الجمهوريـة الجزائـريـة الديمقراطيـة الشعبيـة
Peoples Democratic Republic of Algeria
وزارة التعليـم العالـي والبحـث العلمـي
Ministry of Higher Education and Scientific Research

University of Mila

جــامعة ميلـة

Faculty of Mathematics and Computer Sciences
Order N°: …………………
Registration number: …………….

Department: Mathematics
Field: Mathematics
Specialty: Probability and Statistics

Thesis

Submitted for the degree of
Doctorate LMD

Maximum Likelihood Estimation of locally stationary processes

Presented by: Aissaoui Faris

Jury Members:

Abdelouahab Mohammed-Salah	Pr	University of Mila	Chairman
Khedidja Djaballah-Djeddour	Pr	USTHB Algiers	Supervisor
Boulaarouk Yakoub	Pr	University of Mila	Examiner
Boukaf Samira	MCA	University of Mila	Examiner
Bouden Rabah	Pr	University of Mila	Examiner
Kharfouchi Soumia	Pr	University of Constantine 3	Examiner

University year: 2025/2026

Maximum Likelihood estimation of locally stationary processes

Aissaoui Faris

July 30, 2026

Abstract

Time series analysis has attracted growing interest in the literature over recent years, driven by the need to model nonstationary phenomena whose statistical characteristics evolve over time. Among the most relevant frameworks developed to address this challenge are locally stationary processes, which provide a powerful compromise between flexibility and analytical tractability. These processes assume that, although a time series may be globally nonstationary, it can be well approximated locally by a stationary process in the neighborhood of each time point.

The primary objective of this thesis is to study the estimation of parameters for locally stationary processes, with a particular focus on models featuring time-varying volatility. Since the parameters are allowed to change smoothly over time, classical estimation methods designed for stationary settings become inadequate. To overcome this limitation, we adopt a modeling approach based on time-varying parameter models and employ a nonparametric quasi-maximum likelihood estimation framework that relies on local likelihood approximation.

The core idea consists in approximating the locally stationary process by a stationary one in the vicinity of a fixed rescaled time point. This local approximation justifies the construction of kernel-weighted estimators, where observations closer to the time point of interest receive higher weights. Under suitable regularity conditions, the proposed estimator is shown to be consistent and asymptotically normal, allowing for valid statistical inference on time-varying parameters.

Special attention is devoted to time-varying GARCH models, which play a central role in modeling conditional volatility in financial time series. Classical GARCH models are extended by allowing their parameters to vary over time within the locally stationary framework. In addition to the standard Gaussian assumption, alternative distributions are considered to better capture heavy-tailed behavior commonly observed in financial data. In particular, Student's t and Laplace distributions are incorporated into the quasi-likelihood framework.

Theoretical results show that when the assumed distribution is correctly specified, the estimator converges to the true parameter values. In contrast, under likelihood misspecification, the estimator converges to pseudo-true parameters rather than the true ones. Extensive numerical experiments confirm the theoretical findings and demonstrate the robustness of the proposed methodology. Applications to real financial data further validate the approach, showing a strong agreement between estimated and observed volatility.

Keywords: Maximum likelihood, Kernel estimation, Time-varying, Locally stationary.

ملخـــص

حظيت دراسة السلاسل الزمنية باهتمام متزايد في السنوات الأخيرة، نتيجة الحاجة إلى نماذج قادرة على تمثيل الظواهر غير المستقرة التي تتغير خصائصها الإحصائية مع مرور الزمن. ومن بين أهم هذه النماذج برزت العمليات الثابتة محليًا، التي توفر إطارًا يجمع بين مرونة النمذجة غير المستقرة وسهولة التحليل المرتبط بالعمليات المستقرة. إذ تقوم هذه النماذج على افتراض أن السلسلة الزمنية، رغم كونها غير مستقرة على المدى الطويل، يمكن تقريبها محليًا بعملية مستقرة في جوار كل نقطة زمنية.

تهدف هذه الأطروحة إلى دراسة مسألة تقدير معلمات العمليات الثابتة محليًا، مع التركيز بشكل خاص على النماذج ذات التباين المتغير عبر الزمن. ونظرًا لأن المعلمات تُفترض دوالًا زمنية تتغير بسلاسة، فإن الطرق الكلاسيكية للتقدير تصبح غير مناسبة. ولمواجهة هذه الإشكالية، تم اعتماد نماذج ذات معلمات متغيرة عبر الزمن، إلى جانب أسلوب التقدير شبه الأعظمي للاحتمال غير البارامتري المبني على التقريب المحلي لدالة الاحتمال.

تعتمد المنهجية المقترحة على تقريب العملية الثابتة محليًا بعملية مستقرة في جوار نقطة زمنية معاد تحجيمها، مما يسمح ببناء مقدرات موزونة تعتمد على دوال نواة. وقد أظهرت النتائج النظرية، تحت شروط انتظام مناسبة، أن هذه المقدرات متسقة مع المعلمات الحقيقية وتتمتع بخاصية التقارب الطبيعي، مما يتيح إجراء استدلال إحصائي سليم للمعلمات المتغيرة.

كما أولت الأطروحة اهتمامًا خاصًا بنماذج غارش المتغيرة عبر الزمن، نظرًا لأهميتها في نمذجة التباين الشرطي للسلاسل الزمنية المالية. وتم توسيع هذه النماذج للسماح بتغير معلماتها زمنيًا ضمن إطار العمليات الثابتة محليًا. إضافة إلى ذلك، لم يقتصر التحليل على الحالة الغاوسية، بل شمل توزيعات بديلة مثل توزيع ستودنت وتوزيع لابلاس، لما لها من قدرة أفضل على تمثيل الذّيول الثقيلة في البيانات المالية.

وتبيّن النتائج أن المقدر يتقارب نحو المعلمات الحقيقية في حالة التحديد الصحيح للتوزيع الاحتمالي، بينما في حالة سوء التحديد فإنه يؤول إلى معلمات شبه حقيقية. وقد أكدت التجارب العددية والتطبيقات على بيانات حقيقية صحة النتائج النظرية وفعالية المنهجية المقترحة.

<u>الكلمات المفتاحية:</u> أقصى احتمال، تقدير باستخدام النواة، متغير مع الزمن، ثابت محليًا.

Résumé

L'analyse des séries temporelles a suscité un intérêt croissant ces dernières années, en raison de la nécessité de modéliser des phénomènes non stationnaires dont les propriétés statistiques évoluent au cours du temps. Parmi les cadres théoriques les plus pertinents figurent les processus localement stationnaires, qui offrent un compromis efficace entre la flexibilité des modèles non stationnaires et la simplicité analytique des processus stationnaires. Ces modèles reposent sur l'hypothèse qu'une série temporelle, bien que globalement non stationnaire, peut être localement approchée par un processus stationnaire au voisinage de chaque instant.

Cette thèse a pour objectif principal l'étude de l'estimation des paramètres des processus localement stationnaires, avec un intérêt particulier pour les modèles à volatilité variant dans le temps. Le caractère évolutif des paramètres rend les méthodes d'estimation classiques inadaptées. Pour pallier cette difficulté, nous adoptons un cadre de modélisation à paramètres variant dans le temps et utilisons une approche de quasi-maximum de vraisemblance non paramétrique fondée sur une approximation locale de la fonction de vraisemblance.

La méthodologie proposée repose sur l'approximation locale du processus par un processus stationnaire au voisinage d'un instant reparamétré, ce qui permet de construire des estimateurs pondérés à l'aide de fonctions noyau. Sous des conditions de régularité appropriées, ces estimateurs sont établis comme étant consistants et asymptotiquement normaux, autorisant ainsi une inférence statistique fiable pour les paramètres dépendant du temps.

Une attention particulière est accordée aux modèles GARCH à paramètres variant dans le temps, largement utilisés pour modéliser la volatilité conditionnelle des séries financières. Ces modèles sont étendus au cadre des processus localement stationnaires. En outre, l'analyse ne se limite pas à l'hypothèse gaussienne, mais intègre également des lois alternatives telles que les lois de Student et de Laplace, mieux adaptées aux données à queues épaisses.

Les résultats théoriques montrent que l'estimateur converge vers les vrais paramètres en cas de bonne spécification du modèle, tandis qu'en situation de mauvaise spécification, il converge vers des paramètres pseudo-vrais. Les expériences numériques ainsi que les applications à des données réelles confirment la validité des résultats théoriques et soulignent l'efficacité de l'approche proposée.

Mots-clés: Maximum de vraisemblance, Estimation par noyau, Variable dans le temps, Localement stationnaire.

Dedication

To my family, both near and extended,

and to everyone who supported me throughout this academic journey.

Acknowledgements

I would like to express my sincere gratitude to Professor **Khedidja Djaballah-Djeddour** (USTHB Algiers) for her valuable supervision, guidance, and insightful remarks throughout the completion of this work.

My sincere thanks go to the members of the examination committee, headed by Professor **Abdelouahab Mohammed-Salah** (University of Mila) for chairing the examination committee and for his careful evaluation of this thesis.

I would also like to thank the honorable members of the committee, **Dr. Boulaarouk Yakoub** (University of Mila), **Pr. Kharfouchi Soumia** (University Constantine 3), **Dr. Boukaf Samira** (University of Mila), and **Dr. Bouden Rabah** (University of Mila), for accepting to examine this work and for their constructive comments and valuable suggestions.

Finally, I would like to thank everyone who contributed, directly or indirectly, to the completion of this work.

Contents

Figures

Tables

Introduction

The statistical analysis of time series has a long and rich history, deeply rooted in the development of probability theory and motivated by the need to understand and predict complex dynamic phenomena observed over time. Early contributions by Yule [40] and Slutsky [36] in the 1920s, followed by the axiomatic foundations introduced by Kolmogorov [31], established the conceptual basis for modeling time-indexed observations as realizations of stochastic processes. A major turning point occurred with the formalization of stationary processes in the mid-twentieth century, which provided a mathematically elegant and analytically tractable framework in which probabilistic properties remain invariant under time shifts. This assumption enabled the systematic development of inference and prediction methods and quickly became central to classical time series analysis.

The subsequent decades witnessed rapid methodological progress, driven both by theoretical advances and by pressing empirical demands. The seminal work of Box and Jenkins [10] popularized autoregressive and moving average models, offering practitioners a coherent and practical modeling strategy for stationary data. At the same time, maximum likelihood estimation emerged as a fundamental inferential principle, valued for its optimal asymptotic properties and its natural probabilistic interpretation. These developments, complemented by advances in spectral analysis and linear systems theory, led to a mature statistical theory of stationary time series that has been successfully applied across economics, finance, engineering, and the natural sciences.

However, as increasingly rich and long data sets became available, empirical evidence began to challenge the realism of the global stationarity assumption. Many observed time series exhibit gradual evolution or structural change in their statistical behavior, reflecting economic growth, policy shifts, technological innovation, or environmental change. In particular, characteristics such as the mean level, dependence structure, or volatility often vary over time, making classical stationary models insufficient for accurate description and inference. These empirical observations provided strong motivation for extending the traditional framework toward models capable of accommodating time-varying dynamics.

This need for flexibility, combined with the desire to retain a solid theoretical foundation, has been a key driving force behind modern developments in nonstationary time series analysis. Rather than abandoning the stationary paradigm entirely, contemporary approaches seek to generalize it in a controlled and interpretable manner. Within this historical trajectory, the concept of locally stationary processes represents a natural and influential evolution, offering a framework that reconciles mathematical tractability with the complex, evolving behavior observed in real-world time series.

Despite significant progress in the development of locally stationary models, the statistical literature still reveals substantial gaps in the theory and application of quasi-maximum likelihood estimation (QMLE) for these processes. While time-varying autoregressive models of order p (tvAR(p)) and time-varying volatility models such as tvARCH and tvGARCH have been proposed, rigorous likelihood-based inference under general conditions remains underexplored. Most existing results rely on restrictive assumptions, including Gaussian or light-tailed innovations, strong moment conditions, or smooth functional forms of parameter evolution. Such constraints limit the practical applicability of these methods, especially in economic and financial contexts where heavy-tailed or skewed distributions frequently arise.

Moreover, the interaction between local stationarity and complex innovation structures introduces non-trivial challenges for estimator performance. Questions regarding the robustness of QMLE under misspecified error distributions, the impact of slowly varying parameters on asymptotic bias and variance, and the role of kernel smoothing or bandwidth selection in local likelihood construction remain largely unresolved. Additionally, the extension of QMLE theory to accommodate multivariate time-varying models, mixed orders of dependence, or conditional heteroskedasticity has been addressed only in a handful of specialized studies. This underscores the need for a comprehensive framework that generalizes QMLE to a wider class of locally stationary models, bridging the gap between theoretical developments and practical applications in diverse fields. A seminal contribution to addressing the challenges of likelihood-based inference for locally stationary processes was made by Dahlhaus [12, 13]. He introduced a general and systematic framework that extended maximum likelihood estimation to time-varying models through the concept of a local likelihood. The central idea is to approximate the process as stationary over short time intervals, allowing the parameters to evolve smoothly across time. By doing so, the approach maintains the conceptual clarity and desirable statistical properties of classical MLE, including consistency, asymptotic normality, and efficiency, while accommodating nonstationary dynamics that are commonly observed in practice.

Dahlhaus' methodology provides a flexible blueprint that can be adapted to a wide array of models, including autoregressive, moving average, and volatility processes, without being confined to a specific parametric family. Importantly, it establishes a bridge between fully parametric stationary theory and realistic time-varying phenomena, offering a rigorous yet practically implementable path for quasi-maximum likelihood estimation in locally stationary frameworks. His work has since guided numerous extensions and applications, forming the conceptual foundation for contemporary research in both theoretical development and applied time series analysis.

In this context, the present thesis aims to contribute to the statistical analysis of locally stationary time series by developing a unified quasi-maximum likelihood framework for time-varying models under general innovation structures.

The primary focus is on extending likelihood-based inference to time-varying autoregressive and conditional heteroskedastic models, while allowing for heavy-tailed and potentially misspecified error distributions.

The remainder of this thesis is organized as follows.

Chapter 1 reviews the theoretical foundations of stochastic processes.

Chapter 2 provides a review of the fundamental concepts in likelihood theory.

Chapter 3 introduces the class of time-varying models considered in this work and presents the proposed kernel-based quasi-maximum likelihood estimation methodology. and establishes the main asymptotic properties of the estimators, including consistency and asymptotic normality.

Chapter 4 provides simulation studies and empirical applications that illustrate the finite-sample performance and practical relevance of the proposed methods.

Finally, Conclusion with a discussion of the main findings and outlines directions for future research.

Chapter 1

Time Series Modeling

1.1 Introduction

A time series is a sequence of observations on a specific variable, recorded at regular intervals of time, for example in days, years, months or even seconds, with ability of non-regularity of time interval of observations. We focus on the case where the time series has a regular pattern of time, then, the model of time series in time t is a function of observation in previous time (t-1, t-2, ... etc).

In the early development of time series analysis, the primary goal was to model and predict phenomena, especially in economics and finance. The foundations of modern time series analysis were laid in the early 20th century, beginning with Sir Udny Yule [40], who introduced the autoregressive (AR) model to analyze time series data. This work marked a shift towards systematic approaches for studying stochastic processes. Subsequently, H. Wold made a significant contribution by introducing the Wold decomposition theorem [38], which provided a framework for expressing any stationary time series as a combination of deterministic and stochastic components, thus advancing the study of stationary processes.

In the 1960s and 1970s, time series analysis saw remarkable growth, especially with the work of Clive Granger [24] and Robert Engle. Granger introduced the concept of causality in time series, known as 'Granger causality', a method to test for directional influence between variables over time. Engle [16] developed the autoregressive conditional heteroskedasticity (ARCH) model , which allowed for modeling time-varying volatility in financial time series, a breakthrough particularly relevant to economic and financial applications.

Later, Tim Bollerslev [7] expanded on this by developing the generalized ARCH (GARCH) model, which became a standard tool in financial econometrics for capturing volatility clustering in returns. These models, along with advancements in computation, allowed for greater precision in analyzing time series and facilitated the development of various specialized models, contributing to the rich field of time series analysis as it stands today.

In this chapter, we focus on the fundamental properties, general principles, and key definitions that form the basis of time series analysis. This includes a discussion of the main concepts, types of models, and essential characteristics that will be referenced and applied in subsequent chapters.

1.2 Stochastic Processes

1.2.1 Definition

Let $(\Omega, \mathcal{F}, P)$ be a probability space, a stochastic process is a collection of random variables $\{X_t\}_{t\in\mathbb{Z}}$ defined on the probability space, where $\mathbb{Z}$ is the index set.

We can also refer to stochastic processes as random processes, random functions, or processes.

1.2.2 Theory and Models

Stochastic processes can be modeled by a function $X = f(t, w)$ where t represents time and ω is an element of the underlying probability space that captures the randomness. Such a process describes the evolution of random variables over time, capturing the inherent randomness in systems or phenomena. For example, financial time series, such as stock prices, can be modeled as stochastic processes to represent their unpredictable movements over time. A stochastic process $X(t)$ is characterized by its mean function, covariance structure, and distribution, and can take various forms such as Markov chains, ARMA processes, or diffusion models, depending on the specific application and properties of interest. We present some examples of stochastic processes.

1. White noise

Let (ε_t) be a stochastic process. We say that (ε_t) is a white noise process if it satisfies the following conditions:

- **Zero Mean:** The process (ε_t) has a mean of zero:

$$\mathbb{E}[\varepsilon_t] = 0 \text{ for all } t \in \mathbb{Z}.$$

- **Constant Variance:** The variance of ε_t is constant over time:

$$\mathrm{Var}(\varepsilon_t) = \sigma^2 \text{ for all } t \in \mathbb{Z}..$$

- **No Autocorrelation:** The process has no autocorrelation at non-zero lags:

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t+h}) = 0 \text{ for all } h \neq 0.$$

2. Autoregressive Processes

An autoregressive process of order p, usually denoted by AR(p). Let $(X_t)_{t\in\mathbb{Z}}$ be an AR(p) process that satisfies a relation of the form:

$$X_t + \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t, \quad \forall t \in \mathbb{Z}$$

where $\phi_i \in \mathbb{R}$ and ε_t is white noise with variance σ^2.

This relation can also be written as:

$$\Phi(L)X_t = \varepsilon_t$$

where $\Phi(L) = 1 + \phi_1 L + \phi_2 L^2 + \cdots + \phi_p L^p$ is the Autoregressive Polynomial, where L is the Backshift Operator ($L^p X_t = X_{t-p}$). For the process to be stationary, all roots of the corresponding characteristic equation $\Phi(z) = 0$ must lie outside the unit circle (i.e., $|z| > 1$). This condition ensures that the influence of past shocks diminishes over time, allowing the series to revert to its mean.

3. Moving Average Processes

A moving average process of order q, usually denoted by MA(q), is a process $(X_t)_{t\in\mathbb{Z}}$ defined by:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}, \quad \forall t \in \mathbb{Z}$$

where θ_i are real numbers and ε_t is white noise with variance σ^2.

Similarly to autoregressive processes, this relation can be written as:

$$X_t = (1 + \theta_1 L + \ldots + \theta_q L^q)\,\varepsilon_t = \Theta(L)\varepsilon_t$$

Unlike the AR(p), the definition of MA(q) is explicit.

4. ARMA Processes

The autoregressive moving average process denoted ARMA(p,q) processes extend the concepts of autoregressive and moving average models. These models are particularly practical for modeling real time series, as they require fewer parameters compared to simple AR or MA models.

A process X_t has a minimal ARMA(p, q) representation if it satisfies:

$$X_t + \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \quad \forall t \in \mathbb{Z}$$

or, using previous notations:

$$\Phi(L)X_t = \Theta(L)\varepsilon_t, \quad \forall t \in \mathbb{Z}$$

with ε_t is white noise with variance σ^2, $\phi_p \neq 0$ and $\theta_q \neq 0$, Φ and Θ have no common roots, and their roots must have modulus greater than 1.

If X_t is a stationary process with a minimal ARMA(p, q) representation:

$$\Phi(L)X_t = \Theta(L)\varepsilon_t, \quad \forall t \in \mathbb{Z}$$

1.2.3 Stationarity and Ergodicity

Stationarity is crucial in time series analysis, as it naturally replaces the assumption of iid (independent and identically distributed) observations used in traditional statistics (see [27]). It ensures that increasing the sample size results in a proportional increase in the amount of information, forming the foundation for a general asymptotic theory. There are generally two notions of stationarity.

Strict Stationarity: A stochastic process (X_t) is called strictly stationary if, for any integers k and h, the joint distribution of $(X_1, \ldots, X_k)$ is identical to that of $(X_{1+h}, \ldots, X_{k+h})$.

A related, less stringent concept focuses only on the first two moments of X_t. Unlike strict stationarity, this weaker notion assumes that these moments exist.

Second-Order Stationarity: A stochastic process (X_t) is considered second-order stationary if the following conditions hold:

(i) $\mathbb{E}[X_t^2] < \infty, \quad \forall t \in \mathbb{Z}$;

(ii) $\mathbb{E}[X_t] = m, \quad \forall t \in \mathbb{Z}$;

(iii) $\mathrm{Cov}(X_t, X_{t+h}) = \gamma(h), \quad \forall t, h \in \mathbb{Z}$.

The function $\gamma(\cdot)$ (and respectively $\rho(\cdot) := \gamma(\cdot)/\gamma(0)$) is called the autocovariance function (and the autocorrelation function) of (X_t).

The simplest example of a second-order stationary process is white noise. This process is particularly important, as it serves as the basis for constructing more complex stationary processes.

Ergodicity: A stochastic process (X_t) is said to be ergodic if time averages converge to ensemble averages. For a second-order stationary process this means that

$$\frac{1}{n}\sum_{t=1}^{n} X_t \xrightarrow[n\to\infty]{} \mathbb{E}[X_t], \quad \frac{1}{n}\sum_{t=1}^{n} X_t X_{t+h} \xrightarrow[n\to\infty]{} \mathbb{E}[X_t X_{t+h}].$$

Ergodicity ensures that statistical properties estimated from a single long realization are representative of the process itself, which is crucial for inference in time series analysis [25].

1.3 Heteroskedasticity Processes

Heteroskedasticity, the phenomenon where the variance of a time series is not constant over time, is a common feature in financial data and other economic variables. To model such behavior, Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) processes were developed as effective tools for capturing time-varying volatility. Generally, these kinds of processes are modeled as follows:

$$\begin{cases} X_t = \sigma_t Z_t \quad \{Z_t\} \sim \mathrm{iid}\,\mathcal{N}(0,1) \\ \sigma_t \in \mathcal{F}_{t-1}^{(Z)} \end{cases}$$

where $\mathcal{F}_{t-1}^{(Z)}$ denotes the collection $\{Z_s : s \leq t-1\}$ and $\sigma_t \in \mathcal{F}_{t-1}^{(Z)}$ denotes that σ_t is a function of these random variables, and $\{X_t : t \in \mathbb{Z}\}$ is a stationary, mean zero time series.

1.3.1 ARCH Processes

The ARCH model specifically addresses situations where the conditional variance of a time series depends on past values of the series itself, the variance at time t is modeled as a function of previous squared errors, which allows it to reflect changes in volatility based on historical fluctuations. This makes it particularly useful for modeling heteroskedasticity in financial returns, where volatility often varies unpredictably over time. A process X_t has ARCH(p) representation means we can write it as follow

$$\begin{cases} X_t = \sigma_t Z_t \\ \sigma_t^2 = \omega + \sum_{j=1}^{p} \alpha_j X_{t-j}^2, \end{cases}$$

Where $\{Z_t\} \sim \text{iid}\,\mathcal{N}(0,1)$ and $\omega > 0, \quad \alpha_j \geq 0, \quad j = 1, ..., p$

For the ARCH(p) process to be weakly stationary, the unconditional variance of X_t must be finite. This is guaranteed when the sum of the coefficients α_j satisfies the condition $\sum_{j=1}^{p} \alpha_j < 1$.

