) \\
&= \nu^2\left(\frac{1}{n} + \frac{2\pi\alpha\left(1-\alpha\right)}{m}\exp\left(\left(\Phi^{-1}\left(\alpha\right)\right)^2\right) + \left(\Phi^{-1}\left(\alpha\right)\right)^2\left(1 - g\left(nT\right)^2\right)\right) + O\left(m^{-\frac{3}{4}}\left(\log\left(m\right)\right)^{\frac{3}{4}}\right),
\end{aligned}$$

where we used that $g(nT) \leq 1$ for every $n \in \mathbb{N}$. □

Remark 23. Using the same procedure, it can be seen that, in case (C1), the expectation of the error is equal to

$$\nu\left(\mathbb{E}\left[q_{\alpha,m}\right] - \Phi^{-1}(\alpha)\right) = O\left(m^{-\frac{3}{4}}\left(\log(m)\right)^{\frac{3}{4}}\right),$$

while its variance yields

$$\nu^2\left(\frac{1}{n} + Var\left[q_{\alpha,m}\right]\right) = \nu^2\left(\frac{1}{n} + \frac{2\pi\alpha(1-\alpha)}{m}\exp\left(\left(\Phi^{-1}(\alpha)\right)^2\right)\right) + O\left(m^{-\frac{5}{4}}\left(\log(m)\right)^{\frac{3}{4}}\right).$$

On the other hand, in case (C2), the expectation of the error is the same as in case (C3), i.e.

$$\nu\Phi^{-1}(\alpha)\left(g(nT) - 1\right) + O\left(m^{-\frac{3}{4}}\left(\log(m)\right)^{\frac{3}{4}}\right),$$

while its variance is

$$\begin{aligned}
&\nu^2\left(Var\left[q_{\alpha,m}\right] + \mathbb{E}\left[q_{\alpha,m}\right]^2\left(1 - g(nT)^2\right)\right) \\
&= \nu^2\left(\frac{2\pi\alpha(1-\alpha)}{m}\exp\left(\left(\Phi^{-1}(\alpha)\right)^2\right) + \left(\Phi^{-1}(\alpha)\right)^2\left(1 - g(nT)^2\right)\right) + O\left(m^{-\frac{3}{4}}\left(\log(m)\right)^{\frac{3}{4}}\right),
\end{aligned}$$

removing the factor $\frac{1}{n}$ related to the variance of $\hat{\lambda}_n$.

Finally, let us estimate $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ as $\hat{\mathbb{P}}^{\text{sim}}_{\mathbf{Y}_t,n} := N\left(\hat{\lambda}_{n,m}, \hat{\nu}^2_{n,m}\right)$, for $\hat{\lambda}_{n,m}$ defined in Equation (4.9) and $\hat{\nu}^2_{n,m}$ given in Equation (4.13). We propose to use the α-quantile of $\hat{\mathbb{P}}^{\text{sim}}_{\mathbf{Y}_t,n}$, given by

$$\hat{Q}_{\alpha,\text{P}} = \hat{\lambda}_{n,m} + \Phi^{-1}(\alpha)\,\hat{\nu}_{n,m},$$

as a simulation-based non-parametric estimator for q_α.

Proposition 10. *Consider case (C3). The error of the simulation-based parametric estimator $\hat{Q}_{\alpha,P}$ has expectation $\nu\Phi^{-1}(\alpha)\left(g(nT)\,g(m) - 1\right)$ and variance*

$$\nu^2\left(\frac{1}{n} + \frac{1}{m} + \left(\Phi^{-1}(\alpha)\right)^2\left(1 - g(m)^2\,g(nT)^2\right)\right).$$

Proof. The expectation of the error can be expressed as

$$\begin{aligned}
\mathbb{E}\left[\hat{Q}_{\alpha,\text{P}} - q_\alpha\right] &= \mathbb{E}\left[\hat{\lambda}_{n,m} - \lambda\right] + \mathbb{E}\left[\Phi^{-1}(\alpha)\left(\hat{\nu}_{n,m} - \nu\right)\right] \\
&= \Phi^{-1}(\alpha)\left(\mathbb{E}\left[\hat{\nu}_{n,m}\right] - \nu\right).
\end{aligned}$$

We have that, conditionally on $\hat{\mu}_n$ and $\hat{\sigma}_n$, $\frac{\sqrt{m-1}\hat{\nu}_{n,m}}{\hat{\sigma}_n\sqrt{T}} \sim \chi_{m-1}$. Therefore, similarly to the case of $\hat{\nu}_n$,

$$\mathbb{E}\left[\hat{\nu}_{n,m}|\hat{\mu}_n,\hat{\sigma}_n\right] = \hat{\sigma}_n\sqrt{T}\,g(m) \quad \text{and} \quad Var\left[\hat{\nu}_{n,m}|\hat{\mu}_n,\hat{\sigma}_n\right] = T\hat{\sigma}_n^2\left(1 - g(m)^2\right).$$

By the Law of Total Expectation,

$$\mathbb{E}\left[\hat{\nu}_{n,m}\right] = \mathbb{E}\left[\mathbb{E}\left[\hat{\nu}_{n,m}|\hat{\mu}_n,\hat{\sigma}_n\right]\right] = \mathbb{E}\left[\hat{\sigma}_n\right]\sqrt{T}g\left(m\right) = \nu g\left(nT\right)g\left(m\right),$$

so the expectation of the error $\hat{Q}_{\alpha,\mathrm{P}} - q_\alpha$ is

$$\mathbb{E}\left[\hat{Q}_{\alpha,\mathrm{P}} - q_\alpha\right] = \nu\Phi^{-1}\left(\alpha\right)\left(g\left(nT\right)g\left(m\right) - 1\right).$$

Given $\hat{\mu}_n$ and $\hat{\sigma}_n$, we have that $\hat{\lambda}_{n,m}$ and $\hat{\nu}_{n,m}$ are independent by the normality assumption. Therefore, we can derive the conditional variance of the error as

$$\begin{aligned} Var\left[\hat{Q}_{\alpha,\mathrm{P}} - q_\alpha|\hat{\mu}_n,\hat{\sigma}_n\right] &= Var\left[\hat{\lambda}_{n,m}|\hat{\mu}_n,\hat{\sigma}_n\right] + \left(\Phi^{-1}\left(\alpha\right)\right)^2 Var\left[\hat{\nu}_{n,m}|\hat{\mu}_n,\hat{\sigma}_n\right] \\ &= \frac{T\hat{\sigma}_n^2}{m} + \left(\Phi^{-1}\left(\alpha\right)\right)^2 Var\left[\hat{\nu}_{n,m}|\hat{\sigma}_n\right] \\ &= \frac{T\hat{\sigma}_n^2}{m} + \left(\Phi^{-1}\left(\alpha\right)\right)^2 T\hat{\sigma}_n^2\left(1 - g\left(m\right)^2\right) \\ &= T\hat{\sigma}_n^2\left(\frac{1}{m} + \left(\Phi^{-1}\left(\alpha\right)\right)^2\left(1 - g\left(m\right)^2\right)\right). \end{aligned} \tag{4.17}$$

By the Law of Total Variance, the unconditional variance of $\hat{Q}_{\alpha,\mathrm{P}} - q_\alpha$ is

$$Var\left[\hat{Q}_{\alpha,\mathrm{P}} - q_\alpha\right] = \mathbb{E}\left[Var\left[\hat{Q}_{\alpha,\mathrm{P}} - q_\alpha|\hat{\mu}_n,\hat{\sigma}_n\right]\right] + Var\left[\mathbb{E}\left[\hat{Q}_{\alpha,\mathrm{P}} - q_\alpha|\hat{\mu}_n,\hat{\sigma}_n\right]\right].$$

