الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي

University of Mila

جـــامعة ميلة

Faculty of Mathematics and Computer Sciences
Order N°: .....
Registration number: M133/2022

Department: Mathematics.
Field : Mathematics.
Specialty: Probabilities And Statistic

Thesis

Submitted for the degree of Doctorate LMD

Bayesian approach for the study of spatiotemporal GARCH models

Presented by: Atika Aouri

Jury Members:

Mohammed Salah Abdelouahab	**Pr**	**University of Mila**	**Chairman**
Soumia Kharfouchi	**Pr**	**Salah Boubnider Constantine 3**	**Supervisor**
Mohamed Boukeloua	**MCA**	**The National Polytechnic School of Constantine**	**Co-supervisor**
Yakoub Boularouk	**MCA**	**University of Mila**	**Examiner**
Sarra Leulmi	**MCA**	**The University of Mentouri Brothers Constantine 1**	**Examiner**
Rabah Bououden	**MCA**	**University of Mila**	**Examiner**
Samira Boukaf	**MCA**	**University of Mila**	**Examiner**

University year: 2025/2026

Abstract

Volatility modeling is essential in many fields, including finance, environmental science, and spatial econometrics, where data often show both spatial and temporal dependencies. Classical GARCH models effectively capture temporal volatility but ignore spatial correlations, limiting their applicability to spatially heterogeneous data.

This thesis introduces a spatio-temporal GARCH (STGARCH) model that relaxes the common spatial stationarity assumption by allowing parameters to vary smoothly across space. Observations at each fixed location are modeled as a temporal GARCH process, while spatial variability in parameters is handled through nonparametric estimation.

Three estimation methods are proposed. First, we develop the localized least squares (LLS) and local linear least squares (LLLS) estimators; the latter incorporates linear trends to reduce bias. Second, we introduce an Iterative Localized Expectation-Maximization (LEM) algorithm which adopts an Empirical Bayesian perspective to reconstruct latent volatility. We study the theoretical properties of the kernel-based estimators under two asymptotic settings: (i) fixed locations with increasing time, and (ii) infill asymptotics where both locations and time increase. We establish asymptotic normality and, in the latter case, consistency.

Simulation studies confirm the estimators' robustness and highlight the superior performance of the LEM algorithm in capturing spatial non-stationarity. An application to real ozone data demonstrates the model's ability to capture localized volatility patterns and outperform stationary approaches. This work provides a flexible and statistically sound framework for modeling and forecasting volatility in complex spatio-temporal environments.

Résumé

La modélisation de la volatilité joue un rôle essentiel dans de nombreux domaines, notamment la finance, les sciences de l'environnement et l'économétrie spatiale, où les données présentent souvent à la fois des dépendances spatiales et temporelles. Les modèles GARCH classiques capturent efficacement la volatilité temporelle mais négligent les corrélations spatiales, ce qui limite leur applicabilité aux données spatialement hétérogènes.

Cette thèse introduit un modèle GARCH spatio-temporel (STGARCH) qui assouplit l'hypothèse commune de stationnarité spatiale en permettant aux paramètres de varier de manière lisse dans l'espace. Les observations à chaque site fixe sont modélisées par un processus GARCH temporel, tandis que la variabilité spatiale des paramètres est traitée par une estimation non paramétrique.

Trois méthodes d'estimation sont proposées. Tout d'abord, nous développons les estimateurs des moindres carrés localisés (LLS) et des moindres carrés linéaires locaux (LLLS) ; ce dernier incorpore des tendances linéaires afin de réduire le biais. Ensuite, nous introduisons un algorithme d'Espérance-Maximisation Localisé Itératif (LEM) qui adopte une perspective bayésienne empirique pour reconstruire la volatilité latente. Nous étudions les propriétés théoriques de ces estimateurs à noyau dans deux cadres asymptotiques : (i) un nombre fixe de sites avec une dimension temporelle croissante, et (ii) une asymptotique de remplissage (infill) où le nombre de sites et la taille temporelle augmentent simultanément. Nous établissons la normalité asymptotique et, dans le dernier cas, la consistance des estimateurs.

Des études de simulation confirment la robustesse des estimateurs et mettent en évidence la performance supérieure de l'algorithme LEM pour capturer la non-stationnarité spatiale. Une application à des données réelles de concentration d'ozone démontre la capacité du modèle à représenter les schémas locaux de volatilité et à surpasser les approches stationnaires. Ce travail offre un cadre flexible et statistiquement rigoureux pour la modélisation et la prévision de la volatilité dans des environnements spatio-temporels complexes.

الملخص

تُعد نمذجة التقلب (Volatility) أمراً جوهرياً في العديد من المجالات، بما في ذلك المالية، والعلوم البيئية، والاقتصاد القياسي المكاني، حيث تُظهر البيانات غالباً اعتماداً متبادلاً بين البعدين الزماني والمكاني. تلتقط نماذج GARCH الكلاسيكية التقلب الزمني بفعالية، لكنها تهمل الارتباطات المكانية، مما يحد من قابليتها للتطبيق على البيانات غير المتجانسة مكانياً.

تُقدم هذه الأطروحة نموذج GARCH زماني-مكاني (STGARCH) يُخفف من فرضية الثبات المكاني (Stationarity) الشائعة من خلال السماح للمعلمات بالتغير بسلاسة عبر المكان. يتم نمذجة المشاهدات في كل موقع ثابت كعملية GARCH زمنية، بينما تتم معالجة التغير المكاني في المعلمات باستخدام أساليب التقدير اللامعلمي.

تم اقتراح ثلاث طرق للتقدير في هذا العمل. أولاً، قمنا بتطوير مقدر المربعات الصغرى المموضعة (LLS) ومقدر المربعات الصغرى الخطية المحلية ؛(LLLS) حيث يدمج الأخير اتجاهات خطية لتقليل الانحياز. ثانياً، قدمنا خوارزمية تعظيم التوقع المموضعة التكرارية (LEM) التي تتبنى منظوراً بايزياً تجريبياً (Empirical Bayesian) لإعادة بناء التقلب الكامن. قمنا بدراسة الخصائص النظرية لهذه المقدرات المعتمدة على النواة في إطارين تقريبيين: (1) مواقع ثابتة مع زيادة السلسلة الزمنية، و(2) التقارب المكاني الكثيف (Infill asymptotics) حيث يزداد كل من عدد المواقع والزمن معاً، مع إثبات التوزيع الطبيعي التقريبي، وفي الحالة الثانية، الاتساق .(Consistency)

تؤكد دراسات المحاكاة متانة المقدرات وتبرز الأداء المتفوق لخوارزمية LEM في التقاط عدم الثبات المكاني. كما توضح دراسة تطبيقية على بيانات الأوزون الحقيقية قدرة النموذج على تمثيل أنماط التقلب المحلية وتفوقه على النماذج الثابتة. يوفر هذا العمل إطاراً مرناً وسليماً إحصائياً لنمذجة وتنبؤ التقلب في البيئات الزمانية-المكانية المعقدة.

"The people who don't know they are Bayesian are called non-Bayesian."

- Irving J. Good

Dedication

To my family, for their unwavering love and support,
and to myself, for the perseverance and resilience that carried me through
this journey.

Acknowledgements

The guidance of my supervisor and co-supervisor throughout this research journey is gratefully acknowledged.

I owe special thanks to Professor Philipp Otto, whose invaluable behind-the-scenes guidance, insightful explanations, and generous sharing of expertise greatly shaped both the direction and understanding of this work.

Gratitude is also extended to Professor Mohammed-Salah Abdelouahab, head of the institution, for fostering an academic environment conducive to research and learning.

Declaration

I declare that this thesis was conducted at the University of Mila, within the Department of Mathematics, under the supervision of my supervisor and co-supervisor.

The implementation and application work was carried out by my supervisor.

I further declare that the research presented in this thesis is my own work, except where otherwise acknowledged, and that all sources have been properly cited.

List of Figures

List of Tables

Contents

1 Introduction

1.1 Background and Motivation

Spatial data analysis has become an essential tool in numerous scientific and applied fields, including ecology, environmental sciences, epidemiology, and finance. In many modern applications, data are not only indexed in space but also evolve over time, giving rise to spatio-temporal data. Such data capture the joint influence of location and time on the observed process and are frequently recorded over regular or irregular geographical domains.

Accurate modeling of spatio-temporal data is crucial for understanding underlying mechanisms, forecasting future states, and making informed decisions. Examples include tracking the spread of pollutants across a region, monitoring volatility in interconnected financial markets, and studying weather patterns over time. These applications often involve complex dependencies nonlinear, non-Gaussian, and non-stationary in both space and time that cannot be captured adequately by purely spatial or purely temporal models. Addressing these complexities requires models capable of accommodating intricate interactions between spatial and temporal dimensions [3].

In the temporal domain, the Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) family of models [24, 5] has proven highly successful in modeling time-varying volatility. GARCH models are particularly valued in finance and econometrics for their ability to capture volatility clustering, periods of high variability followed by periods of relative calm, while remaining relatively parsimonious. Numerous extensions have been proposed to account for empirical features such as asymmetry, long memory, and regime changes, including EGARCH and GJR-GARCH [30], FIGARCH [6], and regime-switching GARCH models [36]. Despite their success, these models are inherently temporal and ignore spatial dependence, which is a critical limitation in applications where volatility propagates across space as well as time.

1.2 From Temporal to Spatio-Temporal GARCH Models

The need to incorporate spatial interaction into volatility dynamics has motivated the development of spatio-temporal GARCH-type models. Broadly, existing approaches can be grouped into several classes. A first class consists of multivariate GARCH models that include spatial or network-based lags in the volatility equation but typically assume conditional independence across locations at the same time point (e.g., [11, 15, 37, 12]). A second class includes spatio-temporal GARCH or log-GARCH models that allow for instantaneous spatial dependence in volatility, often through explicit spatial weighting schemes or neighbor structures [57, 58, 50, 51]. A third line of research focuses on stochastic volatility models with spatially correlated innovations [39], where cross-sectional dependence is introduced via the latent volatility process rather than through deterministic recursion. These developments have substantially expanded the scope of volatility modeling beyond the purely temporal setting. However, most existing spatio-temporal GARCH models are formulated for areal data or for a fixed and finite set of spatial locations. This reliance on discrete spatial structures restricts the analysis to pre-defined observation points, thereby precluding the prediction of volatility at unobserved coordinates and limiting the application of infill asymptotic theory. Furthermore, they typically rely on the assumption of spatial stationarity, meaning that model parameters are constant across space. While this assumption simplifies theoretical analysis and estimation, it may be overly restrictive in heterogeneous domains.

1.3 Addressing Spatial Non-stationarity

In many real-world applications, spatial homogeneity of volatility dynamics is unrealistic. Geographic heterogeneity, environmental gradients, institutional differences, or localized socio-economic factors can induce substantial spatial variation in baseline volatility levels, persistence, and responsiveness to shocks. For example, ground-level ozone concentrations may exhibit markedly different volatility patterns near highways or industrial areas compared to rural regions [17]. Housing price volatility can vary sharply across neighborhoods due to zoning regulations, amenities, or demographic composition [34]. In financial networks, firms or regions with different balance

sheet structures or degrees of interconnectedness may display distinct volatility dynamics and spillover effects [2]. Assuming a single global set of GARCH parameters in such settings risks masking important local features, leading to biased estimation and misleading inference. Allowing model parameters to vary smoothly across space provides a more flexible and interpretable representation of spatio-temporal volatility dynamics. This idea is closely related to the concept of locally stationary or smoothly time-varying models in the time series literature [20], and its extension to spatio-temporal volatility modeling is both natural and necessary. Most existing spatio-temporal GARCH models impose spatial stationarity, which is often inadequate in the presence of heterogeneity, asymmetry, or directional effects across space. To date, theoretical and methodological frameworks that explicitly target smooth spatial variation in GARCH dynamics are lacking. Consequently, there is a pressing need for a rigorously defined model that relaxes the stationarity assumption while maintaining computational tractability and statistical interpretability.

1.4 Estimation Strategy

In this thesis, we propose a spatially non-stationary spatio-temporal GARCH (STGARCH) model that extends traditional frameworks by allowing spatially varying parameters and introducing geostatistical dependence via a covariance structure on the innovation process. Specifically, the parameters $\omega_0(s)$, $\alpha_i(s)$, and $\beta_j(s)$ vary smoothly with spatial location $s \in [0,1]^2$. Unlike lattice-based approaches that are restricted to fixed nodes, our model treats the volatility parameters as continuous functions over the spatial domain, allowing for the evaluation of volatility dynamics at any arbitrary location. At each fixed location, the process follows a standard GARCH(p,q) structure in time, while spatial heterogeneity is captured through smoothly varying parameter surfaces. Concurrently, the geostatistical dependence captures contemporaneous cross-sectional correlations that propagate forward in time through the recursive GARCH mechanism, enabling a parsimonious yet flexible representation of spatial volatility spillovers. This formulation preserves the well-understood temporal dependence of GARCH models while allowing for flexible, location-specific volatility dynamics. To estimate the spatially varying parameters, we employ kernel-based local estimation techniques. Specifically, we consider a localized least squares estimator that assumes parameters are approximately constant within a neighborhood of

a target location and weights observations according to spatial proximity. To reduce bias and improve boundary behavior, we further develop a local linear least squares estimator that incorporates first-order spatial variation in the parameters [25]. Both estimators are designed to exploit smoothness across space while accommodating irregular spatial sampling. Regularized variants (ridge-type regularization) are also considered to ensure numerical stability in finite samples. In addition to direct kernel-based least squares estimation, we also develop an alternative estimation approach based on a missing-data formulation of the STGARCH model. By treating the squared returns at an unobserved target location as latent variables, we propose an Iterative Localized Expectation-Maximization (LEM) algorithm. This approach adopts an Empirical Bayesian perspective: the E-step imputes the missing data using spatial best linear prediction (Kriging) and reconstructs the latent volatility path, while the M-step updates the parameters by maximizing a posterior density that incorporates a spatial informative prior derived from neighboring observations. This formulation allows the estimator to "borrow strength" from the local spatial structure, effectively regularizing the estimation in sparse or unobserved regions.

1.5 Thesis Contributions and Structure

The contributions of this thesis are fourfold. First, we introduce a spatially non-stationary spatio-temporal GARCH(p, q) model with smoothly varying parameters, extending the applicability of GARCH-type volatility modeling to heterogeneous spatio-temporal contexts. Second, we develop and analyze kernel-based local least squares and local linear least squares estimators for location-dependent GARCH parameters, and establish their asymptotic properties under different asymptotic regimes. Third, we propose an Empirical Bayesian estimation framework implemented via a Localized EM algorithm. This method provides a robust alternative to standard kernel estimation, particularly for inferring dynamics at unobserved locations by synthesizing temporal reconstruction with spatial regularization. Fourth, we validate the proposed kernel-based methodology through simulation studies and a real-world application, demonstrating its effectiveness in capturing both spatial and temporal volatility structures. The remainder of the thesis is organized as follows. Chapter 2 reviews preliminary concepts, including the GARCH family of models and relevant background on spatio-temporal

processes. Chapter 3 formally defines the proposed spatially non-stationary STGARCH(p, q) model, examines its theoretical properties, and discusses stationarity conditions. Chapter 4 presents the estimation methodology, with a primary focus on kernel-based local estimators and an extension to the EM-based approach. Chapter 5 reports simulation results, while Chapter 6 applies the proposed framework to a real dataset. Finally, Chapter 7 summarizes the main findings and outlines directions for future research.

2 Preliminaries

This chapter provides the foundational concepts necessary to understand and analyze the spatio-temporal GARCH models developed throughout this thesis. It reviews key elements from univariate GARCH theory, spatial and spatio-temporal processes, and the nonparametric tools used in estimation. We also give a brief overview of Bayesian modeling ideas, which are proposed as future extensions. Together, these components form the theoretical backbone for both the methodological and applied contributions of the thesis.

2.1 GARCH models and volatility

Let X_t be a real-valued discrete-time stochastic process, and let $\mathcal{F}_t$ denote the information set available up to time t. According to [5], the process X_t follows a GARCH(p, q) model if:

Definition 2.1.1. *Let X_t be a real-valued discrete-time stochastic process and $\mathcal{F}_t$ the information set up to time t. The process X_t is said to follow a GARCH(p, q) model if*

$$X_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_t^2),$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

where $p \geq 0$, $q > 0$, $\omega > 0$, $\alpha_i \geq 0$ for $i = 1, \ldots, q$, and $\beta_j \geq 0$ for $j = 1, \ldots, p$.

An alternative characterization of the GARCH(p, q) process, presented in [28], distinguishes among three variants-strong, semi-strong, and weak GARCH defined as follows:

Definition 2.1.2. *The process X_t is said to follow a GARCH(p, q) model if it satisfies:*

$$\mathbb{E}(X_t \mid \mathcal{F}_{t-1}) = 0, \quad \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

and falls under one of the following types:

- ***Strong GARCH:*** $\mathrm{Var}(X_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$ *and* $Z_t := X_t/\sigma_t$ *is i.i.d.*
- ***Semi-Strong GARCH:*** $\mathrm{Var}(X_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$.
- ***Weak GARCH:*** $\mathbb{P}(X_t^2 \mid$ *past values of* $X_t, X_t^2) = \sigma_t^2$, *where* $\mathbb{P}$ *denotes the best linear projection.*

Remark 2.1.1. *Define* $Z_t = X_t/\sigma_t$. *For both strong and semi-strong GARCH models, it holds that*

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

These follow from the conditional properties of X_t *and the definition of* σ_t.

Remark 2.1.2. *Assuming* $Z_t \sim \mathcal{N}(0,1)$ *and independence of innovations, it follows that*

$$\mathrm{Cov}(Z_t, Z_{t+s}) = 0 \quad \textit{for all } s \neq 0.$$

Therefore, the sequence $\{Z_t\}$ *is i.i.d. standard normal.*

Bollerslev [5] established the necessary and sufficient conditions for the existence of a weakly stationary solution to the GARCH model. Before stating the result, we first recall the definition of weak stationarity.

Definition 2.1.3. *A stochastic process* X_t, $t \in \mathbb{Z}$, *is said to be* ***weakly stationary*** *if:*

1. $\mathbb{E}(X_t)$ *is constant (i.e., independent of* t*), and*
2. $\mathrm{Cov}(X_{t+h}, X_t)$ *depends only on the lag* h, *not on* t.

In what follows, the term *stationary* will refer to weak stationarity unless otherwise specified.

Theorem 2.1.1. *Let* X_t *be a GARCH(p, q) process as defined in Definition 1. Then* X_t *is stationary with:*

$$\mathbb{E}(X_t) = 0, \quad \mathrm{Var}(X_t) = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}, \quad \mathrm{Cov}(X_t, X_s) = 0 \quad \textit{for } t \neq s,$$

if and only if:

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

The full proof is provided in [5] and is therefore omitted here. However, we outline the main ideas of the proof below using the commonly applied GARCH(1,1) model as an illustrative example.

Example 2.1.1. *Consider the GARCH(1,1) model:*

$$\sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Then the process X_t is stationary if and only if:

$$\alpha_1 + \beta_1 < 1.$$

Proof 2.1.1 (Sketch of Proof)**.** *Assume $X_t = \sigma_t Z_t$ with $Z_t \sim$ i.i.d. $\mathcal{N}(0,1)$. By recursive substitution:*

$$\sigma_t^2 = \omega + (\alpha_1 Z_{t-1}^2 + \beta_1)\sigma_{t-1}^2.$$

Iterating backward:

$$\sigma_t^2 = \omega \sum_{k=0}^{\infty} \left(\prod_{j=1}^{k} (\alpha_1 Z_{t-j}^2 + \beta_1) \right).$$

Taking expectations and using the fact that Z_t is i.i.d. and $\mathbb{E}(Z_t^2) = 1$:

$$\mathbb{E}(\sigma_t^2) = \omega \sum_{k=0}^{\infty} (\alpha_1 + \beta_1)^k = \frac{\omega}{1 - \alpha_1 - \beta_1},$$

which is finite if and only if $\alpha_1 + \beta_1 < 1$.

To empirically illustrate the theoretical properties of GARCH models discussed earlier, we simulate a univariate GARCH(1,1) process using the parameter configuration:

$$\omega = 0.1, \quad \alpha_1 = 0.1, \quad \beta_1 = 0.8.$$

This choice ensures that the stationarity condition $\alpha_1 + \beta_1 < 1$ holds, guaranteeing a finite unconditional variance. The simulation is conducted over 2000 time periods with standard normal innovations. Figure 1 displays the resulting time series of returns X_t and their squared values X_t^2. The return series appears unstructured and exhibits rapid fluctuations around zero, mimicking

the behavior of financial returns. The squared series, however, reveals visible clustering of large values-indicative of *volatility clustering*, a common stylized fact in empirical finance.

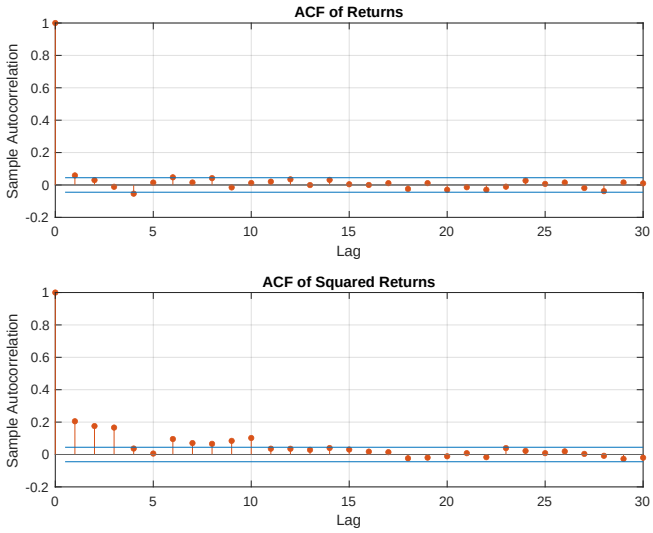

Figure 1: Simulated GARCH(1,1) returns (top) and their squared values (bottom). Volatility clustering is apparent in the squared returns.

To investigate the autocorrelation structure of the series, we compute and plot the sample autocorrelation functions (ACFs) of both the returns

and their squared values, shown in Figure 2. As expected under the GARCH framework, the raw returns exhibit negligible autocorrelation across lags-consistent with the property of conditional mean zero and serial independence. In contrast, the squared returns display significant autocorrelation that decays slowly, which is a key signature of conditional heteroskedasticity.

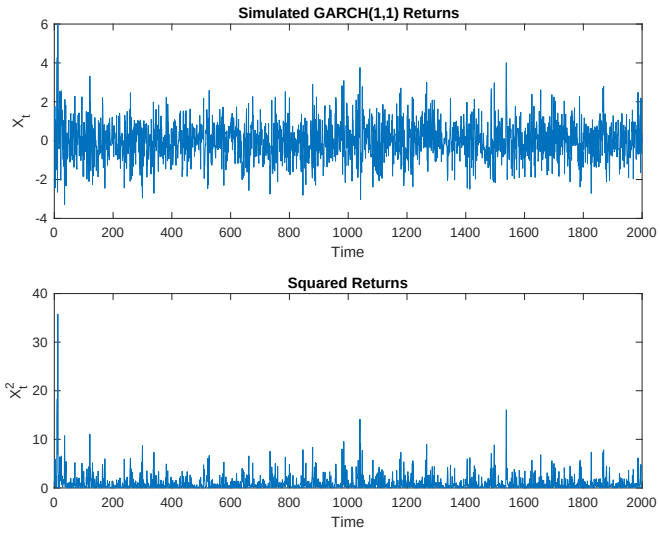

Figure 2: Sample autocorrelation functions of simulated GARCH(1,1) returns (top) and squared returns (bottom). The squared returns exhibit significant persistence.

Finally, Figure 3 compares the empirical return density from the simulated GARCH(1,1) process with a Gaussian distribution having the same mean and variance. The empirical distribution is leptokurtic, displaying a sharper peak at zero and heavier tails than the Gaussian benchmark. This reflects the excess kurtosis commonly observed in financial return series, fur-

ther validating the ability of GARCH models to capture realistic features of market data.

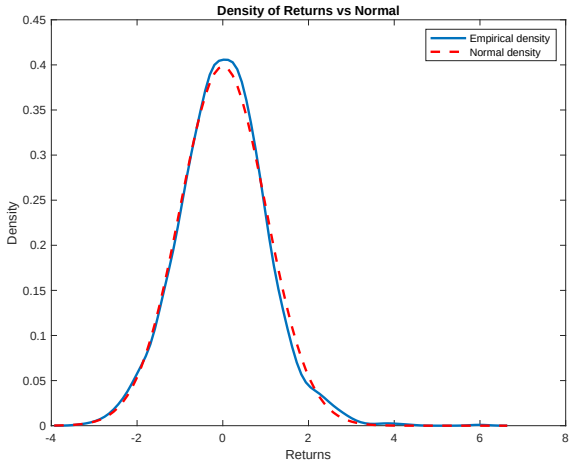

Figure 3: Empirical density of simulated GARCH(1,1) returns (solid) compared with a Gaussian distribution (dashed) with matching mean and variance. The empirical distribution shows leptokurtosis and heavier tails.