Here an example for simulated ARCH(1) process with $\omega = 0.04$ and $\alpha_1 = 0.09$

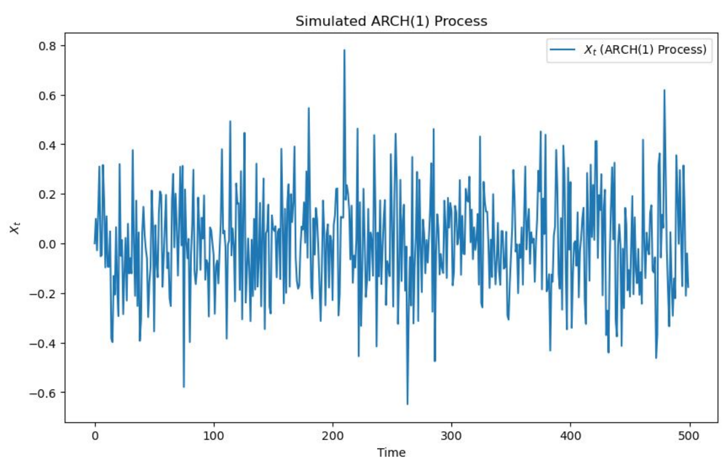


Figure 1.1: Simulated ARCH(1) Process

1.3.2 GARCH Processes

However, the ARCH model can become impractical when capturing longer-term volatility patterns, as it may require many parameters to do so effectively. To remedy this, Tim Bollerslev extended the ARCH model with the introduction of the GARCH model [7]. In a GARCH process, the conditional variance depends not only on past squared errors but also on its own past values, providing a more flexible and parsimonious framework for modeling persistent heteroskedasticity. The GARCH model, with its ability to efficiently handle the persistence of volatility, is better suited for financial time series, where volatility tends to exhibit long memory.

Following the definition of the ARCH(p) model, the GARCH(p, q) model is defined as follows.

$$\begin{cases} X_t = \sigma_t Z_t \\ \sigma_t^2 = \omega + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \end{cases} \tag{1.1}$$

Where $\{Z_t\} \sim \text{iid}\,\mathcal{N}(0,1)$ and $\omega > 0, \quad \alpha_j > 0, \quad \beta_i \geq 0, \quad j = 1, ..., p \quad \text{and} \quad i = 1, ..., q.$

For the GARCH(p, q) process to achieve weak stationarity (second-order stationary), the sum of the coefficients α_j and β_i must meet the condition $\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1$. We can see simulated GARCH(1,1) process with $\omega = 0.09$, $\alpha_1 = 0.1$ and $\beta_1 = 0.7$.

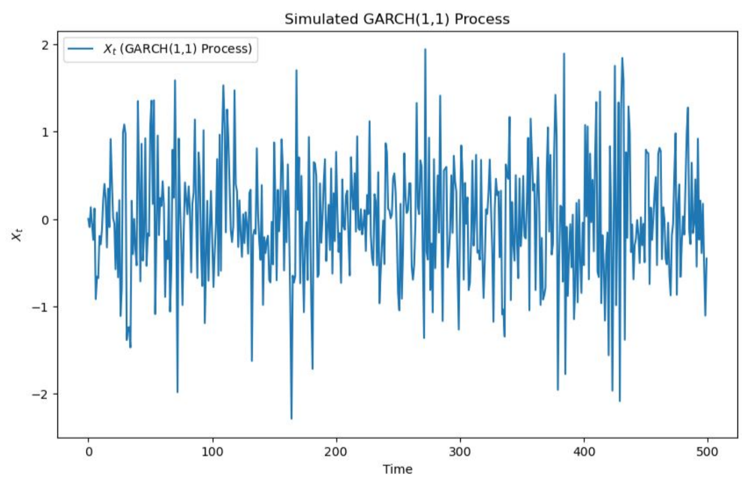


Figure 1.2: Simulated GARCH(1,1) Process

Both ARCH and GARCH processes have become essential tools for analyzing heteroskedastic time series, allowing for improved volatility forecasts, more accurate risk management, and better modeling of financial asset behavior. These models capture the dynamic nature of volatility, making them indispensable for researchers and practitioners dealing with real-world financial and economic data.

1.3.3 Strict Stationarity of GARCH Processes

In practice, many financial time series display characteristics where the parameters fall outside the second-order stationarity region. Even when these series are not square-integrable, they can still satisfy strict stationarity. This motivates the establishment of conditions that guarantee strict stationarity for GARCH processes. Nelson [34] addressed this problem for the GARCH$(1,1)$ model and suggested that the theory of products of random matrices provides a suitable framework for analyzing the general case. Building on this insight, Bougerol and Picard [8] demonstrated that a necessary and sufficient condition for stationarity is that the associated Lyapunov exponent is negative, a fundamental concept in this theoretical framework. before examining the strict stationarity of the GARCH model, we first need to introduce Lyapunov exponent. Let $\|\cdot\|$ a norm on $\mathbb{R}^d$, M is a matrix of $d \times d$.

$$\|M\| = \sup\{\|Mx\|/\|x\|; x \in \mathbb{R}^d \text{and} \quad x \neq 0\}$$

The top Lyapunov exponent associated to a sequence $\{A_t\}_{t\in\mathbb{Z}}$ of i.i.d. random matrices is

$$\begin{aligned}\gamma &= \inf\left\{\mathbb{E}\frac{1}{t+1}\log\|A_0A_{-1}\cdots A_{-t}\| \quad t \in \mathbb{N}\right\}\\ &= \lim_{t\to\infty}\frac{1}{t}\log\|A_0A_{-1}\cdots A_{-t}\|.\end{aligned}$$

For the GARCH(p, q) this vector representation will be very useful $X_t = B_t + A_tX_{t-1}$, we need this notation for the matrix A

$$\begin{aligned}&\tau_t = \left(\beta_1 + \alpha_1 X_t^2, \beta_2, \ldots, \beta_{p-1}\right) \in \mathbb{R}^{p-1}\\ &\xi_t = \left(Z_n^2, 0, \ldots, 0\right) \in \mathbb{R}^{p-1},\\ &\alpha = (\alpha_2, \ldots, \alpha_{q-1}) \in \mathbb{R}^{q-2}\end{aligned}$$

Define the $(p+q-1) \times (p+q-1)$ matrix A_n, written in block form, by

$$A_t = \begin{bmatrix} \tau_t & \beta_p & \alpha & \alpha_q \\ I_{p-1} & 0 & 0 & 0 \\ \xi_t & 0 & 0 & 0 \\ 0 & 0 & I_{q-2} & 0 \end{bmatrix}$$

where I_{p-1} and I_{q-2} are the identity matrices of size $p-1$ and $q-2$, respectively. $\{Z_t, t \in \mathbb{Z}\}$ are random variable i.i.d. distributed with probability law. Therefore, the random matrices $\{A_t, t \in \mathbb{Z}\}$ are i.i.d. follow the same probability law as Z_t has a finite variance, all the coefficients of these matrices are integrable. This implies that $\mathbb{E}\left(\log^+\|A_0\|\right)$ is finite. Thus, the top Lyapunov exponent γ of the sequence $\{A_t, t \in \mathbb{Z}\}$ is well defined.

Theorem 1. *When $\omega > 0$, the GARCH model (defined in 1.1) has a strictly stationary solution if and only if the top Lyapunov exponent γ associate with the matrices $\{A_t\}_{t\in\mathbb{Z}}$ is strictly negative. Moreover, this solution is ergodic. It is the only strictly stationary solution when the Z_t's are given.*

This theorem is proved in P.Bougerol and N. Picard [8].

1.3.4 Second-Order Stationarity of a GARCH Processes

The following result characterizes the second-order stationarity of the GARCH(p, q) model by providing both necessary and sufficient conditions.

Theorem 2. *Assume that a GARCH(p, q) process, as defined in Definition 1.1, exists and is non-anticipative. If the constant parameter satisfies $\omega > 0$ and the process is stationary in the second-order sense, then the model parameters necessarily satisfy*

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1.$$

Conversely, if the above condition holds, the GARCH(p, q) model admits a unique strictly stationary solution. This solution corresponds to a white noise process and is therefore stationary in the second-order sense. No other second-order stationary solutions exist under these assumptions.

The proofs of these results can be found in [21].

1.4 Conclusion

The first chapter provides a comprehensive introduction to the fundamental notions of stochastic processes that are essential for the study of time series models. Particular attention is given to the concepts of stationarity and ergodicity, as they play a central role in ensuring the theoretical validity of statistical inference procedures. The chapter discusses different forms of stationarity and presents the main conditions under which stochastic processes exhibit stable long-run behavior.

In addition to the theoretical framework, the chapter illustrates these concepts through a variety of examples and models commonly encountered in signal processing and financial time series analysis. Special emphasis is placed on models characterized by heteroskedasticity, which are particularly relevant for describing volatility dynamics observed in real-world data. These models serve as a foundation for more advanced developments introduced in the subsequent chapters, especially those related to time-varying volatility and nonstationary behavior.

Chapter 2

Likelihood theory

2.1 Introduction

The theory of likelihood is a foundational concept in statistics and mathematical analysis, crucial for understanding and interpreting data, particularly in the context of parameter estimation and inference. Likelihood theory hinges on the notion of the likelihood function, which measures how well a proposed statistical model explains observed data. Unlike probability, which is concerned with the occurrence of future events given specific parameters, likelihood quantifies the plausibility of parameter values based on observed data.

The method of maximum likelihood, as we know it today, was developed primarily by R. A. Fisher between 1912 and 1922 [18, 19], Fisher also introduced the term 'likelihood', followed by 'maximum likelihood estimates' [19]. formalizing concepts that had appeared only in limited forms previously. Although early versions of parameter estimation were introduced by individuals such as Lambert [33], Daniel Bernoulli [6], and later Gauss in 1809[23], they had no practical impact because the maximum likelihood equation for the error distribution in question was intractable. Fisher's method differed from prior approaches by providing a more rigorous foundation for maximizing the probability of observed data, establishing "likelihood" as the probability model and "maximum likelihood" as the estimation technique.

Consider a parametric model $f(\mathbf{x}; \boldsymbol{\theta})$ representing the probability density of a random variable $\mathbf{X}$ with respect to an appropriate measure. Here, the parameter vector $\boldsymbol{\theta}$ has dimension k, and the observed data $\mathbf{X} = (X_1, \ldots, X_n)$ are typically assumed to be independent and identically distributed.

The likelihood function is defined as a function of $\boldsymbol{\theta}$, proportional to the model density:

$$\mathcal{L}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}; \mathbf{x}) = c\, f(\mathbf{x}; \boldsymbol{\theta}),$$

where c may depend on the observed data $\mathbf{x}$ but not on the parameter $\boldsymbol{\theta}$.

Within this parametric framework, the likelihood function quantifies how plausible different parameter values are, given the fixed observed data. The focus is on $\boldsymbol{\theta}$ as the variable of interest, while $\mathbf{x}$ is treated as fixed. The proportionality constant c ensures that transformations of $\mathbf{X}$ that do not involve $\boldsymbol{\theta}$, such as one-to-one reparameterizations of the data, do not affect inference about the parameter.

2.2 The Likelihood Function

2.2.1 Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample, have a joint density function $f(X_1, X_2, \ldots, X_n \mid \theta)$. Given the observation $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, the the likelihood function is defined as a function of θ defined by:

$$\mathcal{L}(\theta) = \mathcal{L}(\theta \mid x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n \mid \theta) \tag{1}$$

We must note that the likelihood function is not a probability distribution.

2.2.2 Discrete Probability Distribution

Let X be a discrete random variable with a probability mass function p that depends on a parameter θ. Then, the function

$$\mathcal{L}(\theta \mid x) = p_\theta(x) = P_\theta(X = x)$$

is the likelihood function, considered as a function of θ given x of the random variable X. The probability of the measured value x of X for a parameter value θ is denoted by $P(X = x \mid \theta)$ or $P(X = x; \theta)$. The likelihood is the probability that a particular x is observed when the true parameter value is θ_0.

2.2.3 Continuous probability distribution

Let X be a random variable with a continuous probability density function $f_\theta(x)$ depending on a parameter θ. The function

$$\mathcal{L}(\theta \mid x) = f_\theta(x),$$

considered as a function of θ given x, is the likelihood function. Recalling that $\mathcal{L}$ is not a probability density or mass function with respect to θ, since it does not integrate (or sum) to one over the parameter space. It is instead a function of θ obtained by evaluating the probability model at the observed value $X = x$.

2.2.4 Examples

- **Normal distribution:**

  Let $X_1, \ldots, X_n$ be a random sapmle with size n, and Normally distributed $\sim \mathcal{N}(\mu, \sigma)$, so

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$

  The likelihood function is given by:

$$\mathcal{L}(\theta; x_1, \ldots, x_n) = \prod_{j=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_j-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{j=1}^{n}(x_j-\mu)^2}{2\sigma^2}} \tag{2.1}$$

  With $\theta = (\mu, \sigma)'$ is the parameter vector.

- **Bernoulli distribution:**

  Let $X_1, \ldots, X_n$ independent and Bernoulli distributed $\sim \mathcal{B}(p)$ sample, The probability mass function is given by:

$$p(x) = \mathbb{P}\left[X_j = x\right] = \begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0 \end{cases}$$
$$= p^x(1-p)^{1-x} \quad \text{for } x \in \{0, 1\}.$$

  Write the Likelihood Function

$$\mathcal{L}(p; x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} p(x_j)$$

$$= \prod_{j=1}^{n} \left[p^{x_j} (1-p)^{1-x_j} \right]$$

$$\mathcal{L}(p; x_1, x_2, \ldots, x_n) = p^{\sum_{j=1}^{n} x_j} (1-p)^{n - \sum_{j=1}^{n} x_j}$$

2.3 Maximum Likelihood Estimation (MLE)

The principle of maximum likelihood identifies the estimator $\widehat{\theta}$ as the parameter value that maximizes the probability of observing the given data.

2.3.1 Definition

Let $\mathcal{L}$ be a Likelihood function, where

$$\mathcal{L}(\theta|x) = f_\theta(x)$$

The maximum likelihood estimator (MLE), give by

$$\widehat{\theta} = \arg\max_{\theta}(\mathcal{L}(\theta|x))$$

Maximum likelihood estimators offer numerous desirable properties. However, for high dimensional data, the likelihood function can exhibit multiple local maxima, making it challenging to locate the global maximum. In practice, We often resort to maximizing a log-Likelihood function $(\log(\mathcal{L}(\theta|x)))$ to simplify this process. So

$$\widehat{\theta} = \arg\max_{\theta}(\log(\mathcal{L}(\theta|x)))$$

2.3.2 Asymptotic Properties

When discussing Maximum Likelihood Estimator (MLE), it is essential to address the quality of this estimator. In large sample sizes, this quality can be demonstrated by the important properties outlined below.

- **Consistency**

  Suppose that θ_0 is the true parameter, and $\widehat{\theta}_n$ is the maximum likelihood estimator based on n observations, $\widehat{\theta}$ is consistent if

  $$\widehat{\theta}_n \xrightarrow{P} \theta_0 \quad \text{as } n \to \infty.$$

As the sample size increases, the value of the estimated parameter $\widehat{\theta}_n$ approaches the true parameter value θ_0 in probability.

- **Asymptotic normality**

  A central limit theorem holds under certain assumptions, stating that the estimator is asymptotically normal as $n \to \infty$ if

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Sigma).$$

  As the sample size increases, the estimator becomes more concentrated around the true parameter value, and the deviations from the true value, when scaled by $\sqrt{n}$, will be distributed normally with a mean of 0 and a variance of Σ.

2.4 Likelihood ratio test

The likelihood ratio test (LRT) is used to compare the goodness of fit of two nested models, a simpler (null) model and a more complex (alternative) model. This test examines whether additional parameters in the complex model significantly improve the fit.

2.4.1 Likelihood Ratio Statistic

Given: - $\mathcal{L}_0$: the maximum likelihood of the null (restricted) model - $\mathcal{L}_1$: the maximum likelihood of the alternative (unrestricted) model

The likelihood ratio test statistic, Λ, is defined as:

$$\Lambda = -2\log\left(\frac{\mathcal{L}_0}{\mathcal{L}_1}\right) = 2\left(\log\mathcal{L}_1 - \log\mathcal{L}_0\right)$$

2.4.2 Distribution of Λ

Under the null hypothesis, if the sample size is large, Λ approximately follows a chi-square (χ^2) distribution with degrees of freedom equal to the difference in the number of parameters between the alternative and null models.

2.4.3 Hypothesis Testing

1. Null Hypothesis H_0: The simpler model (null model) fits the data as well as the more complex model.

2. Alternative Hypothesis H_1: The more complex model provides a significantly better fit than the null model.

2.4.4 Decision Rule

1. Calculate Λ using the log-likelihoods from both models.

2. Determine the critical value from the χ^2 distribution for a chosen significance level α and degrees of freedom.

3. If Λ exceeds the critical value, reject H_0; otherwise, do not reject H_0.

2.5 Practical Applications

Example 1: $\mathcal{N}(\mu, 1)$

Let $X_1, \ldots, X_n$ be a random sample follows $\mathcal{N}(\mu, 1)$ distribution, we want to estimate the unknown mean μ, so we can put the parameter $\theta = \mu$.

As we mention in the previous section in 2.1, The likelihood function of the Normal distribution with $\sigma = 1$ and unknown μ is given by

$$\mathcal{L}(\theta; x_1, \ldots, x_n) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_j-\mu)^2}{2}} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{\sum_{j=1}^{n}(x_j-\mu)^2}{2}}$$

So

$$\log(\mathcal{L}(\theta; x_1, \ldots, x_n)) = \frac{-n}{2}\log(2\pi) - \frac{1}{2}\sum_{j=1}^{n}(x_j - \mu)^2$$

The estimation of the parameter μ depends on maximizing the likelihood function, achieved by differentiating the log-likelihood since it is analytically more convenient.

$$\frac{\partial}{\partial\mu}\log(\mathcal{L}(\mu; x_1, \ldots, x_n)) = -\frac{1}{2}\sum_{j=1}^{n} 2(x_j - \mu) = n\mu - \sum_{j=1}^{n} x_j$$

When we solve the equation $-\frac{1}{2}\sum_{j=1}^{n} 2(x_j - \mu) = n\mu - \sum_{j=1}^{n} x_j = 0$ we find

$$\widehat{\mu} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

To confirm that $\widehat{\mu}$ maximizes the log-likelihood, we examine the second derivative:

$$\frac{\partial^2}{\partial \mu^2}\log(\mathcal{L}(\mu; x_1, \ldots, x_n)) = -n$$

Since $-n$ is negative, this second derivative indicates that $\widehat{\mu}$ corresponds to a maximum, not a minimum.

Thus, $\widehat{\mu} = \frac{1}{n}\sum_{j=1}^{n} x_j$ maximizes the log-likelihood function.

Example 2: Estimation using the Maximum Likelihood Method for a linear model

A linear model is written as:

$$y = \alpha x + \beta + \varepsilon$$

Thus, the residual element can be written as:

$$\varepsilon_i = y_i - \alpha x_i - \beta \tag{1}$$

The variables ε_i are iid and $\mathcal{N}(0, 1)$, so the density function of the random variable ε is

$$f(\varepsilon) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\varepsilon^2\right)$$

and as it is defined in (1), the density expression is written:

$$= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(y_i - \alpha x_i - \beta)^2\right)$$

According to the definition of the likelihood:

$$\mathcal{L}(x_1, x_2, \ldots, y_1, y_2, \ldots) = \prod_i f(x_i, y_i)$$

Thus:

$$\mathcal{L}(x_1, x_2, \dots, y_1, y_2, \dots) = \frac{1}{\sqrt{2\pi}} \exp\left(\sum_{i=1}^{n} -\frac{1}{2}(y_i - \alpha x_i - \beta)^2\right)$$

To maximize this function, we can use $\log(\mathcal{L})$ since the logarithm function is increasing.

$$\log(\mathcal{L}) = -n\log(\sqrt{2\pi}) - \frac{1}{2}\sum_{i=1}^{n}(y_i - \alpha x_i - \beta)^2$$

To maximize this function, we use:

$$\text{argmax}(\mathcal{L}) = \text{argmax}\left(-n\log(\sqrt{2\pi}) - \frac{1}{2}\sum_{i=1}^{n}(y_i - \alpha x_i - \beta)^2\right)$$

equivalent to:

$$\text{argmax}(\mathcal{L}) = \text{argmax}\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i - \alpha x_i - \beta)^2\right)$$

To find the values of α and β that maximize this function, we take the derivatives of the negative log-likelihood with respect to α and β, and set them to zero.

First Derivative with Respect to α

$$\frac{\partial}{\partial\alpha}\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i - \alpha x_i - \beta)^2\right) = \sum_{i=1}^{n} x_i(y_i - \alpha x_i - \beta) = 0$$

First Derivative with Respect to β

$$\frac{\partial}{\partial\beta}\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i - \alpha x_i - \beta)^2\right) = \sum_{i=1}^{n}(y_i - \alpha x_i - \beta) = 0$$

From the derivative with respect to β, we have:

$$\sum_{i=1}^{n}(y_i - \alpha x_i - \beta) = 0 \quad\Longrightarrow\quad n\beta = \sum_{i=1}^{n} y_i - \alpha\sum_{i=1}^{n} x_i \quad\Longrightarrow\quad \beta = \bar{y} - \alpha\bar{x}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Substituting $\beta = \bar{y} - \alpha\bar{x}$ into the derivative with respect to α, we get:

$$\sum_{i=1}^{n} x_i\left(y_i - \alpha x_i - (\bar{y} - \alpha\bar{x})\right) = 0$$

which simplifies to:

$$\alpha = \frac{\text{COV}(x, y)}{\text{VAR}(x)}$$

where $\text{COV}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ and $\text{VAR}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

This gives us the same function S as the OLS method. Therefore:

$$\hat{\boldsymbol{\theta}} = \begin{cases} \hat{\boldsymbol{\alpha}} = \frac{\text{COV}(x,y)}{\text{VAR}(x)} \\ \hat{\boldsymbol{\beta}} = \bar{y} - \hat{\boldsymbol{\alpha}}\bar{x} \end{cases}$$

To verify that the log-likelihood function reaches a maximum, we calculate the second derivatives of the log-likelihood function. The second derivatives are as follows

Second derivative with respect to α

$$\frac{\partial^2 \log(\mathcal{L})}{\partial \alpha^2} = -\sum_{i=1}^{n} x_i^2$$

Second derivative with respect to β

$$\frac{\partial^2 \log(\mathcal{L})}{\partial \beta^2} = -n$$

Mixed partial derivative

$$\frac{\partial^2 \log(\mathcal{L})}{\partial \alpha \partial \beta} = -\sum_{i=1}^{n} x_i$$

The Hessian matrix H is

$$H = \begin{bmatrix} -\sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & -n \end{bmatrix}$$

To ensure a maximum, H must be negative definite. This is true if $-\sum_{i=1}^{n} x_i^2 < 0$, which is always satisfied as the sum of squared terms is positive. And the determinant of H is positivedet$(H) = (-\sum_{i=1}^{n} x_i^2)(-n) - (-\sum_{i=1}^{n} x_i)^2 > 0$

2.6 Conclusion

This chapter has presented the fundamental principles of likelihood theory, which constitute a cornerstone of modern statistical inference. Beginning with the formal definition of the likelihood function, we clarified its role as a tool for measuring the plausibility of parameter values given observed data, emphasizing its conceptual distinction from probability. Both discrete and continuous frameworks were examined, supported by illustrative examples that highlight the construction and interpretation of likelihood functions in common statistical models.

The chapter then focused on the method of Maximum Likelihood Estimation (MLE), outlining its definition, computational aspects, and key asymptotic properties, namely consistency and asymptotic normality. These properties justify the widespread use of QMLE in large-sample settings and provide a rigorous theoretical basis for statistical inference. Furthermore, the likelihood ratio test was introduced as a powerful hypothesis testing procedure for comparing nested models, with attention given to its asymptotic chi-square distribution and decision rules.

Finally, practical applications were explored through detailed examples, including the estimation of parameters in the normal distribution and linear regression models. These examples demonstrate how likelihood-based methods lead to estimators that coincide with classical results, such as the sample mean and ordinary least squares estimators, while offering a unified inferential framework. Overall, this chapter establishes the theoretical and practical foundations necessary for the likelihood-based methodologies employed in the subsequent chapters, particularly in the context of more complex and time-varying models.

Chapter 3

Time-Varying Models

3.1 Introduction

Time-varying processes are characterized by parameters that evolve smoothly over time, making classical stationary time series models inadequate for their analysis. To address this issue, a flexible framework is required that captures gradual structural changes while preserving tractable probabilistic properties. Following the concept of local stationarity introduced by Dahlhaus [13], a time-varying process can be represented, for example, as

$$X_{t,T} = \sum_{j=-\infty}^{\infty} \psi_{t,T}(j)\, \varepsilon_{t-j},$$

where $X_{t,T}$ denotes the observed process at time t in a sample of size T, $\{\varepsilon_t\}$ is a sequence of independent and identically distributed innovations, and $\psi_{t,T}(j)$ are time-dependent coefficients describing the dynamic structure of the process.

The key idea underlying this formulation is that, although the global behavior of $\{X_{t,T}\}$ is nonstationary, the process can be well approximated locally by a stationary one. More precisely, for rescaled time points $u = t/T$, the coefficients $\psi_{t,T}(j)$ vary smoothly in u, allowing the process to be approximated in a neighborhood of any fixed $u_0 \in (0, 1)$ by a stationary process with frozen parameters. This local approximation property forms the basis for both theoretical analysis and statistical inference.