Note that

$$\begin{aligned} \mathbb{E}\left[\hat{Q}_{\alpha,\mathrm{P}} - q_\alpha|\hat{\mu}_n,\hat{\sigma}_n\right] &= \mathbb{E}\left[\hat{\lambda}_{n,m} - T\mu|\hat{\mu}_n,\hat{\sigma}_n\right] + \mathbb{E}\left[\Phi^{-1}\left(\alpha\right)\left(\hat{\nu}_{n,m} - \sigma\sqrt{T}\right)|\hat{\mu}_n,\hat{\sigma}_n\right] \\ &= T\left(\hat{\mu}_n - \mu\right) + \Phi^{-1}\left(\alpha\right)\left(\mathbb{E}\left[\hat{\nu}_{n,m}|\hat{\mu}_n,\hat{\sigma}_n\right] - \sigma\sqrt{T}\right) \\ &= T\left(\hat{\mu}_n - \mu\right) + \Phi^{-1}\left(\alpha\right)\left(\hat{\sigma}_n\sqrt{T}g\left(m\right) - \sigma\sqrt{T}\right) \\ &= T\left(\hat{\mu}_n - \mu\right) + \Phi^{-1}\left(\alpha\right)\sqrt{T}\left(\hat{\sigma}_n g\left(m\right) - \sigma\right), \end{aligned}$$

so

$$\begin{aligned} Var\left[\hat{Q}_{\alpha,\mathrm{P}} - q_\alpha\right] &= \mathbb{E}\left[T\hat{\sigma}_n^2\left(\frac{1}{m} + \left(\Phi^{-1}\left(\alpha\right)\right)^2\left(1 - g\left(m\right)^2\right)\right)\right] + Var\left[T\left(\hat{\mu}_n - \mu\right) + \Phi^{-1}\left(\alpha\right)\sqrt{T}\left(\hat{\sigma}_n g\left(m\right) - \sigma\right)\right] \\ &= \nu^2\left(\frac{1}{m} + \left(\Phi^{-1}\left(\alpha\right)\right)^2\left(1 - g\left(m\right)^2\right)\right) + \frac{\nu^2}{n} + \nu^2\left(\Phi^{-1}\left(\alpha\right)\right)^2 g\left(m\right)^2\left(1 - g\left(nT\right)^2\right), \end{aligned}$$

where we used the independence between $\hat{\mu}_n$ and $\hat{\sigma}_n$ under the normality assumption to split the variance in the right. Rearranging the terms, we conclude that

$$Var\left[\hat{Q}_{\alpha,\mathrm{P}} - q_\alpha\right] = \nu^2\left(\frac{1}{n} + \frac{1}{m} + \left(\Phi^{-1}\left(\alpha\right)\right)^2\left(1 - g\left(m\right)^2 g\left(nT\right)^2\right)\right).$$

□

Remark 24. In case (C1), the expectation of the error is given by $\nu\Phi^{-1}\left(\alpha\right)\left(g\left(m\right) - 1\right)$, while its variance

is $v^2\left(\frac{1}{n}+\frac{1}{m}+\left(\Phi^{-1}(\alpha)\right)^2\left(1-g(m)^2\right)\right)$. On the other hand, in case (C2), the expectation of the error is still $v\Phi^{-1}(\alpha)\left(g(nT)\,g(m)-1\right)$, while the variance is given by $v^2\left(\frac{1}{m}+\left(\Phi^{-1}(\alpha)\right)^2\left(1-g(m)^2\,g(nT)^2\right)\right)$, which does not include the factor $\frac{1}{n}$ related to the variance of $\hat{\lambda}_n$.

Figure 4.4 presents the ratios $\frac{MAE_{1.000}(\hat{q}_{0.99,\mathrm{NP}})}{q_{0.99}}$ and $\frac{MAE_{1.000}(\hat{Q}_{0.99,\mathrm{NP}})}{q_{0.99}}$ of the non-parametric estimators for $q_{0.99}$. Similarly, Figure 4.6 presents the same metrics for the parametric estimators $\hat{q}_{0.99,\mathrm{P}}$ and $\hat{Q}_{0.99,\mathrm{P}}$. Both figures clearly demonstrate that the simulation-based method yields a much lower MAE than the direct method across all cases. Additionally, a comparison between the two figures shows that the parametric direct estimator achieves a lower MAE compared to the non-parametric direct estimator, suggesting that having knowledge of the distribution is beneficial for quantile estimation. This advantage of knowing the underlying distribution is also evident, though to a lesser extent, when comparing the simulation-based estimators.

In general, the error in quantile estimation depends both on the accuracy of the approximation of the mean and the accuracy of the approximation of the standard deviation. As seen in Section 4.2, the direct and simulation-based methods achieve similar accuracy when estimating the mean. Therefore, the difference in performance between the direct and simulation-based methods for quantile estimation is primarily attributed to the better approximation of the standard deviation provided by the simulation-based method. This aspect is further highlighted when comparing the performance of Case (C2) with Cases (C1) and (C3). For Case (C2) in specific, given that the mean is known, the error depends only on the accuracy of the approximation of the standard deviation. Thus, since the standard deviation is approximated faster than the mean, the simulation-based method has a much better performance in case (C2) than in cases (C1) and (C3). This particular feature of quantile estimation is derived mathematically in Remarks 23 and 24, and is clearly portrayed in Figures 4.4 and 4.6.

Figure 4.5 illustrates the ratio $\frac{MAE_{1.000}(\hat{q}_{0.99,\mathrm{NP}})}{MAE_{1.000}(\hat{Q}_{0.99,\mathrm{NP}})}$, while Figure 4.7 shows the analogous ratio for parametric estimators. Across all tested values of n, the MAE of the simulation-based non-parametric estimator is approximately 30% of the MAE of the direct non-parametric estimator in Cases (C1) and (C3). In Case (C2), this proportion is approximately 5% for $T = 250$ and 10% for $T = 70$. Similarly, the MAE of the simulation-based parametric estimator is about 50% of the MAE of the direct parametric estimator in Cases (C1) and (C3). In Case (C2), this ratio is close to 6% for $T = 250$ and 13% for $T = 70$. It may be tempting to conclude that non-parametric estimators are superior than parametric estimators. However, as mentioned before, the direct parametric estimator performs significantly better than the direct non-parametric estimator. This performance disparity, together with the less significant difference between the simulation-based estimators, explain why the MAEs of the parametric estimators are closer than the ones of the non-parametric estimators.

We conclude that employing simulation-based estimators for quantile estimation is advantageous when dealing with i.i.d. normal data. For both direct and simulation-based estimators, using parametric estimators is preferable when the underlying distribution is known.

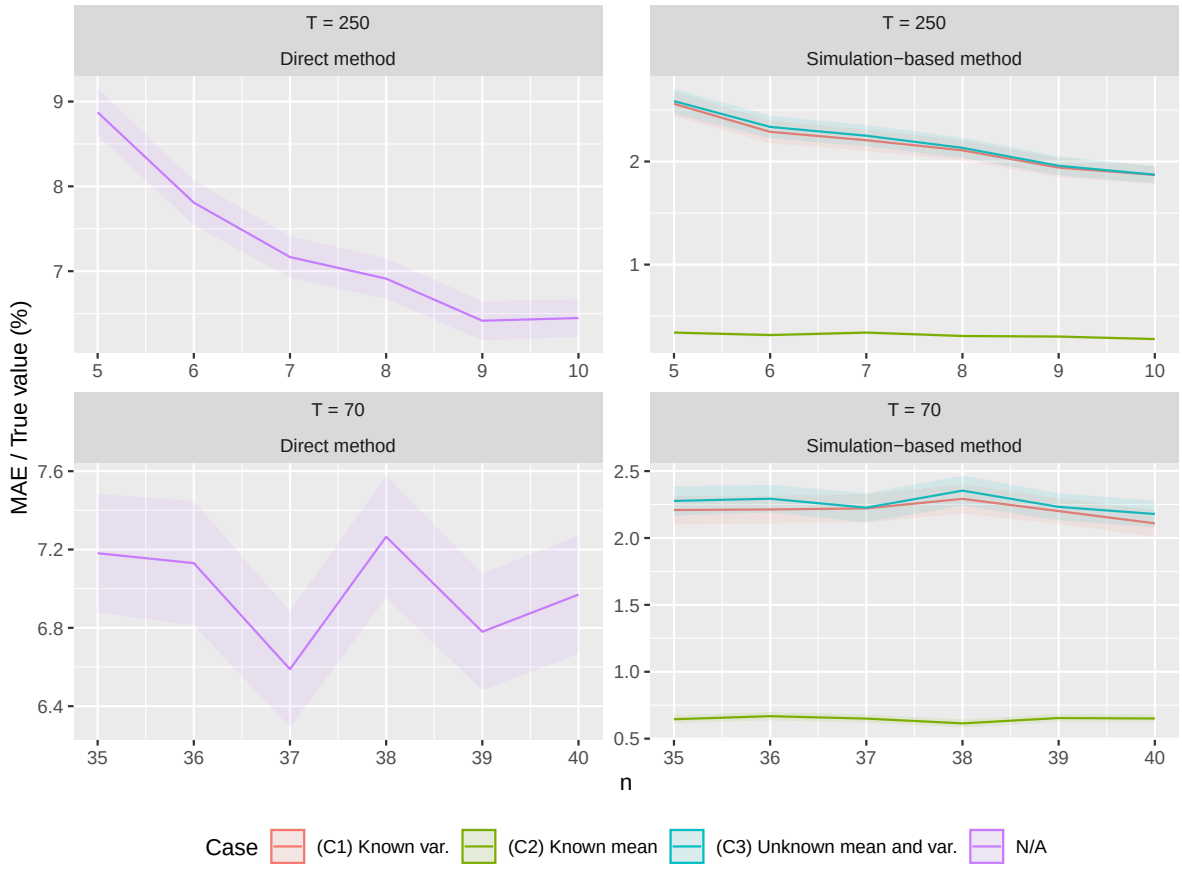


Figure 4.4: Ratios $\frac{MAE_{1.000}(\hat{q}_{0.99,\mathrm{NP}})}{q_{0.99}}$ (left column) and $\frac{MAE_{1.000}(\hat{Q}_{0.99,\mathrm{NP}})}{q_{0.99}}$ (right column) for Cases (C1), (C2) and (C3), and for $T = 250$ (top row) and $T = 70$ (bottom row).