These simulation results confirm the theoretical properties of GARCH models: while the returns appear uncorrelated and Gaussian-like at first glance, their higher-order moments and dependence structures reveal strong temporal patterns-justifying the need for conditional heteroskedastic modeling frameworks in financial econometrics.

2.2 Spatio-temporal Processes

Many observed phenomena evolve over space and time, showing dependence across geo-referenced locations, where the volatility of a given time series is influenced by its neighboring series. This gave rise to the development of spatial and spatio-temporal volatility models that capture spatial interactions in variance dynamics more explicitly [52].

In recent years, spatio-temporal data analysis has gained significant attention, due to the increasing availability of large datasets collected through sensors such as remote sensing platforms, GPS networks, IoT devices, and mobile systems. These developments, coupled with advances in computational power, have enabled the construction of increasingly sophisticated models capable of capturing the complex dynamics of processes that unfold over both space and time [21].

Spatio-temporal modeling encompasses a wide range of frameworks depending on the structure of the spatial domain (continuous or discrete), the nature of spatial dependence, and the temporal dynamics involved.

A useful way to classify spatial and spatio-temporal models is according to the nature of the spatial domain and the way spatial dependence is represented. We may distinguish between models for *continuous spatial data*, which often involve covariance functions and are typically expressed through Gaussian processes; models for *areal or lattice data*, defined over discrete spatial units and incorporating dependence via spatial weights or neighborhood matrices; and *point process models*, where the spatial locations themselves are random, and dependence is modeled through the intensity or interaction functions. In addition, models with *spatially varying coefficients*, such as geographically weighted regression (GWR) are inherently spatial, particularly when they explicitly account for spatial dependence in their structure. This taxonomy highlights the diversity of spatial models and the importance of selecting a framework consistent with the underlying data-generating mechanism.

For processes observed over continuous space, models often rely on spatial covariance functions and Gaussian processes. For data observed over discrete areal units or networks, spatial dependence is typically modeled using adjacency-based weight matrices or neighborhood structures. Temporal dependence may be captured through autoregressive components, while spatial effects can be incorporated either in the mean or variance equations.

To illustrate these modeling principles, we present two representative

examples: a spatially varying autoregressive model and a spatio-temporal GARCH model with spatial weighting in the variance dynamics.

2.2.1 Spatio-Temporal AR Model with Location-Dependent Parameters

A useful way to conceptualize spatio-temporal processes is to fix a spatial location and consider the sequence of observations at that location as a univariate time series. This perspective motivates the modeling of spatio-temporal dynamics via location-specific time series models. Rao (2008) proposes a framework in which each spatial location follows an autoregressive (AR) model in time, while the innovations themselves are spatially structured, enabling spatial dependence to enter through the innovation term.

More formally, the process is defined as a location-dependent autoregressive process $\{X_t(s) : s \in [0,1]^2\}_{t=1}^T$, where for each spatial location $s = (x, y)$, the process satisfies:

$$X_t(s) = \sum_{j=1}^{p} a_j(s) X_{t-j}(u) + \sigma(s)\xi_t(s),$$

with $\{a_j(\cdot)\}_{j=1}^p$ and $\sigma(\cdot)$ representing smooth, nonparametric spatial functions. The innovation process $\{\xi_t(s)\}$ is assumed to be temporally independent and spatially stationary, with zero mean and unit variance:

$$\mathbb{E}[\xi_t(s)] = 0, \quad \text{Var}[\xi_t(s)] = 1.$$

Notably, Rao does not impose any specific distributional assumptions on $\xi_t(s)$, allowing for flexible error structures.

This formulation implies that spatial dependence is induced via the innovation process, while temporal dependence is captured through the local autoregressive terms. If the functions $\{a_j(\cdot)\}$ vary with location, then the process $\{X_t(s)\}$ is spatially nonstationary. The special case where $a_j(s)$ is constant across space corresponds to an integrated spatially stationary AR process, as discussed by Storvik et al. (2002).

Location-dependent AR models have been applied to various fields, including housing price modeling (Gelfand et al., 2003) and environmental data such as ozone concentrations (Gilleland and Nychka, 2005). An alternative treatment of spatial nonstationarity is found in Franke and Gründer (1996), who employed a random coefficient regression model to analyze greenhouse gas data collected across forest sites in Rheinland-Pfalz, Germany.

2.2.2 A Stationary Spatio-Temporal GARCH Model

An alternative way to introduce spatial dependence in spatio-temporal processes is through the conditional variance equation. In this direction, Hølleland and Karlsen (2019) [37] propose a spatio-temporal GARCH framework for data observed over discrete spatial domains, where each location on a spatial lattice evolves according to a GARCH-type process influenced by neighboring values. The general model is defined over an infinite grid $\mathbb{Z}^d$ and takes the form:

$$Y_t(s) = h_t(s)V_t(s), \quad s \in \mathbb{Z}^d,$$

$$h_t^2(s) = \mu + \sum_{i=1}^{p} \sum_{\mathbf{v}\in\Delta_{1i}} \alpha_i(\mathbf{v}) Y_{t-i}^2(s-\mathbf{v}) + \sum_{i=1}^{q} \sum_{\mathbf{v}\in\Delta_{2i}} \beta_i(\mathbf{v}) h_{t-i}^2(s-\mathbf{v}),$$

where $\{V_t(s)\}$ is a sequence of i.i.d. random variables with zero mean and unit variance, and the sets Δ_{1i} and Δ_{2i} define the spatial neighborhood structures associated with the ARCH and GARCH components, respectively.

To address boundary issues that arise from the use of a finite spatial window in applications, the authors propose a *circular modification* of the model by replacing the infinite spatial domain $\mathbb{Z}^d$ with a finite, periodic lattice $\mathcal{R} = \mathbb{Z}^d/(m\mathbb{Z}^d)$. This effectively wraps the spatial grid onto a torus, ensuring that every location has a full neighborhood, regardless of its position.

In the circular version, the model becomes:

$$Y_t(s) = h_t(s)V_t(s), \quad s \in \mathcal{R},$$

$$h_t^2(s) = \mu + \sum_{i=1}^{p} \sum_{\mathbf{v}\in\Delta_{1i}} \alpha_i(\mathbf{v}) Y_{t-i}^2((s-\mathbf{v}) \mid m) + \sum_{i=1}^{q} \sum_{\mathbf{v}\in\Delta_{2i}} \beta_i(\mathbf{v}) h_{t-i}^2((s-\mathbf{v}) \mid m),$$

where modular arithmetic ensures that all spatial operations remain within the finite set $\mathcal{R}$.

This formulation leads to a stationary spatio-temporal GARCH process that is mathematically tractable and avoids the complications of spatial boundaries. The authors study its theoretical properties, including conditions for strict stationarity and ergodicity, and propose a quasi-maximum likelihood estimator for inference. The model is particularly well-suited to settings with grid-based spatial data where volatility clusters both temporally and spatially.

2.3 Nonparametric Estimation: Kernel Smoothing and Local Polynomial Regression

Nonparametric regression provides a flexible alternative to parametric models by estimating the relationship between a dependent variable Y and explanatory variables X without assuming a specific functional form. The general form of a nonparametric regression model is:

$$Y = m(X) + \sigma(X)\varepsilon,$$

where $m(\cdot)$ is the unknown regression function and $\sigma(\cdot)$ denotes the conditional scale (variance) function. Unlike parametric regression, which may suffer from model misspecification, nonparametric approaches impose minimal structural assumptions, relying instead on local approximations of the regression function.

To estimate $m(x)$ at a point x, one widely used method is *local polynomial regression*, particularly *local linear regression*. This method assumes that the regression function can be approximated by a linear function in a neighborhood around x. Specifically, the local linear estimator $\hat{m}_{LL}(x)$ minimizes the following weighted least squares criterion:

$$\min_{a,b} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - a - b^\top (x_i - x)\right)^2 K\left(H^{-1}(x_i - x)\right),$$

where $K(\cdot)$ is a kernel function and H is a bandwidth matrix controlling the size of the neighborhood around x. The estimate $\hat{m}_{LL}(x)$ is given by the fitted intercept a, which captures the local value of the regression function at x.

This approach corrects many of the shortcomings of simpler estimators, such as the *Nadaraya-Watson estimator*, which fits a local constant. The local linear estimator provides better boundary behavior and mitigates the influence of the design density on the bias[18].

At the heart of nonparametric regression is *kernel smoothing*, where observations near the target point x receive more weight. In the case of the local constant estimator, the solution minimizes:

$$\hat{m}_{LC}(x) = \arg\min_{a} \frac{1}{n} \sum_{i=1}^{n} (y_i - a)^2 K\left(H^{-1}(x_i - x)\right),$$

yielding a weighted average of nearby y_i's. Though computationally simple, this method often exhibits boundary bias and is sensitive to the distribution of explanatory variables. These limitations are addressed in local polynomial approaches.

The choice of *kernel function* K (e.g., Gaussian, Epanechnikov) is typically less crucial than the choice of *bandwidth* H, which governs the smoothness of the estimate. Smaller bandwidths lead to lower bias but higher variance, while larger bandwidths smooth out noise at the cost of bias.

2.3.1 Bandwidth Selection

The performance of kernel estimators crucially depends on the selection of the bandwidth parameter, which controls the smoothness of the estimate. An essential step in kernel and local polynomial smoothing is the *selection of bandwidth*, which determines the degree of local averaging and thus the trade-off between bias and variance. Various strategies have been developed to guide this choice, ranging from simple heuristic rules to data-driven optimization criteria.

A *rule-of-thumb* approach provides a quick estimate of the bandwidth by assuming that the underlying distribution is approximately normal. For instance, the *normal scale rule* sets the bandwidth proportional to the sample spread and inversely proportional to the sixth root of the sample size,

$$h_{\mathrm{NS}} = \hat{\sigma} n^{-1/6},$$

where $\hat{\sigma}$ is typically estimated by the average of the marginal standard deviations or interquartile ranges. Although convenient, this rule may oversmooth or undersmooth when the data deviate from normality. The *maximal smoothing principle* provides an alternative, yielding a deliberately large (oversmoothed) bandwidth that can serve as a stable reference for exploratory analysis [22].

More refined, data-driven selectors include the *least squares cross-validation* (LSCV) criterion, which minimizes an unbiased estimate of the mean integrated squared error (MISE):

$$\mathrm{LSCV}(h) = \int \hat{f}_h^2(x)\, dx - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{h,-i}(x_i),$$

where $\hat{f}_{h,-i}$ is the leave-one-out density estimate. Although theoretically appealing, LSCV often produces undersmoothed estimates and can be unstable in regions with sparse observations.

An alternative is the *likelihood cross-validation* (LIK) method, which treats the kernel estimate as a likelihood function and maximizes the average leave-one-out log-likelihood:

$$\text{LIK}(h) = \frac{1}{n} \sum_{i=1}^{n} \log[\hat{f}_{h,-i}(x_i)].$$

This approach typically yields smoother bandwidths than LSCV but remains sensitive to outliers.

In practice, these methods are often combined or used sequentially, for example, employing rule-of-thumb bandwidths as starting values for adaptive, data-driven selectors. Despite their differences, all approaches aim to balance fidelity to local structure with sufficient smoothing to suppress random noise, a principle that remains central to nonparametric estimation in both spatial and spatiotemporal contexts [22].

Once the bandwidth has been appropriately chosen, the local estimator can be viewed more generally as a *local least squares* problem. In this framework, the regression function is estimated by minimizing a loss function over observations within a neighborhood of the target point, weighted by a kernel. Local least squares provides a unifying view of kernel smoothing and local polynomial regression as locally weighted versions of global least squares, tuned to adapt flexibly across the input space.

This local formulation also allows for straightforward extension to robust nonparametric regression via alternative loss functions, such as Huber or quantile-based loss, which can improve estimation in the presence of outliers.

While local polynomial regression and kernel smoothing provide powerful tools for estimating regression functions without parametric assumptions, they remain sensitive to outliers and deviations from ideal error distributions. This motivates the development of *robust nonparametric estimation* techniques, which modify the estimation criterion, often by replacing the squared loss function, to achieve greater resistance to data anomalies. These extensions are explored in detail in the next chapter.

2.4 Weighted Martingales and the Central Limit Theorem

In this section, we recall some definitions and results concerning weighted martingales and a central limit theorem for martingale difference arrays. These results will be used later.

Definition 2.4.1 (Weighted Martingale)**.** *Let $(\mathcal{F}_n)_{n\geq 1}$ be a filtration and $(X_n)_{n\geq 1}$ a sequence of random variables adapted to it. A process $(M_n)_{n\geq 1}$ is called a* weighted martingale *if there exists a sequence of real numbers $(w_n)_{n\geq 1}$ such that*

$$M_n = \sum_{i=1}^{n} w_i X_i, \quad \text{with } E[X_i \mid \mathcal{F}_{i-1}] = 0 \text{ and } E[X_i^2] < \infty.$$

Thus, (M_n) is a martingale with respect to $(\mathcal{F}_n)$ whose increments are weighted by w_i.

Theorem 2.4.1 (Central Limit Theorem for Martingale Difference Arrays [38, Theorem 3.2])**.** *Let $\{S_{n,i}, \mathcal{F}_{n,i}\}$, $1 \leq i \leq k_n, n \geq 1$, be a zero-mean, square-integrable martingale array with differences $X_{n,i}$, and let η^2 be an almost surely finite random variable. Suppose that*

$$\max_i |X_{n,i}| \xrightarrow{p} 0, \tag{1}$$

$$\sum_i X_{n,i}^2 \xrightarrow{p} \eta^2, \tag{2}$$

$$E\Big(\max_i X_{n,i}^2\Big) \text{ is bounded in } n, \tag{3}$$

and that the σ-fields are nested:

$$\mathcal{F}_{n,1} \subseteq \mathcal{F}_{n,2} \subseteq \cdots \subseteq \mathcal{F}_{n,k_n+1}.$$

Then

$$S_{n,k_n} = \sum_i X_{n,i} \xrightarrow{d} Z,$$

where the random variable Z has characteristic function

$$E\big[\exp\big(-\tfrac{1}{2}\eta^2 t^2\big)\big].$$

Hence, Z is conditionally normal given η^2.

Corollary 2.4.1 (Conditional Lindeberg Condition [38, Corollary 3.1])**.** *If conditions* (1) *and* (3) *are replaced by the conditional Lindeberg condition*

$$\sum_i E\big[X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}|>\varepsilon\}} \mid \mathcal{F}_{n,i-1}\big] \xrightarrow{p} 0, \quad \textit{for all } \varepsilon > 0,$$

and if (2) *and the nesting condition of the* σ*-fields hold, then the conclusion of Theorem 3.2 remains true.*

Theorem 3.2 and Corollary 3.1 are adapted from [38], *Martingale Limit Theory and Its Application*, Academic Press.

2.5 Locally Stationary and Approximation Techniques

Non-stationary data present a major challenge in time series modeling since many classical methods rely on the assumption of stationarity. When this assumption is violated, estimators may lose desirable properties such as consistency and asymptotic normality, leading to poor forecasting performance. A useful alternative framework is to model such data as *locally stationary processes*, which allows the data's second-order characteristics to evolve smoothly over time or space while retaining theoretical tractability.

Early work by Priestley [54] extended the spectral theory of stationary processes to nonstationary series via an *evolutionary spectrum*. He showed that any finite-variance process $\{X_t\}$ admits a representation

$$X_t \;=\; \int_{-\pi}^{\pi} e^{i\lambda t}\, A_t(\lambda)\, \mathrm{d}\xi(\lambda),$$

where $A_t(\lambda)$ is a time-varying transfer function and $\xi(\lambda)$ an orthogonal increment process with $E[\mathrm{d}\xi(\lambda)]^2 = \mathrm{d}\mu(\lambda)$. However, Priestley's approach lacked a rigorous asymptotic framework for inference.

Dahlhaus [19] remedied this by introducing an infill asymptotic theory for locally stationary processes: as the sampling span $N \to \infty$, the covariance structure at rescaled time $u = t/N$ varies Hölder-continuously, enabling classical limit theorems to be adapted.

Rao and Sarkar [55] then formulated a very general state-space approximation. Any process of the form

$$X_{t,N} \;=\; A_t\big(t/N\big)\, X_{t-1,N} \;+\; b_t\big(t/N\big),$$

with $\{A_t(u), b_t(u)\}$ Hölder-continuous in u, can for each fixed $u \in (0,1)$ be approximated by the frozen stationary process

$$X_t(u) \;=\; A_t(u)\, X_{t-1}(u) \;+\; b_t(u).$$

They prove the uniform bound

$$\big|X_{t,N} - X_t(u)\big| \;=\; O_p\big(|t/N - u|^{\beta} + N^{-\beta}\big),$$

where $\beta > 0$ is the Hölder exponent. Differentiating in u yields a "derivative process" $\dot{X}_t(u)$, itself geometrically ergodic, and one obtains the full Taylor expansion

$$X_{t,N} = X_t(u) + \big(t/N - u\big)\, \dot{X}_t(u) \;+\; O_p\big(|t/N - u|^{1+\beta} + N^{-1-\beta}\big).$$

Subba Rao [56] extended these ideas to purely spatial variation in AR(p) models. At each location $u \in [0,1]^2$ she "freezes" the AR coefficients $a_j(s)$ and innovation variance $r(u)$, showing that the nonstationary field can be locally approximated by the corresponding stationary AR(p) process. She then proposes both local-constant and local-linear least-squares fits which, under spatial infill asymptotics, achieve consistency and asymptotic normality, with the local-linear version reducing bias from $O(\|u - u_0\|)$ to $O(\|u - u_0\|^2)$.

In this thesis we integrate Rao & Sarkar's time-varying state-space approximation with Subba Rao's spatial-infill methodology to formulate a fully spatio-temporal GARCH model whose parameters evolve smoothly in both time and space. We then leverage the corresponding Taylor-series expansion of the frozen process to derive asymptotic properties for our localized GARCH estimators.

2.6 Bayesian Modeling

Bayesian estimation methods provide a powerful means to handle complex hierarchical models involving numerous unknown parameters and interdependencies. They also offer a systematic way to incorporate prior knowledge about the phenomenon under study into the modeling process. In Bayesian analysis, the focus is driven by the application rather than being restricted by methodological limitations. The uncertainty that is always present in statistical analysis is inherently reflected in the posterior distributions of the

model parameters and, for example, in the distributions of predictions derived from these posteriors. In many situations, the resulting inference is both intuitive and straightforward to interpret.

In general, Bayesian modeling differs from the more conventional maximum likelihood estimation by its use of probability distributions to represent uncertainty at all stages. The approach can be viewed as a process consisting of three main phases. Initially, one specifies a *prior* distribution that expresses beliefs about the parameters before seeing the data. Then, observational evidence in the form of data is used to update this belief. Finally, conclusions often in the form of posterior summaries or predictions are drawn from the updated distribution. This reflects the Bayesian interpretation of probability, where uncertainty is represented directly through distributions.

The origins of the method trace back to Bayes (1763), whose name it bears. According to Bayes' theorem for conditional probabilities, the posterior distribution is given by

$$p(\theta \mid y) = \frac{p(y \mid \theta)\, p(\theta)}{p(y)},$$

where $p(\theta \mid y)$ is the posterior distribution of the unknown parameters θ given the observed data y, $p(y \mid \theta)$ is the likelihood function, and $p(\theta)$ is the prior distribution assigned to θ. The denominator $p(y)$ represents the marginal probability of the data, sometimes called the evidence, and acts as a normalizing constant to ensure the posterior integrates to one. Since $p(y)$ depends only on the observed data, it is often treated as constant, leading to the proportional form:

$$p(\theta \mid y) \propto p(y \mid \theta)\, p(\theta).$$

This unnormalized form is commonly used to reduce computational complexity.

Bayesian analysis also naturally supports prediction. The *prior predictive distribution* which gives the distribution of the data y implied solely by the prior and the model, before observing any actual data is defined as

$$p(y) = \int p(y, \theta)\, \mathrm{d}\theta = \int p(y \mid \theta)\, p(\theta)\, \mathrm{d}\theta.$$

After observing data y, the *posterior predictive distribution* for new observations $\hat{y}$, assuming $\hat{y}$ and y are conditionally independent given θ, is obtained

as

$$p(\hat{y} \mid y) = \int p(\hat{y}, \theta \mid y)\, \mathrm{d}\theta$$

$$= \int p(\hat{y} \mid \theta, y)\, p(\theta \mid y)\, \mathrm{d}\theta$$

$$= \int p(\hat{y} \mid \theta)\, p(\theta \mid y)\, \mathrm{d}\theta.$$

These predictive distributions allow both the unknown parameters and future observations to be described in terms of probability distributions, conveying uncertainty directly.

In real-world applications, the posterior distribution rarely has a closed-form solution, particularly due to the complexity of the evidence term. Therefore, approximation techniques are typically employed. One of the most widely used is *Markov Chain Monte Carlo* (MCMC), which generates samples from the posterior distribution via Markov chains that converge to the target distribution. Such methods often require significant computational resources. This explains why, despite the early origins of Bayesian methods, their widespread use only began with advances in computing technology, especially after the development of the Gibbs sampling algorithm by Geman and Geman (1984). From the 1990s onward, the availability of dedicated software such as BUGS (Bayesian inference using Gibbs sampling, see Lunn et al., 2012) and `Stan` (Stan Development Team, 2024) has made Bayesian modeling much more accessible [32].

While MCMC remains the standard for static parameter estimation, dynamic systems with latent variables, such as the volatility in GARCH models, require specialized recursive techniques. This has led to the development of Sequential Monte Carlo (SMC) methods, also known as Particle Filters. These are inherently Bayesian algorithms that estimate the posterior distribution of latent states sequentially in time [42]. In this thesis, we adopt this modern Bayesian perspective by employing Particle Smoothing within an Expectation-Maximization (EM) framework (Chapter 4). This allows us to leverage the probabilistic power of Bayesian filtering to reconstruct unobserved volatility, while maintaining computational tractability for the spatial parameters. This integration highlights that Bayesian modeling is not just a single tool, but a flexible framework that can be adapted to solve complex inference problems in spatio-temporal statistics.

3 Definition of STGARCH Models and Theoretical Properties

In this chapter, we propose a novel spatio-temporal GARCH model that relaxes spatial stationarity by allowing each location to follow its own GARCH process in time, while still borrowing strength from neighboring sites. Building on Rao's varying-time GARCH methodology, we establish conditions under which the process at any fixed location can be approximated by a stationary GARCH model, thereby facilitating both theoretical analysis and practical inference. We formally define the model, derive its stationarity and ergodicity conditions, and investigate its asymptotic normality. Then, we extend Rao's approximation technique to our spatio-temporal setting, offering insights into the localized behavior of volatility and enabling more effective forecasting across space and time [1].

In all the sequel, $\|.\|_\infty$ denotes the sup-norm of a vector, so that, for any $m \times m$ matrix $E = [e_{ij}]$, $\|E\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{m} \mid e_{ij} \mid$. If A, E and D are matrices such that E is diagonal with all diagonal elements equal in modulus to $\rho(A)$ (the spectral radius of A) and $\rho(D) \leq \rho(E)$, the $p \times p$ matrix A is said to be spectral radius diagonalizable if it is similar to the matrix $\begin{bmatrix} E & 0 \\ 0 & D \end{bmatrix}$. Then, there exists; for A, a $p \times p$ nonsingular matrix P such that $\|P^{-1}AP\|_\infty = \rho_0$ where $\rho_0 = \rho(A)$. Moreover, $C_d = C\big([0,1]^2, \mathbb{R}^d\big)$ denotes the collection of all continuous functions $f : [0,1]^2 \to \mathbb{R}^d$, equipped with the metric $d_{C_d}(f,g) = \sup_{t\in[0,1]^2} \mid f(t) - g(t) \mid$, for $f, g \in C_d$.

3.1 Model Definition

Let $s = (x, y) \in [0,1]^2$ denote spatial location. We define the STGARCH(p,q) field $\{X_t(s)\ , s \in [0,1]^2\}$ by:

$$\begin{cases} X_t(s) = \sigma_t(s)\epsilon_t(s) \\ \sigma_t^2(s) = \omega_0(s) + \sum_{i=1}^{q} \alpha_i(s) X_{t-i}^2(s) + \sum_{j=1}^{p} \beta_j(s)\sigma_{t-j}^2(s)\ , \end{cases} \tag{4}$$

where $X_t(s)$, $\sigma_t(s)$ and $\epsilon_t(s)$ are, respectively, the observed process, the volatility process and the residual process, and $\omega_0(s)$, $\alpha_i(s), i = 1, \ldots, q$ and $\beta_j(s), j = 1, \ldots, p$ are nonparametric functions. In (4) the innovations $\epsilon_t(s)$ are :

- Independent over time, and spatially, strictly stationary process.
- Ergodic over time, with $E(\epsilon_t(s)) = 0$ and $var(\epsilon_t(s)) = 1$.
- Independent of $\sigma_t(s)$.