Within this framework, inference is naturally conducted using local estimation techniques that exploit the approximate stationarity around a fixed time point. In particular, local likelihood and quasi-maximum likelihood methods provide a principled approach for estimating time-varying parameters and studying their asymptotic behavior. This chapter presents the mathematical foundations of locally stationary time-varying models, introduces the associated local likelihood methodology, and establishes the main asymptotic results for the resulting estimators.

3.2 Locally Stationary Processes

3.2.1 Time-Varying Processes

We consider a stochastic process $\{X_t\}$ defined as a stationary solution of

$$X_t = F_\theta(X_{t-k}; \varepsilon_t), \quad k \in \mathbb{N}^*, \quad t \in \mathbb{N}^*,$$

The parameter θ takes values in a predefined subset $\Theta \subset \mathbb{R}^d$ and is treated as deterministic. F_θ is a known function, and ε_t is an independent and identically distributed sequence.

In this classical model, the vector of parameters θ is constant over time, leading to a strictly stationary process. However, in many real-world applications, the data-generating process exhibits time-dependent dynamics, which motivates the introduction of locally stationary processes. In this framework, we replace the constant parameters in the vector θ with a time-varying parameters, which are bounded functions $\theta(u)$, where $u = \dfrac{t}{T}$ represents a rescaled time index (We use this rescaling technique to construct a robust framework that facilitates the development of a meaningful asymptotic theory.). This leads to the time-varying model

$$X_{t,T} = F_{\theta(\frac{t}{T})}(X_{t-k,T}; \varepsilon_t), \quad k \in \mathbb{N}^*, \quad t \in \mathbb{N}^*. \tag{3.1}$$

Here, T refers to the sample size, and ε_t are i.i.d. random variables. The solution of this equation is not necessarily stationary anymore. However, if we want to estimate the parameter curve, we would like to find a family of stationary approximations of the process $X_{t,T}$ in a neighborhood of a fixed time point $u_0 = \frac{t_0}{T}$, which is denoted as

$$\widetilde{X}_t(u_0) = F_\theta(\widetilde{X}_{t-k}(u_0); \varepsilon_t), \quad k \in \mathbb{N}^*, \quad t \in \mathbb{N}^*. \tag{3.2}$$

This stationary approximation satisfies the following theorem:

Theorem 3. *Suppose that $\{X_{t,T}\}$ is a time-varying process satisfying suitable regularity conditions, let*

$\tilde{X}_t(u_0)$ be defined as in (3.2). Then there exists a stationary, ergodic, and positive process $\{Y_t\}$ independent of u_0 with $\mathbb{E}(Y_t) < \infty$ and a constant C such that:

$$\left|X_{t,T} - \tilde{X}_t(u_0)\right| \leq C\left(\left|\frac{t}{T} - u_0\right| + \frac{1}{T}\right) Y_t \quad a.s. \tag{3.3}$$

As a consequence of (3.3), we have

$$X_{t,T} = \tilde{X}_t(u_0) + O_p\left(\left|\frac{t}{T} - u_0\right| + \frac{1}{T}\right) \tag{3.4}$$

The inequality (3.3) allows us to approximate the locally stationary process $X_{t,N}$ by the process $\tilde{X}_t(u_0)$.

Remark 1. *In volatility models of GARCH type, inference is formulated in terms of second-order dynamics, since the conditional variance depends on squared observations. For this reason, local stationarity is naturally studied for the squared process. In particular, Dahlhaus and Rao [14] establish local approximation results of the form*

$$\left|X_{t,T}^2 - \tilde{X}_t^2(u_0)\right| \leq C\left(\left|\frac{t}{T} - u_0\right| + \frac{1}{T}\right) Y_t,$$

reflecting the intrinsic role of X_t^2 in GARCH-type models. The squared version therefore constitutes a model-driven specialization of Theorem 3, rather than a fundamental requirement of local stationarity.

Proof. By the definition of local stationarity in the sense of Dahlhaus, there exists a measurable function

$$F : [0,1] \times \mathbb{R}^{\mathbb{Z}} \longrightarrow \mathbb{R}$$

such that

$$X_{t,T} = F\left(\frac{t}{T}, \mathcal{F}_t\right), \qquad \tilde{X}_t(u_0) = F(u_0, \mathcal{F}_t),$$

where $\mathcal{F}_t = (\dots, \varepsilon_{t-1}, \varepsilon_t)$ and (ε_t) is an i.i.d. sequence.

Assume that $F(u, \cdot)$ is Lipschitz continuous with respect to u, uniformly in $\mathcal{F}_t$. That is, there exist a constant $C > 0$ and a stationary, ergodic, and positive process $\{Y_t\}$ with $\mathbb{E}(Y_t) < \infty$ such that

$$|F(u, \mathcal{F}_t) - F(v, \mathcal{F}_t)| \leq C|u - v| Y_t, \quad \text{a.s.}$$

Then, for any fixed $u_0 \in (0,1)$, we have

$$\begin{aligned}\left|X_{t,T} - \tilde{X}_t(u_0)\right| &= \left|F\left(\frac{t}{T}, \mathcal{F}_t\right) - F(u_0, \mathcal{F}_t)\right| \\ &\leq C\left|\frac{t}{T} - u_0\right| Y_t \quad \text{a.s.}\end{aligned}$$

Moreover, the discretization induced by the rescaled time index t/T introduces an additional approximation error of order T^{-1}. Hence, there exists a (possibly different) constant $C > 0$ such that

$$\left|X_{t,T} - \tilde{X}_t(u_0)\right| \leq C\left(\left|\frac{t}{T} - u_0\right| + \frac{1}{T}\right) Y_t \quad \text{a.s.},$$

which proves (3.3).

Since $\mathbb{E}(Y_t) < \infty$, inequality (3.3) implies

$$X_{t,T} = \tilde{X}_t(u_0) + O_p\left(\left|\frac{t}{T} - u_0\right| + \frac{1}{T}\right),$$

which establishes (3.4) and completes the proof. □

Figure (3.1) illustrates a time-varying process trajectory alongside its stationary approximation trajectory. The vertical dashed lines indicate interval boundaries around u_0, representing local estimation windows. The stationary approximation provides a smoothed representation of the process within each interval, capturing local characteristics of volatility dynamics.

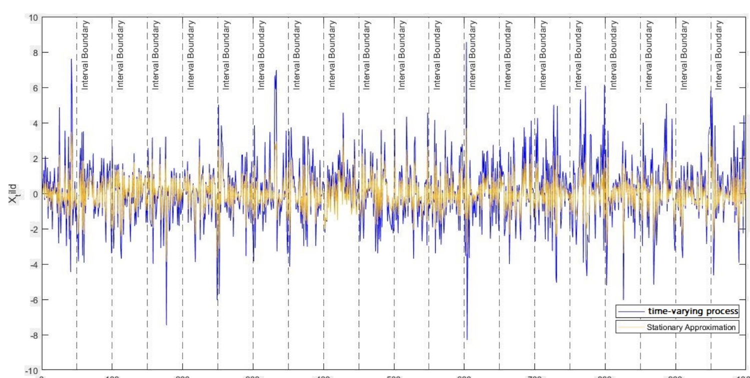


Figure 3.1: Illustration of a time-varying process (blue) with its stationary approximation (yellow).

In the next section, we estimate the parameter vector θ.

First, we introduce θ_0, the true parameter vector. We further assume the following.

Assumption of Local Stationarity (LS(ρ))

There are positifs constants C_θ and $\rho \in (0,1]$, and a continuous function $\theta_0 : [0,1] \to \mathbb{R}^d$, satisfying

$$\|\theta_{t,T} - \theta_0(u)\| \leq C_\theta \left| u - \frac{t}{T} \right|^\rho, \quad \text{for any } T \in \mathbb{N}^* \text{ and } 1 \leq t \leq T.$$

This condition characterizes a Hölder-type approximation of $\theta_{t,T}$ by θ_0. As $\frac{t}{T} \to u$, we have

$$\theta_{t,T} \xrightarrow[T \to +\infty]{} \theta_0(u).$$

Consequently, the behavior of $X_{t,T}$ asymptotically resembles its stationary counterpart.

3.3 Estimation of Time-Varying Parameters

We explore methodologies for estimating time-varying parameters, with a focus on nonparametric approaches. As mentioned in Dahlhaus' published handbook [13], several methods exist for estimating parameters in time-varying models, including nonparametric Whittle-type estimation, Yule-Walker estimation, local polynomial fitting, orthogonal series estimation, and kernel-type estimation. Here, we focus on maximum likelihood estimation based on kernel-type methods.

Nonparametric Whittle-type Estimation

The Whittle likelihood for a time-varying spectral density $f_t(\lambda)$ is given by:

$$Q(f) = \int_{-\pi}^{\pi} \left[\frac{I_T(\lambda)}{f_t(\lambda)} + \log f_t(\lambda) \right] d\lambda, \tag{3.5}$$

where $I_T(\lambda)$ is the periodogram of the process. The estimator minimizes $Q(f)$ over a class of smooth functions.

Yule-Walker Estimation (tvAR(p) - example)

The time-varying autoregressive (tvAR) parameters $\phi_k(t)$ are estimated using:

$$R_t(k) = \sum_{s=1}^{T} w_t(s) X_s X_{s-k}, \tag{3.6}$$

where $R_t(k)$ is the local autocovariance function and $w_t(s)$ is a weighting function emphasizing local information.

Local Polynomial Fitting

The parameter function $\theta(t)$ is approximated by a polynomial expansion:

$$\theta(t) \approx \sum_{j=0}^{p} \beta_j(t_0)(t - t_0)^j, \tag{3.7}$$

where the coefficients $\beta_j(t_0)$ are estimated by minimizing a locally weighted least squares criterion.

Orthogonal Series Estimation

The parameter function is expanded in terms of basis functions $\{\psi_j(t)\}$:

$$\theta(t) = \arg\min \frac{1}{T} \sum_{j=1}^{\infty} c_j \psi_j(\frac{t}{T}), \tag{3.8}$$

where the coefficients c_j are estimated using sample-based approximations.

Kernel-Type Estimation

The kernel-based estimator for a time-varying parameter $\theta(t)$ is given by:

$$\hat{\theta}(t) = \arg\min \frac{1}{T} \sum_{s=1}^{T} K_b(t - s) f_\theta, \tag{3.9}$$

where $K_b(\cdot)$ is a kernel function with bandwidth b, controlling the degree of smoothing.
In the next sections, we focus on the Maximum Likelihood estimation (MLE) based on kernel-type methods.

3.4 Kernel based QMLE

3.4.1 Kernel-based estimation

Kernel-based estimation plays a pivotal role in deriving parameter estimates for time-varying models. In this framework, the kernel serves as a localized weighting function that emphasizes observations near a given point $u_0 \in (0,1)$, while down-weighting those farther away. This localized approach allows for the estimation of parameters at u_0 by effectively focusing on the relevant segment of the data.

For a given u_0, let $t_0 \in \mathbb{N}$ be such that $|u_0 - t_0/T| < 1/T$. The kernel-based estimator is obtained by maximizing a weighted version of the conditional quasi-likelihood, where the weights are determined by the kernel function. The kernel function, typically denoted as K, is a smooth and symmetric function that assigns higher weights to observations closer to u_0 and lower weights to those farther away. Formally, K satisfies $K(x) = 0 \quad \text{for } x \notin \left[-\frac{1}{2}, \frac{1}{2}\right]$, and $\int K(x)\,dx = 1$. The bandwidth parameter $b > 0$, often referred to as the smoothing parameter, governs the degree of localization. A smaller b results in higher localization, focusing on a narrower segment of the data, while a larger b leads to a broader smoothing effect. The choice of b neglect bias and balances variance in the estimation.

In deriving the asymptotic properties of the kernel-based estimator, we rely on the local approximation of the time-varying process $X_{t,T}$ by its stationary version $\tilde{X}_t(u_0)$. This approximation ensures that the kernel-based estimator captures the local behavior of the process, enabling meaningful statistical inference. Importantly, this estimation method achieves a convergence rate of order $\sqrt{Tb}$, reflecting the effective sample size in the local neighborhood determined by the bandwidth, rather than the usual $\sqrt{T}$ rate for global estimators.

3.4.2 Localised QMLE for Time-varying models

Let $X_{t,T}$ be defined in (3.1). with ε_t are i.i.d. random variables. We then define the logarithm of the quasi-likelihood function, $\mathcal{L}_T(\theta(u))$, for a time-varying model, given the realization $X_1, \ldots, X_T$ of the model, as follows:

$$\mathcal{L}_T(\theta(u)) \quad = \quad \frac{1}{Tb} \sum_{t=1}^{T} K\left(\frac{\frac{t}{T} - u}{b}\right) \ell_t(\theta(u)). \tag{3.10}$$

with

$$\ell_t(\theta(u)) = \log\left(f_{\varepsilon_t}\left(X_{t,T} \mid \theta(u)\right)\right),$$

where $f_{\varepsilon_t}(\cdot \mid \theta(u))$ represents a density function of the probability law of the innovation ε_t, depending on the model parameters $\theta(u)$.

Here, $K : \mathbb{R} \to \mathbb{R}$ is a kernel function of bounded variation satisfying: K has a compact support, and

$K(x) = 0$ for $\mid x \mid \geq \dfrac{1}{2}$, and $\int_{\mathbb{R}} K(x)dx = 1$.

The kernel-based estimator of $\theta(u)$ is any measurable solution of the following problem:

$$\widehat{\theta}(u) = \arg\max_{\theta \in \Theta} \left(\frac{1}{Tb} \sum_{j=1}^{T} K\left(\frac{\frac{j}{T} - u}{b} \right) \ell\left(\theta(u)\right) \right), \quad u \in (0,1).$$

If we consider $\ell(x, \theta)$ as the contrast function for the time-varying process, the function ℓ is twice continuously differentiable with respect to θ, and its derivatives decay exponentially.

Clearly, $\mathcal{L}_T(\theta(u))$ represents the conditional likelihood of $X_{t,T}$ given $X_{t-1,T}, \ldots, X_{t-p,T}$ and the parameters θ. Since the conditional likelihood does not correspond to the full likelihood, it is referred to as a quasi-likelihood.

3.5 MLE for locally stationary AR(p) processes

3.5.1 Time-Varying AR(p) processes

Consider the classical AutoResgressive process with order p:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t,$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients, and ε_t is a white noise process with mean zero and variance σ^2.

We obtain the tvAR(p) with applying previous methode explained in section (3.2.1). and we get

$$X_{t,T} = \sum_{i=1}^{p} \phi_i \left(\frac{t}{T} \right) X_{t-i,T} + \varepsilon_t, \tag{3.11}$$

where $\phi_1\left(\frac{t}{T}\right), \phi_2\left(\frac{t}{T}\right), \ldots, \phi_p\left(\frac{t}{T}\right)$ are the time-varying autoregressive parameters, and ε_t is a white noise process with mean zero and variance σ^2.

We assume that $\phi_i(u) = \phi_i(0)$ for $u < 0$, along with additional smoothness conditions on $\phi(\cdot)$. We introduce a family of stationary approximations for the process $X_{t,T}$ in the neighborhood of a fixed time point u_0, as defined in the previous section:

$$\widetilde{X}_t(u_0) = \sum_{i=1}^{p} \phi_i\left(u_0\right) \widetilde{X}_{t-i}(u_0) + \varepsilon_t. \tag{3.12}$$

This stationary approximation satisfies

$$\left| X_{t,T} - \widetilde{X}_t(u_0) \right| = O_p\left(\left| \frac{t}{T} - u_0 \right| + \frac{1}{T} \right),$$

as shown in Theorem 3, and Dahlhaus [13] and [15].

3.5.2 QMLE estimation for tvAR(p)

In this section, we estimate the time-varying parameters of the time-varying AR(p) process using the local quasi-maximum likelihood estimator. We proceed to demonstrate that this estimator is consistent and asymptotically normal. First we introduce the likelihood function for a tvAR(p) process with gaussian noise

$$L(\phi) = \prod_{t=p+1}^{T} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(X_{t,T} - \sum_{i=1}^{p} \phi_i X_{t-i,T})^2 \right).$$

the corresponding log-likelihood function

$$\log L(\phi) = -\frac{T-p}{2}\log(2\pi) - \frac{1}{2}\sum_{t=p+1}^{T}(X_{t,T} - \sum_{i=1}^{p} \phi_i X_{t-i,T})^2.$$

Here $-\frac{T-p}{2}\log(2\pi)$ independent of $X_{t,T}$ so we can use this contrast fucntion

$$\ell_T(\phi) = \sum_{t=p+1}^{T}(X_{t,T} - \sum_{i=0}^{p} \phi_i X_{t-i,T})^2,$$

with $\phi_0 = -1$. Now we define the kernel based likelihood function as

$$\mathcal{L}_T(\phi(u)) = \frac{1}{Tb}\sum_{t=1}^{T} K\left(\frac{\frac{t}{T} - u}{b} \right) \ell_T(\phi(u)), \tag{3.13}$$

Let K be a kernel function satisfying $K(x) = 0$ for $x \notin [-\frac{1}{2}, \frac{1}{2}]$ and $\int K(x), dx = 1$. Replacing $X_{t,T}$ by its stationary approximation $\tilde{X}_t(u)$, we define the kernel-based estimator as follows

$$\hat{\phi} = \arg\min \mathcal{L}_T(\phi(u)). \tag{3.14}$$

3.5.3 Asymptotic behaviour of $\hat{\phi}$ estimator

In this section we discus the consistency and asymptotic normality of our estimator, under this important assumptions

Assumption LS(ρ): In the case of a time-varying AR(p) process (tvAR(p)), we set $\theta = \phi$.

Lemma 1. *Let $\Phi \subset \mathbb{R}^d$ be a compact parameter space. Then for all $\phi \in \Phi$*

$$\mathcal{L}(\phi) \geq \mathcal{L}(\phi_0),$$

with equality if and only if $\phi = \phi_0$.

Proof. Consider the kernel-based contrast difference

$$\mathcal{L}_T(\phi) - \mathcal{L}_T(\phi_0) = \frac{1}{Tb} \sum_{t=1}^{T} K\Big(\frac{t/T - u}{b}\Big) \big[\ell_t(\phi) - \ell_t(\phi_0)\big].$$

We have

$$\ell_t(\phi) - \ell_t(\phi_0) = \varepsilon_t^2(\phi) - \varepsilon_t^2(\phi_0).$$

Adding and subtracting appropriate terms $(-2\varepsilon_t(\phi_0)\varepsilon_t(\phi) + 2\varepsilon_t(\phi_0)\varepsilon_t(\phi) - \varepsilon_t^2(\phi_0) + \varepsilon_t^2(\phi_0))$ yields the identity

$$\ell_t(\phi) - \ell_t(\phi_0) = (\varepsilon_t(\phi) - \varepsilon_t(\phi_0))^2 + 2\varepsilon_t(\phi_0)(\varepsilon_t(\phi) - \varepsilon_t(\phi_0)).$$

Take expectations under the true parameter ϕ_0. Because $\mathbb{E}_{\phi_0}[\varepsilon_t(\phi_0) \mid \mathcal{F}_{t-1}] = 0$, the second term vanishes, leaving

$$\mathbb{E}_{\phi_0}[\ell_t(\phi) - \ell_t(\phi_0)] = \mathbb{E}_{\phi_0}[(\varepsilon_t(\phi) - \varepsilon_t(\phi_0))^2] \geq 0.$$

Equality holds iff $\varepsilon_t(\phi) = \varepsilon_t(\phi_0)$ almost surely, which by identifiability implies $\phi = \phi_0$.

Since the kernel weights are non-negative, multiplying by them and summing preserves the inequality. Letting $T \to \infty$ (or defining $\mathcal{L}(\phi)$ as the probability limit of $\mathcal{L}_T(\phi)$) gives

$$\mathcal{L}(\phi) - \mathcal{L}(\phi_0) \geq 0,$$

with equality only at $\phi = \phi_0$. Hence ϕ_0 is the unique minimizer of $\mathcal{L}(\phi)$. $\square$

Theorem 4. *Let $X_{t,T}$ be the solution of (3.11). Under the assumptions above, for any $u \in (0,1)$, if $p > 1$ and $bT^{1-\frac{1}{p}} \xrightarrow[T\to+\infty]{} \infty$, then the estimator $\hat{\phi}$ converges in probability to the true parameter ϕ_0*

$$\widehat{\boldsymbol{\phi}}(u) \xrightarrow[T\to+\infty]{\mathbb{P}} \boldsymbol{\phi}_0(u), \quad \text{if} \quad b \xrightarrow[T\to+\infty]{} 0 \quad \text{and} \quad bT \xrightarrow[T\to+\infty]{} \infty.$$

Proof. We aim to show that the estimator $\widehat{\phi}$ is consistent for the true parameter value ϕ_0.

By replacing $X_{t,T}$ with its stationary approximation $\tilde{X}_t(u)$, the localized criterion $\mathcal{L}_T(\phi(u))$ defined in (3.13) can be viewed as a kernel-weighted average of a stationary objective function. Under this local stationarity approximation and the moment condition $\mathbb{E}\big(\sup_{\phi\in\Phi}|\ell_t(\phi)|\big) < \infty$, a uniform law of large numbers applies, yielding

$$\sup_{\phi\in\Phi}|\mathcal{L}_T(\phi) - \mathcal{L}(\phi)| \xrightarrow[T\to\infty]{\mathbb{P}} 0,$$

where $\mathcal{L}(\phi) = \mathbb{E}(\ell_t(\phi))$ denotes the limiting (local) criterion.

Indeed, for all $t \in \mathbb{Z}$,

$$\begin{aligned} |\mathcal{L}_t(\phi)| &= \left|K\left(\frac{t/T-u}{b}\right)\ell_t(\phi)\right| \\ &\leq K\left(\frac{t/T-u}{b}\right)\left(\tilde{X}_t(u) - \sum_{i=0}^{p}\phi_i\tilde{X}_{t-i}(u)\right)^2 < \infty, \end{aligned} \tag{3.15}$$

which implies $\mathbb{E}\big(\sup_{\phi\in\Phi}|\mathcal{L}_t(\phi)|\big) < \infty$.

Consequently, the uniform convergence

$$\sup_{\phi\in\Phi}|\mathcal{L}_T(\phi) - \mathcal{L}(\phi)| \to 0 \quad \text{in probability}$$

holds. By Lemma 1, the function $\mathcal{L}(\phi)$ admits a unique minimizer ϕ_0 on the compact parameter space Φ. By uniform convergence of $\mathcal{L}_T(\phi)$ to $\mathcal{L}(\phi)$, with probability tending to one,

$$\inf_{\|\phi-\phi_0\|\geq\varepsilon}\mathcal{L}_T(\phi) > \mathcal{L}_T(\phi_0).$$

Let $\varepsilon > 0$ be arbitrary. Since $\widehat{\phi}$ minimizes $\mathcal{L}_T(\phi)$, it follows that $\|\widehat{\phi} - \phi_0\| < \varepsilon$ with probability tending to one. Hence,

$$\widehat{\phi} \xrightarrow{\mathbb{P}} \phi_0.$$

This completes the proof. ☐

Under the consistency assumptions and Assumption **A1**, we introduce the asymptotic normality theorem.
A1: ℓ is a continuously differentiable function of class C^2.

Theorem 5. *Let $X_{t,T}$ be the solution of (3.11). Under the consistency assumptions and Assumption **A1**, for any $u \in (0,1)$, if $p > 1$ and $bN^{1-\frac{1}{p}} \xrightarrow[T\to+\infty]{} \infty$, then the estimator $\hat{\phi}$ satisfies the asymptotic normality property:*

$$\sqrt{bT}\left(\widehat{\boldsymbol{\phi}}(u) - \boldsymbol{\phi}_0(u)\right) \xrightarrow[T\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, \Sigma(u)\right),$$

where $\Sigma(u)$ is the asymptotic covariance matrix, depending on the local properties of the process.

Proof. The proof follows the classical Taylor expansion argument for kernel-based quasi-maximum likelihood estimators.