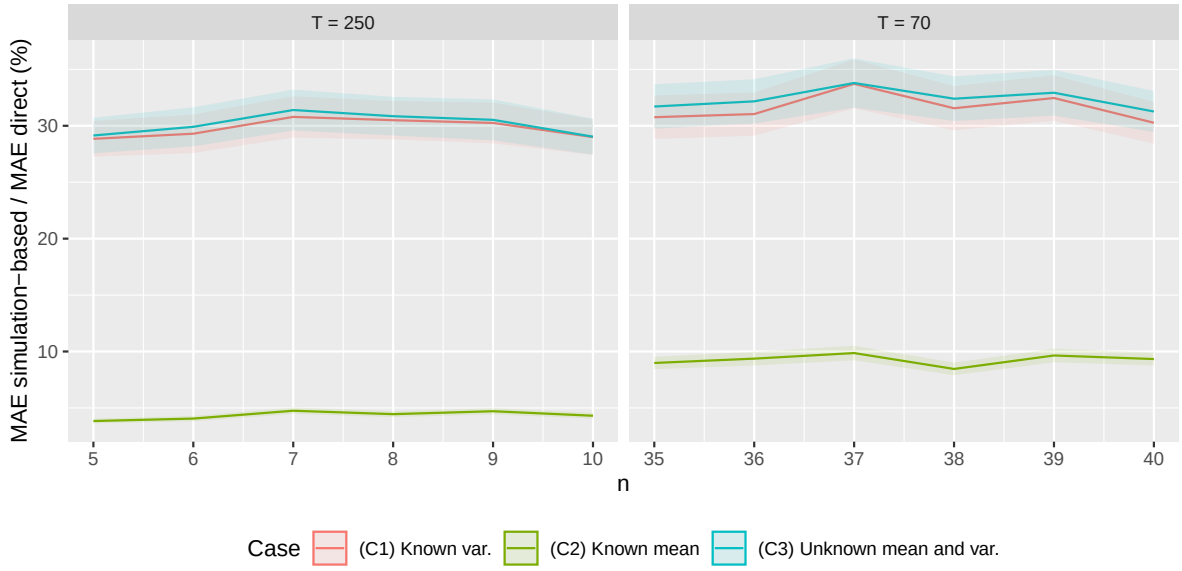


Figure 4.5: Ratio $\frac{MAE_{1.000}(\hat{Q}_{0.99,\mathrm{NP}})}{MAE_{1.000}(\hat{q}_{0.99,\mathrm{NP}})}$ for Cases (C1), (C2) and (C3), and for $T = 250$ (left) and $T = 70$ (right).

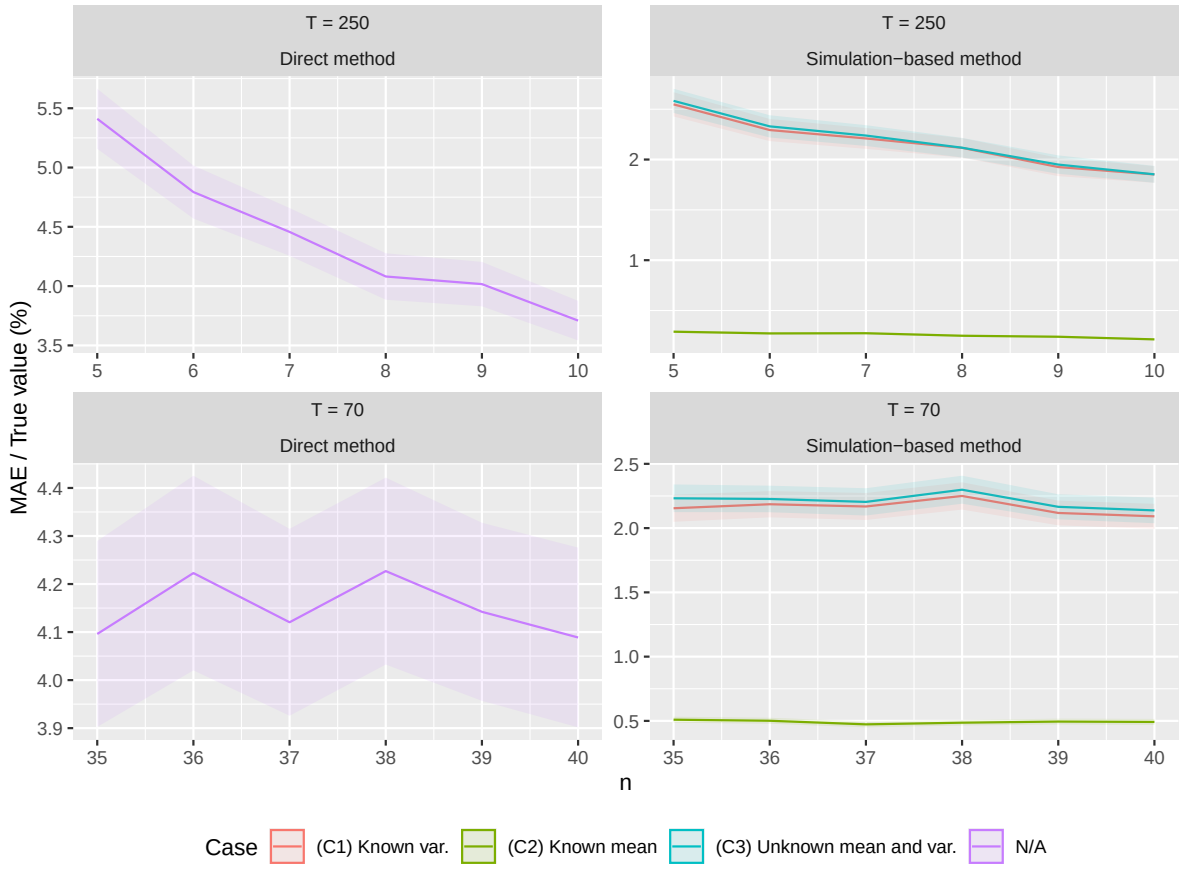


Figure 4.6: Ratios $\frac{MAE_{1.000}(\hat{q}_{0.99,\mathrm{P}})}{q_{0.99}}$ (left column) and $\frac{MAE_{1.000}(\hat{Q}_{0.99,\mathrm{P}})}{q_{0.99}}$ (right column) for Cases (C1), (C2) and (C3), and for $T = 250$ (top row) and $T = 70$ (bottom row).

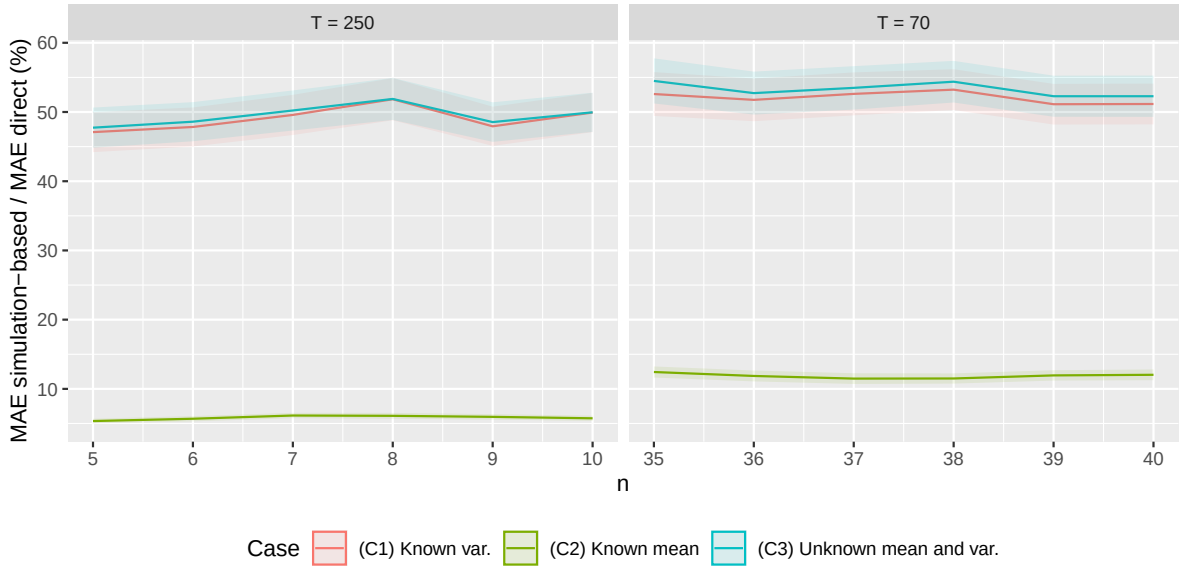


Figure 4.7: Ratio $\frac{MAE_{1.000}(\hat{Q}_{0.99,\mathrm{P}})}{MAE_{1.000}(\hat{q}_{0.99,\mathrm{P}})}$ for Cases (C1), (C2) and (C3), and for $T = 250$ (left) and $T = 70$ (right).