Remark 3.1.1. *The innovation process $\{\epsilon_t(s)\}$ is assumed to be independent over time, spatially strictly stationary, and ergodic, with zero mean and unit variance. Unlike in the purely temporal setting, we allow for spatial dependence through a valid covariance function defined on $[0,1]^2$. In this work, we consider a covariance function belonging to the exponential class, which provides a flexible representation of spatial correlation structures. This geostatistical component induces contemporaneous spatial dependence in the innovations, which then propagates through the recursive volatility equation. Consequently, the model accommodates spatial heterogeneity in both the innovation field and the volatility dynamics, while permitting prediction at unobserved locations. Under mild smoothness conditions on the parameter functions, the spatially varying process $\{X_t(s)\}$ can be locally approximated, at any fixed location $s_0 \in [0,1]^2$, by a spatially stationary GARCH(p, q) process [2].*

Remark 3.1.2. *Although the volatility process $\sigma_t^2(s)$ in (4) does not include instantaneous spatial interactions (that is, terms such as $\sigma_t^2(s')$ with $s' \neq s$), and is thus temporally causal at each location s, it nonetheless inherits spatial dependence through the squared observations $X_{t-i}^2(s)$. Since $X_t(s) = \sigma_t(s)\epsilon_t(s)$ and the innovation process $\epsilon_t(s)$ is spatially correlated across $s \in [0,1]^2$, the process $X_t^2(s)$ also exhibits spatial dependence. More precisely, if $\epsilon_t(s)$ follows a zero-mean, unit-variance Gaussian process with spatial covariance function $C(h)$, then*

$$\mathrm{Cov}\big(\epsilon_t^2(s), \epsilon_t^2(s')\big) = 2\,C(h)^2, \quad \textit{where } h = \|s - s'\|.$$

Hence, the spatial correlation in the squared innovations decays faster than in the innovations themselves, as $C(h)^2$ decreases more rapidly than $C(h)$ with distance. This attenuation effect is transmitted to $X_t^2(s)$ and, subsequently, to $\sigma_t^2(s)$ via the recursive GARCH dynamics. Therefore, even though $\sigma_t^2(s)$ is not explicitly defined through spatial interaction terms, it remains a spatially dependent process whose structure is governed by the spatial properties of the innovation field.

We notice that for the process $\{X_t(s)\ , s \in [0,1]^2\}$ solving the equation 4 the process $\sigma_t^2(s)$ can be considered as a sub-diagonal STBL(p,0,1,q) process having the following representation

$$\sigma_t^2(s) = \omega_0(s) + \sum_{i=1}^{q} \alpha_i(s)\eta_{t+1-i}^2(s)\sigma_{t-i}^2(s) + \sum_{j=1}^{p} \beta_j(s)\sigma_{t-j}^2(s)\ , \qquad (5)$$

where $\eta_t(s) = \epsilon_{t-1}(s)$.Thus, the model is a particular instance of the location-dependent bilinear process introduced in Equation (1.1) of [40]. We can utilize Theorems 2 and 3 from [40] to establish time stationarity at every spatial location for the proposed STGARCH(p,q) process, along with a Central Limit Theorem (CLT). In order to apply these theoretical results, we introduce the following matrices.

$$\underline{\sigma}_t^2(s) = \begin{bmatrix} \sigma_t^2(s) \\ \vdots \\ \sigma_{t-p+1}^2(s) \end{bmatrix}_{p\times 1}, \quad C(s) = \begin{bmatrix} \omega_0(s) \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{p\times 1}$$

$$B(s) = \begin{bmatrix} \beta_1(s) & \beta_2(s) & \cdots & \beta_p(s) \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 1 & 0 \end{bmatrix}_{p\times p}$$

Without loss of generality, we assume that $q < p$ and define $A_j(s), j = 1, \ldots, q$ are $p \times p$ matrices whose entries are all zero except for the element located in first row and j-th column, which is set equal to $\alpha_j(s)$, for example,

$$A_2(s) = \begin{bmatrix} 0 & \alpha_2(s) & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 \end{bmatrix}_{p \times p}$$

So, the bilinear representation given in Equation 5 can be equivalently expressed in the following matrix form:

$$\underline{\sigma}_t^2(s) = C(s) + B(s)\underline{\sigma}_{t-1}^2(s) + \sum_{j=1}^{q} A_j(s)\epsilon_{t-j}^2(s)\underline{\sigma}_{t-1}^2(s). \tag{6}$$

3.2 Stationarity

Stationarity is a fundamental property in time series analysis, ensuring that the probabilistic behaviour of the process does not change over time. In the context of the STGARCH(p, q) model defined in Equation 4, temporal stationarity at each spatial location $s \in [0, 1]^2$ implies that the distribution of $\{X_t(s)\}_{t\in\mathbb{Z}}$ is invariant under time shifts, and that key moments such as $\mathbb{E}[X_t(s)]$ and $\mathbb{E}[X_t^2(s)]$ remain constant over time.

From a practical standpoint, temporal stationarity is crucial for two reasons: It allows the use of long-run statistical averages for parameter estimation and inference at each fixed spatial location. In addition, ensures that the volatility process $\sigma_t^2(s)$ remains bounded and does not drift or explode over time.

In the purely temporal GARCH setting, necessary and sufficient conditions for stationarity are well-established. For example, Theorem 2 of [40] provides a spectral radius condition involving the autoregressive coefficients of the variance equation. In this work, we adapt these results to the spatially varying setting of the STGARCH(p, q) model, where the coefficients themselves are smooth functions of s. The main challenge is that stationarity must hold uniformly across space, not just at a single location.

The following theorem extends the result of [40] to our spatio-temporal framework, providing sufficient conditions for temporal stationarity at each fixed spatial location.

Theorem 3.2.1. *$\{X_t(s),\ s \in [0,1]^2\}$ be as defined in the Equation 4, and let $C(s)$, $A_j(s)$ and $B(s)$ be the matrix coefficients of Equation* (6)*. Assume that $B(s)$ is a spectral radius diagonalizable matrix with $\|P^{-1}(s)B(s)P(s)\|_\infty = \rho_0(s)$, and set $\delta_j(s) = \|P^{-1}(s)A_j(s)P(s)\|_\infty$.*
Assume further that

$$\sup_s E \log\left(\rho_0(s) + \sum_{j=1}^{q} \delta_j(s) \mid \epsilon_{t-j}^2 \mid\right) < 0.$$

Then, for all fixed s, there exists a unique strictly stationary (over time) process $X_t(s)$ satisfying (1)*, such that $\sigma_t^2(s)$ is given by the first component of $\underline{\sigma}_t^2(s)$ which have the following specification*

$$\underline{\sigma}_t^2(s) = C(s) + \sum_{r=1}^{\infty}\prod_{k=0}^{r-1}\left(B(s) + \sum_{j=1}^{q} A_j(s)\epsilon_{t-j-k}^2(s)\right)C(s), \tag{7}$$

with the infinite series being almost surely convergent.

Proof. see Appendix A.2 □

Now, let's provide examples to illustrate this theorem

Example 3.2.1. *Consider the STGARCH $(1,1)$ process*

$$\begin{cases} X_t(s) = \sigma_t(s)\epsilon_t(s)\,, \quad s = (x,y) \in [0,1]^2 \\ \sigma_t^2(s) = \omega(s) + a(s)\, X_{t-1}^2(s) + c(s)\,\sigma_{t-1}^2(s)\,, \end{cases}$$

where $a : [0,1]^2 \to [-1,1]$, $c : [0,1]^2 \to [-1,1]$ and

$$\begin{aligned} a(x,y) &= a_1(5-x)\cos y, \\ c(x,y) &= a_2 \sin x \cos y, \end{aligned}$$

and the innovations $\{\epsilon_t(s)\}$ are independent over time and spatially, strictly stationary and ergodic over time, with $E[\epsilon_t(s)] = 0$ and $\operatorname{var}[\epsilon_t(s)] = 1$. In this case $\underline{\sigma}_t^2(s) = \sigma_t^2(s)$, $B(s) = c(s)$, $C(s) = \omega(s)$ and $A_1(s) = a(s)$. It is clear that $A(s)$ is a spectral radius diagonalizable 1×1 matrix with $\rho_0(s) =\mid c(s) \mid$ and $\delta_1(s) =\mid a(s) \mid$. The sufficient condition for the existence of a strictly stationary (over time) process $\sigma_t^2(s)$ is

$$\sup_s E \log(\mid c(s) \mid + \mid a(s) \mid\mid \epsilon_1^2(s) \mid) < 0,$$

which becomes

$$| a_2 | +5 | a_1 | < 1.$$

Moreover, the solution is given by

$$\sigma_t^2(s) = \omega(s) + \omega(s) \sum_{r=1}^{\infty} \prod_{k=0}^{r-1} [a(s) + c(s)\, \epsilon_{t-1-k}^2(s)].$$

In this STGARCH$(1,1)$ case, the stationarity condition reduces to

$$|a_2| + 5|a_1| < 1,$$

which has a simple geometric interpretation: (a_1, a_2) must lie inside a diamond-shaped region in the parameter plane. The factor 5 arises from the specific spatial variation chosen for $c(x, y)$ and $a(x, y)$ in this example: the constant term $(5 - x)$ in $c(x, y)$ bounds the magnitude of $c(s)$ over the spatial domain, which in turn impacts the spectral radius bound in Theorem 3.2.1.

Intuitively:

- $|a_1|$ controls how strongly local shocks at time $t-1$ influence volatility at time t.
- $|a_2|$ controls the persistence of volatility from the previous period.
- The inequality above ensures that, on average, the multiplicative effect of past shocks and past volatility is small enough for the recursion in $\sigma_t^2(s)$ to dampen over time, guaranteeing stationarity.

Figure 4 illustrates the stationarity region in the (a_1, a_2) plane. Points inside the diamond satisfy $|a_2| + 5|a_1| < 1$ and yield a strictly stationary process (over time) at each fixed s. Points outside the diamond violate the condition and lead to divergence in $\sigma_t^2(s)$.

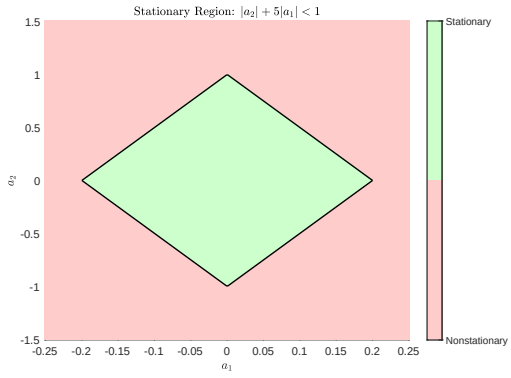

Figure 4: Safe-parameter region for Example 3.2.1. Stationarity holds for (a_1, a_2) inside the diamond defined by $|a_2| + 5|a_1| < 1$.

To demonstrate, we simulate $\sigma_t^2(s)$ at a fixed $s = (0.5, 0.5)$ for $T = 500$:

1. **Stationary case:** $(a_1, a_2) = (0.05, 0.2)$ satisfies $0.2 + 5 \times 0.05 = 0.45 < 1$. Volatility remains bounded and fluctuates around a finite mean (Figure 5).

2. **Nonstationary case:** $(a_1, a_2) = (0.15, 0.5)$ gives $0.5 + 5 \times 0.15 = 1.25 > 1$. Volatility grows without bound over time (Figure 5).

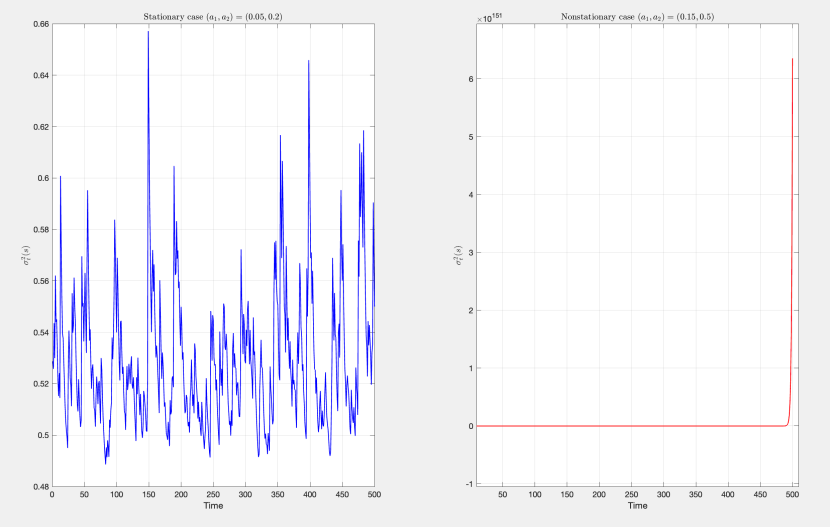

Figure 5: Simulated $\sigma_t^2(s)$ for stationary (left) and nonstationary (right) parameter choices in Example 2.2.1.

Remark 3.2.1. *When the condition* $|a_2| + 5|a_1| < 1$ *is violated, the volatility recursion behaves like a multiplicative process with a growth factor exceeding unity in expectation. This causes* $\sigma_t^2(s)$ *to drift upwards over time, eventually dominating the innovations* $\varepsilon_t(s)$ *and leading to explosive behaviour in* $X_t(s)$*. From an applied perspective, this would correspond to a system where shocks have a compounding and unsustainable effect on volatility, making the process unsuitable for modelling stable real-world systems.*

3.3 Central limit theorem (CLT)

In addition to stationarity, another fundamental probabilistic property in time series analysis is the asymptotic distribution of estimators and sample averages. In particular, establishing a Central Limit Theorem (CLT) for our spatio-temporal process ensures that, as the temporal sample size T grows, the properly normalised sample mean converges in distribution to a Gaussian limit. This property is essential for statistical inference, as it justifies the use of asymptotic confidence intervals and hypothesis tests.

In the present setting, we focus on the temporal CLT for the process $\{X_t(s)\}$ at a fixed spatial location $s \in [0,1]^2$. The spatial dependence of the coefficients adds an extra layer of complexity, as the usual arguments

for GARCH-type models must be adapted to allow for smooth spatial variation. Nevertheless, the core idea remains the same: under suitable moment and dependence conditions, the time series behaves like a weakly dependent process for which a CLT holds.

In the following, we investigate the asymptotic normality of the sample mean derived from our process. We begin by explicitly stating the assumptions under which the CLT can be established.

(H1) **Moment and independence conditions:** $\epsilon_t(s)$ is an i.i.d. sequence with $E(\epsilon_t(s)) = 0$, $E(\epsilon_t(s)^2) = 1$, and $E(\epsilon_t(s)^4)$ finite. Moreover we set $E(\epsilon_t(s)^r) = \mu_r(s)$ for $r = 3, 4$. This assumption controls the tail behaviour of the innovations and ensures that all moments required for the CLT exist.

(H2) **Smooth spatial variation:** $\omega_0(s)$, $\{\alpha_i(s); i = 1, \ldots, p\}$, and $\{\beta_i(s); i = 1, \ldots, q\}$ are continuous functions over $[0, 1]^2$. Continuity allows us to treat parameter functions as elements of C_1 and to use uniform convergence arguments in the spatial dimension.

(H3) **Uniform stationarity and stability:** $\sup_s \rho(B(s) \otimes A(s)) = \rho_1 < 1$ and $\sup_s \rho(G(s)) = \rho_2 < 1$, where

$$G(s) = B(s) \otimes B(s) + \sum_{j=1}^{q} A_j(s) \otimes A_j(s).$$

These spectral radius conditions ensure that the volatility recursion is stable and that the bilinear dependence structure satisfies the conditions of the CLT for bilinear processes (Chanda, 1989).

First we put

$$\wp_{T,m} = \sum_{t=1}^{T} \sum_{u=1}^{m} \frac{\sigma_t^2(s_u) - \mu(s_u)}{(Tm)^{1/2}},$$

where $\mu(u_s)$ denotes the mean of $\sigma_t^2(s)$.

(H1) and (H3) allow us to apply the Chanda Central Limit Theorem [16] associated with bilinear time series on

$$\Sigma_n(s) = \sum_{t=1}^{n} \frac{\sigma_t^2(s) - \mu}{n^{1/2}},$$

while we use (H2) to have $\Sigma_n(\cdot) \in C_1$.

Theorem 3.3.1. *Assume that the conditions (H1), (H2) and (H3) hold. Then*

$$\mathcal{L}\left(\wp_{T,m}\right) \to \mathcal{N}\left(0, \eta^2\right),$$

as $T, m \to \infty$*, where*

$$\eta^2 = \lim_{m\to\infty} \frac{1}{m}\left(\sum_{u=1}^{m} \left[\lim_{n\to\infty} \left\{ E\sigma_{1,n}^4\left(s_u\right) + 2\sum_{v=1}^{n} \left|\sigma_{1,n}^2\left(s_u\right)\sigma_{1+v,n}^2\left(s_u\right)\right| \right\} \right] \right.$$

$$\left. +2 \sum_{1\le i<j\le m} cov\Big(\Sigma_\infty(s_i), \Sigma_\infty(s_j)\Big) \right).$$

Proof. see Appendix A.2 □

Remark 3.3.1. *Theorem 3.3.1 states that the joint temporal-spatial average of the volatility process, when properly scaled, converges to a Gaussian distribution with variance* η^2*. This variance incorporates both temporal autocorrelation and spatial heterogeneity. In the special case where* $\alpha_i(s)$ *and* $\beta_j(s)$ *are constant across* s*, the STGARCH model reduces to a standard GARCH model at each location, and the result simplifies to the classical CLT for the sample mean of a stationary GARCH process.*

3.4 The Approximated Spatially Stationary STGARCH(p,q) Process

In this section, we address an important structural property of our spatio-temporal model: although the STGARCH(p, q) process $X_t(s)$ is, in general, spatially nonstationary, it can be approximated under suitable conditions by a spatially stationary process defined around a fixed reference location s_0.

This approximation is crucial for both theory and applications: it allows us to leverage statistical tools designed for stationary processes, also it provides a local-global link: local stationarity around s_0 can serve as a proxy for the true nonstationary dynamics. We begin by noting that the STGARCH(p, q) process in (4) can be expressed in companion form as

$$\boldsymbol{X}_t(s) = \boldsymbol{A}_t(s)\boldsymbol{X}_{t-1}(s) + \boldsymbol{b}_t(s), \tag{8}$$

where

$$\boldsymbol{X}_t(s)^T = \left(\sigma_t^2(s), \ldots, \sigma_{t-q+1}^2(s), X_{t-1}^2(s), \ldots, X_{t-p+1}^2(s)\right),$$
$$\boldsymbol{b}_t(s)^T = (\omega_0(s), 0, \ldots, 0) \in \mathbb{R}^{p+q-2},$$

and

$$\boldsymbol{A}_t(s) = \begin{pmatrix} \Im_t(s) & \beta_q(s) & \underline{\alpha}(s) & \alpha_p(s) \\ \boldsymbol{I}_{q-1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \underline{\epsilon}_{t-1}^2(s) & 0 & \boldsymbol{0} & 0 \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{p-2} & \boldsymbol{0} \end{pmatrix} \in \mathbb{R}^{(p+q-1)\times(p+q-1)}.$$

$\Im_t(s) = \left(\beta_1(s) + \alpha_1(s)\epsilon_{t-1}^2(s), \beta_2(s), \ldots, \beta_{q-1}(s)\right)$, $\underline{\alpha}(s) = (\alpha_2(s), \ldots, \alpha_{p-1}(s))$ and $\underline{\epsilon}_{t-1}^2(s) = \left(\epsilon_{t-1}^2(s), 0, \ldots, 0\right) \in \mathbb{R}^{q-1}$.
By iterating (8), we obtain the moving-average representation

$$\boldsymbol{X}_t(s) = \sum_{k=0}^{\infty} \prod_{l=0}^{k-1} \boldsymbol{A}_{t-l}(s)\boldsymbol{b}_{t-k}(s). \tag{9}$$

In the sequel, we need the following assumptions:

3.5 Model Assumptions

The STGARCH(p,q) process satisfies the representation (8), and for all $s \in [0,1]^2$ the random matrices $\boldsymbol{A}_t(s)$ and the vector $\boldsymbol{b}_t(s)$ satisfy the next assumptions:

Assumption 1. *For each $v \in \{1, \ldots, V\}^2$ where $V \in \mathbb{N}$, there is a sequence of random matrix $\mathcal{A}_t(v)$, that are identically distributed, positive and such that*

$$\sup_s \mid \boldsymbol{A}_t(s) \mid \leq \mathcal{A}_t(v), \text{ for all } s \in \left[\frac{v-1}{V}, \frac{v}{V}\right)^2.$$

and for some $\delta < 0$, the Lyapunov exponent of those matrices is less than δ.

Assumption 2. *There is a stationary sequence of vectors b_t such that*

$$\sup_s \mid \boldsymbol{b}_t(s) \mid < b_t.$$

Assumption 3. *For some $\lambda > 0$; $E(\|\mathcal{A}_t\|_\infty^\lambda) < \infty$ and $E(\|b_t\|_\infty^\lambda) < \infty$.*

Assumption 4. *For all s , $s' \in [0,1]^2$, there exist a $\iota \in (0,1]$ and $\iota' > 0$ such that*

$$\mid \boldsymbol{A}_t(s) - \boldsymbol{A}_t(s') \mid \leq C \mid s - s' \mid^{\iota} \mathcal{A}_t, \quad \textit{and} \quad E[\|\mathcal{A}_t\|^{\lambda}] < \infty.$$

$$\mid \dot{\boldsymbol{A}}_t(s) - \dot{\boldsymbol{A}}_t(s') \mid \leq C \mid s - s' \mid^{\iota'} \mathcal{A}_t, \quad \textit{with} \quad \sup_s \mid \dot{\boldsymbol{A}}_t(s) \mid \leq C\mathcal{A}_t(v).$$

$$\mid \boldsymbol{b}_t(s) - \boldsymbol{b}_t(s') \mid \leq C \mid s - s' \mid^{\iota} b_t.$$

$$\mid \dot{\boldsymbol{b}}_t(s) - \dot{\boldsymbol{b}}_t(s') \mid \leq C \mid s - s' \mid^{\iota'} b_t, \quad \textit{with} \quad \sup_s \mid \dot{\boldsymbol{b}}_t(s) \mid \leq Cb_t.$$

Furthermore, for results requiring the derivative process, we assume the specific case where $\iota = 1$(Lipschitz continuity), and that the spatial derivatives exist and satisfy. This assumption allows us to use Taylor expansion for approximating the spatio-temporal GARCH model, aligning with the methodology that [55] employed for varying-time GARCH models.
To simplify matters, we assume that for $k \geq 1$, $\prod_{j=0}^{-k} \mathcal{A}_j = I$ where I is the identity matrix.

In the following theorem, we aim to demonstrate that the spatially non-stationary process $X_t^2(s)$ could be approximated by a spatially stationary process $X_{t,s_0}^2(s)$ in a fixed location s_0, where

$$\begin{cases} X_{t,s_0}(s) = \sigma_{t,s_0}(s)\epsilon_t(s) \\ \sigma_{t,s_0}^2(s) = \omega_0(s_0) + \sum_{i=1}^{q} \alpha_i(s_0) X_{t-i,s_0}^2(s) + \sum_{j=1}^{p} \beta_j(s_0)\sigma_{t-j,s_0}^2(s), \end{cases}$$

Theorem 3.5.1. *Let $X_t^2(s)$ be a STGARCH(p,q) model that satisfies the representation 4, (9). We define W_t as follows:*

$$W_t = \sum_{k=0}^{\infty} \sum_{v=1}^{V} \prod_{l=0}^{k-1} \mathcal{A}_{t-l}(v) b_{t-k}.$$

If Assumptions 3.5 hold and if for some $\lambda' > 0$,

$$\sup_s E \mid X_t^2(s) \mid^{\lambda'} < \infty, \qquad \sup_s E \mid W_t \mid^{\lambda'} < \infty,$$

Then, we have

$$|X_t^2(s) - X_{t,s_0}^2(s)| \leq \|s - s_0\|_{\infty} N_t,$$

$$N_t = C \sum_{k=0}^{\infty} \sum_{v=1}^{V} \prod_{l=0}^{k-1} \mathcal{A}_{t-l}(v)(\mathcal{A}_{t-k} W_{t-k-1} + b_{t-k}).$$

N_t is a stationary process in both space and time.

Proof. see A.2 □

The fact that N_t is a stationary process in both space and time is supported by a non-stationary extension of [9] (Theorem 1) and [8] (Theorem 2.5), combined with the almost sure convergence of N_t.