Since $\widehat{\phi}(u)$ is a maximizer, it satisfies the first-order condition

$$\partial_\phi \mathcal{L}_T(\widehat{\phi}(u), u) = 0.$$

Using a second-order Taylor expansion of $\partial_\phi \mathcal{L}_T(\phi, u)$ around the true parameter $\phi_0(u)$, we obtain

$$0 = \partial_\phi \mathcal{L}_T(\phi_0(u), u) + \partial_\phi^2 \mathcal{L}_T(\tilde{\phi}_T(u), u)\left(\widehat{\phi}(u) - \phi_0(u)\right),$$

where $\tilde{\phi}_T(u)$ lies on the segment joining $\widehat{\phi}(u)$ and $\phi_0(u)$. Rearranging terms yields

$$\sqrt{bT}\left(\widehat{\phi}(u) - \phi_0(u)\right) = -\left[\partial_\phi^2 \mathcal{L}_T(\tilde{\phi}_T(u), u)\right]^{-1} \sqrt{bT}\, \partial_\phi \mathcal{L}_T(\phi_0(u), u).$$

We now analyze the two components separately. Under Assumption **A1**, the score function $\partial_\phi \ell_{t,T}(\phi)$ has finite second moments. Using the local stationarity approximation and standard kernel arguments, we obtain

$$\sqrt{bT}\, \partial_\phi \mathcal{L}_T(\phi_0(u), u) = \frac{1}{\sqrt{bT}} \sum_{t=1}^{T} K\left(\frac{t/T - u}{b}\right) \partial_\phi \ell_t(\phi_0(u)) + o_{\mathbb{P}}(1).$$

By a central limit theorem for locally stationary processes and the condition $bT^{1-\frac{1}{p}} \to \infty$, it follows that

$$\sqrt{bT}\, \partial_\phi \mathcal{L}_T(\phi_0(u), u) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\, V(u)),$$

where

$$V(u) = \int K^2(x)\,dx\;\mathbb{E}\big[\partial_\phi \ell_t(\phi_0(u))\partial_\phi \ell_t(\phi_0(u))^\top\big]\,.$$

By consistency (Theorem 4) and Assumption **A1**, we have

$$\partial^2_\phi \mathcal{L}_T(\tilde{\phi}_T(u), u) \xrightarrow{\mathbb{P}} \Gamma(u),$$

where

$$\Gamma(u) = \int K(x)\,dx\;\mathbb{E}\big[\partial^2_\phi \ell_t(\phi_0(u))\big]\,.$$

Moreover, $\Gamma(u)$ is non-singular by the identifiability condition (Lemma 1), implying

$$\Big[\partial^2_\phi \mathcal{L}_T(\tilde{\phi}_T(u), u)\Big]^{-1} \xrightarrow{\mathbb{P}} \Gamma(u)^{-1}.$$

Combining Steps 1 and 2 and applying Slutsky's theorem, we conclude that

$$\sqrt{bT}\left(\widehat{\phi}(u) - \phi_0(u)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0,\,\Gamma(u)^{-1}V(u)\Gamma(u)^{-1}\big)\,.$$

With $\Sigma(u) = \Gamma(u)^{-1}V(u)\Gamma(u)^{-1}$ completes the proof. □

3.6 MLE for locally stationary GARCH processes

3.6.1 locally stationary GARCH processes

The Generalized Autoregressive Conditional Heteroskedasticity GARCH defined in (1.1). It is based on modeling the conditional variance as an affine function of past squared innovations (Francq and Zakoian [20]).

The GARCH(p, q) model is represented as follows:

$$\begin{cases} r_t = \mu + X_t, \\ X_t = \sigma_t Z_t, \\ \sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \end{cases}$$

where (Z_t) is a sequence of i.i.d. random variables with $\mathbb{E}(z_t)=0$ and $\mathbb{E}(z_t^2)=1$.

In the case of a time-varying model, the parameters α_0, α_j, and β_i can be replaced by the following time dependent functions $\alpha_0\left(\frac{t}{T}\right)$, $\alpha_j\left(\frac{t}{T}\right)$ and $\beta_i\left(\frac{t}{T}\right)$. Thus, the time-varying GARCH (tvGARCH(p, q)) model satisfies the following representation:

$$\begin{cases} r_{t,T} = \mu + X_{t,T}, \\ X_{t,T} = \sigma_{t,T} Z_t, \\ \sigma_{t,T}^2 = \alpha_0\left(\dfrac{t}{T}\right) + \displaystyle\sum_{j=1}^{p} \alpha_j\left(\dfrac{t}{T}\right) X_{t-j,T}^2 + \sum_{i=1}^{q} \beta_i\left(\dfrac{t}{T}\right) \sigma_{t-i,T}^2. \end{cases} \tag{3.16}$$

The corresponding stationary approximation of these model at fixed u_0 is

$$\begin{cases} r_t = \mu + \tilde{X}_t(u_0) \\ \tilde{X}_t(u_0) = \sigma_t(u_0) Z_t \\ \sigma_t^2(u_0) = \alpha_0(u_0) + \displaystyle\sum_{j=1}^{p} \alpha_j(u_0)\tilde{X}_{t-j}^2(u_0) + \sum_{i=1}^{q} \beta_i(u_0)\sigma_{t-i}^2(u_0) \end{cases}$$

With

$$\sup_u \left(\sum_{j=1}^{p} \alpha_j(u) + \sum_{i=1}^{q} \beta_i(u) \right) < 1,$$

this approximation satisfies (3.4). Maximum likelihood estimation is used to estimate α_j and β_i based on the local conditional quasi-likelihood. The following approximation of the likelihood function is used when (Z_t) follows a Gaussian distribution.

$$\ell_{t,T}(\boldsymbol{\theta}) = \frac{1}{2}\log\sigma_t(\boldsymbol{\theta}) + \frac{X_{t,T}^2}{2\sigma_t(\boldsymbol{\theta})},$$

About the consistency and asymptotic normality Fryzlewicz, Subba Rao [22] and Bardet [4] prove that for localised approximated conditional quasi-likelihood estimator, their results are

For $u_0 \in [0,1]$, if $p>1$ and $bN^{1-\frac{1}{p}} \underset{T\to+\infty}{\longrightarrow} \infty$, if $b \underset{T\to+\infty}{\longrightarrow} 0$ and $bT \underset{T\to+\infty}{\longrightarrow} \infty$.

i) $\sup \left|\hat{\boldsymbol{\theta}}_T(u) - \boldsymbol{\theta}_0(u)\right| \xrightarrow{\mathbb{P}} 0.$

ii) $\sqrt{bT}\left(\hat{\boldsymbol{\theta}}_T(u_0) - \boldsymbol{\theta}_0(u_0)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Gamma(u_0)^{-1} V(u_0) \Gamma(u_0)^{-1}\right).$

In the next sections, we focus on the process $X_{t,T}$, assuming it has a tvGARCH representation, and study the asymptotic behavior of the estimators in the presence of non-Gaussian noise. More specifically, we analyze tvGARCH models with heavy-tailed distributions using local QMLE.

The key distinction between localized and stationary quasi-maximum likelihood estimation (QMLE) in terms of asymptotic normality lies in their respective convergence rates. In the stationary case, the rate is $\sqrt{T}$, whereas, in the non-stationary case, it is $\sqrt{Tb}$. Moreover, in the non-stationary setting, The corresponding minimax rate is smaller in magnitude, specifically $o(T^{1/3})$ when $\rho = 1$, compared to $O(\sqrt{T})$ in the stationary case, as noted by Bardet, Doukhan, and Wintenberger [3].

3.6.2 QMLE estimation for time-varying GARCH based on t-student distribution

The Student's t-distribution is widely used to model heavy-tailed data due to its ability to capture extreme values more effectively than the Gaussian distribution [39]. Its probability density function is expressed as

$$f_S(t) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{3.17}$$

where ν represents the degrees of freedom, and Γ denotes the gamma function. The variance remains undefined for $\nu \leq 2$ and is given by $\frac{\nu}{\nu-2}$ when $\nu > 2$. Additionally, higher-order moments are not defined for orders exceeding ν. As ν decreases, the distribution exhibits more pronounced heavy tails. Incorporating this distribution into the QMLE framework improves the robustness of volatility estimation by explicitly addressing tail risk, making it particularly valuable for financial modeling.

Let $X_{t,T}$ be defined in (6). We assume that Z_t are i.i.d. random variables following a Student's t-distribution with $\nu > 2$. We then define the quasi- log-likelihood function, $\ell_T(\theta(u))$, for a time-varying GARCH model with Student's t-distributed errors, given the realization $X_1, \ldots, X_n$ of the model, as follows:

$$\mathcal{L}_T(\theta(u)) = \frac{1}{Tb} \sum_{t=1}^{T} K\left(\frac{\frac{t}{T} - u}{b}\right) \ell_t(\theta(u))$$
$$\text{with } \ell_t(\theta(u)) = -\frac{\nu+1}{2} \log\left(1 + \frac{X_t^2}{(\nu-2)\sigma_t^2(\theta(u))}\right) - \frac{1}{2}\log\left(\sigma_t^2(\theta(u))\right)$$

Here, $K : \mathbb{R} \to \mathbb{R}$ is a bounded kernel function with compact support satisfying: $K(x) = 0$ for $x \notin [-\frac{1}{2}, \frac{1}{2}]$, and $\int K(x)dx = 1$.

The kernel-based estimator of $\theta(u)$ is any measurable solution of the following problem:

$$\widehat{\theta}_{td}(u) = \arg\max_{\theta\in\Theta}\left(\widetilde{\mathcal{L}}_T(\theta(u))\right), \quad u \in (0,1)$$

where $\widetilde{\mathcal{L}}$ denotes the same likelihood function, but with $X_{t,T}$ replaced by its stationary approximation to ensure the estimation of the behavior of u_0. And $\widehat{\theta}_{td}$ is the QMLE estimator of t-student ditributed noise.

If we consider $\ell(x,\theta)$ as the contrast function for the time-varying GARCH process, the function ℓ is twice continuously differentiable with respect to θ, and its derivatives decay exponentially. Clearly, $\ell_t(\theta(u))$ represents the conditional likelihood of $X_{t,T}$ given $X_{t-1,T},\ldots,X_{t-p,T}$ and the parameters θ_T, under the assumption that Z_t follows a Student's t-distribution. Since the conditional likelihood does not correspond to the full likelihood, it is referred to as a quasi-likelihood.

Consistency

To establish the strong consistency of the quasi-maximum likelihood estimator (QMLE) for the time-varying GARCH model with t-Student distributed noise, we need the following assumptions:

H1 $\sigma_t^2(.)$ is a Lipschitz function, $\sigma_t^2(\theta) > c$ and $\widehat{\sigma}_t^2(\theta) > c$, a.s. for some deterministic constant $c > 0$.

H2 $a_t \underset{t\to+\infty}{\xrightarrow{a.s.}} 0$ and $a_t Z_t^2 \underset{t\to+\infty}{\xrightarrow{a.s.}} 0$, where $a_t = \sup_{\theta\in\Theta} \mid \widehat{\sigma}_t^2(\theta) - \sigma_t^2(\theta) \mid$.

H3 Moments exist for Z_t such that $\mathbb{E}\left(|Z_t|^\delta\right) < \infty$ for some $\delta > 2$.

H4 $\sigma_t^2(\theta) = \sigma_t^2(\theta_0)$ a.s. if and only if $\theta = \theta_0$.

H5 Θ is a compact set in $\mathbb{R}^d$, and $\theta_0 \in \Theta$.

Theorem 6. *Under **H1-H5** and for any $u \in (0,1)$, the estimator $\widehat{\theta}(u)$ consistently estimates $\theta_0(u)$:*

$$\widehat{\theta}_{td}(u) \underset{T\to+\infty}{\xrightarrow{a.s.}} \theta_0(u), \tag{3.18}$$

if $b \underset{T\to+\infty}{\longrightarrow} 0$ and $Tb \underset{T\to+\infty}{\longrightarrow} \infty$.

Proof. In the first step, a uniform (in $\theta \in \Theta$) strong law of large numbers is established, such that $\widehat{\ell}_T(\theta)$ converges to $\ell(\theta) := \mathbb{E}[\mathcal{L}_0(\theta)]$ for each $\theta \in \Theta$. In the second step, it is shown that $\ell(\theta)$ has a unique maximum at $\theta_0(u)$. These results establish the strong consistency of $\hat{\theta}_{td}(u)$.

If $(\mathcal{L}_t)_{t\in\mathbb{Z}}$ is a stationary ergodic sequence of random elements taking values in $\mathbb{C}(\Theta,\mathbb{R}^m)$, the uniform law of large numbers holds if $\mathbb{E}\left[\sup_\theta |\mathcal{L}_t(\theta)|\right]_\Theta < \infty$. Thus, for all $t \in \mathbb{Z}$:

$$\begin{aligned}
|\mathcal{L}_t(\theta)| &= \left|\frac{1}{Tb}\sum_t \ell_t(\theta) K\left(\frac{\frac{t}{T}-u}{b}\right)\right| \\
&\leq \frac{1}{Tb}\sum_t K\left(\frac{\frac{t}{T}-u}{b}\right)\left[\frac{\nu+1}{2}\left|\log\left(1+\frac{X_t^2}{(\nu-2)\widehat{\sigma}_t^2(\theta)}\right)\right| + \frac{1}{2}\left|\log\left(\widehat{\sigma}_t^2(\theta)\right)\right|\right]. \\
&\leq \frac{1}{Tb}\sum_t K\left(\frac{\frac{t}{T}-u}{b}\right)\left[(\nu+1)\left|\log\left(1+\frac{X_t^2}{(\nu-2)\widehat{\sigma}_t^2(\theta)}\right)\right| + \left|\log\left(\widehat{\sigma}_t^2(\theta)\right)\right|\right]. \\
&\leq \frac{1}{Tb}\sum_t K\left(\frac{\frac{t}{T}-u}{b}\right)\left[\frac{(\nu+1)}{\nu-2}\times\frac{X_t^2}{\sigma} + \left|\log\left(\sigma+\frac{\sigma_t^2}{\sigma}-1\right)\right|\right].
\end{aligned}$$

$$\Longrightarrow \sup_{\theta\in\Theta}|\mathcal{L}_t(\theta)| \leq \frac{1}{Tb}\sum_t K\left(\frac{\frac{t}{T}-u}{b}\right)\left[\frac{\nu+1}{\nu-2}\times\frac{X_t^2}{\sigma} + |\log\sigma| + \frac{\left\|\sigma_t^2\right\|_\Theta}{\sigma}\right]$$

For all $t \in \mathbb{Z}, \mathbb{E}X_t^2 < \infty$, and $\mathbb{E}\left\|\sigma\right\|_\Theta < \infty$. As a consequence

$$\mathbb{E}\left[\sup_{\theta\in\Theta}|\mathcal{L}_t(\theta)|\right] \quad < \quad \infty$$

By assumptions **H1**, **H3**.

Thus, the uniform strong law of large numbers for $(\mathcal{L}_t(\theta))_{t\in\mathbb{Z}}$ holds:

$$\left\|\ell_t(\theta)-\ell(\theta)\right\|_\Theta \underset{T\to+\infty}{\overset{a.s.}{\longrightarrow}} 0, \quad \text{with} \quad \ell(\theta) := \mathbb{E}[\mathcal{L}(\theta_0)]. \tag{3.19}$$

Next, we establish:

$$\left\|\widehat{\ell}_t(\theta)-\ell_t(\theta)\right\|_\Theta \underset{T\to+\infty}{\overset{a.s.}{\longrightarrow}} 0.$$

For all $\theta \in \Theta$ and $t \in \mathbb{N}$:

$$\begin{aligned}
\left|\widehat{\ell}_t(\theta)-\ell_t(\theta)\right| =& \frac{1}{2}\Bigg|\log\widehat{\sigma}_t^2(\theta) + (\nu+1)\log\left(1+\frac{1}{(\nu-2)}\times\frac{X_t^2}{\widehat{\sigma}_t^2(\theta)}\right) \\
&-\log\sigma_t^2(\theta) - (\nu+1)\log\left(1+\frac{1}{(\nu-2)}\times\frac{X_t^2}{\sigma_t^2(\theta)}\right)\Bigg| \\
&\leq \left|\log\widehat{\sigma}_t^2(\theta)-\log\sigma_t^2(\theta)\right| \\
&+\left|(\nu+1)\left[\log\left(1+\frac{1}{(\nu-2)}\times\frac{X_t^2}{\widehat{\sigma}_t^2(\theta)}\right)-\log\left(1+\frac{1}{(\nu-2)}\times\frac{X_t^2}{\sigma_t^2(\theta)}\right)\right]\right| \\
&\leq \left|\widehat{\sigma}_t^2(\theta)-\sigma_t^2(\theta)\right| + \frac{(\nu+1)}{(\nu-2)}\left|\frac{X_t^2}{\widehat{\sigma}_t^2(\theta)}-\frac{X_t^2}{\sigma_t^2(\theta)}\right| \\
&\leq \left|\widehat{\sigma}_t^2(\theta)-\sigma_t^2(\theta)\right| + \frac{(\nu+1)}{(\nu-2)}\left|\frac{1}{\widehat{\sigma}_t^2(\theta)}-\frac{1}{\sigma_t^2(\theta)}\right|X_t^2
\end{aligned}$$

Therefore, we have:

$$\left|\widehat{\ell}_t(\theta) - \ell_t(\theta)\right| \leq \left\|\widehat{\sigma}_t^2(\theta) - \sigma_t^2(\theta)\right\|_\Theta + \frac{X_t^2}{(\nu-2)}\left\|\frac{1}{\widehat{\sigma}_t^2(\theta)} - \frac{1}{\sigma_t^2(\theta)}\right\|_\Theta$$

On the one hand we can write,

$$\left\|\frac{1}{\widehat{\sigma}_t^2(\theta)} - \frac{1}{\sigma_t^2(\theta)}\right\|_\Theta \leq \left\|\frac{1}{\widehat{\sigma}_t^2(\theta)}\right\|_\Theta \left\|\widehat{\sigma}_t^2(\theta) - \sigma_t^2(\theta)\right\|_\Theta \left\|\frac{1}{\sigma_t^2(\theta)}\right\|_\Theta,$$

and

$$\left\|\frac{1}{\sigma_t^2(\theta)}\right\|_\Theta \leq \sigma^{-1}$$

So

$$\sup_{\theta\in\Theta}\left|\widehat{\ell}_t(\theta) - \ell_t(\theta)\right| \leq C\left[\left\|\widehat{\sigma}_t^2(\theta) - \sigma_t^2(\theta)\right\|_\Theta \left(\frac{1}{\sigma^2} + \frac{(\nu+1)}{(\nu-2)}X_t^2\right)\right]$$

From the Hölder and Minkowski inequalities we can write:

$$E\left[\left(|XY|^{2/3}\right)\right] \leq E\left[|X|^{2p/3}\right]\right)^{1/p}\left(E\left[|Y|^{2q/3}\right]\right)^{1/q}$$

If $p = 3$ and $q = 3/2 \Rightarrow$

$$E\left[\left(|XY|^{2/3}\right)\right] \leq E\left[|X|^2\right]\right)^{1/3}(E[|Y|])^{2/3}$$

Therefore

$$\begin{aligned} E\left[\left(\sup_{\theta\in\Theta}\left|\widehat{\ell}_t(\theta) - \ell_t(\theta)\right|\right)^{2/3}\right] &\leq C\left(E\left[\left\|\widehat{\sigma}_t^2(\theta) - \sigma_t^2(\theta)\right\|_\Theta\right]\right)^{\frac{2}{3}} \times \left(E\left[\frac{1}{\sigma^2} + \frac{(\nu+1)}{(\nu-2)}X_t^2\right]\right)^{\frac{1}{3}} \\ &\leq C'\left(\sum_{j=t+1}^{\infty}\alpha_j^{(0)}\right)^{2/3} \end{aligned}$$

for all $\theta \in \Theta$, where $C' > 0$, in view of assumption H3 $\mathbb{E}\left[X_t^2\right] < \infty$. Consider, for $T \in \mathbb{N}^*$,

$$S_T = \sum_{t=1}^{n}\frac{1}{t}\sup_{\theta\in\Theta}\left|\widehat{\ell}_t(\theta) - \ell_t(\theta)\right|.$$

Applying the Kronecker lemma, if $\lim_{T\to\infty} S_T < \infty$ a.s, then $\left\|\widehat{\ell}_T(\theta(u)) - \ell_T(\theta(u))\right\|_\Theta \to 0$ a.s, $Tb \to \infty$.

In the second step, for $\theta \in \Theta$, we use $\log x \leq x - 1$ and assumptions **H1**, **H2**, we have:

$$\begin{aligned}
\ell(\theta) - \ell(\theta_0) =& -\frac{1}{2}K\left(\frac{\frac{t}{T}-u}{b}\right)\mathbb{E}\left[\log\sigma_t^2(\theta) + (\nu+1)\log\left(1+\frac{1}{\nu-2}\times\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}Z_t^2\right)\right]\\
&+\frac{1}{2}K\left(\frac{\frac{t}{T}-u}{b}\right)\mathbb{E}\left[\log\sigma_t^2(\theta_0) + (\nu+1)\log\left(1+\frac{1}{\nu-2}Z_t^2\right)\right]\\
\leq& K\left(\frac{\frac{t}{T}-u}{b}\right)\left[-\mathbb{E}\left(\log\sigma_t^2(\theta) - \log\sigma_t^2(\theta_0)\right)\right.\\
&\left.-(\nu+1)\mathbb{E}\left[\log\left(1+\frac{1}{\nu-2}\times\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}Z_t^2\right) - \log\left(1+\frac{1}{\nu-2}Z_t^2\right)\right]\right]\\
\leq& -K\left(\frac{\frac{t}{T}-u}{b}\right)\mathbb{E}\left[\log\frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)} + (\nu+1)\left\{\log\left(1+\frac{1}{\nu-2}\times\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}Z_t^2\right)\right.\right.\\
&\left.\left.-\log\left(1+\frac{1}{\nu-2}Z_t^2\right)\right\}\right]\\
\leq& -K\left(\frac{\frac{t}{T}-u}{b}\right)\mathbb{E}\left[-\log\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} + \frac{\nu+1}{\nu-2}\times\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}Z_t^2 - \frac{\nu+1}{\nu-2}Z_t^2\right]
\end{aligned}$$

We use $\log x \leq x - 1$

$$\begin{aligned}
&\mathbb{E}\left[-\log\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} + \frac{\nu+1}{\nu-2}\times\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}Z_t^2 - \frac{\nu+1}{\nu-2}Z_t^2\right]\\
&=\mathbb{E}\left[-\log\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} + \frac{\nu+1}{\nu-2}\times\left(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}-1\right)Z_t^2\right]\\
&=\mathbb{E}\left(-\log\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\right) + \mathbb{E}\left(\frac{\nu+1}{\nu-2}\times\left(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}-1\right)\right)\mathbb{E}\left(Z_t^2\right).
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\Bigg[-\log\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} + \frac{\nu+1}{\nu-2}\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}Z_t^2 - \frac{\nu+1}{\nu-2}Z_t^2\Bigg]\\
&\quad=\mathbb{E}\Bigg[-\log\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} + \frac{\nu+1}{\nu-2}\Big(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}-1\Big)Z_t^2\Bigg]\\
&\quad\leq\mathbb{E}\Bigg[-\log\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} + \frac{\nu+1}{\nu-2}\Big(\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}-1\Big)\Bigg],
\end{aligned}$$

which implies

$$\ell(\theta) - \ell(\theta_0) \leq 0 \quad \text{for all } \theta \in \Theta.$$

By assumption **H3** (identifiability), this is true if and only if $\theta = \theta_0$. Hence, θ_0 is the unique maximizer $\sup_{\theta\in\Theta} \ell(\theta) = \ell(\theta_0)$.

The equality $\ell(\theta) - \ell(\theta_0) = 0$ holds if and only if $\sigma_t^2(\theta) = \sigma_t^2(\theta_0)$. By assumption **H4** (identifiability), this is true if and only if $\theta = \theta_0$. □

Asymptotic normality

To provide the asymptotic normality result, we impose the following additional assumptions:

H6 The parameter vector $\theta \in \Theta$ is restricted to a bounded set.

H7 For any $\theta \in \Theta$: $\partial_\theta \sigma^2(\theta)\left(\left(\widetilde{X}_{-k}(u)\right)_{k\in\mathbb{N}}\right) \neq 0 \quad \text{a.s.}$

H8 $(b)_T$ is a sequence of positive numbers such that:

$$b \underset{T\to+\infty}{\longrightarrow} \infty \quad \text{and} \quad Tb^{1+2\rho} \underset{T\to+\infty}{\longrightarrow} 0.$$

H9 The matrices:

- $\Gamma(\theta_0(u)) = E\left(\dfrac{\partial^2}{\partial\theta^2}\ell(\theta_0(u))\right)$,
- $\Sigma(\theta_0(u)) = \int_{\mathbb{R}} K^2(x)dx \sum_{t\in\mathbb{Z}} \operatorname{cov}\left(\frac{\partial}{\partial\theta_i}\ell(\theta_0(u)), \frac{\partial}{\partial\theta_j}\ell(\theta_0(u))\right)_{1\leq i,j}$

are finite, and $\Sigma(\theta_0(u))$ is a definite positive matrix.