4.4.2. Univariate time series setting

Consider cases (C6) and (C7) within the univariate time series setting. In this section, we compare the performance of direct and simulation-based estimators for the quantiles of the low-frequency marginal distribution $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$. Additionally, we examine how well-specified and misspecified models for $\hat{\mathbb{P}}_{\mathbf{X},n}$ impact the performance of simulation-based estimators, as well as how the performance of simulation-based estimators gets better as m increases.

Let $q^{\text{t.s.}}_\alpha$ be the α-quantile of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$ for $\alpha \in (0,1)$. This is defined as $q^{\text{t.s.}}_\alpha := \left(F^{\text{true}}_{\mathbf{Y}_t}\right)^{-1}(\alpha)$, where $F^{\text{true}}_{\mathbf{Y}_t}(x) := \mathbb{P}^{\text{true}}_{\mathbf{Y}_t}(x \leq \mathbf{Y}_t)$ denotes the CDF of $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$, which is unknown. As a direct (non-parametric) estimator for $q^{\text{t.s.}}_\alpha$, we consider the α-empirical quantile of the sample $Y_1,\ldots,Y_n$, given by $\hat{q}^{\text{t.s.}}_{\alpha,\text{NP}} := Y_{n(i)}$ for $\alpha \in \left(\frac{i-1}{n}, \frac{i}{n}\right]$. Naturally, as a simulation-based (non-parametric) estimator for $q^{\text{t.s.}}_\alpha$, we propose the α-empirical quantile of the sample $\hat{Y}_1,\ldots,\hat{Y}_m$. Unlike in previous examples, deriving theoretical results for $\hat{q}^{\text{t.s.}}_{\alpha,\text{NP}}$ and $\hat{Q}^{\text{t.s.}}_{\alpha,\text{NP}}$ is not straightforward, so we focus on simulations to evaluate their performance.

Since the true value of $q^{\text{t.s.}}_\alpha$ is unknown, we use a pseudo-true value to assess the performance of the estimators. This pseudo-true value is derived by simulating a very large number of paths of length T from $\mathbb{P}^{\text{true}}_{\mathbf{X}}$ and aggregating each path individually to form a sample. That is, a modification of the simulation-based method was used, in which the estimated distribution is the true distribution itself, ensuring that the large sample follows the true marginal distribution. We generated $m = 1.000.000$ paths for $T = 250$ and $m = 2.000.000$ paths for $T = 70$.

Figure 4.8 presents the ratios $\frac{MAE_{100}\left(\hat{q}^{\text{t.s.}}_{0.99}\right)}{q^{\text{t.s.}}_{0.99}}$ and $\frac{MAE_{100}\left(\hat{Q}^{\text{t.s.}}_{0.99}\right)}{q^{\text{t.s.}}_{0.99}}$. It is evident that, for both values of T, the simulation-based estimator achieves better performance, in terms of MAE, than the direct estimator. Additionally, a comparison between Cases (C6) and (C7) shows that correctly specifying the model for $\hat{\mathbb{P}}_{\mathbf{X},n}$ results in a better simulation-based estimator, which was an expected outcome. It is worth mentioning that the simulation-based estimator in the misspecified case still performs better than the direct estimator. This may be due to various factors, including the ability of the AR(1)-GARCH(1,1) model to approximate the chosen AR(2)-GARCH(1,1) process.

Figure 4.9 complements these findings, showing that, for $T = 250$, the MAE of the simulation-based estimator is between 20% and 40% of the MAE of the direct estimator in Case (C6), while in Case (C7), the MAE of the simulation-based estimator is between 25% and 45% of the MAE of the direct estimator. Similarly, for $T = 70$, the MAE of the simulation-based estimator is between 20% and 45% of the MAE of the direct estimator in Case (C6), while in Case (C7), the MAE of the simulation-based estimator is between 30% and 55% of the MAE of the direct estimator.

Finally, Figure 4.10 shows the performance of the simulation-based estimator with respect to the direct estimator across different values of m. As expected, both for $T = 250, n = 10$ and $T = 70, n = 40$, the performance of the simulation-based estimator increases as m increases, which is shown as a decrease in the ratio $\frac{MAE_{100}\left(\hat{Q}^{\text{t.s.}}_{\alpha,\text{NP}}\right)}{MAE_{100}\left(\hat{q}^{\text{t.s.}}_{\alpha,\text{NP}}\right)}$. However, it is also apparent that both curves tend to get flatter far from zero, which means that this relative increase in the performance is limited by a parameter that is not m. This parameter is most likely the sample size n, which limits the accuracy of the estimation of $\mathbb{P}^{\text{true}}_{\mathbf{X}_t}$.

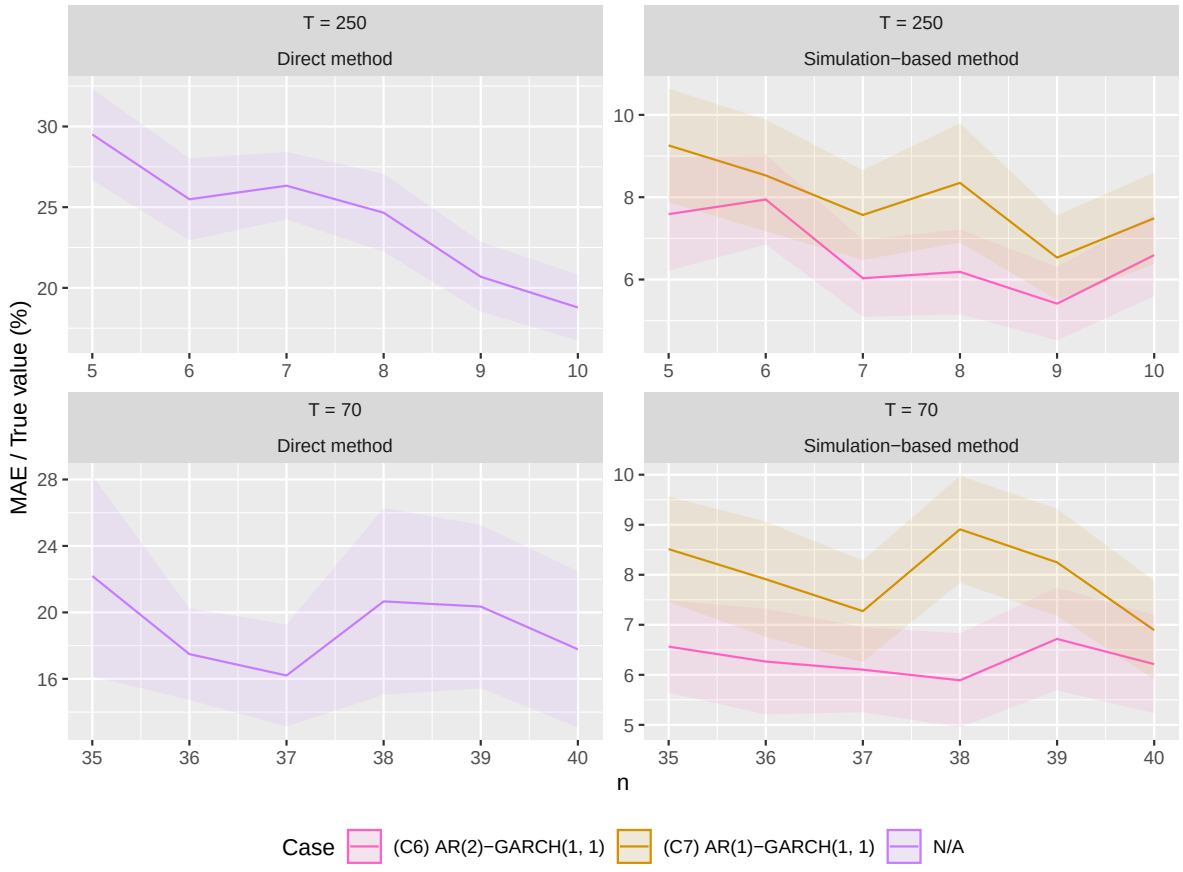


Figure 4.8: Ratios $\frac{MAE_{100}\left(\hat{q}^{\text{t.s.}}_{0.99,\text{NP}}\right)}{q^{\text{t.s}}_{0.99}}$ (left column) and $\frac{MAE_{100}\left(\hat{Q}^{\text{t.s.}}_{0.99,\text{NP}}\right)}{q^{\text{t.s.}}_{0.99}}$ (right column) for Cases (C6) and (C7), and for $T = 250$ (top row) and $T = 70$ (bottom row).