Remark 3.5.1. *Theorem 3.5.1 formalizes the idea that the nonstationary process $X_t^2(s)$ can be well-approximated by a stationary process $X_{t,s_0}^2(s)$ in a neighbourhood of s_0. The error is linear in the spatial distance $\|s - s_0\|_\infty$, with a stationary multiplicative factor N_t. This is the spatial analogue of local stationarity concepts in time series analysis.*

To refine the previous approximation, we begin by introducing the concept of a derivative spatial process. This derivative spatial process corresponds to the derivative of the spatially stationary process $\boldsymbol{X}_{t,s_0}(s)$ with respect to s_0 and it is given by:

$$\dot{\boldsymbol{X}}_t(s) = \dot{\boldsymbol{A}}_t(s)\boldsymbol{X}_{t-1}(s) + \boldsymbol{A}_t(s)\dot{\boldsymbol{X}}_{t-1}(s) + \dot{\boldsymbol{b}}_t(s). \tag{10}$$

Let

$$\boldsymbol{X}_{t,2,s_0}(s)^T = \big(\frac{\partial \boldsymbol{X}_{t,s_0}(s)}{\partial s_0}, \boldsymbol{X}_{t,s_0}(s)\big), \quad \boldsymbol{b}_{t,2}(s_0)^T = \big(\partial_{s_0}\boldsymbol{b}_t(s_0), \boldsymbol{b}_t(s_0)\big).$$

Let also

$$\boldsymbol{b}_{t,2} = \begin{pmatrix} C\boldsymbol{b}_t^T \\ \boldsymbol{b}_t^T \end{pmatrix} \qquad \mathcal{A}_{t,2}(v) = \begin{pmatrix} \mathcal{A}_t(v) & C\mathcal{A}_t(v) \\ 0 & \mathcal{A}_t(v) \end{pmatrix} \text{ and}$$
$$\mathcal{A}_{t,2} = \begin{pmatrix} \mathcal{A}_t & C\mathcal{A}_t \\ 0 & \mathcal{A}_t \end{pmatrix}.$$

We suppose that $\boldsymbol{A}_{t,2}(s)$ and $\mathcal{A}_{t,2}(v)$ have negative Lyapunov exponents. Since $\mid \boldsymbol{A}_{t,2}(s) \mid < \mathcal{A}_{t,2}(v)$ and $\sup_s \mid \boldsymbol{b}_{t,2}(s) \mid < \boldsymbol{b}_{t,2}$, we can rewrite (10) as follows:

$$\begin{pmatrix} \frac{\partial \boldsymbol{X}_t(s)}{\partial s} \\ \boldsymbol{X}_t(s) \end{pmatrix} = \boldsymbol{A}_{t,2}(s)\begin{pmatrix} \frac{\partial \boldsymbol{X}_{t-1}(s)}{\partial s} \\ \boldsymbol{X}_{t-1}(s) \end{pmatrix} + \begin{pmatrix} \partial_s\boldsymbol{b}_t(s) \\ \boldsymbol{b}_t(s) \end{pmatrix}, \tag{11}$$

where

$$\boldsymbol{A}_{t,2}(s) = \begin{pmatrix} \boldsymbol{A}_t(s) & \partial_s\boldsymbol{A}_t(s) \\ 0 & \boldsymbol{A}_t(s) \end{pmatrix}.$$

The primary purpose of defining the process outlined in equation (11) is to enable the estimation of the derivative process, in addition $\boldsymbol{X}_{t,2}(s)$ satisfies all the previous conditions. By adding weak assumptions on the moments of $\boldsymbol{X}_{t,2}(s)$ we can apply directly Theorem 3.5.1. Define

$$W_{t,2} = \sum_{k=0}^{\infty} \sum_{v=1}^{V} \prod_{l=0}^{k-1} \mathcal{A}_{t-l,2}(v) \boldsymbol{b}_{t-k,2}.$$

Corollary 3.5.1. *If all the previous assumptions hold and if for some $\lambda' > 0$,*

$$\sup_s E \mid X_{t,2}(s) \mid^{\lambda'} < \infty \quad \textit{and} \quad \sup_s E \mid Y_{t,2} \mid^{\lambda'} < \infty,$$

then,

$$|X_{t,2} - X_{t,2,s_0}(s)| \leq \|s - s_0\|_{\infty} N_{t,2},$$

and $\| N_t \| \leq \| N_{t,2} \|$, with

$$N_{t,2} = C \sum_{k=0}^{\infty} \sum_{v=1}^{V} \prod_{l=0}^{k-1} \mathcal{A}_{t-l,2}(v) (\mathcal{A}_{t-k,2} W_{t-k-1,2} + \boldsymbol{b}_{t-k,2}).$$

$N_{t,2}$ is a stationary process in both space and time.

Proof. see A.2 □

Corollary 3.5.1 quantifies how the derivative process itself can be approximated by a stationary counterpart around s_0. By utilizing equation (10), we have

$$\dot{\boldsymbol{X}}_t(s) = \dot{\boldsymbol{A}}_t(s) \Big(\boldsymbol{A}_{t-1}(s) \boldsymbol{X}_{t-2}(s) + \boldsymbol{b}_{t-1}(s) \Big) + \boldsymbol{A}_t(s) \left(\dot{\boldsymbol{A}}_{t-1}(s) \boldsymbol{X}_{t-2}(s) + \boldsymbol{A}_{t-1}(s) \dot{\boldsymbol{X}}_{t-2}(s) + \dot{\boldsymbol{b}}_{t-1}(s) \right)$$

$$+ \dot{\boldsymbol{b}}_t(s),$$

By iteration, we conclude that

$$\dot{\boldsymbol{X}}_t(s) = \sum_{k=0}^{\infty} \sum_{v=0}^{k-1} \left[\prod_{j=0}^{v-1} \boldsymbol{A}_{t-j}(s) \right] \dot{\boldsymbol{A}}_{t-v}(s) \left[\prod_{j=v+1}^{k-1} \boldsymbol{A}_{t-j}(s) \right] \boldsymbol{b}_{t-k}(s)$$

$$+ \sum_{k=0}^{\infty} \left[\prod_{j=0}^{k-1} \boldsymbol{A}_{t-j}(s) \right] \dot{\boldsymbol{b}}_{t-k}(s).$$

Now by corollary 3.5.1 and since that all the paths of $X_t^2(s)$ are almost surely Lipschitzian with order $(1+\iota')$, we can write the following Taylor expansion of $X_t^2(s)$ around s_0:

$$X_t^2(s) = X_t^2(s_0) + (s - s_0)\,\dot{X}_{t,s_0}(s) + O_p\left(\|s - s_0\|_\infty^2\right). \tag{12}$$

One notable characteristic, easily verifiable, is that the partial derivatives of $\boldsymbol{X}_{t,s_0}(s)$ concerning s_0 generally yield $\dot{\boldsymbol{X}}_{t,s_0}(s)$. We employ this insight below, illustrating that the spatially nonstationary process $\boldsymbol{X}_t(s)$ can be approximated by a linear combination of two spatially stationary processes.

Theorem 3.5.2. *We suppose that $X_t^2(s)$ satisfies model 4 and Assumption 3.5, we have*

$$\left|X_t^2(s) - X_{t,s_0}^2(s) - (s - s_0)\,\dot{X}_{t,s_0}^2(s)\right| \leq \|s - s_0\|_\infty^2\, N_t.$$

Proof. see A.2 □

We observe that Theorem 3.5.2 is a representation of a spatially nonstationary process in terms of spatially stationary processes. Moreover, if s is in the proximity of s_0, then $X_t^2(s)$ can be approximated by $X_{t,s_0}^2(s) + (s - s_0)\dot{X}_{t,s_0}^2(s)$.

Let us introduce the process $\{\tilde{X}_{t,s_0}^2(s)\}_t$. Considering s_0 as a fixed location, we assume that the process $\{\tilde{X}_{t,u_0}^2(u)\}_t$ conforms to the provided representation:

$$\tilde{\boldsymbol{X}}_{t,s_0}(s) = \tilde{\boldsymbol{A}}_{t,s_0}(s)\tilde{\boldsymbol{X}}_{t-1,s_0}(s) + \tilde{\boldsymbol{b}}_{t,s_0}(s),$$

where $\tilde{\boldsymbol{X}}_{t,s_0}(s)^T = (\tilde{\boldsymbol{X}}_{t,s_0}(s), \ldots, \tilde{\boldsymbol{X}}_{t-p+1,s_0}(s))$. The modified matrix $\tilde{\boldsymbol{A}}_{t,s_0}(s)$ is like $\boldsymbol{A}_t(s)$ but with each $\alpha_j(s)$ and $\beta_j(s)$ substituted by $\alpha_j(s_0) + (s - s_0)\dot{\alpha}_j(s_0)$ and $\beta_j(s_0) + (s - s_0)\dot{\beta}_j(s_0)$ respectively.
Additionally, $\tilde{\boldsymbol{b}}_{t,s_0}(s)^T = (\omega_0(s_0) + (s - s_0)\dot{\omega}_0(s_0), 0, \ldots, 0)$. Consequently, it can be shown that the solution to $\tilde{\boldsymbol{X}}_{t,s_0}(s)$ is almost surely given by

$$\tilde{\boldsymbol{X}}_{t,s_0}(s) = \sum_{k=0}^{\infty}\prod_{l=0}^{k-1}\tilde{\boldsymbol{A}}_{t-l,s_0}(s)\tilde{\boldsymbol{b}}_{t-k,s_0}(s).$$

In the upcoming theorem, we demonstrate that $X_t^2(s)$ can also be represented closely by $\tilde{X}_{t,s_0}(s)^2$.

Theorem 3.5.3. *We suppose that $X_t^2(s)$ satisfies model 4 and Assumption 3.5, Assume that* $\sup_s \mid \tilde{\boldsymbol{A}}_t(s) \mid \leq \tilde{\mathcal{A}}_t(v)$ *where the Lyapunov exponent of* $\tilde{\mathcal{A}}_t(v)$ *is less than certain* $\tilde{\lambda} < 0$ *and that* $\sup_s \mid \tilde{\boldsymbol{b}}_t(s) \mid < \tilde{\boldsymbol{b}}_t$. *Moreover, we assume that for some* $\tilde{\lambda}' > 0$

$$E(\|\tilde{\mathcal{A}}_t\|_\infty^{\tilde{\lambda}'}) < \infty \text{ and } E(\|\tilde{\boldsymbol{b}}_t\|_\infty^{\tilde{\lambda}'}) < \infty.$$

Then,

$$|X_t^2(s) - \tilde{X}_{t,s_0}^2(s)| \leq \|s - s_0\|_\infty \tilde{N}_t,$$

where

$$\tilde{N}_t = C \sum_{k=0}^{\infty} \sum_{v=1}^{V} \prod_{l=0}^{k-1} \tilde{\mathcal{A}}_{t-l}(v)(\tilde{\mathcal{A}}_{t-k}\tilde{W}_{t-k-1} + \tilde{b}_{t-k}).$$

$\tilde{N}_t$ *is a stationary process in both space and time.*

Proof. see A.2 □

Corollary 3.5.2. *We suppose that $X_t^2(s)$ satisfies model 4 and all the conditions of Theorem 3.5.3. We have*

$$|\sigma_t^2(s) - \tilde{\sigma}_{t,s_0}^2(s)| \leq \|s - s_0\|_\infty \tilde{N}_{t,1},$$

where

$$\tilde{N}_{t,1} = C \sum_{k=0}^{\infty} \sum_{v=1}^{V} \prod_{l=0}^{k-1} \tilde{\mathcal{A}}_{t-l}(v) \left(\tilde{\mathcal{A}}_{t-k}\tilde{W}_{t-k-1} + \tilde{b}_{t-k}\right) \frac{1}{\epsilon_t(s)}.$$

$\tilde{N}_{t,1}$ *is a stationary process in both space and time.*

Proof. see A.2 □

Remark 3.1. *Theorems 3.5.1–3.5.3 and their corollaries (3.5.1 and 3.5.2) establish that the spatially nonstationary STGARCH(p, q) process can be locally approximated by spatially stationary processes, both in first-order (Theorem 3.5.1) and second-order (Theorem 3.5.2) forms. The introduction of the derivative process $\dot{\boldsymbol{X}}_{t,s_0}(s)$ provides a principled way to construct local linear approximations in s, while the modified parameter process $\tilde{\boldsymbol{X}}_{t,s_0}(s)$ (Theorem 3.5.3) extends this idea to parameters themselves, capturing local variation in the GARCH coefficients.*

From an applied perspective, these results justify the use of localized estimation techniques, such as kernel-weighted likelihood or local polynomial

regression, since they ensure that, for s close to s_0, the nonstationary process behaves similarly to a stationary one with smoothly varying parameters. This local stationarity framework is essential for deriving asymptotic properties of estimators and constructing consistent inference procedures in the spatio-temporal GARCH setting.

3.6 Spatial Approximation of the STGARCH Process under Gaussian Innovations

In this section, we explore an alternative approximation of the process using the framework of [56]. We first consider the case where the innovations are spatially stationary Gaussian process, independent over time. and satisfy the regularity and moment assumptions used in the estimation chapter (see Assumptions 4.1) [2].
We adopt the same notation as in the model definition: for each $s = (x, y) \in [0, 1]^2$, the STGARCH(p, q) field is defined as Assume further that for each fixed s, the innovations $\{\epsilon_t(s)\}$ are Gaussian and that all conditions of Assumption 4.1 hold. Then for each fixed location s, the squared process $X_t^2(s)$ satisfies the following ARMA-type representation: for $r = \max(p, q)$,

$$X_t^2(s) = \omega_0(s) + \sum_{i=1}^{r} \delta_i(s)\, X_{t-i}^2(s) + \sum_{j=1}^{p} \beta_j(s)\, \zeta_{t-j}(s) + \zeta_t(s), \qquad (13)$$

where

$$\delta_i(s) = \alpha_i(s) + \beta_i(s) \quad (\alpha_i(s) = 0 \text{ for } i > q,\ \beta_i(s) = 0 \text{ for } i > p),$$

and

$$\zeta_t(s) = \sigma_t^2(s)\big(\epsilon_t^2(s) - 1\big).$$

Since $E[\epsilon_t^2(s)] = 1$, it follows that $E[\zeta_t(s) \mid \mathcal{F}_{t-1}] = 0$; hence $\{\zeta_t(s)\}$ is a martingale-difference sequence and uncorrelated over time.

The causal solution of (13) is the MA(∞) expansion

$$X_t^2(s) \;=\; \sum_{j=0}^{\infty} \phi_j(s)\, \zeta_{t-j}(s) + C(s), \qquad (14)$$

where $\{\phi_j(s)\}_{j\geq 0}$ satisfy the recursion

$$\phi_j(s) = -\beta_j(s) + \sum_{i=1}^{j} \delta_i(s)\, \phi_{j-i}(s), \qquad j \geq 0, \qquad (15)$$

with $\beta_0(s) = -1$, $\beta_j(s) = 0$ for $j > p$, and $\phi_j(s) = 0$ for $j < 0$. The constant term is

$$C(s) = \omega_0(s)\Big(1 - \sum_{j=1}^{r} \delta_j(s)\Big)^{-1}.$$

Equivalently, defining $\nu_k(s) := \phi_k(s)$, we may write

$$X_t^2(s) = C(s) + \sum_{k=0}^{\infty} \nu_k(s)\, \zeta_{t-k}(s), \tag{16}$$

and introduce the companion matrix

$$A(s) = \begin{pmatrix} \phi_1(s) & \phi_2(s) & \cdots & \phi_{p-1}(s) & \phi_p(s) \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

$$\nu_k(s) = [A(s)^k]_{1,1}.$$

We assume that for some $K > 0$ and $0 < \lambda < 1$,

$$\sup_{s \in [0,1]^2} |\nu_k(s)| \leq K\lambda^k, \qquad k \geq 0. \tag{17}$$

Using (16) and the smoothness of $s \mapsto \nu_k(s)$, we can locally approximate the spatially nonstationary process $X_t^2(s)$ by a spatially stationary process $\widetilde{X}_t^2(s)$ in a neighbourhood of any fixed s_0. This local approximation plays a crucial role in establishing the asymptotic properties of the estimator.

Lemma 3.6.1. *Let $X_t(s)$ follow the STGARCH(p,q) model and satisfy Assumptions 4.1. Suppose that for some $1 - \gamma < \lambda < 1$, $\nu_k(s) = [A(s)^k]_{1,1}$, and that for all k there exists a finite constant $C > 0$ such that:*

(1) $\sup_s |\nu_k(s)| \leq C\lambda^k$,

(2) $\sup_s E(|X_t^2(s)|^{4+v}) < \infty$,

(3) $\sup_{s,v} |\nu_k(s) - \nu_k(v)| \leq Ck\lambda^{k-1}\|s - v\|_\infty$,

(4) *The second partial derivatives of $\nu_k(s)$ exist and satisfy*

$$\left|\frac{\partial \nu_k(x,y)}{\partial x}\right| \leq Ck\lambda^{k-1}, \quad \left|\frac{\partial \nu_k(x,y)}{\partial y}\right| \leq Ck\lambda^{k-1},$$
$$\left|\frac{\partial^2 \nu_k(x,y)}{\partial x^2}\right| \leq Ck^2\lambda^{k-2}, \quad \left|\frac{\partial^2 \nu_k(x,y)}{\partial y^2}\right| \leq Ck^2\lambda^{k-2}, \quad \left|\frac{\partial^2 \nu_k(x,y)}{\partial x \partial y}\right| \leq Ck^2\lambda^{k-2}.$$

Proof. Since there exists $1-\gamma < \lambda < 1$ and finite C such that for all k,

$$\sup_s \|A(s)^k\|_{\text{spec}} \leq C\lambda^k,$$

we have

$$\|X_t^2(s)\|_2 \leq \sum_{k=0}^{\infty} \|A(s)^k\|_{\text{spec}} \|\zeta_{t-k}(s)\|_2 \leq C \sum_{k=0}^{\infty} \lambda^k \|\zeta_{t-k}(s)\|_2.$$

Then (1) follows directly. Since $E(|\epsilon_t^{4+v}(s)|) < \infty$, (2) holds. For (3), using $\nu_k(s) = [A(s)^k]_{1,1}$ and the Lipschitz continuity of $\omega_0(s)$, $\alpha_i(s)$ and $\beta_j(s)$, we have

$$\|A(s) - A(v)\|_2 \leq C\|s-v\|_\infty,$$

and hence

$$\|A(s)^k - A(v)^k\|_2 \leq Ck\lambda^{k-1}\|s-v\|_\infty,$$

yielding (3). Finally, for (4), differentiating $A(s)^k$ and applying similar bounds gives

$$\Big\| \frac{\partial A(s)^k}{\partial x} \Big\|_2 \leq Ck\lambda^{k-1},$$

implying $\big|\frac{\partial \nu_k(x,y)}{\partial x}\big| \leq Ck\lambda^{k-1}$, and similarly for all other partial derivatives up to second order. □

Theorem 3.6.1. *Let $X_t(s)$ follow the STGARCH(p,q) model satisfying Assumptions 4.1. Define the locally stationary approximation around a fixed point s_0 as*

$$\widetilde{X}_t^2(s) = \sum_{k=0}^{\infty} \nu_k(s_0)\, \zeta_{t-k}(s) + C(s_0).$$

Then

$$|X_t^2(s) - \widetilde{X}_t^2(s)| \leq \|s-s_0\|_\infty W_t(s),$$

where for some $1-\gamma \leq \lambda < 1$,

$$W_t(s) = C\sum_{k=0}^{\infty} (k+1)^2 \lambda^k |\zeta_{t-k}(s)| + C_1,$$

and C, C_1 are finite constants independent of s. $W_t(s)$ is stationary in both space and time with finite variance.

Proof. Using the MA(∞) representation and Lemma 3.6.1(3),

$$|X_t^2(s) - \widetilde{X}_t^2(s)| \leq \sum_{k=0}^{\infty} |\nu_k(s) - \nu_k(s_0)| \, |\zeta_{t-k}(s)| + |C(s) - C(s_0)|.$$

By Lipschitz continuity,

$$|X_t^2(s) - \widetilde{X}_t^2(s)| \leq C\|s - s_0\|_\infty \sum_{k=0}^{\infty} (k+1)^2 \lambda^k |\zeta_{t-k}(s)| + C_1 \|s - s_0\|_\infty,$$

hence the result follows. □

Corollary 3.6.1. *Let $X_t(s)$ follow the STGARCH(p, q) model satisfying Assumptions 4.1, and recall that*

$$X_t^2(s) = \sigma_t^2(s)\epsilon_t^2(s).$$

If

$$|X_t^2(s) - \widetilde{X}_t^2(s)| \leq \|s - s_0\|_\infty W_t(s),$$

then

$$|\sigma_t^2(s) - \widetilde{\sigma}_t^2(s)| \leq \|s - s_0\|_\infty N_t(s),$$

where

$$N_t(s) = \frac{W_t(s)}{\epsilon_t^2(s)}.$$

The process $N_t(s)$ is stationary in both space and time with finite variance.

4 Estimation for the STGARCH(p, q) Model

This chapter focuses on the practical implementation of the spatio-temporal GARCH(p, q) model with location-dependent parameters introduced in Chapter 3. Estimating spatio-temporal GARCH models is challenging because the conditional variance process is unobserved and the parameters may vary smoothly across space. We present two kernel-based estimation methods, a local constant and a local linear approach , designed to capture smooth spatial variation while accounting for temporal dependence. In addition to these nonparametric estimators, a likelihood-based Expectation-Maximization (EM) algorithm is later introduced as an alternative framework that treats the volatility process as latent data. Their finite-sample performance is examined through a Monte Carlo study under various spatial smoothness and bandwidth settings. Simulation designs, implementation details, and diagnostic procedures are described in the main text, while technical derivations are deferred to the Appendix. While exact stationarity ensures constant statistical properties over time and space, real-world data often deviate from these strict conditions. Approximate spatial stationarity, however, allows for slight variations, enabling us to model processes that are not perfectly stationary but still exhibit near-stationary behavior under specific conditions. This approach is essential for practical applications, as it offers flexibility when exact stationarity cannot be achieved. This approximate stationarity assumption underlies both the localized estimators and the EM-based method introduced later in this chapter.

In Chapter 3, we established several local stationarity approximation results that form the theoretical foundation for the estimators developed here. Specifically, the spatially nonstationary process $X_t^2(s)$ can be approximated at a fixed location s_0 by a spatially stationary process $X_{t,s_0}^2(s)$; these bounds justify the use of localized estimators. The local constant estimator is based on the first approximation, treating the process as stationary near s_0, while the local linear estimator exploits the second approximation to reduce bias by including spatial derivatives.

The statistical properties of the proposed estimators are examined under two asymptotic scenarios: when m is fixed and only $T \to \infty$, and when both the temporal sample size T and the number of spatial locations m tend to infinity. The first scenario corresponds to practical cases where the spatial network is fixed but the temporal record grows, while the second reflects situations with increasingly dense spatial and temporal observations.

The remainder of this chapter presents the model assumptions, develops the local constant and local linear estimators, evaluates their asymptotic properties, and concludes with the EM algorithm for spatio-temporal GARCH estimation.

4.1 Model Assumptions

To establish the asymptotic properties of the spatiotemporal GARCH process defined in Equation (4), we impose a set of regularity conditions that guarantee temporal stationarity, control the spatial variation of parameters, and ensure the existence of the required moments. We begin with assumptions on the innovation process and then specify the parameter space and smoothness conditions.

Assumption 5. *The innovations $\epsilon_t(s)$ are independent over time and form a strictly stationary process in space.*

This assumption captures the temporal white-noise behavior of the innovations while allowing for spatial dependence, typically modeled through a valid geostatistical covariance function. Spatial stationarity ensures that the dependence structure of the innovations is invariant across locations, which is essential for spatial prediction and for establishing the consistency of the estimators.

Assumption 6. *Let Ω denote the compact parameter space defined as $\Omega = \Big\{(\omega_0(s), \alpha_i(s), \beta_j(s)) : \sum_{i=1}^{q} \alpha_i(s) + \sum_{j=1}^{p} \beta_j(s) \leq 1-\gamma,\ \rho_1 \leq \omega(s) \leq \rho_2,\ 0 < \rho_1 \leq \rho_2 < \infty,\ \gamma > 0\Big\}$, and assume that for each $s \in (0,1]^2$, the parameter vector $\Theta(s) = (\omega_0(s), \alpha_1(s), \ldots, \alpha_p(s), \beta_1(s), \ldots, \beta_q(s))$ lies in the interior of Ω.*

This assumption guarantees that the local GARCH dynamics are well defined and satisfy a uniform stationarity condition, i.e., $\sum_i \alpha_i(s) + \sum_j \beta_j(s)$ is bounded away from one for all spatial locations. Consequently, the conditional variance $\sigma_t^2(s)$ is uniformly bounded away from zero, preventing degeneracy. Furthermore, smoothness of the parameter functions across space is required to justify local approximations by stationary GARCH models and to ensure the uniform convergence of kernel-based estimators.

Assumption 7. *The functions $\omega_0(s)$, $\alpha_i(s)$, and $\beta_j(s)$ admit continuous second-order partial derivatives with respect to the spatial coordinates, and all such derivatives are uniformly bounded over $[0,1]^2$.*

This condition imposes spatial smoothness on the parameter surfaces, a key requirement for local estimation and for applying Taylor expansions in the asymptotic analysis.