Theorem 7. *Let $(X_{t,T})$ satisfy (3.16). Under **H1-H9** and assumption $LS(\rho)$, for any $u \in (0,1)$ and for all $\nu > 2$, we have:*

$$\sqrt{Tb}\left(\widehat{\theta}_{td}(u) - \theta_0(u)\right) \underset{T\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(0, \Gamma^{-1}(\theta_0(u))\Sigma(\theta_0(u))\Gamma^{-1}(\theta_0(u))\right). \tag{3.20}$$

Proof. The proof of asymptotic normality follows the standard approach involving three steps: 1. Using the difference of martingale sequences. 2. Showing almost sure convergence of certain quantities. 3. Applying the Central Limit Theorem (CLT) for dependent data sequences.

In this proof, we rely on the following facts:

- **1.** $\hat{\theta}_{td}(u)$ is the maximizer of $\mathcal{L}_T(\theta)$ by definition.

- **2.** $\theta_0(u)$ is the maximizer of $\mathcal{L}(\theta)$ by assumption.

- **3.** $\mathcal{L}_T(\theta) \to \mathcal{L}(\theta) = E(\ell \mid \mathcal{F}_0)$ by the Law of Large Numbers.

Applying the Mean Value Theorem:

$$\frac{f(\beta) - f(\alpha)}{\beta - \alpha} = f'(c) \quad \text{or} \quad f(\alpha) = f(\beta) + (\alpha - \beta) f'(c) \quad \text{with} \quad c \in [\alpha, \beta].$$

Let $f(\theta) = \dfrac{\partial \mathcal{L}_T(\theta)}{\partial \theta}$, $a = \hat{\theta}(u)$, and $\beta = \theta_0(u)$. Then:

$$0 = \frac{\partial \mathcal{L}_T(\hat{\theta}_{td}(u))}{\partial \theta} = \frac{\partial \mathcal{L}_T(\theta_0(u))}{\partial \theta} + \frac{\partial^2 \mathcal{L}_T(\bar{\theta}(u))}{\partial \theta^2} (\hat{\theta}_{td}(u) - \theta_0(u)),$$

for some $\bar{\theta}(u) \in [\hat{\theta}_{td}(u), \theta_0(u)]$.

Since $\hat{\theta}(u)$ maximizes $\mathcal{L}_T(\theta)$:

$$\begin{aligned} 0 &= \frac{1}{Tb} \sum_{t=1}^{T} \frac{\partial \ell(\hat{\theta}_{td}(u))}{\partial \theta} K\left(\frac{\frac{t}{T} - u}{b}\right) \\ &= \frac{1}{Tb} \sum_{t=1}^{T} \frac{\partial \ell(\theta_0(u))}{\partial \theta} K\left(\frac{\frac{t}{T} - u}{b}\right) + \frac{1}{Tb} \sum_{t=1}^{T} \frac{\partial^2 \ell(\bar{\theta}(u))}{\partial \theta^2} K\left(\frac{\frac{t}{n} - u}{b}\right) (\hat{\theta}_{td}(u) - \theta_0(u)). \end{aligned}$$

Multiplying by $\sqrt{Tb}$:

$$\begin{aligned} 0 &= \frac{1}{\sqrt{Tb}} \sum_{t=1}^{T} \frac{\partial \ell(\theta_0(u))}{\partial \theta} K\left(\frac{\frac{t}{T} - u}{b}\right) \\ &+ \frac{\sqrt{Tb}}{Tb} \sum_{t=1}^{T} \frac{\partial^2 \ell(\bar{\theta}(u))}{\partial \theta^2} K\left(\frac{\frac{t}{T} - u}{b}\right) (\hat{\theta}_{td}(u) - \theta_0(u)). \end{aligned}$$

From this, we have:

$$\sqrt{Tb}(\hat{\theta}_{td}(u) - \theta_0(u)) = -\sqrt{Tb} \frac{\frac{\partial \mathcal{L}_T(\theta_0(u))}{\partial \theta}}{\frac{\partial^2 \mathcal{L}_T(\bar{\theta}(u))}{\partial \theta^2}}.$$

Since $\theta_0(u)$ maximizes $\mathcal{L}(\theta) = E(\ell \mid \mathcal{F}_0)$, we have:

$$\frac{\partial E(\ell \mid \mathcal{F}_0)}{\partial \theta} = 0. \tag{3.21}$$

By the Law of Large Numbers in the denominator

$$\frac{\partial^2 \mathcal{L}_T(\theta)}{\partial \theta^2} = \frac{1}{Tb} \sum_{t=1}^{T} \frac{\partial^2 \ell(\theta)}{\partial \theta^2} K\left(\frac{\frac{t}{T} - u}{b}\right) \xrightarrow[T \to +\infty]{\mathcal{P}} E\left[\frac{\partial^2 \ell(\theta_0(u))}{\partial \theta^2}\right].$$

Since $\bar{\theta}(u) \in [\hat{\theta}_{td}(u), \theta_0(u)]$ and $\hat{\theta}_{td}(u) \xrightarrow[T \to +\infty]{\mathbb{P}} \theta_0(u)$ by consistency, it follows that:

$$\frac{\partial^2 \mathcal{L}_T(\bar{\theta}(u))}{\partial \theta^2} \xrightarrow[T \to +\infty]{\mathbb{P}} \Gamma(\theta_0(u)),$$

where $\Gamma(\theta_0(u)) = E\left[\frac{\partial^2 \ell(\theta_0(u))}{\partial \theta^2}\right]$.

In the treatment of the numerator, the gradient vector of ℓ constitutes a difference of martingale sequences with respect to the filtration $\{\mathcal{F}_t, t \in \mathbb{Z}\}$. Under conditions similar to Bardet, Doukhan, and Wintenberger, and assuming $E\left[\|\frac{\partial \ell}{\partial \theta}\|^2\right] < \infty$, the multidimensional Central Limit Theorem (CLT) gives:

$$\frac{1}{\sqrt{Tb}} \sum_{t=1}^{T} \frac{\partial \ell(\theta_0(u))}{\partial \theta} K\left(\frac{\frac{t}{T} - u}{b}\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta_0(u))),$$

where

$$\Sigma(\theta_0(u)) = \int_{\mathbb{R}} K^2(x)dx \sum_{t \in \mathbb{Z}} \operatorname{cov}\left(\frac{\partial \ell(\theta_0(u))}{\partial \theta_i}, \frac{\partial \ell(\theta_0(u))}{\partial \theta_j}\right)_{1 \le i,j}.$$

By Slutsky's lemma and (3.21), we obtain:

$$\sqrt{Tb}(\hat{\theta}_{td}(u) - \theta_0(u)) \xrightarrow[T \to +\infty]{\mathcal{L}} \mathcal{N}\left(0, \Gamma^{-1}(\theta_0(u))\Sigma(\theta_0(u))\Gamma^{-1}(\theta_0(u))\right).$$

□

3.6.3 QMLE estimation for time-varying GARCH with Laplace distributed noise

The Laplace distribution, while similar to the normal distribution in density, differs fundamentally in its formulation; the normal distribution involves the squared difference $(x-\mu)^2$, whereas the Laplace distribution uses the absolute difference $|x - \mu|$. Notably, the Laplace distribution is taller at the mean and has fatter tails compared to the Gaussian and Student distributions, making it particularly well-suited for modeling the white noise in financial time series. Its ability to capture extreme events effectively addresses the limitations of Gaussian models and enhances the applicability of the tvGARCH framework. This approach broadens the scope of volatility modeling and strengthens the analytical properties of existing models, offering a robust alternative for analyzing non-stationary or locally stationary time series with heavy-tailed characteristics. By incorporating the Laplace distribution, we provide new insights into the behavior of conditional distributions under non-Gaussian assumptions, with applications in financial modeling, such as derivative pricing and risk management [29].

The Laplace distribution, has the probability density function given by

$$f(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}, \tag{3.22}$$

where μ represents the location parameter and $b > 0$ is the scale parameter. Thus, the logarithm of the conditional quasi-likelihood is expressed as

$$\mathcal{L}_T(\theta(u)) = \frac{1}{Tb} \sum_{t=1}^{T} K\left(\frac{\frac{t}{T} - u}{b}\right) \ell_{t,T}(\theta(u)). \tag{3.23}$$

with

$$\ell_{t,T}(\theta) = \log \sigma_{t,T}(\theta) + \frac{|X_t - \mu|}{b\sigma_{t,T}(\theta)}. \tag{3.24}$$

Using the stationary approximation $\widetilde{X}_t(u_0)$ of the process $X_{t,T}$ to simplify analysis, ensure asymptotic validity, and improve parameter estimation by leveraging the locally stationary behavior of the process, we define the likelihood function as

$$\widetilde{\mathcal{L}}_T(u_0, \theta) = \frac{1}{bT} \sum_{t=1}^{T} K\left(\frac{u_0 - \frac{t}{T}}{b}\right) \widetilde{\ell}_{t,T}(u_0, \theta),$$

where

$$\widetilde{\ell}_{t,T}(u_0, \theta) = \log \widetilde{\sigma}_t(u_0) + \frac{\left|\widetilde{X}_t(u_0) - \mu\right|}{b\widetilde{\sigma}_t(u_0)}.$$

This leads to the local QMLE:

$$\widehat{\theta}_{Lap}(u_0) := \arg\min_{\theta \in \Theta} \widetilde{\mathcal{L}}_T(u_0, \theta).$$

Here, $\widehat{\theta}_{Lap}(u_0)$ denot the QMLE of θ_0 in the point u_0 for Laplace distributed noise.

Asymptotic properties of $\widehat{\theta}_{Lap}$

We introduce the consistency and asymptotic normality theorems , as presented by Aissaoui F and Djaballah K [1], under the same previous asumptions H and $LS(\rho)$.

Consistency

Lemma 2. *Suppose* $\{X_{t,T} : t = 1, \ldots, T\}$ *is a* tvGARCH(p, q) *process satisfying Assumptions above. Let* $\tilde{\mathcal{L}}_T(u_0, \boldsymbol{\theta}), \mathcal{L}(u_0, \boldsymbol{\theta})$ *be the likelihood functions, let the bias caused by the deviation from stationarity* $\mathcal{B}_{t_0,T}$ *defined as* $\mathcal{B}_{t_0,T} = \mathcal{L}_{t_0,T}(\boldsymbol{\theta}) - \tilde{\mathcal{L}}_T(u_0, \boldsymbol{\theta})$ *Then*

$$\sup_{\boldsymbol{\theta}\in\Theta} \left|\tilde{\mathcal{L}}_T(u_0, \boldsymbol{\theta}) - \mathcal{L}(u_0, \boldsymbol{\theta})\right| \xrightarrow{\mathbb{P}} 0,$$

$$\sup_{\boldsymbol{\theta}\in\Theta} |\mathcal{B}_{t_0,T}(\boldsymbol{\theta})| \xrightarrow{\mathbb{P}} 0,$$

where $b \to 0$ *and* $bT \to \infty$ *as* $T \to \infty$.

Proof. i) By definition, we have

$$\widetilde{\mathcal{L}}_T(u_0, \theta) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{b} K\left(\frac{u_0 - \frac{t}{T}}{b}\right) \widetilde{\ell}_{t,T}(\theta),$$

where

$$\widetilde{\ell}_{t,T}(\theta) = \log \widetilde{\sigma}_t(u_0) + \frac{\left|\tilde{X}_t(u_0) - 1\right|}{\widetilde{\sigma}_t(u_0)}.$$

ii) Consider the difference

$$\widetilde{\mathcal{L}}_T(u_0, \theta) - \mathcal{L}(u_0, \theta) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{b} K\left(\frac{u_0 - \frac{t}{T}}{b}\right) \widetilde{\ell}_{t,T}(\theta) - \mathbb{E}(\widetilde{\ell}_0(u_0, \theta)).$$

Define the deviation term

$$z_t(\theta, u) = \widetilde{\ell}_{t,T}(\theta) - \mathbb{E}(\widetilde{\ell}_0(u, \theta)).$$

Applying the law of large numbers, and using the results from the previous lemma, we obtain

$$\frac{1}{T}\sum_{t=1}^{T} z_{t,T} \overset{T\to\infty}{\to} \mathbb{E}(\widetilde{\ell}_0(u, \theta)).$$

Thus, it follows that

$$\widetilde{\mathcal{L}}_T(u_0, \theta) \to \mathcal{L}(u_0, \theta).$$

□

Theorem 8. *Under condition of lemma 2, and assumption $LS(\rho)$, and for any $u \in (0,1)$, the estimator $\widehat{\theta}(u)$ consistently estimates $\theta_0(u)$:*

$$\widehat{\theta}_{Lap}(u) \xrightarrow[T\to+\infty]{a.s.} \theta_0(u), \tag{3.25}$$

if $b \underset{T\to+\infty}{\longrightarrow} 0$ and $Tb \underset{T\to+\infty}{\longrightarrow} \infty$.

Proof. From the result of Lemma 2, we have pointwise convergence:

$$\mathcal{L}_{t_0,T}(\boldsymbol{\theta}) \xrightarrow{\mathbb{P}} \mathcal{L}(u_0, \boldsymbol{\theta}),$$

where $\widehat{\boldsymbol{\theta}}_{Lap}(u_0) = \operatorname{argmin}_{\boldsymbol{\theta}} \mathcal{L}(u_0, \boldsymbol{\theta})$. It follows that:

$$\mathcal{L}_{t_0,T}(\widehat{\boldsymbol{\theta}}_{Lap,t_0,T}) \leq \mathcal{L}_{t_0,T}(\boldsymbol{\theta}_0(u_0)) \xrightarrow{\mathbb{P}} \mathcal{L}(u_0, \boldsymbol{\theta}_0(u_0)) \leq \mathcal{L}(u_0, \widehat{\boldsymbol{\theta}}_{Lap,t_0,T}).$$

Using Lemma 2 again, we obtain:

$$\mathcal{L}_{t_0,T}(\widehat{\boldsymbol{\theta}}_{Lap,t_0,T}) \xrightarrow{\mathbb{P}} \mathcal{L}(u_0, \boldsymbol{\theta}_0(u_0)).$$

From the continuity of $\mathcal{L}(u_0, \cdot)$ and the compactness of the parameter space, we conclude:

$$\widehat{\boldsymbol{\theta}}_{Lap,t_0,T} \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0(u_0),$$

provided that $\mathcal{L}(u_0, \cdot)$ has a unique minimizer. Note that $\mathcal{L}(u_0, \cdot)$ corresponds to the stationary case considered in [28].

□

Asymptotic normality

Theorem 9. *Let* $(X_{t,T})$ *satisfy* (3.16). *Under consistency conditions and assumption* $LS(\rho)$*, for any* $u \in (0,1)$*, we have:*

$$\sqrt{Tb}\left(\widehat{\theta}_{Lap}(u) - \theta_0(u)\right) \underset{T\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(0, \Gamma^{-1}(\theta_0(u))\Sigma(\theta_0(u))\Gamma^{-1}(\theta_0(u))\right). \tag{3.26}$$

Proof. The asymptotic normality of the estimator is established through a Taylor expansion of the quasi-likelihood function around $\theta_0(u)$.

The derivative $\frac{\partial \mathcal{L}_T(\theta)}{\partial \theta}$ generates δ-functions, which are differentiable almost everywhere on Θ (excluding a subset of measure zero). For a Laplace-distributed variable, the likelihood function features the term:

$$\ell_{t,T}(\theta) = \log \sigma_t + \frac{|X_{t,N} - \mu|}{b\sigma_t},$$

which behaves similarly to the Least Absolute Deviation $|X_{t,N} - \mu|$. This function is differentiable except at $X_t = \mu$. Since we assume $P(X_t = \mu) = 0$, the derivative can still be computed in a generalized sense. The absolute value function introduces a discontinuity at $X_t = 1$, where its derivative is given by $\text{sign}(X_t - 1)$. However, the sign function $\text{sign}(X_t - \mu)$ is not continuous at $X_t = \mu$, making it non-differentiable at this point in the usual sense. To handle this, we use the fact that differentiating $\text{sign}(x)$ results in a Dirac delta function:

$$\frac{\partial}{\partial x}\text{sign}(x) = 2\delta(x).$$

Thus, at $X_t = \mu$, the derivative generates a Dirac delta function, which concentrates probability mass at that point. Therefore, the derivative of the likelihood function $\mathcal{L}_{t,T}(u, \theta)$ is written as:

$$\frac{\partial \mathcal{L}_{t,T}(u,\theta)}{\partial \theta} = \frac{1}{bT}\sum_{t=1}^{T} K(\frac{u_0 - \frac{t}{T}}{b})\frac{\partial \ell_{t,T}(u,\theta)}{\partial \theta}.$$

This formulation allows us to incorporate the non-differentiability of the Laplace(1,1) likelihood while maintaining a valid derivative expression in a generalized sense. For more details, see [9] and [2].

We have $\dfrac{\partial \widetilde{\ell}_{t,T}(u_0, \boldsymbol{\theta}_0)}{\partial \theta} = \dfrac{\partial}{\partial \theta}\left(\log \sigma_t(u_0) + \frac{|\tilde{X}_t(u) - \mu|}{b\sigma_t(u_0)}\right)$ that satisfies a multidimensional Central Limit

Theorem, We notice that $\dfrac{\partial \mathbb{E}\left(\widetilde{\ell}_{t,T}(u_0,\boldsymbol{\theta}_0)|\mathcal{F}_0\right)}{\partial \theta} = 0$ as $(\boldsymbol{\theta_0})$ is the unique minimiser of $\mathbb{E}\left(\widetilde{\ell}_{t,T}(u_0,\boldsymbol{\theta}_0)|\mathcal{F}_0\right)$ over $\mathring{\Theta}$.

$\theta \in \Theta \to \mathbb{E}\left(\widetilde{\ell}_{t,T}(\boldsymbol{\theta}_0)|\mathcal{F}_0\right)$ is differentiable under the condition $\left|\dfrac{\partial \widetilde{\ell}_{t,T}(u_0,\boldsymbol{\theta}_0)}{\partial \theta}\right| \in \mathrm{Lip}_p(\Theta)$. Thus $\dfrac{\partial \widetilde{\ell}_{t,T}(u_0,\boldsymbol{\theta}_0)}{\partial \theta}$ constitutes a differences of martingale sequence. We also have $\mathbb{E}\left[\left\|\dfrac{\partial \widetilde{\ell}_{t,T}(u_0,\boldsymbol{\theta}_0)}{\partial \theta}\right\|^2\right] < \infty$. Consequently, the Central Limit Theorem can be applied to the differences in martingale sequences.

$$\frac{1}{\sqrt{Tb}}\sum_{t=1}^{N}\frac{\partial \widetilde{\ell}_{t,T}(u,\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}K\left(\frac{\frac{t}{T}-u}{b}\right) \xrightarrow[T\to+\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma(\theta_0(u)))$$

$$\text{with } \Sigma(\theta_0(u)) = \int_{\mathbb{R}} K^2(x) \times \sum_{t\in\mathbb{Z}}\left(\mathrm{Cov}\left[\frac{\partial \widetilde{q}_{t,N}(u,\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i},\frac{\partial \widetilde{q}_{t,N}(u,\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_j}\right]\right)_{1\le i,j\le d}$$

Now we use a Taylor-Lagrange expansion for establishing

$$\begin{aligned}\frac{1}{\sqrt{Tb}}\sum_{t=1}^{T}\frac{\partial \widetilde{\ell}_{t,T}(u,\boldsymbol{\theta}_0)K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial \boldsymbol{\theta}} &= \frac{1}{\sqrt{Tb}}\sum_{t=1}^{T}\frac{\partial \widetilde{\ell}_{t,T}(u,\boldsymbol{\theta}_0)K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial \boldsymbol{\theta}} \\ &+\sqrt{Tb}\times\frac{1}{Tb}\sum_{t=1}^{T}\frac{\partial^2 \widetilde{\ell}_{t,T}(u,\bar{\boldsymbol{\theta}})K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial \boldsymbol{\theta}^2}\left(\widehat{\boldsymbol{\theta}}(u)-\boldsymbol{\theta}_0(u)\right),\end{aligned} \tag{3.27}$$

with $\bar{\boldsymbol{\theta}}$ in the segment bounded by $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$This leads to:

$$\begin{aligned}\left(\widehat{\boldsymbol{\theta}}(u)-\boldsymbol{\theta}_0(u)\right) &= \frac{\sum_{t=1}^T\frac{\partial \widetilde{\ell}_{t,T}(u,\widehat{\boldsymbol{\theta}})K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial \boldsymbol{\theta}} - \sum_{t=1}^T\frac{\partial \widetilde{\ell}_{t,T}(u,\boldsymbol{\theta_0})K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial \boldsymbol{\theta}}}{\sum_{t=1}^T\frac{\partial^2 \widetilde{\ell}_{t,T}(u,\bar{\boldsymbol{\theta}})K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial \boldsymbol{\theta}^2}} \\ &= \frac{-\sum_{t=1}^T\frac{\partial \widetilde{\ell}_{t,T}(u,\boldsymbol{\theta_0})K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial \boldsymbol{\theta}}}{\sum_{t=1}^T\frac{\partial^2 \widetilde{\ell}_{t,T}(u,\bar{\boldsymbol{\theta}})K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial \boldsymbol{\theta}^2}},\end{aligned}$$

Multiplying by $\sqrt{Tb}$ results in

$$\sqrt{Tb}\left(\widehat{\boldsymbol{\theta}}(u)-\boldsymbol{\theta}_0(u)\right)=-\sqrt{Tb}\frac{\sum_{t=1}^{T}\dfrac{\partial\widetilde{\ell}_{t,T}(u,\boldsymbol{\theta_0})K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial\boldsymbol{\theta}}}{\sum_{t=1}^{T}\dfrac{\partial^2\widetilde{\ell}_{t,T}(u,\bar{\boldsymbol{\theta}})K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial\boldsymbol{\theta}^2}}, \tag{3.28}$$

So by law of large numbers, the denominator of (3.28) deliver the following information

$$\frac{\partial^2\tilde{\mathcal{L}}_{t,T}(u,\theta)}{\partial\theta^2}=\frac{1}{Tb}\sum_{t=1}^{T}\frac{\partial^2\widetilde{\ell}_{t,T}(u,\boldsymbol{\theta})}{\partial\boldsymbol{\theta^2}}K\left(\frac{u-\frac{t}{T}}{b}\right)\underset{T\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}}\mathbb{E}\left[\frac{\partial^2\widetilde{\ell}_{t,T}(u,\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta^2}}\right]$$

Since $\bar{\theta}$ is in the segment bounded by $\hat{\theta}$ and θ_0 and by the consistency result of the previous section, we have $\widehat{\boldsymbol{\theta}}(u)\underset{T\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}}\boldsymbol{\theta_0}(u)$ and

$$\frac{1}{\sqrt{Tb}}\sum_{t=1}^{T}\left(\frac{\partial^2\widetilde{\ell}_{t,T}(u,\bar{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta^2}}\right)K\left(\frac{u-\frac{t}{T}}{b}\right)\underset{T\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}}\Gamma\left(\boldsymbol{\theta}_0(u)\right).$$

with

$$\Gamma\left(\boldsymbol{\theta}_0(u)\right)=\mathbb{E}\left[\frac{\partial^2\widetilde{\ell}_{t,T}(u,\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}^2}\right].$$

Applying the CLT to the numerator of (3.28) gives

$$\frac{1}{\sqrt{Tb}}\sum_{t=1}^{T}\left(\frac{\partial\widetilde{\ell}_{t,T}(u,\bar{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\right)K\left(\frac{u-\frac{t}{T}}{b}\right)\underset{T\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}}\mathcal{N}\left(0,\Sigma\left(\boldsymbol{\theta}_0(u)\right)\right).$$

Combining these results and using Slutsky's lemma, we obtain

$$-\sqrt{Tb}\frac{\sum_{t=1}^{T}\dfrac{\partial\widetilde{\ell}_{t,T}(u,\boldsymbol{\theta_0})K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial\boldsymbol{\theta}}}{\sum_{t=1}^{T}\dfrac{\partial^2\widetilde{\ell}_{t,T}(u,\bar{\boldsymbol{\theta}})K\left(\frac{u-\frac{t}{T}}{b}\right)}{\partial\boldsymbol{\theta}^2}}\Gamma\left(\boldsymbol{\theta}_0(u)\right)\underset{T\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}}\mathcal{N}\left(0,\Sigma\left(\boldsymbol{\theta}_0(u)\right)\right). \tag{3.29}$$

which implies from (3.29) that

$$\sqrt{Tb}\left(\hat{\boldsymbol{\theta}}(u)-\boldsymbol{\theta}_0(u)\right)\underset{T\to+\infty}{\mathcal{L}}\mathcal{N}\left(0,\Gamma^{-1}\left(\boldsymbol{\theta}_0(u)\right)\Sigma\left(\boldsymbol{\theta}_0(u)\right)\Gamma^{-1}\left(\boldsymbol{\theta}_0(u)\right)\right). \tag{3.30}$$

$\square$

3.6.4 Misspecification of the Distribution in Locally Stationary Time-Varying Models

The specification of the innovation distribution plays a central role in likelihood-based estimation of conditionally heteroskedastic models. In the classical stationary framework, it is well known that an incorrect specification of the distribution generally leads to inconsistency of the maximum likelihood estimator. This phenomenon is rigorously established in the seminal work of Straumann, and constitutes a fundamental limitation of likelihood-based inference under misspecification.