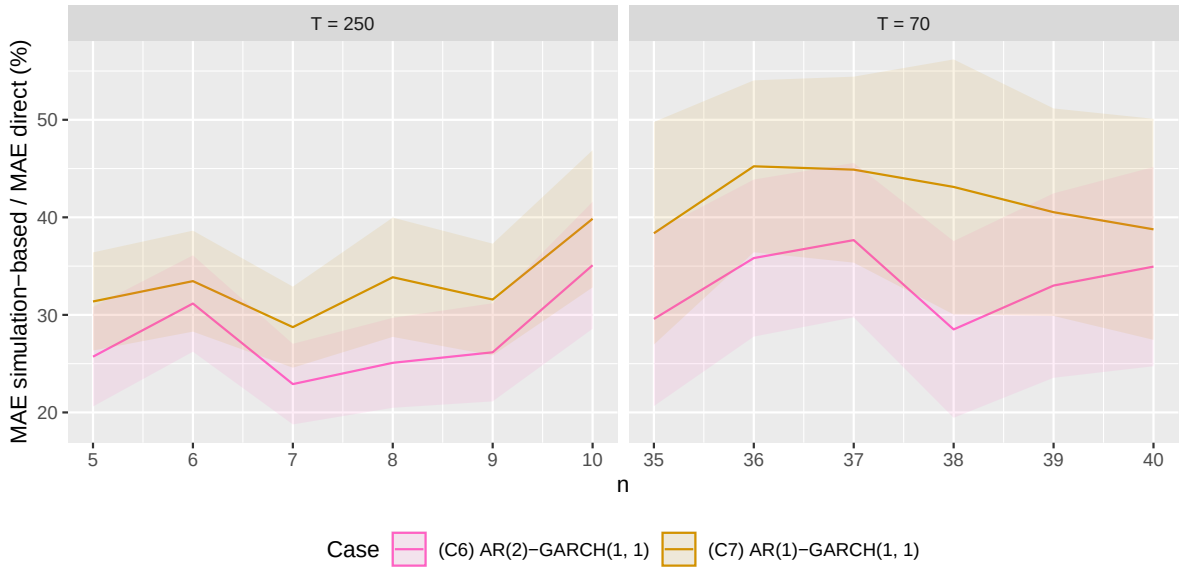


Figure 4.9: Ratio $\frac{MAE_{100}\left(\hat{Q}^{\text{t.s.}}_{\alpha,\text{NP}}\right)}{MAE_{100}\left(\hat{q}^{\text{t.s.}}_{\alpha,\text{NP}}\right)}$ for Cases (C6) and (C7), and for $T = 250$ (left) and $T = 70$ (right).

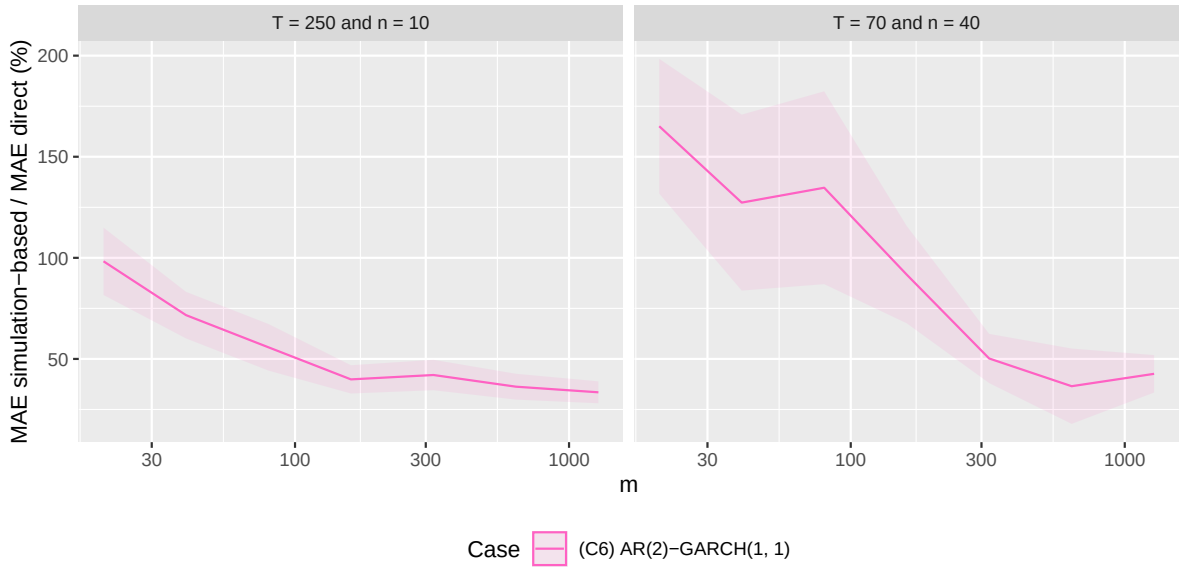


Figure 4.10: Ratio $\frac{MAE_{100}\left(\hat{Q}^{\text{t.s.}}_{\alpha,\text{NP}}\right)}{MAE_{100}\left(\hat{q}^{\text{t.s.}}_{\alpha,\text{NP}}\right)}$ across different values of m for Case (C6), for different values of m, and for $T = 250, n = 10$ and $T = 70, n = 40$.

4.5. Correlation

Let us consider cases (C4) and (C5) within the bivariate normal setting. The objective of this section is to compare the performance of direct and simulation-based estimators for the correlation ρ of the low-frequency marginal distribution $\mathbb{P}^{\text{true}}_{\mathbf{Y}_t}$.

As a direct estimator for ρ, we consider the sample correlation of $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, given by

$$\hat{\rho}_n := \frac{\sum_{i=1}^n \left(Y_{i,1} - \hat{\lambda}_{n,1}\right)\left(Y_{i,2} - \hat{\lambda}_{n,2}\right)}{\sqrt{\sum_{i=1}^n \left(Y_{i,1} - \hat{\lambda}_{n,1}\right)^2}\sqrt{\sum_{i=1}^n \left(Y_{i,2} - \hat{\lambda}_{n,2}\right)^2}},$$

where $\hat{\boldsymbol{\lambda}}_n := \left(\hat{\lambda}_{n,1}, \hat{\lambda}_{n,2}\right)$ is the sample mean of $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$.

Proposition 11. *The error of the direct estimator $\hat{\rho}_n$ has expectation*

$$\frac{\rho\left(\rho^2 - 1\right)}{2n} + O\left(\frac{1}{n^2}\right)$$

and variance

$$\frac{\left(1-\rho^2\right)^2}{n} + O\left(\frac{1}{n^2}\right).$$

Moreover, the scaled error satisfies $\sqrt{n}\left(\hat{\rho}_n - \rho\right) \rightsquigarrow N\left(0, \left(1-\rho^2\right)^2\right)$.

Proof. Hotelling [11] derived the first six moments of the coefficient of correlation of an i.i.d. bivariate normal sample around its true value, as well as the first six moments around its expectation. The first moment around the true value is

$$\begin{aligned}\mathbb{E}\left[\hat{\rho}_n - \rho\right] &= \left(1-\rho^2\right)\left(-\frac{\rho}{2n} + \frac{\rho - 9\rho^3}{8n^2} + \frac{\rho + 42\rho^3 - 75\rho^5}{16n^3} + \ldots\right) \\ &= \frac{\rho\left(\rho^2 - 1\right)}{2n} + O\left(\frac{1}{n^2}\right),\end{aligned}$$

which is as claimed. Similarly, the second moment of $\hat{\rho}_n$ around its mean $\mathbb{E}\left[\hat{\rho}_n\right]$ is

$$\begin{aligned}Var\left[\hat{\rho}_n\right] &= \left(1-\rho^2\right)^2\left(\frac{1}{n} + \frac{11\rho^2}{2n^2} + \frac{-24\rho^2 + 75\rho^4}{2n^3} + \ldots\right) \\ &= \frac{\left(1-\rho^2\right)^2}{n} + O\left(\frac{1}{n^2}\right),\end{aligned}$$

giving us the variance of the error $Var\left[\hat{\rho}_n - \rho\right] = Var\left[\hat{\rho}_n\right]$.