Assumption 8. *For some $v > 0$, the innovation process satisfies*

$$E\big(|\epsilon_t(s)|^{4+v}\big) < \infty.$$

While the existence of fourth moments is sufficient for standard GARCH processes, the locally weighted quasi-likelihood estimation employed here involves martingale difference arrays, for which a slightly stronger moment condition is required [38][see][Theorem 3.2].

Assumptions 6–8 collectively ensure the asymptotic normality of the estimators introduced below. In practice, if $\epsilon_t(s)$ follows a Gaussian process with mean zero and a exponential covariance function, all these assumptions are automatically satisfied since all moments exist and the spatial smoothness conditions hold.

4.2 Localized Least Squares Estimation

The estimation of spatially varying parameters in spatio-temporal GARCH-type models presents two main challenges. First, the parameters $\omega_0(s)$, $\alpha_j(s)$, and $\beta_j(s)$ are unknown smooth functions over space, but data are available only at a finite set of locations $\{s_u\}_{u=1}^m$. Second, for an arbitrary unobserved site s_0, there are no direct observations $\{X_t^2(s_0)\}_{t=1}^T$, so standard global estimation methods such as maximum likelihood or global least squares cannot be applied directly.

To address these challenges, we adopt *localized nonparametric* estimation methods. The key idea is to exploit the smoothness of the parameter functions: if s_u is close to s_0, then $\mathcal{B}(s_u)$ should be close to $\mathcal{B}(s_0)$. Thus, information from neighboring sites can be borrowed, with observations weighted according to their spatial proximity to s_0. This is achieved through a *kernel weighting scheme*, which assigns larger weights to closer locations and smaller weights to more distant ones.

Localized least squares is a natural choice in this context for several reasons:

- It provides a computationally efficient, closed-form solution similar to ordinary least squares, making it feasible for large spatio-temporal datasets.
- It adapts automatically to local variation in the parameters without requiring a parametric model for spatial dependence.
- It performs well in small-sample settings, where quasi-maximum likelihood methods may be unstable or biased.

Moreover, the local least squares framework can be extended from a *local constant* form, which assumes parameters are constant within the neighborhood of s_0, to a *local linear* form, which uses a first-order Taylor expansion in s to reduce bias and improve performance near the boundaries of the spatial domain [25]. This flexibility makes it a powerful tool for our STGARCH(p, q) setting, where parameter surfaces may change smoothly but non-uniformly over space.

The starting point is the representation of the STGARCH(p, q) model:

$$X_t^2(s) = \omega_0(s) + \sum_{j=1}^{p} \alpha_j(s) X_{t-j}^2(s) + \sum_{j=1}^{q} \beta_j(s) \sigma_{t-j}^2(s) + \sigma_t^2(s)(\epsilon_t^2(s) - 1). \quad (18)$$

In the local constant approach, we approximate the parameter vector $\mathcal{B}(s)$ by a constant $\mathcal{B}(s_0)$ within a neighborhood of s_0. This leads to the *localized least squares criterion*:

$\mathcal{L}_{1,T}(s_0, \mathcal{B}) =$

$$\frac{1}{mT'} \sum_{t=p+1}^{T} \sum_{u=1}^{m} W_b(s_0 - s_u) \left\{ X_t^2(s_u) - \left(\omega_0(s_u) + \sum_{j=1}^{p} \alpha_j(s_u) X_{t-j}^2(s_u) + \sum_{j=1}^{p} \beta_j(s_u) \sigma_{t-j}^2(s_u) \right) \right\}^2, \quad (19)$$

where $T' = T - p - 1$ and $K : \mathbb{R}^2 \to \mathbb{R}$ is a kernel function with

$$\int K(x) \mathrm{d}x = 1, \quad W(s) = K(x)K(y) \text{ and } W_b(.) = \frac{1}{b^2} W\left(\frac{\cdot}{b}\right).$$

Let $\mathcal{B}(s_0)' = (\omega_0(s_0), \alpha_1(s_0) \ldots, \alpha_p(s_0), \beta_1(s_0), \ldots \beta_p(s_0))$, where $(\alpha_{q+1} = 0, \alpha_{q+2} = 0, \ldots, \alpha_p = 0)$. We use $\hat{\mathcal{B}}_{1,T}(s_0)$ as an estimator of $\mathcal{B}(s_0)$, where

$$\hat{\mathcal{B}}_{1,T}(s_0) = \arg\min_{\mathcal{B}} \mathcal{L}_{1,T}(s_0, \mathcal{B}).$$

$\hat{\mathcal{B}}_{1,T}\left(s_0\right)' = \left(\hat{\omega}_{1,0}\left(s_0\right), \hat{\alpha}_{1,1}(s_0) \ldots, \hat{\alpha}_{1,p}\left(s_0\right), \hat{\beta}_{1,1}(s_0), \ldots \hat{\beta}_{1,p}(s_0)\right)$ is the local least squares estimator.
By evaluating $\hat{\mathcal{B}}_{1,T}(s_0)$ we get :

$$\hat{\mathcal{B}}_{1,T}(s_0) = \hat{J}_{1,T}^{-1}(\underline{s})\ \hat{j}_{1,T}(\underline{s}),\ \text{where}\ \ \underline{s} = (s_0, \ldots, s_m)$$

and

$$\hat{J}_{1,T}(\underline{s}) = \frac{1}{mT'} \sum_{u=1}^{m} W_b\left(s_0 - s_u\right) \otimes \sum_{t=p+1}^{T} \underline{\Theta}_{t-1}(s_u)\ \underline{\theta}_{t-1}^{(2)'}(s_u)$$

$$\hat{j}_{1,T}(\underline{s}) = \frac{1}{mT'} \sum_{u=1}^{m} W_b\left(s_0 - s_u\right) \otimes \sum_{t=p+1}^{T} X_t^2(s_u) \underline{\Theta}_{t-1}(s_u) \tag{20}$$

while $\otimes$ denotes the Kronecker product,

$$\underline{\Theta}_{t-1}(s_u) = (1,\ \underline{\theta}_{t-1}^{(1)},\ \underline{\theta}_{t-1}^{(3)})',$$

$$\underline{\theta}_{t-1}^{(1)} = (\sigma_{t-1}^2(s_u)\epsilon_{t-j}^2(s_u), \ldots, \sigma_{t-p}^2(s_u)\epsilon_{t-j}^2(s_u))',$$

$$\underline{\theta}_{t-1}^{(2)} = (\sigma_{t-1}^2(s_u)(\epsilon_{t+1-j}^2(s_u) + 1) + \frac{1}{p}, \ldots, \sigma_{t-p}^2(s_u)(\epsilon_{t+p-j}^2(s_u) + 1) + \frac{1}{p})',$$

.

$$\underline{\theta}_{t-1}^{(3)} = (\sigma_{t-1}^2(s_u), \ldots, \sigma_{t-p}^2(s_u)).$$

This estimator is equivalent to a locally weighted OLS fit of the conditional variance equation, with weights $W_b(s_0 - s_u)$ emphasizing sites near s_0.

While the local constant approach is simple, it may suffer from bias if the parameters vary significantly within the neighborhood of s_0. A standard bias reduction technique in nonparametric regression is to allow a first-order Taylor expansion in s (Fan, 1993) [25].

Under Assumption (iii), we expand:

$$\omega_0(s_u) = \omega_0(s_0) + (s_0 - s_u)\dot{\omega}_0(s_0) + O(\|s_u - s_0\|_\infty^2),$$

$$\alpha_j(s_u) = \alpha_j(s_0) + (s_0 - s_u)\dot{\alpha}_j(s_0) + O(\|s_u - s_0\|_\infty^2),$$
$$\beta_j(s_u) = \beta_j(s_0) + (s_0 - s_u)\dot{\beta}_j(s_0) + O(\|s_u - s_0\|_\infty^2).$$

Substituting these expansions into the local least squares criterion yields:

$$\mathcal{L}_{2,T}(s_0,\mathcal{B}_2)=\frac{1}{mT'}\sum_{t=p+1}^{T}\sum_{u=1}^{m}W_b(s_0-s_u)\Bigg\{X_t^2(s_u)-\Big((\omega_0(s_0)+(s_0-s_u)\dot{\omega}_0(s_0))$$

$$+\sum_{j=1}^{p}(\alpha_j(s_0)+(s_0-s_u)\dot{\alpha}_j(s_0))X_{t-j}^2(s_u)+\sum_{j=1}^{p}(\beta_0(s_0)+(s_0-s_u)\dot{\beta}_0(s_0))\sigma_{t-j}^2(s_u)]\Big)\Bigg\}^2,$$

where $\dot{\omega}_0'=(\dot{\omega}_{10},\dot{\omega}_{20})$, $\dot{\alpha}_j'=(\dot{\alpha}_{1j},\dot{\alpha}_{2j})$,$\dot{\beta}_j'=(\dot{\beta}_{1j},\dot{\beta}_{2j})$. and

$$\mathcal{B}_2(s_0)'=(\omega_0(s_0),\dot{\omega}_0(s_0),\underline{\alpha}(s_0),\dot{\alpha}_1(s_0),\ldots,\dot{\alpha}_p(s_0),\underline{\beta}(s_0),\dot{\beta}_1(s_0),\ldots,\dot{\beta}_p(s_0))$$

$\mathcal{B}_2$ is a $9\times p$-dimensional vector. In addition $\underline{\alpha}'(s_0)=(\alpha_1(s_0)\ldots,\alpha_p(s_0))$ and $\underline{\beta}'(s_0)=(\beta_1(s_0),\ldots\beta_p(s_0))$. We use $\hat{\mathcal{B}}_{2,T}(s_0)$ as an estimator of $\mathcal{B}_2(s_0)$, where

$$\hat{\mathcal{B}}_{2,T}(s_0)=\arg\min_{\mathcal{B}_2}\mathcal{L}_{2,T}\left(s_0,\mathcal{B}_2\right).$$

$$\hat{\mathcal{B}}_{2,T}(s_0)'=(\hat{\omega}_{2,0}(s_0),\hat{\dot{\omega}}_{2,0}(s_0),\underline{\hat{\alpha}}_{2,T}(s_0),\hat{\dot{\alpha}}_{2,1}(s_0),\ldots,\hat{\dot{\alpha}}_{2,p}(s_0),\underline{\hat{\beta}}_{2,T}(s_0),\hat{\dot{\beta}}_{2,1}(s_0),\ldots,\hat{\dot{\beta}}_{2,p}(s_0)),$$

with $\underline{\hat{\alpha}}_{2,T}(s_0)$ is an estimator of $\underline{\alpha}(s_0)$ and $\underline{\hat{\beta}}_{2,T}(s_0)$ is an estimator of $\underline{\beta}(s_0)$.
We call $\hat{\mathcal{B}}_{2,T}(s_0)$ the linear local least squares estimator.
This approach estimates both the level and gradient of $\mathcal{B}(s)$ at s_0, leading to reduced bias, especially near boundaries or in regions where the parameters change rapidly. By evaluating $\hat{\mathcal{B}}_{2,T}(s_0)$, we get:

$$\hat{\mathcal{B}}_{2,T}(s_0)=\hat{J}_{2,T}^{-1}(\underline{s})\;\hat{j}_{2,T}(\underline{s}),\;\text{where}\;\;\underline{s}=(s_0,\ldots,s_m),$$

$$\hat{J}_{2,T}(\underline{s})=\frac{1}{mT'}\sum_{u=1}^{m}W_b\left(s_0-s_u\right)S(u)\;S'(u)\otimes\sum_{t=p+1}^{T}\underline{\Theta}_{t-1}(s_u)\;\underline{\theta}_{t-1}^{(2)'}(s_u)$$

$$\hat{\hat{j}}_{2,T}(\underline{s})=\frac{1}{mT'}\sum_{u=1}^{m}W_b\left(s_0-s_u\right)S(u)\otimes\sum_{t=p+1}^{T}X_t^2(s_u)\underline{\Theta}_{t-1}(s_u),\qquad(21)$$

with $S(u) = (1, s_u - s_0)'$.
Upon the study of equations (20) and (21), when the observed locations are closely grouped, the matrices $\hat{J}_{1,T}(\underline{s})$ $\hat{J}_{2,T}(\underline{s})$ possibly approach singularity. To prevent this problem, we add to the Hessian matrix a small diagonal matrix for ensuring the stability of the estimator.

This common technique in linear regression maintains the invertibility, and guarantees the minimum eigenvalue of the diagonal matrix to be less than the minimum of the obtained Hessian matrix. We apply this technique on the previous matrices, which gives the following criterions .

$$\hat{\mathcal{B}}_{3,T}(s_0) = [\hat{J}_{1,T}(\underline{s}) + \gamma_1^{\nu} I]^{-1} \hat{j}_{1,T}(\underline{s})$$
$$\hat{\mathcal{B}}_{4,T}(s_0) = [\hat{J}_{2,T}(\underline{s}) + \gamma_2^{\nu} I]^{-1} \hat{j}_{2,T}(\underline{s}),$$

where γ_1 and γ_2 are small positive constants (ridge parameters) that ensure the minimum eigenvalue of the inverted matrix is bounded away from zero. This guarantees the existence of the estimator without affecting its asymptotic consistency, provided $\gamma \to 0$ as the sample size increases. I is the identity matrix and $\nu > 0$. Let

$$\hat{\mathcal{B}}_{3,T}(s_0) = \left(\hat{\omega}_{3,0}(s_0), \hat{\alpha}_{3,1}(s_0) \ldots, \hat{\alpha}_{3,p}(s_0), \hat{\beta}_{3,1}(s_0), \ldots \hat{\beta}_{3,p}(s_0)\right)$$

$$\hat{\mathcal{B}}_{4,T}(s_0) = \left(\hat{\omega}_{4,0}(s_0), \hat{\dot{\omega}}_{4,0}(s_0), \underline{\hat{\alpha}}_{4,T}(s_0), \hat{\dot{\alpha}}_{4,1}(s_0), \ldots, \hat{\dot{\alpha}}_{4,p}(s_0), \underline{\hat{\beta}}_{4,T}(s_0), \hat{\dot{\beta}}_{4,1}(s_0), \ldots, \hat{\dot{\beta}}_{4,p}(s_0)\right)$$

$\underline{\hat{\alpha}}_{4,T}(s_0)$ is an estimator of $\underline{\alpha}(s_0)$ and $\underline{\hat{\beta}}_{4,T}(s_0)$ is an estimator of $\underline{\beta}(s_0)$. This technique simplifies the calculation of the estimator and ensure that the resulting estimator is consistent. The resulting estimators $\hat{\mathcal{B}}_{1,T}(s_0)$ and $\hat{\mathcal{B}}_{2,T}(s_0)$ form the basis for our asymptotic analysis in the next section.

4.3 Properties of the estimators

In this section, we examine the asymptotic properties of the estimators introduced in the previous subsection. Our aim is to understand their bias, variance, and limiting distribution under different asymptotic regimes. In particular, we distinguish between the following two settings:

(i) The temporal sample size $T \to \infty$ while the number of spatial locations m is fixed.

(ii) Both $T \to \infty$ and $m \to \infty$ (infill asymptotics).

The first regime corresponds to the classical time series framework, in which we observe long sequences over a fixed number of sites. In contrast, the second regime arises naturally in spatial statistics, where additional sites are sampled within a fixed spatial domain as T grows. This is known as the infill asymptotic framework (see, for example,[46]).

We begin by considering the fixed m setting. Let s_0 be a fixed location in the spatial domain $[0,1]^2$. To analyze local behavior around s_0, we define the approximated spatially stationary process $X^2_{t,s_0}(s)$ by

$$\begin{cases} X_{t,s_0}(s) = \sigma_{t,s_0}(s)\epsilon_t(s) \\ \sigma^2_{t,s_0}(s) = \omega_0(s_0) + \sum_{i=1}^{q} \alpha_i(s_0) X^2_{t-i,s_0}(s) + \sum_{j=1}^{p} \beta_j(s_0)\sigma^2_{t-j,s_0}(s) \end{cases} \tag{22}$$

where the parameters $\omega_0(s_0)$, $\alpha_j(s_0)$, and $\beta_j(s_0)$ are kept constant over space. By construction, the process $\{X^2_{t,s_0}(s) : s \in [0,1]^2\}$ is spatially stationary.

We then define $\mathscr{B}_{t,1}(s_0, \mathcal{B}) = \frac{1}{mT'} \sum_{t=p+1}^{T} \sum_{u=1}^{m} W_b(s_0 - s_u) \times$

$$\left(\left\{ X^2_t(s_u) - \left(\omega_0(s_u) + \sum_{j=1}^{p} \alpha_j(s_u) X^2_{t-j}(s_u) + \sum_{j=1}^{p} \beta_j(s_u)\sigma^2_{t-j}(s_u) \right) \right\}^2 \right.$$
$$\left. - \left\{ X^2_{t,s_0}(s_u) - \left(\omega_0(s_0) + \sum_{j=1}^{p} \alpha_j(s_0) X^2_{t-j,s_0}(s_u) + \sum_{j=1}^{p} \beta_j(s_0)\sigma^2_{t-j,s_0}(s_u) \right) \right\}^2 \right). \tag{23}$$

We observe that for a fixed s_0, the parameters $\omega_0(s_0), \alpha_j(s_0),$ and $\beta_j(s_0)$ remain constant over space. Therefore, the process $\{X^2_{t,s_0}(s) : s \in [0,1]^2\}$ is spatially stationary. Now, if $X^2_t(s) = X^2_{t,s_0}(s)$ for all $s \in [0,1]^2$, even with a fixed m, the estimator $\hat{\mathcal{B}}_{1,T}(s_0)$ is asymptotically unbiased and $\mathscr{B}_{t,1}(s_0, \mathcal{B}) = 0$. However, in the general case when $X^2_t(s) \neq X^2_{t,s_0}(s)$, we demonstrate in Theorem 4.3.1 that the bias of the estimator arises from the difference $X^2_t(s) - X^2_{t,s_0}(s)$.

Theorem 4.3.1. *Let $X_t(s)$ be the process defined model in 4 which satisfies Assumption 4.1. We have*

$$(i)\left(\hat{\mathcal{B}}_{1,T}(s_0)-\mathcal{B}(s_0))\right)=\frac{1}{2}\Lambda_1(s_0)^{-1}\left(\frac{-4}{m}\sum_{u=1}^{m}W_b\left(s_0-s_u\right)\mathcal{B}(s_0)\mathbb{E}\left[\dot{\underline{\Theta}}_{t-i,s_0}(s_u)\underline{\Theta}_{t-j,s_0}(s_u)^T\right]\right)$$
$$+O_p\left(\frac{C}{m}\sum_{u=1}^{m}W_b\left(s_0-s_u\right)\|s_u-s_0\|_\infty^2\right). \tag{24}$$

(ii)

$$\sqrt{T'}\left(\left(\hat{\mathcal{B}}_{1,T}(s_0)-\mathcal{B}(s_0))\right)+\frac{1}{2}\Lambda_1(s_0)^{-1}\nabla\mathscr{B}_{t_1}\left(s_0,\mathcal{B}\right)\right)\quad\xrightarrow{D}\mathcal{N}\left(0,\Lambda_1(s_0)\Upsilon(s_0)\Lambda_1(s_0)^{-1}\right), \tag{25}$$

where

$$\Lambda_1(s_0)=\frac{2}{m}\sum_{u=1}^{m}W_b(s_0-s_u)\ \mathbb{E}(\underline{\Theta}_0(s_u)\underline{\Theta}_0(s_u)^T) \tag{26}$$

and

$$\Upsilon(s_0)=\frac{4}{m^2}\sum_{u=1}^{m}W_b^2(s_0-s_u)\ var(\underline{\epsilon}_t^2(u_s))\ \mathbb{E}(\underline{\Theta}_0(s_u)\underline{\Theta}_0(s_u)^T)\mathbb{E}(\sigma_0^4(s_u)) \tag{27}$$

Proof. see A.3 □

Remark 4.3.1. *Part (i) of Theorem 4.3.1 identifies the leading bias term in the fixed-m setting, which depends on spatial heterogeneity through $\mathscr{B}_{t,1}(s_0,\mathcal{B})$. The $O_p(\cdot)$ remainder term reflects the additional bias induced by spatial distances $\|s_u-s_0\|_\infty^2$ weighted by the kernel $W_b(\cdot)$. Part (ii) establishes asymptotic normality, with $\Lambda_1(s_0)$ serving as a local information matrix and $\Upsilon(s_0)$ as the variance of the local score function. This structure parallels standard results in local polynomial regression*

The next result provides analogous properties for the linear local least squares estimator $\hat{\mathcal{B}}_{2,T}(s_0)$.

Theorem 4.3.2. *Let $X_t(s)$ be the process defined in 4 which satisfies Assumption 4.1.*
We have

$$\left\{\hat{\mathcal{B}}_{2,T}(s_0)-\mathcal{B}(s_0)\right\}=O_p\left\{\Lambda_2\left(s_0\right)^{-1}\left(\frac{1}{m}\sum_{u=1}^{m}W_b\left(s_0-s_u\right)\|s_u-s_0\|_\infty\right)^2\right\}, \tag{28}$$

where

$$\Lambda_2(s_0) = \frac{1}{m}\sum_{u=1}^{m} W_b(s_0 - s_u)\, S(u)\; S'(u) \otimes \mathbb{E}(\underline{\Theta}_0(s_u)\; \underline{\theta}_0^{(2)'}(s_u)). \tag{29}$$

Proof. see A.3 □

Remark 4.3.2. *Theorem 4.3.2 shows that, for fixed m, the bias of the linear local least squares estimator is of order $\|s_u - s_0\|_\infty^2$ on average, again weighted by the kernel function. The matrix $\Lambda_2(s_0)$ plays the same role as $\Lambda_1(s_0)$ in Theorem 4.3.1, encoding the local design and variability of the regressors. It is important to highlight that the asymptotic normality of $\hat{\mathcal{B}}_{2,T}(s_0)$ can be established, and the method of the proof is the same as that of Theorem 4.3.1*

In the initial case, we have demonstrated that as $T \to \infty$ while m remains fixed, the estimator can exhibit a significant bias. Infill asymptotics are commonly employed in spatial statistics, where the number of locations grows, to establish the consistency of an estimator (see Mukherjee and Lahiri, 2004 [46]). To demonstrate the consistency of $\hat{\mathcal{B}}_{1,T}(s_0)$ as an estimator of $\mathcal{B}(s_0)$, we employ a similar infilling argument. Theorem 4.3.3 addresses the scenario when $m \to \infty$ (the number of sites grows) as $T \to \infty$. In particular, we demonstrate that $\hat{\mathcal{B}}_{1,T}(s_0)$ is a consistent estimator of $\mathcal{B}(s_0)$. Analogous results hold for $\hat{\mathcal{B}}_{2,T}(s_0)$.

Theorem 4.3.3. *Let $X_t(s)$ be the process defined in 4 which satisfies Assumptions 4.1.*
We have

$$\hat{\mathcal{B}}_{1,T}(s_0) \xrightarrow{\mathcal{P}} \mathcal{B}(s_0), \tag{30}$$

when $T \to \infty$ and $m \to \infty$.

Proof. see A.3 □

4.4 Empirical Bayesian Inference via Iterative Localized EM

While the localized least squares estimators developed in the previous section provide consistent point estimates, they do not explicitly account for the probabilistic uncertainty associated with the unobserved location. To address this, we propose an Empirical Bayesian framework implemented via an Iterative Localized Expectation-Maximization (LEM) algorithm.

In this framework, the unobserved squared returns at the target location s_0 are treated as latent variables. Unlike standard Maximum Likelihood Estimation (MLE), which relies solely on the target data, our approach constructs a spatial informative prior derived from neighboring observations. This allows us to estimate the posterior mode of the parameters by borrowing strength from the local spatial structure.

We define the estimation problem at a fixed target location s_0 using the following components:

- **Observed Information ($\mathcal{X}$):** The set of squared returns at the m neighboring locations, which form the basis of the empirical prior:

$$\mathcal{X} = \{X_t^2(s_u) : t = 1, \ldots, T; u = 1, \ldots, m\}.$$

- **Latent Variables ($\mathcal{Z}$):** The unobserved squared returns at the target location, which constitute the missing likelihood component:

$$\mathcal{Z} = \{X_t^2(s_0) : t = 1, \ldots, T\}.$$

- **Parameter Vector (Θ):** The local GARCH parameters at s_0:

$$\Theta(s_0) = (\omega(s_0), \alpha_1(s_0), \ldots, \alpha_q(s_0), \beta_1(s_0), \ldots, \beta_p(s_0))'.$$

We formulate the problem as finding the Maximum A Posteriori (MAP) estimator. The posterior density is proportional to the product of the likelihood and the prior:

$$P(\Theta \mid \mathcal{Z}, \mathcal{X}) \propto P(\mathcal{Z} \mid \Theta) \cdot \pi(\Theta \mid \mathcal{X}), \tag{31}$$

where $P(\mathcal{Z} \mid \Theta)$ represents the complete-data likelihood of the target location, and $\pi(\Theta \mid \mathcal{X})$ is the **empirical spatial prior**.