The purpose of this section is to revisit this issue in the context of locally stationary time-varying volatility models, with particular emphasis on the use of Laplace and Student's t distributions as quasi-likelihoods. Our analysis is motivated by the apparent tension between consistency results derived for time-varying models and the classical stationary theory. We show that, once properly interpreted, the two frameworks are fully compatible, and that the inconsistency result of Straumann continues to hold in an appropriate sense.

To fix ideas, consider a time-varying GARCH-type process defined in 3.16

In the stationary setting, likelihood-based estimation relies on maximizing the expected log-likelihood

$$E\big[\ell_t(\theta)\big], \qquad \ell_t(\theta) = \log\left[k\left(\frac{X_t}{\sigma_t(\theta)}\right)\right] - \log(\sigma_t(\theta))\,, \tag{3.31}$$

where k denotes the assumed innovation density. Consistency of the estimator requires that this criterion be uniquely maximized at the true parameter θ_0. A necessary condition for this property is that the expected score vanishes at θ_0, that is,

$$E[\partial_\theta \ell_t(\theta_0)] = 0. \tag{3.32}$$

Differentiating $\ell_t(\theta)$ with respect to θ yields

$$\partial_\theta \ell_t(\theta) = -\partial_\theta \sigma_t(\theta) \frac{1}{\sigma_t(\theta)^2}\left(\frac{X_t}{\sigma_t(\theta)} \frac{k'(X_t/\sigma_t(\theta))}{k(X_t/\sigma_t(\theta))} + 1\right). \tag{3.33}$$

Evaluated at the true parameter θ_0, where $X_t/\sigma_t(\theta_0) = Z_t$, the expectation of the score factorizes as

$$E[\partial_\theta \ell_t(\theta_0)] = E\left(-\frac{\partial_\theta \sigma_t(\theta_0)}{\sigma_t(\theta_0)^2}\right) E\left(Z_t \frac{k'(Z_t)}{k(Z_t)} + 1\right). \tag{3.34}$$

Consequently, the vanishing of the expected score requires

$$E\left(Z_t \frac{k'(Z_t)}{k(Z_t)}\right) = -1. \tag{3.35}$$

As shown in Theorems 6.2.1 and 6.2.3 of [37], this condition is highly restrictive. In particular, it holds for all standardized random variables if and only if the density k is Gaussian. Therefore, when a non-Gaussian density such as the Laplace or Student's t distribution is employed, the likelihood-based estimator is generally inconsistent for the true parameter θ_0 whenever the distribution is misspecified [17].

This result has a crucial implication, under strict stationarity, the use of Laplace or Student's t likelihoods does not merely affect efficiency, but fundamentally alters the probability limit of the estimator. In such cases, the estimator converges, if at all, to a parameter value different from the true one, thereby violating consistency. This phenomenon is not an artifact of a particular model specification, but a structural consequence of likelihood misspecification.

At first glance, this conclusion appears to contradict recent consistency results obtained for time-varying volatility models estimated by local quasi-maximum likelihood methods. However, this apparent contradiction stems from an implicit confusion between consistency with respect to the true parameter and convergence toward a pseudo-true parameter induced by the misspecified likelihood. Clarifying this distinction is essential for a correct interpretation of local QMLE results.

In the locally stationary framework considered in this thesis, estimation is based on kernel-weighted local quasi-likelihoods rather than on a single global likelihood function. As a result, the estimator is no longer tied to the stationary score condition (3.35). Nevertheless, the fundamental insight of Straumann remains valid under distributional misspecification, likelihood-based estimators are not consistent for the true parameter. The appropriate question is therefore not whether inconsistency disappears in the time-varying setting, but rather toward which parameter the estimator converges.

In locally stationary models, the local QMLE converges to a time-varying pseudo-true parameter curve, whose definition and asymptotic properties naturally extend the stationary misspecification theory to the time-varying case. Building on the stationary inconsistency result, we now consider the estimation of time-varying volatility models under local stationarity. Let $u \in (0,1)$ denote a fixed rescaled time point, and recall the local quasi-log-likelihood

$$\mathcal{L}_T(\theta, u) = \frac{1}{Tb} \sum_{t=1}^{T} K\Big(\frac{t/T - u}{b_T}\Big)\, \ell_t(\theta(u)), \tag{3.36}$$

where K is a kernel function, $b \to 0$ is a bandwidth sequence, and $\ell_t(\theta(u))$ is the quasi-log-likelihood function corresponding to either the Laplace or Student's t density. The kernel weighting ensures that observations further from u contribute less to the local criterion, effectively producing a localized stationary approximation.

In this framework, the local estimator

$$\hat{\theta}(u) = \arg\max_{\theta \in \Theta} \mathcal{L}_T(\theta, u) \tag{3.37}$$

converges, in probability, to a pseudo-true parameter curve $\theta^*(u)$, defined as the maximizer of the localized expected quasi-likelihood:

$$\theta^*(u) = \arg\max_{\theta \in \Theta} E\Big[K\Big(\frac{t/T - u}{b_T}\Big)\, \ell_t(\theta)\Big]. \tag{3.38}$$

Unlike the stationary case, where the expected score factorizes and inconsistency arises whenever

$$\mathbb{E}\bigg(Z_t \frac{k'(Z_t)}{k(Z_t)}\bigg) \neq -1,$$

the local score does not admit such a global decomposition due to the kernel weighting and the time-varying nature of $\sigma_t(\theta)$. Consequently, the vanishing of the expected global score is no longer a prerequisite for convergence. Instead, the pseudo-true parameter $\theta^*(u)$ captures the best local approximation of the true conditional scale under the chosen quasi-likelihood.

It follows that, under Laplace or Student's t quasi-likelihoods, the local estimator is generally inconsistent with respect to the true parameter $\theta_0(u)$ in the sense of Straumann. Mathematically,

$$\hat{\theta}(u) \xrightarrow{P} \theta^*(u) \neq \theta_0(u) \tag{3.39}$$

whenever the quasi-likelihood does not match the true innovation density. This aligns with the classical stationary theory, the estimator converges to a pseudo-true parameter rather than the actual parameter whenever distributional misspecification occurs.

Nevertheless, the concept of a pseudo-true parameter is constructive. The curve $\theta^*(u)$ provides a time-varying target that is internally consistent within the localized likelihood framework. The discrepancy between $\theta^*(u)$ and $\theta_0(u)$ quantifies the impact of misspecification, and its magnitude depends on both the choice of quasi-likelihood and the properties of the true innovations. Heavy-tailed quasi-likelihoods such as

Laplace or Student's t typically produce pseudo-true curves that are robust to extreme observations and transient deviations from Gaussianity, providing practical advantages in financial and economic time series.

In summary, local quasi-maximum likelihood estimation in time-varying models extends Straumann's stationary inconsistency results in a coherent manner. The estimator remains consistent for the pseudo-true curve $\theta^*(u)$, while distributional misspecification continues to imply inconsistency with respect to the actual parameter $\theta(u)$. This distinction is critical for correctly interpreting local QMLE results, the estimator's probability limit reflects the local maximization of the quasi-likelihood, not a universal global score condition. The next part addresses the asymptotic properties of $\hat{\theta}(u)$ around the pseudo-true curve, and the associated efficiency implications under misspecification.

Having established that the local QMLE converges to a pseudo-true parameter curve $\theta^*(u)$ under distributional misspecification, we now analyze its asymptotic behavior and efficiency properties in the time-varying setting. Let $\hat{\theta}(u)$ denote the local estimator at rescaled time $u \in (0,1)$. Standard arguments in kernel-based quasi-likelihood theory yield the asymptotic expansion

$$\sqrt{Tb}\,\big(\hat{\theta}(u) - \theta^*(u)\big) \implies \mathcal{N}\Big(0, \mathcal{V}^*(u)\Big), \tag{3.40}$$

where b is the bandwidth and $\mathcal{V}^*(u)$ is the asymptotic covariance matrix evaluated around the pseudo-true parameter. This result holds provided that the innovations possess finite second moments and that $\theta(u)$ is sufficiently smooth in u.

The covariance matrix $\mathcal{V}^*(u)$ adopts the classical sandwich form

$$\mathcal{V}^*(u) = \mathcal{I}^*(u)^{-1}\,\mathcal{J}^*(u)\,\mathcal{I}^*(u)^{-1}, \tag{3.41}$$

where $\mathcal{I}^*(u)$ is the local expected Hessian of the quasi-log-likelihood and $\mathcal{J}^*(u)$ is the variance of the local score evaluated at $\theta^*(u)$. In the special case of correctly specified Gaussian likelihood, $\mathcal{I}^*(u) = \mathcal{J}^*(u)$ and the estimator attains the Cramér–Rao lower bound. However, for Laplace or Student's t quasi-likelihoods, these matrices generally differ, leading to a loss of efficiency relative to the Gaussian benchmark.

It is important to emphasize that, while the estimator is generally inconsistent with respect to the true parameter curve $\theta_0(u)$, it is consistent for $\theta^*(u)$, which provides a coherent target for inference in locally stationary models. Mathematically, this implies

$$\hat{\theta}(u) - \theta(u) = \big(\hat{\theta}(u) - \theta^*(u)\big) + \big(\theta^*(u) - \theta(u)\big), \tag{3.42}$$

where the first term converges to zero at the $\sqrt{Tb}$ rate, while the second term represents the bias induced by distributional misspecification. The latter quantifies the deviation from the true parameter curve due to the choice of quasi-likelihood.

The heavy-tailed nature of Laplace and Student's t quasi-likelihoods ensures robustness to extreme observations. The influence function of the estimator is bounded for large standardized residuals in the Laplace case and moderated by the degrees of freedom in the Student's t case. Consequently, even under distributional misspecification, the estimator exhibits stable finite-sample performance in environments characterized by volatility clustering or abrupt changes, common in financial and economic time series.

From a practical perspective, the time-varying asymptotic variance $\mathcal{V}^*(u)$ reflects both the local information content of the data and the distributional properties of the innovations. Unlike stationary models, the inefficiency induced by misspecification is itself time-dependent, adapting to the changing local variance and the pseudo-true parameter. This highlights the flexibility of the local QMLE framework in accommodating slowly evolving tail behavior without explicitly modeling time-varying higher-order moments.

Finally, we note that the asymptotic results hinge on standard kernel regularity conditions and the shrinking bandwidth $b \to 0$. In this asymptotic regime, the local QMLE provides a mathematically rigorous and robust approach to inference in locally stationary volatility models, simultaneously respecting Straumann's inconsistency results in the stationary case and extending them to time-varying pseudo-true targets.

In conclusion, distributional misspecification in locally stationary time-varying volatility models primarily affects. Consistency with respect to the true parameter $\theta_0(u)$ (in line with Straumann), Asymptotic efficiency of the local estimator relative to the Gaussian benchmark. Nevertheless, the estimator remains consistent for the pseudo-true curve $\theta^*(u)$ and enjoys stable asymptotic normality. This provides a coherent framework for employing Laplace or Student's t quasi-likelihoods in time-varying GARCH models, reconciling classical stationary theory with modern locally stationary inference.

3.7 Conclusion

This chapter was devoted to the study of time-varying models within the framework of locally stationary processes. After introducing the general motivation for time-varying modeling, the concept of local stationarity was presented as a flexible and theoretically sound approach for capturing gradual structural changes in stochastic processes over time. This framework allows nonstationary processes to be approximated locally by stationary ones, thereby enabling the use of classical inference tools in a localized manner.

The chapter then focused on the estimation of time-varying parameters, with particular emphasis on kernel-based methods. Kernel-based maximum likelihood estimation was introduced as a powerful nonparametric technique for exploiting local information around a fixed rescaled time point. Both kernel-based estimation and localized quasi-maximum likelihood estimation (QMLE) were discussed, highlighting their suitability for handling time-dependent dynamics while preserving desirable asymptotic properties.

Subsequently, likelihood-based inference was developed for locally stationary autoregressive processes. Time-varying AR(p) models were examined, and QMLE procedures were presented along with their asymptotic behavior, providing insight into the consistency and limiting distribution of the estimators. These results illustrate how classical time series models can be extended to accommodate temporal evolution in their parameters.

The final part of the chapter addressed locally stationary GARCH-type models, which are particularly relevant for modeling time-varying volatility. Estimation procedures based on QMLE were developed for time-varying GARCH processes under different distributional assumptions, including Student-t and Laplace innovations. Special attention was given to the impact of distributional misspecification, emphasizing its consequences on inference and motivating the use of robust estimation techniques.

Overall, this chapter establishes a comprehensive likelihood-based framework for modeling and estimating time-varying processes. The theoretical results and estimation methodologies developed here form a crucial foundation for the subsequent chapters, where more refined models and inferential procedures are further analyzed and applied.

Chapter 4

Numerical experiments

4.1 Introduction

This section is devoted to assessing the empirical performance of the proposed kernel-type likelihood estimator for locally stationary processes. We carry out Monte Carlo simulations, and real financial data applications to evaluate the method under controlled and practical conditions.

The Monte Carlo study is designed to shed light on how the estimator behaves in finite samples. We simulate time series from known time-varying models and examine how well the method recovers the underlying parameter trajectories. To understand the role of sample size, we repeat the experiments over different lengths T, as increasing T tends to improve local approximation and reduce variability in estimation. Performance metrics such as bias, variance, and root mean squared error (RMSE) are reported.

A critical component in kernel-based inference is the bandwidth b, which governs the degree of local smoothing in time. In our study, we do not rely on data-driven methods such as cross-validation or plug-in rules. Instead, we adopt a theoretically grounded choice inspired by the work of Bardet, Doukhan, and Wintenberger, who proposed a deterministic bandwidth of the form

$$b = T^{-\lambda},$$

where T is the sample size and $\lambda \in (0,1)$ is a tuning exponent. Following their recommendation (inducing $Tb^3 \xrightarrow[T\to\infty]{} 0$, which is the consistency and the asymptotic normality condition required for $\delta = 1$-Hölder function $\boldsymbol{\theta}_0$), we set $\lambda = 0.35$, which provides a balanced trade-off between bias and variance in locally stationary settings. This choice is particularly well-suited for models exhibiting mild smoothness in time variation and avoids the variability often introduced by fully data-driven bandwidth selection.

Another important aspect of our method is the specification of the contrast function, which depends on the assumed or empirically estimated distribution of the innovations ε_t. For example, when the errors are assumed Gaussian, the contrast reduces to the usual negative log-likelihood:

$$\rho(x,\theta) = \frac{1}{2\sigma^2(\theta)}\left(x-\mu(\theta)\right)^2 + \frac{1}{2}\log\left(2\pi\sigma^2(\theta)\right),$$

where $\mu(\theta)$ and $\sigma^2(\theta)$ denote time-varying conditional mean and variance functions, respectively.

Whereas for Laplace-distributed errors, a robust alternative is the absolute deviation contrast:

$$\rho(x,\theta) = \frac{1}{\beta(\theta)}|x-\mu(\theta)| + \log(2\beta(\theta)).$$

In addition, we consider the Student-t distribution, which is particularly suitable for capturing extreme values and heavier tails. The corresponding contrast function for a Student-t distribution with degrees of freedom $\nu > 2$ is given by:

$$\rho(x,\theta) = \frac{\nu+1}{2}\log\left(1+\frac{1}{\nu}\left(\frac{x-\mu(\theta)}{\sigma(\theta)}\right)^2\right) + \log\left(\sigma(\theta)\sqrt{\nu\pi}\,\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right),$$

which reduces to the Gaussian contrast as $\nu \to \infty$.

We test all three settings in our simulations and demonstrate how the choice of contrast affects estimation accuracy—especially in the presence of outliers or heavy-tailed noise. The Student-t contrast provides a flexible middle ground between Gaussian efficiency and Laplace robustness.

Finally, we compare our locally adaptive estimator with conventional stationary methods, which assume constant parameters throughout the sample. These comparisons clearly illustrate the gain in flexibility and responsiveness to structural change offered by our approach. In real data applications, such as financial returns and environmental time series, we find that the locally stationary method captures evolving dynamics that stationary models fail to reflect.

Overall, the numerical experiments support the theoretical properties of the estimator and provide practical insights into bandwidth selection, contrast specification, and sample size effects.

4.2 Monte Carlo simulations

4.2.1 Time-varying AR(1)-processes(tvAR(1))

We simulate a time-varying AR(1) process of the form

$$X_{t,T} = a_1(u)X_{t-1,T} + \xi_t,$$

the autoregressive time-varying coefficient

$$a_1(u) = 0.4 + 0.3\sin(\log(u)).$$

The innovations $\{\xi_t\}$ are consists of i.i.d. random variables, with $\mathbb{E}(\xi_t) = 0$.

We use sample sizes for T = 1000, 5000, and 10000. Figure 4.1 illustrates one trajectory of the simulated process. Subsequently, we estimate the time-varying coefficient $a_1(u)$ and compare the estimated function $\hat{a}_1(u)$ with the true curve. Figure 4.2 displays the path of the true function (blue) alongside its estimate (red). Table 4.1 reports the square root of the MISE for Laplace-based maximum likelihood estimation compared with Gaussian-based estimation.

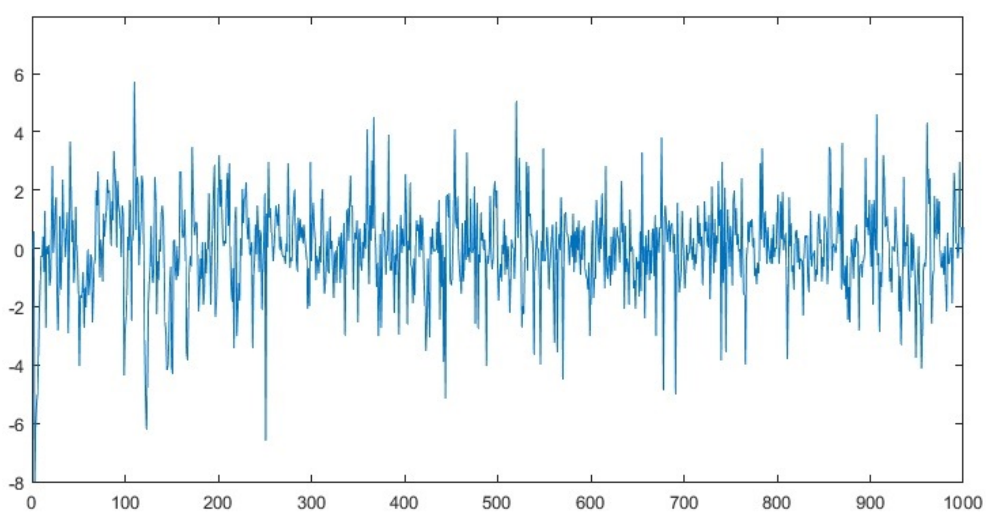


Figure 4.1: Trajectory of a simulated tvAR(1) process with Laplace-distributed noise ($T = 5000$).

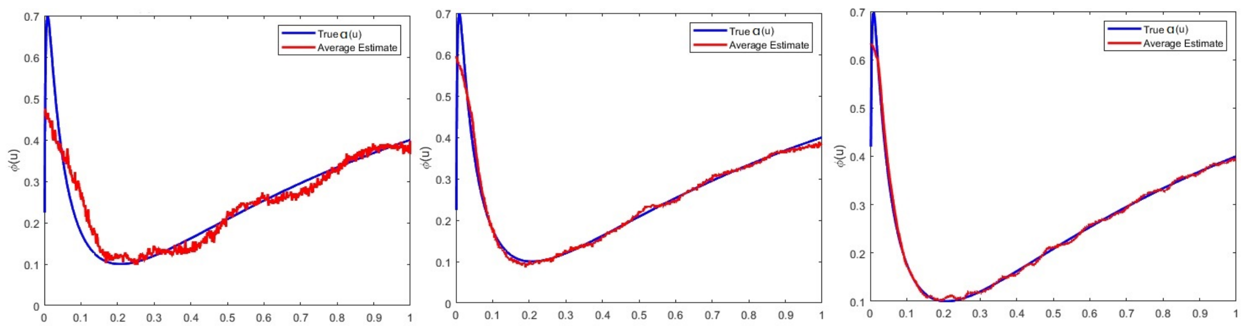


Figure 4.2: True function $a_1(u)$ (blue) and its estimate $\hat{a}_1(u)$ (red) for $T = 1000, 5000$, and 10000.

Table 4.1: Square root of the MISE for the tvAR(1) process under Laplace QMLE and Gaussian QMLE, for $T = 1000$, 5000, and 10000.

T	Laplace	$\mathcal{N}$
1000	0.07303	0.079273
5000	0.05062	0.054353
10000	0.04442	0.047262

4.2.2 tvARMA(1,1) with tvARCH(1) errors

We simulate the process defined by the general tvARMA(p,q)–tvARCH(r) model

$$X_{t,T} = \sum_{k=1}^{p} a_k\left(\frac{t}{T}\right) X_{t-k,T} + \sum_{\ell=1}^{q} b_\ell\left(\frac{t}{T}\right) \varepsilon_{t-\ell,T} + \varepsilon_{t,T}, \tag{4.1}$$

$$\varepsilon_{t,T} = \sigma_{t,T}\xi_t, \tag{4.2}$$

$$\sigma_{t,T}^2 = c_0\left(\frac{t}{T}\right) + \sum_{m=1}^{r} c_m\left(\frac{t}{T}\right) \varepsilon_{t-m,T}^2, \tag{4.3}$$

where $\{\xi_t\}$ is an i.i.d. sequence with $\mathbb{E}[\xi_t] = 0$ and $\mathbb{E}[\xi_t^2] = 1$. with $p = q = 1$. The true parameters of the tvARMA(1,1) part are

$$a_1(u) = 0.5 + 0.3\sin(2u), \quad b_1(u) = 0.4 + 0.5\cos(2u),$$

and for $m = 1$ the tvARCH(1) part:

$$c_0(u) = 0.5 + 0.3\sin(2\pi u), \quad c_1(u) = 0.4 + 0.3u.$$

To analyze the behavior of our estimator, we generate datasets of size $T = 1000, 5000$, and 10000 in

MATLAB using the parameters above. Figure 4.3 illustrates one trajectory of the simulated process. We then apply our estimator to these datasets to evaluate its performance under the given time-varying conditions.

Figure 4.4 displays the paths of the true functions (a_1, b_1, c_0, c_1) (in blue) alongside their estimated counterparts $(\hat{a}_1, \hat{b}_1, \hat{c}_0, \hat{c}_1)$ (in red) for sample sizes $T = 1000, 5000$, and 10000. Table 4.2 reports the square root of the mean integrated squared error (MISE) between the estimated and true parameters, comparing maximum likelihood estimation based on the Laplace distribution with that based on the Gaussian distribution.

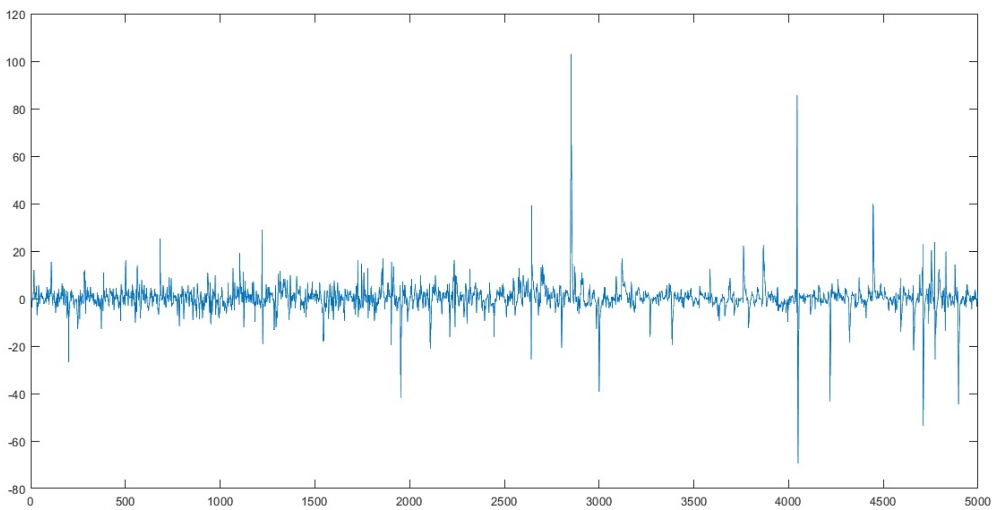


Figure 4.3: Trajectory of a simulated tvARMA(1,1)–tvARCH(1) process with Laplace errors ($T = 5000$).