Finally, since the true distribution is bivariate normal, by Example 3.6 of van der Vaart [22] we have

$$\sqrt{n}\left(\hat{\rho}_n - \rho\right) \rightsquigarrow N\left(0, \left(1-\rho^2\right)^2\right).$$

□

Remark 25. An interesting observation is that the moments of the sample correlation for i.i.d. bivariate

normal distributions depend exclusively on the true correlation, irrespective of the other parameters of the distribution.

Now, let us recall that the marginals of $\mathbf{X}$ have the same correlation as the marginals of $\mathbf{Y}$. Therefore, we propose using $\hat{r}_n$ as an estimator of ρ.

Proposition 12. *The error of the estimator $\hat{r}_n$ has expectation*

$$\frac{\rho\left(\rho^2-1\right)}{2nT}+O\left(\frac{1}{n^2}\right)$$

and variance

$$\frac{\left(1-\rho^2\right)^2}{nT}+O\left(\frac{1}{n^2}\right).$$

Moreover, the scaled error satisfies $\sqrt{nT}\left(\hat{r}_n-\rho\right)\rightsquigarrow N\left(0,\left(1-\rho^2\right)^2\right)$ as $n\to\infty$.

Proof. Analogous to Proposition 11 considering a sample of size nT and the same true value ρ. □

Finally, for completeness, we propose a simulation-based estimator for ρ, given by the sample correlation of $\hat{\mathbf{Y}}_1,\ldots,\hat{\mathbf{Y}}_m$, which reads

$$\hat{\rho}_{n,m}:=\frac{\sum_{i=1}^m\left(\hat{Y}_{i,1}-\hat{\lambda}_{n,m,1}\right)\left(\hat{Y}_{i,2}-\hat{\lambda}_{n,m,2}\right)}{\sqrt{\sum_{i=1}^m\left(\hat{Y}_{i,1}-\hat{\lambda}_{n,m,1}\right)^2}\sqrt{\sum_{i=1}^m\left(\hat{Y}_{i,2}-\hat{\lambda}_{n,m,2}\right)^2}},$$

where $\hat{\boldsymbol{\lambda}}_{n,m}=\left(\hat{\lambda}_{n,m,1},\hat{\lambda}_{n,m,2}\right)$ is the sample mean of $\hat{\mathbf{Y}}_1,\ldots,\hat{\mathbf{Y}}_m$.

Figure 4.11 presents the ratios $\frac{MAE_{1.000}(\hat{\rho}_n)}{\rho}$, $\frac{MAE_{1.000}(\hat{\rho}_{n,m})}{\rho}$ and $\frac{MAE_{1.000}(\hat{r}_n)}{\rho}$. It is clear that the simulation-based method achieves significantly better results than the direct method for Cases (C4) and (C5), and for both values of T. The difference between these cases does not seem substantial, which seems to be due to the fact that the variance of the error is dependent only on the true correlation (which is unknown in both cases), and not in the rest of the parameters. From the same figure, it can be seen that $\hat{r}_n$ presents a lower MAE than the simulation-based estimator. This was expected, as $\hat{\rho}_{n,m}$ is constructed from a distribution that uses $\hat{r}_n$ and is therefore biased towards it. Essentially, $\hat{\rho}_{n,m}$ approximates $\hat{r}_n$, and $\hat{r}_n$ approximates ρ. Figure 4.12 complements these findings, illustrating that, for $T=250$, the MAEs of $\hat{\rho}_{n,m}$ and $\hat{r}_n$ fall between 5% and 7.5% of the MAE of $\hat{\rho}_n$. On the other hand, for $T=70$, the MAE of $\hat{r}_n$ is around $10\%-13$ of the MAE of $\hat{\rho}_n$, while the MAE of $\hat{\rho}_{n,m}$ accounts for $12.5\%-15\%$ of the MAE of $\hat{\rho}_n$.

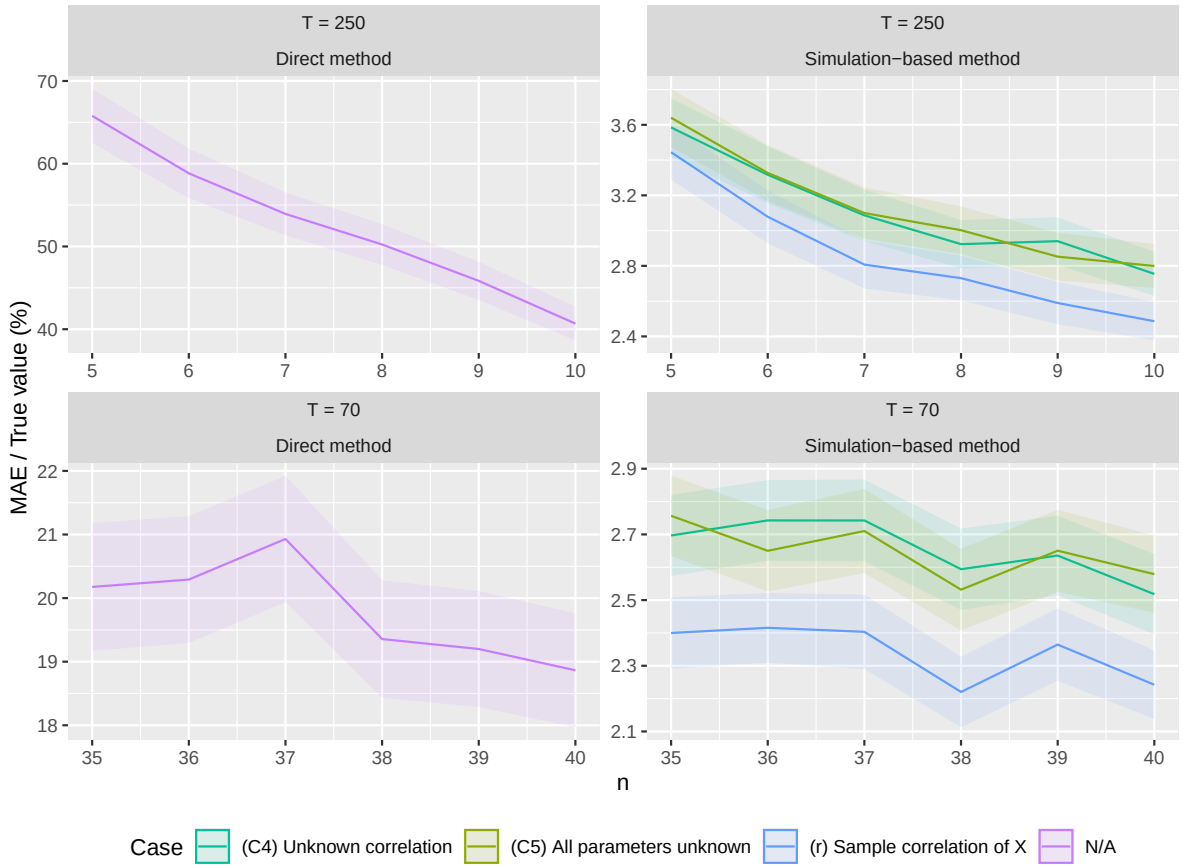


Figure 4.11: Ratios $\frac{MAE_{1.000}(\hat{\rho}_n)}{\rho}$ (left column) and $\frac{MAE_{1.000}(\hat{\rho}_{n,m})}{\rho}$ (right column) for Cases (C4) and (C5), and for $T = 250$ (top row) and $T = 70$ (bottom row). The right column also includes the ratio $\frac{MAE_{1.000}(\hat{r}_n)}{\rho}$.