To enforce the hypothesis of local stationarity, we construct this prior as a weighted composite likelihood of the neighboring observations. Mathematically, the log-prior is defined as:

$$\ln \pi(\Theta \mid \mathcal{X}) \propto \sum_{u=1}^{m} W_b(s_0 - s_u) \ln L(s_u; \Theta), \tag{32}$$

where $L(s_u; \Theta)$ is the likelihood of the u-th neighbor evaluated at the candidate parameter Θ, and $W_b(\cdot)$ is a spatial kernel function. This formulation ensures that the prior density is highest in regions of the parameter space that are consistent with the surrounding spatial field.

Since the target process $\mathcal{Z}$ is unobserved, we cannot maximize the posterior directly. Instead, we utilize the EM algorithm to maximize the *expected* log-posterior density. The algorithm alternates between imputing the latent variables (E-step) and updating the parameter estimates (M-step).

Step 0: Initialization. To ensure the algorithm begins in a high-probability region of the posterior, we elicit a starting value $\Theta^{(0)}(s_0)$ using the Linear Local Least Squares estimator $\hat{\mathcal{B}}_{2,T}(s_0)$ derived in Section 4.2.

Step 1: The E-Step (Bayesian Imputation). In the k-th iteration, we compute the conditional expectation of the latent variables given the observed spatial data and the current parameter estimates. Since the squared process X_t^2 admits a linear spatial representation [2], we obtain the latent squared returns via the Best Linear Predictor (Spatial Kriging):

$$\hat{X}_t^{2,(k)}(s_0) = \mathbb{E}\left[X_t^2(s_0) \mid \mathcal{X}, \Theta^{(k)}\right]. \tag{33}$$

Simultaneously, we reconstruct the latent conditional variance. Due to the recursive non-linearity of the GARCH process, exact evaluation of the expected log-likelihood is intractable. We therefore employ a deterministic filter approximation:

$$\hat{\sigma}_t^{2,(k)}(s_0) = \omega^{(k)} + \sum_{i=1}^{q} \alpha_i^{(k)} \hat{X}_{t-i}^{2,(k)}(s_0) + \sum_{j=1}^{p} \beta_j^{(k)} \hat{\sigma}_{t-j}^{2,(k)}(s_0). \tag{34}$$

This step generates the imputed "pseudo-data" necessary to evaluate the target likelihood.

Step 2: The M-Step (Posterior Maximization). In the M-step, we update the parameter vector by minimizing the negative expected log-posterior. The objective function, $Q(\Theta \mid \Theta^{(k)})$, balances the fit to the imputed target data against the information provided by the spatial neighbors:

$$\Theta^{(k+1)}(s_0) = \arg\min_{\Theta\in\Omega} \left\{ \sum_{t=p+1}^{T} \left(\ln \sigma_t^2(s_0,\Theta) + \frac{\widehat{X}_t^{2,(k)}(s_0)}{\sigma_t^2(s_0,\Theta)} \right) + \sum_{u=1}^{m} W_b(s_0 - s_u) \sum_{t=p+1}^{T} \left(\ln \sigma_t^2(s_u,\Theta) + \frac{X_t^2(s_u)}{\sigma_t^2(s_u,\Theta)} \right) \right\}. \tag{35}$$

where $\sigma_t^2(s,\Theta)$ denotes the volatility calculated at location s using the *candidate* parameters Θ.

This formulation explicitly justifies the Empirical Bayesian interpretation: the first term updates beliefs based on the latent target dynamics, while the second term (the prior) penalizes deviations from the local spatial structure.

The E-step and M-step are iterated until the parameter estimates stabilize:

$$\|\Theta^{(k+1)}(s_0) - \Theta^{(k)}(s_0)\| < \epsilon, \tag{36}$$

where ϵ is a pre-defined tolerance level. The resulting estimator $\hat{\Theta}_{MAP}(s_0)$ represents the mode of the posterior distribution.

By embedding the spatial information within the M-step as an empirical prior, the algorithm ensures that the parameters are not only consistent with the imputed temporal dynamics at s_0 but are also regularized by the surrounding spatial field. The iterative nature allows the imputation (E-step) and the parameter estimation (M-step) to mutually reinforce each other, providing a robust solution to the missing data problem.

5 Simulation

In this chapter, we empirically validate the theoretical properties of the proposed estimators, we conduct a Monte Carlo simulation study. The objective is to evaluate the finite-sample performance of the Localized Least Squares (LLS), Local Linear Least Squares (LLLS), and the proposed Iterative Localized Expectation-Maximization (EM) estimators in recovering the true structural parameters of the STGARCH process.

We generate the spatio-temporal process on a spatial domain $S = [0, 1]^2$. The data generating process (DGP) is defined by a STGARCH(1,1) structure where the parameters vary smoothly across space according to the following deterministic functions:

$$\begin{aligned}
\omega(x, y) &= 0.05 + 0.04 \sin(0.5x) \cos(0.9y), \\
\alpha(x, y) &= 0.25 \left(1 + \sin(0.7x) \cos(0.1y)\right), \\
\beta(x, y) &= 0.25 \left(1 + \sin(0.08x) \cos(0.2y)\right).
\end{aligned}$$

The innovation process $\varepsilon_t(s)$ is generated from a multivariate Gaussian distribution with mean zero and unit marginal variance. To introduce realistic spatial dependence, we impose a spatial covariance structure defined by an exponential structure :

$$\text{Cov}(\varepsilon_t(s_i), \varepsilon_t(s_j)) = \exp\left(-\frac{\|s_i - s_j\|}{0.3}\right), \tag{37}$$

where the spatial range parameter is set to 0.3.

The first component of the simulation study assesses the consistency and accuracy of the estimators. We evaluate the models under two asymptotic regimes: a moderate sample size S1(training data) ($T = 200, n_1 = 50$) and a larger sample size ($T = 500, n_1 = 100$).

For each scenario, we perform $MC = 100$ independent replications. In each replication, estimation is performed at $n_2 = 5$ randomly selected from S_2 locations. This setup rigorously tests the ability of the models to predict parameter surfaces at locations where no data was collected.

Performance is measured using the Root Mean Squared Error (RMSE) averaged over S_2:

$$\text{RMSE}_\theta = \sqrt{\frac{1}{MC \cdot n_2} \sum_{k=1}^{MC} \sum_{i=1}^{n_2} \left(\hat{\theta}_k(s_i) - \theta_{true}(s_i)\right)^2}, \tag{38}$$

where $\theta \in \{\omega, \alpha, \beta\}$.

The bandwidths for the kernel-based methods are selected using a standard rule-of-thumb, $h \propto n^{-1/6}$, adjusted to $h_{LLS} = 0.40 n_1^{-1/6}$ and $h_{LLLS} = 0.65 n_1^{-1/6}$ to balance bias and variance.

Table 1 summarizes the simulation results for the three parameters across the different sample sizes.

Table 1: RMSE of Parameter Estimation for LLS, LLLS, and EM methods at unobserved locations.

Sample Size		**Localized LSE (LLS)**			**Local Linear (LLLS)**			**Proposed EM**		
T	n_1	ω	α	β	ω	α	β	ω	α	β
200	50	0.0126	0.0593	0.1053	0.0125	0.0577	0.0911	**0.0094**	**0.0441**	**0.0793**
500	50	0.0091	0.0433	0.0785	0.0085	0.0432	0.0688	**0.0060**	**0.0305**	**0.0500**
200	100	0.0115	0.0541	0.0947	0.0107	0.0528	0.0826	**0.0088**	**0.0432**	**0.0731**
500	100	0.0083	0.0405	0.0684	0.0079	0.0366	0.0605	**0.0060**	**0.0272**	**0.0473**

The results in Table 1 highlight several key findings. Most notably ,the proposed Iterative Localized EM estimator consistently yields the lowest RMSE for all parameters across all sample sizes. For example, in the largest sample setting ($T = 500, n = 100$), the EM method reduces the error for the β parameter to 0.0473, compared to 0.0605 for the Local Linear method. This improvement is attributed to the EM algorithm's ability to iteratively reconstruct the latent volatility process, rather than relying on noisy squared returns as proxies. Consistent with nonparametric regression theory, the Local Linear (LLLS) estimator outperforms the Local Constant (LLS) estimator, particularly for the β parameter. The LLLS method is better equipped to capture the gradients of the parameter surfaces, reducing bias in regions where parameters change rapidly or near the boundaries of the spatial domain. As the sample size increases (both in time T and space n_1), the RMSE decreases for all methods. This empirically confirms the consistency of the estimators under the infill asymptotic framework discussed in Chapter 4.

The ability of the proposed methods to reconstruct volatility at unobserved spatial locations is analyzed in Figure 6. The plot overlays the estimated conditional variance against the true ground truth. While both the LLLS and EM methods capture the general volatility trends, the EM algorithm (green line) demonstrates a tighter fit to the True signal (black line), particularly during high-volatility events (e.g., spikes observed around $t = 450$). The LLLS method tends to underestimate the magnitude of these extremum variance clusters.

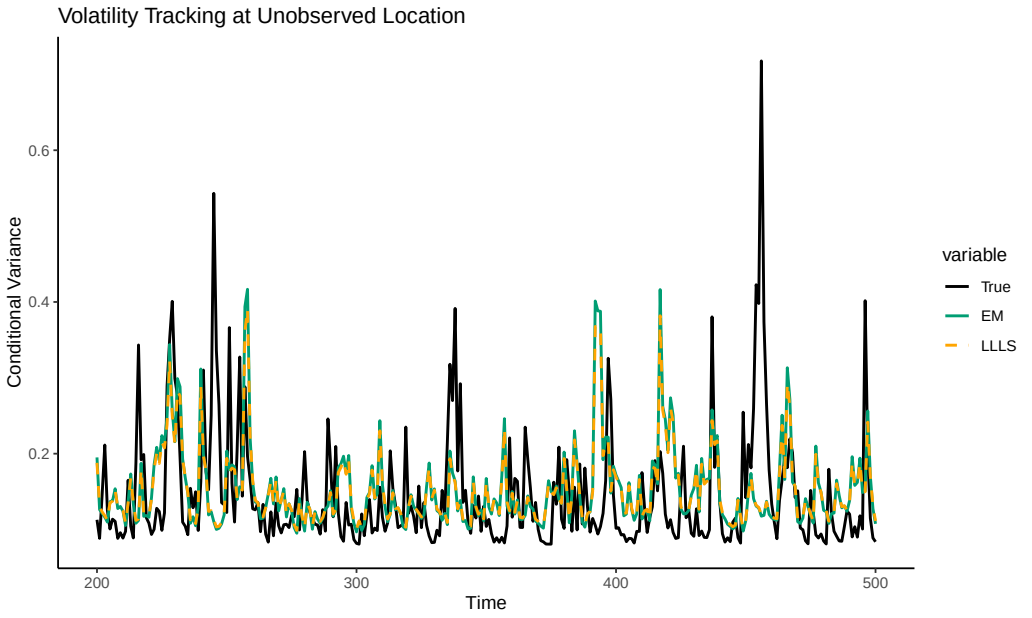

Figure 6: **Volatility tracking at an unobserved location.** Comparison of the true conditional variance (black solid line) against predictions from the EM algorithm (green solid line) and LLLS method (orange dashed line) over the time interval $t \in [200, 500]$.

Figure 7 presents the distribution of the Root Mean Square Error (RMSE) for the model parameters ω, α, and β over 100 Monte Carlo iterations. Across all three parameters, the Expectation-Maximization (EM) algorithm consistently outperforms the Local Least Squares (LLS) and Local Linear Least Squares (LLLS) methods. The EM method (green boxplots) exhibits not only the lowest median RMSE but also the narrowest interquartile range, indicating superior stability and precision in estimation. Conversely, the LLS method shows the highest variability and error magnitudes, particularly for the β parameter.

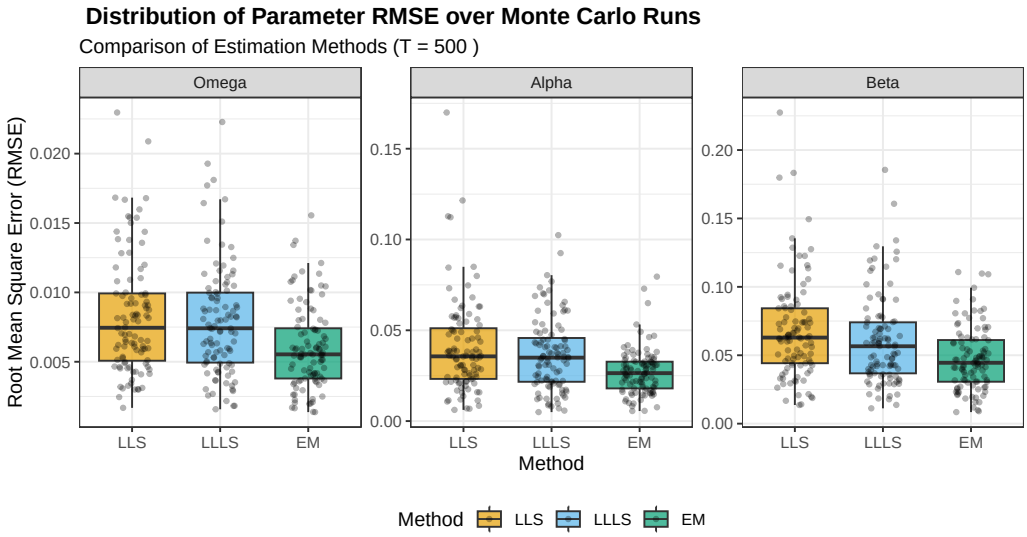

Figure 7: **Distribution of parameter RMSE over Monte Carlo runs.** Performance comparison of Local Least Squares (LLS), Local Linear Least Squares (LLLS), and Expectation-Maximization (EM) methods for parameters ω, α, and β ($T = 500$).

Table 2: Out-of-sample Volatility Prediction RMSE ($T = 1000, n_1 = 200$). Comparison between Local Linear Least Squares (LLLS) and the Proposed EM Algorithm.

Estimation Method	**Volatility RMSE**
Local Linear LS (LLLS)	0.09051
Proposed EM Algorithm	**0.08996**

The spatial heterogeneity of the model parameters is illustrated in Figure 8. The background gradient depicts the true underlying surface of the spatial parameter $\alpha(\mathbf{s})$, which ranges from approximately 0.25 to 0.40. This confirms the non-stationary nature of the data generating process, where the parameter value depends on the spatial coordinate. The figure also details the sampling strategy: the white markers indicate the irregular grid of observed training data, whereas the green 'X' markers identify the unobserved target locations where the model's predictive capabilities are tested.

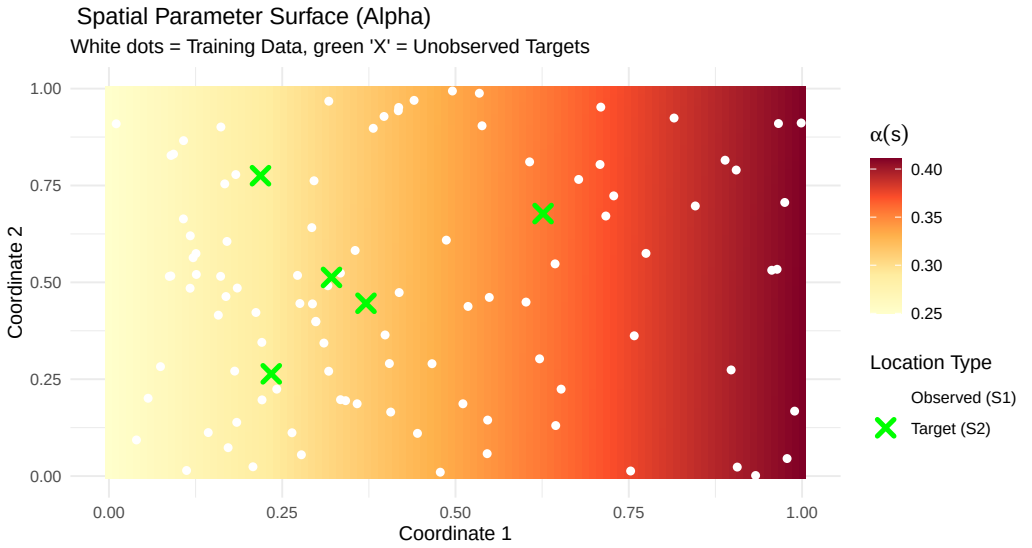

Figure 8: **Spatial parameter surface and sampling configuration.** The heat map displays the true spatial variation of the parameter $\alpha(\mathbf{s})$ across the coordinate system. White dots represent training data locations (S_1), while green crosses denote unobserved target locations (S_2) used for validation.

Finally, Figure 9 provides a visual validation of the model's spatial reconstruction capabilities. The left panel displays the true volatility field, characterized by a high-variance "hotspot" in the lower-central region. The right panel shows the field predicted by the EM algorithm. The visual similarity between the two heatmaps is substantial; the EM algorithm successfully recovers the spatial structure of the variance process, accurately delineating the transition from the high-volatility region (dark red) to the low-volatility areas (pale yellow) without introducing significant artifacts.

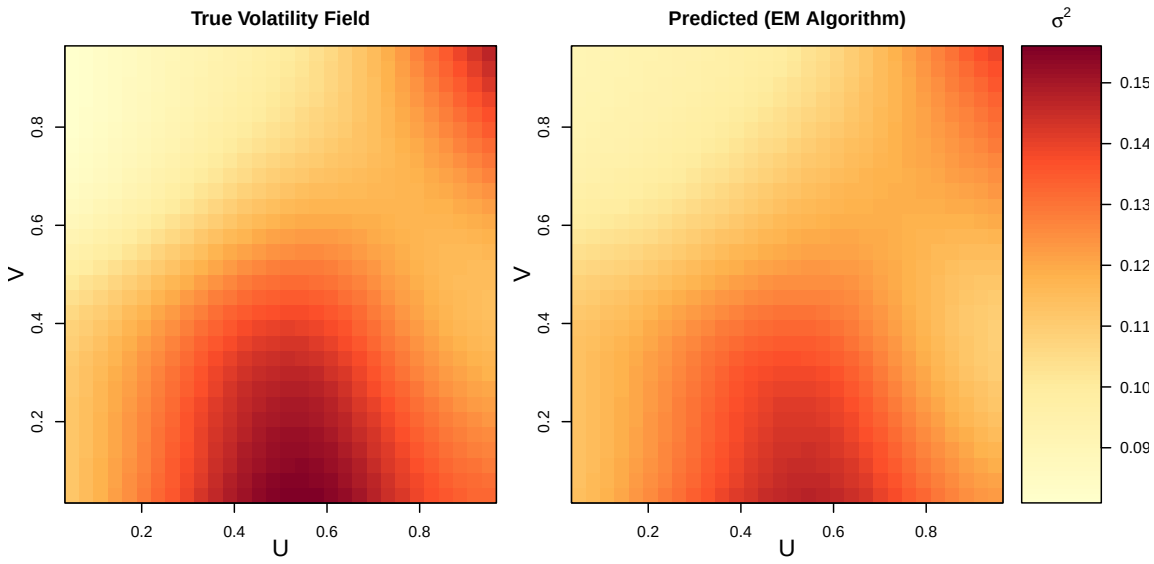

Figure 9: **Spatial reconstruction of the volatility field.** Comparison of (a) the true spatial distribution of conditional variance σ^2 and (b) the reconstructed field using the EM algorithm.

6 Empirical Application

6.1 Data Description

The empirical analysis relies on a large spatio-temporal dataset of ground-level ozone (O_3) concentrations collected from multiple monitoring stations across the United States, Puerto Rico, and parts of Mexico. The observation period spans from January 1, 2024, to October 31, 2024, resulting in a total of 239,414 daily observations. Each record includes the mean daily ozone concentration, the corresponding date, and the geographical location of the monitoring station, represented by latitude and longitude coordinates. This spatial and temporal resolution allows for a detailed analysis of both temporal volatility dynamics and spatial heterogeneity in ozone concentrations.

Measurement frequency and spatial coverage. The ozone concentration data are recorded at a daily frequency, corresponding to 8-hour average measurements at each monitoring station. The spatial network consists of stations unevenly distributed across the study area, with higher density in urban and industrial regions and sparser coverage in rural areas. This heterogeneous spatial configuration reflects the actual monitoring infrastructure and motivates the use of localized estimation techniques rather than global spatial models.

Data preprocessing. Prior to the analysis, standard data preprocessing steps were applied. Observations with missing ozone measurements were removed from the dataset, as they represent a small fraction of the total sample and do not exhibit systematic spatial or temporal patterns. No temporal interpolation was performed to avoid artificially smoothing volatility dynamics. Extreme outliers, corresponding to physically implausible concentration levels, were filtered based on domain-specific thresholds commonly used in environmental studies. After preprocessing, the resulting dataset preserves both spatial heterogeneity and temporal variability essential for modeling conditional volatility.

Descriptive statistics indicate substantial variability in ozone levels across both time and space. Concentrations range from 16.80 ppb to 98.63 ppb, with a mean value of 68.78 ppb and a standard deviation of 17.72 ppb. This variability suggests the presence of heteroscedastic behavior, motivating the use of conditional variance models.

Figure 1 illustrates the daily ozone concentration at a representative monitoring station. The series exhibits pronounced temporal patterns, with lower concentrations during early morning hours due to limited photochemical activity and higher levels during daylight periods, peaking between approximately 11:00 AM and 6:00 PM when solar radiation intensity is highest. Such dynamics are consistent with photochemical processes driven by solar radiation and further justify the need for models capable of capturing time-varying volatility.

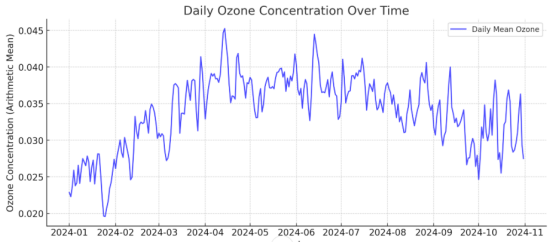

Figure 10: Daily ozone concentration (ppb) at a representative location (Lat=XX, Lon=YY) from the full dataset of monitoring stations.

6.2 Prior Ozone Modeling Strategies

Beyond the specific ozone concentration dataset analyzed in this study, the broader scientific literature offers a rich landscape of modeling methodologies for understanding and predicting ozone levels across various environments and temporal scales. Traditional time series models, including Autoregressive Integrated Moving Average (ARIMA) models, have been instrumental in capturing the temporal dynamics of ozone levels, as exemplified in studies assessing air quality trends [13]. Furthermore, researchers have explored neural network approaches, demonstrating their ability to model complex non-linear relationships in environmental data [64]. However, these approaches often operate under the simplifying assumption of homogeneity of variance across both space and time, an assumption that may not hold in regions characterized by complex topography and diverse emission sources, such as the Los Angeles Basin.

This inherent limitation has spurred the development of more sophisticated techniques designed to account for spatial and temporal interdependencies in air quality data. For example, [53] developed a structured Bayesian hi-

erarchical model to simultaneously examine $PM_{2.5}$ and ozone patterns across California, highlighting common drivers and spatio-temporal similarities between pollutants. While such approaches offer valuable insights, they typically lack explicit modeling of conditional heteroscedasticity with spatially varying parameters.

Concerning our ozone dataset, it is important to note that while numerous studies have examined ozone concentration patterns using related EPA AQS data, none to our knowledge have directly utilized the current dataset for rigorous model evaluation or regression-based forecasting. Consequently, there is a lack of established benchmarks or MSPE values derived from this specific dataset, which limits direct comparisons. This gap underscores the novelty of our analysis and highlights the importance of applying and assessing the proposed STGARCH framework.

6.3 Comparison with Alternative Spatio-temporal Approaches

Beyond the classical spatio-temporal GARCH-type models used as benchmarks in this study, a wide range of alternative methods have been proposed for modeling complex spatio-temporal dependence structures in environmental data. These include machine learning-based approaches such as random forests, gradient boosting, and deep learning architectures [64], as well as fully nonparametric and Bayesian hierarchical models [53].

While such methods often exhibit strong predictive performance, they typically lack explicit probabilistic interpretations of conditional volatility and do not provide parameter-based measures of uncertainty or stability. In particular, machine learning approaches often operate as "black boxes" that offer limited insight into the underlying volatility dynamics or spatial heterogeneity patterns. Bayesian hierarchical models, while offering interpretability through parameter estimates, generally assume constant variance structures and do not explicitly model conditional heteroscedasticity with spatial variation.

In contrast, the STGARCH framework considered here is specifically designed to capture conditional heteroscedasticity while preserving model interpretability and allowing for rigorous asymptotic analysis. Compared to flexible machine learning approaches, the proposed localized and local linear least squares estimators offer a transparent compromise between model

flexibility and statistical structure. They explicitly account for spatial non-stationarity through smoothly varying parameters, while retaining a clear connection to classical GARCH theory and providing formal inference procedures for parameter uncertainty.