Table 4.2: Square root of the MISE for tvARMA(1,1)-tvARCH(1) processes with local Laplace QMLE and local Gaussian QMLE, for $T = 1000$, 5000, and 10000.

	$\widehat{a}_1$		$\widehat{b}_1$		$\widehat{c}_0$		$\widehat{c}_1$	
T	Laplace	$\mathcal{N}$	Laplace	$\mathcal{N}$	Laplace	$\mathcal{N}$	Laplace	$\mathcal{N}$
1000	0.0374	0.0717	0.0470	0.0665	0.0944	0.1035	0.1700	0.1895
5000	0.0191	0.0255	0.0233	0.0303	0.0544	0.0582	0.0885	0.0984
10000	0.0140	0.0186	0.0181	0.0271	0.0431	0.0463	0.0644	0.0714

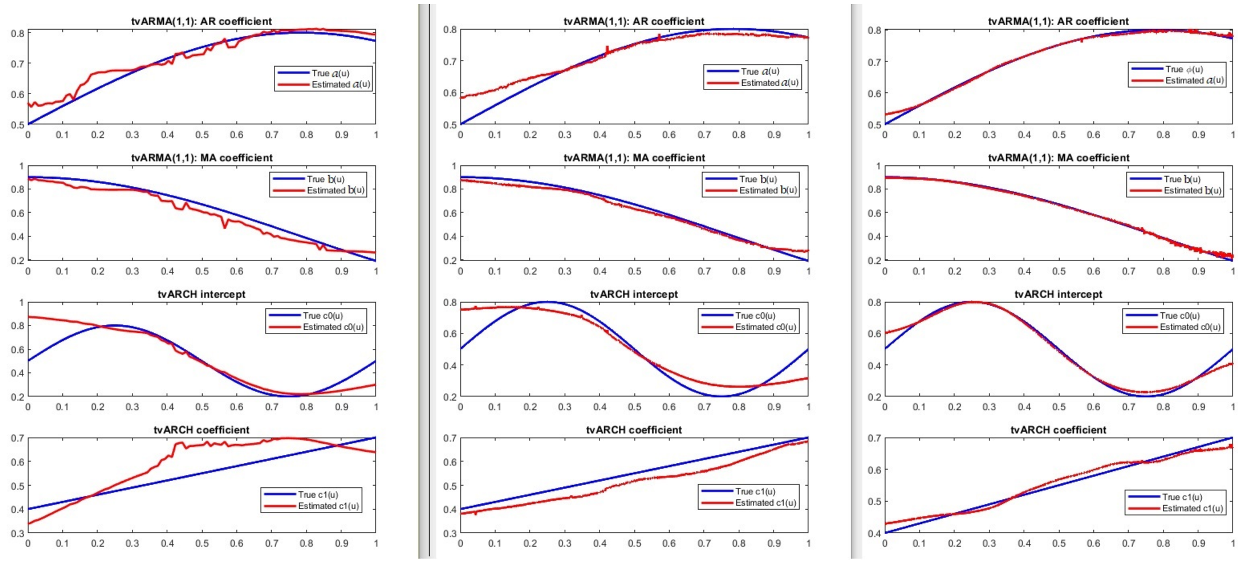


Figure 4.4: True functions (a_1, b_1, c_0, c_1) (blue) and their estimates $(\hat{a}_1, \hat{b}_1, \hat{c}_0, \hat{c}_1)$ (red) for $T = 1000, 5000$, and 10000.

4.2.3 tvGARCH(1,1) with heavy tails

Assuming that the true time-varying parameters are

$$\omega(\frac{t}{T}) = 1 + 0.5 \sin\left(5\frac{t}{T}\right), \quad \alpha(\frac{t}{T}) = 0.1 + 0.4 \cos^2\left(4\frac{t}{T}\right), \quad \beta(\frac{t}{T}) = 0.1 + 0.4\frac{t}{T},$$

for all $1 \le t \le T$. As a consequence, $\omega(u) = 1 + 0.5 \sin(5u)$, $\alpha_1(u) = 0.1 + 0.4 \cos^2(4u)$, and $\beta_1(u) = 0.1 + 0.4u$. Moreover. Independent realizations of the process were generated for sample sizes $T = 1000$, 5000, and 10000. For each case, we computed the Student and Laplace QMLEs, yielding the estimators $\widehat{\omega}(u)$, $\widehat{\alpha}_1(u)$, and $\widehat{\beta}_1(u)$, which were then compared with the Gaussian QMLE. Finally, we use the Gaussian kernel $G(x) = \frac{1}{\sqrt{\pi}} e^{-x^2/2} \mathbb{I}_{x \in [-1,1]}$ and the Uniform kernel $U(x) = \frac{1}{2} \mathbb{I}_{x \in [-1,1]}$, denoted respectively by $\widehat{\theta}^G(u)$ and $\widehat{\theta}^U(u)$.

Figure 4.6 presents the results of the simulations described above, we calculated the root square of the mean integrated squared error (RSMISE) for the Student's QMLE, compared with the Gaussian QMLE in Table 4.3. Figure 4.5 illustrates an example of generated data trajectories. for $T = 1000$

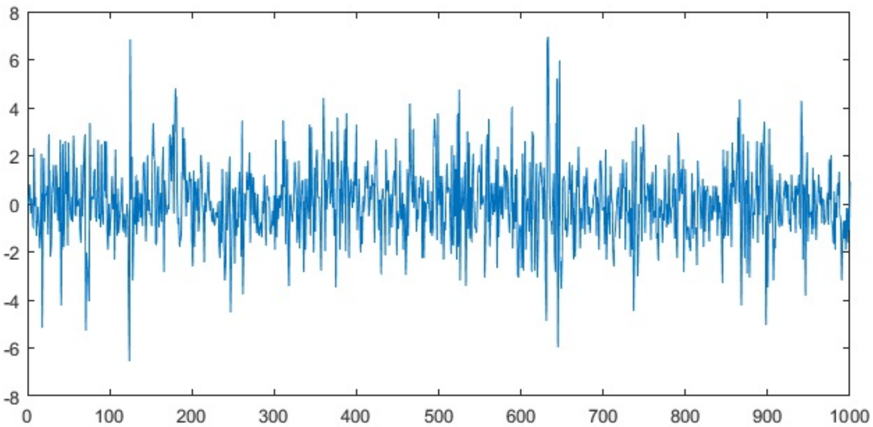


Figure 4.5: trajectory of a Time-varying GARCH(1,1) process for T = 1000

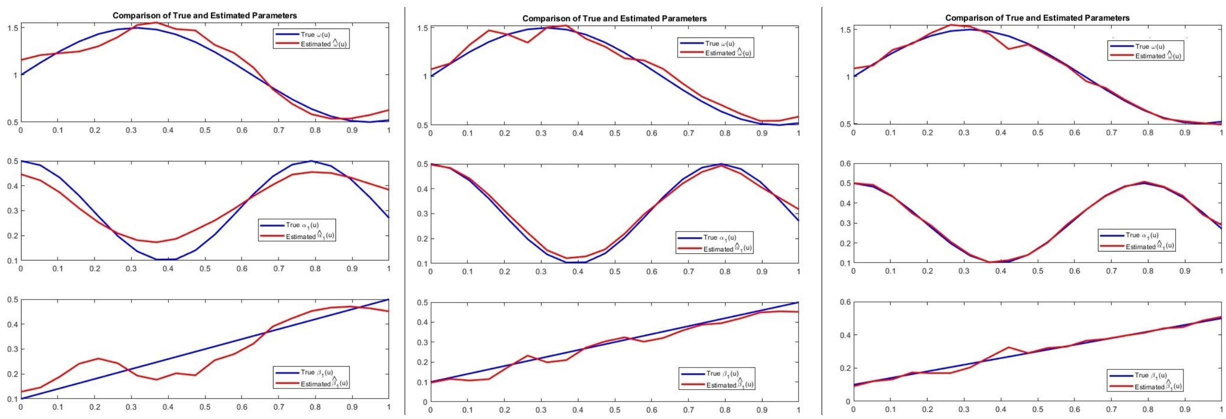


Figure 4.6: paths of functions ω, α and β (in blue), and paths of estimated functions $\hat{\omega}, \hat{\alpha}$ and $\hat{\beta}$ (in red) for T = 1000, T=5000 and T = 10000 respectively

G(x)	$\hat{\omega}^G$			$\hat{\alpha}^G$			$\hat{\beta}^G$		
T	t	L	$\mathcal{N}$	t	L	$\mathcal{N}$	t	L	$\mathcal{N}$
1000	0.0615	0.218	0.455	0.0316	0.074	0.122	0.0052	0.102	0.208
5000	0.0094	0.146	0.323	0.0057	0.039	0.077	9.8466e-04	0.071	0.146
10000	1.1397e-04	0.112	0.224	0.0022	0.025	0.048	6.6498e-04	0.051	0.101
U(x)	$\hat{\omega}^U$			$\hat{\alpha}^U$			$\hat{\beta}^U$		
T	t	L	$\mathcal{N}$	t	L	$\mathcal{N}$	t	L	$\mathcal{N}$
1000	0.2704	0.308	0.3449	0.1181	0.198	0.2778	0.1914	0.221	0.2495
5000	0.1755	0.212	0.2494	0.1009	0.102	0.1035	0.1232	0.170	0.2165
10000	1.3321e-04	0.125	0.2501	0.0042	0.019	0.033	7.415e-04	0.0085	0.0164

Table 4.3: RSMISE for tvGARCH(1,1) processes using Student-t, Laplace and Gaussian QMLE for $T = 1000, 5000$, and 10000.

4.2.4 tvGJR-GARCH(1,1) with Laplace-distributed noise

We simulate a time-varying GJR-GARCH(1,1) process of the form

$$\begin{cases} X_{t,T} = \sigma_{t,T} Z_t, \\ \sigma^{\delta}_{t,T} = \omega\left(\frac{t}{T}\right) + \alpha_1\left(\frac{t}{T}\right)\left(|X_{t-1,T}| - \gamma_1\left(\frac{t}{T}\right) X_{t-1,T}\right)^{\delta} + \beta_1\left(\frac{t}{T}\right)\sigma^{\delta}_{t-1,T}, \end{cases}$$

where (Z_t) is an i.i.d. sequence with zero mean and unit variance. We take the time-varying parameter functions

$$\omega(u) = 0.4 + 0.2\sin(2u), \quad \alpha_1(u) = 0.3 + 0.2u, \quad \beta_1(u) = 0.4 + 0.3\cos^2(3u), \quad \gamma_1(u) = 0.2 + 0.1u^2,$$

with $\delta = 2$.

For comparability, we use the same sample sizes as in the previous examples. Figure 4.7 illustrates one trajectory of the simulated process.

We then estimate the functions $(c_0, c_1, \beta, \gamma)$ under Laplace QMLE and compare them with their true counterparts. Figure 4.10 presents the true paths (blue) against their estimates (red). Table 4.4 reports the square root of the mean integrated squared error (MISE) for Laplace-based maximum likelihood estimation compared with Gaussian-based estimation.

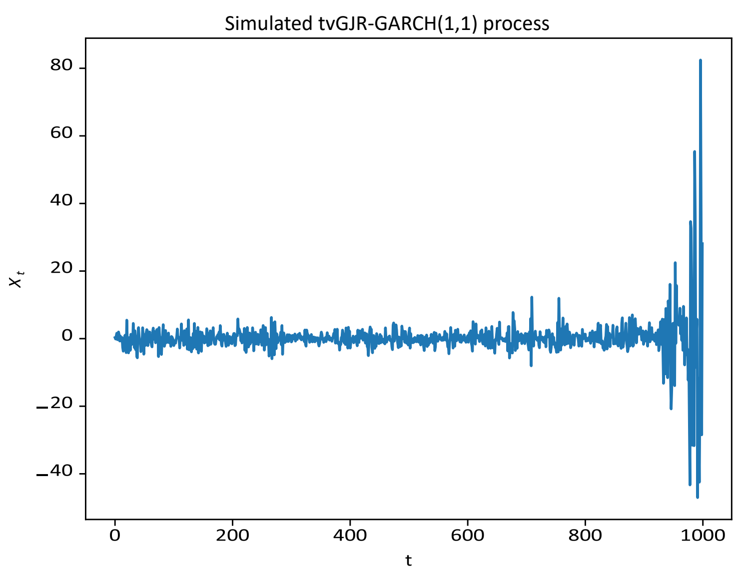


Figure 4.7: Trajectory of a simulated tvGJR-GARCH(1,1) process with Laplace-distributed noise.

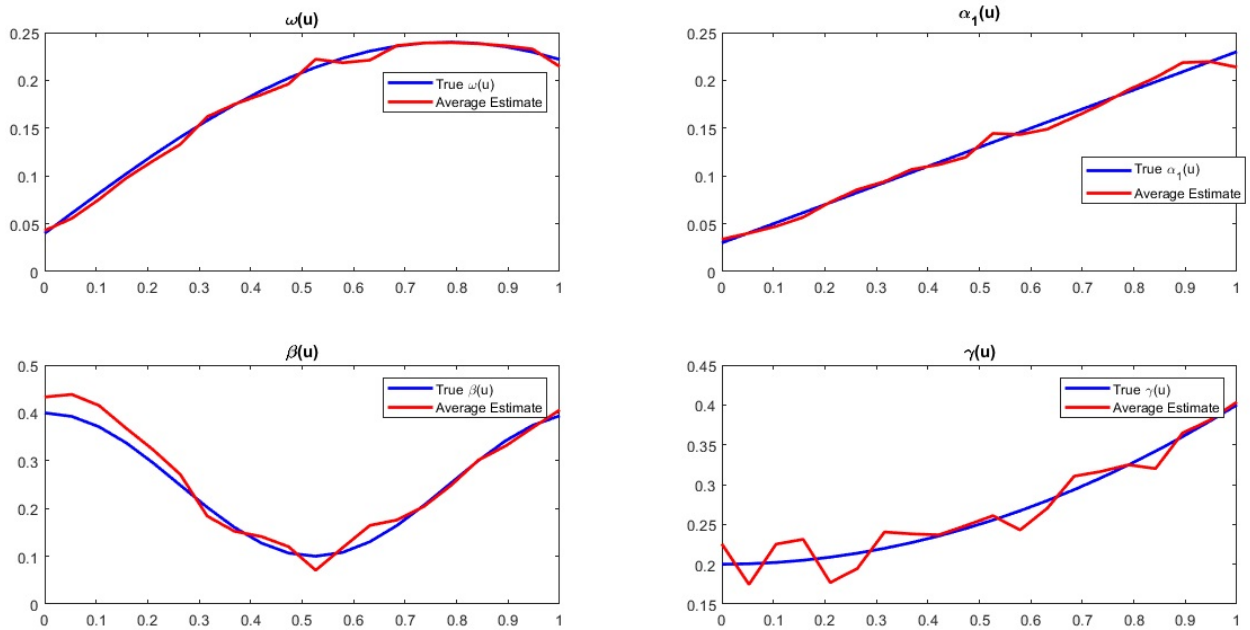


Figure 4.8: True functions $(c_0, c_1, \beta, \gamma)$ (blue) and their estimates $(\hat{c}_0, \hat{c}_1, \hat{\beta}, \hat{\gamma})$ (red).

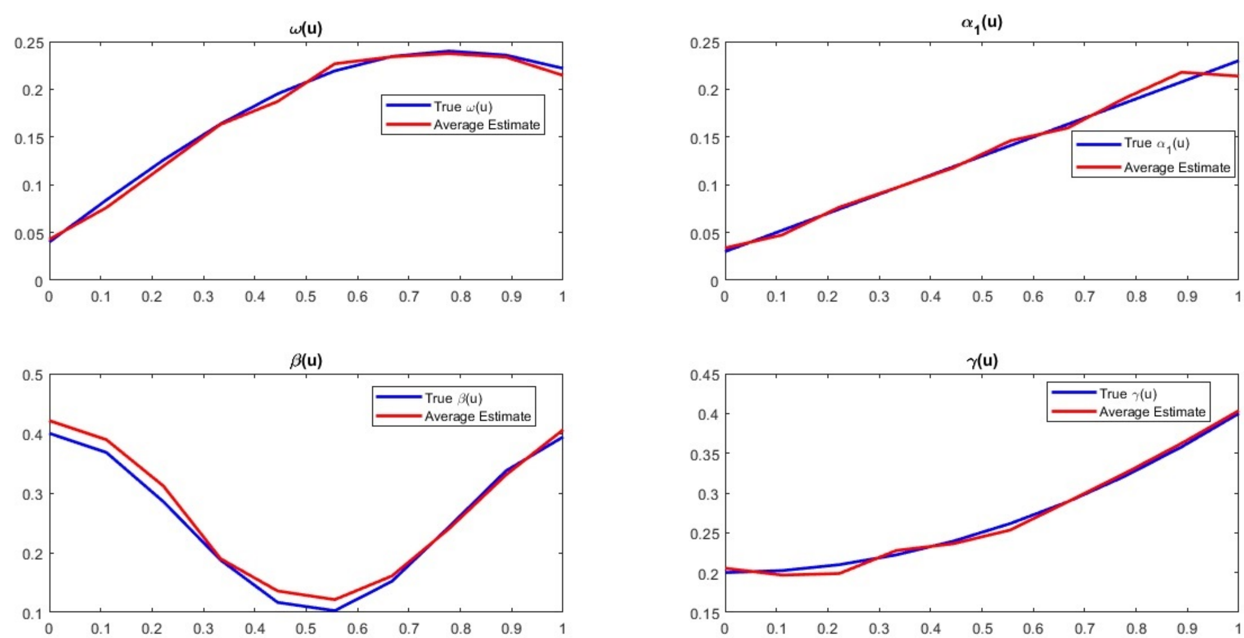


Figure 4.9: True functions $(c_0, c_1, \beta, \gamma)$ (blue) and their estimates $(\hat{c}_0, \hat{c}_1, \hat{\beta}, \hat{\gamma})$ (red).

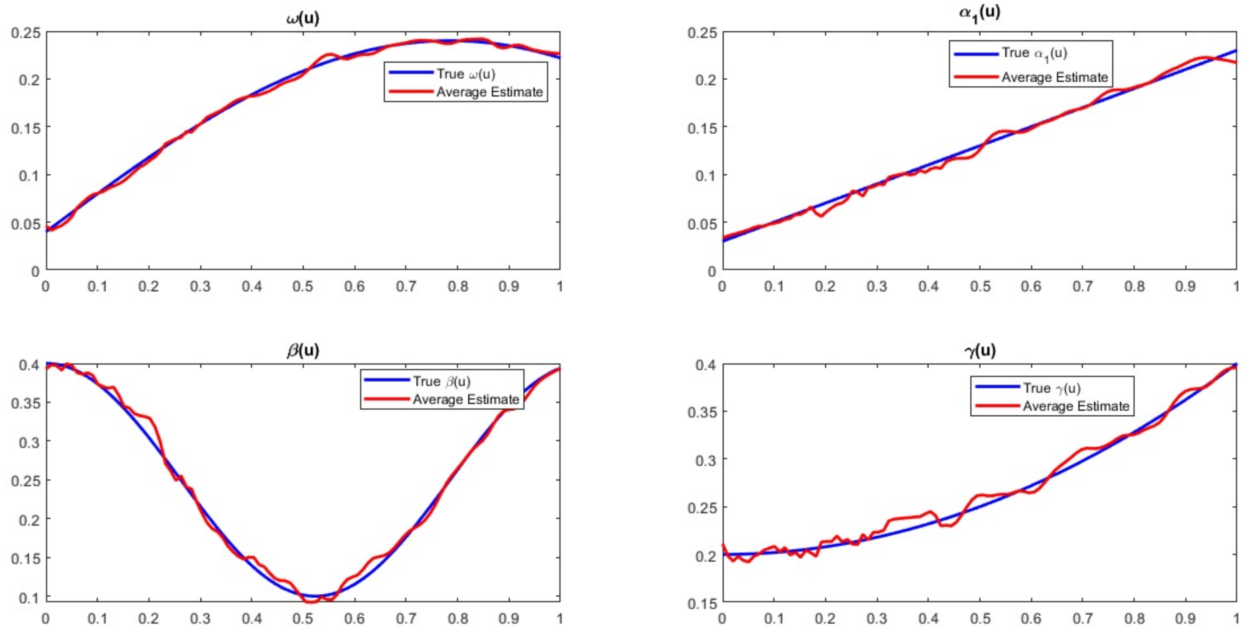


Figure 4.10: True functions $(c_0, c_1, \beta, \gamma)$ (blue) and their estimates $(\hat{c}_0, \hat{c}_1, \hat{\beta}, \hat{\gamma})$ (red).

Table 4.4: Square root of the MISE for the tvGJR-GARCH(1,1) process under Laplace QMLE and Gaussian QMLE.

	$\widehat{c_0}$		$\widehat{c_1}$		$\widehat{\beta}$		$\widehat{\gamma}$	
T	Laplace	$\mathcal{N}$	Laplace	$\mathcal{N}$	Laplace	$\mathcal{N}$	Laplace	$\mathcal{N}$
1000	0.201	0.288	0.145	0.213	0.176	0.249	0.092	0.157
5000	0.118	0.205	0.083	0.141	0.097	0.183	0.054	0.102
10000	0.079	0.166	0.057	0.119	0.071	0.152	0.036	0.084

4.2.5 Misspecified tvGARCH(1,1)

Assuming that the true time-varying parameters are

$$\omega\left(\frac{t}{T}\right) = 1 + 0.5\sin\left(5\frac{t}{T}\right), \quad \alpha\left(\frac{t}{T}\right) = 0.1 + 0.4\cos^2\left(4\frac{t}{T}\right), \quad \beta\left(\frac{t}{T}\right) = 0.1 + 0.4\frac{t}{T},$$

With Z_t follows a light-tailed distribution, we apply our approach to the generated data (example for T = 1000 in Figure 4.11) and obtain the results shown in Figure 4.12, and compare with gaussian QMLE in table 4.5.