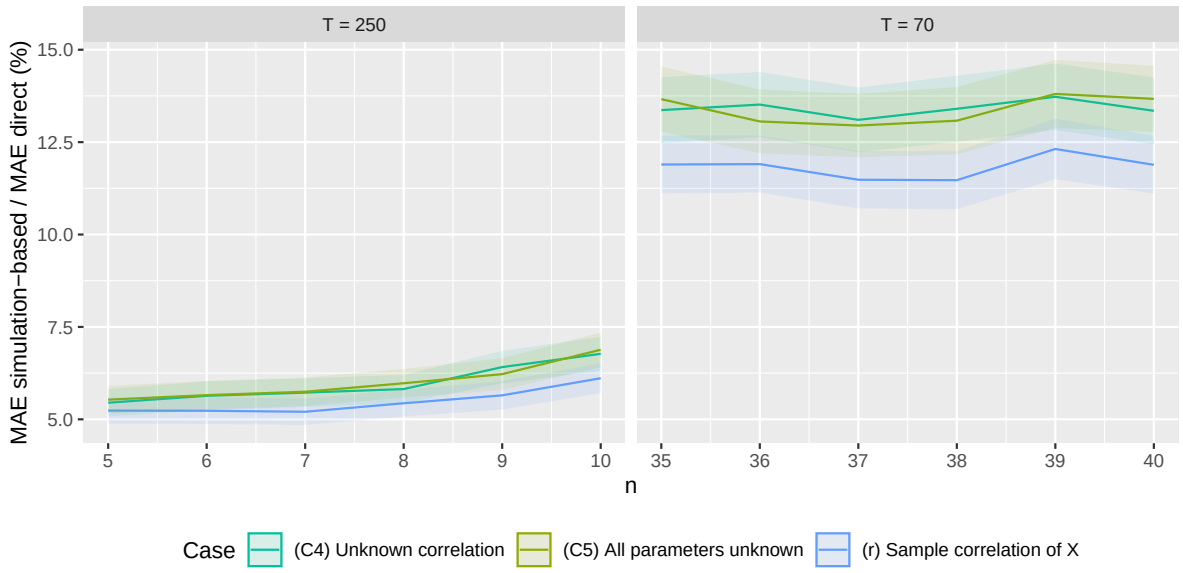


Figure 4.12: Ratio $\frac{MAE_{1.000}(\hat{\rho}_{n,m})}{MAE_{1.000}(\hat{\rho}_n)}$ for Cases (C4) and (C5), and ratio $\frac{MAE_{1.000}(\hat{r}_n)}{MAE_{1.000}(\hat{\rho}_n)}$, for $T = 250$ (left) and $T = 70$ (right).

4.6. Copula degrees-of-freedom

Let us consider the bivariate time series setting. The purpose of this section is to compare the performance of direct and simulation-based estimators for the degrees-of-freedom f of a Student's-t copula $C^{\mathrm{t}}_{\rho,f}$ that models the dependency between the components of $\mathbf{Y}_t$.

The true dependency among the components of $\mathbf{Y}_t$ remains unknown, potentially leading to model misspecification. Nevertheless, our objective is to identify the Student's t-copula that best models this dependency. Since the optimal approximation is also undetermined, we use a pseudo-true value for the degrees-of-freedom to evaluate the accuracy of both direct and simulation-based methods. This pseudo-true value is derived as described in Section 4.4.2: by simulating a large number of paths of length T from the true distribution and fitting a copula to the aggregated values.

From the sample $\mathbf{Y}_1,\dots,\mathbf{Y}_n$, and for each component $j \in \{1,2\}$, we calculate its pseudo-observations $y_{1,j}$, $\dots, y_{n,j}$. Recall that these pseudo-observations are defined as $y_{i,j} = \frac{R_{i,j}}{n+1}$, where $R_{i,j}$ is the rank of $Y_{i,j}$ within the univariate path $Y_{1,j},\dots,Y_{n,j}$. We define a direct estimator for f, which we denote $\hat{f}_n$, as the MLE of f based on the pseudo-observations $(y_{1,1},y_{1,2}),\dots,(y_{n,1},y_{n,2})$. Analogously, from the sample $\hat{\mathbf{Y}}_1,\dots,\hat{\mathbf{Y}}_m$, and for each coordinate $j \in \{1,2\}$ of this sample, we calculate the pseudo-observations $(\hat{y}_{1,1},\hat{y}_{1,2}),\dots,(\hat{y}_{m,1},\hat{y}_{m,2})$, and define our simulation-based estimator for f, denoted by $\hat{f}_{n,m}$, as the MLE of f based on these pseudo-observations.

Figure 4.13 displays the ratios $\frac{MAE_{100}(\hat{f}_n)}{f}$ and $\frac{MAE_{100}(\hat{f}_{n,m})}{f}$. The simulation results show that, for both values of T, the MAE of the direct method consistently equals or exceeds the pseudo-true value, indicating a significant deviation from its pseudo-true value. Contrarily, the MAE of the simulation-based method does not exceed 60% of the pseudo-true value, for both values of T. Although the error may still be considerable, this represents a substantial improvement over the direct method. Figure 4.14 strengthen these findings showing that, for $T = 250$, the MAE of the simulation-based estimator is less than 60% of the MAE of the direct estimator, while for $T = 70$, the MAE of the simulation-based estimator is less than 45% of the MAE of the direct estimator.

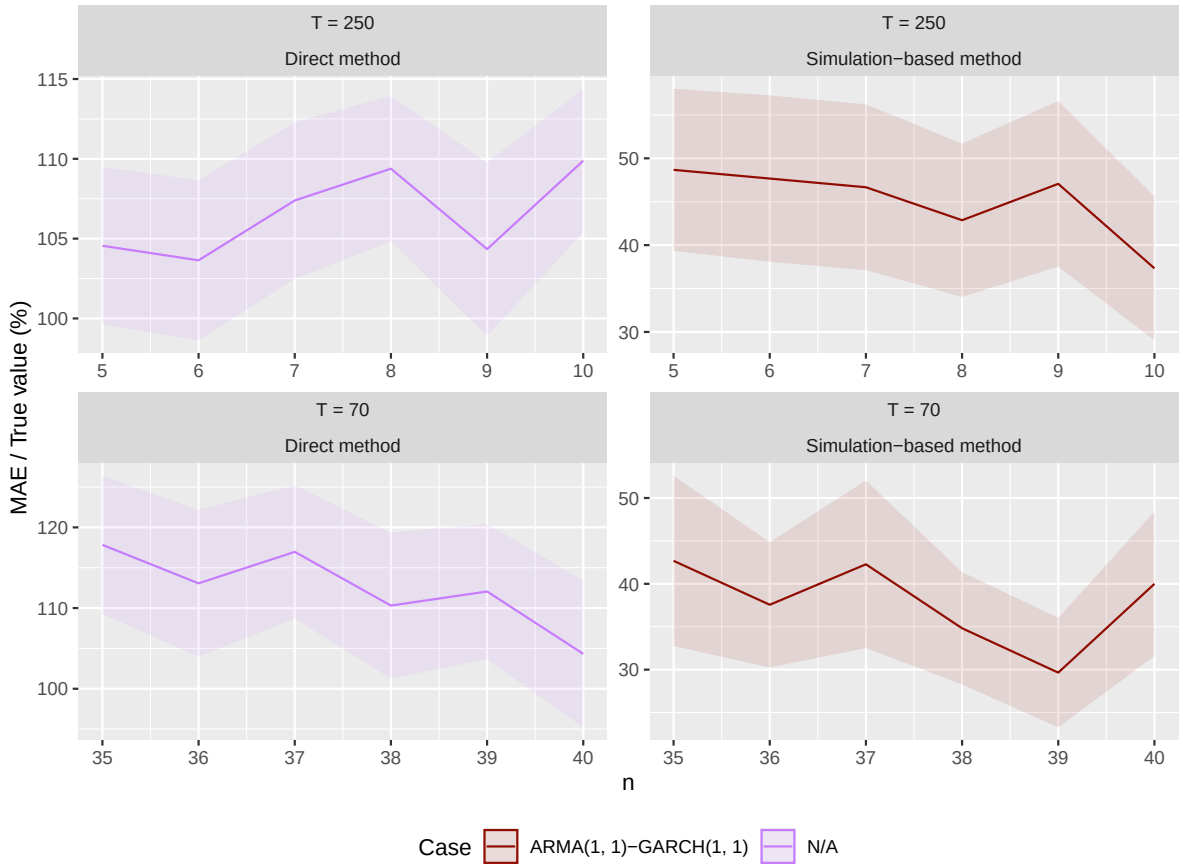


Figure 4.13: Ratios $\frac{MAE_{100}(\hat{f}_n)}{f}$ (left column) and $\frac{MAE_{100}(\hat{f}_{n,m})}{f}$ (right column) for $T = 250$ (top row) and $T = 70$ (bottom row).