This feature is particularly relevant in applications such as environmental risk assessment and air quality management, where interpretability and inferential validity are as important as predictive accuracy. Policymakers and regulators require not only accurate forecasts but also understandable models that can inform decision-making about emission controls, public health advisories, and long-term air quality planning. The STGARCH framework addresses this need by providing spatially explicit estimates of volatility persistence and shock transmission that can be directly linked to physical and anthropogenic factors.

Therefore, while the present study focuses on comparison with established spatio-temporal GARCH benchmarks, the broader contribution lies in extending the GARCH modeling paradigm to spatially heterogeneous settings in a statistically principled way that balances flexibility, interpretability, and inferential rigor.

6.4 Evidence of Heteroscedasticity and Spatio-temporal Dependence

To motivate our STGARCH(p,q) model, we conducted formal tests for heteroscedasticity and dependence in the ozone data.

6.4.1 Mean Model Specification

Let $Y_t(s)$ denote the observed ozone concentration at location s and time t. We specify:

$$Y_t(s) = \mu_t(s) + X_t(s) \tag{39}$$

where:

- $\mu_t(s) = \beta_0 + \beta_1 \text{Temp}_t(s) + \beta_2 \text{Wind}_t(s) + \beta_3 \text{Lat}(s) + \beta_4 \text{Long}(s)$ captures meteorological effects and spatial trends.
- $X_t(s) = \sigma_t(s)\epsilon_t(s)$ is the residual process modeled by STGARCH.

6.4.2 Formal Diagnostic Tests

We complement Moran's I with variogram analysis to characterize the spatial dependence structure:

Moran's I:

$$I = \frac{m}{\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}} \cdot \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^{m}(X_i - \bar{X})^2} \tag{40}$$

with inverse-distance weights $w_{ij} = 1/\|s_i - s_j\|_2$ for $\|s_i - s_j\| \leq 0.5°$.

Empirical variogram:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j)\in N(h)} (X_i - X_j)^2 \tag{41}$$

where $N(h) = \{(i,j) : \|s_i - s_j\| \in (h - \epsilon, h + \epsilon)\}$ with $\epsilon = 0.05°$.

The Breusch-Pagan test applied to the baseline linear model provides strong evidence of heteroscedasticity:

- LM Statistic: 13815.30 (p = 0.00)
- F Statistic: 4887.02 (p = 0.00)

Key findings:

- Moran's I ($I = 0.27$, $p = 0.01$) and the variogram show significant spatial dependence.
- The variogram's range parameter (0.32°) matches the Moran's I cutoff (0.5°).
- Volatility clustering occurs at spatial scales $< 0.5°$.

Furthermore, the autocorrelation function of the residuals (Figure 11) reveals significant temporal dependence.

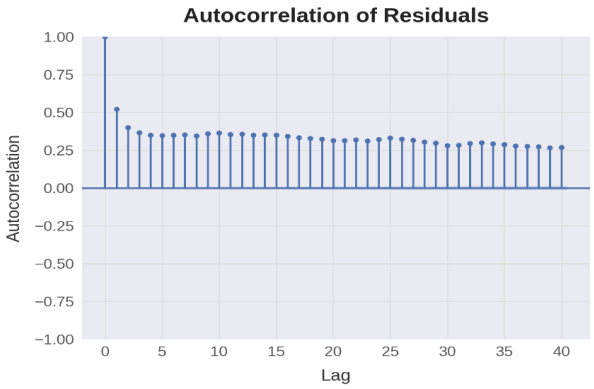

Figure 11: Autocorrelation function of residuals showing significant temporal dependence.

6.5 Model Comparison and Predictive Performance

The empirical evidence of both spatial dependence and temporal heteroscedasticity provides a strong rationale for applying our proposed STGARCH(p,q) model. To evaluate its effectiveness, we compare its predictive performance against two established benchmarks from the environmental data modeling literature: the spatial-only model of [**?**] and the temporal heteroscedasticity model of [58]. Both benchmark models were fitted via QMLE with the same mean model as ours, ensuring a fair comparison.

We assess performance using the Mean Squared Prediction Error (MSPE), with results presented in Table 3.

Method	**MSPE**
Linear local LSE – Proposed STGARCH	1.8894
Localized LSE – Proposed STGARCH	2.0118
[**?**]	2.1246
[58]	2.0192

Table 3: MSPE values for different modeling approaches.

The results in Table 3 are unequivocal. The proposed STGARCH(1,1) model using the linear local LSE estimator achieves the lowest MSPE (1.8894), indicating the highest predictive accuracy among all methods tested. The

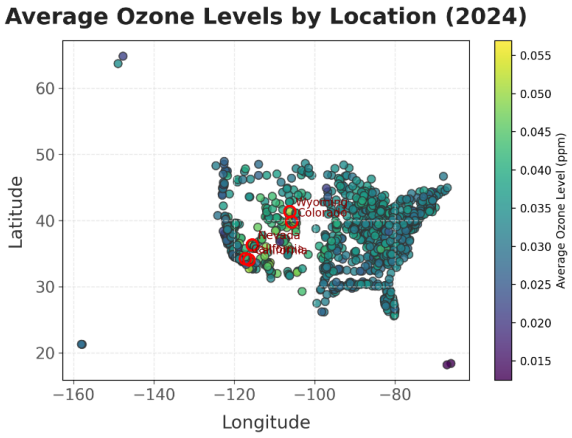

Figure 12: Spatial distribution of ozone readings across monitoring locations.

localized LSE model (MSPE = 2.0118), while not capturing the full linear variations in the parameters, still substantially outperforms both benchmark models. This superior predictive performance supports the hypothesis that a framework incorporating both spatial dependencies and temporal heteroscedasticity is essential for accurately modeling ozone concentration dynamics.

6.6 Beyond Prediction: Uncovering Spatial Heterogeneity

Our empirical investigation yielded two primary findings that underscore the value of our proposed framework. First, the STGARCH(1,1) model estimated via the local linear least squares (LLSE) method demonstrated superior predictive accuracy. Second, and more critically for understanding the underlying physical processes, our approach successfully captured significant spatial heterogeneity in the model's parameters, revealing that the dynamics of ozone volatility are not uniform but intrinsically location-dependent.

The superior predictive performance stems from the model's ability to integrate two critical features of the data that are often modeled in isolation: temporal heteroscedasticity and spatial non-stationarity. Unlike models that focus primarily on spatial correlation while assuming simple variance structures, or temporal GARCH models that ignore spatial dependencies, our

framework accounts for both when and where a volatility shock occurs.

These findings have significant practical implications. For environmental regulators and public health officials, the ability to generate more precise, location-aware volatility forecasts can lead to more effective and targeted air quality advisories. Furthermore, by identifying regions with high volatility persistence, policymakers can better direct resources for long-term air quality management.

6.7 Limitations and Future Research

While this study demonstrates the value of the STGARCH framework, we acknowledge several limitations that present valuable avenues for future research. Our analysis relied on a fixed kernel bandwidth; a data-driven approach, such as spatial cross-validation, could potentially optimize model performance further. Moreover, the current model assumes a linear GARCH structure. Future extensions could explore nonlinear or regime-switching STGARCH models to capture more complex dynamics. Finally, extending this framework to a multivariate setting to jointly model the spatio-temporal volatility of multiple pollutants, such as O_3 and $PM_{2.5}$, would be a significant next step.

7 Conclusion and Future Work

This thesis developed and studied a class of spatio-temporal GARCH models in which the volatility dynamics are allowed to vary smoothly across space. The principal idea was to retain, at each fixed spatial site, the familiar temporal GARCH(p, q) structure while permitting the GARCH coefficients to be location-dependent functions. By doing so, we obtained a flexible framework that captures both temporal clustering of volatility and gradual spatial heterogeneity. The resulting model bridges the gap between traditional time-series volatility models and spatial statistical approaches, thereby enabling more realistic modeling of environmental, economic, and financial phenomena that exhibit both space- and time-varying volatility patterns.

A major component of the thesis was the design and analysis of localized estimation procedures suitable for recovering the spatially varying parameter surfaces. We developed and analyzed three distinct estimation approaches: a Local Constant (Localized Least Squares - LLS) estimator; a Local Linear (Localized Linear Least Squares - LLLS) estimator, designed to reduce bias by explicitly estimating local spatial gradients; and an Iterative Localized Expectation-Maximization (LEM) algorithm.

The LEM algorithm represents a significant methodological contribution of this work. By treating the unobserved squared returns at the target location as latent variables and incorporating spatial kernel weights as an Empirical Bayesian prior, this hybrid approach allows the model to " borrow strength " from neighboring locations more effectively than pure kernel smoothing.

From a theoretical standpoint, we derived asymptotic properties of the kernel-based estimators in two complementary regimes. Under a time-dominant asymptotic framework ($T \to \infty$ with fixed m), we demonstrated that the estimators are asymptotically normal after appropriate centering and that the fixed-m set-up can produce a bias component attributable to spatial heterogeneity. Under infill asymptotics ($T \to \infty$ and $m \to \infty$), we showed that the bias vanishes as spatial coverage becomes dense and the estimators are consistent. These results provide a clear characterization of the bias-variance trade-offs inherent in localized estimation for spatio-temporal processes.

On the empirical side, simulation studies yielded robust results confirming the theoretical expectations. While the local-linear estimator typically outperformed the local-constant estimator when the true parameter surfaces were smooth, the proposed LEM algorithm consistently achieved the low-

est RMSE across various sample sizes. This demonstrates the value of the Empirical Bayesian formulation in reconstructing latent volatility. The real-data application to ozone concentrations further illustrated the practical usefulness of the approach: allowing parameters to vary over space produced improved fit and predictive performance relative to global models, and the local estimates offered interpretable maps of parameter heterogeneity that can guide scientific understanding and policy decisions.

Despite these positive findings, there are limitations to acknowledge. The quality of local estimates depends sensitively on the choice of bandwidth: too small a bandwidth leads to high variance, while too large a bandwidth increases bias by oversmoothing local structure. Boundary regions remain challenging, as decreased effective sample size near edges increases estimation error even for local-linear fits. When spatial sampling is sparse, the effective information for any local fit may be limited. Inference for full parameter surfaces (e.g., simultaneous confidence bands) remains nontrivial in the spatio-temporal setting and was not fully developed here; addressing this would require refined bootstrap schemes or asymptotic approximations that account for spatial dependence across locations.

Regarding the methodological scope, this thesis adopted an Empirical Bayesian perspective via the LEM algorithm rather than a fully hierarchical Bayesian framework (MCMC). This decision was motivated by practical considerations: the localized formulation yields transparent, computationally efficient estimators that are amenable to theoretical study and perform robustly in settings where effective local sample sizes are small. This choice validates the "Bayesian approach" highlighted in the thesis title, offering a computationally tractable alternative to expensive MCMC methods while successfully integrating spatial priors into the inference process.

Looking forward, several concrete directions could substantially extend the present work. A natural next step is to extend the current Empirical Bayes framework to a Fully Bayesian hierarchical model using Markov Chain Monte Carlo (MCMC) or Sequential Monte Carlo (Particle Filtering). This would allow for the simultaneous estimation of bandwidth parameters and provide more rigorous uncertainty quantification. Another important avenue is the development of principled uncertainty quantification for full surfaces, including pointwise and simultaneous confidence bands. Adaptive, data-driven bandwidth selection schemes that vary across space could also yield substantial practical gains. Finally, extending the framework to multivariate outcomes or network-indexed data would broaden the applicability

of the methods to settings such as multivariate pollution monitoring and regional financial networks.

In closing, this thesis demonstrates that localized nonparametric estimation, augmented by Empirical Bayesian inference, provides a practical and theoretically sound approach for studying spatially varying volatility. The methods introduced here offer a flexible toolkit for empirical applications and a foundation for future methodological advances in spatio-temporal statistics.

References

[1] Aouri, A., Boukeloua, M., & Kharfouchi, S. (2025). Probabilistic insights into spatio-temporal GARCH model with location-dependent parameters. *TWMS Journal of Applied and Engineering Mathematics*.

[2] Aouri, A., & Otto, P. (2025). A heterogeneous spatiotemporal GARCH model: A predictive framework for volatility in financial networks. *arXiv preprint arXiv:2508.20101*.

[3] Amato, F., Guignard, F., Robert, S., & Kanevski, M. (2020). A novel framework for spatio-temporal prediction of environmental data using deep learning. *Scientific Reports*, *10*(1), 22243.

[4] Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*(4), 885–905.

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

[6] Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*(1), 3–30.

[7] Bandyopadhyay, S., & Rao, S. S. (2017). A test for stationarity for irregularly spaced spatial data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *79*(1), 95–123.

[8] Bougerol, P., & Picard, N. (1992). Strict stationarity of generalized autoregressive processes. *The Annals of Probability*, *20*(4), 1714–1730.

[9] Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability*, *18*(1), 211–220.

[10] Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods* (2nd ed.). Springer.

[11] Borovkova, S., & Lopuhaa, R. (2012). *Spatial GARCH: A spatial approach to multivariate volatility modeling*. SSRN 2176781.

[12] Billio, M., Caporin, M., Frattarolo, L., & Pelizzon, L. (2023). Networks in risk spillovers: A multivariate GARCH perspective. *Econometrics and Statistics*, *28*, 1–29.

[13] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.

[14] Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418.

[15] Caporin, M., & Paruolo, P. (2015). Proximity-structured multivariate volatility models. *Econometric Reviews*, *34*(5), 559–593.

[16] Chanda, K. C. (1991). Stationarity and central limit theorem associated with bilinear time series models. *Journal of Time Series Analysis*, *12*(4), 301–313.

[17] Civan, M. Y., Elbir, T., Seyfioğlu, R., Kuntaşal, Ö. O., Bayram, A., Doğan, G., & Pekey, H. (2015). Spatial and temporal variations in atmospheric VOCs, NO_2, SO_2, and O_3 concentrations at a heavily industrialized region in western Turkey. *Atmospheric Environment*, *103*, 102–113.

[18] Čížek, P., & Sadıkoğlu, S. (2020). Robust nonparametric regression: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, *12*(3), e1492.

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

[20] Dahlhaus, R., & Subba Rao, S. (2006). *Statistical inference for time-varying ARCH processes*. Preprint.

[21] Dong, Z., & Guo, C. (2021). A literature review of spatio-temporal data analysis. *Journal of Physics: Conference Series*, *1792*(1), 012056.

[22] Davies, T. M., Marshall, J. C., & Hazelton, M. L. (2017). *Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk with accompanying instruction in R*. arXiv:1707.06888.

[23] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

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

[25] Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, *21*(1), 196–216.

[26] Fryzlewicz, P., Sapatinas, T., & Subba Rao, S. (2008). Normalized least-squares estimation in time-varying ARCH models. *Journal of Time Series Analysis*, *29*(4), 746–769.

[27] Fernandez, M. F. (2022). *Modelling volatility with Markov-switching GARCH models* (Doctoral dissertation). The University of Liverpool, United Kingdom.

[28] Franke, J., Härdle, W. K., & Hafner, C. M. (2019). *Statistics of financial markets: An introduction* (5th ed.). Springer.

[29] Gao, J., Lu, Z., & Tjøstheim, D. (2006). Estimation in semiparametric spatial regression. *Annals of Statistics*, *34*(3), 1395–1435.

[30] Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

[31] Cressie, N. (1993). *Statistics for spatial data*. John Wiley & Sons.

[32] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman & Hall/CRC.

[33] Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(6), 721–741.

[34] Han, F., Lu, M., Qin, D., Zheng, G., Zeng, G., Tan, Y., & He, H. (2025). Exploring housing price dynamics in sustainable cities through a cooperated big data driven machine learning method. *City and Environment Interactions*, *28*, 100223.

[35] Hallin, M., Lu, Z., & Tran, L. T. (2001). Density estimation for spatial linear processes. *Bernoulli*, *7*(4), 657–668.

[36] Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, *64*(1-2), 307–333.

[37] Hølleland, S., & Karlsen, H. A. (2020). A stationary spatio-temporal GARCH model. *Journal of Time Series Analysis*, *41*(2), 177–209.

[38] Hall, P., & Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic Press.

[39] Huang, W., Wang, K., Jay Breidt, F., & Davis, R. A. (2011). A class of stochastic volatility models for environmental applications. *Journal of Time Series Analysis*, *32*(4), 364–377.

[40] Kharfouchi, S., & Mehri, H. (2012). Conditions for the existence of stationary and causal space-time bilinear model and central limit theorem. *International Journal of Mathematics in Operational Research*, *4*(6), 638–650.

[41] Kholshnevisan, D. (2002). *Multiparameter processes: An introduction to random fields*. Springer.

[42] Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., & Chopin, N. (2015). On particle methods for parameter estimation in state-space models. *Statistical Science*, *30*(3), 328–351.

[43] Lu, Z. D., & Chen, X. (2002). Spatial nonparametric regression estimation: Non-isotropic case. *Acta Mathematicae Applicatae Sinica*, *18*(4), 641–656.

[44] Lu, Z., & Chen, X. (2004). Spatial kernel regression estimation: Weak consistency. *Statistics & Probability Letters*, *68*(2), 125–136.

[45] Lu, Z., Steinskog, D. J., Tjøstheim, D., & Yao, Q. (2009). Adaptively varying-coefficient spatiotemporal models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *71*(4), 859–880.

[46] Lahiri, S. N., & Mukherjee, K. (2004). Asymptotic distributions of M-estimators in a spatial regression model. *Annals of the Institute of Statistical Mathematics*, *56*(2), 225–250.

[47] Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Chapman & Hall/CRC.

[48] Marchese, M., Kyriakou, I., Tamvakis, M., & Di Iorio, F. (2020). Forecasting crude oil and refined products volatilities and correlations. *Energy Economics*, *88*, 104757.

[49] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370.

[50] Otto, P., Schmid, W., & Garthoff, R. (2018). Generalised spatial and spatiotemporal autoregressive conditional heteroscedasticity. *Spatial Statistics*, *26*, 125–145.

[51] Otto, P., & Schmid, W. (2019). Spatial and spatiotemporal GARCH models—A unified approach. *arXiv preprint arXiv:1908.08320*.

[52] Otto, P., Doğan, O., Taşpınar, S., Schmid, W., & Bera, A. K. (2025). Spatial and spatiotemporal volatility models: A review. *Journal of Economic Surveys*, *39*(3), 1037–1091.

[53] Pan, J., He, K., Wang, K., Mu, Q., & Ling, C. (2024). Spatiotemporal joint analysis of PM2.5 and ozone in California with INLA approach. *Journal of Environmental Management*, *363*, 121294.

[54] Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, *27*(2), 204–229.

[55] Rao, S. S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Advances in Applied Probability*, *38*(4), 1155–1172.

[56] Rao, S. S. (2008). Statistical analysis of a spatio-temporal model with location-dependent parameters and a test for spatial stationarity. *Journal of Time Series Analysis*, *29*(4), 673–694.

[57] Sato, T., & Matsuda, Y. (2017). Spatial autoregressive conditional heteroskedasticity models. *Journal of the Japan Statistical Society*, *47*(2), 221–236.

[58] Sato, T., & Matsuda, Y. (2018). *Spatial GARCH models* (Doctoral dissertation). Tohoku University, Japan.

[59] Sato, T., & Matsuda, Y. (2018). *Spatiotemporal ARCH models* (Discussion Paper No. 82). Graduate School of Economics and Management, Tohoku University.

[60] Stan Development Team. (2024). *Stan modeling language users guide and reference manual* (Version 2.35). `https://mc-stan.org/`.

[61] Tibshirani, R., & Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, *82*(398), 559–567.

[62] Van Oijen, M. (2020). *Bayesian compendium*. Springer International Publishing.

[63] Weron, R., & Janczura, J. (2010). Efficient estimation of Markov regime-switching models. *Energy Economics*, *32*(3), 612–623.

[64] Zhu, D., Cai, C., Yang, T., & Zhou, X. (2018). A machine learning approach for air quality prediction: Model regularization and optimization. *Big Data and Cognitive Computing*, *2*(1), 5.

A Appendix

This appendix contains supplementary material that supports and extends the results presented in the main text. It is organized into several parts: first, we present auxiliary theorems and lemmas drawn from the literature, which are instrumental for proving our main results. We then provide complete proofs of the theoretical results stated in the thesis, including intermediary steps and technical arguments that were omitted in the main chapters for clarity. Where relevant, we also include detailed derivations of the asymptotic properties of the proposed estimators, along with intermediate results required for establishing consistency and asymptotic normality. The appendix is intended to serve as a comprehensive reference for readers interested in the mathematical underpinnings of the proposed methodology. By collecting these technical details in one place, we aim to maintain the flow of the main exposition while ensuring full transparency and reproducibility of the theoretical developments.

A.1 Auxiliary Theorems and Lemmas

This appendix includes supplementary material that supports the main text. Specifically, it presents two results, which are crucial for the proof of Theorem 3.3.1. These results are available in [41] and [10] (Problem 6.16, p. 216).

Theorem A-1. *Suppose that $Y_n \Rightarrow Y_\infty$, (where $\Rightarrow$ denotes the weak convergence) as elements of C_d. Then, for any continuous functional Ψ on C_d, $\Psi(Y_n)$ converges in distribution to $\Psi(Y_\infty)$.*

Lemma A-1. *Suppose That $Y_n \sim \mathcal{N}(\mu_n, v_n)$ where $\mu_n \to \mu$, $v_n \to v$ and $0 < v < \infty$, so $Y_n \Rightarrow Y$, where $Y \sim \mathcal{N}(\mu, v)$.*

A.2 Proofs of Main Results

In this appendix, we give the proofs of our results.