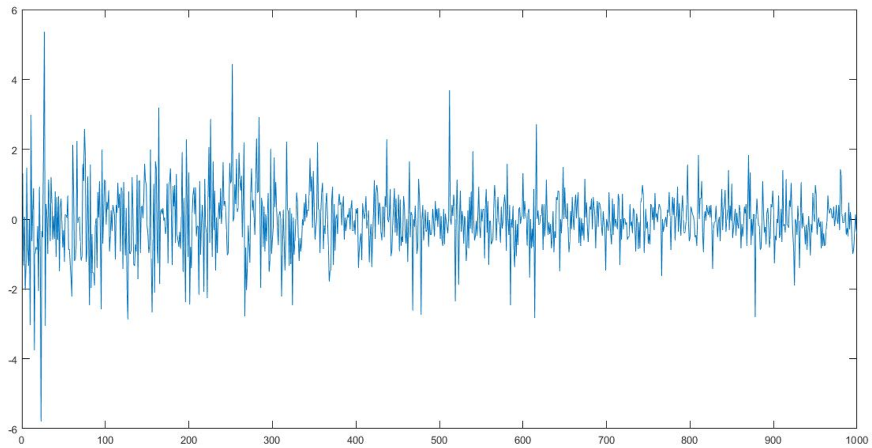


Figure 4.11: Simulated tvGARCH(1,1) process with T = 1000

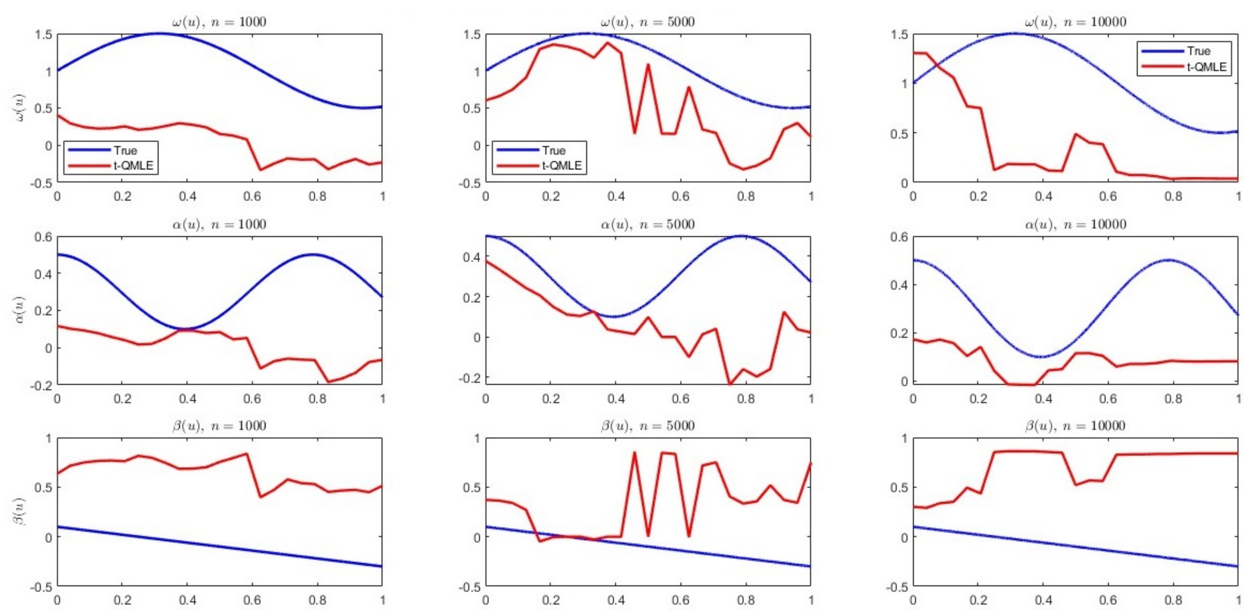


Figure 4.12: path of functions $\omega, \alpha_1, \alpha_2$ and β and paths of estimated values of $\hat{\omega}, \hat{\alpha_1}, \hat{\alpha_2}$ and $\hat{\beta}$ for T = 1000, T = 5000 and T = 10000 respectively

G(x)	$\hat{\omega}^G$		$\hat{\alpha}^G$		$\hat{\beta}^G$	
T	t	$\mathcal{N}$	t	$\mathcal{N}$	t	$\mathcal{N}$
1000	7.956e-01	7.449e-01	4.599e-01	4.569e-01	6.009e-01	5.198e-01
5000	8.891e-01	7.776e-01	2.980e-01	2.309e-01	8.571e-01	7.208e-01
10000	3.980e-01	2.615e-01	3.026e-01	2.465e-01	4.032e-01	2.258e-01
U(x)	$\hat{\omega}^U$		$\hat{\alpha}^U$		$\hat{\beta}^U$	
T	t	$\mathcal{N}$	t	$\mathcal{N}$	t	$\mathcal{N}$
1000	9.456e-01	6.152e-01	4.026e-01	3.445e-01	6.379e-01	6.083e-01
5000	5.265e-01	3.579e-01	3.494e-01	1.919e-01	4.633e-01	4.013e-01
10000	6.199e-01	4.568e-01	3.068e-01	2.199e-01	4.071e-01	3.434e-01

Table 4.5: RSMISE of tvGARCH$(1,1)$ parameter estimates using Student-t QMLE, compared with Gaussian QMLE estimations, using Gaussian and Uniform kernels. for the misspecification case

4.3 Application to real-world data and comparison

4.3.1 Application to real-world data

This section examines the China Securities Index 800 (CSI 800) over the period from September 02, 2015 to September 02, 2025, with 2428 daily observations. The log-returns series was obtained by differencing the logarithm of the closing prices (the data path is shown in Figure 4.13).

As shown in Figure 4.14, the mean of the log-returns is close to zero. At the 5% significance level, the Jarque–Bera statistic is greater than the critical value (see Table 4.6), indicating that the series does not follow a normal distribution. This suggests that the distribution of returns exhibits a “fat tail” feature.

Table 4.6: Normality tests for log-returns

Test	Statistic	Prob.	Decision
Jarque–Bera	4488.3873	0.0010	Reject H_0
Anderson–Darling	32.9623	0.0005	Reject H_0
Kolmogorov–Smirnov	0.0819	0.0000	Reject H_0

The presence of heteroskedasticity is checked using the ARCH–LM test. As reported in Table 4.7, $n \times R^2 = 205.42 > \chi^2_{0.1}(1)$, which confirms that the series has conditional heteroskedasticity.

As shown in Table 4.7, the ARCH-LM test gives an F-statistic of 18.61 with a p-value $< 4.23 \times 10^{-39}$,

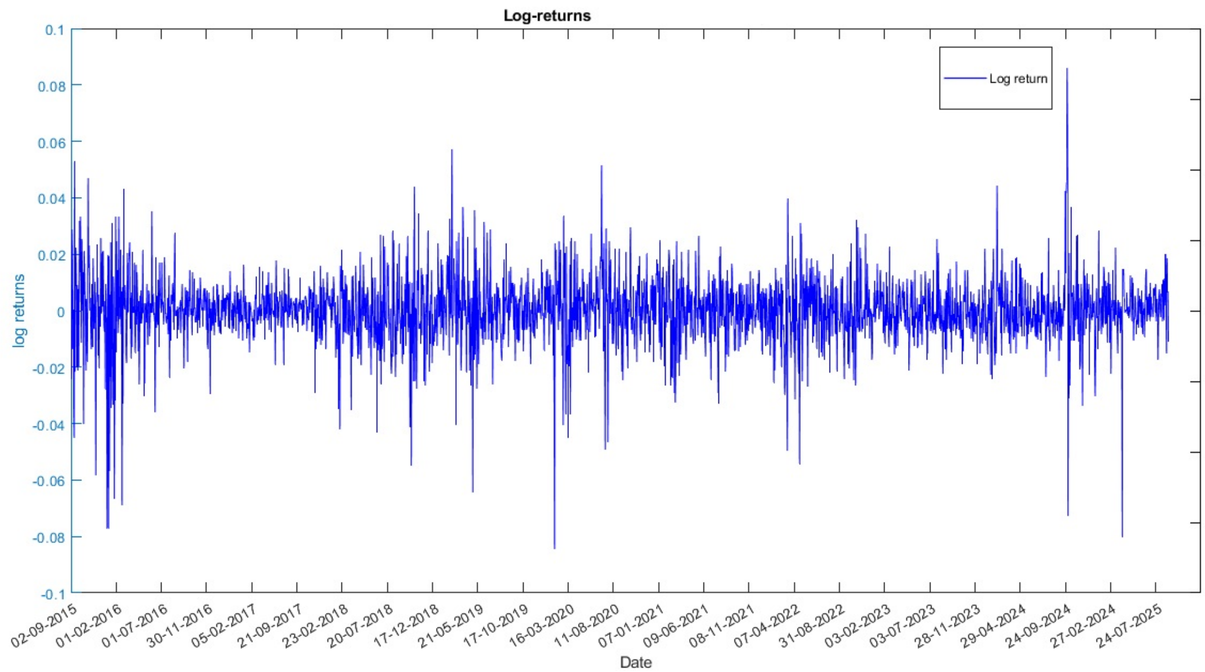


Figure 4.13: Path of log-returns of the CSI800 index closing prices

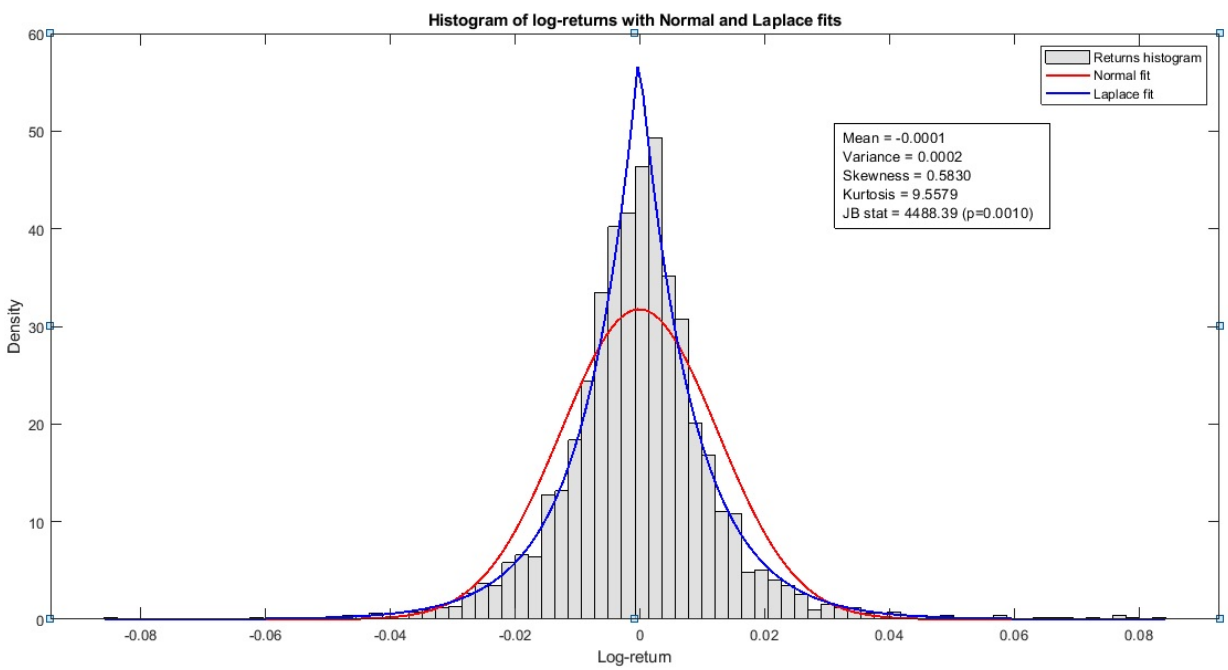


Figure 4.14: Histogram (CSI800 index)

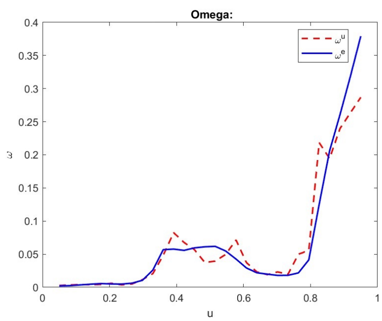
and an Obs*R-squared statistic of 205.42 with a p-value $< 2.48 \times 10^{-37}$. Both tests provide strong evidence against the null hypothesis of no ARCH effects, confirming that the log return series exhibits conditional

Table 4.7: ARCH–LM test forlog returns

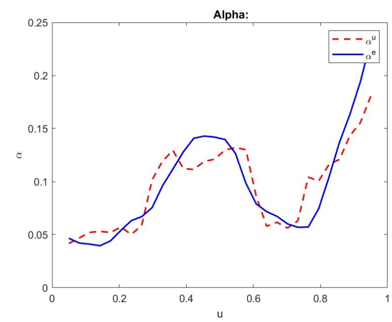
	Statistic	p-value	Decision (5% level)
Obs*R^2	205.42	2.48×10^{-37}	Reject H_0
F-statistic	18.61	4.23×10^{-39}	Reject H_0

heteroskedasticity.

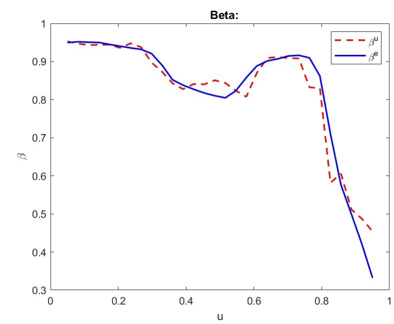
Given the presence of volatility clustering, the GARCH(1,1) model is specified for log returns of closing prices. Assuming the innovations Z_t follow the Laplace distribution, the parameter curves was estimated using Laplacian QMLE with Epaneshnikov and Uniform kernels in MATLAB, denoted θ^E and θ^U respectively.



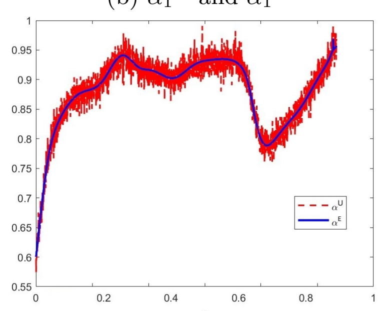
(a) $\hat{\omega}^E$ and $\hat{\omega}^U$



(b) $\hat{\alpha_1}^E$ and $\hat{\alpha_1}^U$



(c) $\hat{\beta}^E$ and $\hat{\beta}^U$



(d) $\hat{\alpha}^E$ and $\hat{\beta}^U$

Figure 4.15: Estimators of ω, α_1 and β_1 for the log-ret of closing prices of the CSI800 index from September 2015 to September 2025.

For comparison, the estimation was also performed under the Gaussian QMLE. Table 4.8 reports the biases, sample standard deviations (SD), and asymptotic standard deviations (AD) of the estimated parameters under both assumptions. The results show that when $Z_t \sim$ Laplace, the bias, SD, and AD are consistently smaller than those obtained under centred Gaussian assumption. This indicates that the Laplace distribution provides a better fit to the CSI 800 returns than the Gaussian distribution.

Therefore, the Laplace distribution provides a more suitable specification than the Gaussian distribution

Table 4.8: Comparison of parameter estimation accuracy: Gaussian vs Laplace QMLE

Innovation	Bias	SD	AD
$N(0,1)$	0.015	0.082	0.076
Laplace(1,1)	0.008	0.061	0.053

for capturing the heavy-tailed behavior commonly observed in financial return series, such as those of the CSI 800 index. This observation supports the adoption of a Laplace-based QMLE within the tvGARCH(1, 1) framework for volatility modeling.

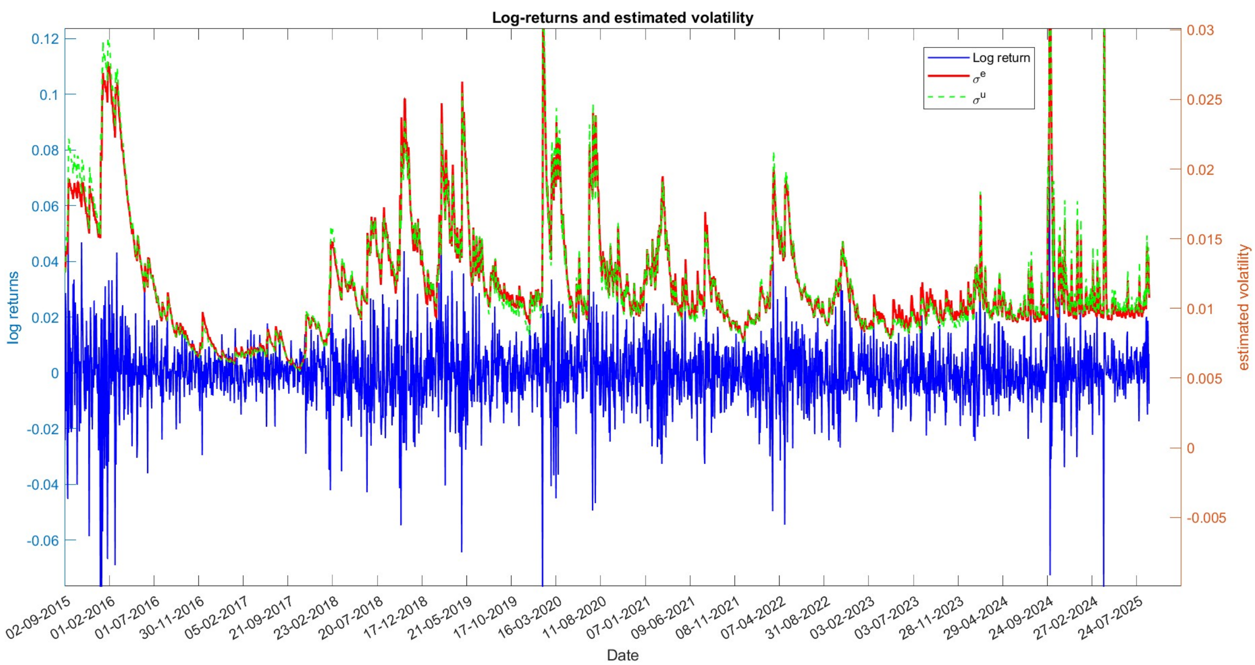


Figure 4.16: Log returns with estimated volatility using 3 kernels.

4.3.2 Compare with classical approach

For the class of time-varying processes under consideration, the Akaike information criterion (AIC) is defined in [35] as follows:

$$\mathrm{AIC} = -2\ell(\hat{\theta}(u)) + 2(1 + p + q) \tag{4.4}$$

and the Bayesian information criterion (BIC) is defined in [32] as:

$$\mathrm{BIC} = -2\ell(\hat{\theta}(u)) + (1 + p + q)\log(T), \tag{4.5}$$

where $\ell(\hat{\theta}(u))$ represents the log-likelihood evaluated at $\hat{\theta}(u)$.We apply this method to compare the previously discussed models with classical estimations on the same dataset. Lower AIC and BIC values indicate a better model fit [5]. Tables (4.9 and 4.10) below shows the comparison. The uniform kernel and Gaussian kernel are used for estimating time-varying models.

Table 4.9: Comparison of AIC values across different simulated processes for $T = 1000$.

	Uniform Kernel	Gaussian Kernel	Classical model
tvARMA(1,1)-tvARCH	44.94	44.92	101821157453.37
tvAR(1)	7.74	7.67	3515.33
tvGJR-GARCH(1,1)	14.10	11.33	737080015.50

Table 4.10: Comparison of BIC values across different simulated processes for $T = 1000$.

	Uniform Kernel	Gaussian Kernel	Classical model
tvARMA(1,1)-tvARCH	54.76	54.73	101821157463.19
tvAR(1)	17.56	17.48	3525.14
tvGJR-GARCH(1,1)	21.15	23.91	737080025.32

4.4 Interpretation of Results

In order to assess the practical performance of the proposed estimators, we applied them to several time-varying models under different simulation settings. The numerical results consistently indicate that the estimator exhibits good consistency properties across the considered scenarios. In particular, the accuracy of the estimation improves as the sample size increases, confirming the asymptotic behavior established in the theoretical results.

Moreover, the simulations highlight the impact of the assumed innovation distribution in the presence of heavy-tailed data. The use of Student-t and Laplace-based estimators proves to be more appropriate in such contexts, yielding more stable and reliable estimates compared to Gaussian-based approaches. This illustrates the importance of accounting for heavy-tailed behavior when modeling financial and similar data.

In addition, the results show that, in cases of distributional misspecification, the estimator remains consistent for pseudo-true parameters. This empirical finding aligns with the theoretical framework of quasi-maximum likelihood estimation and confirms the robustness of the proposed methodology under incorrect distributional assumptions.

Finally, the application to real financial data demonstrates a strong agreement between the observed peaks in the data and the estimated volatility dynamics. The estimated time-varying volatility successfully captures periods of heightened variability, providing empirical evidence of the relevance and effectiveness of the proposed models. Overall, these numerical and empirical findings strongly support the theoretical results derived in the previous section.

Conclusion

This thesis investigates maximum likelihood estimation for locally stationary processes, with particular emphasis on time-varying volatility models. The central idea relies on approximating a locally stationary process by a stationary one in a neighborhood of a fixed rescaled time point. This local approximation framework allows the construction of estimators that are consistent and asymptotically normal under suitable regularity conditions.

To address the presence of heavy-tailed behavior commonly observed in financial return series, the likelihood contrast function is modified in several settings. In particular, alternative contrast functions based on the Student's t and Laplace distributions are considered. Within this framework, special attention is devoted to time-varying GARCH models with innovations following either a Student's t or a Laplace distribution. These choices provide increased robustness against heavy tails and outliers, a feature that is clearly illustrated through applications to real financial data.

We show that the resulting estimators retain consistency and asymptotic normality across a wide class of contrast functions, provided that certain assumptions are satisfied. These assumptions may vary depending on the specific distribution underlying the chosen contrast function, highlighting the flexibility of the proposed framework.

Finally, the scope of numerical experiments is substantially extended to include several time-varying volatility models, such as tvGARCH and tvGJR-GARCH processes, under different likelihood specifications. After extensive simulation studies, the proposed models are compared with classical stationary counterparts using standard information criteria, namely AIC and BIC. The results confirm the relevance of the locally stationary approach and its advantages in capturing time-varying dynamics.

Bibliography

[1] Aissaoui, F., & Djeddour-Djaballah, K. (2025). Nonparametric estimation for locally stationary GARCH processes with Laplace-distributed noise. *Austrian Journal of Statistics*, 55(2), 61–78.

[2] Bardet J.-M. Boularouk Y. Djaballah K. Asymptotic behavior of the Laplacian quasi-maximum likelihood estimator of affine causal processes. *Electron. J. Statist.* 11 (1) 452 - 479, 2017.

[3] Bardet, J.-M., Doukhan, P., & Wintenberger, O. (2022). Contrast estimation for locally stationary processes. *Journal of Time Series Analysis*, 43(5), 1234–1260.

[4] Bardet J.-M., Doukhan P., Wintenberger O. (2020). Contrast estimation of general locally stationary processes using coupling. ffhal-02586009v2.

[5] Bandyopadhyay, P. S., & Forster, M. R., eds. (2011). Philosophy of Statistics. *North-Holland Publishing.*

[6] Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5, 175–192.

[7] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.

[8] Bougerol, P., & Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, 52, 115–127.

[9] Boularouk, Y. (2023). Standard Laplace Quasi-Maximum Likelihood Estimator for GARCH Processes. *Communications in Statistics - Simulation and Computation*, 52(4), 1490–1503.

[10] Box, G. (2013). Box and Jenkins: Time Series Analysis, Forecasting and Control. In: A Very British Affair. Palgrave Advanced Texts in Econometrics. Palgrave Macmillan, London.

[11] Charlton, M., & Caimo, A. (2012). Time Series Analysis. ESPON, FHAL-03609303.

[12] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1), 1–37.

[13] Dahlhaus, R. (2012). Locally Stationary Processes. Handbook of Statistics, Vol. 30, pp. 351–413. Elsevier.

[14] Dahlhaus, R., & Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Annals of Statistics*, 34(3), 1075–1114.

[15] Dahlhaus, R. (1996). Maximum likelihood estimation and model selection for locally stationary processes. *Journal of Nonparametric Statistics*, 6(2–3), 171–191.

[16] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.

[17] Fiorentini, G., Sentana, E., and Calzolari, G. (2014). On the validity of the quasi-maximum likelihood estimation of GARCH models. *Econometric Theory*, 30, 1–29.

[18] Fisher, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3–32.

[19] Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A*, 222, 309–368.

[20] Francq, C., & Zakoian, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA–GARCH processes. *Bernoulli*, 10(4), 605–637.

[21] Francq, C., & Zakoian, J.-M. (2019). GARCH Models: Structure, Statistical Inference and Financial Applications (3rd ed.). Wiley.

[22] Fryzlewicz, P., & Subba Rao, S. (2011). Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli*, 17(1), 320–346.

[23] Gauss, C. F. (1809). Theoria motus corporum coelestium in sectionibus conicis solem ambientium. Hamburg: Perthes and Besser.

[24] Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424–438.

[25] Gray, R. (2009). Probability, Random Processes, and Ergodic Properties. *Springer*, London.

[26] Hald, A. (1999). On the history of maximum likelihood in relation to inverse probability and least squares. *Statistical Science*, 14(2), 214–222.

[27] Harald C. (1940), On the Theory of Stationary Random Processes, *Annals of Mathematics, Second Series*, Vol. 41, No. 1 (Jan., 1940), pp. 215-230

[28] Haiyan, X., Lixin, S., Muhammad, A., & Yongxia, S. (2017). Quasi-maximum likelihood estimator of Laplace (1, 1) for GARCH models. *Open Mathematics*, 15, 1539–1548.

[29] Hsu DA (1979). Long-Tailed Distributions for Position Errors in Navigation. *Journal Of The Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 62–72.

[30] Kerar, S., & Djeddour, K. (2019). Quasi-maximum likelihood estimation of GARCH with Student-distributed noise. *Communications in Statistics - Simulation and Computation*, 50(5), 1249–1271.

[31] Kolmogorov, A. N. (1933). Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer.

[32] Konishi, S., & Kitagawa, G. (2008). Information Criteria and Statistical Modeling. *Springer Series in Statistics.*

[33] Lambert, J. H. (1760). Anmerkungen und Zusätze zur Entwerfung der Land- und Himmelskugeln. Berlin.

[34] Nelson, D. B. (1988). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 4(2), 17–35.

[35] Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike Information Criterion Statistics. Part I. *D. Reidel.*

[36] Slutzky, E. (1927). The summation of random causes as the source of cyclic processes. In: Problems of Economic Conditions, Vol. 3, No. 1, Moscow, pp. 34–64 (English summary pp. 156–160).

[37] Straumann, D. (2005). Estimation in Conditionally Heteroskedastic Time Series Models. Springer.

[38] Wold, H. (1938). A study in the analysis of stationary time series. Stockholm: Almqvist & Wiksell.

[39] Wong, C. S., W. S. Chan, and P. L. Kam. (2009). A Student t-mixture autoregressive model with appli-cations to heavy-tailed financial data. *Biometrika* 96 (3):75160.

[40] Yule, U. (1926). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society A*, 226, 267–298.