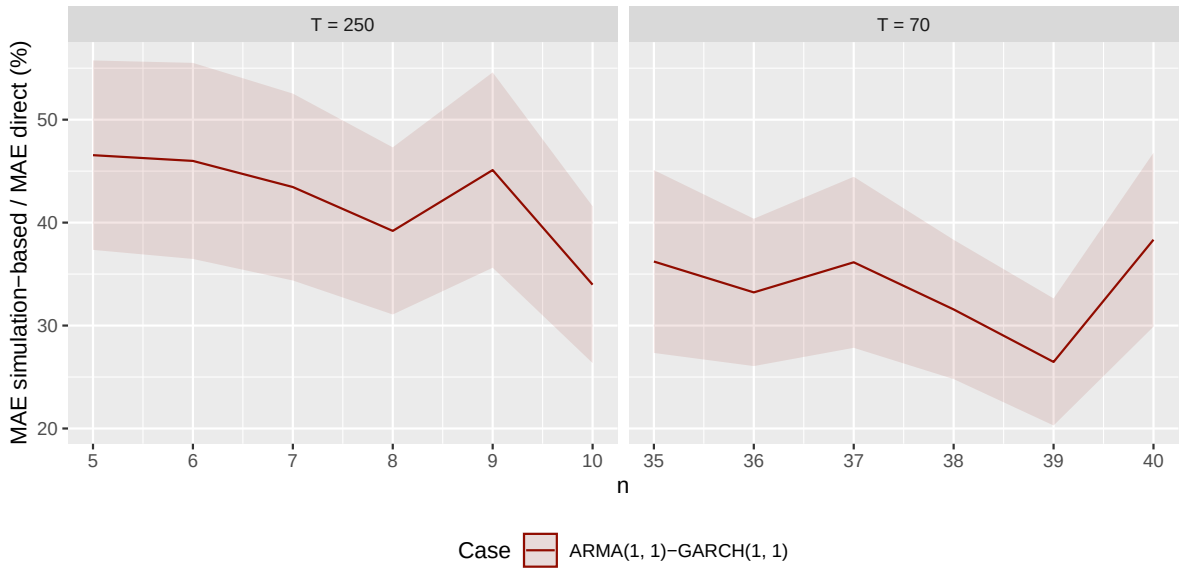


Figure 4.14: Ratio $\frac{MAE_{100}(\hat{f}_{n,m})}{MAE_{100}(\hat{f}_n)}$ for $T = 250$ (left) and $T = 70$ (right).

5

Conclusion

This thesis introduced a novel method for estimating parameters of the distribution of low-frequency stochastic processes. The proposed simulation-based methodology offers a robust and accurate solution to the challenges posed by limited non-overlapping observations of such processes. By leveraging higher-frequency data to simulate low-frequency observations, this method provides more stable and accurate estimators than the traditional direct method across various scenarios and objectives.

The proposed methodology was analyzed through two main approaches:

- Theoretical results were derived for three families of estimators based on estimations of higher-frequency distributions. Given certain conditions, the strong consistency, weak convergence, and asymptotic normality of estimators constructed through continuous and Hadamard-differentiable maps were demonstrated. Additionally, the consistency and rate of convergence of the estimators were established when classified as Minimum Distance Estimators.
- An extensive experimental framework verified that the proposed simulation-based method in most cases significantly outperformed the direct method. Through various parameter estimations in univariate and bivariate settings, the simulation-based method yielded much lower Mean Absolute Errors and more reliable parameter estimates, particularly when the available low-frequency data was scarce.

As for the future research, one may explore the extension of the proposed method to stochastic processes characterized by more complex dependencies and non-linear models, which may present greater challenges in approximation and thus complicate their correct specification when modeling higher-frequency processes. Additionally, the application of the framework to high-dimensional time series could be considered, introducing further complexities in terms of approximation. An extensive sensitivity analysis could be conducted by varying the family of the underlying distribution, its true parameters, and the models fitted to the higher-frequency process. Such analysis would help to understand the robustness of the proposed framework under different modeling conditions and parameter settings.

The simulation-based method could also be tested for estimating parameters of the joint distributions of low-frequency processes. This approach would enable the estimation of relationships between different time points of the low-frequency process, such as their auto-correlation. Such an analysis is particularly valuable for predicting future observations based on current or past values. As previously specified, this could be achieved by simulating longer independent paths from the estimated higher-frequency distribution. Further extensions could involve adapting the method to low-frequency processes that are derived not from the sum of higher-frequency processes, but from other operations, such as taking the maximum within an interval. This adjustment would necessitate modifications to the operators introduced in the theoretical part of this thesis. Finally, the simulation-based method could be tested using real-world data. In such cases, the underlying true distribution of the low-frequency process would be unknown. However, as observed in this investigation, the simulation-based method has shown robustness under moderate model misspecification, suggesting that satisfactory results can be anticipated.

Bibliography

[1] Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Vol. 55. US Government printing office, 1948.

[2] Basel Committee on Banking Supervision. *Minimum capital requirements for market risk*. Accessed: 2024-07-22. Jan. 2019. URL: https://www.bis.org/bcbs/publ/d457.htm.

[3] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[4] Peter J. Brockwell and Richard A. Davis. *Time series: theory and methods*. Springer science & business media, 1991.

[5] Djalil Chafaï and Didier Concordet. "On the strong consistency of asymptotic M-estimators". In: *Journal of statistical planning and inference* 137.9 (2007), pp. 2774–2783.

[6] Alexis Derumigny, Lucas Girard, and Yannick Guyonvarch. "On the construction of confidence intervals for ratios of expectations". In: *arXiv preprint arXiv:1904.07111* (2019).

[7] Constantine A. Drossos and Andreas N. Philippou. "A note on minimum distance estimates". In: *Annals of the Institute of Statistical Mathematics* 32 (1980), pp. 121–123.

[8] Richard M. Dudley. *Real analysis and probability*. Chapman and Hall/CRC, 2018.

[9] Fabrizio Durante and Carlo Sempi. *Principles of copula theory*. Vol. 474. CRC press Boca Raton, FL, 2016.

[10] I.M. Ghani and H.A. Rahim. "Modeling and Forecasting of Volatility using ARMA-GARCH: Case Study on Malaysia Natural Rubber Prices". In: *IOP Conference Series: Materials Science and Engineering*. Vol. 548. 1. IOP Publishing. 2019, p. 012023.

[11] Harold Hotelling. "New light on the correlation coefficient and its transforms". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 15.2 (1953), pp. 193–232.

[12] Eric Jondeau and Michael Rockinger. "The copula-garch model of conditional dependencies: An international stock market application". In: *Journal of international money and finance* 25.5 (2006), pp. 827–853.

[13] Helmut Lütkepohl. "Forecasting aggregated time series variables: A survey". In: *OECD Journal: Journal of Business Cycle Measurement and Analysis* 2010.2 (2011), pp. 1–26.

[14] William C. Parr and William R. Schucany. "Minimum distance and robust estimation". In: *Journal of the American Statistical Association* 75.371 (1980), pp. 616–624.

[15] Benjamin Poignard and Jean-David Fermanian. “The finite sample properties of sparse M-estimators with pseudo-observations”. In: *Annals of the Institute of Statistical Mathematics* 74.1 (2022), pp. 1–31.

[16] Feng Qi and Bai-Ni Guo. “Wendel’s and Gautschi’s inequalities: Refinements, extensions, and a class of logarithmically completely monotonic functions”. In: *Applied Mathematics and Computation* 205.1 (2008), pp. 281–290.

[17] Esther Ruiz and María Rosa Nieto. “Direct versus iterated multiperiod Value-at-Risk forecasts”. In: *Journal of Economic Surveys* 37.3 (2023), pp. 915–949.

[18] Robert J. Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, 2009.

[19] Robert H. Shumway and David S. Stoffer. *Time series analysis and its applications*. Vol. 3. Springer, 2000.

[20] Andrea Silvestrini and David Veredas. “Temporal aggregation of univariate and multivariate time series models: a survey”. In: *Journal of Economic Surveys* 22.3 (2008), pp. 458–497.

[21] Paulo Teles and Paulo S.A. Sousa. “The effect of temporal aggregation on the estimation accuracy of ARMA models”. In: *Communications in Statistics-Simulation and Computation* 47.10 (2018), pp. 2865–2885.

[22] A.W. Van der Vaart. *Asymptotic statistics*. Vol. 3. Cambridge university press, 2000.

[23] A.W. Van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media, 2013.

[24] Qingsong Wen et al. “Time series data augmentation for deep learning: A survey”. In: *arXiv preprint arXiv:2002.12478* (2020).

[25] Jacob Wolfowitz. “The minimum distance method”. In: *The Annals of Mathematical Statistics* (1957), pp. 75–88.