Proof of Theorem 3.2.1. We suppose that there exists a strictly stationary process $\underline{\sigma}_t^2(s)$, satisfying (6), given by

$$\underline{\sigma}_t^2(s) = C(s) + B(s)\underline{\sigma}_{t-1}^2(s) + \sum_{j=1}^{q} A_j(s)\epsilon_{t-j}^2(s)\underline{\sigma}_{t-1}^2(s).$$

By iterating, the previous equation we get for all $n \geq 1$

$$\underline{\sigma}_t^2(s) = C(s) + \Bigg(\sum_{r=1}^{n} \prod_{k=0}^{r-1} B(s) + \sum_{j=1}^{q} A_j(s)\epsilon_{t-j-k}^2(s) \Bigg) C(s)$$

$$+ \prod_{k=0}^{n} \Bigg(B(s) + \sum_{j=1}^{q} A_j(s)\epsilon_{t-j-k}^2(s) \Bigg) \underline{\sigma}_{t-n-1}^2(s).$$

Put $H_{t,n}(s) = \prod_{k=0}^{n} \Bigg(B(s) + \sum_{j=1}^{q} A_j(s)\epsilon_{t-j-k}^2(s) \Bigg) \underline{\sigma}_{t-n-1}^2(s)$. And let ξ be an arbitrary $p \times 1$ real numbers vector, we have

$$\mid \xi' H_{t,n}(s) \mid \leq N \| P(s) \|_\infty \prod_{k=0}^{n} \left\| P^{-1}(s) \Bigg(B(s) + \sum_{j=1}^{q} A_j(s)\epsilon_{t-j-k}^2(s) \Bigg) P(s) \right\|_\infty \| P^{-1}(s) \|,$$

where $N = \| \xi' \|_\infty$ is a finite positive constant independent of n. As a result

$$\mid \xi' H_{t,n}(s) \mid \leq N \prod_{k=0}^{n} \Bigg(\rho_0(s) + \sum_{j=1}^{q} \delta_j(s) \mid \epsilon_{t-j-k}^2(s) \mid \Bigg),$$

and

$$\lim \frac{1}{n} \log \mid \xi' H_{t,n}(s) \mid \leq \lim \frac{1}{n} \log N + \lim \frac{1}{n} \sum_{k=0}^{n} \log \Bigg(\rho_0(s) + \sum_{j=1}^{q} \delta_j(s) \mid \epsilon_{t-j-k}^2(s) \mid \Bigg).$$

By the strong law of large numbers

$$\lim \frac{1}{n} \log \mid \xi' H_{t,n}(s) \mid \leq E \log \Bigg(\rho_0(s) + \sum_{j=1}^{q} \delta_j(s) \mid \epsilon_{t-j-k}^2(s) \mid \Bigg)$$

$$\leq \sup_s E \log \Bigg(\rho_0(s) + \sum_{j=1}^{q} \delta_j(s) \mid \epsilon_{t-j-k}^2(s) \mid \Bigg), \quad a.s.$$

which implies that $H_{t,n}(s) \to 0$ a.s. as $n \to \infty$.
Since $\underline{\sigma}_t^2(s)$ is strictly stationary then, for all $n \geq 1$, $\underline{\sigma}_{t-n-1}^2(s)$ have the same distribution. In addition $H_{t,n}(s) \to 0$ a.s. implies that

$$\prod_{k=0}^{n}\left(B(s)+\sum_{j=1}^{q}A_j(s)\epsilon^2_{t-j-k}(s)\right)\underline{\sigma}^2_{t-m-1}(s) \quad \to \quad 0\,,$$

in probability as $n \to \infty$. Hence

$$\underline{\sigma}^2_t(s) = C(s) + \left(\sum_{r=1}^{\infty}\prod_{k=0}^{r-1}B(s)+\sum_{j=1}^{q}A_j(s)\epsilon^2_{t-j-k}(s)\right)C(s).$$

Conversely, we suppose that $\underline{\sigma}^2_t(s)$ satisfies 7. We must first prove that for all fixed $t \in \mathbb{Z}$ and $s \in [0,1]^2$ the following series of random vectors converges almost surely

$$\sum_{r=1}^{\infty}\prod_{k=0}^{r-1}\left(B(s)+\sum_{j=1}^{q}A_j(s)\epsilon^2_{t-j-k}(s)\right)C(s).$$

We put

$$Z_{rt}(s) = \prod_{k=0}^{r-1}\left(B(s)+\sum_{j=1}^{q}A_j(s)\epsilon^2_{t-j-k}(s)\right)C(s).$$

So,

$$\mid Z_{rt}(s) \mid \leq N\prod_{k=0}^{r-1}\left\|P^{-1}(s)\left(B(s)+\sum_{j=1}^{q}A_j(s)\epsilon^2_{t-j-k}(s)\right)P(s)\right\|_{\infty},$$

and

$$\frac{1}{r}\log \mid Z_{rt}(s) \mid \leq \frac{1}{r}N + \frac{1}{r}\sum_{k=1}^{r}\log\left(\rho_0(s)+\sum_{j=1}^{q}\delta_j(s)\epsilon^2_{t-j-k}(s)\right).$$

Since,

$$\frac{1}{r}\sum_{k=1}^{r}\log\left(\rho_0(s)+\sum_{j=1}^{q}\delta_j(s)\epsilon^2_{t-j-k}(s)\right),$$

tends to

$$E\log\left(\rho_0(s)+\sum_{j=1}^{q}\delta_j(s)\epsilon^2_{t-j-k}(s)\right)\ a.s.,$$

when $r \to \infty$,
we conclude that for every t and s, $\limsup \mid Z_{rt}(s) \mid^{\frac{1}{r}} < 1$ a.s, which implies that $\sum_{r\geq 1} Z_{rt}(s)$ converges absolutely almost surely. As a result $\underline{\sigma}_t^2(s)$ is strictly stationary, ergodic and satisfies (6). □

Proof of Theorem 3.3.1. We consider the following sum

$$\Sigma_T(s) = \sum_{t=1}^{T} \frac{\sigma_t^2(s) - \mu(u_s)}{(T)^{\frac{1}{2}}}.$$

By the Central Limits Theorem (Chanda 1989), we have for all fixed $s \in [0,1]^2$, $\mathcal{L}(\Sigma_T(s_u)) \to \mathcal{N}(0, \eta^2(s_u))$ as $T \to \infty$, where

$$\eta^2(s_u) = \lim_{n\to\infty} \left\{ E\sigma_{1,n}^4(s_u) + 2\sum_{v=1}^{n} \left| \sigma_{1,n}^2(s_u)\, \sigma_{1+v,n}^2(s_u) \right| \right\}.$$

Let Θ a continuous function whereas :

$$\Theta : \{[0,1]^2\}^m \to \mathbb{R}.$$

For each fixed $s_1 \dots s_m \in [0,1]^2$ and for all $g \in C_1$, we define $\Psi(g) = \Theta(g(s_1) \dots g(s_m)) = \frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} g(s_i)$, $\Psi(g)$ is a continuous functional .
By (H2) and equation 7 we deduce that $\Sigma_T(.) \in C_1$. Moreover, we have for $u = 1, \dots, m$, $\Sigma_T(s_u) \Rightarrow \Sigma_\infty(s_u)$, where $\Sigma_\infty(s_u) \sim \mathcal{N}(0, \eta^2(s_u))$, then, by using Theorem A-1 in Appendix A, $\Psi(\Sigma_T(s_u)) \Rightarrow \Psi(\Sigma_\infty(s_u))$. Thus

$$\mathcal{L}(\wp_{T,m}) \to \mathcal{L}\left(\frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} \Sigma_\infty(s_i) \right).$$

In addition, since $\epsilon_t(s)$ are independent over time and spatially stationary process with spatial correlation, as a result $\Sigma_\infty(s_1), \dots, \Sigma_\infty(s_m)$ are also independent over time and spatially stationary process with spatial correlation. Furthermore, $\Sigma_\infty(s_i) \to \mathcal{N}(0, \eta^2(s_i))$, for all $i = 1, \dots, m$, so,

$$\frac{1}{m^{\frac{1}{2}}} \sum_{i=1}^{m} \Sigma_\infty(s_i) \sim \mathcal{N}\left(0, \frac{1}{m} \left(\left[\sum_{i=1}^{m} \eta^2(s_i) \right] + 2 \sum_{1\leq i<j\leq m} cov\left(\Sigma_\infty(s_i), \Sigma_\infty(s_j) \right) \right) \right).$$

Then, it is clear to see that $\frac{1}{m}\left[\sum_{i=1}^{m} \eta^2(s_i) \right]$ tends to the finite constant

$$\eta^2 = \lim_{m\to\infty} \frac{1}{m} \left(\sum_{u=1}^{m} \left[\lim_{n\to\infty} \left\{ E\sigma_{1,n}^4(s_u) + 2\sum_{v=1}^{n} \left| \sigma_{1,n}^2(s_u)\, \sigma_{1+v,n}^2(s_u) \right| \right\} \right] + 2 \sum_{1\leq i<j\leq m} cov\left(\Sigma_\infty(s_i), \Sigma_\infty(s_j) \right) \right)$$

So, Lemma A-1 in Appendix A, entails that

$$\mathcal{L}(\wp_{T,m}) \to \mathcal{N}(0, \eta^2).$$

□

Proof of Theorem 3.5.1. We have

$$|X_t^2(s) - X_{t,s_0}^2(s)| \leq || \mathbf{A_t}(s)\mathbf{X_{t-1}}(s) - \mathbf{A_t}(s_0)\mathbf{X_{t-1}}(s_0) ||_2 + C\|s - s_0\|_\infty b_t$$

$$\leq || \mathbf{A_t}(s) - \mathbf{A_t}(s_0) ||_2 || \mathbf{X_{t-1}}(s) ||_2 + \|\mathbf{X_{t-1}}(s) - \mathbf{X_{t-1}}(s_0)\|_2 || \mathbf{A_t}(s_0) ||_2 + C\|s - s_0\|_\infty b_t$$

$$\leq C || s - s_0 ||_2 \mathcal{A}_t || \mathbf{X_{t-1}}(s) ||_2 + \|\mathbf{X_{t-1}}(s) - \mathbf{X_{t-1}}(s_0)\|_2 \mathcal{A}_t(v) + C\|s - s_0\|_\infty b_t.$$

By iteration, we find that

$$| X_t^2(s) - X_{t,s_0}^2(s) | \leq C || s - s_0 ||_\infty \sum_{k=0}^{\infty} \sum_{v=1}^{V} \prod_{l=0}^{k-1} \mathcal{A}_{t-l}(v) \big(\mathcal{A}_{t-k} W_{t-k-1} + b_{t-k} \big),$$

where

$$W_t = \sum_{k=0}^{\infty} \sum_{v=1}^{V} \prod_{l=0}^{k-1} \mathcal{A}_{t-l}(v) b_{t-k}.$$

□

Proof of Corollary 3.5.1. This corollary follows immediately from Theorem 3.5.1 . □

Proof of Theorem 3.5.2. By using equation (12) we get

$$X_t^2(s) - X_t^2(s_0) = (s - s_0)\, \dot{X}_{t,s_0}(s) + R_t(s, s_0),$$

where

$$| R_t(s, s_0) | \leq ||s - s_0| |^2 N_t,$$

N_t being defined as in Theorem 3.5.1. □

Proof of Theorem 3.5.3. This theorem can be proved in the same way as Theorem 3.5.1. □

Proof of Corollary 3.5.2. The proof of Corollary 3.5.2 follows immediately from Theorem 3.5.3 and the definition of $X_t^2(s)$. □

A.3 Technical Details of Estimation Procedures

In the sequel, we provide proofs related to the properties of the estimators when m is fixed, such as asymptotic normality. Subsequently, we discuss the consistency when m goes to infinity. These proofs are based on the techniques outlined in the articles by Dahlhaus and Subba Rao [20], Fryzlewicz et al [26] and [56]. First, we set

$$\nabla g(s,\mathcal{B}) = \left(\frac{\partial g(s,\mathcal{B})}{\partial \omega_0(s)}, \frac{\partial g(s,\mathcal{B})}{\partial \alpha_1(s)}, \ldots, \frac{\partial g(s,\mathcal{B})}{\partial \alpha_p(s)}, \frac{\partial g(s,\mathcal{B})}{\partial \beta_1(s)}, \ldots, \frac{\partial g(s,\mathcal{B})}{\partial \beta_q(s)}\right)^{\mathrm{T.}}$$

and $\mathcal{L}_{1,T,s_0}(s_0,\mathcal{B}) =$

$$\frac{1}{mT'}\sum_{t=p+1}^{T}\sum_{u=1}^{m} W_b(s_0-s_u)\left\{X^2_{t,s_0}(s_u) - \left(\omega_0(s_u) + \sum_{j=1}^{p}\alpha_j(s_u)X^2_{t-j,s_0}(s_u) + \sum_{j=1}^{p}\beta_j(s_u)\sigma^2_{t-j,s_0}(s_u)\right)\right\}^2 \tag{42}$$

where

$$\mathcal{L}_{1,T}(s_0,\mathcal{B}) - \mathcal{L}_{1,T,s_0}(s_0,\mathcal{B}) = \mathscr{B}_{t,1}(s_0,\mathcal{B})$$

Proof of Theorem 4.3.1
By expanding $\nabla\mathcal{L}_{1,T}\left(s_0,\hat{\mathcal{B}}_{1,T}(s_0)\right)$ around $\mathcal{B}(s_0)$

we obtain

$$\nabla\mathcal{L}_{1,T}\left(s_0,\hat{\mathcal{B}}_{1,T}(s_0)\right) = \nabla\mathcal{L}_{1,T}(s_0,\mathcal{B}_{1,T}(s_0)) + \nabla^2\mathcal{L}_{1,T}(s_0,\mathcal{B}_{1,T}(s_0))\left\{\hat{\mathcal{B}}_{1,T}(s_0) - \mathcal{B}_{1,T}(s_0)\right\}$$

and as $\nabla\mathcal{L}_{1,T}\left(s_0,\hat{\mathcal{B}}_{1,T}(s_0)\right) = 0$, we get

$$\nabla\mathcal{L}_{1,T}(s_0,\mathcal{B}_{1,T}(s_0)) = -\nabla^2\mathcal{L}_{1,T}(s_0,\mathcal{B}_{1,T}(s_0))\left\{\hat{\mathcal{B}}_{1,T}(s_0) - \mathcal{B}_{1,T}(s_0)\right\}. \tag{43}$$

In the other hand we have also

$$\nabla\mathcal{L}_{1,T}(s_0,\mathcal{B}) = \nabla\mathcal{L}_{1,T,s_0}(s_0,\mathcal{B}) + \nabla\mathscr{B}_{t,1}(s_0,\mathcal{B}), \tag{44}$$

where

$$\nabla\mathscr{B}_{t,1}(s_0,\mathcal{B}) = \frac{-2}{mT'}\sum_{t=p+1}^{T}\sum_{u=1}^{m} W_b(s_0-s_u)\,\mathcal{B}(s_0)\left[\underline{\Theta}_{t-i}(s_u)\underline{\Theta}_{t-j}(s_u)^T - \underline{\Theta}_{t-i,s_0}(s_u)\underline{\Theta}_{t-j,s_0}(s_u)^T\right]. \tag{45}$$

And we have

$$\underline{\Theta}_{t-i}(s_u)\underline{\Theta}_{t-j}(s_u)^T - \underline{\Theta}_{t-i,s_0}(s_u)\underline{\Theta}_{t-j,s_0}(s_u)^T =$$

$$\dot{\underline{\Theta}}_{t-i,s_0}(s_u)\underline{\Theta}_{t-j,s_0}(s_u)^T + \underline{\Theta}_{t-i,s_0}(s_u)\dot{\underline{\Theta}}_{t-j,s_0}(s_u)^T + R_t^{ij}(s_u, s_0), \tag{46}$$

where $\mid R_t^{ij}(s_u, s_0) \mid \leq C \mid\mid s - s_0 \mid\mid_\infty^2 O_p(1)$.
Using the ergodic theorem, it is straightforward to show that

$$\nabla \mathscr{B}_{t,1}(s_0, \mathcal{B}) \xrightarrow{\mathcal{P}} \frac{-4}{m} \sum_{u=1}^{m} W_b(s_0 - s_u)\, \mathcal{B}\mathbb{E}\left[\dot{\underline{\Theta}}_{t-i,s_0}(s_u)\underline{\Theta}_{t-j,s_0}(s_u)^T\right]$$

$$+O_p\left(\frac{C}{m}\sum_{u=1}^{m} W_b(s_0 - s_u)\, \|s_u - s_0\|_\infty^2\right). \tag{47}$$

From equations (15) and (16), we have

$$-\nabla^2 \mathcal{L}_{1,T}(s_0, \mathcal{B}_{1,T}(s_0))\left\{\hat{\mathcal{B}}_{1,T}(s_0) - \mathcal{B}_{1,T}(s_0)\right\} = \nabla \mathcal{L}_{1,T,s_0}(s_0, \mathcal{B}) + \nabla \mathscr{B}_{t,1}(s_0, \mathcal{B}),$$

while $\nabla^2 \mathcal{L}_{1,T}(s_0, \mathcal{B}_{1,T}(s_0)) = \frac{2}{mT'}\sum_{t=p+1}^{T}\sum_{u=1}^{m} W_b(s_0 - s_u)\, \underline{\Theta}_{t-1}(s_u)\underline{\Theta}_{t-1}(s_u)^T$.
By the ergodic theorem, we get

$$\nabla^2 \mathcal{L}_{1,T}(s_0, \mathcal{B}_{1,T}(s_0)) \to \Upsilon(s_0).$$

Then

$$\left\{\hat{\mathcal{B}}_{1,T}(s_0) - \mathcal{B}_{1,T}(s_0)\right\} = -\frac{1}{2}\left\{\nabla \mathscr{B}_{t,1}(s_0, \mathcal{B}) + \nabla \mathcal{L}_{1,T,s_0}(s_0, \mathcal{B})\right\}\Upsilon(s_0)^{-1},$$

$$\left\{\hat{\mathcal{B}}_{1,T}(s_0) - \mathcal{B}_{1,T}(s_0)\right\} + \frac{1}{2}\nabla \mathscr{B}_{t,1}(s_0, \mathcal{B})\,\Upsilon(s_0)^{-1} = -\frac{1}{2}\nabla \mathcal{L}_{1,T,s_0}(s_0, \mathcal{B})\,\Upsilon(s_0)^{-1}.$$

As a result, the asymptotic normality can be deduced by applying the central limit theorem (Hall and Heyde (1980) [38]) on $\nabla \mathcal{L}_{1,T,s_0}(s_0, \mathcal{B})$, where

$$\nabla \mathcal{L}_{1,T,s_0}(s_0, \mathcal{B}) = -\frac{2}{mT'}\sum_{t=p+1}^{T}\sum_{u=1}^{m} W_b(s_0 - s_u)\, \underline{\Theta}_{t-1}(s_u)\left(X_{t,s_0}^2(s_u) - \mathcal{B}\underline{\Theta}_{t-1}(s_u)\right). \blacksquare$$

Proof of Theorem 4.3.2.
To study $\hat{\mathcal{B}}_{2,T}(s_0)$, we define the following quantities

$$\hat{J}_{2,T,s_0}(\underline{s}) = \frac{1}{mT'} \sum_{u=1}^{m} W_b\left(s_0 - s_u\right) S(u)\ S'(u) \otimes \sum_{t=p+1}^{T} \underline{\Theta}_{t-1,s_0}(s_u)\ \underline{\theta}^{(2)'}_{t-1,s_0}(s_u)$$

$$\hat{\underline{j}}_{2,T,s_0}(\underline{s}) = \frac{1}{mT'} \sum_{u=1}^{m} W_b\left(s_0 - s_u\right) S(u) \otimes \sum_{t=p+1}^{T} X^2_{t,s_0}(s_u) \underline{\Theta}_{t-1,s_0}(s_u) \qquad (48)$$

with $S(u) = (1, s_u - s_0)'$,
and $\tilde{\mathcal{L}}_{1,T,s_0}\left(s_0, \mathcal{B}\right) =$

$$\frac{1}{mT'} \sum_{t=p+1}^{T} \sum_{u=1}^{m} W_b\left(s_0 - s_u\right) \left\{ \tilde{X}^2_{t,s_0}(s_u) - \left(\omega_0(s_u) + \sum_{j=1}^{p} \alpha_j(s_u) \tilde{X}^2_{t-j,s_0}(s_u) + \sum_{j=1}^{p} \beta_j(s_u) \tilde{\sigma}^2_{t-j,s_0}(s_u)\right) \right\}^2$$

By the ergodic theorem, we obtain

$$\begin{aligned}
\nabla \tilde{\mathcal{L}}_{1,T,s_0}\left(s_0, \mathcal{B}\right) &= \hat{\underline{j}}_{2,T,s_0}(\underline{s}) - \hat{J}_{2,T,s_0}(\underline{s}) \mathcal{B}(\underline{s}) \\
&= -\frac{2}{mT'} \sum_{t=p+1}^{T} \sum_{u=1}^{m} W_b\left(s_0 - s_u\right) \left(\epsilon^2_t(s_u) - 1\right) \tilde{\sigma}^2_{t,s_0}(s_u) S(u) \otimes \underline{\Theta}_{t-1,s_0}(s_u). \\
\nabla \tilde{\mathcal{L}}_{1,T,s_0}\left(s_0, \mathcal{B}\right) &\xrightarrow{\mathcal{P}} 0, \quad \text{as } T' \to \infty. \qquad (49)
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\hat{J}_{2,T}(\underline{s})^{-1} \hat{\underline{j}}_{2,T}(\underline{s}) - \hat{J}_{2,T,s_0}(\underline{s})^{-1} \hat{\underline{j}}_{2,T,s_0}(\underline{s}) =& \hat{J}_{2,T,}(\underline{s})^{-1} \left\{ \left[\hat{J}_{2,T}(\underline{s}) - \hat{J}_{2,T,s_0}(\underline{s}) \right] \mathcal{B}\left(s_0\right) + \left[\hat{\underline{j}}_{2,T}(\underline{s}) - \hat{\underline{j}}_{2,T,s_0}(\underline{s}) \right] \right\} \\
&+ \hat{J}_{2,T}(\underline{s})^{-1} \left\{ \hat{J}_{2,T}(\underline{s}) - \hat{J}_{2,T,s_0}(\underline{s}) \right\} \left\{ \hat{J}_{2,T,s_0}(\underline{s})^{-1} \hat{\underline{j}}_{2,T}(\underline{s}) - \mathcal{B}\left(s_0\right) \right\}
\end{aligned}$$

Next, and by considering the estimator $\hat{\mathcal{B}}_{2,T}$ we have

$$\begin{aligned}
\hat{\mathcal{B}}_{2,T} - \mathcal{B}_2\left(s_0\right) =& \Lambda_2(\underline{s})^{-1} \mathscr{B}_{2,T}\left(s_0\right) + \left\{ \hat{J}_{2,T}(\underline{s})^{-1} - \Lambda_2(\underline{s})^{-1} \right\} \mathscr{B}_{2,T}\left(s_0\right), \\
&+ \hat{J}_{2,T}(\underline{s})^{-1} \left\{ \hat{J}_{2,T}(\underline{s}) - \hat{J}_{2,T,s_0}(\underline{s}) \right\} \hat{J}_{2,T,s_0}(\underline{s})^{-1} \left\{ \hat{\underline{j}}_{2,T,s_0}(\underline{s}) - \hat{J}_{2,T,s_0}(\underline{s}) \mathcal{B}_2\left(s_0\right) \right\} \\
&+ \hat{J}_{2,T,s_0}(\underline{s})^{-1} \left\{ \hat{\underline{j}}_{2,T,s_0}(\underline{s}) - \hat{J}_{2,T,s_0}(\underline{s}) \mathcal{B}_2(s_0) \right\}
\end{aligned}$$

where

$$\mathcal{B}_2(s_0)' = (\omega_0(s_0), \dot{\omega}_0(s_0), \underline{\alpha}(s_0), \dot{\alpha}_1(s_0), \ldots, \dot{\alpha}_p(s_0), \underline{\beta}(s_0), \dot{\beta}_1(s_0), \ldots, \dot{\beta}_p(s_0))$$

and

$$\mathscr{B}_{2,T}\left(s_0\right)=\left\{\hat{\underline{j}}_{2,T}(\underline{s})-\hat{\underline{j}}_{2,T,s_0}(\underline{s})\right\}+\left\{\hat{J}_{2,T,s_0}(\underline{s})-\hat{J}_{2,T,s_0}(\underline{s})\right\}\mathcal{B}_2(s_0).$$

Now, and by using equation (21) and $\hat{J}_{2,T}(\underline{s})\rightarrow\Lambda_2(\underline{s})$ as $T'\rightarrow\infty$ we have

$$\left\{\hat{J}_{2,T}(\underline{s})^{-1}-\Lambda_2(\underline{s})^{-1}\right\}\mathscr{B}_{2,T}\left(s_0\right)\xrightarrow{\mathcal{P}}0,$$

$$\hat{J}_{2,T}(\underline{s})^{-1}\left\{\hat{J}_{2,T}(\underline{s})-\hat{J}_{2,T,s_0}(\underline{s})\right\}\hat{J}_{2,T,s_0}(\underline{s})^{-1}\left\{\hat{\underline{j}}_{2,T,s_0}(\underline{s})-\hat{J}_{2,T,s_0}(\underline{s})\mathcal{B}_2\left(s_0\right)\right\}\xrightarrow{\mathcal{P}}0,$$

$$\hat{J}_{2,T,s_0}(\underline{s})^{-1}\left\{\hat{\underline{j}}_{2,T,s_0}(\underline{s})-\hat{J}_{2,T,s_0}(\underline{s})\mathcal{B}_2(s_0)\right\}\xrightarrow{\mathcal{P}}0.$$

Next, we use Theorem 3.5.3 and Corollary S9. to bound $\mathscr{B}_{2,T}\left(s_0\right)$ than by applying the ergodic theorem we obtain

$$\|\mathscr{B}_{2,T}\left(s_0\right)\|_1\leq C\frac{1}{m}\sum_{u=1}^{m}W_b\left(s_0-s_u\right)\|s_u-s_0\|_\infty^2$$

when $T'\rightarrow\infty$ ■

.

Proof of Theorem 4.3.3

We have

$$\hat{\mathcal{B}}_{1,T}(s_0)=\hat{J}_{1,T}^{-1}(\underline{s})\,\hat{\underline{j}}_{1,T}(\underline{s})$$

where

$$\hat{J}_{1,T}(\underline{s})=\frac{1}{mT'}\sum_{u=1}^{m}W_b\left(s_0-s_u\right)\otimes\sum_{t=p+1}^{T}\underline{\Theta}_{t-1}(s_u)\,\underline{\theta}_{t-1}^{(2)'}(s_u),$$

$$and\ \hat{\underline{j}}_{1,T}(\underline{s})=\frac{1}{mT'}\sum_{u=1}^{m}W_b\left(s_0-s_u\right)\otimes\sum_{t=p+1}^{T}X_t^2(s_u)\underline{\Theta}_{t-1}(s_u).$$

By using the ergodic theorem and Slutsky theorem we obtain

$$\hat{J}_{1,T}(\underline{s})\xrightarrow{\mathcal{P}}\mathbb{E}\left(\underline{\Theta}_0(s_0)\,\underline{\theta}_0^{(2)'}(s_0)\right),$$

$$\hat{\underline{j}}_{1,T}(\underline{s})\xrightarrow{\mathcal{P}}\mathbb{E}\left(X_t^2(s_0)\underline{\Theta}_{t-1}(s_0)\right).$$

Now we need to prove that

$$\mathcal{B}(s_0)=\mathbb{E}\left(\underline{\Theta}_0(s_0)\,\underline{\theta}_0^{(2)'}(s_0)\right)^{-1}\mathbb{E}\left(X_t^2(s_0)\underline{\Theta}_{t-1}(s_0)\right).$$

By using equation (2) we have

$$X_t^2(s_0) = \mathcal{B}(s_0)\underline{\Theta}_{t-1}(s_0) + \sigma_t^2(s_0)(\epsilon_t^2(s_0) - 1),$$

then

$$X_t^2(s_0)\underline{\theta}_0^{(2)'}(s_0) = \mathcal{B}(s_0)\underline{\Theta}_{t-1}(s_0)\underline{\theta}_0^{(2)'}(s_0) + \sigma_t^2(s_0)(\epsilon_t^2(s_0) - 1)\underline{\theta}_0^{(2)'}(s_0),$$

by taking expectations on both sides, we find the desired result. ■